

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

Amazon.com, Inc. and Amazon.com Services LLC.,

Petitioner

v.

LightGuide Inc.,

Patent Owner.

IPR2023-01423

U.S. Patent No. 10,528,036

DECLARATION OF JAMES L. MULLINS, Ph.D.

TABLE OF CONTENTS (cont.)

	<u>Page</u>
I. INTRODUCTION	1
II. QUALIFICATIONS	2
III. LEGAL STANDARDS	4
A. Authentication.....	5
B. Public Accessibility	5
C. The Critical Date.....	6
D. The Person of Ordinary Skill in the Art.....	6
IV. BACKGROUND OF LIBRARY CATALOG RECORDS AND OTHER RESOURCES	7
A. Monograph Publications	14
B. Periodical Publications.....	15
C. Cataloging / Indexing.....	15
V. OPINIONS REGARDING AUTHENTICITY AND PUBLIC ACCESSIBILITY OF REFERENCES.....	19
A. Exhibit 1005: Rick Kjeldsen, Claudio Pinhanez, Gopal Pingali, Jacob Hartman, Tony Levas and Mark Podlaseck. “Interacting with Steerable	

TABLE OF CONTENTS (cont.)

	<u>Page</u>
Projected Displays.” Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition, Washington, DC, USA, May 20-21, 2002, pp. 402-407	19
1. Authentication.....	20
2. Public Accessibility	22
B. Exhibit 1006: Claudio Pinhanez, Rick Kjeldsen, Anthony Levas, Gopal Pingali, Mark Podlaseck, and Paul Chou. “Ubiquitous Interactive Graphics.” IBM Research Report RC22495 (W0205-143) (2002). (“Pinhanez-Ubiquitous”).....	26
1. Authenticity.....	26
2. Public Accessibility	28
C. Exhibit 1014: Claudio Pinhanez, et al., “Applications of Steerable Projector-Camera Systems.” Proceedings of the IEEE International Workshop on Projector-Camera Systems at ICCV.....	32
1. Authentication.....	33
2. Public Accessibility	34
D. Exhibit 1015: Claudio Pinhanez and Gopal Pingali, “Projector-Camera Systems for Telepresence.” IBM Research Report RC23308 (W0408-	

TABLE OF CONTENTS (cont.)

	<u>Page</u>
130) (2004); also in Proceedings of the 2004 ACM SIGMM Workshop on Effective Telepresence, October 15, 2004, New York, NY	38
1. Authentication.....	39
2. Public Accessibility	41
E. Exhibit 1016: Tamotsu Machino, Yoshito Nanjo, Yoshimasa Yanagihara, Hiroaki Kawata, Satoshi Iwaki, and Ken-ichiro Shimokura. “Robot-Augmented Communication: A Remote- collaboration System based on a Shared Field of View in Real Space.” In Proceedings 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, Edmonton, AB, Canada, August 2-6, 2005, pp. 3467-3473	49
1. Authentication.....	50
2. Public Accessibility	52
F. Exhibit 1017: Tamotsu Machino, Satoshi Iwaki, Hiroaki Kawata, Yoshimasa Yanagihara, Yoshito Nanjo, and Ken-ichiro Shimokura. “Remote-Collaboration System Using Mobile Robot with Camera and Projector.” In Proceedings 2006 IEEE International Conference on Robotics and Automation, Orlando, Florida, May 15-18, 2006. ICRA 2006 on pages 4063-4068, published by IEEE Robotics and	

TABLE OF CONTENTS (cont.)

	<u>Page</u>
Automation Society, 2006. (“Machino-Remote-Collaboration”).....	57
1. Authentication.....	58
2. Public Accessibility	60
G. Exhibit 1018: Tetsuo Iyoda, Tsutomu Abe, Kiwame Tokai, Shoji Sakamoto, Jun Shingu, Hiroko Onuki, Meng Shi, and Shingo Uchihashi. “LightCollabo: Distant Collaboration Support System for Manufacturers.” In Advances in Multimedia Modeling: 14th International Multimedia Modeling Conference, MMM 2008, Kyoto, Japan, January 9-11, 2008, Proceedings, vol. 4903, p. 369. Springer Science & Business Media, 2007 (“Iyoda”)	67
1. Authentication.....	68
2. Public Accessibility	70
VI. RIGHT TO SUPPLEMENT	75

TABLE OF APPENDIXES

Appendix No.	Description
Appendix A-1	Scan of Kjeldsen et al., <i>Interacting with Steerable Projected Displays</i> (“Kjeldsen”) from Linda Hall Library, Kansas City, MO
Appendix A-2	Kjeldsen downloaded from IEEE Xplore
Appendix A-3	WorldCat entry for Kjeldsen
Appendix A-4	Linda Hall Library OPAC record of Kjeldsen
Appendix A-5	Kim et al., <i>A Development of Interactive Game “Ting Ting” Using Real and Virtual Objects</i> (“Kim”)
Appendix B-1	Pinhanez et al., <i>Ubiquitous Interactive Graphics</i> (“Pinhanez-Ubiquitous”) downloaded from CiteSeer
Appendix B-2	Bibliography of Claudio Pinhanez (https://researcher.watson.ibm.com/researcher/view_person_publications.php?person=br-csantosp&t=1&mhsrc=ibmsearch_a&mhq=RC22495)
Appendix B-3	Personal website of Claudio Pinhanez (http://www.pinhanez.com/claudio/)
Appendix B-4	Pinhanez-Ubiquitous downloaded from Pinhanez’s personal website (http://www.pinhanez.com/claudio/publications/rc22495.pdf)
Appendix B-5	Google Scholar entries citing Pinhanez-Ubiquitous
Appendix B-6	Flagg et al., <i>Projector-Guided Painting</i> (“Flagg”)
Appendix B-7	Bhasker et al., <i>Registration Techniques for Using Imperfect and Partially Calibrated Devices in Planar Multi-Projector Displays</i> (“Bhasker”)
Appendix B-8	Borkowski et al., <i>User-Centric Design of a Vision System for Interactive Applications</i> (“Borkowski”)
Appendix C-1	Pinhanez et al., <i>Applications of Steerable Projector-Camera Systems</i> (“Pinhanez-Applications”) downloaded from CiteSeer

Appendix C-2	Google Scholar record of Claudio Pinhanez
Appendix C-3	Google Scholar entries citing Pinhanez-Applications
Appendix C-4	Butz, et al., <i>Searchlight - A Lightweight Search Function for Pervasive Environments</i> (“Butz”)
Appendix C-5	Nakada, et al., <i>A Supporting System for Finding Lost Objects by Spotlight</i> (“Nakada”)
Appendix C-6	Kuren et al., <i>An Intelligent Disassembly Assistant for Man-Machine Demanufacturing</i> (“Kuren”)
Appendix D-1	Scan of Pinhanez et al., <i>Projector-camera Systems for Telepresence</i> (“Pinhanez-Projector-Camera”) from Library of Congress
Appendix D-2	Pinhanez-Projector-Camera downloaded from ACM Digital Library
Appendix D-3	WorldCat entry for Pinhanez-Projector-Camera
Appendix D-4	Library of Congress OPAC record of Pinhanez-Projector-Camera
Appendix D-5	Library of Congress OPAC MARC tags of Pinhanez-Projector-Camera
Appendix D-6	<i>Google Scholar</i> entries citing Pinhanez-Projector-Camera
Appendix D-7	Schmitt, et al., <i>Towards Ambient Business: Enabling Open Innovation in a World of Ubiquitous Computing</i> (“Schmitt”)
Appendix D-8	Alex Olwal, <i>Unobtrusive Augmentation of Physical Environments: Interaction Techniques, Spatial Displays, and Ubiquitous Sensing</i> (“Olwal”)
Appendix D-9	Karsten Weber, <i>Mobile Devices and a New Understanding of Presence</i> (“Weber”)
Appendix E-1	Scan of Machino et al., <i>Robot-Augmented Communication: A Remote-collaboration System based on a Shared Field of View</i>

	<i>in Real Space</i> (“Machino-Robot-Augmented”) from Library of Congress
Appendix E-2	Machino-Robot-Augmented downloaded from IEEE Xplore
Appendix E-3	<i>WorldCat</i> entry for Machino-Robot-Augmented
Appendix E-4	Library of Congress OPAC record of Machino-Robot-Augmented
Appendix E-5	Library of Congress OPAC MARC tags of Machino-Robot-Augmented
Appendix E-6	Google Scholar entries citing Machino-Robot-Augmented
Appendix E-7	Matsumaru et al., <i>Step-on Interface on Mobile Robot to Operate by Stepping on Projected Button</i> (“Matsumaru-1”)
Appendix E-8	Matsumaru et al., <i>Functions of Mobile-Robot Step-On Interface</i> (“Matsumaru-2”)
Appendix F-1	Scan of Machino et al., <i>Remote-Collaboration System Using Mobile Robot with Camera and Projector</i> (“Machino-Remote-Collaboration”) from Western University Libraries, London, Ontario, Canada
Appendix F-2	Machino-Remote-Collaboration downloaded from IEEE Xplore
Appendix F-3	<i>WorldCat</i> entry for Machino-Remote-Collaboration
Appendix F-4	Western University OPAC record of Machino-Remote-Collaboration
Appendix F-5	IEEE Xplore record of Machino-Remote-Collaboration
Appendix F-6	<i>Google Scholar</i> entries citing Machino-Remote-Collaboration
Appendix F-7	Park et al., <i>Robots with Projectors: An Alternative to Anthropomorphic HRI</i> (“Park”)

Appendix F-8	Kawata, et al., <i>Field Trial of Asynchronous Communication Using Network-Based Interactive Child Watch System for the Participation of Parents in Day-Care Activities</i> (“Kawata”)
Appendix F-9	Niitsuma et al., <i>Spatial Memory for Collaboration Support System in Intelligent Space-User Identification Based on Spatial Memory</i> (“Niitsuma”)
Appendix G-1	Scan of Iyoda et al., <i>LightCollabo: Distant Collaboration Support System for Manufacturers</i> (“Iyoda”) from Cornell University Libraries
Appendix G-2	Iyoda downloaded from Springer Link
Appendix G-3	<i>WorldCat</i> entry for Iyoda
Appendix G-4	Cornell University Libraries OPAC record of Iyoda
Appendix G-5	Cornell University Libraries OPAC Marc tags of Iyoda
Appendix H	<i>Curriculum Vitae</i> of James L. Mullins., Ph.D.

I. INTRODUCTION

I, James L. Mullins, am above the age of majority and hereby declare under penalty of perjury:

1. I have been retained by counsel for Petitioners Amazon.com, Inc. and Amazon.com Services LLC to analyze and opine on the authenticity and public accessibility of certain documents referenced herein.

2. I am presently Dean Emeritus of Libraries and Esther Ellis Norton Professor Emeritus at Purdue University. My career as a professional and academic/researcher spanned more than 44 years, including library positions at Indiana University, Villanova University, Massachusetts Institute of Technology, and Purdue University. Appendix H is a true and correct copy of my curriculum vitae describing my background and experience.

3. In 2018, I founded the firm Prior Art Documentation Librarian Services, LLC, now located at 205 Saint Cuthbert, Williamsburg, VA 23188. Further information about my firm, Prior Art Documentation Librarian Services, LLC (PADLS), is available at www.priorartdoclib.com.

4. For my services in connection with this case, I bill at my usual and customary hourly fee of \$275.00. I have no stake in the outcome of this proceeding

or any related litigation or administrative proceedings, and my compensation in no way depends on the substance of my testimony or the outcome of this proceeding.

5. As part of my preparation for completing this report, I have reviewed and considered the materials identified in this report.

II. QUALIFICATIONS

6. I received a Bachelor of Arts degree in History, Religion and Political Science in 1972 as well as a Master of Arts degree in Library Science in 1973, both from the University of Iowa. I received my Ph.D. in Academic Library Management in 1984 from Indiana University. Over forty-four years, I have held various positions, and I am considered a leader in the field of library and information sciences.

7. I am currently Dean Emeritus of Libraries and Esther Ellis Norton Professor Emeritus at Purdue University, 2018 – present. I was previously employed as follows:

i) Dean of Libraries and Esther Ellis Norton Professor, Purdue University, West Lafayette, IN, 2004 – 2017;

ii) Assistant/Associate Director for Administration, Massachusetts Institute of Technology (MIT) Libraries, Cambridge, MA 2000 – 2004;

iii) University Librarian and Director, Falvey Memorial Library, Villanova University, Villanova, PA, 1996 – 2000;

iv) Director of Library Services, Indiana University South Bend, South Bend, IN, 1978 – 1996;

v) Part-time Instructor, School of Library and Information Science, Indiana University, Bloomington, IN, 1979 – 1996;

vi) Associate Law Librarian, and associated titles, Indiana University School of Law, Bloomington, IN, 1974 – 1978;

vii) Catalog Librarian, Assistant Professor, Georgia Southern College (now University), Statesboro, GA, 1973 – 1974.

8. I am a member of the American Library Association (“ALA”), where I served as the chair of the Research Committee of the Association of College and Research Libraries (“ACRL”). My service to ALA included service on the editorial board of the most prominent library journal, *College and Research Libraries*. I also served on the Standards Committee, College Section of the ACRL, where I was instrumental in developing a re-issue of the *Standards for College Libraries* in 2000.

9. I am an author of numerous publications in the field of library science and have given presentations in library sciences at national and international conferences. During more than 44 years as an academic librarian and library science scholar, I have gained extensive experience with techniques and systems for cataloging and indexing printed materials in the library setting, including catalog records and online library management systems built using Machine-Readable

Cataloging (“MARC”) standards. As an academic library administrator, I have had responsibility to ensure that students were educated to identify, locate, assess, and integrate information garnered from research library resources. I have also facilitated the research of faculty colleagues either directly or through the provision of and access to the requisite print and/or digital materials and services at the universities where I worked.

10. Based on my experience identified above and detailed in my curriculum vitae, which is attached hereto as Appendix H, I consider myself to be an expert in the field of library science and academic library administration. I have previously offered my opinions on the public accessibility and authenticity of documents in over 118 cases. I have been deposed in one case.

III. LEGAL STANDARDS

11. I am not a lawyer. My understanding of the legal standards to apply in reaching the conclusions in this declaration is based on discussions with counsel for Petitioner, my experience applying similar standards in other patent-related matters, and my reading of the documents related to this proceeding. In preparing this declaration, I have tried to faithfully apply these legal standards to the authenticity of the documents referenced herein and whether, when, and how these documents were publicly accessible.

A. Authentication

12. I have been informed by counsel and understand that an item is considered authentic if there is sufficient evidence to support a finding that the item is what it is claimed to be by the proponent of the item. I have also been informed that authenticity can be established based on the properties of the item itself, such as the appearance, content, substance, internal patterns, or other distinctive characteristics of the item.

B. Public Accessibility

13. I have been informed by counsel and understand that a given reference qualifies as publicly accessible if it were disseminated or otherwise made available such that persons interested and ordinarily skilled in the relevant subject matter could locate it through the exercise of ordinary diligence.

14. I have been informed by counsel and understand that the determination of public accessibility under the foregoing standard rests on a case-by-case analysis of the facts particular to an individual publication. However, I also understand that one way that public accessibility can be shown is if a publication is cataloged and indexed by a library such that a person interested in the relevant subject matter could locate it. In other words, I understand that cataloging and indexing by a library, such that a person interested in the relevant subject matter could locate the reference, is sufficient to render it publicly accessible, though there are other ways that a given

reference may qualify as publicly accessible. One manner of sufficient indexing is indexing according to subject matter. I have been informed by counsel and understand that it is not necessary to prove someone actually looked at the reference in order to show it was publicly accessible by virtue of a library's cataloging and indexing thereof. I understand that cataloging and indexing of a particular printed publication can indicate that the publication is publicly accessible.

15. I also understand that there are other ways that a given reference may be deemed publicly accessible, including where that reference is cited by others in the field.

C. The Critical Date

16. I have been informed by counsel that for purposes of this declaration I should consider evidence of public accessibility prior to October 1, 2010. As described further below, I have concluded that each of the references discussed herein was publicly accessible, including to a Person of Ordinary Skill in the Art, prior to October 1, 2010.

D. The Person of Ordinary Skill in the Art

17. I have been informed by counsel that a person of ordinary skill in the art at the time of the inventions ("POSITA") is a hypothetical person who is presumed to be familiar with the relevant field and its literature at the time of the inventions.

This hypothetical person is also a person of ordinary creativity, capable of understanding the scientific principles applicable to the pertinent field.

18. In addition, I have been informed by counsel that as relevant to these proceedings, the POSITA would have the skills of one or more individuals with experience in the fields of interactive computer applications. These individuals would have at least a four-year degree in electrical engineering, mechanical engineering, computer science, or a closely related field, and at least a year of experience working in interactive computer applications with integrated cameras. Lack of work experience can be remedied by additional education, and vice versa.

IV. BACKGROUND OF LIBRARY CATALOG RECORDS AND OTHER RESOURCES

19. Some background on MARC formatted records, the Online Computer Library Center (“OCLC”), and WorldCat is helpful to understand the library catalog records discussed in this report. I am fully familiar with the library cataloging standard known as the MARC standard, which is an industry-wide standard method of storing and organizing library catalog information.¹ MARC practices have been consistent since the MARC format was developed by the Library of Congress in the

¹ The full text of the standard is available from the Library of Congress at <http://www.loc.gov/marc/bibliographic/>.

1960s, and by the early 1970s became the U.S. national standard for disseminating bibliographic data. By the mid-1970s, the MARC format became the international standard, and it persists as such through the present. A MARC-compatible library is one that has a catalog consisting of individual MARC records for each of its items. The underlying MARC format underpins the online public access catalog (“OPAC”) that is available to library users to locate a particular holding of a library. Today, MARC is the primary communications protocol for the transfer and storage of bibliographic metadata in libraries.² The MARC practices discussed below were in place during the time frames relevant to the documents referenced herein.

20. OCLC is a not-for-profit worldwide consortium of libraries. Similar to MARC standards, OCLC’s practices have been consistent since the 1970s through the present. Accordingly, the OCLC practices discussed below were in place during the time frame discussed in my opinions section. OCLC was created “to establish, maintain and operate a computerized library network and to promote the evolution of library use, of libraries themselves, and of librarianship, and to provide processes and

² Almost every major library in the world uses a catalog that is MARC-compatible. *See, e.g.,* Library of Congress, *MARC Standards Frequently Asked Questions (FAQ)*, available at <https://www.loc.gov/marc/faq.html> (last visited Jan. 24, 2018). “MARC is the acronym for MACHine-Readable Cataloging. It defines a data format that emerged from a Library of Congress-led initiative that began nearly forty years ago. It provides the mechanism by which computers exchange, use, and interpret bibliographic information, and its data elements make up the foundation of most library catalogs used today.” *Id.* MARC is the ANSI/NISO Z39.2-1994 (reaffirmed 2009) standard for Information Interchange Format.

products for the benefit of library users and libraries, including such objectives as increasing availability of library resources to individual library patrons and reducing the rate of rise of library per-unit costs, all for the fundamental public purpose of furthering ease of access to and use of the ever-expanding body of worldwide scientific, literary and educational knowledge and information.”³ Among other services, OCLC and its members are responsible for maintaining the WorldCat database (<http://www.worldcat.org/>), used by libraries throughout the world.

21. Libraries worldwide use the machine-readable MARC format for catalog records. MARC-formatted records include a variety of subject access points based on the content of the document being cataloged. A MARC record for a particular work comprises several fields, each of which contains specific data about the work. Each field is identified by a standardized, unique, three-digit code corresponding to the type of data that follows. For example, a work’s title is recorded in field 245, the primary author of the work is recorded in field 100, a work’s International Standard Book Number (“ISBN”) is recorded in field 020, and the work’s Library of Congress call number (assigned by the Library of Congress) is

³ OCLC Online Computer Library Center, Inc., Amended Articles of Incorporation of OCLC Online Computer Library Center, Inc., Third Article (OCLC, Dublin, Ohio) Revised June 23, 2017, available at <https://www.oclc.org/content/dam/oclc/membership/articles-of-incorporation.pdf>.

recorded in field 050. Some fields can contain subfields, which are indicated by letters. For example, a work's publication date is recorded in field 260 under the subfield "c."

22. The MARC Field 040, subfield "a," identifies the library or other entity that created the catalog record in the MARC format. MARC Field 008 identifies the date when this first MARC record was created.

23. MARC records also include several fields that include subject matter classification information. An overview of MARC record fields is available through the Library of Congress at <http://www.loc.gov/marc/bibliographic/>. For example, 6XX fields are termed "Subject Access Fields."⁴ Among these, for example, is the 650 field; this is the "Subject Added Entry - Topical Term (R)" field. See <http://www.loc.gov/marc/bibliographic/bd650.html>. The 650 field is a "[s]ubject added entry in which the entry element is a topical term." *Id.* The 650 field entries "are assigned to a bibliographic record to provide access according to generally accepted thesaurus-building rules (e.g., *Library of Congress Subject Headings* (LCSH), *Medical Subject Headings* (MeSH))." *Id.* Thus, a researcher can easily discover material relevant to a topic of interest with a search using the terms employed in MARC Fields 6XX.

⁴ See <http://www.loc.gov/marc/bibliographic/bd6xx.html>.

24. Further, MARC records include call numbers, which themselves include a classification number. For example, the 050 field is dedicated as the “Library of Congress Call Number (R)”⁵ as assigned by the Library of Congress. A defined portion of the Library of Congress Call Number is the classification number; the “source of the classification number is *Library of Congress Classification* and the *LC Classification-Additions and Changes*.” *Id.* (emphases in original). Thus, included in the 050 field is a subject matter classification. As an example: TK5105.59 indicates books on computer networks - security measures. When a local library assigns a classification number -- most often a Library of Congress derived classification number created by a local library cataloger, but sometimes a Dewey Decimal classification number such as, for example, “005.8, computer networks - security measures” -- that classification number appears in the 090 field. In either scenario, the MARC record includes a classification number in the call number field that represents a subject matter classification.

25. The 9XX fields, which are not part of the standard MARC 21 format⁶, were defined by OCLC for use by the Library of Congress in processing or holding notes for a local library, and for internal OCLC use. For example, the 955 field is

⁵ See <http://www.loc.gov/marc/bibliographic/bd050.html>.

⁶ See <https://www.oclc.org/bibformats/en/9xx.html>.

reserved for use by the Library of Congress to track the progress of a new acquisition from the time it is submitted for Cataloging in Publication (“CIP”) review until it is published and fully cataloged and publicly available for use within the Library of Congress. Fields 901-907, 910, and 945-949 have been defined by OCLC for local use and will pass OCLC validation. Fields 905 and 910 are often used by an individual library for internal processing purposes, for example, the date of cataloging and/or the initials of the cataloger.

26. *WorldCat* is the world’s largest public online catalog. It is maintained by the OCLC, a not-for-profit international library consortium, and built with the records created by the thousands of libraries that are members of OCLC. OCLC provides bibliographic and abstract information to the public based on MARC-compliant records through its OCLC *WorldCat* database. *WorldCat* requires no knowledge of MARC tags and codes and does not require a login or password. *WorldCat* is easily accessible through the World Wide Web to all who wish to search it; there are no restrictions to be a member of a particular community, etc. The date a given catalog record was created (corresponding to the MARC Field 008) appears in some detailed *WorldCat* records as the Date of Entry but not necessarily all. *WorldCat* does not provide a view of the underlying MARC format for a specific *WorldCat* record. In order to see the underlying MARC format, the researcher must locate the book in a holding library listed among those shown in *WorldCat*, and search the online

public catalog (OPAC) of a holding library. Whereas *WorldCat* records are widely available, the availability of library-specific MARC formatted records varies from library to library. When a specific library wishes to make the underlying MARC format available, there will be a link from the library's OPAC display, often identified as an option to click on the MARC record or use the librarian/staff view.

27. When a MARC record is created by the Library of Congress or an OCLC member institution, the date of creation for that record is automatically populated in the fixed field (008), with characters 00 through 05 in year, month, day format (YYMMDD).⁷ Therefore, the MARC record creation date reflects the date on which the publication associated with the record was first cataloged. Thereafter, the local library's computer system may automatically update the date in field 005 every time the library updates the MARC record (e.g., to reflect that an item has been moved to a different shelving location within the library, or a reload of the bibliographic data with the introduction of a new library management system that creates and manages the OPAC).

⁷ Some of the newer library catalog systems also include hour, minute, second (HHMMSS).

A. Monograph Publications

28. Monograph publications are written on a single topic, presented at length, and distinguished from an article. They include books, dissertations, proceedings of a conference, and technical reports. A library typically creates a catalog record when the monograph is acquired by the library. First, it will search the OCLC to determine if a record has already been created by the Library of Congress or another OCLC institution.

29. If a record is found in the OCLC, the record is downloaded into the library's management system, which typically includes the OPAC. Once the item is downloaded into the library's management system, the library adds its identifier to the OCLC database so when a search is completed on *WorldCat*, the library will be indicated as possessing the publication. Once a record is created in a library's management system, it is searchable and viewable through the library's OPAC, typically by author, title, and subject heading, at that library and from anywhere in the world through the internet by accessing that library's OPAC. The OPAC also connects with the circulation function of the library, which typically indicates whether the record is available, in circulation, etc., with its call number and location in a specific departmental/disciplinary library, if applicable. The OPAC not only provides immediate bibliographic access on-site, but it also facilitates the interlibrary loan process, which is when one publication is loaned from one library to another.

B. Periodical Publications

30. A periodical publication is a published work in a series, for example, a journal article. A library typically creates a catalog record for a periodical publication when the library receives its first issue. When the institution receives subsequent issues or volumes of the periodical, the issues or volumes are checked in, often using a date stamp; added to the institution's holding records; and made available very soon thereafter, normally within a few days of receipt or at most within a few weeks of receipt.

C. Cataloging / Indexing

31. A researcher may discover material relevant to a topic of interest in a variety of ways. One common means of discovery is to search for relevant information in an index of periodicals and other publications. Having found relevant material, the researcher will then normally obtain it online, look for it in libraries, or purchase it from a publisher, a bookstore, a document delivery service, or another provider.

32. Indexing services use a wide variety of controlled vocabularies to provide subject access and other means of discovering the content of documents. The formats in which these access terms are presented vary from service to service.

33. Online indexing services commonly provide bibliographic information, abstracts, and full-text copies of the indexed publications, along with a list of the

documents cited in the indexed publication. These services also often provide lists of publications that cite a given document. A citation of a document is evidence that the document was publicly available and in use by researchers no later than the publication date of the citing document. A few examples are provided below.

34. **ACM Digital Library** - Association for Computing Machinery. The ACM was founded in 1947 and is the world's largest scientific and educational computing society. The portal has two main sections: the ACM Digital Library and the ACM Guide to Computing Literature. The ACM Digital Library provides access to the journals, newsletters and conference proceedings of the organization. The Guide to Computing Literature is a bibliography of computing literature beginning in the 1950s with more than a million entries. See ACM Digital Library home page, available at <https://dl.acm.org/>.

35. **CiteSeer^x**. – CiteSeer is an evolving scientific literature digital library and search engine that has focused primarily on literature in computer and information science. CiteSeer was developed in 1997 at the NEC Research Institute (now known as NEC Laboratories America, Inc.) in Princeton, New Jersey. The service transitioned to Pennsylvania State University's College of Information Sciences and Technology in 2003. Since its inception, CiteSeer has grown to an index of over 750,000 documents and served over 1.5 million requests daily. See CiteSeer home page, available at <http://csxstatic.ist.psu.edu/home>.

36. **Google Scholar** - This web search engine indexes full text or metadata of scholarly literature, covering numerous formats and disciplines. The size of the database is not published by Google, but researchers have estimated that it contained approximately 160 million items in 2014. See Orduña-Malea, Enrique; Ayllón, Juan Manuel; Martín-Martín, Alberto; Delgado López-Cózar, Emilio, “About the size of Google Scholar: playing the numbers” Granada: EC3 Working Papers, 18:23 July 2014, available at <https://arxiv.org/ftp/arxiv/papers/1407/1407.6239.pdf>.

37. **IEEE Xplore** - The Institute of Electrical and Electronics Engineers is the world’s largest organization for the advancement of technology with some 430,000 members in 160 countries. Known by its acronym IEEE, it has created a database, IEEE Xplore, that provides access to its hundreds of publications and those of its publishing partners. This includes the content of over 170 journals, more than 1,400 conference proceedings, some 5,100 technical standards, 2,000 eBooks and 400 educational courses. In all, more than 3 million documents, dating from 1872 on, are searchable and available for purchase either through subscription or individually. Many of these records are accessible via Google Scholar. See IEEE Xplore home page, available at <https://ieeexplore.ieee.org/Xplore/home.jsp>.

38. **ResearchGate** – ResearchGate is the professional network for scientists and researchers. Over 15 million members from all over the world use it to share,

discover, and discuss research. See ResearchGate's "About" page, available at <https://www.researchgate.net/about>.

39. **ScienceDirect** – ScienceDirect is a website which provides subscription-based access to a large database of scientific and medical research. It hosts over 12 million pieces of content from 3,500 academic journals and 34,000 eBooks. The journals are grouped into four main sections: *Physical Sciences and Engineering*, *Life Sciences*, *Health Sciences*, and *Social Sciences and Humanities*. Article abstracts are freely available, but access to their full texts (in PDF and, for newer publications, also HTML) generally requires a subscription or pay-per-view purchase. See ScienceDirect webpage, available at <https://www.elsevier.com/solutions/sciencedirect>.

40. **SPIE Digital Library** - The SPIE Digital Library is the database of the International Society for Optical Engineering (originally the Society of Photographic Instrumentation Engineers). The technical papers, journals, eBooks, and conference proceedings of the organization are included in the database. More than 500,000 technical papers make up the database, with more than 18,000 new technical papers and 25 eBooks added each year. See SPIE Digital Library "About" page, available at <https://www.spiedigitallibrary.org>.

41. **Springer Link** - Springer Link provides researchers with access to millions of scientific documents from journals, books, series, protocols, reference

works and proceedings. See SpringerLink home page, available at <https://link.springer.com/>.

42. **Wisconsin TechSearch (“WTS”)** - WTS is a set of services offered by the University of Wisconsin Libraries. WTS offers an array of article delivery and research services to any individual or organization who requests the specialized skills of WTS staff in locating and retrieving information, regardless of whether the individual is affiliated with the University of Wisconsin. See WTS home page, available at <https://wts.wisc.edu/>.

V. OPINIONS REGARDING AUTHENTICITY AND PUBLIC ACCESSIBILITY OF REFERENCES

A. Exhibit 1005: Rick Kjeldsen, Claudio Pinhanez, Gopal Pingali, Jacob Hartman, Tony Levas and Mark Podlaseck. “Interacting with Steerable Projected Displays.” *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*, Washington, DC, USA, May 20-21, 2002, pp. 402-407. IEEE Computer Society, 2002. (“Kjeldsen”)

43. I have been asked to opine on the authenticity and public access of an article, Exhibit 1005, authored by Rick Kjeldsen, et al., titled “Interacting with Steerable Projected Displays,” published by the IEEE Computer Society in *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*, Washington, DC, USA, May 20-21, 2002, on pages 402-407 (“Kjeldsen”).

44. For consistency and clarification, Kjeldsen, when used in this declaration, refers not only to the Kjeldsen article but also the monograph in which it was published: *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*, Washington, DC, USA, May 20-21, 2002.

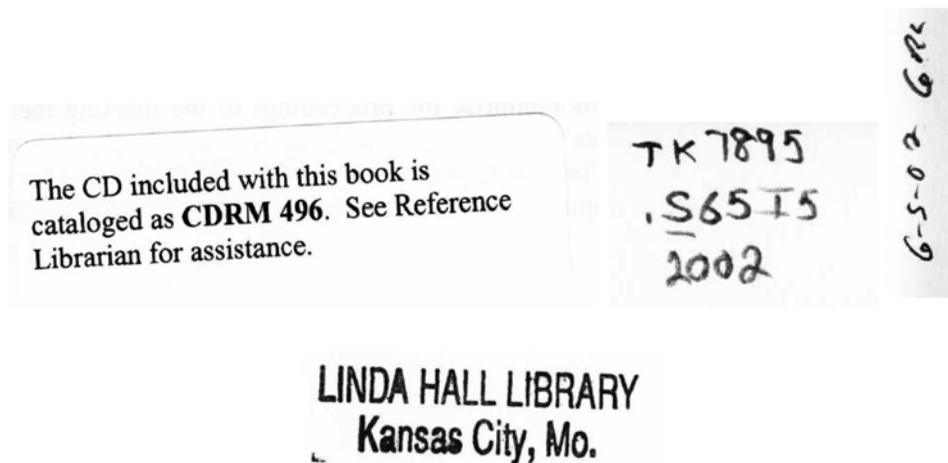
1. Authentication

45. I have evaluated the Kjeldsen reference in three ways: (1) by assessing Exhibit 1005, provided to me by counsel; (2) by requesting and assessing scans of Kjeldsen from the Linda Hall Library through its Document Delivery Services on June 8, 2023, appended to this declaration as Appendix A-1⁸; and (3) by retrieving and reviewing a download I made from *IEEE Xplore*, a proprietary database provided by IEEE, on June 8, 2023, through Purdue University Libraries.

46. Appendix A-1 is a scan of Kjeldsen received from the Linda Hall Library that also includes the title page of *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*, Washington, DC, USA, May 20-21, 2002. As shown below, the version of Kjeldsen attached as Appendix A-1 also includes (1) a copy of the proceedings title page including a copy of a sticker that says,

⁸ The Linda Hall Library, located in Kansas City, Missouri, is a pre-eminent research library in the areas of science, engineering and technology.

“See Reference Librarian for assistance.” (2) a handwritten call number on the title page that reads “TK7895.S65 I5 2002”; (3) the Table of Contents with a handwritten date written on the left side of the page that reads “6-5-02” (June 5, 2002); and (4) a stamp that reads “LINDA HALL LIBRARY, KANSAS CITY, MO.”



Appendix A-1 at 1-3. The table of contents shows that Kjeldsen was published starting on pp. 402 of these proceedings:

Interacting with Steerable Projected Displays	402
<i>R. Kjeldsen, C. Pinhanez, G. Pingali, J. Hartman, T. Levas, and M. Podlaseck</i>	

Id. at 6. The Kjeldsen article itself appears on pp. 7-12 of Exhibit A-1.

47. All identifying characteristics on Appendix A-1, such as the handwritten call number, date and stamp, are consistent with library practice and procedure that I have observed during my career as a professional librarian, specifically with those I have seen over the last five years with items held by the Linda Hall Library. I have no cause for concern about the authenticity or accuracy of these

identifying attributes. Appendix A-1 was found within the custody of a library, the Linda Hall Library, one of the most likely locations for an authentic publication.

48. Kjeldsen is available on-line through the *IEEE Xplore* database. I downloaded Appendix A-2 from *IEEE Xplore* through Purdue University Libraries by my emeritus faculty position at this URL: <https://ieeexplore-ieee-org.ezproxy.lib.purdue.edu/stamp/stamp.jsp?tp=&arnumber=1004187>. Kjeldsen is available for a fee at: <https://ieeexplore.ieee.org/abstract/document/1004187/citations?tabFilter=papers#citations>.

49. I have compared Appendix A-1 and Appendix A-2 and have found no difference between the two. Additionally, I have compared Appendix A-1 and Appendix A-2 with Exhibit 1005, provided to me by counsel, and have found no differences in content among the three. Given the above, I conclude that each of Exhibit 1005, Appendix A-1, and Appendix A-2 contains an authentic, true, and correct copy of Kjeldsen.

2. Public Accessibility

50. Appendix A-3 is a true and correct copy of the WorldCat entry for Kjeldsen, Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition. Kjeldsen was included among the papers published in

Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition on pages 402-407.

51. I obtained Appendix A-3 by completing a search on WorldCat on June 8, 2023, *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*.

52. The document indexed in this record is Kjeldsen as verified by title (*Fifth IEEE International Conference on Automatic Face Gesture Recognition, Proceedings*); publisher and date (IEEE in 2002); and ISBN (0-7695-1602-5).

53. When undertaking the search of *WorldCat* as described above using the geographical location of Missouri, the Linda Hall Library was 1st on the list among the 52 libraries that hold Kjeldsen (*Fifth IEEE International Conference on Automatic Face Gesture Recognition, Proceedings*) world-wide. Consistent with *WorldCat* standard practice, the entry for Kjeldsen would have been created in response to cataloging by a member library, no later than 2002 (as indicated below by the Linda Hall Library's cataloging of this document).

54. As a result, by 2010, the *WorldCat* record of Kjeldsen, *Fifth IEEE International Conference on Automatic Face Gesture Recognition, Proceedings*, could have been retrieved by searching title (*Fifth IEEE International Conference on Automatic Face Gesture Recognition, Proceedings*), and/or subject headings (e.g.,

Face perception; Gesture; Human computer interaction; Image processing digital techniques; Optical pattern recognition; among many others).

55. Appendix A-4 is a download I made from the Linda Hall Library OPAC (online catalog) on June 11, 2023. The document cataloged in this record is Kjeldsen as verified by title (*Fifth IEEE International Conference on Automatic Face Gesture Recognition, Proceedings*); and by ISBN (0-7695-1602-5). The “Creation date” for this catalog entry is indicated as 2002.

56. Thus, after the proceedings were catalogued in 2002, Kjeldsen could have been located by searching on the Linda Hall Library OPAC (online catalog) for the title (*Fifth IEEE International Conference on Automatic Face Gesture Recognition, Proceedings*), and/or subject headings (e.g., *Optical pattern recognition – Congresses; Face perception automation - Congresses; Gesture - Congresses; Image processing digital techniques - Congresses; Human computer interaction – Congresses*).

57. Evidence that the Linda Hall Library owns and received Kjeldsen (*Fifth IEEE International Conference on Automatic Face Gesture Recognition, Proceedings*) is detailed below.

58. On Appendix A-1 the ownership stamp on the first page of *Table of Contents*, reading “LINDA HALL LIBRARY,” and the handwritten date of receipt, “6-5-02” (June 5, 2002), are shown below.

LINDA HALL LIBRARY
Kansas City, Mo.

6-5-02 GR

59. Consistent with standard library policy and practice, Kjeldsen (*Fifth IEEE International Conference on Automatic Face Gesture Recognition, Proceedings*) would have been accessible on the shelf in the Linda Hall Library within a week to ten days after cataloging, labeling, and transfer to the shelf, that is, no later than June 15, 2002.

60. Evidence that Kjeldsen was publicly accessible is further confirmed by articles that cite Kjeldsen prior to October 1, 2010.

61. Appendix A-5 is an article by Si-Jung Kim, Moon-Sung Jang, Hong-Sic Kim and Tae-Yong Kuc, titled “A Development of Interactive Game ‘Ting Ting’ Using Real and Virtual Objects” published in the *2004 IEEE International Conference on Systems, Man and Cybernetics*, The Hague, Netherlands, October 10-13, 2004, in volume 2, on pages 1191-1197 (“Kim”).

62. Appendix A-5 was downloaded by me on June 11, 2023, enabled by my emeritus faculty position at Purdue University, through the Purdue University Libraries at this URL: <https://ieeexplore-ieee.org.ezproxy.lib.purdue.edu/stamp/stamp.jsp?tp=&arnumber=1399786>. Kim is available for a fee at this URL: <https://ieeexplore.ieee.org/abstract/document/1399786>.

63. On page 1197 of Kim is a citation to Kjeldsen, as shown below:

[10] Rick Kjeldsen, Claudio Pinhanez, Gopal Pingali, Jacob Hartman, Tony Levasand Mark Podlaseck, "Interacting with Steerable Projected Displays", Proc. Of 5th International Conference on Automatic Face and Gesture Recognition(FG'02). 2002.

64. The citation to Kjeldsen in Kim is evidence that Kjeldsen was known and accepted as a scholarly reference when Kim was presented as a paper at the *2004 IEEE International Conference on Systems, Man and Cybernetics* held in the Hague, Netherlands, between October 10-13, 2004.

65. Based on the above, I conclude that Kjeldsen, submitted in this proceeding as Exhibit 1005, is an authentic document that was publicly accessible before October 1, 2010.

B. Exhibit 1006: Claudio Pinhanez, Rick Kjeldsen, Anthony Levas, Gopal Pingali, Mark Podlaseck, and Paul Chou. "Ubiquitous Interactive Graphics." IBM Research Report RC22495 (W0205-143) (2002). ("Pinhanez-I")

66. I have been asked to opine on the authenticity and public accessibility of a document authored by Claudio Pinhanez, et al., titled "Ubiquitous Interactive Graphics" and published as *IBM Research Report RC22495 (W0205-143)* in 2002 ("Pinhanez-I").

1. Authenticity

67. I have evaluated the Pinhanez-I reference several ways: (1) by assessing Exhibit 1006, provided to me by counsel; (2) by retrieving and assessing a copy of Pinhanez-I which I downloaded from the *CiteSeer* database, stewarded by the Penn State University, College of Information Science and Technology, appended to this declaration as Appendix B-1; (3) by accessing and verifying a copy of Pinhanez-I from the list of *IBM Research Reports* available through an open access website, Appendix B-2; (4) by retrieving and reviewing a download I made on May 11, 2023 of Pinhanez-I from the personal website of Claudio Pinhanez, Appendix B-4; and (4) retrieving and assessing articles that cite Pinhanez-I prior to October 1, 2010, discussed in the Public Accessibility section of Part V.B, *infra*.

68. Appendix B-1 is a copy of Pinhanez-I I downloaded from the *CiteSeer* database, stewarded by the Penn State University, College of Information Science and Technology. *CiteSeer* is a trusted academic research repository and as such would only provide access to authentic research documents. Appendix B-1 was downloaded by me on May 11, 2023 from this URL:

<https://citeseerx.ist.psu.edu/pdf/391893bda0d70ab64e8b0ad90d375b80d9075970>

69. Appendix B-2 is a bibliography of the publications of Claudio Pinhanez published between 1995 and 2020. It is downloaded from IBM Research at this URL:

https://researcher.watson.ibm.com/researcher/view_person_pubs.php?person=br-

[csantosp&t=1&mhsrc=ibmsearch_a&mhq=RC22495](#). The following citation is shown by scrolling to year 2002:

Ubiquitous interactive graphics 
C. Pinhanez, R. Kjeldsen, T. Levas, G. Pingali, M. Podlaseck, P. Chou
IBM Research Report RC22495 (W0205-143), 2002

70. The hyperlink from the title in Appendix B-2, labeled “Ubiquitous interactive graphics,” does not link to the document, only to a “404” missing link. However, the fact that Pinhanez-I is linked from the IBM Research Report website provides evidence that it is a research publication issued under the auspices of IBM.

71. Appendix B-3 is the personal website of Claudio Pinhanez accessible at the URL <http://www.pinhanez.com/claudio/>. Appendix B-4 is a copy of Pinhanez-I I downloaded from Pinhanez’s personal website at this URL: <http://www.pinhanez.com/claudio/publications/rc22495.pdf>.

72. Having compared Appendix B-1 and Appendix B-4, I found no difference between the two. I then compared Appendix B-1 and Appendix B-4 with Exhibit 1006, provided to me by counsel, and found no difference among the three. Given the above, I conclude that each of Exhibit 1006, Appendix B-1, and Appendix B-4 contains an authentic, true, and correct copy of Pinhanez-I.

2. Public Accessibility

73. To determine the public accessibility of Pinhanez-I, I researched articles that cite Pinhanez-I prior to October 1, 2010.

74. Appendix B-5 is a true and correct download of the *Google Scholar* entries that cite Pinhanez-I and were published between 2004 and 2007. The citation list in Appendix B-5 consists of 10 documents. Of those 10 documents which cite Pinhanez-I, 3 documents are as follows: Appendix B-6, Flagg; Appendix B-7, Bhasker; and Appendix B-8, Borkowski. I completed the download of this list of entries from *Google Scholar* on May 7, 2023, at this URL:

https://scholar.google.com/scholar?hl=en&as_sdt=5%2C47&sciodt=0%2C47&cites=8433351548041826889&scipsc=&as_ylo=2004&as_yhi=2007.

75. Appendix B-6 is an article by Matthew Flagg and James M. Rehg titled “Projector-Guided Painting,” published in the *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology*, Montreux, Switzerland, October 15-18, 2006, on pages 235-243 (“Flagg”).

76. Flagg was downloaded by me from the *ACM Digital Library* on April 25, 2023, enabled by my emeritus faculty position at Purdue University, through the Purdue University Libraries at this URL: <https://dl-acm-org.ezproxy.lib.purdue.edu/doi/abs/10.1145/1166253.1166290>. Flagg is available for a fee at this URL: <https://dl.acm.org/doi/abs/10.1145/1166253.1166290>.

77. Appendix B-6, Flagg, at page 243, shows Pinhanez-I in the References section, which lists citations, as shown below:

17. C. Pinhanez, R. Kjeldsen, T. Levas, G. Pingali, M. Podlaseck, and P. Chou. Ubiquitous Interactive Graphics. Technical Report RC22495 (W0205-143), IBM Research Report, May 2002.

78. The citation to Pinhanez-Ubiquitous in Flagg is evidence that Pinhanez-I was known and accepted as a scholarly reference when Flagg was presented as a paper at *19th Annual ACM Symposium on User Interface Software and Technology* in Montreux, Switzerland on October 15-18, 2006 and was subsequently published in its *Proceedings*.

79. Appendix B-7 is an article by E. Bhasker, R. Juang and A. Majumder titled “Registration Techniques for Using Imperfect and Partially Calibrated Devices in Planar Multi-Projector Displays” published in *IEEE Transactions on Visualization and Computer Graphics*, volume 13, number 6, November – December 2007, on pages 1368-1375 (“Bhasker”).

80. Bhasker was downloaded by me from *IEEE Xplore* on April 25, 2023, enabled by my emeritus faculty position at Purdue University, through the Purdue University Libraries at this URL: <https://ieeexplore-ieee.org.ezproxy.lib.purdue.edu/stamp/stamp.jsp?tp=&arnumber=4376163&tag=1>.

Bhasker is available for a fee at this URL: <https://ieeexplore.ieee.org/abstract/document/4376163/references#references>.

81. Appendix B-7, Bhasker, at page 1375, shows Pinhanez-I as the 25th citation in its References, as shown below:

[25] C. S. Pinhanez, F. C. Kjeldsen, A. Levas, G. S. Pingali, M. E. Podlaseck, and P. B. Chou. Ubiquitous interactive graphics. Technical Report RC22495 (W0205-143), IBM Research, 2002.

82. The citation to Pinhanez-I in Bhasker is evidence that Pinhanez-I was known and accepted as a scholarly reference when Bhasker was published as a paper in *IEEE Transactions on Visualization and Computer Graphics*, volume 13, number 6, November – December 2007, on pages 1368-1375.

83. Appendix B-8 is an article by S. Borkowski, J. L. Crowley, J. Letessier and F. Berard, "User-Centric Design of a Vision System for Interactive Applications," *Fourth IEEE International Conference on Computer Vision Systems (ICVS'06)*, New York, NY, USA, January 4-7, 2006 on pages [9-9] ("Borkowski").

84. Borkowski was downloaded by me from *IEEE Xplore* on April 26, 2023, enabled by my emeritus faculty position at Purdue University, through the Purdue University Libraries at this URL: <https://ieeexplore-ieee.org.ezproxy.lib.purdue.edu/search/searchresult.jsp?searchWithin=%22Publication%20Number%22:10542&searchWithin=%22Document%20Title%22:User%20Centric%20Design%20of%20a%20Vision%20System%20for%20Interactive%20Applications>. Borkowski is available for a fee at this URL: <https://ieeexplore.ieee.org/abstract/document/1578697/references#references>.

85. Appendix B-8, Borkowski, at page [9], shows Pinhanez-I as the 9th citation in its References, as shown below:

[9] C. Pinhanez, R. Kjeldsen, T. Levas, G. Pingali, M. Podlaseck, and P. Chou. Ubiquitous interactive graphics. In *IBM Research Report RC22495 (W0205-143)*, May 2002.

86. The citation to Pinhanez-I in Borkowski is evidence that Pinhanez-I was known and accepted as a scholarly reference when Borkowski was presented as a paper at the *Fourth IEEE International Conference on Computer Vision Systems (ICVS'06)*, New York, NY, USA, January 4-7, 2006.

87. Based on the above, I conclude that Pinhanez-I, submitted in this proceeding as Exhibit 1006, is an authentic document that was publicly accessible before October 1, 2010.

C. Exhibit 1014: Claudio Pinhanez, et al., “Applications of Steerable Projector-Camera Systems.” *Proceedings of the IEEE International Workshop on Projector-Camera Systems at ICCV*. Nice, France. October 12, 2003 (“Pinhanez-II”).

88. I have been asked to opine on the authenticity and public accessibility of a paper authored by Claudio Pinhanez, et al., “Applications of Steerable Projector-Camera Systems” from the Proceedings of *the IEEE International Workshop on Projector-Camera Systems at International Conference on Computer Vision (ICCV)* held in Nice, France on October 12, 2003 (“Pinhanez-II”). The Proceedings were published by the ICCV 2003 committee in a CD-ROM distributed to all ICCV’03

conference, associated workshops, and courses participants,⁹ and online at this URL:
<http://lear.inrialpes.fr/people/triggs/events/iccv03/cdrom/index.php>.

1. Authentication

89. I have evaluated Pinhanez-II in three ways: (1) by assessing the version of this document attached as Exhibit 1014, provided to me by counsel; (2) by accessing and reviewing a download of Pinhanez-II that I obtained from the open access digital repository *CiteSeer* on April 25, 2023, appended to this declaration as Appendix C-1; and (3) by retrieving and reviewing articles that cite Pinhanez-II, published prior to October 1, 2010, discussed further below.

90. Exhibit 1014 is a copy of Pinhanez-II provided to me by counsel. I have been informed that it was downloaded from the ICCV 2003 website at this URL:
http://lear.inrialpes.fr/people/triggs/events/iccv03/cdrom/procams03/pinhanez_kjeldsen_levas_pingali_podlaseck_sukaviriya.pdf.

91. Appendix C-1 is a copy of Pinhanez-II that I downloaded from *CiteSeer*, a respected database stewarded by Penn State University, College of

⁹ See Information for Authors, ICCV'03,
<http://lear.inrialpes.fr/people/triggs/events/iccv03/author-info.php> (last visited May 25, 2023)
("We will also be producing a scratch CD-ROM containing all of the ICCV main conference papers, many of the workshop papers and course notes, and other supplementary materials. The CD-ROM will be distributed to all ICCV main conference, workshop and course participants, and most of it will also be made available on the web.")

Information Science and Technology, on April 25, 2023, at this URL:
<https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=6e4804c182b0ed9e2204b5e229efaea0ba96f2a8>.

92. In order to verify that Pinhanez-II was indeed an article by Claudio Pinhanez, I located his Google Scholar record. Pinhanez's *Google Scholar* record is attached as Appendix C-2 and can be accessed digitally at:
<https://scholar.google.com/citations?user=YPq3ax4AAAAJ&hl=en>.

93. Pinhanez-II is listed in Pinhanez's *Google Scholar* record as shown below and on the third page of Appendix C-2:

Applications of steerable projector-camera systems 41 2003
C Pinhanez, R Kjeldsen, A Levas, G Pingali, M Podlaseck, N Sukaviriya
Proceedings of the IEEE International Workshop on Projector-Camera Systems ...

94. I have compared Exhibit 1014, a copy of Pinhanez-I provided by counsel, to Appendix C-1, the digital version I retrieved from *CiteSeer*. I have concluded that they are the same, which supports my opinion that both Exhibit 1014 and Appendix C-1 are authentic, true, and correct copies of Pinhanez-Applications.

95. Additionally, based on my review of Appendix C-2, *Pinhanez's Google Scholar* record, I have no reason to question that this is an authentic document presented by Pinhanez at the *IEEE International Workshop on Projector-Camera Systems at ICCV* in Nice, France, on October 12, 2003.

96. My finding of authenticity is further supported based on finding Pinhanez-II in digital format on *CiteSeer*, which is where a publication such as Pinhanez-II is likely to be archived and accessible.

97. Therefore, for the reasons I have described above, I conclude that Exhibit 1014 contains an authentic, true, and correct copy of Pinhanez-II.

2. Public Accessibility

98. To determine the public accessibility of Pinhanez-II, I researched articles that cite Pinhanez-II prior to October 1, 2010. Appendix C-3 is a true and correct screen capture of the Google Scholar entries published between 2003 and 2005 that cite Pinhanez-II. The screen capture, which I conducted on May 15, 2023, can be accessed here:

https://scholar.google.com/scholar?hl=en&as_sdt=5%2C47&sciodt=0%2C47&cites=2989178860379338159&scipsc=&as_ylo=2003&as_yhi=2005.

99. Appendix C-3 shows nine citations to Pinhanez-II between 2003 and 2005. Two citations, discussed below, demonstrate the public accessibility of Pinhanez-I prior to October 1, 2010.

100. Appendix C-4 is an article by Andreas Butz, Michael Schneider, and Mira Spassova, titled: “Searchlight - A Lightweight Search Function for Pervasive Environments” published in *Pervasive*, in volume 3001, April 2004, on pages 351-356 (“Butz”). Butz is listed as citing Pinhanez-II in Appendix C-3.

101. I downloaded Butz on April 24, 2023 through the Purdue University Libraries at the following URL: https://link-springer-com.ezproxy.lib.purdue.edu/chapter/10.1007/978-3-540-24646-6_26. Butz may be downloaded for a fee at this link: https://link.springer.com/chapter/10.1007/978-3-540-24646-6_26.

102. Appendix C-4, which is a true and correct copy of Butz, lists Pinhanez-II as a reference, as shown below:

4. C. Pinhanez, R. Kjeldsen, A. Levas, G. Pingali, M. Podlaseck, and N. Sukaviriya. Applications of steerable projector-camera systems. In *Proceedings of the IEEE International Workshop on Projector-Camera Systems at ICCV 2003*, Nice Acropolis, Nice, France, October 12 2003. IEEE Computer Society Press.

103. Based on this, I conclude that Butz is evidence that Pinhanez-II was known and accepted as a scholarly reference when Butz was published in *Pervasive* in its April 2004 issue, which is prior to October 1, 2010.

104. Appendix C-5 is an article published in Japanese that appears on Appendix C-3. The citation in English is: Toyohisa Nakada, Hideaki Kanai, and Susumu Kunifuji, “A Supporting System for Finding Lost Objects by Spotlight,” IPSJ SIG Technical Report 2005/3/17, pages 25-30 by Information Processing Society of Japan, March 17, 2005 (“Nakada”).

105. I retrieved Nakada, appended as Appendix C-5, on May 15, 2023 from the following open access URL:

https://ipsj.ixsq.nii.ac.jp/ej/?action=repository_uri&item_id=42046&file_id=1&file_no=1.

106. Appendix C-5 is a true and correct copy of Nakada. Page 30 of Nakada includes Pinhanez-II in its list of references as shown below:

[7] C. Pinhanez, R. Kjeldsen, A. Levas, G. Pingali, M. Podlaseck, N. Sukaviriya, "Applications of steerable projector-camera systems", IEEE International Workshop on Projector-Camera Systems at ICCV 2003, 2003

107. Nakada's citation to Pinhanez-II provides further evidence that Pinhanez-II was known and accepted as a scholarly reference when Nakada was published as a Technical Report by the Information Processing Society of Japan on March 17, 2005.

108. Appendix C-6 is an article by M. B. Kuren and J. Soltani, "An Intelligent Disassembly Assistant for Man-Machine Demanufacturing," published in the *Proceedings, 2005 IEEE/ASME International Conference on Advanced Intelligent Mechatronics.*, Monterey, CA, USA, July 24-28, 2005 on pages 1109-1113 ("Kuren").

109. Kuren was downloaded by me on April 25, 2023, enabled by my emeritus faculty position at Purdue University, through the Purdue University Libraries at this URL:

<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1511158>. Kuren is available for a fee at this URL: <https://ieeexplore.ieee.org/abstract/document/1511158>.

110. Appendix C-6 is a true and correct copy of the Kuren paper, showing Pinhanez-II as the 3rd item in the list of references on page 1113, as shown below:

[3] C. Pinhanez, R. Kjeldsen, A. Levas, G. Pingali, M. Podlaseck, N. Sukaviriya, Applications of Steerable Projector-Camera Systems, IEEE International Workshop on Projector-Camera Systems (2003).

111. The Kuren article is evidence that Pinhanez-II was known and accepted as a scholarly reference when Kuren was presented as a paper at the *2005 IEEE/ASME International Conference on Advanced Intelligent Mechatronics.*, Monterey, CA, USA, July 24-28, 2005.

112. Based on the above, I conclude that Pinhanez-II, submitted in this proceeding as Exhibit 1014, is an authentic document that was publicly accessible before October 1, 2010.

D. Exhibit 1015: Claudio Pinhanez and Gopal Pingali, “Projector-Camera Systems for Telepresence.” IBM Research Report RC23308 (W0408-130) (2004); also in Proceedings of the 2004 ACM SIGMM Workshop on Effective Telepresence, October 15, 2004, New York, NY. Sponsored by SIGMM, SIGCOMM, & SIGGRAPH, 2004. pp. 63-66 (“Pinhanez-III”).

113. I have been asked to opine on an article by Claudio Pinhanez and Gopal Pingali, titled “Projector-Camera Systems for Telepresence,” published as *IBM*

Research Report RC23308 (W0408-130) in 2004 and in *Proceedings of the 2004 ACM SIGMM Workshop on Effective Telepresence* on pages 63-66 (“Pinhanez-III”). Pinhanez-III was presented as a paper at the 2004 ACM SIGMM Workshop on Effective Telepresence held on October 15, 2004, in New York, NY. The workshop was sponsored by SIGMM, SIGCOMM, & SIGGRAPH.

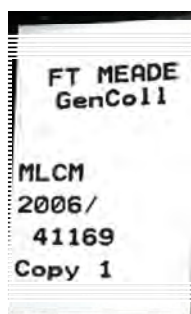
114. For consistency and clarification, Pinhanez-III, when used in this report, refers not only to the Pinhanez-III article but also to the publication in which it appears, *Proceedings of the 2004 ACM SIGMM Workshop on Effective Telepresence*.

1. Authentication

115. I have evaluated the Pinhanez-III reference four ways: (1) by assessing Exhibit 1015, provided to me by counsel; (2) by assessing scans of a copy of Pinhanez-III owned by the Library of Congress, Appendix D-1; (3) by retrieving and assessing a copy of Pinhanez-III I downloaded from the *ACM Digital Library*, Appendix D-2; and (4) by retrieving and assessing articles that cite Pinhanez-III prior to October 1, 2010 (discussed in the Public Accessibility section of Part V.D, *infra*).

116. Appendix D-1 is a scan of the copy of Pinhanez-III owned by the Library of Congress, provided to me at my request by the WTS on April 25, 2023. Appendix D-1 includes: front cover with a label and a shelving number that reads, “FT MEADE, Gen Coll, MLCM 2006/41169 Copy 1,” as shown below; title page;

verso of the title page (copyright); Foreword with a stamp that reads, “LIBRARY OF CONGRESS, NOV 10 2004, COPYRIGHT OFFICE,” as shown below; Table of Contents; and the Claudio Pinhanez article, titled: “Projector-camera Systems for Telepresence” on pages 63-66.



117. All identifying characteristics, such as stamps and notations, on Appendix D-1, are consistent with library practice and procedure that I have observed during my career as a professional librarian, specifically with those items held by the Library of Congress. I have no cause for concern about the authenticity or accuracy of these identifying attributes. In addition, Pinhanez-III was found within the custody of a library, the Library of Congress, one of the most likely locations for an authentic publication to be located.

118. Evidence of the ownership of Pinhanez-III by the Library of Congress will be further explored and detailed in Part V.D’s section on public accessibility, *infra*.

119. Appendix D-2 is a download I made of Pinhanez-III, by my emeritus faculty position at Purdue University, through the Purdue University Libraries on April 25, 2023, from the ACM Digital Library at this URL: <https://dl-acm-org.ezproxy.lib.purdue.edu/doi/pdf/10.1145/1026776.1026795>. Pinhanez-III is available for a fee at this URL: <https://dl.acm.org/doi/abs/10.1145/1026776.1026795>.

120. After comparing Appendix D-1 and Appendix D-2, I found no difference between the two. I then compared Appendix D-1 and Appendix D-2 with Exhibit 1015, provided to me by counsel, and have found no differences among the three. Given the above, I conclude that each of Exhibit 1015, Appendix D-1, and Appendix D-2 contains an authentic, true, and correct copy of Pinhanez-III.

2. Public Accessibility

121. Appendix D-3 is a true and correct copy of the *WorldCat* entry for Pinhanez-III. I obtained Appendix D-3 by completing a search on *WorldCat* on May 7, 2023 by searching District of Columbia (Library of Congress).

122. Appendix D-3 shows that Pinhanez-III is the document associated with this *WorldCat* entry, as verified by title (*Proceedings of the 2004 ACM SIGMM Workshop on Effective Telepresence*); publisher and publication date (ACM Press in 2004); and ISBN (1581139330). Consistent with *WorldCat* practice, this entry would have been created at or around the time of cataloging by the Library of Congress (April 12, 2005, as discussed below).

123. As a result, by 2010, Pinhanez-III could have been located by searching for the title (*Proceedings of the 2004 ACM SIGMM Workshop on Effective Telepresence*); or by searching the subject headings (e.g., *Computer Science – Conference papers and proceedings*; *Multimedia systems*; *Multimedia systems – Congresses*; *Video telephone*; and/or *Video telephone – Congresses*; among others).

124. When I searched *WorldCat* for holdings of Pinhanez-III in the District of Columbia, the Library of Congress was second on the list among the 304 shown as holding Pinhanez-III worldwide.

125. By 2010, the search discussed above could have been performed anywhere in the world by anyone who accessed *WorldCat* and its predecessor database through an OCLC member library.

126. Appendix D-4 is a download I made from the Library of Congress OPAC (online catalog) on April 24, 2023. The document cataloged in this record is Pinhanez-III as verified by the OPAC record fields including main title (*Proceedings of the 2004 ACM SIGMM Workshop on Effective Telepresence*); publisher and publication (ACM Press in 2004); and ISBN (1581139330).

127. I compared the Library of Congress Classification (call number), MLCM 2006/41169, with that shown on the cover. It is the same on both, except for the inclusion of “Copy 1,” which was a shelf indicator location and is not indicated in the OPAC or MARC records.

128. Pinhanez-III could have been located in the Library of Congress OPAC by searching for title (*Proceedings of the 2004 ACM SIGMM Workshop on Effective Telepresence*); and/or by searching the subject headings (e.g., *Video telephone – Congresses*; and/or *Multimedia systems – Congresses*).

129. The date stamped on Appendix D-1, as shown below, indicates receipt of Pinhanez-III by the Library of Congress on November 10, 2004. After such a receipt, it would have been sent to the cataloging department for cataloging/indexing. After that, it would have been immediately retrievable through the Library of Congress OPAC.



130. Appendix D-5 is the MARC record I downloaded from the Library of Congress OPAC on April 24, 2023. The MARC format provides information about the processing of Pinhanez-III by the Library of Congress. As mentioned above, the 9XX field in the MARC format is allocated to local libraries to enter information specific to that library. The Library of Congress has reserved the 955 field to indicate date of receipt of the published book and cataloging. The MARC 955 Field in this record reads:

955 — |a pv21 2005-04-12 to ASCD |a jx00 2006-01-27 |a ja15 2006-08-03 z-processor |a jf01 2006-08-25 to Ft. Meade UNASSESSED; |a xh00 2008-11-12 SEND TO DISCARD COPY NUMBERS GREATER THAN 2 UNLESS REFERENCE OR OTHER REQUEST TO BE KEPT

131. The 955 MARC Field indicates that Pinhanez-III was received by the catalog/index department of the Library of Congress on “2005-04-12” (April 12, 2005), and that it finished cataloging and was sent to the “Ft. Meade” location on “2006-08-25” (August 25, 2006).

132. The print copy of Pinhanez-III at the Library of Congress would have been available for public access one week to ten days after it was sent to the Fort Meade location on August 25, 2006.

133. Therefore Pinhanez-III would have been accessible to the public in the Library of Congress no later than September 4, 2006.

134. Once Pinhanez-III was entered into the general collection of the Library of Congress, members of the public could access the book by having it brought to either the Jefferson or Adams Reading Rooms.

135. As of 2006, members of the public could read, study, and make notes about a selected work in the Library of Congress Reading Rooms. Further, members of the public were permitted to make photocopies of portions of the work while in the Reading Rooms. Accordingly, Pinhanez-III was accessible to interested researchers and the general public when it was available at the Library of Congress, no later than September 4, 2006.

136. Evidence that Pinhanez-III was publicly accessible to researchers is further confirmed by articles that cite Pinhanez-III prior to October 1, 2010, thereby also confirming authenticity as an acceptable scholarly publication in the field.

137. Appendix D-6 is a true and correct download of the *Google Scholar* entries that cite Pinhanez-III. The citation list on this download consists of seven documents. Of those seven documents that cite Pinhanez-III, three are listed as follows: Appendix D-7 (Schmitt); Appendix D-8 (Olwal); and Appendix D-9 (Weber). I completed the download of Appendix D-6 from *Google Scholar* on May 7, 2023, at this URL: https://scholar.google.com/scholar?cites=15663163250938915955&as_sdt=5,47&sciodt=0,47&hl=en.

138. Appendix D-7 is an article by Christian Schmitt, Detlef Schoder, Kai Fischbach, and Steffen Muhle, titled “Towards Ambient Business: Enabling Open Innovation in a World of Ubiquitous Computing.” This article was published at Chapter 10, pages 251-279, in *Advances in Ubiquitous Computing: Future Paradigms and Directions*, which was edited by Soraya Kouadri Mostefaoui, Zakaria Maamar, and George M. Giaglis, and published by IGI Global in 2008. (“Schmitt”).

139. Schmitt was downloaded by me on April 26, 2023, enabled by my emeritus faculty position at Purdue University, through the Purdue University Libraries at this URL:

<https://www-igi-global-com.ezproxy.lib.purdue.edu/gateway/book/53>.

Schmitt is available for a fee at this URL within the published book:

<https://www-igi-global-com.ezproxy.lib.purdue.edu/gateway/book/53>.

140. Appendix D-7, Schmitt, shows Pinhanez-III in the alphabetical listing in References on page 276 as shown below:

Pinhanez, C., & Pingali, G. (2004). Projector-camera systems for telepresence. In G. Pingali, F. Nack, H. Sundaram (Ed.), *Proceedings of the 2004 ACM SIGMM workshop on Effective telepresence* (pp. 63-66). New York, NY: ACM Press.

141. The citation to Pinhanez-III in Schmitt is evidence that Pinhanez-III was known and accepted as a scholarly reference when Schmitt was published in the *Advances in Ubiquitous Computing: Future Paradigms and Directions* in 2008.

142. Appendix D-8 is a thesis by Alex Olwal titled “Unobtrusive Augmentation of Physical Environments: Interaction Techniques, Spatial Displays, and Ubiquitous Sensing”, submitted in fulfillment of the requirements for the Doctor of Philosophy degree (Ph.D.) on June 5, 2009 at KTH, Royal Institute of Technology, Stockholm, Sweden (“Olwal”).

143. Olwal was downloaded by me on April 26, 2023 from the open access link provided by DiVA portal. See DiVA portal “About” page, available at <https://www.diva-portal.org/smash/aboutdiva.jsf>. DiVA portal is “a finding tool for research publications and student theses written at . . . 50 universities and research

institutions” in Sweden. *Id.* I downloaded Olwal at this URL: <https://www.diva-portal.org/smash/get/diva2:217821/FULLTEXT02.pdf>.

144. In Appendix D-8, Olwal, Pinhanez-III is cited in the alphabetical listing in “References” on page 70 as shown below:

Pinhanez, C. and Pingali, G. (2004). Projector-camera systems for telepresence. *ETP '04: Proceedings of the 2004 ACM SIGMM workshop on Effective telepresence*, 63–66.

145. The citation to Pinhanez-III in Olwal is evidence that Pinhanez-III was known and accepted as a scholarly reference when Olwal submitted his thesis on June 5, 2009 at KTH, Royal Institute of Technology in Stockholm, Sweden.

146. Appendix D-9 is an article by Karsten Weber, titled “Mobile Devices and a New Understanding of Presence,” published in *Workshop Paper from SISSI2010 at the 12th Annual ACM International Conference on Ubiquitous Computing*, Copenhagen, Denmark, September 26-29, 2010, on pages 101-109 (“Weber”).

147. Weber was downloaded by me on April 26, 2023 from this URL: https://d1wqtxts1xzle7.cloudfront.net/50467482/20101215053038_Proceedings-SISSI-2010-libre.pdf?1479784028=&response-content-disposition=inline%3B+filename%3DA_Framework_for_Communication_and_Intera.pdf&Expires=1694618300&Signature=Es7kd6qs8tCJYdA7UqNFoM~J4wtOx2ETamx68famdICkKgnHq1r6OQJCtdDdAnNIpDQv-

[rpveYavEQagNPN8h38ki51oNZStTD4j6PMMFsFifxBBGWT1MhGCHr9OPFxDk
ochlhMtB7Y5dCrdfNMBK02pGaPuDGLw~4BetuYNOE-
2mOeW5LigBIKWonka0VejJ7-8COAylLAICEbj2I295kBJR3k094-
Hx05hfrdHKyV4cCz26Uq4IQyaZoEz7eB0FXfGk-
QdErep2Oa5MFL9mNQKJckiUEHqVbh8xZ~29cfKbFhhahAzW-
VcpOIiwTCr~Fzq2GysF3z~5MuUVh-w__&Key-Pair-
Id=APKAJLOHF5GGSLRBV4ZA](http://rpveYavEQagNPN8h38ki51oNZStTD4j6PMMFsFifxBBGWT1MhGCHr9OPFxDk
ochlhMtB7Y5dCrdfNMBK02pGaPuDGLw~4BetuYNOE-
2mOeW5LigBIKWonka0VejJ7-8COAylLAICEbj2I295kBJR3k094-
Hx05hfrdHKyV4cCz26Uq4IQyaZoEz7eB0FXfGk-
QdErep2Oa5MFL9mNQKJckiUEHqVbh8xZ~29cfKbFhhahAzW-
VcpOIiwTCr~Fzq2GysF3z~5MuUVh-w__&Key-Pair-
Id=APKAJLOHF5GGSLRBV4ZA)

148. Appendix D-9, Weber, shows Pinhanez-III as the 17th citation in References on page 109 as shown below:

17. Pinhanez, Cl., Pingali, G. Projector-camera systems for telepresence. Proc. of the 2004 ACM SIGMM Workshop on Effective Telepresence, ACM (2004), 63-66.

149. The citation to Pinhanez-III in Weber is evidence that Pinhanez-III was known and accepted as a scholarly reference when Weber was presented as a paper at the *12th Annual ACM International Conference on Ubiquitous Computing*, Copenhagen, Denmark, September 26-29, 2010.

150. Based on the above, I conclude that Pinhanez-III, submitted in this proceeding as Exhibit 1015, is an authentic document that was publicly accessible before October 1, 2010.

- E. Exhibit 1016: Tamotsu Machino, Yoshito Nanjo, Yoshimasa Yanagihara, Hiroaki Kawata, Satoshi Iwaki, and Ken-ichiro Shimokura. “Robot-Augmented Communication: A Remote-collaboration System based on a Shared Field of View in Real Space.” In *Proceedings 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Edmonton, AB, Canada, August 2-6, 2005, pp. 3467-3473 (“Machino-I”).**

151. I have been asked to opine on the authenticity and public accessibility of an article authored by Tomotsu Machino, et al., titled “Robot-Augmented Communication: A Remote-collaboration System Based on a Shared Field of View in Real Space,” published in *Proceedings 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Edmonton, AB, Canada, August 2-6, 2005, pp. 2203-2209, by IEEE in 2005 (“Machino-I”).

152. For consistency and clarification, Machino-I, when used in this report, refers not only to the Machino article but also to the publication in which it appears, *Proceedings 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*.

1. Authentication

153. I have evaluated Machino-I in four ways: (1) assessing Exhibit 1016, provided to me by counsel; (2) requesting a scan of the print copy owned by the Library of Congress through WTS, appended to this petition as Appendix E-1; (3) reviewing a download of Machino-I that I obtained from the *IEEE Xplore* on July 7,

2023, appended as Appendix E-2; and (4) by retrieving and reviewing articles that cite Machino-I, which were published prior to October 1, 2010.

154. Appendix E-1 is a scan of the print copy of Machino-I owned by the Library of Congress, which was obtained by WTS at my request on July 10, 2023.

155. These scans show: inside front cover and flyleaf, with the endpapers indicating ownership by the Library of Congress, as shown below; title page of Volume 4; verso of the title page (copyright) with a handwritten call number that reads “TJ210.3 .I44 2005, v.4, copy 1”, as shown below; Table of Contents, pages iii-liii; and the Machino article on pages 3467-3473.



156. All identifying characteristics of Appendix E-1, including endpapers and handwritten call numbers, are consistent with library practice and procedure that I have observed during my career as a professional librarian, specifically with those applied by the Library of Congress, and support a finding of authenticity. I have no cause for concern about the authenticity or accuracy of these attributes. Indeed,

Machino-I was found within the custody of a library, the Library of Congress, one of the most likely locations for an authentic publication to be located.

157. Appendix E-2 is a copy of Machino-I that I downloaded through the Purdue University Libraries on July 7, 2023 from the *IEEE Xplore* at the following URL:

<https://ieeexplore-ieee-org.ezproxy.lib.purdue.edu/stamp/stamp.jsp?tp=&arnumber=1545507>

Machino-I is also available for a fee at the following link:

<https://ieeexplore.ieee.org/abstract/document/1545507>

158. I compared Exhibit 1016 with Appendix E-1 and Appendix E-2. I found no difference between Exhibit 1016 and Appendices E-1 and E-2, except Exhibit 1016 has ‘color’, whereas the scans and download are black and white. But the content is the same. Therefore, given that I have located Machino-I in the Library of Congress and found it in a publisher proprietary database, I conclude that each of Exhibit 1016, Appendix E-1, and Appendix E-2 contains an authentic, true, and correct copy of Machino-I.

2. Public Accessibility

159. Appendix E-3 is a true and correct copy of the *WorldCat* entry for Machino-I. I obtained Appendix E-3 by completing a search on *WorldCat* on July 7, 2023 with Washington, D.C., as a geographical limitation. As shown in Appendix E-3, Machino-I is the document associated with this *WorldCat* entry, as verified by title

and ISBN. Consistent with *WorldCat* practice, this entry would have been created at or around August 6, 2006, the time of cataloging by the Library of Congress as discussed below.

160. When I searched *WorldCat* for holdings of Machino-I with the geographical designation of Washington, D.C., the Library of Congress was fourth on the list among 298 libraries shown to have Machino-I. As of 2010, this search could have been performed anywhere in the world by anyone who accessed *WorldCat* or its predecessor database through an OCLC member library in the early 2000s through today.

161. As of 2010, Machino-I could have been located in *WorldCat* by searching for the title (*Proceedings 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*) and/or by searching the subject headings (e.g., *Artificial intelligence*; *Intelligent control systems*; and/or *Robotics*, among others).

162. Appendix E-4 is a download from the Library of Congress OPAC on July 7, 2023. The document cataloged in this record is verified by the OPAC record fields including main title, publisher and publication date, and ISBN.

163. I compared the Library of Congress Classification call number, TJ210.3.I44 2005, with the call number shown in Exhibit 1016. It is the same except for the inclusion of “v.4 Copy 1,” which was a shelf indicator location and is not indicated in the OPAC or MARC records.

164. Machino-I could have been located in the Library of Congress OPAC by searching for title (*Proceedings 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*) and/or by searching the subject headings (e.g., *Robotics-Congresses*; *Artificial intelligence – Congresses*; and/or *Intelligent control systems – Congresses*).

165. Appendix E-5 is the MARC record I downloaded from the Library of Congress OPAC on July 7, 2023. The MARC format provides information about the processing of Machino-I by the Library of Congress. The 9XX field in the MARC format is allocated to local libraries to enter information specific to that library. The Library of Congress has reserved the 955 field to indicate date of receipt of the published book and cataloging. The MARC 955 Field in this record reads:

955 __ |a jf09 2006-08-01 z-processor |i jf09 2006-08-01 to sl |e jf09 2006-08-01 to Dewey |a aa07 2006-08-04

166. The 955 field above indicates that Machino-I was received by the Library of Congress on “2006-08-01” (August 1, 2006) and that cataloging was completed on “2006-08-04” (August 4, 2006).

167. The physical copy of Machino-I at the Library of Congress would have been available for public access one week to ten days after it was fully cataloged on August 4, 2006, received labelling, and was transferred to the stacks. Therefore,

Machino-I would have been accessible to the public in the Library of Congress no later than August 14, 2006.

168. Once Machino-I was entered into the general collection of the Library of Congress, members of the public could access Machino-I by having it brought to either the Jefferson or Adams Reading Rooms.

169. Evidence that Machino-I was publicly accessible is further confirmed by articles that cite Machino-I prior to October 1, 2010.

170. Appendix E-6 is a true and correct screen capture of two Google Scholar entries that cite Machino-I in 2009. I completed the screen capture on July 7, 2023, which can be accessed through the following URL:

https://scholar.google.com/scholar?cites=4216749433296779189&as_sdt=5,47&sciodt=0,47&hl=en.

171. Appendix E-7 is an article by Takafumi Matsumaru and Kosuke Akai, titled “Step-on Interface on Mobile Robot to Operate by Stepping on Projected Button,” published in *The Open Automation and Control Systems Journal*, volume 2, number 1, in 2009 on pages 85-95 (“Matsumaru – 1”).

172. I downloaded Matsumaru – 1 on July 7, 2023 from this open access URL: <https://benthamopen.com/contents/pdf/TOAUTOCJ/TOAUTOCJ-2-85.pdf>

173. In “References” on page 95 of Appendix E-7, Matsumaru – 1, Machino-I is shown as 20th in the list of references, as shown below. (Note that the pagination of the article is incorrectly entered.)

- [20] T. Machino, Y. Nanjo, Y. Yanagihara, H. Kawata, S. Iwaki, and K. Shimokura, “Robot-augmented communication: a remote-collaboration system based on a shared field of view in real space”, In: *2005 IEEE/RSJ Int’l Conf. on Intelligent Robots and Systems (IROS 2005)*, Aug 2-6, 2005, Edmonton, Alberta, Canada. Piscataway, NJ: IEEE, 2005, pp. 2203-2209.

174. The Matsumaru – 1 citation to Machino-I is evidence that Machino-I was known, publicly accessible, and accepted as a scholarly reference when Matsumaru – 1 was published in 2009 in the *Open Automation and Control Systems Journal*, volume 2, number 1.

175. Appendix E-8 is an article by Takafumi Matsumaru and Kosuke Akai, titled, “Functions of Mobile-Robot Step-On Interface” published in the *Journal of Robotics and Mechatronics*, volume 21, number 2 in 2009, on pages 267-276 (“Matsumaru – 2”).

176. I downloaded Matsumaru - 2 on July 7, 2023 through Purdue University Libraries at this URL: https://www.fujipress.jp/main/wp-content/themes/Fujipress/pdf_subscribed.php. It is available without a fee through *Research Gate* at this URL: https://www.researchgate.net/publication/263275136_Functions_of_Mobile-Robot_Step-On_Interface.

177. In “References” on page 275 of Appendix E-8, Matsumaru – 2, Machino-I is shown as 24th in the list of references, as shown below. (Note that the pagination is incorrectly entered.)

[24] T. Machino, Y. Nanjo, Y. Yanagihara, H. Kawata, S. Iwaki, and K. Shimokura, “Robot-augmented communication: a remote-collaboration system based on a shared field of view in real space,” IEEE/RSJ IROS 2005, pp. 2203-2209, 2005.

178. The Matsumaru – 2 citation to Machino-I is evidence that Machino-I was known, publicly accessible, and accepted as a scholarly reference when Matsumaru – 2 was published in 2009 in the *Journal of Robotics and Mechatronics*, volume 21, number 2 in 2009,

179. Based on the above, I conclude that Machino-I, submitted in this proceeding as Exhibit 1016, is an authentic document that was publicly accessible before October 1, 2010.

F. Exhibit 1017: Tamotsu Machino, Satoshi Iwaki, Hiroaki Kawata, Yoshimasa Yanagihara, Yoshito Nanjo, and Ken-ichiro Shimokura. “Remote-Collaboration System Using Mobile Robot with Camera and Projector.” In Proceedings 2006 IEEE International Conference on Robotics and Automation, Orlando, Florida, May 15-18, 2006. ICRA 2006 on pages 4063-4068, published by IEEE Robotics and Automation Society, 2006 (“Machino-II”).

180. I have been asked to opine on the authenticity and public accessibility of an article authored by Tomotsu Machino, et al., titled “Remote-Collaboration System Using Mobile Robot with Camera and Projector,” published in *Proceedings*

2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006 on pages 4063-4068. These proceedings were published by the IEEE Robotics and Automation Society in 2006 (“Machino-II”).

181. For consistency and clarification, Machino-II, when used in this report, refers not only to the Machino article but also to the publication in which it appears, *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006*.

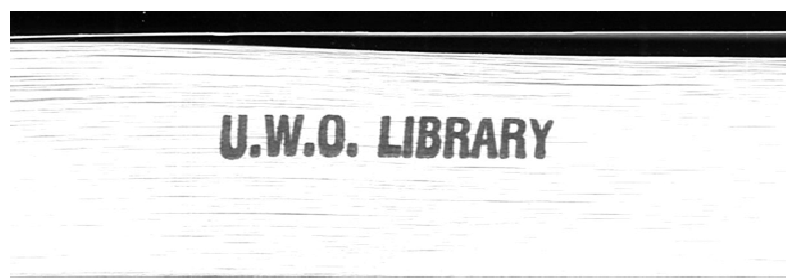
1. Authentication

182. I have evaluated the Machino-II reference in four ways: (1) by assessing a copy of Machino-II provided to me by counsel, Exhibit 1017; (2) by requesting a scan from Western University Libraries (London, Ontario) through WTS, appended to this declaration as Appendix F-1; (3) by accessing and assessing a download of Machino-II that I obtained from the *IEEE Xplore database* on April 25, 2023, enabled by my emeritus faculty status at Purdue University through the Purdue University Libraries, Appendix F-2; and (4) by retrieving and reviewing articles that cite Machino-II, published prior to October 1, 2010, discussed in Part V.F’s section on public accessibility, *infra*.

183. Appendix F-1 is a scan of the print copy of Machino-II owned by the Western University Libraries, obtained for me by WTS on April 11, 2023. Print copies of *Proceedings 2006 IEEE International Conference on Robotics and*

Automation, 2006. ICRA 2006 are rare, because it was a ten-volume print set, published nearly at the same time as the digital version; libraries frequently opted for the digital version and did not acquire the print copy. However, as discussed further in Part V.F's section on public accessibility, *infra*, access to Machino-II was available through the digital version.

184. Appendix F-1 is composed of: title page of the 2006 International Conference on Robotics and Automation (ICRA), with a bar code partially obscuring the title that reads, "Western Libraries UWD", as shown below; title page of volume 9 of 10; verso of the title page for volume 9; a print of the fore edge of that reads, "U.W.O. Library", as shown below; Table of Contents on pages iii-lxvii; and the Machino article on pages 4063-4068.



185. The identifying ownership barcode and fore edge ownership stamp are consistent with library practice and procedure that I have observed during my career as a professional librarian. I have no cause for concern about the authenticity or

accuracy of this ownership attribute. Machino-II was found within the custody of a library, the Western University Libraries, one of the most likely locations for an authentic publication to be located.

186. Evidence of the ownership of Machino-II by the Western University Libraries will be further explored and detailed below in section on public accessibility, *infra*.

187. Appendix F-2 is a download I made of Machino-II, enabled by my emeritus faculty position at Purdue University, through the Purdue University Libraries on April 25, 2023 from the *IEEE Xplore* at this URL: <https://ieeexplore-ieee-org.ezproxy.lib.purdue.edu/stamp/stamp.jsp?tp=&arnumber=1642326>.

Machino-II is also available for a fee at this URL: <https://ieeexplore.ieee.org/document/1642326>.

188. I compared Appendix F-1 and Appendix F-2, and I found no difference between the two. I then compared Appendix F-1 and Appendix F-2 with Exhibit 1017, provided to me by counsel, and found no differences among the three. Given the above, I conclude that each of Exhibit 1017, Appendix F-1, and Appendix F-2 contains an authentic, true, and correct copy of Machino-II.

2. Public Accessibility

189. Appendix F-3 is a true and correct copy of the *WorldCat* entry for Machino-II. I obtained Appendix F-3 by completing a search on *WorldCat* on May

9, 2023, by setting Williamsburg, Virginia as the geographical limiter. Only three holding libraries appear anywhere in the world: Western University, Canada; Technische Informationsbibliothek (TIB), Germany; and Kanagawa Institute of Technology, Japan.

190. Appendix F-3 shows that Machino-II is the document associated with this *WorldCat* entry, as verified by title (*2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006; Orlando, Florida, 15-19, 2006*) and ISBN (0780395050). Consistent with standard *WorldCat* practice, entry for these conference proceedings would have been created on or about July 14, 2006 at the time of original cataloging by a member library (as discussed below relative to the MARC record for the proceedings).

191. As a result, by 2010, Machino-II could have been located by searching for the title (*2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006; Orlando, Florida, 15-19*), and/or by searching the subject headings (e.g., *Automation; Automation-Congresses; and/or Robotics*).

192. By 2010, the search discussed above could have been performed anywhere in the world by anyone who accessed *WorldCat* and its predecessor database through an OCLC member library in the early 2000s.

193. Appendix F-4 is a screen capture I made from the Western University Libraries OPAC (online catalog) on May 9, 2023. The document cataloged in this

record is Machino-II as verified by the OPAC record fields including main title (*2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006; Orlando, Florida, 15-19, 2006*); publisher and publication (IEEE Robotics and Automation Society in 2006); and ISBN (0780395050).

194. Machino-II could have been located in the Western University Libraries OPAC by searching for title (*2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006; Orlando, Florida, 15-19, 2006*); and/or subject headings (e.g., *Robotics – Congresses*; and/or *Automation-Congresses*).

195. Appendix F-10 is the MARC record that I downloaded from the Western University library OPAC. The 008 MARC field reads “008 060714s2006 njua b 101 0 eng d.” This indicates that the proceedings including Machino-II were originally cataloged on July 14, 2006. It is unclear from the MARC record which library performed the original cataloging; nevertheless, on or around this date, the proceedings would have been added to *WorldCat*, as discussed above.

196. Appendix F-5 is a download from *IEEE Xplore* of the webpage that provides access to Machino-II. On Appendix F-5 toward the lower middle of the page it reads:

Date of Conference: 15-19 May 2006

Date Added to IEEE Xplore: 26 June 2006

197. The date above of June 26, 2006, provides evidence that Machino-II was accessible and retrievable to anyone who had access to *IEEE Xplore* as of June 26, 2006 through a personal membership in IEEE, as well as to anyone who had access through a subscribing institution to *IEEE Xplore*, which most research libraries would provide.

198. Evidence that Machino-II was publicly accessible to researchers is confirmed by articles that cite Machino-II prior to October 1, 2010, thereby also confirming authenticity as an acceptable scholarly publication in the field.

199. Appendix F-6 is a true and correct screen capture of the *Google Scholar* entries that cite Machino-II between the years 2007 and 2009. This citation list consists of six documents that cite Machino-II during that period of time. Three of these six documents are listed as follows: Appendix F-7, Park; Appendix F-8, Kawata; and Appendix F-9, Niitsuma. I completed the screen capture from *Google Scholar* on May 9, 2023, at this URL: https://scholar.google.com/scholar?hl=en&as_sdt=5%2C47&scioldt=0%2C47&cites=10317471256020941969&scipsc=&as_ylo=2007&as_yhi=2009.

200. Appendix F-7 is an article by Jongkyeong Park and Gerard Jounghyun Kim, titled “Robots with Projectors: An Alternative to Anthropomorphic HRI,” published in *Proceedings of the 4th ACM/IEEE International Conference on Human*

Robot Interaction, La Jolla, California, March 9-13, 2009 on pages 221-222. 2009 (“Park”).

201. Park, Appendix F-7, was downloaded by me on April 25, 2023 from the *ACM Digital Library*, enabled by my emeritus faculty position at Purdue University, through the Purdue University Libraries at this URL: <https://dl-acm-org.ezproxy.lib.purdue.edu/doi/abs/10.1145/1514095.1514146>. Park is available for a fee at this URL: <https://dl.acm.org/doi/abs/10.1145/1514095.1514146>.

202. Appendix F-7, a true and correct copy of the Park paper, shows Machino-II as the 9th citation in the “References” on page 222, as shown below:

[9] Machino, T., Iwaki, S., Kawata, H., Yanagihara, Y., Nanjo, Y. and Shimokura, K., Remote Collaboration System Using Mobile Robot with Camera and Projector, *Proc. of IEEE Intl. Conf. on Robotics and Automation*, pp. 4063-4068, 2006.

203. The citation to Machino-II in Park is evidence that Machino-II was known, publicly accessible, and accepted as a scholarly reference when Park was presented as a paper at the *4th ACM/IEEE International Conference on Human Robot Interaction*, La Jolla, California, March 9-13, 2009, and later published in its proceedings.

204. Appendix F-8 is an article by Hiroaki Kawata, Yosuke Takano, Yoshiyuki Iwata, Naoyoshi Kanamaru, Ken-ichiro Shimokura and Yoshihiro Fujita, titled “Field Trial of Asynchronous Communication Using Network-Based

Interactive Child Watch System for the Participation of Parents in Day-Care Activities,” published in the *Proceedings, 2008 IEEE International Conference on Robotics and Automation*, Pasadena, CA, USA May 19 – 23, 2008, on pages 2558-2563 (“Kawata”).

205. Appendix F-8, Kawata, was downloaded by me on April 25, 2023, from *IEEE Xplore*, enabled by my emeritus faculty position at Purdue University, through the Purdue University Libraries at this URL:

<https://ieeexplore-ieee-org.ezproxy.lib.purdue.edu/stamp/stamp.jsp?tp=&arnumber=4543598>.

Kawata is available for a fee at this URL:

<https://ieeexplore.ieee.org/abstract/document/4543598>.

206. Appendix F-8, a true and correct copy of the Kawata paper, shows Machino-II as the 7th citation in the “References” on page 2563, as shown below:

[7] T. Machino, et al. "Remote-Collaboration Sysytem Using Mobile Robot with Camera and Projector, " *Proc. IEEE Int. Conf. on Robotics and Automation*, 2006, pp. 4063- 4068.

207. The citation to Machino-II in Kawata is evidence that Machino-II was known, publicly accessible, and accepted as a scholarly reference when Kawata was presented as a paper at the *2008 IEEE International Conference on Robotics and Automation*, Pasadena, CA, USA, May 19 – 23, 2008.

208. Appendix F-9 is an article by Mihoko Niitsuma and Hideki Hashimoto, titled “Spatial Memory for Collaboration Support System in Intelligent Space-User Identification Based on Spatial Memory,” published in the proceedings of the *SICE Annual Conference 2007*, Takamatsu, Japan, September 17-20, 2007, on pages 3073-3079 (“Niitsuma”).

209. Niitsuma, Appendix F-9, was obtained for me by the Purdue University Libraries Interlibrary Loan Department from the print copy owned by the University of Texas at Dallas on April 25, 2023. Niitsuma is available for a fee at this URL: <https://ieeexplore.ieee.org/abstract/document/4421519>.

210. Appendix F-9, a true and correct copy of the Niitsuma paper, shows Machino-II as the 5th citation in “References,” on page 3079 as shown below:

[5] Tamotsu Machino, Yoshito Nanjo, Yoshimasa Yanagihara, Hiroaki Kawata, Satoshi Iwaki and Ken-ichiro Shimokura, “A Remote-collaboration System using a Mobile Robot with a Camera and a Projector”, Proc. 2006 IEEE Int. Conf. on Robotics and Automation, pp. 4063- 4068, 2006.

211. The citation to Machino-II in the Niitsuma article is evidence that Machino-II was known, publicly accessible, and accepted as a scholarly reference when Niitsuma was presented as a paper at the *SICE Annual Conference 2007*, Takamatsu, Japan, September 17-20, 2007, and subsequently published in its Proceedings.

212. Based on the above, I conclude that Machino-II, submitted in this proceeding as Exhibit 1017, is an authentic document that was publicly accessible before October 1, 2010.

G. Exhibit 1018: Tetsuo Iyoda, Tsutomu Abe, Kiwame Tokai, Shoji Sakamoto, Jun Shingu, Hiroko Onuki, Meng Shi, and Shingo Uchihashi. “LightCollabo: Distant Collaboration Support System for Manufacturers.” In *Advances in Multimedia Modeling: 14th International Multimedia Modeling Conference, MMM 2008, Kyoto, Japan, January 9-11, 2008, Proceedings*, vol. 4903, p. 369. Springer Science & Business Media, 2007 (“Iyoda”).

213. I have been asked to opine on an article by Tetsuo Iyoda, et al., titled “LightCollabo: Distant Collaboration Support System for Manufacturers,” published in *Advances in Multimedia Modeling: 14th International Multimedia Modeling Conference, MMM 2008, Kyoto, Japan, January 9-11, 2008, Proceedings*, volume 4903, page 369, Springer Science & Business Media, 2007 (“Iyoda”).

214. For consistency and clarification, Iyoda, when used in this report, refers not only to the Iyoda article but also to the publication in which it appears, *Advances in Multimedia Modeling: 14th International Multimedia Modeling Conference, MMM 2008, Kyoto, Japan, January 9-11, 2008, Proceedings*.

1. Authentication

215. I have evaluated the Iyoda reference four ways: (1) by assessing a copy of Iyoda provided to me by counsel, Exhibit 1018; (2) by assessing scans of a copy of Iyoda owned by the Cornell University Libraries, appended to this declaration as

Appendix G-1; and (3) by retrieving and assessing a copy of Iyoda I downloaded from the *Springer Link*, Appendix G-2.

216. Appendix G-1 is a scan provided to me at my request by the WTS on April 11, 2023 from the copy owned by Cornell University Libraries. Appendix G-1 includes: front cover; inside front cover with handwritten call number that reads, “QA76.575.I333 2008”, as shown below; a date due slip with a stamp that reads, Cornell University Libraries, Engineering Library, as shown below, and a date stamped above the “Date Due” slip that reads “MAR 28 2008”, as shown below; title page; verso of the title page (copyright) with a handwritten call number that reads, “ENGR QA75.575.I333 2008,” and a faint stamp of the Cornell University Libraries, as shown below; Table of Contents; and the Tetsuo Iyoda article, titled, “LightCollabo: Distant Collaboration Support System for Manufacturers.”



217. All identifying characteristics, such as stamps and notations, on Appendix G-1, are consistent with library practice and procedure that I have observed during my career as a professional librarian, specifically with those items I have had retrieved from the Cornell University Libraries. I have no cause for concern about the authenticity or accuracy of these identifying attributes. In addition, Iyoda was found within the custody of a library, the Cornell University Libraries, one of the most likely locations for an authentic publication to be located.

218. Evidence of the ownership of Iyoda by the Cornell University Libraries will be further explored and detailed in Part V.G's section on public accessibility, *infra*.

219. Appendix G-2 is a download I made of Iyoda, by my emeritus faculty position at Purdue University, through the Purdue University Libraries on May 2, 2023 from the *Spring Link digital database* at this URL: https://link-springer-com.ezproxy.lib.purdue.edu/chapter/10.1007/978-3-540-77409-9_35. Iyoda is available for a fee at this URL: <https://link.springer.com/book/10.1007/978-3-540-77409-9>.

220. I compared Appendix G-1 and Appendix G-2, and I found no difference between the two. I then compared Appendix G-1 and Appendix G-2 with Exhibit 1018, provided to me by counsel, and found no difference among the three. Given

the above, I conclude that each of Exhibit 1018, Appendix G-1, and Appendix G-2 contains an authentic, true, and correct copy of Iyoda.

2. Public Accessibility

221. Appendix G-3 is a true and correct copy of the *WorldCat* entry for Iyoda. I obtained Appendix G-3 by completing a search on *WorldCat* on May 10, 2023 by setting New York as the geographical limiter.

222. Appendix G-3 shows that Iyoda is the document associated with this *WorldCat* entry, as verified by title (*Advances in Multimedia Modeling: 14th International Multimedia Modeling Conference, MMM 2008, Kyoto, Japan, January 9-11, 2008, Proceedings*); publisher and publication date (Springer in 2008); and ISBN (3-540-77407-6). Consistent with *WorldCat* practice, the entry for Iyoda would have been added to the *WorldCat* database on, or soon thereafter, the date of cataloging by Cornell University Libraries on March 28, 2008.

223. Iyoda could have been located by searching for the title (*Advances in Multimedia Modeling: 14th International Multimedia Modeling Conference, MMM 2008, Kyoto, Japan, January 9-11, 2008, Proceedings*); and/or by searching the subject headings (e.g., *Computer graphics*; *Computer graphics – Congresses*; *Multimedia systems*; *Multimedia systems – Congresses*; *Visual communication* *Digital techniques*; and/or *Visual communication Digital techniques – Congresses*).

224. When I searched *WorldCat* for holdings of Iyoda in New York, Cornell University was on the list among the 497 libraries shown as holding Iyoda worldwide.

225. The search discussed above could have been performed anywhere in the world by anyone who accessed *WorldCat* or its predecessor database through an OCLC member library in the early 2000s through today.

226. Appendix G-4 is a download I made from the Cornell University Libraries OPAC (online catalog) on May 10, 2023. The document cataloged in this record is Iyoda as verified by the OPAC record fields including main title (*Advances in Multimedia Modeling: 14th International Multimedia Modeling Conference, MMM 2008, Kyoto, Japan, January 9-11, 2008, Proceedings*); publisher and publication date (Springer in 2008); and ISBN (3540774076).

227. I compared the Library of Congress Classification (call number), QA76.575 .I333 2008, with that shown inside the front cover and on the verso of the cover page. It is the same on both.

228. Iyoda could have been located in the Cornell University Libraries Library by searching for title (*Advances in Multimedia Modeling: 14th International Multimedia Modeling Conference, MMM 2008, Kyoto, Japan, January 9-11, 2008, Proceedings*); and/or by searching the subject headings (e.g., *Multimedia systems – Congresses*; *Computer graphics – Congresses*; and/or *Visual communication – Digital techniques – Congresses*).

229. The date stamped on Appendix G-1 of March 28, 2008 indicates receipt of Iyoda by the Cornell University Engineering Library, as shown below:



230. The date, March 28, 2008, is consistent with the MARC 948 Field shown on Appendix G-5. Cornell University's catalog department designates the 948 MARC field (as provided in the MARC protocol) for local holding notes. Cornell University Libraries staff insert the date of acquisition in the 948 field for each item added to the collections. The 948 fields read as shown below:

948 3 ⑆a 20080320 ⑆h YANKEE ⑆i YANK0311.dat

948 1 ⑆a 20080320 ⑆b f ⑆d fl31 ⑆e lts

231. The date in the 948 MARC FIELD reads: 20080320 (March 20, 2008). From my experience of over forty-four years as an academic and research librarian, the typical time for an item to be labeled and processed in an academic research

library is within a week to ten days. Therefore, the date of March 28, 2008 is consistent with the shelving within the Engineering Library.

232. Therefore, more likely than not, Iyoda would have been accessible to the public in the Engineering Library at Cornell University no later than March 28, 2008.

233. Based on the above, I conclude that Iyoda, submitted in this proceeding as Exhibit 1018, is an authentic document that was publicly accessible before October 1, 2010.

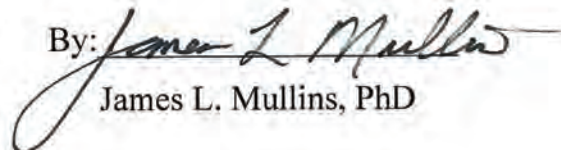
VI. RIGHT TO SUPPLEMENT

234. I reserve the right to supplement my opinions in the future to respond to any arguments that Patent Owner or its expert(s) or counsel may raise, and to take into account new information as it becomes available to me.

235. I declare that all statements made herein of my knowledge are true, and that all statements made on information and belief are believed to be true, and that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

Executed on: September 13, 2023

By:


James L. Mullins, PhD

Appendix A-1

Proceedings

Fifth IEEE International Conference on Automatic Face and Gesture Recognition

20–21 May 2002 • Washington, D.C.

The CD included with this book is
cataloged as **CDRM 496**. See Reference
Librarian for assistance.

Sponsored by

IEEE Computer Society TC PAMI



Los Alamitos, California

Washington • Brussels • Tokyo

TK 7895
S65 T5
2002

Copyright © 2002 by The Institute of Electrical and Electronics Engineers, Inc.
All rights reserved

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries may photocopy beyond the limits of US copyright law, for private use of patrons, those articles in this volume that carry a code at the bottom of the first page, provided that the per-copy fee indicated in the code is paid through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

Other copying, reprint, or republication requests should be addressed to: IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, P.O. Box 133, Piscataway, NJ 08855-1331.

The papers in this book comprise the proceedings of the meeting mentioned on the cover and title page. They reflect the authors' opinions and, in the interests of timely dissemination, are published as presented and without change. Their inclusion in this publication does not necessarily constitute endorsement by the editors, the IEEE Computer Society, or the Institute of Electrical and Electronics Engineers, Inc.

IEEE Computer Society Order Number PR01602

ISBN 0-7695-1602-5

Library of Congress Number 2002104474

Additional copies may be ordered from:

IEEE Computer Society
Customer Service Center
10662 Los Vaqueros Circle
P.O. Box 3014
Los Alamitos, CA 90720-1314
Tel: +1 714 821 8380
Fax: +1 714 821 4641
<http://computer.org/>
csbooks@computer.org

IEEE Service Center
445 Hoes Lane
P.O. Box 1331
Piscataway, NJ 08855-1331
Tel: +1 732 981 0060
Fax: +1 732 981 9667
[http://shop.ieee.org/store/](http://shop.ieee.org/store/customer-service@ieee.org)
customer-service@ieee.org

IEEE Computer Society
Asia/Pacific Office
Watanabe Bldg., 1-4-2
Minami-Aoyama
Minato-ku, Tokyo 107-0062
JAPAN
Tel: +81 3 3408 3118
Fax: +81 3 3408 3553
tokyo.ofc@computer.org

Editorial production by A. Denise Williams

Cover art production by Joseph Daigle/Studio Productions

Printed in the United States of America by The Printing House



6-5-02 624

Table of Contents

Fifth IEEE International Conference on Automatic Face and Gesture Recognition

Foreword	ix
Conference Organizers	x
Program Committee	xi
1.1 Face Analysis	
Eigen Light-Fields and Face Recognition across Pose	3
<i>R. Gross, I. Matthews, and S. Baker</i>	
Real-Time, Fully Automatic Upper Facial Feature Tracking	10
<i>A. Kapoor and R.W. Picard</i>	
A Unified Learning Framework for Real Time Face Detection and Classification.....	16
<i>G. Shakhnarovich, P.A. Viola, and B. Moghaddam</i>	
1.2 Detection and Recognition	
Consistency and Coupling in Human Model Likelihoods	27
<i>C. Sminchisescu</i>	
An Investigation into the Use of Partial-Faces for Face Recognition.....	33
<i>S. Gutta, V. Philomin, and M. Trajovic</i>	
Human Detection Using Geometrical Pixel Value Structures	39
<i>A. Utsumi and N. Tetsutani</i>	
An Appearance-Based Framework for 3D Hand Shape Classification and Camera Viewpoint Estimation	45
<i>V. Athitsos and S. Sclaroff</i>	
Poster Session I	
The CMU Pose, Illumination, and Expression (PIE) Database	53
<i>T. Sim, S. Baker, and M. Bsat</i>	
Smiling Faces Are Better for Face Recognition	59
<i>Y. Yacoob and L.S. Davis</i>	
Face Database Retrieval Using Pseudo 2D Hidden Markov Models	65
<i>S. Eickeler</i>	
Pose-Invariant Face Recognition with Parametric Linear Subspaces.....	71
<i>K. Okada and C. von der Malsburg</i>	
Recognition of Human Body Posture from a Cloud of 3D Data Points Using Wavelet Transform Coefficients	77
<i>N. Werghi and Y. Xiao</i>	
Facial Component Extraction and Face Recognition with Support Vector Machines.....	83
<i>D. Xi, I.T. Podolak, and S-W. Lee</i>	
Head Tracking by Active Particle Filtering	89
<i>Z. Zeng and S. Ma</i>	

An Automatic Method for Determining Quaternion Field Boundaries for Ball-and-Socket Joint Limits	95
<i>L. Herda, R. Urtasun, P. Fua, and A. Hanson</i>	
Non-Contact Eye Gaze Tracking System by Mapping of Corneal Reflections.....	101
<i>D.H. Yoo, J.H. Kim, B.R. Lee, and M.J. Chung</i>	
Multi-View Face Pose Estimation Based on Supervised ISA Learning.....	107
<i>S.Z. Li, X.H. Peng, X.W. Hou, H.J. Zhang, and Q.S. Cheng</i>	
Robust Facial Expression Recognition Using a State-Based Model of Spatially-Localised Facial Dynamics	113
<i>F. Bourel, C.C. Chibelushi, and A.A. Low</i>	
Mode-Based Multi-Hypothesis Head Tracking Using Parametric Contours.....	119
<i>Y. Chen, Y. Rui, and T. Huang</i>	
Hierarchical Wavelet Networks for Facial Feature Localization.....	125
<i>R.S. Feris, J. Gemmell, K. Toyama, and V. Krüger</i>	
Subpixel Eye Gaze Tracking	131
<i>J. Zhu and J. Yang</i>	
Baseline Results for the Challenge Problem of Human ID Using Gait Analysis	137
<i>P.J. Phillips, S. Sarkar, I. Robledo, P. Grother, and K.W. Bowyer</i>	
Algorithms for Inference in Specialized Maps for Recovering 3D Hand Pose	143
<i>R. Rosales and S. Sclaroff</i>	
Real-Time Multi-View Face Detection	149
<i>Z.Q. Zhang, L. Zhu, S.Z. Li, and H.J. Zhang</i>	
Gait Analysis for Recognition and Classification.....	155
<i>L. Lee and W.E.L. Grimson</i>	
Robust Full-Motion Recovery of Head by Dynamic Templates and Re-Registration Techniques	163
<i>J. Xiao, T. Kanade, and J.F. Cohn</i>	
An HMM-Based Temporal Difference Learning with Model-Updating Capability for Visual Tracking of Human Communicational Behaviors.....	170
<i>M.A.T. Ho, Y. Yamada, and Y. Umetani</i>	
On Probabilistic Combination of Face and Gait Cues for Identification	176
<i>G. Shakhnarovich and T. Darrell</i>	
Exemplar-Based Face Recognition from Video	182
<i>V. Krüger and S. Zhou</i>	
On Importance of Nose for Face Tracking	188
<i>D.O. Gorodnichy</i>	
1.3 Face Recognition	
Face Recognition Using Kernel Based Fisher Discriminant Analysis.....	197
<i>Q. Liu, R. Huang, H. Lu, and S. Ma</i>	
Face Identification across Different Poses and Illuminations with a 3D Morphable Model.....	202
<i>V. Blanz, S. Romdhani, and T. Vetter</i>	
Facial Asymmetry Quantification for Expression Invariant Human Identification.....	208
<i>Y. Liu, K.L. Schmidt, J.F. Cohn, and R.L. Weaver</i>	

Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods	215
<i>M-H. Yang</i>	
Face Recognition from Video: A CONDENSATION Approach	221
<i>S. Zhou, V. Krüger, and R. Chellappa</i>	
2.1 Evaluation	
Evaluation of Gabor-Wavelet-Based Facial Action Unit Recognition in Image Sequences of Increasing Complexity	229
<i>Y-L. Tian, T. Kanade, and J.F. Cohn</i>	
Meta-Analysis of Face Recognition Algorithms	235
<i>P.J. Phillips and E.M. Newton</i>	
A Comparison of Gabor Filter Methods for Automatic Detection of Facial Landmarks	242
<i>I.R. Fasel, M.S. Bartlett, and J.R. Movellan</i>	
2.2 Tracking and Motion	
Adaptive Color Space Switching for Face Tracking in Multi-Colored Lighting Environments	249
<i>H. Stern and B. Efron</i>	
Model-Based Head Pose Tracking with Stereovision	255
<i>R. Yang and Z. Zhang</i>	
An Adaptive Fusion Architecture for Target Tracking	261
<i>G. Loy, L. Fletcher, N. Apostoloff, and A. Zelinsky</i>	
Motion-Based Recognition of People in EigenGait Space	267
<i>C. BenAbdelkader, R. Cutler, and L. Davis</i>	
Poster Session II	
Two-Handed Gesture Tracking Incorporating Template Warping with Static Segmentation	275
<i>Y. Huang, T.S. Huang, and H. Niemann</i>	
Generative and Discriminative Face Modelling for Detection	281
<i>R.R. Wang, T. Huang, and J. Zhong</i>	
Model-Based Face Tracking for View-Independent Facial Expression Recognition	287
<i>S.B. Gokturk, J-Y. Bouguet, C. Tomasi, and B. Girod</i>	
A PCA/MDA Scheme for Hand Posture Recognition	294
<i>J. Deng and H.T. Tsui</i>	
A Modification of Kernel-Based Fisher Discriminant Analysis for Face Detection	300
<i>T. Kurita and T. Taguchi</i>	
A New Robust Face Detection in Color Images	306
<i>S. Srisuk and W. Kurutach</i>	
A SRN/HMM System for Signer-Independent Continuous Sign Language Recognition	312
<i>G. Fang and W. Gao</i>	
Hand Gesture Recognition Following the Dynamics of a Topology-Preserving Network	318
<i>F. Flórez, J.M. García, J. García, and A. Hernández</i>	

Multi-View Face Alignment Using Direct Appearance Models.....	324
<i>S.Z. Li, S.C. Yan, H.J. Zhang, and Q.S. Cheng</i>	
An Efficient Method for Recognition of Human Faces Using Higher Orders Pseudo Zernike Moment Invariant	330
<i>J. Haddadnia, M. Ahmadi, and K. Faez</i>	
Gait-Based Recognition of Humans Using Continuous HMMs.....	336
<i>A. Kale, A.N. Rajagopalan, N. Cuntoor, and V. Krüger</i>	
Robust Face Detection at Video Frame Rate Based on Edge Orientation Features	342
<i>B. Fröba and C. Küblbeck</i>	
Recognition of Arm Movements	348
<i>Z. Duric, F. Li, and H. Wechsler</i>	
Clustering Contextual Facial Display Sequences	354
<i>J. Hoey</i>	
An Approach to Automatic Recognition of Spontaneous Facial Actions.....	360
<i>B. Braathen, M.S. Bartlett, G. Littlewort, E. Smith, and J.R. Movellan</i>	
Silhouette-Based Human Identification from Body Shape and Gait	366
<i>R.T. Collins, R. Gross, and J. Shi</i>	
Stride and Cadence as a Biometric in Automatic Person Identification and Verification	372
<i>C. BenAbdelkader, R. Cutler, and L. Davis</i>	
Detection of People Carrying Objects: A Motion-Based Recognition Approach	378
<i>C. BenAbdelkader and L. Davis</i>	
On Modeling Variations for Face Authentication.....	384
<i>X. Liu, T. Chen, and B.V.K. Vijaya Kumar</i>	
Fast Stereo-Based Head Tracking for Interactive Environments.....	390
<i>L-P. Morency, A. Rahimi, N. Checka, and T. Darrell</i>	
Visual Prosody: Facial Movements Accompanying Speech	396
<i>H.P. Graf, E. Cosatto, V. Strom, and F.J. Huang</i>	
Interacting with Steerable Projected Displays	402
<i>R. Kjeldsen, C. Pinhanez, G. Pingali, J. Hartman, T. Levas, and M. Podlaseck</i>	
2.3 Gesture	
An Approach Based on Phonemes to Large Vocabulary Chinese Sign Language Recognition.....	411
<i>C. Wang, W. Gao, and S. Shan</i>	
Human Action Recognition in Smart Classroom.....	417
<i>H. Ren and G. Xu</i>	
Hand Gesture Recognition Using Multi-Scale Colour Features, Hierarchical Models and Particle Filtering.....	423
<i>L. Bretzner, I. Laptev, and T. Lindeberg</i>	
Real-Time Tracking of Multiple Fingertips and Gesture Recognition for Augmented Desk Interface Systems.....	429
<i>K. Oka, Y. Sato, and H. Koike</i>	
Author Index.....	435

Interacting with Steerable Projected Displays

Rick Kjeldsen, Claudio Pinhanez, Gopal Pingali, Jacob Hartman,
Tony Levas, Mark Podlaseck

IBM T.J. Watson Research Center

30 Sawmill River Road, Hawthorne, NY 10532 USA

fcmk / pinhanez / gpingali / jhartman / levas / podlasec@us.ibm.com

Abstract

When computer vision is combined with a steerable projector, any surface in an environment can be turned into an interactive interface, without having to modify or wire the surface. Steerable projected displays offer rich opportunities and pose new challenges for interaction based on gesture recognition. In this paper, we present real-time techniques for recognizing "touch" and "point" gestures on steerable projected displays produced by a new device called the Everywhere Displays projector (ED-projector). We demonstrate the viability of our approach through an experiment involving hundreds of users interacting with projected interfaces.

1. Introduction

Most displays today are tethered to special devices such as monitors. Projectors make it possible to display on any surface, but are static and typically tied to one designated surface in an environment. We have developed a new device called the *Everywhere Displays projector* [9], or ED-projector, that uses a computer-controlled mirror to steer a projected display on to any surface in an environment, while correcting for distortion caused by oblique projection. With the ED-projector, any surface can become a display and everyday objects can be enriched with information without having to wire them.

This ability to steer projected displays introduces the challenge of interacting naturally with such displays. If steerable displays could actually become steerable interfaces, it is possible to bring computer access to where the user is located or wherever the user desires. Indeed, an entire environment can act as the output device of an invisible computer to which a user's natural gestures and actions are the input. This lends another dimension to the vision of ubiquitous computing [12].

The ED-projector is composed of an LCD projector, a computer-controlled pan/tilt mirror, and a pan/tilt/zoom camera. The projector is connected to the display output of a host computer that also controls the mirror and the camera. The camera is steered to observe the projected display. Figure 1 shows two prototypes of ED-projectors

built with off-the-shelf components — rotating mirrors used in theatrical/discotheque lighting, steerable cameras and LCD projectors.

The use of projection systems to augment reality has been demonstrated by researchers in many different situations [7, 8, 10, 11, 13, 14, 16]. However, these systems were constrained to a fixed projector that could only project information on a limited area of an environment. Projected images can become interactive by using a vision system to detect the users' hand gestures [2],

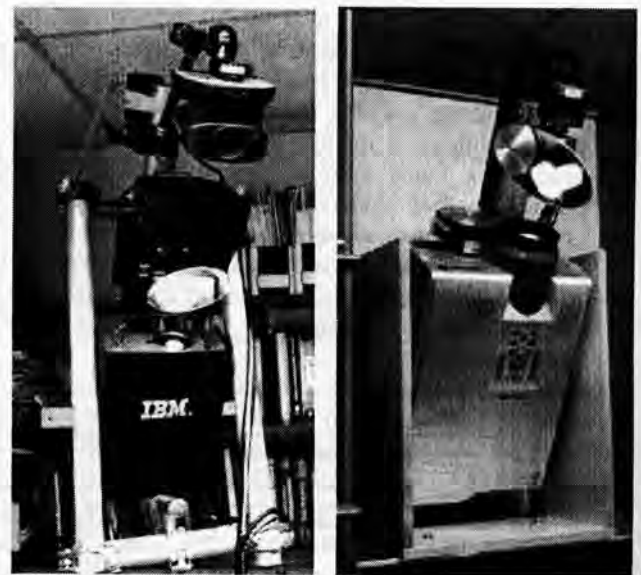


Fig. 1: Two prototypes of the Everywhere Displays projector.

moving objects [11], or the users' body position [6].

In this paper we document our first attempts at interacting with displays produced by the ED-projector. Users interact with the projected displays by touching them with their hands. These hand gestures are recognized by a computer vision system attached to the steerable camera of the ED-projector. Gesture recognition poses new challenges in the context of projected displays as the appearance of the user's hand changes drastically as it moves through the projection, and the projected light overwhelms the inherent color of the surfaces.

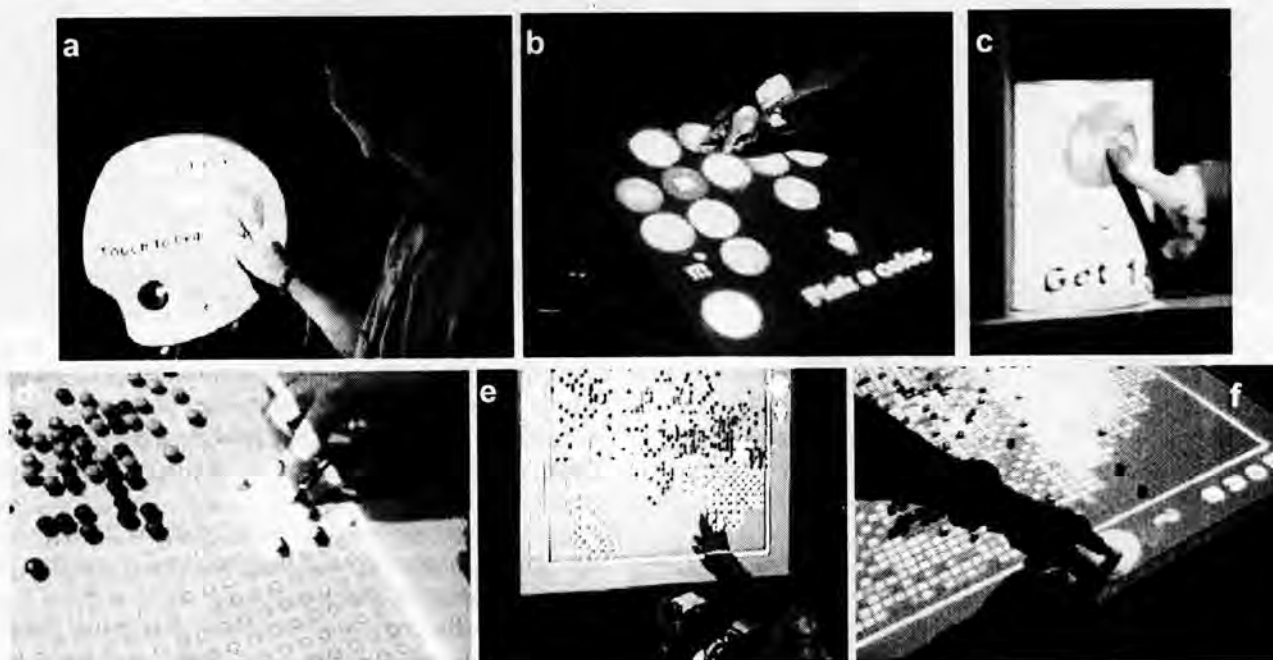


Fig. 2: Steps followed by each user when placing the M&M's in the SIGGRAPH'01 demo: a) the entrance panel is touched to start the process; b) a color is selected; c) the bucket containing the selected color is projected upon and the "done" button bucket is touched after the M&M's are picked up; d) the user places the M&M's on highlighted pixels; e) finger-painting to reveal the full image; and f) a "yes" button is touched after all interaction is done. (c&e from system camera)

2. An experiment on interaction with projected displays

We designed an experiment to both demonstrate and test the concept of transforming ordinary surfaces into interactive touch-screens. The experiment, which engaged hundreds of users in an augmented reality assembly task, was conducted in the Emerging Technologies area of the computer graphics conference, SIGGRAPH'01. As in any assembly task, each individual contributes to the assembly of an object by executing specific parts of the assembly process. In keeping with the theme of entertainment at our venue, we chose the assembled object to be a 60x50 picture made of M&M[®]1's (multi-colored sugar-coated chocolates) where each M&M is regarded as a "pixel" of the picture. Figure 3 shows an example, based on a portrait by Van Gogh.

In the front of our experimental space is a table; on the table lies a flat transparent Plexiglas board coated with double-sided transparent stick tape. This board is used as a mounting "canvas" for the M&M's and after assembly, as the hanging support for the pictures. On the back walls of the space are shelves with unlabeled buckets, each containing one of the 18 different colors of M&M's used.

Each bucket is covered with paper of different color and texture. Two other surfaces are also used for projection: a painter's palette mounted on an easel and a wood board first covered with one inch thick foam and then topped with white fabric. On the surrounding walls are completed M&M pictures, giving users a notion of the final goal of the assembly task.

2.1 The user experience

As the visitor arrives in the space, she sees the image of an M&M projected on the painter's palette with the instruction "Touch to begin", as shown in Fig. 2a. When the visitor touches the M&M "button," the system responds by projecting an arrow onto the palette and a message inviting the visitor to "Come in." The ED-projector is then steered towards the foam-covered-board where images of different colored M&Ms are projected along with the invitation "Pick a color" (see fig. 2b). Here, the user can select which color to place in this iteration by simply touching the M&M of the desired color.

As soon as the system detects the user selecting the color she wants to place, it projects a message on the foam covered-board, asking the user to get a certain number of M&M's from a highlighted can. The ED-projector then rotates the mirror so it highlights the bucket that contains the M&Ms of that color. It displays the message "Get n ", (where n is the number of M&Ms to be picked up from the

¹ "M&M" is a registered trademark of Mars, Inc..

bucket) on the bucket's frontal face. In addition, a button with the word "done" on it, *i.e.*, an image of an M&M underlined by a pointing finger (see Fig. 2c), is also displayed on the can. After retrieving the appropriate number of M&M's from inside the bucket, the visitor has to touch this "done" button to communicate that the picking has been completed (Fig. 2c). Immediately afterward an instruction to "Go to the table" is projected on the bucket and the ED-projector is redirected to the Plexiglas board.



Fig. 3: A picture made of M&M's

The ED-projector then indicates the precise location where the M&M's should be placed (Figs. 2d) by projecting an image of the target locations on the canvas. It also displays the instructions "Place the M&M's", and "Done ?" and an M&M button labeled "yes". After the user places the M&M's, she communicates the end of her action by touching this button. At this point, the user is invited to "Finger paint." The vision system tracks the position of the user's hand and fills in circles of the appropriate color in the vicinity of the fingertip, interactively completing the picture (Fig. 2e). At any time the user can stop this activity by touching the "yes" button to the right of the image (Fig. 2f). The system times out after 40 seconds and shows the complete image to the user.

3. Gesture recognition for user input

A user's interactions with the Everywhere Display are detected with a single pan/tilt/zoom camera that is steered to follow the projected image. A vision system examines the video stream for user actions and generates events for the application software. The camera is located adjacent to the mirror assembly so that the camera's view of the user's hand interacting with the display is generally occluded only if she occludes the projected image itself, a situation

that is apparent to the users and which we assume she will quickly correct.

The user interface is composed of individual interactive components (widgets), similar to the way current GUIs are composed of scroll bars, buttons, menus and the like. Each widget provides a basic type of interaction, such as triggering an event or controlling the value of a parameter. A widget need not have a visible representation on the display. Just as with current interfaces, when the task changes the set of active widgets changes as well. For insights into the design philosophy of the user interface see [5].

The vision system makes no attempt to model the user's activity. Instead it examines portions of the video stream for image events that indicate a user has interacted with a widget, and generates an application event in response.

We have implemented two types of widgets for this system: a *button* the user touches, and a *tracking region* where the location the user is pointing at is determined. In both cases the vision system first determines when and where the user is pointing. For buttons it then looks for a button touch motion in the trajectory of the pointing fingertip. For tracking it passes the fingertip position through a transfer function similar to that described in [4] to obtain a pleasing pointing dynamics, and translates to the coordinate system of the region.

3.1 Pointing Detection

Detecting the user's hand in the context of an ED-projector presents several challenges. For one the appearance of the user's hand changes drastically as it moves through the projection. This makes techniques based on color or appearance unusable. Similarly, techniques based on background subtraction often give unreliable results, as the projected image can completely overwhelm the inherent color of the moving surface. Hence, many of the techniques used in other gesture recognition systems [15] will not work with projected displays.

Even though the appearance of an object will change as it moves across a projected image, it will create a region of changed pixels that retains the basic shape of the moving object (see Figs. 4a and 4b). Therefore each video frame is subtracted from the frame before it, and noise is removed with computational morphology.

When people reach out to touch or point at something, they use a variety of hand shapes, but almost invariably there is a finger that extends further than the rest by some amount leading the way. We assume this fingertip provides the most accurate estimate of where a user is pointing.



Fig. 4. a) Camera view of an interaction with a bucket; b) image difference data; c) overlay of search region (square), button active area (circle), and the fingertip template shown at the pointing location.

We find fingertip-like shapes by convolving a fingertip template (Fig. 5) over the difference image using a matching function tuned so that the gray level in the template determines the desirability of finding a changed pixel at that location, and changed pixels in black regions of the template are penalized more severely than unchanged pixels in the white regions. If the template does not match well anywhere in the image, we assume the user is not pointing.

The fingertip template is deliberately kept short so that it will match fingertips that extend only slightly beyond their neighbors and will match fingertips within a wider range of angles. As a result the template often matches well at several points in the image. We resolve between these hypotheses by using the "fingertip" furthest from the user.

This approach supports a tracking rate between 5 and 30 frames/second on a 500 MHz workstation for a 320x240 image, depending on the size of the search region and the size of the fingertip template (as determined by the expected size of the user's hand).

3.2 Button Touch Detection

A button touch event is defined to occur when a fingertip is in the vicinity of a button, travels away from the user to a point within the button and then returns toward the user. Button touches are detected by examining the hand trajectory for several specific patterns that indicate this type of motion. If more than one button is present in a configuration, and there is any ambiguity about which has been touched, only the button furthest from the user is allowed to generate an event.

This algorithm works very well on interactions where the user is asked to touch one button at a time. Importantly, it also resists generating an event when the user's hand moves through a button on the way to or from another location. The algorithm can fail when the user touches several buttons without retracting their hand, or "flies" their finger around in the image before touching a button.

Notice that although our vision system was built to recognize the touching of a button, the image of a button

tends to elicit from the user a slightly different gesture, the pressing of the button. We will address the consequences of this difference later.

3.3 Calibration

For each surface where the user will interact with the display, the vision system requires knowledge about the location and size of the user's hand. While this information could be inferred dynamically from the shape and motion of the changed region, to keep the implementation simple we obtain it during calibration by sizing and rotating a hand icon to match the image of the hand. This approach assumes that the user will approach the interaction area from a consistent location each time. In practice this assumption worked well for most surfaces, but failed often for others.

During calibration we also show the system the location and size of buttons and tracking regions. A search region is identified for each surface in order to help the system ignore extraneous movement in the image and speed response (see Fig. 4c).

The optimal recognition behavior varies from surface to surface, depending on the requirements of the task. For example on the selection screen (Figure 2b) false positive button presses are very disruptive to the demo, while false negatives, where the user must press again, are less disruptive. Conversely on the buckets of M&Ms the false negative rate is "naturally" higher because the user's pointing behavior is less consistent. Here false positives are less disruptive to the demo, while false negative results make the problem worse.

Several parameters can be used to adjust the inherent tradeoffs in the recognition system to match these desired characteristics. In the case of the selection screen we have



Fig. 5: Fingertip template.

the system emphasize a good fingertip template match, which has the effect of lowering the false positive rate, but increasing the false negative rate. For the buckets, we place less emphasis on a good template match, as users tend to point at the can buttons from a wide range of orientations. We also ensure the search region extends well above the button so that when the user reaches into the can for M&Ms the trajectory is more likely to fly through the button and not stop inside.

All calibration information for a surface is saved in a configuration file that is loaded each time the display switches to that surface.

4. Experimental Results

During the 6 days of SIGGRAPH '01 more than 650 people went through our demo and 4 complete M&M pictures were composed. The camera video stream was recorded onto tape and the users' selections were logged.

One objective of the SIGGRAPH experiment was to use the large number of subjects to help determine the main issues associated with the usage of projected displays, which could then be explored more deeply in controlled studies. Since most of our users were confronting a novel interaction paradigm, we were not shy of giving them instruction as needed.

A problem that we detected early on was the 2 to 9 second delay required to move the ED-projector's mirror and adjust focus between surfaces. Even the fastest transitions seem to be perceived as too long by the users of the system. The problem seems to be exacerbated by the lack of familiarity with the application, and by not knowing on which surface of the environment the display would appear next.

To minimize the perceived delay, the interaction was changed so that the user always received some information about what to expect next before the time-consuming mirror movement. For instance, after the user touched the "Done" button on a bucket (see Fig. 2c), the instruction "Go to the table" was immediately projected. Although this helped to alleviate some of the anxiety, confusion and frustration we observed before the change, the delays were still uncomfortably long.

Considering that the vision system and the projection system were integrated for the first time only in the weeks preceding the exhibition, the combined system worked remarkably well. A sample with 130 consecutive users with 621 button touch events (touching gestures or false detections) yielded correct detection of touching gestures up to 81%, with 12% of false negatives and 7% of false positives, as shown in table 1. If the buckets are excluded from the count, the performance exceeds 89%.

The buckets yielded a high number of errors for several reasons. The biggest problem was that while picking M&M's from the bucket, the user often partially or completely occluded the display with their head or back.

	Success	false positives	false negatives
Entrance	88%	9%	3%
select board	94%	2%	4%
buckets	55%	33%	12%
board (place)	91%	8%	2%
board (paint)	80%	2%	18%
total w/o buckets	89%	6%	5%
total	81%	12%	7%

Table 1: Performance of the vision system (130 users).

The user's motion in such situations sometimes triggered false positives for the button projected on the bucket.

Some false negative errors were due to the fact that the system was tuned to detect a straight, forward, out-and-back touch motion and the user sometimes assumed otherwise. For example, when picking a color, some users would fly their hand over several buttons, then press one with a vertical motion that was not readily visible to the camera. Others tried to wave their hands through the buttons or made similar non-touching actions. The later problem seemed particularly acute on non-standard interaction surfaces like the buckets.

The assumption that a user will not interact with an occluded button held true for most interactions, but there was one instance when it was repeatedly violated. After picking out their M&Ms from a bucket, a user would often begin to walk toward the table, then realize that they had not touched the "Done" button. Returning to the bucket they would occlude the projection (and camera) with their head or back just before they arrived, but they still continued to reach out and touch the now dark bucket.

Finally, we credit the high number of false negatives during finger painting to the reduced frame rate when simultaneously tracking over a large area and detecting touch events. Occasionally the brief interval when the finger was inside the button would fall between video frames.

Either when trying to click a button on the hard surface of the Plexiglas board or the soft foam board, our users almost always applied pressure to the projected "buttons." When a button touch was not detected by the vision system, the most common reaction was to try to press again with more pressure. Other common reactions were to use the whole hand instead of a finger, and to change the touch motion to a wave or some other action.

To reduce the excessive pressure users would often apply to the surfaces, we changed the wording of the instruction from "Press to begin" to "Touch to begin" (Fig. 7a). Although this helped somewhat, it seems that people expect buttons, even projected buttons, to react to pressure.

Users also seemed to react differently to the various surfaces on which the display was projected. Our users

looked more tentative interacting with the fabric-covered soft board and, especially, the buckets than with the painter's palette in the entrance and the Plexiglas board on the table. We often had to tell users to "Click the bucket" after picking out the M&Ms, while interaction with the more traditional surfaces came naturally. Are buckets expected to be less interactive than tables? Our experience at SIGGRAPH suggests an interesting hypothesis: people seem to attribute different interactive capabilities to surfaces transformed into touch-screens according to the nature of the surfaces themselves. If true, this observation may have implications in a variety of areas, e.g. [3]

5. Conclusions

The ED-projector introduces a new concept: projecting an interactive display onto arbitrary surfaces in the environment. This technology can be used as a generic input/output device that can replace, in many situations, current displays and interactive devices. Instead of being limited to a fixed display, an application can now move to wherever it is needed in the environment, such as a phone book by the phone, or a database of papers by the file cabinet. Computer and information access can be provided in spaces where traditional displays can be broken or stolen. An interactive display can also be brought to the proximity of a user without requiring the user to move. In particular, the ED-projector can facilitate the access and use of computers by people with locomotive disabilities. For instance, it can project an interactive display on a hospital bed sheet without patient contact with any device.

Projected displays also enable a new set of applications where a computer acts on the physical world, almost like a robotic arm made of light. These applications can use the ED-projector to point to physical objects, show connections among them, and project patterns to indicate movement or change in the real world. The experiment described here is an example of this class of applications. We also see the Everywhere Displays projector as a potential enabler of a new generation of games that happen not in the virtual world but are projected into the physical, everyday world where we live [1].

In this paper, the viability of the ED concept was demonstrated in an experiment where we guided several hundred first-time users through an assembly task. The experiment was too simple to enable us to draw definitive conclusions about how people react to a system that transforms everyday surfaces into touch-screens. However, it has alerted us about some limitations of the technology and has suggested guidelines for the design of such applications. For example, it is evident that the switching time between surfaces has to be either shortened or filled with some sort of user feedback.

The analysis of our errors in detecting user input points out many ways in which we can improve the performance of the vision system. It has become clear that,

like in many other computer vision systems, domain knowledge is essential to obtaining good performance. This might include knowledge of where the user is located during the interaction, what types of activities she is likely to perform outside the context of the interaction, and what types of errors are more or less disruptive to the interaction.

Some of the most interesting research concerns what interaction mechanisms are best suited for a display projected on an arbitrary surface. Is a "button" the right metaphor for interaction? As we demonstrated, we can detect button-pressing actions with reasonable accuracy, even with a single camera, but is this the right paradigm from the user's point of view for a widget that detects the hand position, not pressure? How can more flexible and expressive hand gestures play a role in these interfaces? Should other body gestures and natural actions by the user also play a role?

References

1. Björk, S., et al. Pirates! - Using the Physical World as a Game Board. In Proc. of Interact'01. 2001. Tokyo, Japan.
2. Crowley, J.L., et al., Things that See. Communications of the ACM, 2000. 43(3): p. 54-64.
3. Ishii, H. and B. Ullmer. Tangible Bits: Towards Seamless Interfaces between People, Bits, and Atoms. In Proc. of CHI'97. 1997. Atlanta, Georgia.
4. Kjeldsen, R., "Head Gestures for Computer Control", in the Proceedings of the Workshop on Recognition And Tracking of Face and Gesture - Real Time Systems (RATFG-RTS), Vancouver, BC, Canada, July 2001
5. Kjeldsen, R. and Hartman, J, Design Issues for Vision-based Computer Interaction Systems. In Proc. of the Workshop on Perceptual User Interfaces. 2001. Orlando, Florida, USA.
6. Keays, B. and R. Macneil. metaField Maze. In Proc. of SIGGRAPH'99. 1999. Los Angeles, California.
7. Krueger, M.W., Artificial Reality II. 1990: Addison-Wesley.
8. Morishima, S., et al. HyperMask: Talking Head Projected Onto Real Objects. In Proc. of Multimedia Modeling (MMM'00). 2000: World Scientific.
9. Pinhanetz, C. The Everywhere Displays Projector: A Device to Create Ubiquitous Graphical Interfaces. In Proc. of Ubicomp'01. 2001. Atlanta, Georgia.
10. Rekimoto, J. A Multiple Device Approach for Supporting Whiteboard-based Interactions. In Proc. of CHI'98. 1998. Los Angeles, CA.
11. Underkoffler, J., et al. Emancipated Pixels: Real-World Graphics in the Luminous Room. In Proc. of SIGGRAPH'99. 1999. Los Angeles, CA.
12. Weiser, M., The Computer for the Twenty-First Century. Scientific American, 1991: p. 94-100.
13. Welch, G., et al., Projected Imagery in Your "Office of the Future". IEEE Computer Graphics and Applications, 2000(July/August): p. 62-67.
14. Wellner, P., Interacting with Paper on the DigitalDesk. Communications of the ACM, 1993. 36(7).
15. Wu, Y. and T. Huang, Vision-Based Gesture Recognition: A Review. Lecture Notes in Artificial Intelligence, 1999. 1739.
16. Yang, R. and G. Welch. Automatic and Continuous Projector Display Surface Calibration Using Every-Day Imagery. In Proc. of 9th International Conf. in Central Europe in Computer Graphics, Visualization, and Computer Vision. 2001. Plzen, Czech Republic.

Appendix A-2

Interacting with Steerable Projected Displays

Rick Kjeldsen, Claudio Pinhanez, Gopal Pingali, Jacob Hartman,
Tony Levas, Mark Podlaseck
IBM T.J. Watson Research Center
30 Sawmill River Road, Hawthorne, NY 10532 USA
fcmk | pinhanez | gpingali | jhartman | levas | podlasec@us.ibm.com

Abstract

When computer vision is combined with a steerable projector, any surface in an environment can be turned into an interactive interface, without having to modify or wire the surface. Steerable projected displays offer rich opportunities and pose new challenges for interaction based on gesture recognition. In this paper, we present real-time techniques for recognizing “touch” and “point” gestures on steerable projected displays produced by a new device called the Everywhere Displays projector (ED-projector). We demonstrate the viability of our approach through an experiment involving hundreds of users interacting with projected interfaces.

1. Introduction

Most displays today are tethered to special devices such as monitors. Projectors make it possible to display on any surface, but are static and typically tied to one designated surface in an environment. We have developed a new device called the *Everywhere Displays projector* [9], or ED-projector, that uses a computer-controlled mirror to steer a projected display on to any surface in an environment, while correcting for distortion caused by oblique projection. With the ED-projector, any surface can become a display and everyday objects can be enriched with information without having to wire them.

This ability to steer projected displays introduces the challenge of interacting naturally with such displays. If steerable displays could actually become steerable interfaces, it is possible to bring computer access to where the user is located or wherever the user desires. Indeed, an entire environment can act as the output device of an invisible computer to which a user’s natural gestures and actions are the input. This lends another dimension to the vision of ubiquitous computing [12].

The ED-projector is composed of an LCD projector, a computer-controlled pan/tilt mirror, and a pan/tilt/zoom camera. The projector is connected to the display output of a host computer that also controls the mirror and the camera. The camera is steered to observe the projected display. Figure 1 shows two prototypes of ED-projectors

built with off-the-shelf components — rotating mirrors used in theatrical/discotheque lighting, steerable cameras and LCD projectors.

The use of projection systems to augment reality has been demonstrated by researchers in many different situations [7, 8, 10, 11, 13, 14, 16]. However, these systems were constrained to a fixed projector that could only project information on a limited area of an environment. Projected images can become interactive by using a vision system to detect the users’ hand gestures [2],



Fig. 1: Two prototypes of the Everywhere Displays projector.

moving objects [11], or the users’ body position [6].

In this paper we document our first attempts at interacting with displays produced by the ED-projector. Users interact with the projected displays by touching them with their hands. These hand gestures are recognized by a computer vision system attached to the steerable camera of the ED-projector. Gesture recognition poses new challenges in the context of projected displays as the appearance of the user’s hand changes drastically as it moves through the projection, and the projected light overwhelms the inherent color of the surfaces.

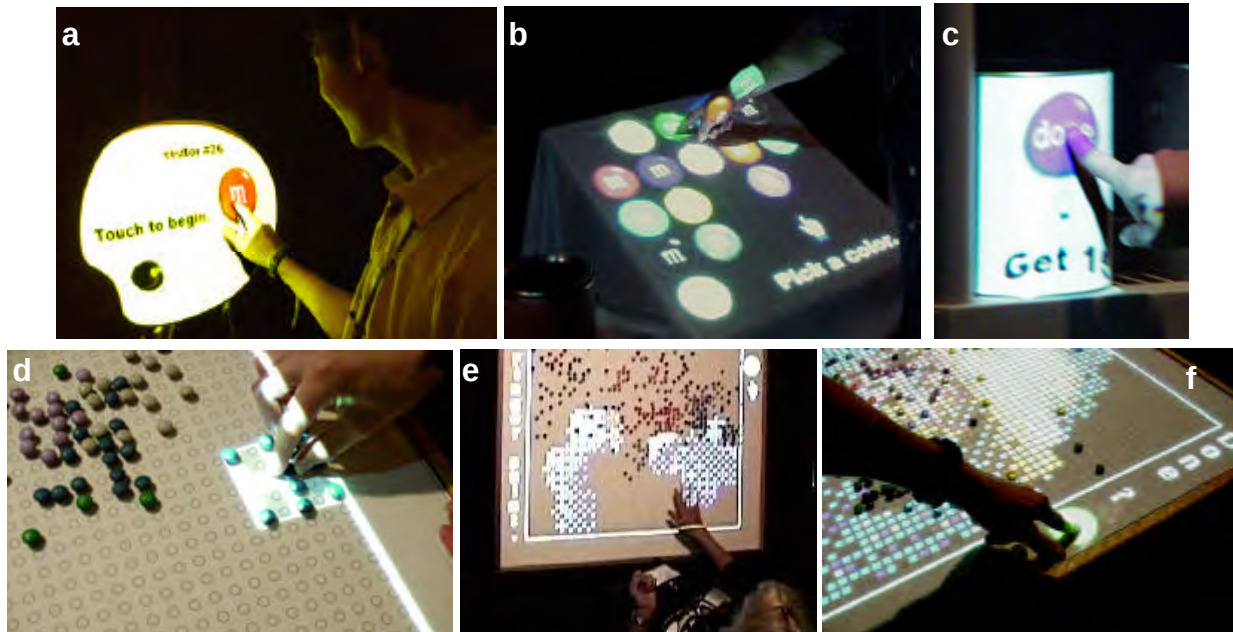


Fig. 2: Steps followed by each user when placing the M&M's in the SIGGRAPH'01 demo: a) the entrance panel is touched to start the process; b) a color is selected; c) the bucket containing the selected color is projected upon and the "done" button bucket is touched after the M&M's are picked up; e) the user places the M&M's on highlighted pixels; e) finger-painting to reveal the full image; and f) a "yes" button is touched after all interaction is done. (c&e from system camera)

2. An experiment on interaction with projected displays

We designed an experiment to both demonstrate and test the concept of transforming ordinary surfaces into interactive touch-screens. The experiment, which engaged hundreds of users in an augmented reality assembly task, was conducted in the Emerging Technologies area of the computer graphics conference, SIGGRAPH'01. As in any assembly task, each individual contributes to the assembly of an object by executing specific parts of the assembly process. In keeping with the theme of entertainment at our venue, we chose the assembled object to be a 60x50 picture made of M&M[®]'s (multi-colored sugar-coated chocolates) where each M&M is regarded as a "pixel" of the picture. Figure 3 shows an example, based on a portrait by Van Gogh.

In the front of our experimental space is a table; on the table lies a flat transparent Plexiglas board coated with double-sided transparent stick tape. This board is used as a mounting "canvas" for the M&M's and after assembly, as the hanging support for the pictures. On the back walls of the space are shelves with unlabeled buckets, each containing one of the 18 different colors of M&M's used.

Each bucket is covered with paper of different color and texture. Two other surfaces are also used for projection: a painter's palette mounted on an easel and a wood board first covered with one inch thick foam and then topped with white fabric. On the surrounding walls are completed M&M pictures, giving users a notion of the final goal of the assembly task.

2.1 The user experience

As the visitor arrives in the space, she sees the image of an M&M projected on the painter's palette with the instruction "Touch to begin", as shown in Fig. 2a. When the visitor touches the M&M "button," the system responds by projecting an arrow onto the palette and a message inviting the visitor to "Come in." The ED-projector is then steered towards the foam-covered-board where images of different colored M&Ms are projected along with the invitation "Pick a color" (see fig. 2b). Here, the user can select which color to place in this iteration by simply touching the M&M of the desired color.

As soon as the system detects the user selecting the color she wants to place, it projects a message on the foam covered-board, asking the user to get a certain number of M&M's from a highlighted can. The ED-projector then rotates the mirror so it highlights the bucket that contains the M&Ms of that color. It displays the message "Get n ", (where n is the number of M&Ms to be picked up from the

¹ "M&M" is a registered trademark of Mars, Inc..

bucket) on the bucket's frontal face. In addition, a button with the word "done" on it, *i.e.*, an image of an M&M underlined by a pointing finger (see Fig. 2c), is also displayed on the can. After retrieving the appropriate number of M&M's from inside the bucket, the visitor has to touch this "done" button to communicate that the picking has been completed (Fig. 2c). Immediately afterward an instruction to "Go to the table" is projected on the bucket and the ED-projector is redirected to the Plexiglas board.

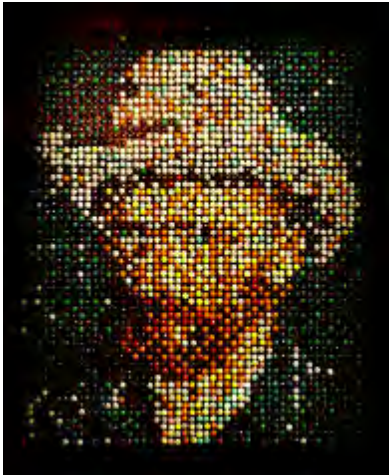


Fig. 3: A picture made of M&M's

The ED-projector then indicates the precise location where the M&M's should be placed (Figs. 2d) by projecting an image of the target locations on the canvas. It also displays the instructions "Place the M&M's", and "Done ?" and an M&M button labeled "yes". After the user places the M&M's, she communicates the end of her action by touching this button. At this point, the user is invited to "Finger paint." The vision system tracks the position of the user's hand and fills in circles of the appropriate color in the vicinity of the fingertip, interactively completing the picture (Fig. 2e). At any time the user can stop this activity by touching the "yes" button to the right of the image (Fig. 2f). The system times out after 40 seconds and shows the complete image to the user.

3. Gesture recognition for user input

A user's interactions with the Everywhere Display are detected with a single pan/tilt/zoom camera that is steered to follow the projected image. A vision system examines the video stream for user actions and generates events for the application software. The camera is located adjacent to the mirror assembly so that the camera's view of the user's hand interacting with the display is generally occluded only if she occludes the projected image itself, a situation

that is apparent to the users and which we assume she will quickly correct.

The user interface is composed of individual interactive components (widgets), similar to the way current GUIs are composed of scroll bars, buttons, menus and the like. Each widget provides a basic type of interaction, such as triggering an event or controlling the value of a parameter. A widget need not have a visible representation on the display. Just as with current interfaces, when the task changes the set of active widgets changes as well. For insights into the design philosophy of the user interface see [5].

The vision system makes no attempt to model the user's activity. Instead it examines portions of the video stream for image events that indicate a user has interacted with a widget, and generates an application event in response.

We have implemented two types of widgets for this system: a *button* the user touches, and a *tracking region* where the location the user is pointing at is determined. In both cases the vision system first determines when and where the user is pointing. For buttons it then looks for a button touch motion in the trajectory of the pointing fingertip. For tracking it passes the fingertip position through a transfer function similar to that described in [4] to obtain a pleasing pointing dynamics, and translates to the coordinate system of the region.

3.1 Pointing Detection

Detecting the user's hand in the context of an ED-projector presents several challenges. For one the appearance of the user's hand changes drastically as it moves through the projection. This makes techniques based on color or appearance unusable. Similarly, techniques based on background subtraction often give unreliable results, as the projected image can completely overwhelm the inherent color of the moving surface. Hence, many of the techniques used in other gesture recognition systems [15] will not work with projected displays.

Even though the appearance of an object will change as it moves across a projected image, it will create a region of changed pixels that retains the basic shape of the moving object (see Figs. 4a and 4b). Therefore each video frame is subtracted from the frame before it, and noise is removed with computational morphology.

When people reach out to touch or point at something, they use a variety of hand shapes, but almost invariably there is a finger that extends further than the rest by some amount leading the way. We assume this fingertip provides the most accurate estimate of where a user is pointing.



Fig. 4. a) Camera view of an interaction with a bucket; b) image difference data; c) overlay of search region (square), button active area (circle), and the fingertip template shown at the pointing location.

We find fingertip-like shapes by convolving a fingertip template (Fig. 5) over the difference image using a matching function tuned so that the gray level in the template determines the desirability of finding a changed pixel at that location, and changed pixels in black regions of the template are penalized more severely than unchanged pixels in the white regions. If the template does not match well anywhere in the image, we assume the user is not pointing.

The fingertip template is deliberately kept short so that it will match fingertips that extend only slightly beyond their neighbors and will match fingertips within a wider range of angles. As a result the template often matches well at several points in the image. We resolve between these hypotheses by using the “fingertip” furthest from the user.

This approach supports a tracking rate between 5 and 30 frames/second on a 500 MHz workstation for a 320x240 image, depending on the size of the search region and the size of the fingertip template (as determined by the expected size of the user's hand).

3.2 Button Touch Detection

A button touch event is defined to occur when a fingertip is in the vicinity of a button, travels away from the user to a point within the button and then returns toward the user. Button touches are detected by examining the hand trajectory for several specific patterns that indicate this type of motion. If more than one button is present in a configuration, and there is any ambiguity about which has been touched, only the button furthest from the user is allowed to generate an event.

This algorithm works very well on interactions where the user is asked to touch one button at a time. Importantly, it also resists generating an event when the user's hand moves through a button on the way to or from another location. The algorithm can fail when the user touches several buttons without retracting their hand, or “flies” their finger around in the image before touching a button.

Notice that although our vision system was built to recognize the touching of a button, the image of a button

tends to elicit from the user a slightly different gesture, the pressing of the button. We will address the consequences of this difference later.

3.3 Calibration

For each surface where the user will interact with the display, the vision system requires knowledge about the location and size of the user's hand. While this information could be inferred dynamically from the shape and motion of the changed region, to keep the implementation simple we obtain it during calibration by sizing and rotating a hand icon to match the image of the hand. This approach assumes that the user will approach the interaction area from a consistent location each time. In practice this assumption worked well for most surfaces, but failed often for others.

During calibration we also show the system the location and size of buttons and tracking regions. A search region is identified for each surface in order to help the system ignore extraneous movement in the image and speed response (see Fig. 4c).

The optimal recognition behavior varies from surface to surface, depending on the requirements of the task. For example on the selection screen (Figure 2b) false positive button presses are very disruptive to the demo, while false negatives, where the user must press again, are less disruptive. Conversely on the buckets of M&Ms the false negative rate is “naturally” higher because the user's pointing behavior is less consistent. Here false positives are less disruptive to the demo, while false negative results make the problem worse.

Several parameters can be used to adjust the inherent tradeoffs in the recognition system to match these desired characteristics. In the case of the selection screen we have

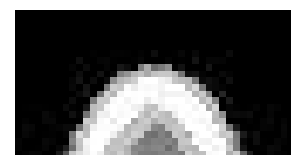


Fig. 5: Fingertip template.

the system emphasize a good fingertip template match, which has the effect of lowering the false positive rate, but increasing the false negative rate. For the buckets, we place less emphasis on a good template match, as users tend to point at the can buttons from a wide range of orientations. We also ensure the search region extends well above the button so that when the user reaches into the can for M&Ms the trajectory is more likely to fly through the button and not stop inside.

All calibration information for a surface is saved in a configuration file that is loaded each time the display switches to that surface.

4. Experimental Results

During the 6 days of SIGGRAPH '01 more than 650 people went through our demo and 4 complete M&M pictures were composed. The camera video stream was recorded onto tape and the users' selections were logged.

One objective of the SIGGRAPH experiment was to use the large number of subjects to help determine the main issues associated with the usage of projected displays, which could then be explored more deeply in controlled studies. Since most of our users were confronting a novel interaction paradigm, we were not shy of giving them instruction as needed.

A problem that we detected early on was the 2 to 9 second delay required to move the ED-projector's mirror and adjust focus between surfaces. Even the fastest transitions seem to be perceived as too long by the users of the system. The problem seems to be exacerbated by the lack of familiarity with the application, and by not knowing on which surface of the environment the display would appear next.

To minimize the perceived delay, the interaction was changed so that the user always received some information about what to expect next before the time-consuming mirror movement. For instance, after the user touched the "Done" button on a bucket (see Fig. 2c), the instruction "Go to the table" was immediately projected. Although this helped to alleviate some of the anxiety, confusion and frustration we observed before the change, the delays were still uncomfortably long.

Considering that the vision system and the projection system were integrated for the first time only in the weeks preceding the exhibition, the combined system worked remarkably well. A sample with 130 consecutive users with 621 button touch events (touching gestures or false detections) yielded correct detection of touching gestures up to 81%, with 12% of false negatives and 7% of false positives, as shown in table 1. If the buckets are excluded from the count, the performance exceeds 89%.

The buckets yielded a high number of errors for several reasons. The biggest problem was that while picking M&M's from the bucket, the user often partially or completely occluded the display with their head or back.

	Success	false positives	false negatives
Entrance	88%	9%	3%
select board	94%	2%	4%
buckets	55%	33%	12%
board (place)	91%	8%	2%
board (paint)	80%	2%	18%
total w/o buckets	89%	6%	5%
total	81%	12%	7%

Table 1: Performance of the vision system (130 users).

The user's motion in such situations sometimes triggered false positives for the button projected on the bucket.

Some false negative errors were due to the fact that the system was tuned to detect a straight, forward, out-and-back touch motion and the user sometimes assumed otherwise. For example, when picking a color, some users would fly their hand over several buttons, then press one with a vertical motion that was not readily visible to the camera. Others tried to wave their hands through the buttons or made similar non-touching actions. The later problem seemed particularly acute on non-standard interaction surfaces like the buckets.

The assumption that a user will not interact with an occluded button held true for most interactions, but there was one instance when it was repeatedly violated. After picking out their M&Ms from a bucket, a user would often begin to walk toward the table, then realize that they had not touched the "Done" button. Returning to the bucket they would occlude the projection (and camera) with their head or back just before they arrived, but they still continued to reach out and touch the now dark bucket.

Finally, we credit the high number of false negatives during finger painting to the reduced frame rate when simultaneously tracking over a large area and detecting touch events. Occasionally the brief interval when the finger was inside the button would fall between video frames.

Either when trying to click a button on the hard surface of the Plexiglas board or the soft foam board, our users almost always applied pressure to the projected "buttons." When a button touch was not detected by the vision system, the most common reaction was to try to press again with more pressure. Other common reactions were to use the whole hand instead of a finger, and to change the touch motion to a wave or some other action.

To reduce the excessive pressure users would often apply to the surfaces, we changed the wording of the instruction from "Press to begin" to "Touch to begin" (Fig. 7a). Although this helped somewhat, it seems that people expect buttons, even projected buttons, to react to pressure.

Users also seemed to react differently to the various surfaces on which the display was projected. Our users

looked more tentative interacting with the fabric-covered soft board and, especially, the buckets than with the painter's palette in the entrance and the Plexiglas board on the table. We often had to tell users to "Click the bucket" after picking out the M&Ms, while interaction with the more traditional surfaces came naturally. Are buckets expected to be less interactive than tables? Our experience at SIGGRAPH suggests an interesting hypothesis: people seem to attribute different interactive capabilities to surfaces transformed into touch-screens according to the nature of the surfaces themselves. If true, this observation may have implications in a variety of areas, e.g. [3]

5. Conclusions

The ED-projector introduces a new concept: projecting an interactive display onto arbitrary surfaces in the environment. This technology can be used as a generic input/output device that can replace, in many situations, current displays and interactive devices. Instead of being limited to a fixed display, an application can now move to wherever it is needed in the environment, such as a phone book by the phone, or a database of papers by the file cabinet. Computer and information access can be provided in spaces where traditional displays can be broken or stolen. An interactive display can also be brought to the proximity of a user without requiring the user to move. In particular, the ED-projector can facilitate the access and use of computers by people with locomotive disabilities. For instance, it can project an interactive display on a hospital bed sheet without patient contact with any device.

Projected displays also enable a new set of applications where a computer acts on the physical world, almost like a robotic arm made of light. These applications can use the ED-projector to point to physical objects, show connections among them, and project patterns to indicate movement or change in the real world. The experiment described here is an example of this class of applications. We also see the Everywhere Displays projector as a potential enabler of a new generation of games that happen not in the virtual world but are projected into the physical, everyday world where we live [1].

In this paper, the viability of the ED concept was demonstrated in an experiment where we guided several hundred first-time users through an assembly task. The experiment was too simple to enable us to draw definitive conclusions about how people react to a system that transforms everyday surfaces into touch-screens. However, it has alerted us about some limitations of the technology and has suggested guidelines for the design of such applications. For example, it is evident that the switching time between surfaces has to be either shortened or filled with some sort of user feedback.

The analysis of our errors in detecting user input points out many ways in which we can improve the performance of the vision system. It has become clear that,

like in many other computer vision systems, domain knowledge is essential to obtaining good performance. This might include knowledge of where the user is located during the interaction, what types of activities she is likely to perform outside the context of the interaction, and what types of errors are more or less disruptive to the interaction.

Some of the most interesting research concerns what interaction mechanisms are best suited for a display projected on an arbitrary surface. Is a "button" the right metaphor for interaction? As we demonstrated, we can detect button-pressing actions with reasonable accuracy, even with a single camera, but is this the right paradigm from the user's point of view for a widget that detects the hand position, not pressure? How can more flexible and expressive hand gestures play a role in these interfaces? Should other body gestures and natural actions by the user also play a role?

References

1. Björk, S., et al. Pirates! - Using the Physical World as a Game Board. In Proc. of Interact'01. 2001. Tokyo, Japan.
2. Crowley, J.L., et al., Things that See. Communications of the ACM, 2000. 43(3): p. 54-64.
3. Ishii, H. and B. Ullmer. Tangible Bits: Towards Seamless Interfaces between People, Bits, and Atoms. In Proc. of CHI'97. 1997. Atlanta, Georgia.
4. Kjeldsen, R., "Head Gestures for Computer Control", in the Proceedings of the Workshop on Recognition And Tracking of Face and Gesture - Real Time Systems (RATFG-RTS), Vancouver, BC, Canada, July 2001
5. Kjeldsen, R. and Hartman, J. Design Issues for Vision-based Computer Interaction Systems. In Proc. of the Workshop on Perceptual User Interfaces. 2001. Orlando, Florida, USA.
6. Keays, B. and R. Macneil. metaField Maze. In Proc. of SIGGRAPH'99. 1999. Los Angeles, California.
7. Krueger, M.W., Artificial Reality II. 1990: Addison-Wesley.
8. Morishima, S., et al. HyperMask: Talking Head Projected Onto Real Objects. In Proc. of Multimedia Modeling (MMM'00). 2000: World Scientific.
9. Pinhanez, C. The Everywhere Displays Projector: A Device to Create Ubiquitous Graphical Interfaces. In Proc. of Ubicomp'01. 2001. Atlanta, Georgia.
10. Rekimoto, J. A Multiple Device Approach for Supporting Whiteboard-based Interactions. In Proc. of CHI'98. 1998. Los Angeles, CA.
11. Underkoffler, J., et al. Emancipated Pixels: Real-World Graphics in the Luminous Room. In Proc. of SIGGRAPH'99. 1999. Los Angeles, CA.
12. Weiser, M., The Computer for the Twenty-First Century. Scientific American, 1991: p. 94-100.
13. Welch, G., et al., Projected Imagery in Your "Office of the Future". IEEE Computer Graphics and Applications, 2000(July/August): p. 62-67.
14. Wellner, P., Interacting with Paper on the DigitalDesk. Communications of the ACM, 1993. 36(7).
15. Wu, Y. and T. Huang, Vision-Based Gesture Recognition: A Review. Lecture Notes in Artificial Intelligence, 1999. 1739.
16. Yang, R. and G. Welch. Automatic and Continuous Projector Display Surface Calibration Using Every-Day Imagery. In Proc. of 9th International Conf. in Central Europe in Computer Graphics, Visualization, and Computer Vision. 2001. Plzen, Czech Republic.

Appendix A-3



Fifth IEEE International Conference on Automatic Face and
Gesture Recognition : proceedings : 20-21 May, 2002, Washington,

[Show more](#) ▾

Authors: [IEEE International Conference on Automatic Face & Gesture
Recognition](#), [IEEE Computer Society Technical Committee on Pattern Analysis](#)

[Show more](#) ▾

 Print Book, English, ©2002

Edition: [View all formats and editions](#)

Publisher: IEEE Computer Society, Los Alamitos, Calif., ©2002

Genre:	Conference papers and proceedings
--------	--------------------------------------

Physical Description:	xi, 436 pages : illustrations ; 28 cm + 1 CD-ROM
-----------------------	--

ISBN:	9780769516028, 0769516025
-------	------------------------------

OCLC Number / Unique Identifier:	49937766
----------------------------------	----------

Subjects:

[Automatische](#)
[Identifikation](#)
[Conference papers and](#)
[proceedings](#)
[Face perception](#)
[Automation Congresses](#)
[Gesicht](#)
[Gestes Congrès](#)
[Gestik](#)
[Gesture](#)
[Gesture Congresses](#)
[Human-computer](#)
[interaction](#)
[Human-computer](#)
[interaction Congresses](#)
[Image processing Digital](#)
[techniques](#)
[Image processing Digital](#)
[techniques Congresses](#)
[Kongress](#)
[Mustererkennung](#)
[Optical pattern](#)
[recognition](#)
[Optical pattern](#)
[recognition Congresses](#)
[Perception des visages](#)
[Automatisation Congrès](#)
[Person](#)
[Reconnaissance optique](#)
[des formes](#)
[\(Informatique\) Congrès](#)
[Traitement d'images](#)
[Techniques numériques](#)
[Congrès](#)
[Washington \(DC, 2002\)](#)
[Show less ^](#)

 3 editions in 53 libraries

Featured libraries

All libraries

Showing 37 libraries near [Missouri, United States](#)

☐ Only public libraries



Special or
Corporate Library

Linda Hall Library



[166 miles](#) from your current location.
5109 Cherry Street, Kansas City, MO, 64110,
United States

Borrow 



Favorite Library



Academic

University of Arkansas - Fayetteville

University Libraries



[184 miles](#) from your current location.
University Libraries, 365 N. McIlroy Ave., Fayetteville,
AR, 72701, United States

Borrow 



Favorite
Library

Appendix A-4

Screen capture made by J L Mullins from the Linda Hall Library OPAC record for Kjeldsen
(*Fifth IEEE International Conference on Automatic Face Gesture Recognition, Proceedings*)
on June 11, 2023



CONFERENCE PROCEEDING

[Fifth IEEE International Conference on Automatic Face and Gesture Recognition : proceedings : 20-21 May, 2002, Washington, D.C. / sponsored by IEEE Computer Society TC PAMI](#)
IEEE International Conference on Automatic Face & Gesture Recognition (5th : 2002 : Washington, D.C.); IEEE Computer Society. Technical Committee on Pattern Analysis and Machine Intelligence.
c2002
Available at Linda Hall Library Books - 2nd Floor (TK7895 .S65 I5 2002) and other locations >

Send to

Email

QR

Permalink

Citation

EasyBib

Export RIS

Export to Excel

Get It

LOCATIONS

Linda Hall Library
Available, Books - 2nd Floor ; TK7895 .S65 I5 2002
Go to Location Items >

Linda Hall Library
Available, Closed Stacks - Books Digital Media ; CDRM 496
Go to Location Items >

Title

Fifth IEEE International Conference on Automatic Face and Gesture Recognition : proceedings : 20-21 May, 2002, Washington, D.C. / sponsored by IEEE Computer Society TC PAMI

Other title

[Fifth International Conference on Automatic Face and Gesture Recognition](#)
[Automatic face and gesture recognition](#)

Creator/Contributor

[IEEE International Conference on Automatic Face & Gesture Recognition \(5th : 2002 : Washington, D.C.\)](#) >
[IEEE Computer Society. Technical Committee on Pattern Analysis and Machine Intelligence.](#) >

Subjects

[Optical pattern recognition -- Congresses](#) >
[Face perception -- Automation -- Congresses](#) >
[Gesture -- Congresses](#) >
[Image processing -- Digital techniques -- Congresses](#) >
[Human-computer interaction -- Congresses](#) >

Identifier

ISBN : 9780769516028
ISBN : 0769516025
OCLC : (OCoLC)49937766

Publisher

Los Alamitos, California : IEEE Computer Society

Creation date

c2002

Format

xi, 436 p. : ill. ; 28 cm. + 1 CD-ROM.

General notes

"IEEE Computer Society Order Number PR01602"--verso of T.p.
Accompanying CD-ROM cataloged separately as CDRM 496.
Includes bibliographic references and author index.

Amazon Ex. 1008.0107

Appendix A-5

A Development of Interactive Game "Ting Ting" Using Real and Virtual Objects*

Si-Jung Kim, Moon-Sung Jang, Hong-Sic Kim
R&D Center
Korea Visuals Corporation
107-7 Samsung-dong, Gangnam-gu, Seoul, Korea
hisijung@korea.com

Tae-Yong, Kuc
Intelligent Control & Dynamic Simulation Lab.
Sungkyunkwan University
Chunchun-dong, Suwon, Kyungki-do, Korea
tykuc@yurim.skku.ac.kr

Abstract - This paper presents an interactive virtual reality game using real and virtual objects with general beam projector and computer vision. We show how the computer recognizes the objects and makes the virtual objects move after detecting the real objects in the game space created by a beam projector. To achieve the goals, we developed two methods: Find Objects Distance (FOD) method and Find Reflection Angle (FRA) method. After detecting the real and virtual objects, the methods predict the next moving direction of virtual objects. This interactive virtual reality (VR) game is named as "Ting Ting" which contains various fire-shaped virtual objects and bamboo-shaped real objects. "Ting Ting" game's core technology can be used in various application area as an interactive communication engine. These include but not limited to, for example, an education game in schools, or an entertainment application in museums and/or science halls.

Keywords: interaction, virtual reality, game, collision, vision, HCI, projector

1 Introduction

Over the decades, various kinds of devices such as computer vision, projector and sensor were developed by scientists and according to the development of each device, other researchers and engineers are developing new interface to play a game or study on the Internet through the e-learning system. More recently, digital electronic devices controlled by a computer have been developed in the interactive application area such as interactive game and bilateral e-learning system. For instance, the aggregation of a projector, a digital camcorder, a digital camera and sound devices with a

computer provides us a new level of interactivity which is natural and human friendly. In developing interactive systems, "Videodesk" and "Digitaldesk" are used not only for a desk but also for the integrated computer system that adopts digital electronic devices such as touch screen and camera with a computer. In another research of natural interactive interface, Kenji and Jun Rekimoto worked for the content of computer with 3D hand interaction and full body interaction [2][3][13]. As for the input interface, a "Laser Shot system" was developed by ourselves. "The Laser Shot System" is a non-touchable multiple input device which was developed by using general laser pointers. This input device was used as remote interaction device for indication and selection of target object in a virtual aquarium. In the virtual aquarium, this input device was useful in selecting the menus, expanding the information and controlling the program from a long distance replacing the keyboard and mouse of computer. The laser pointer as an input device enables a user to obtain information from a long distance. In this research, the input device has been developed and used to solve the real problem and make more contents easily operate [5][8][9][12][14]. But one of important problems in implementing the user interface is how to track objects involved in the interaction. This means that one needs to solve the object recognition task in a limited time interval. The time limitation enters into conflict with the details and quality of the recognition.

This paper presents how Find Objects Distances (FOD) between real and virtual object and Find Reflection Angle (FRA) of the virtual objects can be found. Our algorithms calculate the distance between the virtual and real objects and determine a new path for each virtual objects. The interactive game "Ting Ting" was implemented to operate on a large-sized floor by using the FOD and the FRA with a general beam projector and vision.

* 0-7803-8566-7/04/\$20.00 © 2004 IEEE.

2 Problem description and elimination

To interact with real and virtual objects in the real environment, several problems must first be tackled. These include the display problem, to display the scene fully with avoiding obstacles; fast real objects recognition problem without image distortion from the computer vision system; realistic motion generation problem like acting in real world. All of these problems have input and output condition and require internal function to perform exactly. The considered problems are listed into five categories from 2.1 to 2.5. However we deal only with three problems under 2.3, 2.4 and 2.5 in this paper.

2.1 Game Space Configuration

In interactive contents, the setup aspect becomes a challenge as there is a need to determine the game space in which the real and the virtual object are displayed and meet each other through an interaction device such as a computer vision system. The game space may consist of various elements as shown in figure 1: front type, front rear type, bottom type and bottom rear type. Many researchers have researched how to display the scene fully under the circumstances of perturbing obstacles. This problem has been researched by Rick Kjeldsen who used mirror or mechanical frames named "Everywhere Displays Projector" [4][7][10]. In this paper, we adopt figure 1, (c) as the game space and we premise that there are no obstacles which prevents the displaying in front of the screen. Also in this configuration, the projector is set up vertically as a displayer and along with a camera that is also set vertically.

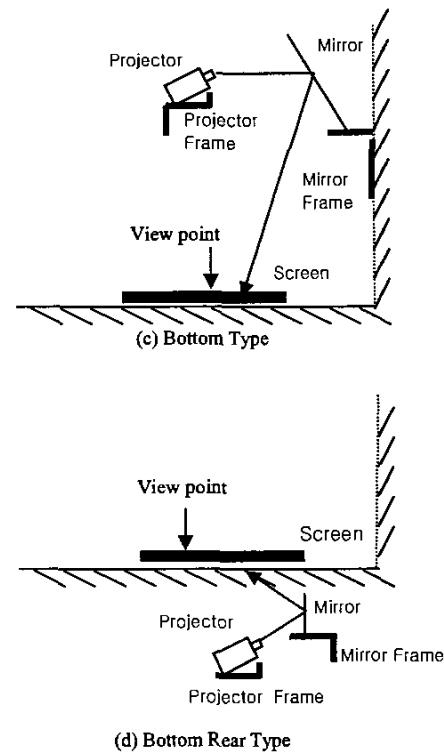
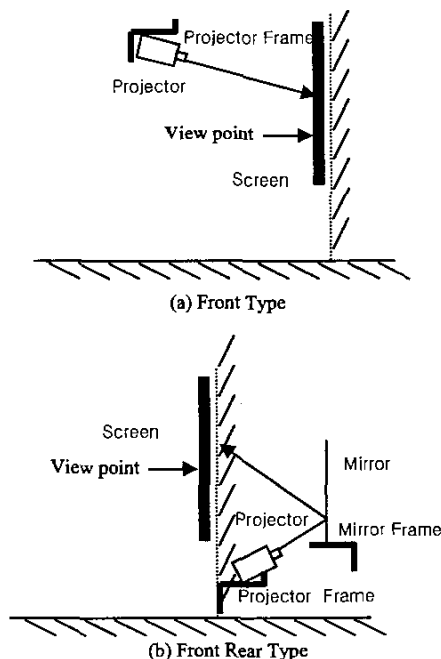


Figure 1. Four Types of Game Space Configuration

2.2 Space Recognition Problem

After the setup of the physical environment, one more set up problem remains. We must obtain the whole scene to find the real and the virtual objects in the game space. It is difficult for a computer vision to capture the whole scene without any distortion as the projector and camera are not set up in an exact parallel along with the screen. To settle the display exactly on the game space inputted by the computer vision, first we select four points before starting the program. We set the four regions of interest areas as shown in figure 2.: top left, top right, bottom left, and bottom right. Using this step, the computer can know the interest area of the game space exactly and calculate the object's movement and process when occurred some events [9].

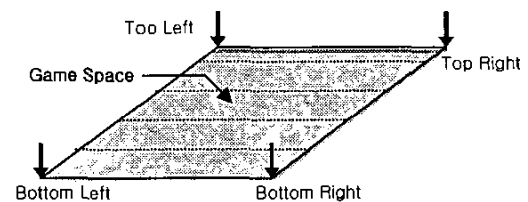


Figure 2. Define the Interest Region

2.3 Real and Virtual Objects Detection

In an interactive related game, real or virtual objects move according to the moving rule. For instance, a virtual environment, like a walkthrough, creates a computer-generated world filled with virtual objects. These objects should not pass through each object, and things should move as expected when pushed, pulled or grasped. As well known, these actions are the output of collision between two objects as shown below in figure 3. In this point we concentrate more on the real time response than on accurate detection. Collision detection has played a fundamental and important role in computer animation, physically based modeling, game, and simulation. In these applications, interactions between moving objects are modeled by dynamic constraints and contact analysis [8]. The objects' motions are constrained by various interactions, including collisions. In this paper, to achieve this we detected both the real and the virtual objects' position from the computer vision, and calculated each object's length while comparing both the real and the virtual object's color.

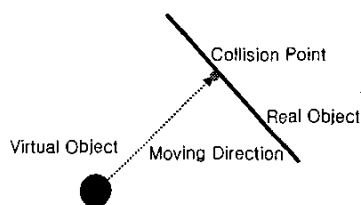


Figure 3. Collision Detection

2.4 Finding Objects Distance (FOD)

Calculating the distance between the real and the virtual object is for the usage of deciding the collision stage or approaching stage of the real and the virtual objects. If there is a real object in the direction of the virtual object, we obtain the position of the two objects and compare it. To calculate the distance and compare between the real object and virtual object we use the Find Object Distance (FOD) method, further described in section 3.

2.5 Finding Reflection Angle (FRA)

When we play recent games, we feel that most of the objects played behave like objects in a real world in terms of movement and rotation as most games are developed under physical rules. Like in a game, we follow the physical rules when the virtual object is reflected by the real object. More specifically, we adapt the law of reflection as shown in figure 4 based on the laws of reflection in physics. In the law of reflection, if we know the incoming angle, we will know the outgoing angle as they are the same.

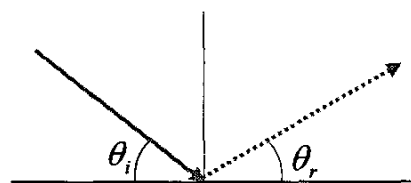


Figure 4. The Law of Reflection

The angle of reflection θ_r of a ray or beam is the angle measured from the reflected ray to the normal surface. From the law of reflection, $\theta_i = \theta_r$, where θ_i is the angle of incidence. θ_r is the measure between the ray and the normal of the surface that intersects the surface at the same point as the ray. Section 3 provides further description of the Finding Reflection Angle (FRA) [15].

3 Problem Solving

The following section consists of solving the problems mentioned on section 2. The specific problems are: finding two real and virtual objects in the game space, calculating the distance between the real and the virtual object and finding the reflection angle when the virtual object is reflected by the real object. To solve these problems, our approach is to use the simple and intrinsic method of geographical and geometry method rather than to adapt a complicated mathematical method.

3.1 Detection of Real Objects

As we know, each moving object has its own information such as color, motion vector and speed as shown in figure 5. Generally, there are many detection methods such as in facial recognition and fingerprint recognition. However these methods are highly complex so that using them makes it difficult to formulate fast real-time applications in interactive response games. A characteristic in detecting the real object is that the real object moves and has its own color. But we also know the virtual object's position and moving angle. These are key factors to escape from the complexity of the problem we have nominated as the method "from the point of view of the virtual object". This method compares some regular pixels according to the direction of the virtual object to predefine the color table (CT) of the real object. CT was saved before starting the program by computer vision. Through these steps we can detect the real object. And we can know where the real object is on the game space [1][6] [11].

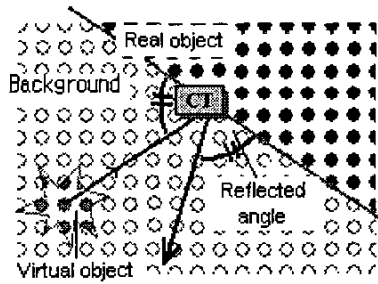


Figure 5. Real and Virtual Object Detection

3.2 Find Object Distance (FOD)

Find Object Distance method is needed in order to find out the distance between the real and the virtual objects as shown in figure 6. In figure 6, P_0 is a center point of the virtual object, P_c is the first collision point of the real object on the line drawn from point P_0 in given direction of the virtual object's movement. Also we define the β as the moving direction of the virtual object and l_{xpixel} as the length of each pixel of the game space. As we know in a bit-mapped image described by a matrix, conjunctions are described by ordered pair of x and y location. We described the real object and the virtual object's position as P_0 and P_c on a bitmap plane individually in figure 6.

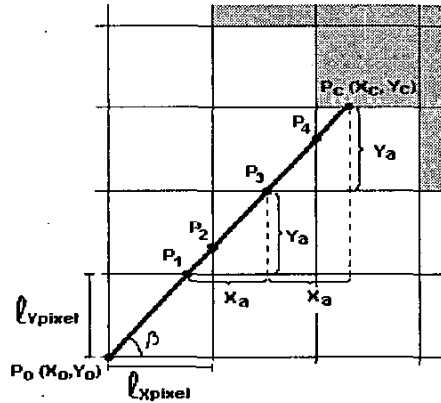


Figure 6. Distance Calculation Between Objects

First of all find the square cell where the imaginary line from point P_0 in given direction β intersects the real object's center point P_c . To check whether this line has intersected with a square cell or not, it is sufficient to check the grid intersection points at P_1, P_2, P_3, P_4 and the real object P_c . To do this, we need to check any grid intersection points that are encountered by the line. And then see if there is a real object color behind the grid or not. When there is a real object color behind either on a vertical or on a horizontal intersection, the checking process stops. The distance between both intersection

points is compared and we choose the nearest distance. In brief, the method of finding intersections with horizontal grid lines can be described as in the following steps below.

Step1: Find Y_a , the y size of the image pixel. If the line is turning up, Y_a will be negative.

Step2: Find X_a using the equation 1, where X_a is the length of x size of the pixel.

$$X_a = \frac{l_{ypixel}}{\tan(\beta)} \quad (1)$$

Step3: Check the grid at the intersection point. If there is a real object color behind the grid, stop and calculate the distance according the equation 1.

Step4: If there is no real object color, extend to the next intersection point. Set up the coordinates of the next intersection point to be X_{new} and Y_{new} , where X_{new} is calculated by X_{old} plus X_a , and Y_{new} is calculated by Y_{old} plus Y_a .

When this calculation comes to an end the coordinate of the P_c point will be approximately the same as (X_{new}, Y_{new}) . We can calculate the distance between them like absolute value of $X_{new} - X_0 / \cos(\beta)$. Finally, come back from the grid to the pixel map to get the distance in pixel's size. To take the real size in pixels we must divide this calculated value on l_{xpixel} .

3.3 Find Angle Reflection (FRA)

After finding the distance of the two objects, the next potential problem is to find out the virtual objects' reflection direction as shown in figure 7. In figure 7, line ED is the length of the real object, B is the virtual object's position and β is its moving direction. The line BC is the imaginary moving path of virtual object B and line CR is the reflected path after collision with the real object. Also α and d_a is the constant factor appeared by drawing the line BD to find the reflection angle $\angle OCR$.

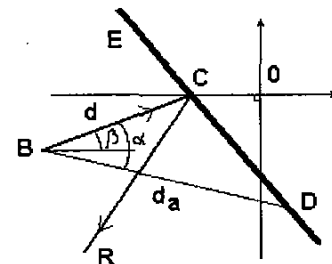


Figure 7: Find Reflection Angle

We must find the direction of the CR line which is the reflected direction of the BC line. We calculate the unknown direction of the CR line as defined by equation 2.

$$\angle OCR = \beta + 2\angle BCD \quad (2)$$

To find out the angle $\angle BCD$ in equation 2, we draw the additional line BD from point B to D on the line ED, where angel α is given. Then we calculate the length of the new line BD using the same method of FOD described in the above section. Let the length of the line CB and BD be d and d_a respectively. Now we can determine $\angle BCD$ defined by two given sides BC, BD and angle α between them. We show the result in equation 3.

$$\angle BCD = 90^\circ - \alpha + \arctan\left(\frac{d_a}{d \cdot \sin(\alpha)} - \frac{1}{\tan(\alpha)}\right) \quad (3)$$

The method of finding the reflection angle is briefly described below.

Step1: Draw the imaginary line BD to the line ED according to the optional angle α , where α is a constant and given out temporally.

Step2: Find the length of line BD according the Find Object Distance (FOD) method.

Step3: Draw the per-pendicular line from point C to line BD and calculate its length. Calculate the two angle caused by the per-pendicular line using the parameters d , d_a and α and trigonometry.

4 Implementation

The interactive game "Ting Ting" which is played by the real and the virtual object was implemented under IBM compatible Pentium4 PC with computer vision and beam projector. We used Visual Studio dot Net of Microsoft with Direct Draw to calculate the motion information and draw the scene. Microrobot's vision board (30 frame/sec) was used to acquire real and virtual objects from the game space. We adapted the 30 frames over processing vision board that cover the human motion to make real time processing into image processing. Last, we also used Sharp beam projector to represent virtual objects, generated by the PC.

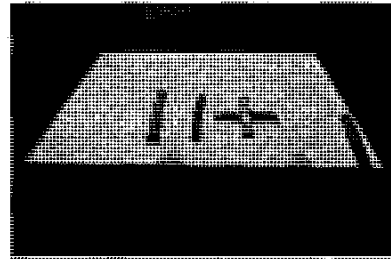
4.1 System configuration

On the physical game space, we made the five meter sides square environment according to the general hardware configuration in figure 1 (c). The game spaces consist of four main parts: the first one is the beam

projector that displays the game scene; the second part is the computer vision in order to detect the real and virtual objects on the game space; the third is the computer part to control the overall device and to analyze the images, and the last part is the real object to play the interactive game. We showed the setted game space in figure 8.



(a) Beam Projector and Camera Setup



(b) Game Space Setup (Real Environment)

Figure 8: System Configuration

4.2 System block

The interactive game "Ting Ting" needs mainly six functions: 1) capture module for capturing the real object, 2) real object finding module for calculating the real object's position, 3) virtual object finding module for loading current virtual object's status, 4) FOD for determining the real and the virtual object collision or not, 5) FRA for calculating the next moving direction, and 6) Renderer for traversing the scene graph. We showed the six main functions in figure 9.

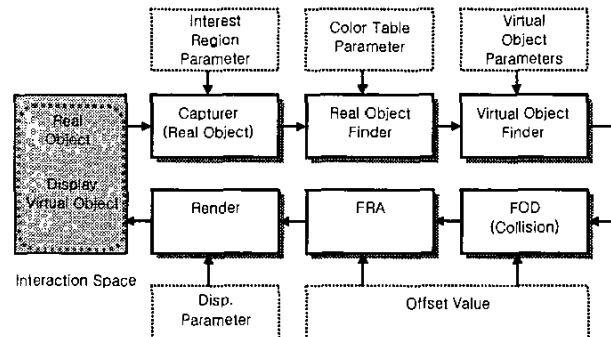
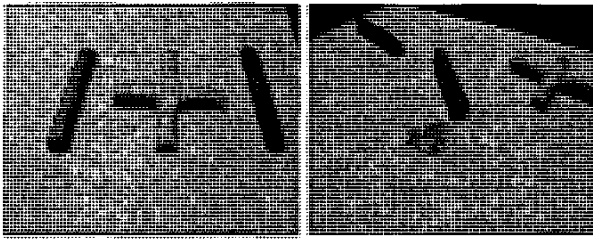


Figure 9: Functional Block Diagram

4.3 Game configuration

"Ting Ting" game is about capturing the virtual object by the real objects. The virtual objects are drawn on the surface of the game space by the projector and the real objects are used by users at the real environment. Real objects types are that of two bamboo stick shaped images and virtual object consists of several one-type fire-shaped image as shown in figure 10.



(a)Implemented Real Object (b)Generated Virtual Object

Figure 10. Real and Virtual Object

5 Conclusion

The Find Object Distance (FOD) method and Find Reflection Angle (FRA) method are implemented completely. The interactive game "Ting Ting" was developed using two electronic devices, the beam projector and computer vision system with personal computer. We have tested this system using two virtual objects, which is fire-shaped and a real objects, which are two bamboo shaped. Under computer control, the virtual objects were moving around and changed their moving direction after collision with the real object. We showed this in figure 11. The interaction system based on the general beam projector and computer vision and controlled under by a personal computer combines the best simplicity: the natural interface and the power of computational resources. The resulted interactive interface is easy to use so that this can be used by anyone with minimal training. However, potential problems may appear as that of when we recognized the real objects in the dark level light circumstances.

In this paper, we presented the real and the virtual object communication interface including two related processing methods: the FOD and the FRA. The FOD and FRA modules can be useful to make a variety of interactive applications including entertainment and training related contents with fast result.

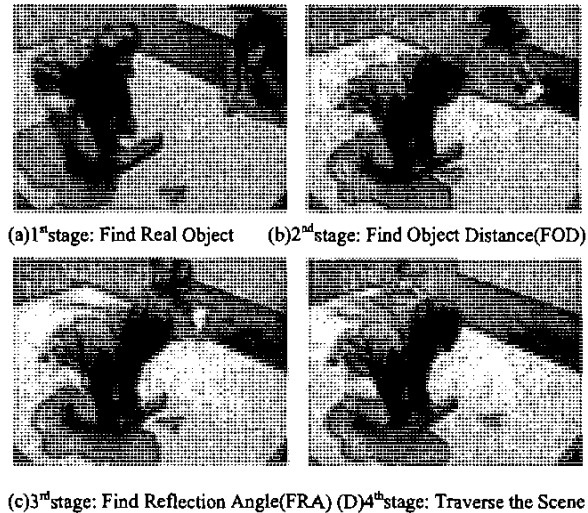


Figure 11. Snapshots

References

- [1] Hand D. J., Discrimination and Classification. Wiley and Sons, New York, 1981.
- [2] Jay Summet, Mathew Flagg, James M.Rehg, Gregory M. Corso, Gregory D. Abowd, "Increasing the Usability of Virtual Rear Projection Displays".
- [3] Jun Rekimoto, "SmartSkin: An Infrastructure for Freehand Manipulation on Interactive Surfaces", CHI 2002.
- [4] Kenji Oka, Yoichi Sato, Hideki Koike, "Real-time Tracking of Multiple Fingertips and Gesture Recognition for Augmented Desk Interface Systems".
- [5] Krueger M.W., Artificial Reality II. Addison-Wesley, 1991.
- [6] Nadler M. and Smith E. P., Pattern Recognition Engineering, Wiley and Sons, New York, 1993.
- [7] Rahul Sukthankar, Robert G.Stockton and Matthew D. Mullin, "Smarter Presentations: Exploiting Homography in Camera-Projector Systems", Proceedings of International Conference on Computer Vision, 2001.
- [8] Rauterberg M., Steiger P., "Pattern recognition as a key technology for the next generation of user interfaces", In Proc. of SMC'96, Volume 4, IEEE Catalog Number: 96CH35929, pp. 2805-2810, 1996.

- [9] Rehg J.M. and Kanade T., "Visual tracking of high dof articulated structures", Third European Conf. On Computer Vision, Volume 2, Stockholm, Sweden, Springer-Verlag, pp. 35-46, May 1994.
- [10] Rick Kjeldsen, Claudio Pinhanez, Gopal Pingali, Jacob Hartman, Tony Levasand Mark Podlaseck, "Interacting with Steerable Projected Displays", Proc. Of 5th International Conference on Automatic Face and Gesture Recognition(FG'02). 2002.
- [11] Schettini R., "A general-purpose procedure for complex graphic symbol recognition. Cybernetics and Systems", Volume 27, pp. 353-365, 1996.
- [12] Si-Jung Kim, Moon-Sung Jang, Tae-Yong Kuc, "An Interactive User Interface for Computer-Based Education: The "Laser Shot System", ED-MEDIA 2004
- [13] Wellner P., "Interacting with paper on the DigitalDesk", In Com. of the ACM. 36(7), pp. 86-96, 1993.
- [14] Wren C., Azarbayejany A., Darrel T. and Pentland A., Pfinder "Real-Time Tracking of the Human Body", IEEE Trans. On PAMI, 19 (7), pp. 780-785, 1997.
- [15] <http://scienceworld.wolfram.com>

Appendix B-1

IBM Research Report

Ubiquitous Interactive Graphics

**Claudio S. Pinhanez, Frederik C. Kjeldsen, Anthony Levas,
Gopal S. Pingali, Mark E. Podlaseck, Paul B. Chou**

IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598



Research Division

Almaden - Austin - Beijing - Delhi - Haifa - India - T. J. Watson - Tokyo - Zurich

LIMITED DISTRIBUTION NOTICE: This report has been submitted for publication outside of IBM and will probably be copyrighted if accepted for publication. It has been issued as a Research Report for early dissemination of its contents. In view of the transfer of copyright to the outside publisher, its distribution outside of IBM prior to publication should be limited to peer communications and specific requests. After outside publication, requests should be filled only by reprints or legally obtained copies of the article (e.g., payment of royalties). Copies may be requested from IBM T. J. Watson Research Center, P. O. Box 218, Yorktown Heights, NY 10598 USA (email: reports@us.ibm.com). Some reports are available on the internet at <http://domino.watson.ibm.com/library/CyberDig.nsf/home>.

Amazon Ex. 1008.0117

Ubiquitous Interactive Graphics

Claudio Pinhanez, Rick Kjeldsen, Gopal Pingali, Anthony Levas, Mark Podlaseck, Paul Chou

IBM Research, T.J. Watson

pinhanez@us.ibm.com

Abstract

This paper presents a device that creates ubiquitous interactive graphics in a real world environment without the need for pre-wiring surfaces or requiring users to carry or wear any special devices. Called the *Everywhere Displays projector*, or *ED-projector*, this device employs *steerability* as the key principle to achieve the promise of ubiquitous computing and augmented reality. The ED-projector uses an LCD projector with a steerable mirror to project graphics onto any pre-calibrated surface, while warping the projected image to correct for oblique distortion. Three different methods to determine this warping function are given. To sense user interaction with the projected image, the ED-projector employs a pan/tilt/zoom camera and a motion-based computer vision system to track the user's hand and touch-like gestures. We propose a system software structure that ties together steering, display, and sensing. Three implemented applications demonstrate the notion of ubiquitously "painting" the real world with interactive graphics, in the context of a futuristic office, an augmented assembly task, and a ubiquitous computer game. The observation of hundreds of users performing the assembly task suggests that the traditional desktop paradigm is inadequate for ubiquitous interaction and that new conceptual widgets and interaction paradigms must be developed.

Keywords: ubiquitous computing; augmented reality; gesture recognition; interactive projections; ubiquitous games.

1 INTRODUCTION

The extraordinary success of computer graphics technology in recent years has been responsible for the creation of amazing new worlds in movies, interactive computer games, and virtual reality (VR) environments. However, this very success poses the question of when the real world, where we live and work, is going to be as rich with imagery and information as movies and games today.

This paper describes technology that can be used to ubiquitously augment, correct, and enhance reality. Although the ultimate goal is to create interactive pixels anywhere in a space as seamlessly as it happens in movies (thanks to CG special effects), this particular work focus on how to "paint" interactive graphics on the surfaces of an environment without requiring any kind of fixed display devices, special wiring of surfaces or the use of encumbering headsets.

The system described here is based on a new device, the *Everywhere Displays projector* (*ED-projector*), composed of a steerable projection system and a pan/tilt camera. A prototype of the projection system has been built using a pan/tilt mirror that directs imagery generated by a standard LCD projector. To capture user interaction with the projected graphics, the images gathered by the camera are processed using computer vision techniques. In the current prototype, it is possible to determine the user's hand position in reference to a surface and to detect simple, touch-like

gestures, effectively mapping the user's hand gestures into mouse-like events.

For instance, we have developed a computer game, called "Fro...og!", where a computer generated image of a frog initially appears "sitting" on the side of a file cabinet (see Figure 1). If a player tries to catch it, the vision system detects the player's hand approaching and then shows an image of frog jumping (see Figure 1). This is followed by moving the image of the frog to another surface, for instance the top of a table by appropriately steering the mirror. All this happens in an environment under normal lighting conditions, ultimately creating the illusion that the frog jumps around the space to avoid being caught by the player.

Our goals, however, go beyond transforming the real world into a magical entertaining place. In particular, we envision the creation of seamless interfaces to devices on surface of walls and everyday objects such as tables, chairs, couches, doors, cabinets, and carpets. Currently, only very "special" surfaces such as TVs, monitors, and projected screens support the creation of computer graphics, and usually require external devices such as mice and joysticks to make them interactive. In our paradigm, any surface is regarded as generic support for interactive graphics. By projecting imagery and detecting user interaction with a camera, it is possible to transform every surface into a virtual touch-screen.

These ideas have been explored in three application prototypes

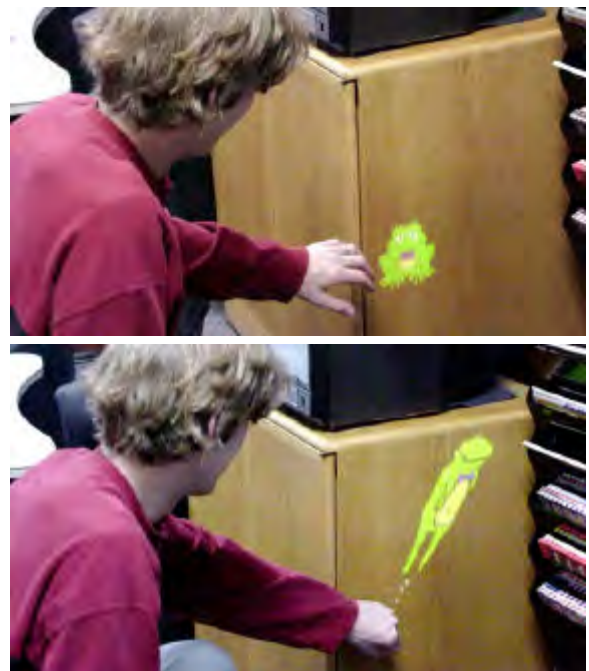


Figure 1. Trying to catch a projected frog.

described in the paper. The first is an attentive/interactive office space. The second is in the context of augmenting the environment of an assembly task by projecting interactive graphics directly onto the surfaces of the objects involved. Ultimately we aim to realize Mark Weiser's paradigm of ubiquitous computing (see [1]) without having to change the fabric of the world by embedding touch-screens on every surface. The third application is the ubiquitous computer game "Fro...og!".

This paper starts by comparing our approach and solutions with previous work, followed by a description of the technology used in the prototypes we have built. Although initial ideas on the ED-projector have been reported in a previous publication [2], this paper presents the device comprehensively including mathematical formulation, computer vision based interaction, and three implemented applications. Specifically, this paper describes the mathematical formulation of the method for correcting oblique distortion and presents two new calibration methods for the first time. Further, in this paper we describe an integrated, working computer vision system that we implemented to track hands and detect touches, as well as a way to structure the interface to application software. While [2] elaborates only on conceptual mock-ups, this paper describes three implemented applications using the ED-projector, followed by a discussion of how novice users — in one case, more than 600 of them — have experienced and interacted with the application prototype. We conclude by examining other applications that can take special advantage of the ED-projector and discuss the challenges these new interfaces and applications create for HCI research.

2 RELATED WORK

In his seminal paper that defined the field of ubiquitous computing, Weiser [1] envisioned computers embedded everywhere — ubiquitous and transparent. In his view, interaction with computers would be accomplished both through direct mechanisms linking sensors and actuators and by the widespread presence of cheap interactive screens (referred as "pads") embedded on the surfaces of walls, furniture, and objects. Since the publication of that paper, different devices and methods to implement the functionality envisioned for Weiser's pads have been proposed. In general, those devices and methods can be divided into four categories: head-mounted, portable, embedded, and projection-based.

A good survey of the current research on head-mounted devices and their uses in ubiquitous computing is provided by Barfield and Caudell [3], describing both wearable computers and the more traditional approach of off-body computer power. We see two main challenges in this approach. First, in order to really connect to the physical world, it is necessary to precisely register the position and attitude of the head of the user with the environment in real time. An examination of the state-of-art of the research in the area shows that this is currently achievable only in very constrained environments (see, for example, Raskar et al. [4]).

The second challenge with head-mounted methods for ubiquitous displays is that they do not work well in social contexts and collaborative tasks. If multiple users are looking at the same surface (for instance, a whiteboard), it is necessary to render exactly the same graphics on each user's head-mounted display simultaneously. This creates considerable difficulties in terms of accommodating different resolutions, rendering speeds, etc. of different brands and models of head-mounted displays. In practice, it is very hard to guarantee that all participants are seeing exactly the same images.

The second method being studied to create ubiquitous displays utilizes portable devices such as laptops, PDAs, and mobile

phones. Like head-mounted displays, portable devices do not work well in social and collaboration tasks (except in very unusual conditions such as in multi-player games [4]), and are a nuisance to carry and power.

The third method involves embedding screens or similar interaction devices into the objects themselves, as proposed initially by Weiser [1] and more recently by Ishii [5]. However, even if screens became extremely cheap, this approach requires a widespread change in the way everyday objects are manufactured and installed. Power and network access (or wireless bandwidth) has to be provided to even the simplest object in a home or workplace. In other words, this approach requires a drastic change of the very fabric of the real world.

Our approach follows the fourth method of creating ubiquitous displays that is based on projection devices. Interestingly, the use of projectors to change and augment reality was pioneered by media artists such as Tony Oursler [6] and Michael Naimark [7]. The use of projectors to access computers have been initially proposed by Bolt [8] and Wellner [9]. Recently, more radical applications of projectors have been proposed by Morishima et al. [10] that enhance the face of an actor, and by Raskar et al. [11] to change the color and texture of real objects such as vases and architectural models. Although our system does not currently track moving surfaces as in the former example, or project onto 3D objects as the latter, it can clearly incorporate both features in future versions.

Deviceless interaction with projections was pioneered by Krueger [12], using camera-based systems to detect the users' hand and body. The fact that most vision techniques put strict constraints on lighting conditions and on the background surfaces has restricted the use of these systems to entertainment applications such as Keays and McNeil's *metaField* [13]. To avoid lighting constraints, Omojola et al. [14], Ju [15], and the RED project at Xerox Parc [16], among others, embedded sensors on the surface where the images are projected. As in the embedded approach described before, the widespread adoption of sensor-loaded surfaces requires a dramatic change in the objects of the real world. Underkoffler and Ishii [17] and Rekimoto [18] used visual tags to detect the position and the identity of the objects on a surface and to overlay the interaction based on these attributes. To avoid the need of tagging all objects in an environment as in those two cases, a possible option is the use of object recognition techniques (such as those described by Schiele and Crowley [19]). However, those techniques currently work only in very constrained situations. We believe, though, that incorporating object recognition capabilities into our system is an important direction to follow, and we plan to do so for specific applications.

Vision-based gesture recognition has received a great deal of attention in recent years [20], but only a little of that work has attended to the problems of gesture recognition in the context of a projected display. Because the projected image drastically changes the appearance of the hand, most approaches to hand detection are based on either background subtraction or motion (frame-frame differencing). Our experience agrees with that of von-Hardenberg and Berard [21] in that background subtraction does not give reliable results in the presence of strong projected images.

Gesture-based interaction methods tend to fall into one of two categories: either variations of a point-and-click paradigm [21, 22] or application-specific pose or motion gestures [23]. Both these approaches have disadvantages. The point-and-click approach is often not well suited to hand pointing due to limited resolution and the lack of a natural "click". On the other hand, a proliferation of application-specific gestures will make for a complex gestural interface that a user will have to learn and execute largely without feedback, as discussed by Kjeldsen in [24].

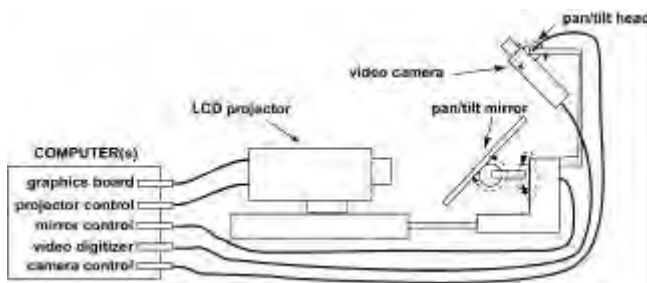


Figure 2. Diagram of the components of the ED-projector.



Figure 3. Prototype of the ED-projector: a) ED-projector without the camera mounted; b) a detail of a complete prototype, including the pan/tilt camera.

A major difference between our approach and the cited work on the use of projection-based devices to create ubiquitous interactive displays is *steerability*. By directing the projection and vision subsystems to a relevant position, it is possible to create new “interactive” surfaces in a space without any hardware installation, addition, or change. Unlike Raskar et al. [25], who use mirrors to increase the resolution of multiple-projector installations, we steer the image from a projector to realize interactive displays everywhere in the workspace. By doing so, we opportunistically use the surfaces of the objects present in the environment. We believe that steerability is a key feature that creates a real alternative to realize the promises of ubiquitous computing. However, it has constraints and technical challenges of its own, such as obstruction, oblique distortion, and surface texture interference. These issues are discussed in the next section where solutions to some of these problems are proposed.

3 THE ED-PROJECTOR

Our approach to ubiquitous interactive graphics uses a device — the *Everywhere Displays projector* (the *ED-projector*) — that can steer graphic displays onto any surface and can also sense user interaction with these steered graphical displays. The ED-projector is composed of an LCD projector, a computer-controlled pan/tilt mirror, and a pan/tilt/zoom camera. The projector is connected to the display output of a host computer that also controls the mirror and the camera (see Figure 2). The video output of the camera is sent to a digitizer board in the host computer where the imagery is processed, as described later.

Figure 3 shows a prototype of the ED-projector built with off-the-shelf components — a rotating mirror used in theatrical/discotheque lighting, a standard LCD projector, and a pan/tilt camera. The projector’s light can be directed anywhere

within a cone defined by approximately 60 degrees in the vertical axis and 230 degrees in the horizontal axis. The camera has a maximum vertical viewing range of 88 degrees and a horizontal range of 249 degrees, and is thus able to view the projected displays as they are steered around.

In general, a projector needs to be able to project a white pattern approximately 10 times brighter than its surroundings to create the illusion of contrast. Our current prototype employs a 3000 lumen projector, which we have found to be sufficiently bright in typical home and office lighting conditions to provide good contrast in the projected displays. Of course, the perceived brightness and contrast is heavily dependent on the color and material of the projected surface. Although our best results involve projection on white surfaces, we have also obtained good results projecting on carpets, black objects, and even translucent backlit surfaces. However, highly textured surfaces such as natural wood and surfaces with high reflectance are, in general, less suited for projected displays.

The three main issues with the ED-projector are 1) correction of the projected image to compensate for distortion due to oblique projection and surface characteristics; 2) sensing the user interactions with the projected displays; and 3) a system software structure that enables applications to easily define interactive display surfaces. These issues are discussed in detail in the rest of this section.

3.1 Correcting the Projected Image

Standard LCD projectors are designed to project light in a direction orthogonal to the projected surface. In the ED-projector, the pan/tilt mirror can deflect the light to multiple surfaces in the room, but most of the times the projected surface is oblique to the direction of projection. The result is a distorted image.

To correct the oblique projection distortion, we pre-warp the image to be projected to compensate for the distortion. It is always possible to correct the projected image, as long as the projected surface does not have a geometry that occludes the cone of projection (see [2] for details). However, the process of warping the image normally uses only a fraction of the projector’s image plane, causing loss of resolution. We will discuss this issue later; first, we address the problem of determining the warping function that corrects the oblique distortion.

Although there are many different ways to determine and implement the warping function, it is important to use a method that allows fast computation. We choose to implement it using 3D computer graphics hardware to texture map the images to be displayed onto a mesh in 3D space. If the mesh is appropriately warped and/or positioned in 3D space and the virtual camera correctly positioned, the projection of the resulting rendered image is free of distortion.

The basic structure of the problem, in the case of planar surfaces, is illustrated in Figure 4. We start by defining the *application plane*, a coordinate system where images to be projected are positioned within the area of a rectangular mesh M , and the *projector image plane*, the coordinate system associated with the LCD matrix of the projector. We also define a *projected surface* to be the real world surface on to which images are projected. In addition, we have the camera image plane, the coordinate system associated with the CCD matrix of the camera.

The goal is to define a transformation H from the application plane into the projector image plane so the transformation that maps the support mesh M into the projected surface is the identity I , and, therefore, appears free of distortion from an observer orthogonal to the projected surface. Since the process of projecting the image can be modeled as a projective transformation P (as

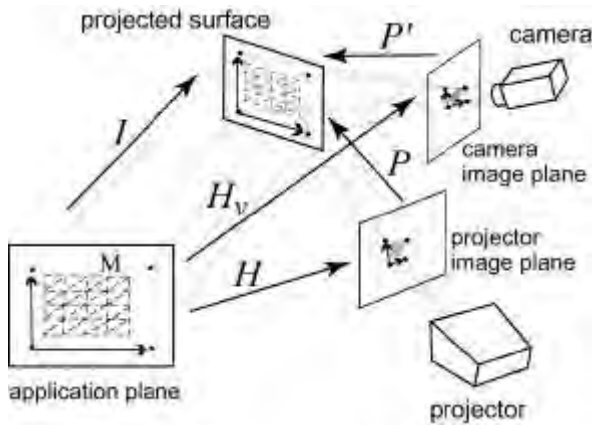


Figure 4. The transformations between the application plane, the projector's image plane, the camera's image plane and the projected surface.

observed in [26, 27]), we can represent the relationships among the planes as matrices in planar homogeneous coordinates, $I = PH$. Since H and P are projective transformations and therefore invertible, we obtain:

$$H = P^{-1}$$

Similarly if P' is the projective transformation between the camera image plane and the projected surface, and H_v is the transformation from the application plane to the camera image plane, we have $I = P'H_v$. Although there is no easy direct way to compute the projection function P , we developed three different methods to determine the actual value of H for a given surface. Of course, H and H_v are different for each planar surface in an environment.

1st method: Using an Equivalent 3D Virtual CG World

This method, first described in [2], is based on the fact that, geometrically, projection is the inverse of camera viewing. So, it is possible to model the projection by simulating the camera view by rotating, scaling, and viewing the support mesh M in 3D space.

Given the n mesh points in 2D projective space, $m_i = [m_i^x \ m_i^y \ 1]$, $1 \leq i \leq n$, let M be the matrix of all n points, $M = [m_1^T m_2^T \dots m_n^T]$. This mesh can be "positioned" in the XY plane of the 3D space by multiplying it by the matrix

$$K = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad KM = \begin{bmatrix} m_1^x & m_2^x & \dots & m_n^x \\ m_1^y & m_2^y & \dots & m_n^y \\ 0 & 0 & \dots & 0 \\ 1 & 1 & \dots & 1 \end{bmatrix}$$

We can then simulate the process of projecting the mesh as the inverse of the process of "viewing" the appropriately positioned mesh in 3D space. That is,

$$P^{-1}M = P_f R_x R_y R_z S_x S_y T_x T_y R_z K M$$

where R_x, R_y, R_z are rotation matrixes in 3D homogeneous coordinates, S_x, S_y are scale matrixes, T_x, T_y are translation matrixes, and P_f is the 3 4 projective matrix with focus point f . Since $H = P^{-1}$, we obtain

$$H = P_f R_x R_y R_z S_x S_y T_x T_y R_z K$$

Notice that all the matrices that compose H , with the exception of K , correspond to standard projection parameters of computer graphics boards. Therefore, to implement this method it is only necessary to determine these individual parameters, load them into

the hardware, and command the graphics board to render the scene.

To determine the projection parameters for a given surface we start by texture mapping a calibration pattern to the mesh M . Starting off with default values for the projection parameters that render the mesh on the center of the projector image plane without warping (i.e., $H \approx I$), we iteratively change each of the projection parameters until the projection of the pattern on the projected surface is observed to be identical to the original mesh. We then record these parameters and reload them whenever the ED-projector is aimed at that surface. Although conceptually simple, this iterative process of determining the right projection parameters is typically very time consuming.

2nd method: Manual Determination of Corresponding Points

This and the following method are based on the fact that projective transformations can be completely determined if the corresponding coordinates of four points in each 2D projective space are known (as shown by Faugeras [28]).

In this method, four points $a_i = [a_i^x \ a_i^y \ 1]$, $i = 1, 2, 3, 4$, are directly rendered on the projector image plane. We then iteratively move these points over the projector image space with a pointing device such as a mouse until their projections overlap the location of four known points on the projected surface, $b_i = [b_i^x \ b_i^y \ 1]$, $i = 1, 2, 3, 4$. For greater accuracy, we can choose these four points to be the corners of the mesh M .

It is easy to see that if the projected points are aligned with the projected surface points then $PA = B$, where $A = [a_1^T \ a_2^T \ a_3^T \ a_4^T]$ and $B = [b_1^T \ b_2^T \ b_3^T \ b_4^T]$. By computing the pseudo-inverse,

$$PA = B \rightarrow PAA^T = BA^T \rightarrow P = BA^T (AA^T)^{-1}$$

Since $H = P^{-1}$, we obtain $H = AA^T (BA^T)^{-1}$. To implement this procedure, we simply construct a warped mesh M' in 3D projective space by computing $M' = KHM$, texture map the image to be projected on it, and use the graphics engine to render it in real time.

As we see, the calibration procedure of this method is quite simple and fast, requiring only a few minutes for the manual alignment of the four projected points. Although similar to the method proposed by [29], here the use of the 3D mesh M and the computer graphics hardware yields much faster rendering speeds.

3rd method: Using a Camera and a Paper Pattern

In this method, we employ the ED-projector camera so we can avoid any manual calibration procedure. In [30], Yang et al. have proposed a method to automatically determine the warping function using a projector/camera pair. Unfortunately, their approach requires precise 3D calibration of the relative attitude of the camera to the projector and, in practice, a considerable baseline distance between them. In particular, it does not work with the ED-projector that aligns the projector and the camera axis as much as possible to avoid user's obstruction of the camera's view.

Instead, we propose here a method that employs a paper pattern positioned on the surface to be calibrated so it is aligned with the desired coordinate system. No relative 3D or 2D calibration is necessary, although we assume that the camera has been positioned in such a way that it has a complete view of the projected surface. Under these conditions, we can define a projective transformation P' between the camera image and the projected surface (see Figure 4). This transformation can be determined automatically by processing an image of the pattern and extracting the camera image coordinates of the corners, $c_i = [c_i^x \ c_i^y \ 1]$, $i = 1, 2, 3, 4$. Assuming the coordinates of the

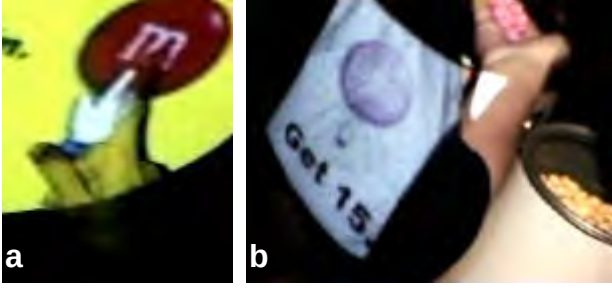


Figure 5. Difficulties faced by the vision system: a) the user's hand is partially hidden by the strong projection; b) the interactive surface is occluded by the user's shoulder as he moves to gather objects from inside a bucket.



Figure 6. Detecting a touch: a) camera view of an interaction with a bucket; b) image difference data; c) overlay of search region (square), button active area (circle), and the fingertip template shown at the pointing location.

pattern corners to be identical to the mesh M in application space, $b_i = [b_i^x \ b_i^y \ 1]$, $i = 1, 2, 3, 4$, we obtain $P'C = B$, and computing the pseudo-inverse,

$$P' = BC^T(CC^T)^{-1}$$

To complete the computation of the transformation H , it is necessary to determine the correspondence between the projector and the camera image planes. This can be done by projecting a pattern with four non-collinear points rendered on the projector image plane $d_i = [d_i^x \ d_i^y \ 1]$, $i = 1, 2, 3, 4$, and obtaining four points on the projected surface, $e_i = [e_i^x \ e_i^y \ 1]$, $i = 1, 2, 3, 4$. Then, by image processing the image of the pattern, it is possible to determine the coordinates of its four corners on the camera image plane, $f_i = [f_i^x \ f_i^y \ 1]$, $i = 1, 2, 3, 4$.

Since $E = [e_1^T \ e_2^T \ e_3^T \ e_4^T]^T$ corresponds both to the projection of $D = [d_1^T \ d_2^T \ d_3^T \ d_4^T]^T$ by the projector, $PD = E$, and to the captured image of the pattern, $P'F = E$, we obtain $PD = P'F$. Computing the pseudo-inverse, we obtain $P = P'FD^T(DD^T)^{-1}$ and therefore

$$H = DD^T(P'FD^T)^{-1}$$

As in the previous method, this transformation is implemented by computing a warped mesh $M' = KHM$.

One problem with all these three distortion-correcting methods is that they project displays with resolution lower than the projector's resolution. The distortion-correction process normally fits an irregular quadrangle into the 4:3 viewing area of typical displays. A considerable amount of display area, and thus resolution, is lost in the process. We have employed standard 1024x768 XGA projectors in our prototypes. Due to the loss of display area created by the distortion correction process, the final display resolution has approximately corresponded to VGA, i.e., 640x480 pixels.

3.2 Sensing Interaction on Any Surface

To support the ability to interact on any surface we use a pan/tilt camera feeding custom image processing and gesture recognition engines. This process has been detailed in another publication [31], but will be outlined here for completeness.

Image processing in this domain is both simplified and made more difficult by the strong light of the projected image. It is simplified in the sense that the lighting environment is very consistent because the dynamic range of the projection swamps all but major changes in ambient light. It is made more difficult because the appearance of the user's hand can be severely distorted as it passes through the projected image (see Figure 5.a).

Although parts of the hand can be hard to detect in a static image, by looking at the differences between adjacent images in the video stream, we can still detect the shape of a moving object as shown in Figure 6.a and b. The use of motion for segmentation instead of background subtraction techniques (such as [32]) also makes it easier to accommodate the variability in the projected image caused by the texture of the surface it is projected on. This motion mask forms the basis of subsequent image understanding.

To provide interaction in a flexible and consistent manner we use the notion of *interface widgets*. Each widget generates computer or application events in response to a particular user motion in a region of the image. We currently use three types of widgets: *buttons* that respond to touch-like motions; *sliders*, 1-dimensional tracking regions, and *touch pads*, which are 2-dimensional tracking regions.

These three widget types rely on tracking the user's fingertip. Convolution template matching is used to find fingertip candidates in the motion mask image described above. The candidates are evaluated using spatial clustering and domain specific heuristics, such as the distance from the user, to determine the best estimate of whether there is a fingertip in the image and where it is currently pointing. The trajectory created by this point over a sequence of video frames is examined by each widget to determine when it should trigger an event.

Touch pad widgets smooth the fingertip trajectory and map it into the touch pad coordinate system. Motion events are only generated when the fingertip is within the touch pad's borders. Sliders are simply a 1-dimensional version of touch pads.

Buttons, however, are more complex widgets. What buttons attempt to detect is when a surface in the environment is touched. Inferring the distance of the fingertip to the surface using the shadow proved to be difficult and unreliable. Fortunately, people tend to interact with real-life buttons by moving in a characteristic forward/pause/backward pattern, and this seems to carry over well to projected buttons. Our implementation of buttons takes advantage of this movement pattern.

To determine when a user has touched a button we look for a roughly linear sequence of fingertip locations that persists for 250ms or longer, with speed within a set range, that either ends or begins with a pause (i.e. the fingertip template did not find a match in the motion mask). A button generates its event when the pause in such a trajectory occurs within it (see Figure 6.c). In other words, we are looking for the user's fingertip to travel in one direction for $\frac{1}{4}$ second and stop within the button, or start moving from the button and move in a consistent direction for $\frac{1}{4}$ second. These approximations for the motion of a fingertip in the moments before and after it touches a button seem to be universal: almost every button touch either starts or ends that way. Interestingly though, many button touches only have one or the other of these motions. Of course, adjustments are needed in the case of people with tremors or other motor disabilities.

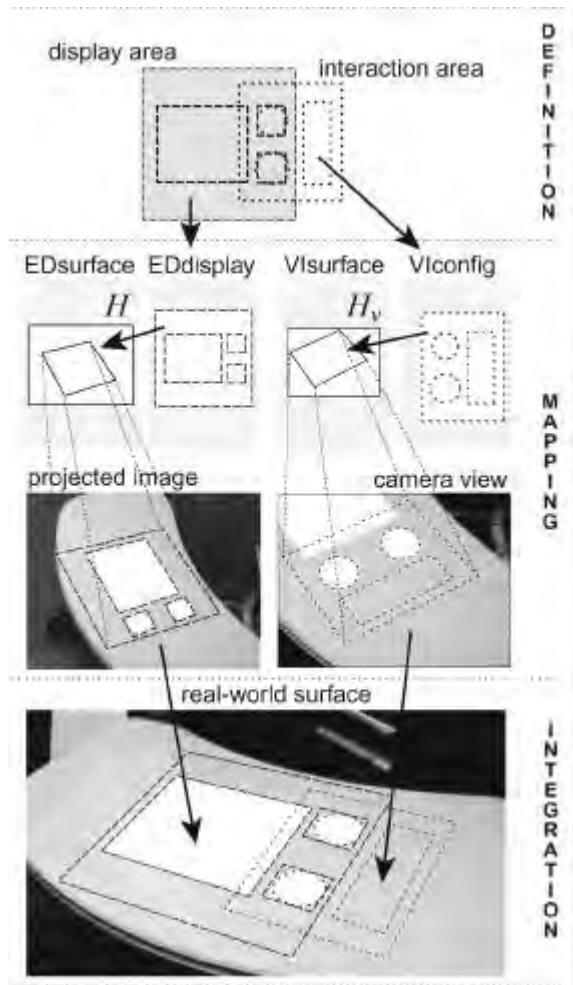


Figure 7. Process of creating an interactive surface on a table.

One of the interesting problems in this domain occurs when users occlude buttons as shown in Figure 5.b. Our experiments with users, even novices, have shown that they naturally avoid obstructing the projection if they are attending to it, but they are quite bad at staying out of the way when they are working with the physical world instead. Buttons have inherent resistance to triggering in these types of situations because any motion must pass several tests before it generates an event. It must resemble a fingertip for several frames; during that time it has to travel with a consistent direction and velocity and stop within a button. Such a sequence sometimes occur as a result of a chance combination of errors, contributing to a somewhat higher false positive rate for some buttons, particularly those which the user occludes for a period of time rather than just passes through.

By using both the shape of the moving region as well as an analysis of its trajectory, this touch detection method gives far better accuracy than using either one alone. Because it relies directly on image data, rather than building an explicit model of the hand, it is computationally efficient enough to run at high frame rates on standard hardware.

Widgets can be combined in various ways in order to perform a complete interaction with the user. However, configurations of widgets are defined independently of the surface where they are used. That is, if an application interface using a set of buttons is moved to different surfaces, it is not necessary to define each button for each surface. Instead, the location of a widget within the

camera image is determined at run time by appropriately mapping the configuration onto the camera image plane. In this case the surface refers to a set of camera parameters. These parameters, like those of the projector, have to be calibrated in advance. In particular, the calibration process has to determine the transformation H_v between the widget coordinate system and the camera image (as shown in Figure 4). Calibration of surfaces is performed using methods analogous to the 2nd and 3rd methods of projection calibration, although information about the expected location of the user, and the size of their hand, has to be obtained during an extra, special calibration step (see [31] for details).

3.3 Integrating Displaying and Sensing

The two preceding sections described the techniques and subsystems that provide the display (output) and sensing (input) functionality of the ED-projector. This section discusses the integration of these separate subsystems in a manner that affords application development and human interaction. One of our primary objectives here is to enable a specific user interaction to occur on any calibrated surface in a room without requiring the customization of this interaction for every surface. This leads to a clear separation of the abstract definition of an interaction from the actual surface upon which it occurs.

Our widgets are constructed using the standard Model View Controller (MVC) design pattern structure [33]. In our case, the projector subsystem is responsible for rendering the *view*, while the vision subsystem provides the *controller* input events. The *definition* layer of Figure 7 shows the separation of the display and interaction areas that are responsible for the definition of the view and the interaction respectively. Display and interaction will often overlap as shown by the square overlapping buttons on the right of the *display area*. Provision is also made, however, for allowing interaction to occur in a region where nothing is displayed, as shown in the large rectangular region on the right of the *interaction area*. Notice that the display area here corresponds to the mesh M of the application plane described before (see also Figure 4).

Every interaction surface in a physical space is named and initially calibrated by the projection and vision subsystems yielding the transforms H and H_v respectively. H corrects the projected image while H_v provides the transformation to the camera image, for a specific surface as described in the prior sections. Figure 7 demonstrates the flow of definitions for display images and interaction data (widget data) from the *definition* layer to the *mapping* layer. During run-time, the system software automatically maps (through the warping functions H and H_v) the *definition* layer data into the projected image and the camera view associated with the current surface. The result is a projected image that generates application events when user gestures occur in the areas defined in the camera view, as shown in the *integration* layer of Figure 7.

4 PAINTING WITH INTERACTIVE GRAPHICS

To better understand how to use projected interactive graphics and how users relate to them, we have designed and implemented three application prototypes using the ED-projector. In the first application, an office space is augmented with dynamic wallpapers, visual notifications of messages, and computer access from multiple surfaces. The second prototype explores an assembly task where the instructions are projected directly onto the surfaces of the objects involved in the task. The third prototype is the “catch the frog” game described in the introduction to this paper.

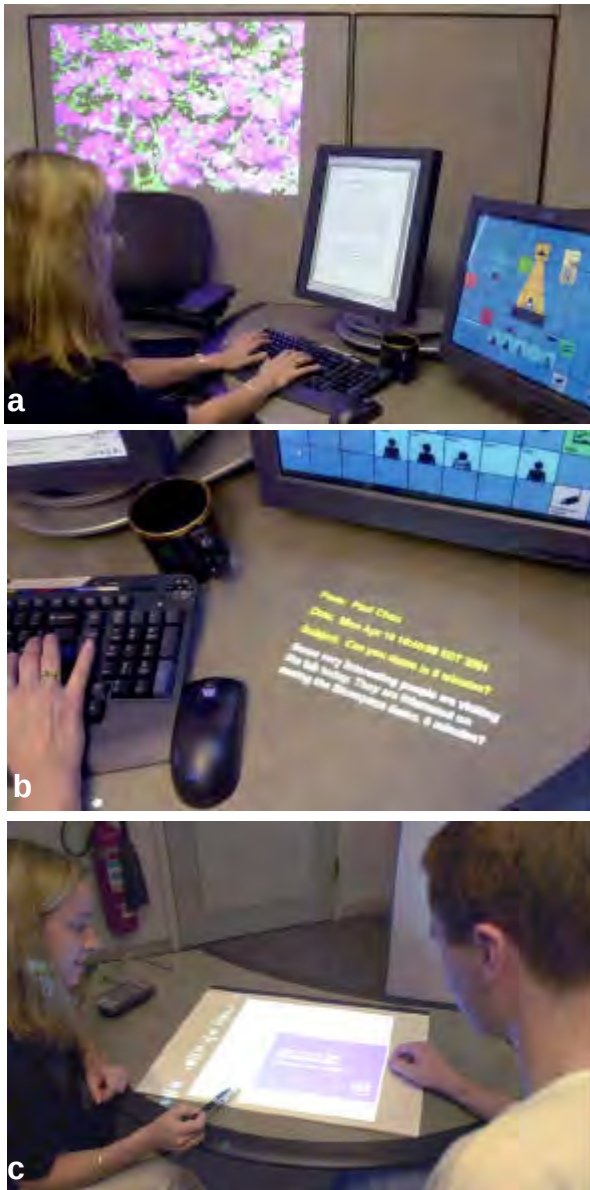


Figure 8. Uses of the ED-projector in an office application: a) to create dynamic wallpapers; b) to provide silent notification of e-mail; and c) to create a computer desktop on a working surface.

4.1 An Office Application

The first prototype application in which we have deployed the ED-projector is a futuristic personalized and context-aware workspace. The workspace has an 8' by 10' footprint and incorporates several sensors for measuring ambient lighting, temperature, humidity and noise level. Users and visitors to the workspace wear active badges that facilitate presence detection and identification. The desk chair is equipped with a pressure sensor connected to a wireless micro-controller, which detects if a person is sitting in the chair. This environment incorporates an ED-projector, strategically located to maximize coverage of the workspace. In this setting, we have used a non-interactive version of the ED-projector to facilitate greater personalization of the environment and to free its users from the confines of the desktop displays.

The presence of an ED-projector in the office tranfigures the space into an extended graphical display and enables the user to have a favorite painting on a wall, receive important notifications on surfaces that catch their attention, or use a tabletop as a display when collaborating with a colleague (see also the video figure). Figure 8.a shows an example of a dynamic wallpaper created by the ED-projector. The user can similarly have a virtual window that is connected to a live webcam looking at her favorite scenery (effectively making it a window office) or monitoring her child at home. The ED-projector can automatically switch to a neutral, less personal image when a co-worker comes to the office, based on the active badge.

Figure 8.b illustrates an urgent notification to the user appearing on a surface close to the user's current location. Here, the smart office detects that the user is seated at the desk and steers the notification right to the surface of the desk. By determining the position of the user, the system is assured that the user will see the urgent message without having to rely on the traditional sound alerts which tend to be disruptive to people working in a shared environment.

Another example of the utility of a steerable display is seen in Figure 8.c where two colleagues have created a computer desktop display on the working surface of their choice. Here, interactive capabilities, provided by a camera system, are planned to be incorporated to facilitate collaborative work.

4.2 Augmenting an Assembly Task

The ED-projector can be seen as a device for augmenting reality – a beam of light that can add information on to real world objects by highlighting them, adding text or images, flashing etc. For example, the ED-projector can show the user an object they are searching for, or indicate the procedure for an assembly task, and even where to place a part in an assembly.

We developed a prototype assembly task to highlight these themes of augmented reality and interaction anywhere. In this prototype application, which was experienced by hundreds of novice users in a major technical exhibition, the object to be assembled is a picture made of M&M's (multi-colored sugar-coated chocolates) where each M&M is regarded as a pixel of the picture.

The theme of interactivity anywhere is demonstrated several times in this prototype application (see also the video figure). Figure 9.a shows a white fabric on a table transformed into an interactive menu for color selection. The user simply touches the color of his choice to select it. Figure 9.b shows a "clickable" button appearing on a paint bucket that contains the M&M's of the selected color. Figure 9.b and c highlight the augmented reality and user assistance aspects of the system. In Figure 9.b the system highlights the bucket that contains the M&M's of the selected color, and also provides additional information on how many M&M's to pick. In Figure 9.c, the system points the exact places on the picture board where the M&M's should be placed. Figure 9.d shows another combination of interaction and augmented reality: the user moves her hand over the surface to interactively reveal the rest of the picture being built.

More than 650 people participated in this augmented assembly over 6 days and completed 4 different M&M pictures. Considering that the vision system and the projection system were integrated for the first time only in the weeks preceding the exhibition, the combined system worked remarkably well. A sample with 130 consecutive users with 621 button touch events (touching gestures or false detections) yielded correct detection of touching gestures in 81% of the events, with 12% false negatives and 7% false positives. If the bucket events are excluded from the count, the

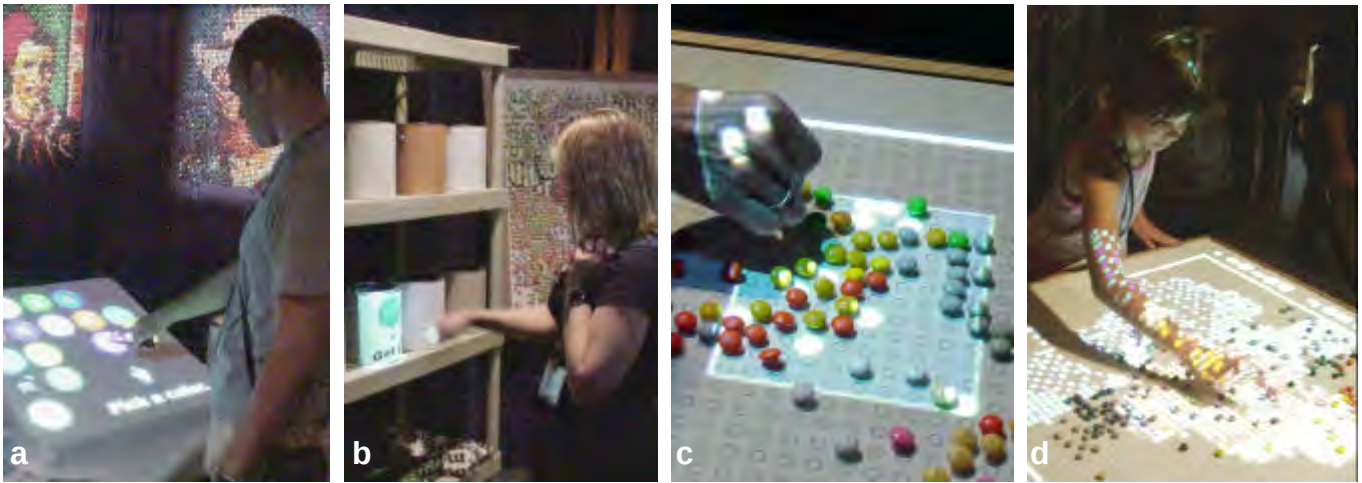


Figure 9. The M&M picture assembly task: a) choosing a color; b) “clicking” a bucket after getting the M&Ms; c) placing the M&Ms on the highlighted areas; and d) “finger painting” to reveal the complete picture.

performance exceeds 89%. A detailed account of the reasons why buckets yielded more errors is given in [31].

The experiment also revealed that users are readily able to relate to the notion of steerable interactive graphics. We also observed the value of the projection creating a shared experience between the users and the people observing the demo: in many cases, users learned the interaction process by first watching other users; and frequently, observers helped users suggesting what to do next. As mentioned before, a major advantage of projection over goggles is precisely this “social” nature of projection.

An interesting observation is that our users were less hesitant to interact with the fabric-covered interactive color selection menu (Figure 9.a) and the M&M placement board (Figure 9.c,d) than with the paint buckets (Figure 9.b). When they went to a bucket to pick up M&M’s, many of them did not realize that they had to click the big “done” button on it (see also Figure 6.a). We had expected that, since this was the third occasion for touch interaction in the demo, users would quite naturally press the button on the bucket as instructed. Most of the time, this was not the case. To help the users in that situation we often gave them verbal instructions such as “Click the bucket,” aiming to make clearer that there was, in fact, a button on the bucket.

4.3 “Fro...og!”

The ED-projector’s unique characteristic of being able to quickly move the projected image to different surfaces is being explored in a computer game for young children. In this game, called “Fro...og!”, an image of a frog is projected on a surface. Whenever the child tries to touch the frog, the camera detects her approach, and the frog “jumps” to another surface, that is, the projector moves the image to a wall or the top of a chair.

Figure 1 shows two pictures of a user playing with the current prototype of “Fro...og!”. The basic idea is to define surfaces in the environment where the animated frog can jump to. Whenever the user’s hand approaches the frog, the next surface is randomly chosen, and an image of the frog jumping in that direction is briefly shown. After landing on the new surface, a simple animation of the frog jumping up and down is projected, so it becomes easier to find the frog in the environment.

In the next development phase of this game we plan to integrate it to a system that tracks the position of multiple children in the room, so the frog can jump to surfaces which are far from most players. Tracking is also particularly important to avoid

projecting the frog directly onto the players which, in this scenario, can easily break the illusion of the game.

5 DISCUSSION

Having shown how we implemented a working prototype of an ED-projector and deployed it in three different applications, it is interesting to explore other possible applications and scenarios, and discuss how to design efficient interfaces for them.

5.1 Other Applications

We see two classes of applications of the ED-projector to create ubiquitous interactive graphics. The first involves the creation of computer interfaces anywhere, while the second deals with augmented reality in a variety of public and private spaces.

In the first class of applications, we regard the ED-projector as a generic input/output device that can replace, in many situations, current displays and interactive devices by creating computer desktop-like interactive displays on non-tethered surfaces. For example, a desktop application can be projected directly on any surface, as previously suggested by Wellner [9] and as demonstrated in subsection 4.1. Unlike the interactive whiteboard described in [34], the projected application can be easily moved around the room, for instance, from a whiteboard on the wall to the top of a desk for more detailed reading. Similarly, the ED-projector can be used to bring information to the physical location where it is used or needed. For example, a database application managing reports can be projected on top of the file cabinet with hard copies of the reports.

In the second class of applications the ED-projector can augment reality by pointing to physical objects, showing connections among them, and projecting patterns to indicate movement or change in the real world. The prototype assembly task described in subsection 4.2 is a typical example of this class of applications. Another example applies to a library in which directions are provided to a user looking for a book by projecting arrows on the floor, walls and shelves, and finally by highlighting the desired book.

Steerable graphics produced by the ED-projector can also aid in collaborative work by providing large displays on convenient surfaces, and allowing easy reconfiguration of meeting spaces for different functions and teamwork styles, in similar ways to the office project described in subsection 4.1.

The ED-projector can also be used to provide computer and information access in environments where traditional displays are

difficult to install, secure, or maybe unsafe to operate, such as public spaces and areas subject to harsh environmental conditions. The device also permits an interactive display to be brought to the proximity of a user, eliminating the need for the user to approach a traditional, fixed display. In particular, the ED-projector can facilitate the access and use of computers by people with locomotive disabilities. For instance, it can project an interactive display on a hospital bed sheet without requiring the patient to contact any device.

We also see the ED-projector as a potential enabler of a new generation of games where interaction happens not in the virtual world but in the physical world such as the “Fro...og!” game described above. Unlike games based on phones and portable computers [4], the use of the ED-projector provides high-resolution displays where characters and fantastic objects can move from surface to surface, creating a game that surrounds the players. Using video input, various user actions can be detected depending on the game needs, for example, hand gestures, body movements, foot action, and/or facial expressions. Above all, a single ED-projector installed on the ceiling of a living room or an arcade, using just software commands, can completely reconfigure the space to create different games according to the interests, age, and motor skills of the players.

The scenarios examined so far are chiefly concerned in creating ways for humans to command computers and/or obtain information. However, we see the projecting of interactive graphics ubiquitously as a way to provide computer agents the ability to “reach and touch” the real world and the people that inhabit it. By using projected patterns, it is possible to make computers not only interact with people, but act and coordinate them. For instance, a computer agent can control people in a line by projecting red lines and green arrows around the people on the line. A steerable projector system is, thus, able to create a “body”, composed of light, for an agent. From this perspective, we may be creating an interface to the real world for computers.

5.2 Human Interface Issues

As foreseen by Weiser [1], ubiquitous interactive graphics do not seem to follow the standard paradigm of interaction based on the desktop metaphor. In our M&M demonstration, we had the opportunity to observe more than 650 novice users experiencing a scenario of true ubiquitous computing, in the context of an assembly task. On one side, we observed that they quite naturally accepted the idea of augmenting reality with information directly projected on objects. Moreover, we saw they naturally change their body’s position to minimize problems like the shadow of their arms and body over the board. People seemed not only to accept projected graphics naturally but also to compensate for its limitations.

However, when trying to click a button, the users almost always applied pressure onto the projected surface with their fingers. Also, whenever a button touch was not detected by the vision system, the most common reaction was to press again with more pressure and, if still not successful, using the whole hand instead of a finger. It seems that people expect buttons, even virtual buttons, to react to pressure and not to touch, and to fail if enough pressure is not applied on them.

This observation clearly illustrates the limitations of the hand-as-a-mouse paradigm used in the M&M demo. Although it is easier for users to relate the ubiquitous experience to their everyday contact with desktop interfaces, the metaphor proved to be misleading to the true nature of the interface. Unlike a mouse that has appropriate haptic feedback for clicking gestures, touching a projected button is devoid of feedback movement.

Similar observations prompt us to believe that a new paradigm for interaction with projected screens must be developed, with new kinds of widgets, tuned for the natural capabilities and limitations of projection and gesture-based interaction, just as the point and click paradigm is tuned for the characteristics of a physical mouse. Touch-screen-like interactions do translate better to the projected interface domain. Although our widgets are not based on surface contact, the widgets we have found useful so far (buttons, scroll bars and touch pads) are all based on motions people naturally make when touching.

One of the exciting aspects of this work is that it is an excellent environment to explore more adventurous gesture-based interaction styles. By giving the user feedback as to what gestural interactions the system is expecting and where in the environment they may be performed, the cognitive load on the user can be reduced. Without such feedback a user would have to memorize what amounts to a set of complex dance steps to simply use a gesture-based interface.

This investigation of new human-computer interface paradigms is even more important in the case of multi-user interaction. Unlike desktops, wearable computers, or PDAs, projected interactive graphics naturally allow many users to simultaneously interact with an application. Although the applications we have developed so far are all single-user, it seems clear multi-user interaction will soon become an important area of research.

Another intriguing question is related to the different degrees of difficulty experienced by users when interacting with graphics projected on different supporting objects. Clicking a paint bucket seems to be harder to accept than clicking a table. Of course the design of our system and the simplicity of our experiment allow us neither to definitively state that the phenomenon does happen nor to draw conclusions about its causes. However, one hypothesis is that people seem to attribute different interactive capabilities to projected interactive graphics according to the nature of the surfaces themselves. Are the paint buckets “functionally fixed”, that is, if something looks and acts like a paint bucket, is it less likely to be perceived as something else – namely, a touch-screen? We are currently preparing a study to determine the validity of this hypothesis.

Finally, our research so far has addressed only visual input and output. A promising area of research seems to be the application of the concept to steerability of other stimuli such as sound, using, for instance, the audio spotlights described by Pompei [35]. What kind of interfaces and applications can be created by a full multimodal steerable system?

6 CONCLUSION

This paper describes a new method and device for creating ubiquitously interactive graphics on non-tethered surfaces. We have shown three different methods to determine the warping function that corrects for oblique projection distortion and a computer vision system that tracks the user’s hand over the projected images and detects touch-like gestures. We have also implemented three applications that realize ubiquitous computing concepts such as everywhere computer access and augmented reality. Based on the experience of running the M&M demo with hundreds of novice users, we have also detected a need to re-examine the current interaction paradigms, and possibly, investigate and propose new ones.

We see this work as an important contribution to realize the vision of ubiquitous computing, without asking people to wear headsets or wiring the surfaces of everyday objects. However, we do not advocate projection-based solutions as a panacea for ubiquitous applications: in many cases, it is more desirable and efficient to use head-mounted displays. A good analogy is to

compare headsets and projectors to Walkmans and stereo sound systems. Walkmans are unsurpassable as a way to provide music when people are moving around or in public spaces such as trains and airplanes. They are very effective in creating a personal experience, albeit not able to create social interaction. Similarly, we keep loving stereo systems because they render music to a whole space, nurturing a collective experience that is attached to the space itself. We see similar advantages and limitations when comparing headsets and steerable projectors as a way to render interactive graphics onto the real world.

Acknowledgements

“M&M” is a registered trademark of Mars, Inc. “Walkman” is a registered trademark of Sony Corporation.

References

1. Weiser, M., *The Computer for the Twenty-First Century*. Scientific American, 1991. **265**(3): p. 94-100.
2. Pinhanez, C. *The Everywhere Displays Projector: A Device to Create Ubiquitous Graphical Interfaces*. in *Proc. of Ubiquitous Computing 2001 (UbiComp'01)*. 2001. Atlanta, Georgia.
3. Barfield, W. and T. Caudell, eds. *Fundamentals of Wearable Computers and Augmented Reality*. 2001, Lawrence Erlbaum Assoc. 728.
4. Raskar, R., et al. *The Office of the Future: A Unified Approach to Image-Based Modeling and Spatially Immersive Displays*. in *Proc. of SIGGRAPH'98*. 1998. Orlando, Florida.
5. Ishii, H. and B. Ullmer. *Tangible Bits: Towards Seamless Interfaces between People, Bits, and Atoms*. in *Proc. of CHI'97*. 1997. Atlanta, Georgia.
6. Bellour, R., et al., *Tony Oursler*. 2001: Ediciones Poligrafa. 208.
7. Naimark, M., *Spatial Correspondence in Motion Picture Display*. SPIE Optics in Entertainment II, 1984. **462**: p. 78-81.
8. Bolt, R., *Put That There: Voice and Gesture at the Graphics Interface*. ACM Computer Graphics, 1980. **14**(3): p. 262-270.
9. Wellner, P., *Interacting with Paper on the DigitalDesk*. Communications of the ACM, 1993. **36**(7).
10. Morishima, S., et al. *HyperMask: Talking Head Projected Onto Real Objects*. in *Proc. of Multimedia Modeling (MMM'00)*. 2000: World Scientific.
11. Raskar, R., et al. *Shader Lamps: Animating Real Objects with Image-Based Illumination*. in *Proc. of 12th Eurographics Workshop on Rendering*. 2001. London, England.
12. Krueger, M.W., *Artificial Reality II*. 1990: Addison-Wesley.
13. Keays, B. and R. Macneil. *metaField Maze*. in *Proc. of SIGGRAPH'99*. 1999. Los Angeles, California.
14. Omojola, O., et al., *An Installation of Interactive Furniture*. IBM Systems Journal, 2000. **39**(3&4): p. 861-879.
15. Ju, W. *Origami Desk*. in *Proc. of SIGGRAPH'01 - Conference Abstracts and Applications*. 2001. Los Angeles, California.
16. Back, M., et al., *Designing Interactive Reading Experiences for a Museum Exhibition*. IEEE Computer Magazine, 2001. **34**(1): p. 1-8.
17. Underkoffler, J., B. Ullmer, and H. Ishii. *Emancipated Pixels: Real-World Graphics in the Luminous Room*. in *Proc. of SIGGRAPH'99*. 1999. Los Angeles, CA.
18. Rekimoto, J. *A Multiple Device Approach for Supporting Whiteboard-based Interactions*. in *Proc. of CHI'98*. 1998. Los Angeles, CA.
19. Schiele, B. and J.L. Crowley, *Recognition without Correspondence using Multidimensional Receptive Field Histograms*. IJCV, 2000. **36**(1): p. 31-52.
20. Wu, Y. and T. Huang, *Vision-Based Gesture Recognition: A Review*. Lecture Notes in Artificial Intelligence, 1999. **1739**.
21. Hardenberg, C.v. and F. Berard. *Bare-hand human-computer interaction*. in *Proc. of Workshop on Perceptive User Interfaces, PUT'01*. 2001. Orlando, Florida.
22. Quek, F., T. Mysliwiec, and M. Zhao. *Finger mouse: A freehand pointing interface*. in *Proc. of International Workshop on Automatic Face- and Gesture-Recognition*. 1995. Zurich, Switzerland.
23. Segen, J. *GestureVR: Vision-Based 3D Hand Interface for Spatial Interaction*. in *Proc. of ACM Multimedia Conference*. 1998. Bristol, England.
24. Kjeldsen, F., *Visual Recognition of Hand Gesture as a Practical Interface Modality*. 1997, Columbia University: New York, New York.
25. Raskar, R., et al. *Multi-Projector Displays Using Camera-Based Registration*. in *Proc. of IEEE Visualization'99*. 1999. San Francisco, CA.
26. Pinhanez, C., F. Nielsen, and K. Binsted. *Projecting Computer Graphics on Moving Surfaces: A Simple Calibration and Tracking Method*. in *Proc. of SIGGRAPH'99*. 1999. Los Angeles, California.
27. Raskar, R. *Oblique Projector Rendering on Planar Surfaces for a Tracked User*. in *Proc. of SIGGRAPH'99*. 1999. Los Angeles, California.
28. Faugeras, O., *Three-Dimensional Computer Vision: A Geometric Viewpoint*. 1993, Cambridge, Massachusetts: The MIT Press.
29. Sukthankar, R., R. Stockton, and M. Mullin. *Smarter Presentations: Exploiting Homography in Camera-Projector Systems*. in *Proc. of ICCV'01*. 2001. Vancouver, Canada.
30. Yang, R. and G. Welch. *Automatic and Continuous Projector Display Surface Calibration Using Every-Day Imagery*. in *Proc. of 9th International Conf. in Central Europe in Computer Graphics, Visualization, and Computer Vision*. 2001. Plzen, Czech Republic.
31. Kjeldsen, R., et al. *Interacting with Steerable Projected Displays*. in *Submitted to Face and Gesture'02*. 2002.
32. Wren, C., et al., *Pfinder: Real-Time Tracking of the Human Body*. IEEE Trans. Pattern Analysis and Machine Intelligence, 1997. **19**(7): p. 780-785.
33. Buschmann, F., et al., *A System of Patterns - Pattern Oriented Software Architecture*. 1996, New York, New York: John Wiley & Sons.
34. Crowley, J.L., J. Coutaz, and F. Berard, *Things that See*. Communications of the ACM, 2000. **43**(3): p. 54-64.
35. Pompei, F.J. *The Use of Airborne Ultrasonics for Generating Audible Sound Beams*. in *Proc. of 105th Audio Engineering Society Convention*. 1998. San Francisco, CA.

Appendix B-2



IBM Research

Research Areas ▾

Labs ▾

Disciplines

Blog

Claudio S. Pinhanez


[feedback](#)

Senior Manager, Conversational Intelligence
 IBM Research - Brazil , Sao Paulo
csantosp@br.ibm.com

Professional Associations: [ACM](#) | [ACM SIGCHI](#) | [Brazilian Computer Society](#) | [IBM Academy of Technology](#) | [IEI](#)

More information: [LinkedIn](#)

[Profile](#)
[Publications](#)
[Patents](#)

2020

Creating Corpora for Seq2Seq Tone Rephrasing Using Social Media Posts

Paulo Cavalin, [Marisa Vasconcelos](#), Marcelo Grave and Claudio Pinhanez

International Joint Conference on Neural Networks, 2020

[From Disjoint Sets to Parallel Data to Train Seq2Seq Models for Sentiment Transfer](#)

Paulo R. Cavalin, [Marisa Vasconcelos](#), Marcelo Grave, [Claudio S. Pinhanez](#), Victor Henrique Alves Ribeiro

Findings of the Association for Computational Linguistics: EMNLP 2020, pp. 689-698

artificial neural network, semantic similarity, machine translation, machine learning, computer science, exploit, disjoint grammar, fluency, artificial intelligence

[Improving Out-of-Scope Detection in Intent Classification by Using Embeddings of the Word Graph S](#)

Paulo R. Cavalin, Victor Henrique Alves Ribeiro, Ana Paula Appel, [Claudio S. Pinhanez](#)

Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 3952-3961

[Abstract](#) graph embedding, theoretical computer science, percentage point, computer science, inverse, discrete s

[HCI Research Challenges for the Next Generation of Conversational Systems](#)

[Claudio S. Pinhanez](#)

Proceedings of the 2nd Conference on Conversational User Interfaces, 2020

psychology, human computer interaction

Unveiling Practices and Challenges of Machine Teachers of Customer Service Conversational Systems

Heloisa Candello, Mairieli Santos Wessel, [Claudio S. Pinhanez](#)

Proceedings of the 2nd Conference on Conversational User Interfaces, 2020

knowledge management, customer service, computer science

Intentions, Meanings, and Whys: Designing Content for Voice-based Conversational Museum Guides

Fabrecio Barth, Heloisa Candello, [Paulo R. Cavalin](#), [Claudio S. Pinhanez](#)

Proceedings of the 2nd Conference on Conversational User Interfaces, 2020

linguistics, computer science

Co-designing a conversational interactive exhibit for children

Heloisa Candello, Mauro Pichiliani, [Claudio S. Pinhanez](#), Sara Vidon, Mairieli Santos Wessel

Proceedings of the 2020 ACM Interaction Design and Children Conference: Extended Abstracts, pp. 326-331

human computer interaction, computer science

2019

Can direct address affect user engagement with chatbots embodied in physical spaces

[Heloisa Candello](#), [Claudio Pinhanez](#), Mauro Pichiliani, [Marisa Vasconcelos](#), Haylla Conde

Proceedings of the 1st International Conference on Conversational User Interfaces, 2019

vocative case, visitor pattern, user experience design, social relation, psychology, performing arts, observational study, empirical research, embodied cognition

Machine Teaching by Domain Experts: Towards More Humane, Inclusive, and Intelligent Machine Learning

[Claudio S. Pinhanez](#)

arXiv preprint arXiv:1908.08931, 2019

Abstract teaching technology, open source software, mediation, intelligent machine, human computer interaction, approach

Live It Up: Analyzing Emotions and Language Use in Tweets during the Soccer World Cup Finals

[Marisa Vasconcelos](#), Jussara Almeida, [Paulo Cavalin](#), [Claudio Pinhanez](#)

10th ACM Conference on Web Science, 2019

Teaching Robots to Act and Converse in Physical Spaces: Participatory Design Fictions with Museum

[Heloisa Candello](#), Mauro Pichiliani, Mairieli Wessel, [Claudio Pinhanez](#), Michael Muller

Proceedings of the Halfway to the Future Symposium 2019 on

visual arts, the imaginary, sociology, social dilemma, robot, participatory design, emerging technologies, dilemma, c

The Effect of Audiences on the User Experience with Conversational Interfaces in Physical Spaces

[Heloisa Candello](#), [Claudio Pinhanez](#), Mauro Pichiliani, [Paulo Cavalin](#), Marisa Vasconcelos, Haylla Conde

CHI 2019

2018

[Recovering from Dialogue Failures Using Multiple Agents in Wealth Management Advice](#) 

[Heloisa Candello](#), [Claudio Pinhanez](#)

Studies in Conversational UX Design, Springer, 2018

Moore, R.J., Szymanski, M.H., Arar, R., Ren, G.-J. (Eds.)

[The Role of Dialogue User Data in the Information Interaction Design of Conversational Systems](#) 

[Heloisa Candello](#), [Claudio Pinhanez](#)

DUXU'2018, Springer

Specifying and Implementing Multi-Party Conversation Rules with Finite-State-Automata

Maira Gatti de Bayser, Melina Alberio Guerra, [Paulo R. Cavalin](#), [Claudio S. Pinhanez](#)

AAAI Workshops, pp. 713-719, 2018

machine learning, finite state machine, conversation, computer science, artificial intelligence

[Having an Animated Coffee with a Group of Chatbots from the 19 th Century](#) 

Candello, Heloisa and Pinhanez, Claudio and Pichiliani, Mauro Carlos and Guerra, Melina Alberio and Gatti de Bayse

Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, pp. D206

[Abstract](#)

[Revealing Physicians Referrals from Health Insurance Claims Data](#) 

de Santana, Vagner Figueredo and Appel, Ana Paula and Moyano, Luis Gregorio and Ito, Marcia and Pinhanez, Claud

Big Data Research, Elsevier, 2018

[Abstract](#)

2017

[Personality Archetypes, Pantomime, Comics: New Methods to Design Interfaces with Human Traits](#) 

[Claudio S. Pinhanez](#)

International Conference on Computer-Human Interaction Research and Applications, 1-19, 2017

[Abstract](#) service recovery, personality, human computer interaction, humanoid robot, engineering design process, comics, archetype, animation

[Modeling Task Deviations as Eccentricity Distribution Peaks](#) 

de Santana, Vagner Figueredo and de Paula, Rogério Abreu and Pinhanez, Claudio Santos

Proceedings of the XVI Brazilian Symposium on Human Factors in Computing Systems, pp. 36, 2017

[Abstract](#) (Honorable Mention Paper)

[Typefaces and the Perception of Humanness in Natural Language Chatbots](#)

Candello, Heloisa and Pinhanez, Claudio and Figueiredo, Flavio

Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 3476--3487

[Abstract](#)

[A Hybrid Architecture for Multi-Party Conversational Systems](#)

Gatti de Bayser, Maira and Cavalin, Paulo and Souza, Renan and Braz, Alan and Candello, Heloisa and Pinhanez, Claudio

arXiv preprint arXiv:1705.01214, 2017

[Abstract](#)

Shaping the Experience of a Cognitive Investment Adviser

Candello, Heloisa and Pinhanez, Claudio and Millen, David and Andrade, Bruna Daniele

International Conference of Design, User Experience, and Usability, pp. 594--613, 2017

[Bottester: Testing Conversational Systems with Simulated Users](#)

Marisa Vasconcelos, [Heloisa Candello](#), [Claudio Pinhanez](#)

IHC 2017, ACM

Evaluating the Conversation Flow and Content Quality of a Multi-bot Conversational System

Heloisa Candello, Marisa Vasconcelos, [Claudio Pinhanez](#)

HCI in Practice (IHC na Pratica) Workshop (co-located with the Brazilian Symposium about Human Factors in Computing Systems)

[Computer Interfaces to Organizations: Perspectives on Borg-Human Interaction Design](#)

Pinhanez, Claudio

arXiv preprint arXiv:1712.03012, 2017

[Abstract](#)

2016

[Service Design Tools to Explore Financial Services for Poor Microbusiness Owners](#)

Candello, Heloisa and Pinhanez, Claudio and Millen, David and Bianchi, Silvia

Service Design Geographies. Proceedings of the ServDes. 2016 Conference, pp. 192--204

[Abstract](#)

2015

[Using Visual Analytics to Diagnose Productivity and Quality Issues on IT Service Pools](#)

Cavalcante, Victor and Candello, Heloisa and Pinhanez, Claudio and Braz, Alan and Amorin, Franklin and Nauata, Neiva

Maximizing Management Performance and Quality with Service Analytics, pp. 207--232, IGI Global, 2015

[Abstract](#)

Understanding Fiado : Informal Credit in Brazil

Heloisa Candello, David Millen, Silvia Bianchi, Rogerio de Paula, [Claudio Pinhanez](#)

Proceedings of the 2015 Annual Symposium on Computing for Development, 147-147

Abstract northeast brazil, microfinance, marketing, financial application, database transaction, biology

A scalable architecture for real-time analysis of microblogging data

Cavalin, Paulo Rodrigo and Gatti, Ma{\'i}ra A de C and Moraes, TGP and Oliveira, Fabio Silva and Pinhanez, Claudio S
Paula, Rog{\'e}rio Abreu

IBM Journal of Research and Development 59(2/3), 16--1, IBM, 2015

Abstract

Collaboration and social computing in emerging financial services

Millen, David R and Pinhanez, Claudio and Kaye, Jofish and Bianchi, Silvia Cristina Sardela and Vines, John

Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing, ,

Abstract

2014

Borg-Human Interaction Design

[Claudio S. Pinhanez](#)

ServDes.2014 Service Future; Proceedings of the fourth Service Design and Service Innovation Conference; Lancaster April 2014 pp. 99, 100-109

Abstract service design, personality, narrative, human computer interaction, human interaction, engineering, emb methods

A CIENCIA DA INOVACAO EM SERVICOS: ESTUDO EXPLORATORIO SOBRE OS INTERESSES E PRIOR PESQUISA NO BRASIL

Paulo Cesar Calabria, Roberto Carlos Bernardes, Eduardo Raupp de Vargas, [Claudio S. Pinhanez](#)

RAI: Revista de Administracao e Inovacao 10(4), 110-135, 2014

Abstract

Unsupervised annotation of usability test videos

de Santana, Vagner Figueredo and Pinhanez, Claudio Santos and de Paula, Rog{\'e}rio Abreu

Proceedings of the 13th Brazilian Symposium on Human Factors in Computing Systems, pp. 207--215, 2014

Handling Big Data on Agent-based Modeling of Online Social Networks with Mapreduce

Maira A. de C. Gatti, Marcos R. Vieira, Joao Paulo F. de Melo, [Paulo Rodrigo Cavalin](#), [Claudio Santos Pinhanez](#)

Proceedings of the 2014 Winter Simulation Conference, pp. 851--862, IEEE Press

Towards Personalized Offers by Means of Life Event Detection on Social Media and Entity Matching

[Paulo Rodrigo Cavalin](#), Maira Gatti, [Claudio Santos Pinhanez](#)

HT (Doctoral Consortium / Late-breaking Results / Workshops), 2014

Abstract world wide web, software framework, social media, life events, infosphere, ibm, computer science, bloss

2013

[An Awareness Raising System to Support Cross-department Coordination in Matrix Organizations](#) 

[Vagner Figueredo de Santana](#), [Rogerio Abreu de Paula](#), [Claudio Santos Pinhanez](#), Silvia Cristina Sardela Bianchi, Elaine Na Costa, Carina Missae

Proceedings of Brazilian Symposium on Collaborative Systems, pp. 136, 2013

[Visual summarization of incidents in the context of service delivery factories](#) 

[Vagner Figueredo de Santana](#), [Rogerio Abreu de Paula](#), [Claudio S Pinhanez](#), Silvia Cristina Sardela Bianchi, Elaine Na Carina Missae

Proceedings of the 12th Brazilian Symposium on Human Factors in Computing Systems, pp. 300--303, 2013

[Data-Driven Analytical Tools for Characterization of Productivity and Service Quality Issues in IT Serv](#)

Victor F. Cavalcante, [Claudio S. Pinhanez](#), Rogerio A. de Paula, Carolina S. Andrade, Cleidson R. B. de Souza, Ana Pa web science, pp. 295-310, 2013

Abstract service quality, service product management, service level requirement, service level objective, service g delivery framework, operations research, operations management, incident management, computer science

[Reaction times for user behavior models in microblogging online social networks](#) 

Samuel Martins Barbosa Neto, Maira Athanazio De Cerqueira Gatti, [Paulo Rodrigo Cavalin](#), [Claudio Santos Pinhanez](#), [Paula Appel](#)

Proceedings of the 2103 workshop on Data-driven user behavioral modelling and mining from social media, pp. 17--.

[Data-driven analytical tools for characterization of productivity and service quality issues in IT service](#)

Cavalcante, Victor F and Pinhanez, Claudio S and de Paula, Rog{e}rio A and Andrade, Carolina S and de Souza, Cleid *Journal of Service Research*, 1094670513480559, SAGE Publications, 2013

Abstract

[A Simulation-based Approach to Analyze the Information Diffusion in Microblogging Online Social Ne](#)

Maira A de C Gatti, Ana Paula Appel, Cicero Nogueira dos Santos, [Claudio Santos Pinhanez](#), [Paulo Rodrigo Cavalin](#), S *Winter Simulation Conference, WSC'13*, 2013

accepted for publication

[Large-Scale Multi-Agent-based Modeling and Simulation of Microblogging-based Online Social Netw](#)

Maira Gatti, [Ana Paula Appel](#), [Claudio Pinhanez](#), Cicero dos Santos, Daniel Gribel, [Paulo Cavalin](#), and Samuel Barbos *The 14th International Workshop on Multi-Agent-based Simulation (MABS, AAMAS 2013)*

2012

Designing an enterprise social tool for cross-boundary communication, coordination, and information de Souza, [Cleudson Ronald B and Pinhanez](#), Claudio S and Cavalcante, Victor F and Aluani, Fernando and Daros, Vini Paula, Rogério A

Proceedings of the 30th ACM international conference on Design of communication, pp. 55--64, 2012

[Abstract](#)

Using Social Analytics for Studying Work-Networks: A Novel, Initial Approach [↗](#)

Rogério A de Paula, Ana Paula Appel, [Claudio S Pinhanez](#), Victor F Cavalcante, Carolina S Andrade

Collaborative Systems (SBSC), 2012 Brazilian Symposium on, pp. 146--153

Scheduling with Preemption for Incident Management: When Interrupting Tasks is not Such a Bad Idea

Marcos D. Assuncao, Victor F. Cavalcante, Maira A. de C. Gatti, Marco A. S. Netto, [Claudio S. Pinhanez](#), and Cleudson

Proceedings of the Winter Simulation Conference (WSC '12), pp. 12, ACM, 2012

Characterizing Time-Bounded Incident Management Systems [↗](#)

V Cavalcante, [Claudio S Pinhanez](#), CB de Souza, RA de Paula, Ana Paula Appel, Carolina S Andrade

SRII Global Conference (SRII), 2012 Annual, pp. 750--759

2011

Knowledge and information and needs of system administrators in IT service factories [↗](#)

De Souza, [Cleudson RB and Pinhanez](#), Claudio S and Cavalcante, Victor

Proceedings of the 10th Brazilian Symposium on Human Factors in Computing Systems and the 5th Latin American C Interaction, pp. 81--90, 2011

[Abstract](#)

Information needs of system administrators in information technology service factories [↗](#)

C.R.B. de Souza, [C.S. Pinhanez](#), V.F. Cavalcante

Proceedings of the 5th ACM Symposium on Computer Human Interaction for Management of Information Technology,

Human aspects of internet services: considering the needs of users and providers [↗](#)

[C.S. Pinhanez](#)

Journal of Internet Services and Applications 1(3), 155--164, Springer, 2011

2010

Designing the interaction with service systems [↗](#)

[C. Pinhanez](#)

Proceedings of the 28th ACM International Conference on Design of Communication, pp. 253--253, 2010

2009

[Services as Customer-Intensive Systems](#)

[Claudio Pinhanez](#)

Design Issues 25(2), 3-13, 2009

Abstract type of service, service system, service provider, service product management, service design, service design process management, marketing, information system, business

[A Service Science Perspective on Human-Computer Interface Issues of Online Service Applications](#) [

[C. Pinhanez](#)

International Journal of Information Systems in the Service Sector (IJISSS) 1(2), 17--35, IGI Global, 2009

[Humans inside as the key characteristic of service systems](#)

[C. Pinhanez](#)

Proceedings of QUIS, pp. 515--524, 2009

[Multimedia Chat for Helpdesks: A Requirements Study, a Practical SOA Architecture, and an Interface](#)

Z.Y. Shae, T. Bergstrom, [C. Pinhanez](#), [M. Podlaseck](#)

IEEE International Symposium on Object/Component/Service-Oriented Real-Time Distributed Computing ISORC'09, p

2008

[Aplicando conceitos e tecnicas de servicos em IHC](#)

[Claudio Pinhanez](#)

IHC 08 Proceedings of the VIII Brazilian Symposium on Human Factors in Computing Systems, pp. 346-347, 2008

Abstract philosophy, humanities

[Multimedia Chat for Helpdesks: a Practical SOA Architecture](#)

Z.Y. Shae, T. Bergstrom, [C. Pinhanez](#), [M. Podlaseck](#)

IEEE Congress on Services. Part I, pp. 75-77, 2008

[Service systems as customer-intensive systems and its implications for service science and engineer](#)

[C. Pinhanez](#)

Hawaii International Conference on System Sciences, Proceedings of the 41st Annual, pp. 117--117, 2008

[A service science perspective for interfaces of online service applications](#)

[C. Pinhanez](#)

Proceedings of the VIII Brazilian Symposium on Human Factors in Computing Systems, pp. 11--20, 2008

[A Proposal for a Service Science Discipline Classification System](#)

[C. Pinhanez](#), P. Kontogiorgis

Frontiers in Service conference, 2008

2007

[Ubiquitous Services](#)

[C. Pinhanez](#)

Services Computing, 2007. SCC 2007. IEEE International Conference on, pp. 146--153

[A Services Theory Approach to Online Service Applications](#)

[C. Pinhanez](#)

Services Computing, 2007. SCC 2007. IEEE International Conference on, pp. 395--402

2006

[From Users to Customers: HCI for Online Services](#)

[C. Pinhanez](#)

Information & Communications Technology, 2006. ICICT'06. ITI 4th International Conference on, pp. 1--1

2005

[A study on the manipulation of 2D objects in a projector/camera-based augmented reality environme](#)

S. Voida, [M. Podlaseck](#), R. Kjeldsen, [C. Pinhanez](#)

Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 611--620, 2005

[To frame or not to frame: The role and design of frameless displays in ubiquitous applications](#)

[C. Pinhanez](#), [M. Podlaseck](#)

UbiComp 2005: Ubiquitous Computing, 903-903, Springer

[Using symbiotic displays to view sensitive information in public](#)

S. Berger, R. Kjeldsen, C. Narayanaswami, [C. Pinhanez](#), [M. Podlaseck](#), M. Raghunath

3rd IEEE International Conference on Pervasive Computing and Communications, PerCom, pp. 139-148, 2005

2004

[A portable system for anywhere interactions](#)

N. Sukaviriya, R. Kjeldsen, [C. Pinhanez](#), L. Tang, A. Levas, G. Pingali, [M. Podlaseck](#)

CHI'04 extended abstracts on Human factors in computing systems, pp. 789--790, 2004

[Projector-camera systems for telepresence](#) [C. Pinhanez](#), G. Pingali*Proceedings of the 2004 ACM SIGMM workshop on Effective telepresence*, pp. 63--66[Dynamically reconfigurable vision-based user interfaces](#) R. Kjeldsen, A. Levas, [C. Pinhanez](#)*Machine Vision and Applications* 16(1), 6--12, Springer, 2004

2003

[Interval scripts: a programming paradigm for interactive environments and agents](#) [C.S. Pinhanez](#), A.F. Bobick*Personal and Ubiquitous Computing* 7(1), 1--21, Springer, 2003[An Architecture and Framework for Steerable Interface Systems](#) A. Levas, [C. Pinhanez](#), G. Pingali, R. Kjeldsen, [M. Podlaseck](#), N. Sukaviriya*UbiComp 2003: Ubiquitous Computing*, pp. 333--348[Ubiquitous interactive displays in a retail environment](#) [C. Pinhanez](#), [M. Podlaseck](#), R. Kjeldsen, A. Levas, G. Pingali, N. Sukaviriya*Proceedings of SIGGRAPH Sketches* 2(3), Citeseer, 2003[Applications of steerable projector-camera systems](#) [C. Pinhanez](#), R. Kjeldsen, A. Levas, G. Pingali, M. Podlaseck, N. Sukaviriya*ICCV 2003: Proceedings of the IEEE International Workshop on Projector-Camera Systems*[Augmenting a Retail Environment Using Steerable Interactive Displays](#) Noi Sukaviriya, Mark Podlaseck, Rick Kjeldsen, Anthony Levas, Gopal Pingali, [Claudio Pinhanez](#)*CHI '03 extended abstracts on Human factors in computing systems*, pp. 978--979, ACM, 2003

Abstract 608035

[Ubiquitous computing in the living room: Concept sketches and an implementation of a persistent user interface](#)S.S. Intille, V. Lee, [C. Pinhanez](#)*Proc. of the Video Session of Ubi-Comp3*, 12--15, 2003[On interfaces projected onto real-world objects](#) [M. Podlaseck](#), [C. Pinhanez](#), N. Alvarado, M. Chan, E. Dejesus*CHI'03 extended abstracts on Human factors in computing systems*, pp. 802--803, 2003[Creating touch-screens anywhere with interactive projected displays](#) [C. Pinhanez](#), R. Kjeldsen, L. Tang, A. Levas, [M. Podlaseck](#), N. Sukaviriya, G. Pingali*Proceedings of the eleventh ACM international conference on Multimedia*, pp. 460--461, 2003

[Embedding interactions in a retail store environment: The design and lessons learned](#)

N. Sukaviriya, [M. Podlaseck](#), R. Kjeldsen, A. Levas, G. Pingali, [C. Pinhanez](#)

Proc. of the Ninth IFIP International Conference on Human-Computer Interaction, Zurich, Switzerland, 2003

[Steerable interfaces for pervasive computing spaces](#)

G. Pingali, [C. Pinhanez](#), A. Levas, R. Kjeldsen, [M. Podlaseck](#), H. Chen, N. Sukaviriya

Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, PerCom, pp. 31

[Fostering a symbiotic handheld environment](#)

M. Raghunath, C. Narayanaswami, [C. Pinhanez](#)

Computer 36(9), 56--65, IEEE Computer Society Press, 2003

2002

[HyperMask --- projecting a talking head onto a real object](#)

T. Yotsukura, S. Morishima, F. Nielsen, K. Binsted, [C. Pinhanez](#)

The Visual Computer 18(2), 111-120, 2002

Abstract computer graphics, ibm, watson, computer graphics images, artificial neural network, computer vision, cc artificial intelligence

[It/I: a theater play featuring an autonomous computer character](#)

[Claudio S. Pinhanez](#), Aaron F. Bobick

Presence: Teleoperators & Virtual Environments 11(5), 536-548, 2002

Abstract scripting language, virtual machine, narrative, multimedia, computer science

[" It/I": A Theater Play Featuring an Autonomous Computer Character](#)

[C.S. Pinhanez](#), A.F. Bobick

Presence-Teleoperators and Virtual Environments 11(5), 536--548, Cambridge, MA: MIT Press, c1992-, 2002

[HyperMask--projecting a talking head onto a real object](#)

T. Yotsukura, S. Morishima, F. Nielsen, K. Binsted, [C. Pinhanez](#)

The Visual Computer 18(2), 111--120, Springer, 2002

[That's entertainment! Designing streaming, multimedia web experiences](#)

C.M. Karat, J. Karat, J. Vergo, [C. Pinhanez](#), D. Riecken, T. Cofino

International Journal of Human-Computer Interaction 14(3-4), 369--384, Taylor & Francis, 2002

[Ubiquitous interactive graphics](#)

[C. Pinhanez](#), R. Kjeldsen, T. Levas, G. Pingali, [M. Podlaseck](#), P. Chou

IBM Research Report RC22495 (W0205-143), 2002

[Creating ubiquitous interactive games using everywhere displays projectors](#)

[C. Pinhanez](#)

Proc. of the International Workshop on Entertainment Computing (IWEC'02), pp. 149--156, 2002

[User-Following Displays](#)

G. Pingali, [C. Pinhanez](#), A. Levas, F. Kjeldsen, [M. Podlaseck](#)

Proc. of the IEEE International Conference on Multimedia and Expo 2002 (ICME'02), pp. 845-848

[Interacting with Steerable Projected Displays](#)

R. Kjeldsen, [C. Pinhanez](#), G. Pingali, J. Hartman, A. Levas, [M. Podlaseck](#).

Proc. of the 5th International Conference on Automatic Face and Gesture Recognition (FG'02), pp. 402--407, 2002

[BlueSpace: personalizing workspace through awareness and adaptability](#)

Jennifer Lai, Anthony Levas, Paul Chou, [Claudio Pinhanez](#), Marisa Viveros

International Journal of Human-Computer Studies 57(5), 415--428, Elsevier, 2002

2001

["Less clicking, more watching": an option for entertainment on the web?](#)

[Claudio Pinhanez](#), Clare Marie Karat, John Karat, John Vergo, Renee Arora

CHI 01 Extended Abstracts on Human Factors in Computing Systems, pp. 347-348, 2001

Abstract web design, mashup, entertainment, world wide web, multimedia, narrative, computer science

Hyper Mask-Projecting Virtual Face onto a Moving Real Object

S. Morishima, T. Yotsukura, F. Nielsen, K. Binsted, [C. Pinhanez](#)

Proc. of Eurographics' 01, 2001

[Can Web entertainment be passive](#)

[C. Pinhanez](#), C.M. Karat, J. Vergo, J. Karat, R. Arora, D. Riecken, T. Cofino

Proc., 2001

[Augmenting reality with projected interactive displays](#)

[C. Pinhanez](#)

Proceedings from Virtual and Augmented Architecture, 93--100, 2001

[Less clicking, more watching: Results of the iterative design and evaluation of entertaining web exper](#)

C.M. Karat, [C. Pinhanez](#), J. Karat, R. Arora, J. Vergo

Proceedings of IFIP TC, pp. 455--463, 2001

[Less clicking, more watching: an option for entertainment on the web?](#)

[C. Pinhanez](#), C.M. Karat, J. Karat, J. Vergo, R. Arora

CHI'01 extended abstracts on Human factors in computing systems, pp. 347--348, 2001

[Transforming Surfaces into Touch-Screens](#)

[C. Pinhanez](#), R. Kjeldsen, A. Levas, G. Pingali, J. Hartman, V. Kwatra, M. Podlaseck, P. Chou
submitted to CHI, Citeseer, 2001

[" Less Clicking, More Watching": Results from the User-Centered Design of a Multi-Institutional Web](#)

J. Vergo, C.M. Karat, J. Karat, [C. Pinhanez](#), R. Arora, T. Cofino, D. Riecken, [M. Podlaseck](#)
2001 - ERIC, ERIC

[BlueSpace: Creating a personalized and context-aware workspace](#)

Paul Chou, Marco Gruteser, Jennifer Lai, Anthony Levas, Scott McFaddin, [Claudio Pinhanez](#), Marisa Viveros, Danny W
Relat\orio t\ecnico1, 10--75, 2001

[Using a steerable projector and a camera to transform surfaces into interactive displays](#)

[C. Pinhanez](#)
CHI'01 extended abstracts on Human factors in computing systems, pp. 369--370, 2001

[The everywhere displays projector: A device to create ubiquitous graphical interfaces](#)

[C. Pinhanez](#)
Ubicomp 2001: Ubiquitous Computing, pp. 315--331

2000

[HyperMask: Talking Head Projected Onto Real Objects](#)

S. Morishima, T. Yotsukura, K. Binsted, F. Nielsen, [C. Pinhanez](#)
Proc. of Multimedia Modeling (MMM'00), 2000

[Perceptual user interfaces: the KidsRoom](#)

A.F. Bobick, S.S. Intille, J.W. Davis, F. Baird, [C.S. Pinhanez](#), L.W. Campbell, Y.A. Ivanov, A. Sch\utte, A. Wilson
Communications of the ACM 43(3), 60--61, ACM, 2000

[Physically interactive story environments](#)

[CS Pinhanez](#), JW Davis, S. Intille, MP Johnson, AD Wilson, AF Bobick, B. Blumberg
IBM Systems Journal 39(3.4), 438--455, IBM, 2000

1998

[Scripting method based on temporal intervals for designing interactive systems](#)

K. Mase, [CS Pinhanez](#), AF Bobick
Trans. of IPSJ 39(5), 1403--1413, 1998

1997

[Pnf propagation and the detection of actions described by temporal intervals](#) 

[C.S. Pinhanez](#), A.F. Bobick

Proc. of the DARPA Image Understanding Workshop, 1997

1995

[Scripts in machine understanding of image sequences](#) 

[C. Pinhanez](#), A. Bobick

AAAI Fall Symposium Series on Computational Models for Integrating Language and Vision, 1995

Year Unknown

[Design Methods for Personified Interfaces](#) 

Pinhanez, Claudio S

researchgate.net, 0

[Abstract](#)

[Supporting Remote and Simultaneous Participation in Meetings](#) 

[C. Pinhanez](#)

fxpal.com, 0

Share this page



Projects and Groups

- [Conversational Intelligence Group in IBM Research - Brazil](#)
- [IBM Research @ CHI 2021](#)

Technical Areas

- [Human Computer Interaction and Data Visualization](#)
 - [Knowledge](#)
 - [Mobile Computing](#)
 - [Natural Language Processing PIC at IBM Research AI](#)
 - [Reasoning](#)
 - [Speech](#)
-

Appendix B-3

Claudio Pinhanez - Professional Website

Claudio Pinhanez

IBM Research - Brazil
Manager, Social Data Analytics

Ph.D. in Media Arts and Sciences

M.I.T. Media Laboratory, 1999

Research interests:

Social Computing,
Service Science, HCI,
Ubiquitous Computing,
Computer Vision, AI.

E-mail: [claudio at pinhanez dot com](mailto:claudio@pinhanez.com)
Twitter: [@cinhanez](https://twitter.com/cinhanez)
Linkedin: br.linkedin.com/in/pinhanez/
Lattes: lattes.cnpq.br/5774595361715876



© Claudio Pinhanez, 2004

Last update: Jan 04, 2016

Appendix B-4

IBM Research Report

Ubiquitous Interactive Graphics

**Claudio S. Pinhanez, Frederik C. Kjeldsen, Anthony Levas,
Gopal S. Pingali, Mark E. Podlaseck, Paul B. Chou**

IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598



Research Division

Almaden - Austin - Beijing - Delhi - Haifa - India - T. J. Watson - Tokyo - Zurich

LIMITED DISTRIBUTION NOTICE: This report has been submitted for publication outside of IBM and will probably be copyrighted if accepted for publication. It has been issued as a Research Report for early dissemination of its contents. In view of the transfer of copyright to the outside publisher, its distribution outside of IBM prior to publication should be limited to peer communications and specific requests. After outside publication, requests should be filled only by reprints or legally obtained copies of the article (e.g., payment of royalties). Copies may be requested from IBM T. J. Watson Research Center, P. O. Box 218, Yorktown Heights, NY 10598 USA (email: reports@us.ibm.com). Some reports are available on the internet at <http://domino.watson.ibm.com/library/CyberDig.nsf/home>.

Amazon Ex. 1008.0147

Ubiquitous Interactive Graphics

Claudio Pinhanez, Rick Kjeldsen, Gopal Pingali, Anthony Levas, Mark Podlaseck, Paul Chou

IBM Research, T.J. Watson

pinhanez@us.ibm.com

Abstract

This paper presents a device that creates ubiquitous interactive graphics in a real world environment without the need for pre-wiring surfaces or requiring users to carry or wear any special devices. Called the *Everywhere Displays projector*, or *ED-projector*, this device employs *steerability* as the key principle to achieve the promise of ubiquitous computing and augmented reality. The ED-projector uses an LCD projector with a steerable mirror to project graphics onto any pre-calibrated surface, while warping the projected image to correct for oblique distortion. Three different methods to determine this warping function are given. To sense user interaction with the projected image, the ED-projector employs a pan/tilt/zoom camera and a motion-based computer vision system to track the user's hand and touch-like gestures. We propose a system software structure that ties together steering, display, and sensing. Three implemented applications demonstrate the notion of ubiquitously "painting" the real world with interactive graphics, in the context of a futuristic office, an augmented assembly task, and a ubiquitous computer game. The observation of hundreds of users performing the assembly task suggests that the traditional desktop paradigm is inadequate for ubiquitous interaction and that new conceptual widgets and interaction paradigms must be developed.

Keywords: ubiquitous computing; augmented reality; gesture recognition; interactive projections; ubiquitous games.

1 INTRODUCTION

The extraordinary success of computer graphics technology in recent years has been responsible for the creation of amazing new worlds in movies, interactive computer games, and virtual reality (VR) environments. However, this very success poses the question of when the real world, where we live and work, is going to be as rich with imagery and information as movies and games today.

This paper describes technology that can be used to ubiquitously augment, correct, and enhance reality. Although the ultimate goal is to create interactive pixels anywhere in a space as seamlessly as it happens in movies (thanks to CG special effects), this particular work focus on how to "paint" interactive graphics on the surfaces of an environment without requiring any kind of fixed display devices, special wiring of surfaces or the use of encumbering headsets.

The system described here is based on a new device, the *Everywhere Displays projector* (*ED-projector*), composed of a steerable projection system and a pan/tilt camera. A prototype of the projection system has been built using a pan/tilt mirror that directs imagery generated by a standard LCD projector. To capture user interaction with the projected graphics, the images gathered by the camera are processed using computer vision techniques. In the current prototype, it is possible to determine the user's hand position in reference to a surface and to detect simple, touch-like

gestures, effectively mapping the user's hand gestures into mouse-like events.

For instance, we have developed a computer game, called "Fro...og!", where a computer generated image of a frog initially appears "sitting" on the side of a file cabinet (see Figure 1). If a player tries to catch it, the vision system detects the player's hand approaching and then shows an image of frog jumping (see Figure 1). This is followed by moving the image of the frog to another surface, for instance the top of a table by appropriately steering the mirror. All this happens in an environment under normal lighting conditions, ultimately creating the illusion that the frog jumps around the space to avoid being caught by the player.

Our goals, however, go beyond transforming the real world into a magical entertaining place. In particular, we envision the creation of seamless interfaces to devices on surface of walls and everyday objects such as tables, chairs, couches, doors, cabinets, and carpets. Currently, only very "special" surfaces such as TVs, monitors, and projected screens support the creation of computer graphics, and usually require external devices such as mice and joysticks to make them interactive. In our paradigm, any surface is regarded as generic support for interactive graphics. By projecting imagery and detecting user interaction with a camera, it is possible to transform every surface into a virtual touch-screen.

These ideas have been explored in three application prototypes



Figure 1. Trying to catch a projected frog.

described in the paper. The first is an attentive/interactive office space. The second is in the context of augmenting the environment of an assembly task by projecting interactive graphics directly onto the surfaces of the objects involved. Ultimately we aim to realize Mark Weiser's paradigm of ubiquitous computing (see [1]) without having to change the fabric of the world by embedding touch-screens on every surface. The third application is the ubiquitous computer game "Fro...og!".

This paper starts by comparing our approach and solutions with previous work, followed by a description of the technology used in the prototypes we have built. Although initial ideas on the ED-projector have been reported in a previous publication [2], this paper presents the device comprehensively including mathematical formulation, computer vision based interaction, and three implemented applications. Specifically, this paper describes the mathematical formulation of the method for correcting oblique distortion and presents two new calibration methods for the first time. Further, in this paper we describe an integrated, working computer vision system that we implemented to track hands and detect touches, as well as a way to structure the interface to application software. While [2] elaborates only on conceptual mock-ups, this paper describes three implemented applications using the ED-projector, followed by a discussion of how novice users — in one case, more than 600 of them — have experienced and interacted with the application prototype. We conclude by examining other applications that can take special advantage of the ED-projector and discuss the challenges these new interfaces and applications create for HCI research.

2 RELATED WORK

In his seminal paper that defined the field of ubiquitous computing, Weiser [1] envisioned computers embedded everywhere — ubiquitous and transparent. In his view, interaction with computers would be accomplished both through direct mechanisms linking sensors and actuators and by the widespread presence of cheap interactive screens (referred as "pads") embedded on the surfaces of walls, furniture, and objects. Since the publication of that paper, different devices and methods to implement the functionality envisioned for Weiser's pads have been proposed. In general, those devices and methods can be divided into four categories: head-mounted, portable, embedded, and projection-based.

A good survey of the current research on head-mounted devices and their uses in ubiquitous computing is provided by Barfield and Caudell [3], describing both wearable computers and the more traditional approach of off-body computer power. We see two main challenges in this approach. First, in order to really connect to the physical world, it is necessary to precisely register the position and attitude of the head of the user with the environment in real time. An examination of the state-of-art of the research in the area shows that this is currently achievable only in very constrained environments (see, for example, Raskar et al. [4]).

The second challenge with head-mounted methods for ubiquitous displays is that they do not work well in social contexts and collaborative tasks. If multiple users are looking at the same surface (for instance, a whiteboard), it is necessary to render exactly the same graphics on each user's head-mounted display simultaneously. This creates considerable difficulties in terms of accommodating different resolutions, rendering speeds, etc. of different brands and models of head-mounted displays. In practice, it is very hard to guarantee that all participants are seeing exactly the same images.

The second method being studied to create ubiquitous displays utilizes portable devices such as laptops, PDAs, and mobile

phones. Like head-mounted displays, portable devices do not work well in social and collaboration tasks (except in very unusual conditions such as in multi-player games [4]), and are a nuisance to carry and power.

The third method involves embedding screens or similar interaction devices into the objects themselves, as proposed initially by Weiser [1] and more recently by Ishii [5]. However, even if screens became extremely cheap, this approach requires a widespread change in the way everyday objects are manufactured and installed. Power and network access (or wireless bandwidth) has to be provided to even the simplest object in a home or workplace. In other words, this approach requires a drastic change of the very fabric of the real world.

Our approach follows the fourth method of creating ubiquitous displays that is based on projection devices. Interestingly, the use of projectors to change and augment reality was pioneered by media artists such as Tony Oursler [6] and Michael Naimark [7]. The use of projectors to access computers have been initially proposed by Bolt [8] and Wellner [9]. Recently, more radical applications of projectors have been proposed by Morishima et al. [10] that enhance the face of an actor, and by Raskar et al. [11] to change the color and texture of real objects such as vases and architectural models. Although our system does not currently track moving surfaces as in the former example, or project onto 3D objects as the latter, it can clearly incorporate both features in future versions.

Deviceless interaction with projections was pioneered by Krueger [12], using camera-based systems to detect the users' hand and body. The fact that most vision techniques put strict constraints on lighting conditions and on the background surfaces has restricted the use of these systems to entertainment applications such as Keays and McNeil's *metaField* [13]. To avoid lighting constraints, Omojola et al. [14], Ju [15], and the RED project at Xerox Parc [16], among others, embedded sensors on the surface where the images are projected. As in the embedded approach described before, the widespread adoption of sensor-loaded surfaces requires a dramatic change in the objects of the real world. Underkoffler and Ishii [17] and Rekimoto [18] used visual tags to detect the position and the identity of the objects on a surface and to overlay the interaction based on these attributes. To avoid the need of tagging all objects in an environment as in those two cases, a possible option is the use of object recognition techniques (such as those described by Schiele and Crowley [19]). However, those techniques currently work only in very constrained situations. We believe, though, that incorporating object recognition capabilities into our system is an important direction to follow, and we plan to do so for specific applications.

Vision-based gesture recognition has received a great deal of attention in recent years [20], but only a little of that work has attended to the problems of gesture recognition in the context of a projected display. Because the projected image drastically changes the appearance of the hand, most approaches to hand detection are based on either background subtraction or motion (frame-frame differencing). Our experience agrees with that of von-Hardenberg and Berard [21] in that background subtraction does not give reliable results in the presence of strong projected images.

Gesture-based interaction methods tend to fall into one of two categories: either variations of a point-and-click paradigm [21, 22] or application-specific pose or motion gestures [23]. Both these approaches have disadvantages. The point-and-click approach is often not well suited to hand pointing due to limited resolution and the lack of a natural "click". On the other hand, a proliferation of application-specific gestures will make for a complex gestural interface that a user will have to learn and execute largely without feedback, as discussed by Kjeldsen in [24].

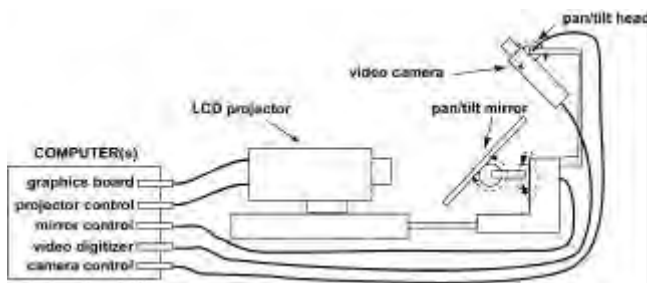


Figure 2. Diagram of the components of the ED-projector.



Figure 3. Prototype of the ED-projector: a) ED-projector without the camera mounted; b) a detail of a complete prototype, including the pan/tilt camera.

A major difference between our approach and the cited work on the use of projection-based devices to create ubiquitous interactive displays is *steerability*. By directing the projection and vision subsystems to a relevant position, it is possible to create new “interactive” surfaces in a space without any hardware installation, addition, or change. Unlike Raskar et al. [25], who use mirrors to increase the resolution of multiple-projector installations, we steer the image from a projector to realize interactive displays everywhere in the workspace. By doing so, we opportunistically use the surfaces of the objects present in the environment. We believe that steerability is a key feature that creates a real alternative to realize the promises of ubiquitous computing. However, it has constraints and technical challenges of its own, such as obstruction, oblique distortion, and surface texture interference. These issues are discussed in the next section where solutions to some of these problems are proposed.

3 THE ED-PROJECTOR

Our approach to ubiquitous interactive graphics uses a device — the *Everywhere Displays projector* (the *ED-projector*) — that can steer graphic displays onto any surface and can also sense user interaction with these steered graphical displays. The ED-projector is composed of an LCD projector, a computer-controlled pan/tilt mirror, and a pan/tilt/zoom camera. The projector is connected to the display output of a host computer that also controls the mirror and the camera (see Figure 2). The video output of the camera is sent to a digitizer board in the host computer where the imagery is processed, as described later.

Figure 3 shows a prototype of the ED-projector built with off-the-shelf components — a rotating mirror used in theatrical/discotheque lighting, a standard LCD projector, and a pan/tilt camera. The projector’s light can be directed anywhere

within a cone defined by approximately 60 degrees in the vertical axis and 230 degrees in the horizontal axis. The camera has a maximum vertical viewing range of 88 degrees and a horizontal range of 249 degrees, and is thus able to view the projected displays as they are steered around.

In general, a projector needs to be able to project a white pattern approximately 10 times brighter than its surroundings to create the illusion of contrast. Our current prototype employs a 3000 lumen projector, which we have found to be sufficiently bright in typical home and office lighting conditions to provide good contrast in the projected displays. Of course, the perceived brightness and contrast is heavily dependent on the color and material of the projected surface. Although our best results involve projection on white surfaces, we have also obtained good results projecting on carpets, black objects, and even translucent backlit surfaces. However, highly textured surfaces such as natural wood and surfaces with high reflectance are, in general, less suited for projected displays.

The three main issues with the ED-projector are 1) correction of the projected image to compensate for distortion due to oblique projection and surface characteristics; 2) sensing the user interactions with the projected displays; and 3) a system software structure that enables applications to easily define interactive display surfaces. These issues are discussed in detail in the rest of this section.

3.1 Correcting the Projected Image

Standard LCD projectors are designed to project light in a direction orthogonal to the projected surface. In the ED-projector, the pan/tilt mirror can deflect the light to multiple surfaces in the room, but most of the times the projected surface is oblique to the direction of projection. The result is a distorted image.

To correct the oblique projection distortion, we pre-warp the image to be projected to compensate for the distortion. It is always possible to correct the projected image, as long as the projected surface does not have a geometry that occludes the cone of projection (see [2] for details). However, the process of warping the image normally uses only a fraction of the projector’s image plane, causing loss of resolution. We will discuss this issue later; first, we address the problem of determining the warping function that corrects the oblique distortion.

Although there are many different ways to determine and implement the warping function, it is important to use a method that allows fast computation. We choose to implement it using 3D computer graphics hardware to texture map the images to be displayed onto a mesh in 3D space. If the mesh is appropriately warped and/or positioned in 3D space and the virtual camera correctly positioned, the projection of the resulting rendered image is free of distortion.

The basic structure of the problem, in the case of planar surfaces, is illustrated in Figure 4. We start by defining the *application plane*, a coordinate system where images to be projected are positioned within the area of a rectangular mesh M , and the *projector image plane*, the coordinate system associated with the LCD matrix of the projector. We also define a *projected surface* to be the real world surface on to which images are projected. In addition, we have the camera image plane, the coordinate system associated with the CCD matrix of the camera.

The goal is to define a transformation H from the application plane into the projector image plane so the transformation that maps the support mesh M into the projected surface is the identity I , and, therefore, appears free of distortion from an observer orthogonal to the projected surface. Since the process of projecting the image can be modeled as a projective transformation P (as

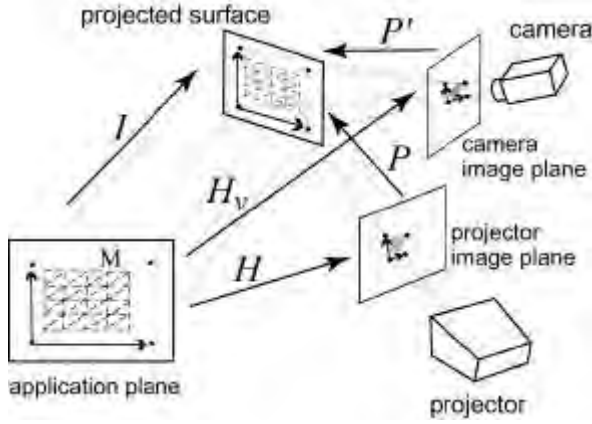


Figure 4. The transformations between the application plane, the projector's image plane, the camera's image plane and the projected surface.

observed in [26, 27]), we can represent the relationships among the planes as matrices in planar homogeneous coordinates, $I = PH$. Since H and P are projective transformations and therefore invertible, we obtain:

$$H = P^{-1}$$

Similarly if P' is the projective transformation between the camera image plane and the projected surface, and H_v is the transformation from the application plane to the camera image plane, we have $I = P'H_v$. Although there is no easy direct way to compute the projection function P , we developed three different methods to determine the actual value of H for a given surface. Of course, H and H_v are different for each planar surface in an environment.

1st method: Using an Equivalent 3D Virtual CG World

This method, first described in [2], is based on the fact that, geometrically, projection is the inverse of camera viewing. So, it is possible to model the projection by simulating the camera view by rotating, scaling, and viewing the support mesh M in 3D space.

Given the n mesh points in 2D projective space, $m_i = [m_i^x \ m_i^y \ 1]$, $1 \leq i \leq n$, let M be the matrix of all n points, $M = [m_1^T m_2^T \dots m_n^T]$. This mesh can be "positioned" in the XY plane of the 3D space by multiplying it by the matrix

$$K = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad KM = \begin{bmatrix} m_1^x & m_2^x & \dots & m_n^x \\ m_1^y & m_2^y & \dots & m_n^y \\ 0 & 0 & \dots & 0 \\ 1 & 1 & \dots & 1 \end{bmatrix}$$

We can then simulate the process of projecting the mesh as the inverse of the process of "viewing" the appropriately positioned mesh in 3D space. That is,

$$P^{-1}M = P_f R_x R_y R_z S_x S_y T_x T_y R_z K M$$

where R_x, R_y, R_z are rotation matrixes in 3D homogeneous coordinates, S_x, S_y are scale matrixes, T_x, T_y are translation matrixes, and P_f is the 3 4 projective matrix with focus point f . Since $H = P^{-1}$, we obtain

$$H = P_f R_x R_y R_z S_x S_y T_x T_y R_z K$$

Notice that all the matrices that compose H , with the exception of K , correspond to standard projection parameters of computer graphics boards. Therefore, to implement this method it is only necessary to determine these individual parameters, load them into

the hardware, and command the graphics board to render the scene.

To determine the projection parameters for a given surface we start by texture mapping a calibration pattern to the mesh M . Starting off with default values for the projection parameters that render the mesh on the center of the projector image plane without warping (i.e., $H \approx I$), we iteratively change each of the projection parameters until the projection of the pattern on the projected surface is observed to be identical to the original mesh. We then record these parameters and reload them whenever the ED-projector is aimed at that surface. Although conceptually simple, this iterative process of determining the right projection parameters is typically very time consuming.

2nd method: Manual Determination of Corresponding Points

This and the following method are based on the fact that projective transformations can be completely determined if the corresponding coordinates of four points in each 2D projective space are known (as shown by Faugeras [28]).

In this method, four points $a_i = [a_i^x \ a_i^y \ 1]$, $i = 1, 2, 3, 4$, are directly rendered on the projector image plane. We then iteratively move these points over the projector image space with a pointing device such as a mouse until their projections overlap the location of four known points on the projected surface, $b_i = [b_i^x \ b_i^y \ 1]$, $i = 1, 2, 3, 4$. For greater accuracy, we can choose these four points to be the corners of the mesh M .

It is easy to see that if the projected points are aligned with the projected surface points then $PA = B$, where $A = [a_1^T \ a_2^T \ a_3^T \ a_4^T]$ and $B = [b_1^T \ b_2^T \ b_3^T \ b_4^T]$. By computing the pseudo-inverse,

$$PA = B \rightarrow PAA^T = BA^T \rightarrow P = BA^T (AA^T)^{-1}$$

Since $H = P^{-1}$, we obtain $H = AA^T (BA^T)^{-1}$. To implement this procedure, we simply construct a warped mesh M' in 3D projective space by computing $M' = KHM$, texture map the image to be projected on it, and use the graphics engine to render it in real time.

As we see, the calibration procedure of this method is quite simple and fast, requiring only a few minutes for the manual alignment of the four projected points. Although similar to the method proposed by [29], here the use of the 3D mesh M and the computer graphics hardware yields much faster rendering speeds.

3rd method: Using a Camera and a Paper Pattern

In this method, we employ the ED-projector camera so we can avoid any manual calibration procedure. In [30], Yang et al. have proposed a method to automatically determine the warping function using a projector/camera pair. Unfortunately, their approach requires precise 3D calibration of the relative attitude of the camera to the projector and, in practice, a considerable baseline distance between them. In particular, it does not work with the ED-projector that aligns the projector and the camera axis as much as possible to avoid user's obstruction of the camera's view.

Instead, we propose here a method that employs a paper pattern positioned on the surface to be calibrated so it is aligned with the desired coordinate system. No relative 3D or 2D calibration is necessary, although we assume that the camera has been positioned in such a way that it has a complete view of the projected surface. Under these conditions, we can define a projective transformation P' between the camera image and the projected surface (see Figure 4). This transformation can be determined automatically by processing an image of the pattern and extracting the camera image coordinates of the corners, $c_i = [c_i^x \ c_i^y \ 1]$, $i = 1, 2, 3, 4$. Assuming the coordinates of the

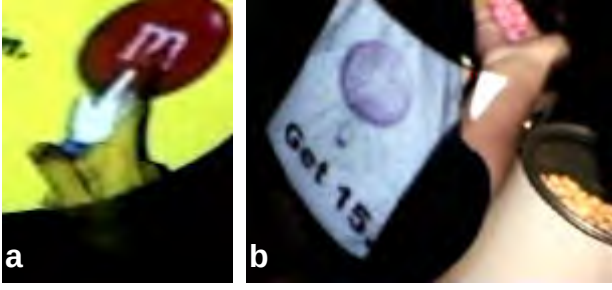


Figure 5. Difficulties faced by the vision system: a) the user's hand is partially hidden by the strong projection; b) the interactive surface is occluded by the user's shoulder as he moves to gather objects from inside a bucket.



Figure 6. Detecting a touch: a) camera view of an interaction with a bucket; b) image difference data; c) overlay of search region (square), button active area (circle), and the fingertip template shown at the pointing location.

pattern corners to be identical to the mesh M in application space, $b_i = [b_i^x \ b_i^y \ 1]$, $i = 1, 2, 3, 4$, we obtain $P'C = B$, and computing the pseudo-inverse,

$$P' = BC^T(CC^T)^{-1}$$

To complete the computation of the transformation H , it is necessary to determine the correspondence between the projector and the camera image planes. This can be done by projecting a pattern with four non-collinear points rendered on the projector image plane $d_i = [d_i^x \ d_i^y \ 1]$, $i = 1, 2, 3, 4$, and obtaining four points on the projected surface, $e_i = [e_i^x \ e_i^y \ 1]$, $i = 1, 2, 3, 4$. Then, by image processing the image of the pattern, it is possible to determine the coordinates of its four corners on the camera image plane, $f_i = [f_i^x \ f_i^y \ 1]$, $i = 1, 2, 3, 4$.

Since $E = [e_1^T \ e_2^T \ e_3^T \ e_4^T]^T$ corresponds both to the projection of $D = [d_1^T \ d_2^T \ d_3^T \ d_4^T]^T$ by the projector, $PD = E$, and to the captured image of the pattern, $P'F = E$, we obtain $PD = P'F$. Computing the pseudo-inverse, we obtain $P = P'FD^T(DD^T)^{-1}$ and therefore

$$H = DD^T(P'FD^T)^{-1}$$

As in the previous method, this transformation is implemented by computing a warped mesh $M' = KHM$.

One problem with all these three distortion-correcting methods is that they project displays with resolution lower than the projector's resolution. The distortion-correction process normally fits an irregular quadrangle into the 4:3 viewing area of typical displays. A considerable amount of display area, and thus resolution, is lost in the process. We have employed standard 1024x768 XVGA projectors in our prototypes. Due to the loss of display area created by the distortion correction process, the final display resolution has approximately corresponded to VGA, i.e., 640x480 pixels.

3.2 Sensing Interaction on Any Surface

To support the ability to interact on any surface we use a pan/tilt camera feeding custom image processing and gesture recognition engines. This process has been detailed in another publication [31], but will be outlined here for completeness.

Image processing in this domain is both simplified and made more difficult by the strong light of the projected image. It is simplified in the sense that the lighting environment is very consistent because the dynamic range of the projection swamps all but major changes in ambient light. It is made more difficult because the appearance of the user's hand can be severely distorted as it passes through the projected image (see Figure 5.a).

Although parts of the hand can be hard to detect in a static image, by looking at the differences between adjacent images in the video stream, we can still detect the shape of a moving object as shown in Figure 6.a and b. The use of motion for segmentation instead of background subtraction techniques (such as [32]) also makes it easier to accommodate the variability in the projected image caused by the texture of the surface it is projected on. This motion mask forms the basis of subsequent image understanding.

To provide interaction in a flexible and consistent manner we use the notion of *interface widgets*. Each widget generates computer or application events in response to a particular user motion in a region of the image. We currently use three types of widgets: *buttons* that respond to touch-like motions; *sliders*, 1-dimensional tracking regions, and *touch pads*, which are 2-dimensional tracking regions.

These three widget types rely on tracking the user's fingertip. Convolution template matching is used to find fingertip candidates in the motion mask image described above. The candidates are evaluated using spatial clustering and domain specific heuristics, such as the distance from the user, to determine the best estimate of whether there is a fingertip in the image and where it is currently pointing. The trajectory created by this point over a sequence of video frames is examined by each widget to determine when it should trigger an event.

Touch pad widgets smooth the fingertip trajectory and map it into the touch pad coordinate system. Motion events are only generated when the fingertip is within the touch pad's borders. Sliders are simply a 1-dimensional version of touch pads.

Buttons, however, are more complex widgets. What buttons attempt to detect is when a surface in the environment is touched. Inferring the distance of the fingertip to the surface using the shadow proved to be difficult and unreliable. Fortunately, people tend to interact with real-life buttons by moving in a characteristic forward/pause/backward pattern, and this seems to carry over well to projected buttons. Our implementation of buttons takes advantage of this movement pattern.

To determine when a user has touched a button we look for a roughly linear sequence of fingertip locations that persists for 250ms or longer, with speed within a set range, that either ends or begins with a pause (i.e. the fingertip template did not find a match in the motion mask). A button generates its event when the pause in such a trajectory occurs within it (see Figure 6.c). In other words, we are looking for the user's fingertip to travel in one direction for $\frac{1}{4}$ second and stop within the button, or start moving from the button and move in a consistent direction for $\frac{1}{4}$ second. These approximations for the motion of a fingertip in the moments before and after it touches a button seem to be universal: almost every button touch either starts or ends that way. Interestingly though, many button touches only have one or the other of these motions. Of course, adjustments are needed in the case of people with tremors or other motor disabilities.

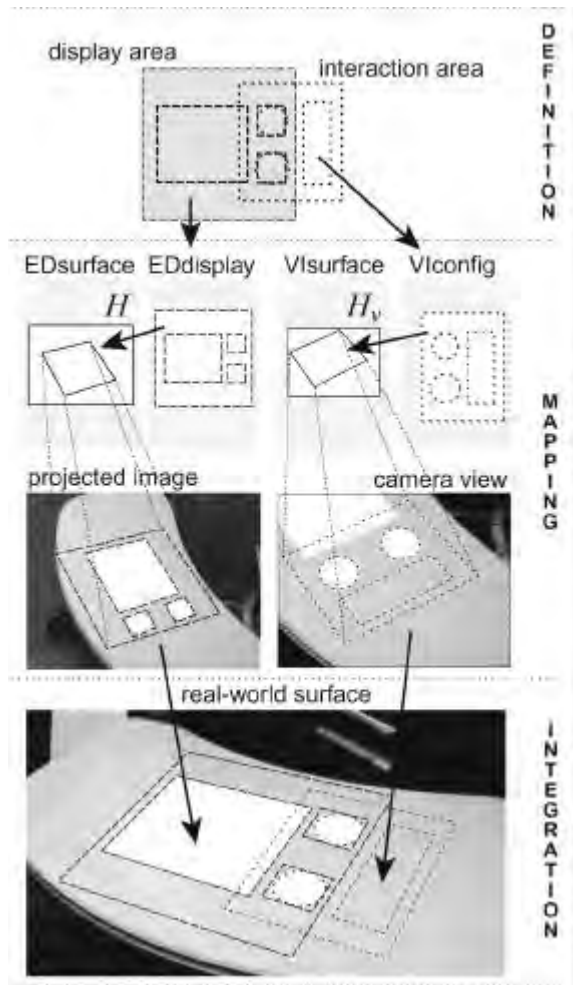


Figure 7. Process of creating an interactive surface on a table.

One of the interesting problems in this domain occurs when users occlude buttons as shown in Figure 5.b. Our experiments with users, even novices, have shown that they naturally avoid obstructing the projection if they are attending to it, but they are quite bad at staying out of the way when they are working with the physical world instead. Buttons have inherent resistance to triggering in these types of situations because any motion must pass several tests before it generates an event. It must resemble a fingertip for several frames; during that time it has to travel with a consistent direction and velocity and stop within a button. Such a sequence sometimes occur as a result of a chance combination of errors, contributing to a somewhat higher false positive rate for some buttons, particularly those which the user occludes for a period of time rather than just passes through.

By using both the shape of the moving region as well as an analysis of its trajectory, this touch detection method gives far better accuracy than using either one alone. Because it relies directly on image data, rather than building an explicit model of the hand, it is computationally efficient enough to run at high frame rates on standard hardware.

Widgets can be combined in various ways in order to perform a complete interaction with the user. However, configurations of widgets are defined independently of the surface where they are used. That is, if an application interface using a set of buttons is moved to different surfaces, it is not necessary to define each button for each surface. Instead, the location of a widget within the

camera image is determined at run time by appropriately mapping the configuration onto the camera image plane. In this case the surface refers to a set of camera parameters. These parameters, like those of the projector, have to be calibrated in advance. In particular, the calibration process has to determine the transformation H_v between the widget coordinate system and the camera image (as shown in Figure 4). Calibration of surfaces is performed using methods analogous to the 2nd and 3rd methods of projection calibration, although information about the expected location of the user, and the size of their hand, has to be obtained during an extra, special calibration step (see [31] for details).

3.3 Integrating Displaying and Sensing

The two preceding sections described the techniques and subsystems that provide the display (output) and sensing (input) functionality of the ED-projector. This section discusses the integration of these separate subsystems in a manner that affords application development and human interaction. One of our primary objectives here is to enable a specific user interaction to occur on any calibrated surface in a room without requiring the customization of this interaction for every surface. This leads to a clear separation of the abstract definition of an interaction from the actual surface upon which it occurs.

Our widgets are constructed using the standard Model View Controller (MVC) design pattern structure [33]. In our case, the projector subsystem is responsible for rendering the *view*, while the vision subsystem provides the *controller* input events. The *definition* layer of Figure 7 shows the separation of the display and interaction areas that are responsible for the definition of the view and the interaction respectively. Display and interaction will often overlap as shown by the square overlapping buttons on the right of the *display area*. Provision is also made, however, for allowing interaction to occur in a region where nothing is displayed, as shown in the large rectangular region on the right of the *interaction area*. Notice that the display area here corresponds to the mesh M of the application plane described before (see also Figure 4).

Every interaction surface in a physical space is named and initially calibrated by the projection and vision subsystems yielding the transforms H and H_v respectively. H corrects the projected image while H_v provides the transformation to the camera image, for a specific surface as described in the prior sections. Figure 7 demonstrates the flow of definitions for display images and interaction data (widget data) from the *definition* layer to the *mapping* layer. During run-time, the system software automatically maps (through the warping functions H and H_v) the *definition* layer data into the projected image and the camera view associated with the current surface. The result is a projected image that generates application events when user gestures occur in the areas defined in the camera view, as shown in the *integration* layer of Figure 7.

4 PAINTING WITH INTERACTIVE GRAPHICS

To better understand how to use projected interactive graphics and how users relate to them, we have designed and implemented three application prototypes using the ED-projector. In the first application, an office space is augmented with dynamic wallpapers, visual notifications of messages, and computer access from multiple surfaces. The second prototype explores an assembly task where the instructions are projected directly onto the surfaces of the objects involved in the task. The third prototype is the “catch the frog” game described in the introduction to this paper.

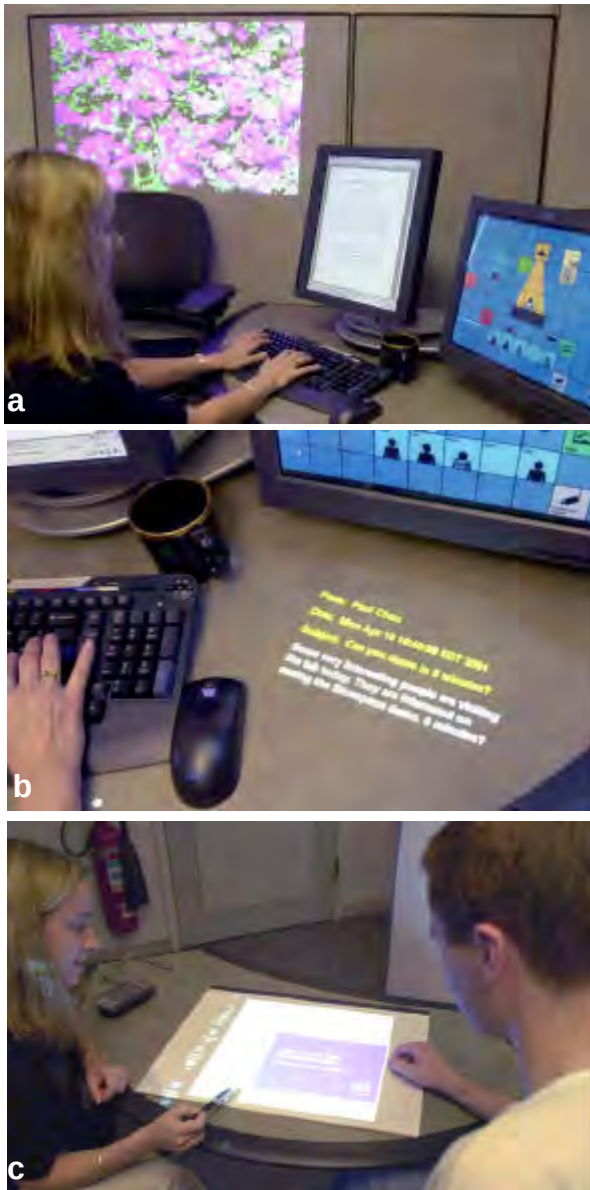


Figure 8. Uses of the ED-projector in an office application: a) to create dynamic wallpapers; b) to provide silent notification of e-mail; and c) to create a computer desktop on a working surface.

4.1 An Office Application

The first prototype application in which we have deployed the ED-projector is a futuristic personalized and context-aware workspace. The workspace has an 8' by 10' footprint and incorporates several sensors for measuring ambient lighting, temperature, humidity and noise level. Users and visitors to the workspace wear active badges that facilitate presence detection and identification. The desk chair is equipped with a pressure sensor connected to a wireless micro-controller, which detects if a person is sitting in the chair. This environment incorporates an ED-projector, strategically located to maximize coverage of the workspace. In this setting, we have used a non-interactive version of the ED-projector to facilitate greater personalization of the environment and to free its users from the confines of the desktop displays.

The presence of an ED-projector in the office tranfigures the space into an extended graphical display and enables the user to have a favorite painting on a wall, receive important notifications on surfaces that catch their attention, or use a tabletop as a display when collaborating with a colleague (see also the video figure). Figure 8.a shows an example of a dynamic wallpaper created by the ED-projector. The user can similarly have a virtual window that is connected to a live webcam looking at her favorite scenery (effectively making it a window office) or monitoring her child at home. The ED-projector can automatically switch to a neutral, less personal image when a co-worker comes to the office, based on the active badge.

Figure 8.b illustrates an urgent notification to the user appearing on a surface close to the user's current location. Here, the smart office detects that the user is seated at the desk and steers the notification right to the surface of the desk. By determining the position of the user, the system is assured that the user will see the urgent message without having to rely on the traditional sound alerts which tend to be disruptive to people working in a shared environment.

Another example of the utility of a steerable display is seen in Figure 8.c where two colleagues have created a computer desktop display on the working surface of their choice. Here, interactive capabilities, provided by a camera system, are planned to be incorporated to facilitate collaborative work.

4.2 Augmenting an Assembly Task

The ED-projector can be seen as a device for augmenting reality – a beam of light that can add information on to real world objects by highlighting them, adding text or images, flashing etc. For example, the ED-projector can show the user an object they are searching for, or indicate the procedure for an assembly task, and even where to place a part in an assembly.

We developed a prototype assembly task to highlight these themes of augmented reality and interaction anywhere. In this prototype application, which was experienced by hundreds of novice users in a major technical exhibition, the object to be assembled is a picture made of M&M's (multi-colored sugar-coated chocolates) where each M&M is regarded as a pixel of the picture.

The theme of interactivity anywhere is demonstrated several times in this prototype application (see also the video figure). Figure 9.a shows a white fabric on a table transformed into an interactive menu for color selection. The user simply touches the color of his choice to select it. Figure 9.b shows a "clickable" button appearing on a paint bucket that contains the M&M's of the selected color. Figure 9.b and c highlight the augmented reality and user assistance aspects of the system. In Figure 9.b the system highlights the bucket that contains the M&M's of the selected color, and also provides additional information on how many M&M's to pick. In Figure 9.c, the system points the exact places on the picture board where the M&M's should be placed. Figure 9.d shows another combination of interaction and augmented reality: the user moves her hand over the surface to interactively reveal the rest of the picture being built.

More than 650 people participated in this augmented assembly over 6 days and completed 4 different M&M pictures. Considering that the vision system and the projection system were integrated for the first time only in the weeks preceding the exhibition, the combined system worked remarkably well. A sample with 130 consecutive users with 621 button touch events (touching gestures or false detections) yielded correct detection of touching gestures in 81% of the events, with 12% false negatives and 7% false positives. If the bucket events are excluded from the count, the

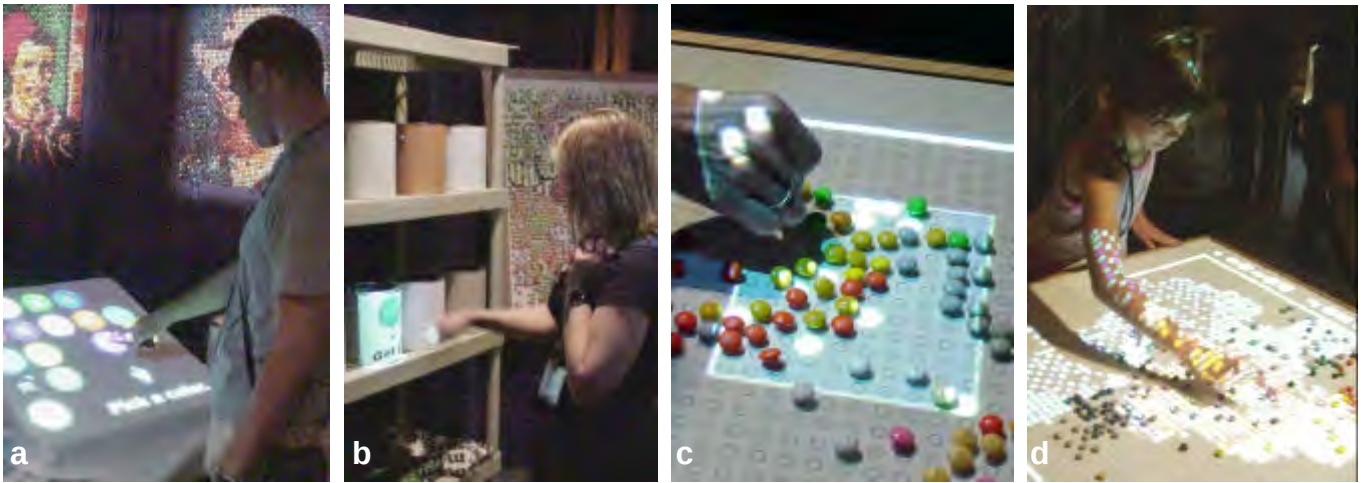


Figure 9. The M&M picture assembly task: a) choosing a color; b) “clicking” a bucket after getting the M&Ms; c) placing the M&Ms on the highlighted areas; and d) “finger painting” to reveal the complete picture.

performance exceeds 89%. A detailed account of the reasons why buckets yielded more errors is given in [31].

The experiment also revealed that users are readily able to relate to the notion of steerable interactive graphics. We also observed the value of the projection creating a shared experience between the users and the people observing the demo: in many cases, users learned the interaction process by first watching other users; and frequently, observers helped users suggesting what to do next. As mentioned before, a major advantage of projection over goggles is precisely this “social” nature of projection.

An interesting observation is that our users were less hesitant to interact with the fabric-covered interactive color selection menu (Figure 9.a) and the M&M placement board (Figure 9.c,d) than with the paint buckets (Figure 9.b). When they went to a bucket to pick up M&M’s, many of them did not realize that they had to click the big “done” button on it (see also Figure 6.a). We had expected that, since this was the third occasion for touch interaction in the demo, users would quite naturally press the button on the bucket as instructed. Most of the time, this was not the case. To help the users in that situation we often gave them verbal instructions such as “Click the bucket,” aiming to make clearer that there was, in fact, a button on the bucket.

4.3 “Fro...og!”

The ED-projector’s unique characteristic of being able to quickly move the projected image to different surfaces is being explored in a computer game for young children. In this game, called “Fro...og!”, an image of a frog is projected on a surface. Whenever the child tries to touch the frog, the camera detects her approach, and the frog “jumps” to another surface, that is, the projector moves the image to a wall or the top of a chair.

Figure 1 shows two pictures of a user playing with the current prototype of “Fro...og!”. The basic idea is to define surfaces in the environment where the animated frog can jump to. Whenever the user’s hand approaches the frog, the next surface is randomly chosen, and an image of the frog jumping in that direction is briefly shown. After landing on the new surface, a simple animation of the frog jumping up and down is projected, so it becomes easier to find the frog in the environment.

In the next development phase of this game we plan to integrate it to a system that tracks the position of multiple children in the room, so the frog can jump to surfaces which are far from most players. Tracking is also particularly important to avoid

projecting the frog directly onto the players which, in this scenario, can easily break the illusion of the game.

5 DISCUSSION

Having shown how we implemented a working prototype of an ED-projector and deployed it in three different applications, it is interesting to explore other possible applications and scenarios, and discuss how to design efficient interfaces for them.

5.1 Other Applications

We see two classes of applications of the ED-projector to create ubiquitous interactive graphics. The first involves the creation of computer interfaces anywhere, while the second deals with augmented reality in a variety of public and private spaces.

In the first class of applications, we regard the ED-projector as a generic input/output device that can replace, in many situations, current displays and interactive devices by creating computer desktop-like interactive displays on non-tethered surfaces. For example, a desktop application can be projected directly on any surface, as previously suggested by Wellner [9] and as demonstrated in subsection 4.1. Unlike the interactive whiteboard described in [34], the projected application can be easily moved around the room, for instance, from a whiteboard on the wall to the top of a desk for more detailed reading. Similarly, the ED-projector can be used to bring information to the physical location where it is used or needed. For example, a database application managing reports can be projected on top of the file cabinet with hard copies of the reports.

In the second class of applications the ED-projector can augment reality by pointing to physical objects, showing connections among them, and projecting patterns to indicate movement or change in the real world. The prototype assembly task described in subsection 4.2 is a typical example of this class of applications. Another example applies to a library in which directions are provided to a user looking for a book by projecting arrows on the floor, walls and shelves, and finally by highlighting the desired book.

Steerable graphics produced by the ED-projector can also aid in collaborative work by providing large displays on convenient surfaces, and allowing easy reconfiguration of meeting spaces for different functions and teamwork styles, in similar ways to the office project described in subsection 4.1.

The ED-projector can also be used to provide computer and information access in environments where traditional displays are

difficult to install, secure, or maybe unsafe to operate, such as public spaces and areas subject to harsh environmental conditions. The device also permits an interactive display to be brought to the proximity of a user, eliminating the need for the user to approach a traditional, fixed display. In particular, the ED-projector can facilitate the access and use of computers by people with locomotive disabilities. For instance, it can project an interactive display on a hospital bed sheet without requiring the patient to contact any device.

We also see the ED-projector as a potential enabler of a new generation of games where interaction happens not in the virtual world but in the physical world such as the “Fro...og!” game described above. Unlike games based on phones and portable computers [4], the use of the ED-projector provides high-resolution displays where characters and fantastic objects can move from surface to surface, creating a game that surrounds the players. Using video input, various user actions can be detected depending on the game needs, for example, hand gestures, body movements, foot action, and/or facial expressions. Above all, a single ED-projector installed on the ceiling of a living room or an arcade, using just software commands, can completely reconfigure the space to create different games according to the interests, age, and motor skills of the players.

The scenarios examined so far are chiefly concerned in creating ways for humans to command computers and/or obtain information. However, we see the projecting of interactive graphics ubiquitously as a way to provide computer agents the ability to “reach and touch” the real world and the people that inhabit it. By using projected patterns, it is possible to make computers not only interact with people, but act and coordinate them. For instance, a computer agent can control people in a line by projecting red lines and green arrows around the people on the line. A steerable projector system is, thus, able to create a “body”, composed of light, for an agent. From this perspective, we may be creating an interface to the real world for computers.

5.2 Human Interface Issues

As foreseen by Weiser [1], ubiquitous interactive graphics do not seem to follow the standard paradigm of interaction based on the desktop metaphor. In our M&M demonstration, we had the opportunity to observe more than 650 novice users experiencing a scenario of true ubiquitous computing, in the context of an assembly task. On one side, we observed that they quite naturally accepted the idea of augmenting reality with information directly projected on objects. Moreover, we saw they naturally change their body’s position to minimize problems like the shadow of their arms and body over the board. People seemed not only to accept projected graphics naturally but also to compensate for its limitations.

However, when trying to click a button, the users almost always applied pressure onto the projected surface with their fingers. Also, whenever a button touch was not detected by the vision system, the most common reaction was to press again with more pressure and, if still not successful, using the whole hand instead of a finger. It seems that people expect buttons, even virtual buttons, to react to pressure and not to touch, and to fail if enough pressure is not applied on them.

This observation clearly illustrates the limitations of the hand-as-a-mouse paradigm used in the M&M demo. Although it is easier for users to relate the ubiquitous experience to their everyday contact with desktop interfaces, the metaphor proved to be misleading to the true nature of the interface. Unlike a mouse that has appropriate haptic feedback for clicking gestures, touching a projected button is devoid of feedback movement.

Similar observations prompt us to believe that a new paradigm for interaction with projected screens must be developed, with new kinds of widgets, tuned for the natural capabilities and limitations of projection and gesture-based interaction, just as the point and click paradigm is tuned for the characteristics of a physical mouse. Touch-screen-like interactions do translate better to the projected interface domain. Although our widgets are not based on surface contact, the widgets we have found useful so far (buttons, scroll bars and touch pads) are all based on motions people naturally make when touching.

One of the exciting aspects of this work is that it is an excellent environment to explore more adventurous gesture-based interaction styles. By giving the user feedback as to what gestural interactions the system is expecting and where in the environment they may be performed, the cognitive load on the user can be reduced. Without such feedback a user would have to memorize what amounts to a set of complex dance steps to simply use a gesture-based interface.

This investigation of new human-computer interface paradigms is even more important in the case of multi-user interaction. Unlike desktops, wearable computers, or PDAs, projected interactive graphics naturally allow many users to simultaneously interact with an application. Although the applications we have developed so far are all single-user, it seems clear multi-user interaction will soon become an important area of research.

Another intriguing question is related to the different degrees of difficulty experienced by users when interacting with graphics projected on different supporting objects. Clicking a paint bucket seems to be harder to accept than clicking a table. Of course the design of our system and the simplicity of our experiment allow us neither to definitively state that the phenomenon does happen nor to draw conclusions about its causes. However, one hypothesis is that people seem to attribute different interactive capabilities to projected interactive graphics according to the nature of the surfaces themselves. Are the paint buckets “functionally fixed”, that is, if something looks and acts like a paint bucket, is it less likely to be perceived as something else – namely, a touch-screen? We are currently preparing a study to determine the validity of this hypothesis.

Finally, our research so far has addressed only visual input and output. A promising area of research seems to be the application of the concept to steerability of other stimuli such as sound, using, for instance, the audio spotlights described by Pompei [35]. What kind of interfaces and applications can be created by a full multimodal steerable system?

6 CONCLUSION

This paper describes a new method and device for creating ubiquitously interactive graphics on non-tethered surfaces. We have shown three different methods to determine the warping function that corrects for oblique projection distortion and a computer vision system that tracks the user’s hand over the projected images and detects touch-like gestures. We have also implemented three applications that realize ubiquitous computing concepts such as everywhere computer access and augmented reality. Based on the experience of running the M&M demo with hundreds of novice users, we have also detected a need to re-examine the current interaction paradigms, and possibly, investigate and propose new ones.

We see this work as an important contribution to realize the vision of ubiquitous computing, without asking people to wear headsets or wiring the surfaces of everyday objects. However, we do not advocate projection-based solutions as a panacea for ubiquitous applications: in many cases, it is more desirable and efficient to use head-mounted displays. A good analogy is to

compare headsets and projectors to Walkmans and stereo sound systems. Walkmans are unsurpassable as a way to provide music when people are moving around or in public spaces such as trains and airplanes. They are very effective in creating a personal experience, albeit not able to create social interaction. Similarly, we keep loving stereo systems because they render music to a whole space, nurturing a collective experience that is attached to the space itself. We see similar advantages and limitations when comparing headsets and steerable projectors as a way to render interactive graphics onto the real world.

Acknowledgements

“M&M” is a registered trademark of Mars, Inc. “Walkman” is a registered trademark of Sony Corporation.

References

1. Weiser, M., *The Computer for the Twenty-First Century*. Scientific American, 1991. **265**(3): p. 94-100.
2. Pinhanez, C. *The Everywhere Displays Projector: A Device to Create Ubiquitous Graphical Interfaces*. in *Proc. of Ubiquitous Computing 2001 (UbiComp'01)*. 2001. Atlanta, Georgia.
3. Barfield, W. and T. Caudell, eds. *Fundamentals of Wearable Computers and Augmented Reality*. 2001, Lawrence Erlbaum Assoc. 728.
4. Raskar, R., et al. *The Office of the Future: A Unified Approach to Image-Based Modeling and Spatially Immersive Displays*. in *Proc. of SIGGRAPH'98*. 1998. Orlando, Florida.
5. Ishii, H. and B. Ullmer. *Tangible Bits: Towards Seamless Interfaces between People, Bits, and Atoms*. in *Proc. of CHI'97*. 1997. Atlanta, Georgia.
6. Bellour, R., et al., *Tony Oursler*. 2001: Ediciones Poligrafa. 208.
7. Naimark, M., *Spatial Correspondence in Motion Picture Display*. SPIE Optics in Entertainment II, 1984. **462**: p. 78-81.
8. Bolt, R., *Put That There: Voice and Gesture at the Graphics Interface*. ACM Computer Graphics, 1980. **14**(3): p. 262-270.
9. Wellner, P., *Interacting with Paper on the DigitalDesk*. Communications of the ACM, 1993. **36**(7).
10. Morishima, S., et al. *HyperMask: Talking Head Projected Onto Real Objects*. in *Proc. of Multimedia Modeling (MMM'00)*. 2000: World Scientific.
11. Raskar, R., et al. *Shader Lamps: Animating Real Objects with Image-Based Illumination*. in *Proc. of 12th Eurographics Workshop on Rendering*. 2001. London, England.
12. Krueger, M.W., *Artificial Reality II*. 1990: Addison-Wesley.
13. Keays, B. and R. Macneil. *metaField Maze*. in *Proc. of SIGGRAPH'99*. 1999. Los Angeles, California.
14. Omojola, O., et al., *An Installation of Interactive Furniture*. IBM Systems Journal, 2000. **39**(3&4): p. 861-879.
15. Ju, W. *Origami Desk*. in *Proc. of SIGGRAPH'01 - Conference Abstracts and Applications*. 2001. Los Angeles, California.
16. Back, M., et al., *Designing Interactive Reading Experiences for a Museum Exhibition*. IEEE Computer Magazine, 2001. **34**(1): p. 1-8.
17. Underkoffler, J., B. Ullmer, and H. Ishii. *Emancipated Pixels: Real-World Graphics in the Luminous Room*. in *Proc. of SIGGRAPH'99*. 1999. Los Angeles, CA.
18. Rekimoto, J. *A Multiple Device Approach for Supporting Whiteboard-based Interactions*. in *Proc. of CHI'98*. 1998. Los Angeles, CA.
19. Schiele, B. and J.L. Crowley, *Recognition without Correspondence using Multidimensional Receptive Field Histograms*. IJCV, 2000. **36**(1): p. 31-52.
20. Wu, Y. and T. Huang, *Vision-Based Gesture Recognition: A Review*. Lecture Notes in Artificial Intelligence, 1999. **1739**.
21. Hardenberg, C.v. and F. Berard. *Bare-hand human-computer interaction*. in *Proc. of Workshop on Perceptive User Interfaces, PUT'01*. 2001. Orlando, Florida.
22. Quek, F., T. Mysliwiec, and M. Zhao. *Finger mouse: A freehand pointing interface*. in *Proc. of International Workshop on Automatic Face- and Gesture-Recognition*. 1995. Zurich, Switzerland.
23. Segen, J. *GestureVR: Vision-Based 3D Hand Interface for Spatial Interaction*. in *Proc. of ACM Multimedia Conference*. 1998. Bristol, England.
24. Kjeldsen, F., *Visual Recognition of Hand Gesture as a Practical Interface Modality*. 1997, Columbia University: New York, New York.
25. Raskar, R., et al. *Multi-Projector Displays Using Camera-Based Registration*. in *Proc. of IEEE Visualization'99*. 1999. San Francisco, CA.
26. Pinhanez, C., F. Nielsen, and K. Binsted. *Projecting Computer Graphics on Moving Surfaces: A Simple Calibration and Tracking Method*. in *Proc. of SIGGRAPH'99*. 1999. Los Angeles, California.
27. Raskar, R. *Oblique Projector Rendering on Planar Surfaces for a Tracked User*. in *Proc. of SIGGRAPH'99*. 1999. Los Angeles, California.
28. Faugeras, O., *Three-Dimensional Computer Vision: A Geometric Viewpoint*. 1993, Cambridge, Massachusetts: The MIT Press.
29. Sukthankar, R., R. Stockton, and M. Mullin. *Smarter Presentations: Exploiting Homography in Camera-Projector Systems*. in *Proc. of ICCV'01*. 2001. Vancouver, Canada.
30. Yang, R. and G. Welch. *Automatic and Continuous Projector Display Surface Calibration Using Every-Day Imagery*. in *Proc. of 9th International Conf. in Central Europe in Computer Graphics, Visualization, and Computer Vision*. 2001. Plzen, Czech Republic.
31. Kjeldsen, R., et al. *Interacting with Steerable Projected Displays*. in *Submitted to Face and Gesture'02*. 2002.
32. Wren, C., et al., *Pfinder: Real-Time Tracking of the Human Body*. IEEE Trans. Pattern Analysis and Machine Intelligence, 1997. **19**(7): p. 780-785.
33. Buschmann, F., et al., *A System of Patterns - Pattern Oriented Software Architecture*. 1996, New York, New York: John Wiley & Sons.
34. Crowley, J.L., J. Coutaz, and F. Berard, *Things that See*. Communications of the ACM, 2000. **43**(3): p. 54-64.
35. Pompei, F.J. *The Use of Airborne Ultrasonics for Generating Audible Sound Beams*. in *Proc. of 105th Audio Engineering Society Convention*. 1998. San Francisco, CA.

Appendix B-5

Ubiquitous interactive graphics

☐ Search within citing articles

Projector-guided painting

M Flagg, JM Rehg - *Proceedings of the 19th annual ACM symposium on ...*, 2006 - [dl.acm.org](#)

This paper presents a novel interactive system for guiding artists to paint using traditional media and tools. The enabling technology is a multi-projector display capable of controlling ...

☆ Save  Cite Cited by 102 Related articles All 11 versions

Registration techniques for using imperfect and partially calibrated devices in planar multi-projector displays

E Bhasker, R Juang, A Majumder - *IEEE Transactions on ...*, 2007 - [ieeexplore.ieee.org](#)

Multi-projector displays today are automatically registered, both geometrically and photometrically, using cameras. Existing registration techniques assume pre-calibrated ...

☆ Save  Cite Cited by 75 Related articles All 14 versions

A personal projected display

M Ashdown, P Robinson - *Proceedings of the 12th annual ACM ...*, 2004 - [dl.acm.org](#)

User interfaces using windows, keyboard and mouse have been in use for over 30 years, but only offer limited facilities to the user. Conventional displays are small, at least compared ...

☆ Save  Cite Cited by 50 Related articles All 15 versions 

To frame or not to frame: The role and design of frameless displays in ubiquitous applications

C Pinhanez, M Podlasek - *... , UbiComp 2005, Tokyo, Japan, September 11 ...*, 2005 - Springer

A frameless display is a display with no perceptible boundaries; it appears to be embodied in the physical world. Frameless displays are created by projecting visual elements on a ...

☆ Save  Cite Cited by 50 Related articles All 15 versions

User-centric design of a vision system for interactive applications

S Borkowski, JL Crowley, J Letessier... - *... on Computer Vision ...*, 2006 - [ieeexplore.ieee.org](#)

Despite great promise of vision-based user interfaces, commercial employment of such systems remains marginal. Most vision-based interactive systems are one-time," proof of ...

☆ Save  Cite Cited by 18 Related articles All 12 versions

The dream of Peter Pan: user interaction arts using computer vision techniques

K Kim, W Kwon, E Choi, J Park, M Jeon, H Ju... - *Proceedings of the ...*, 2007 - [dl.acm.org](#)

In this paper, we present the user interaction arts in mixed reality environment using computer vision techniques. In most of the mixed reality environment, user has to wear ...

☆ Save  Cite Cited by 4 Related articles All 3 versions

Projection-based assistance for ultrasonographic diagnosis

A Prayitno - 2004 - [Citeseer](#)

In the modern age, medical examinations tend to use visual information about the internal condition of the human body for more accurate diagnosis. This information is especially ...

☆ Save  Cite Cited by 2 Related articles All 4 versions 

Oil painting assistance using projected light: Bridging the gap between digital and physical art

M Flagg, JM Rehg - 2005 - [smartech.gatech.edu](#)

This paper presents a novel interactive system for guiding artists to paint using traditional media and tools. The enabling technology is a multi-projector display capable of controlling ...

☆ Save  Cite Cited by 2 Related articles All 3 versions 

유비쿼터스 환경의 디스플레이를 위한 실시간 기하보정

경동욱, 임현규, 정기철 - 한국정보과학회 학술발표논문집, 2005 - dbpia.co.kr

최근 프로젝터를 기반으로 사용자가 원하는 위치에 영상을 제공하기 위한 많은 연구가 진행중에 있다. 프로젝터는 투사방향이 정확하지 않으면 영상왜곡이 발생되며, 영상왜곡은 ...

☆ Save  Cite Cited by 2 Related articles All 2 versions

실시간 보정을 지원하는 Everywhere Display

최현철, 경동욱, 한은정, 양종열, 정기철 - 한국정보과학회 학술발표 ..., 2006 - dbpia.co.kr

최근 유비쿼터스 환경에서 프로젝터를 기반으로 사용자가 원하는 위치에 영상을 제공하는 유비쿼터스 디스플레이 연구가 진행 중이다. 프로젝터는 투사방향에 따라 영상의 왜곡이 ...

☆ Save  Cite Related articles All 2 versions

Appendix B-6

Projector-Guided Painting

Matthew Flagg and James M. Rehg
GVU Center, College of Computing,
Georgia Institute of Technology, Atlanta, GA
{mflagg,rehg}@cc.gatech.edu

ABSTRACT

This paper presents a novel interactive system for guiding artists to paint using traditional media and tools. The enabling technology is a multi-projector display capable of controlling the appearance of an artist's canvas. This display-on-canvas guides the artist to construct the painting as a series of layers. Our process model for painting is based on classical techniques and was designed to address three main issues which are challenging to novices: (1) positioning and sizing elements on the canvas, (2) executing the brushstrokes to achieve a desired texture and (3) mixing pigments to make a target color. These challenges are addressed through a set of interaction modes. Preview and color selection modes enable the artist to focus on the current target layer by highlighting the areas of the canvas to be painted. Orientation mode displays brushstroke guidelines for the creation of desired brush texture. Color mixing mode guides the artist through the color mixing process with a user interface similar to a color wheel. These interaction modes allow a novice artist to focus on a series of manageable subtasks in executing a complex painting. Our system covers the gamut of the painting process from overall composition down to detailed brushwork. We present the results from a user study which quantify the benefit that our system can provide to a novice painter.

ACM Classification H5.2 [Information interfaces and presentation]: User Interfaces. - Graphical user interfaces.

General Terms Design, Human Factors, Algorithms

INTRODUCTION

In this paper, we present an interactive painting system to support novice artists in the creation of paintings using traditional media and tools. In contrast with other computer painting systems that provide tools for digital image synthesis, this work focuses on guiding users to paint according to classical techniques for creating *paintings as objects*, not just images. By employing multiple projectors to create an interactive display on the artist's canvas, our system is capable of augmenting a painting in progress with visualization features to assist with technical details such as brushstroke position, orientation and color mixing. We present a set of interaction modes designed to address common challenges for novices.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

UIST'06, October 15–18, 2006, Montreux, Switzerland.
Copyright 2006 ACM 1-59593-313-1/06/0010 ...\$5.00.

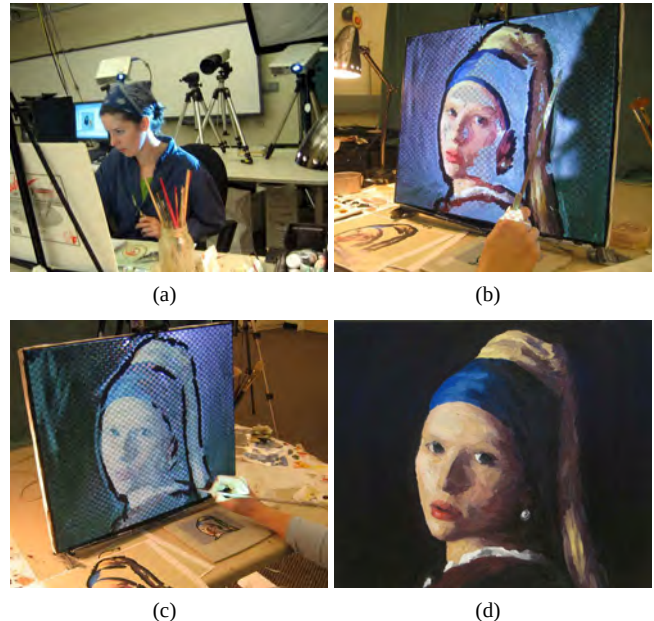


Figure 1: (a) Painting studio setup (b) Viewing a layer in *preview mode* (c) Painting the black regions of a layer in *color selection mode* (d) Final painting created using our system

Our system is part of a long history of tools for reflecting and capturing the perspective and shading of a scene as a first step in establishing the structure of a painting. Perhaps the most well-known of such tools is the *camera obscura* [22]. The modern descendant of the camera obscura is the *art projector*, a projector sold in most major art stores. The art projector is commonly used by professional painters to make accurate perspective sketches as a first step in copying originals. By replacing the art projector with video projectors and a camera, we propose using visual feedback to guide the user through the steps that follow toward creating a final desired painting.

Our system employs a multi-projector adaptive display for minimal intrusion on the traditional painter's environment in order to assist with painting technique. The display is formed on the canvas by aligning multiple overlapping front projectors. As shown in Figure 1(a), projectors are positioned on both sides of the artist to minimize the occlusion of the projected light by the artist's body. The design of our system is based on our previous work on shadow removal using redundant projection [23, 7]. Following a geometric calibration step, each pixel on the display may be illuminated by two rays of light. In addition to reducing shadows, this redundancy facilitates adaptation to the painted canvas.

An issue that arises in projecting onto a painted canvas is the need to photometrically adapt the projected image to compensate for the varying reflectance properties of the paint. We use visual feedback to incrementally adjust the projected light for each canvas pixel so as to minimize the difference between the observed and desired display color. This visual feedback approach has the advantage of avoiding the need for an explicit model of paint reflectance. Our domain is a challenging one for adaptation because the canvas is continuously changing throughout the painting process and the presence of wet paint results in specular surfaces. We note that projecting onto the canvas from behind is not an option when opaque paints such as oil and acrylic are used.

This paper makes the following contributions:

- A novel interactive system for guiding artists to paint using traditional methods and tools.
- A set of interaction modes for a canvas-centric interface which supports a traditional layer-based approach to painting.
- A novel method for creating virtual paint on a previously painted canvas, which allows artists to paint each layer without being distracted by the layers underneath.

We have conducted a user study which quantifies the benefits provided by our system over a standard art projector in assisting novice painters. We show that our system provides a significantly greater improvement in the confidence of novice painters and enables them to produce higher quality artifacts as measured by a panel of experts.

Related Work

While there is a long line of work in painterly rendering for digital images [11, 15, 8, 20, 13, 12] and the simulation of traditional media and tools for art [4, 3], to our knowledge there have only been a few systems for creating art as objects using physical tools. A robotic painting system named AARON [14] has been actively developed for three decades and its paintings have been featured in several museums and art galleries. The Everywhere Display (ED) demonstration at Siggraph in 2001 [17] guided users to create “paintings” on a table out of a collection of M&Ms. In contrast, our system is designed to support a fundamentally creative process within the framework of traditional painting practice.

Other researchers have used projector-camera systems to support spatially augmented reality [18]. Shader lamps [19], for example, used multiple projectors to illuminate complex models, such as the Taj Mahal, or animate static objects, such as toy cars. As an application of real painting augmentation, [24] use projectors for restoring colors in damaged paintings. In the work of [5], a film was applied to a painting to make the surface more amenable to displaying visualization information, such as x-ray images, directly on the canvas for display in art museums. A system described in [1] used a single portable projector to create displays that spanned multiple non-coplanar surfaces. Multiple projectors were used in [6] to boost the dynamic range of projected light in a view-dependent stereoscopic display that featured photometric adaptation to the projection surface. In contrast to these applications, the goal of our system is to support a *creative process* through the construction of a dynamic display

on the work of art as it is being developed. Our compensation approach is based on the method of [16], which addressed the problem of projecting onto photometrically non-uniform surfaces using feedback.

MOTIVATION AND GOALS

Existing tools that support the creation of works of art include books and manuals, demonstration videos, and art projectors. These tools have many deficiencies which limit their effectiveness. While books and manuals can demonstrate technique and illustrate the elements of painting style, it is difficult for the novice painter to bridge the gap between the printed page and the canvas. For example, painting manuals such as [2] demonstrate how to decompose a complex painting into a sequence of layers, but do not provide any direct support for the execution of each layer on the canvas. The execution of individual layers affects the overall composition in ways that are difficult for a novice painter to predict. It is also hard for novices to detect deviations between the painted layer and a reference picture.

In contrast to books and manuals, art projectors do provide direct support for the execution of the initial stages of a painting through their ability to project light directly onto the canvas. Art projectors are commonly used to trace or sketch onto the canvas the rough geometric forms that comprise a painting. Unfortunately, art projectors are unable to provide support to the artist throughout the entire painting process. They are also cumbersome to work with, as they lack the ability to calibrate and adapt automatically to the canvas and its contents.

The motivation for this work is the opportunity to leverage recent advances in projection and camera-based sensing technologies which make it possible to create a new projector-based painting aid. Our painting system has a number of novel features: it supports interaction directly on the canvas, it incorporates a model of the painting process, and it supports adaptation to an evolving painting through sensing and analysis of the painted content.

The target audience for our system is novices who would like to have an enjoyable painting experience while creating a high-quality work of art. Our system is designed to enable people who have never painted before to create a piece of artwork that they can be proud of. We accomplish this task by addressing, in the design of our system, the primary difficulties and pitfalls which novices encounter when working with oil paints. Note that we are not claiming to teach people how to paint or to provide a substitute for formal art instruction. Our system is part of a spectrum of painting aids which include traditional art projectors as well as instruction books and videos. Unlike these passive aids, however, our system is interactive and adaptive and incorporates a computational model of the painting process.

LAYER-BASED PROCESS MODEL FOR PAINTING

The versatility of oil and acrylic paint as an artistic medium has led to the development of highly refined processes and techniques for expressing a wide variety of artistic styles. These processes range from the quick freeform application of paint in impressionist styles to the painstaking 7-layer Flem-

ish method, a classically realist style with drying periods of months between each layer [21]. Coupled to these wide variations in the strategy for building up a painting is an equally wide range of methods for applying paint to canvas, from precisely controlled strokes or dots to the 'drip and splash' approach that characterized Jackson Pollock's work.

The design of our system is based on the following two assumptions about the process of painting in acrylics and oils:

- We assume that paintings will be planned out in advance of execution and decomposed into a series of layers that can be applied to the canvas in linear order (back to front).
- When painting each layer, we assume that the artist will want to apply paint to the canvas in an orderly, sequential manner and will base their palette on the color scheme for the layer and the desired amount of manipulation.

Under these assumptions, a painting consists of the execution of a series of layers. Therefore, our system adapts a layer-based representation. In painting each layer we assume that the artist will mix one or more desired paint colors, select an appropriate brush or knife, and load it with paint. The artist then chooses a region of the layer corresponding to the selected color, applies the paint, and manipulates it until the desired effect has been achieved. Thus a layer is painted region-by-region until it is complete. The interaction modes described on the next page are designed to support each of these tasks. While our approach can apply equally to acrylic paints with minimal modification, for the purpose of our discussion and experiments we will confine ourselves to painting in oil.

Our model of the painting process is consistent with a number of classical and popular painting styles. Traditional layer-based oil painting methods are characterized by long drying times, which can be accelerated by thinning the paint heavily with turpentine, as in the "fat over lean" approach [10]. In these layer-based methods, wet paint is applied on top of existing layers of dry paint. An alternative is the *Alla Prima* method of wet-into-wet, which was popularized by Impressionist painters like Edouard Manet. In this approach paints are applied in rapid succession without intervening drying times. The goal is to achieve decisive, clean strokes which remain visible in the final painting. Although the *Alla Prima* method is not limited to a layer-based painting approach, it is compatible with our layer-based computational representation. Brushstrokes will often follow a progression in both size and color as early layers tend to consist of broad general strokes, followed by a progression of detail. Cézanne's advice, cited in [10], was to "start with the broom and end with the needle!". Similarly, the more opaque light paints are frequently applied after the darks. For the purpose of testing and evaluating our painting system, we have adopted the *Alla Prima* method, as the shorter drying times facilitate experimentation and the immediacy of the technique is appealing to beginners.

The layers which our painting model requires can come from a variety of sources. One possibility is to capture the painting process of an experienced artist. In addition, layers can be scanned in from painting manuals and registered. We can also use painterly rendering techniques to create layers auto-

matically from photographs or 3-D models. One of the example paintings from our video submission uses the method of Hays and Essa [12] to generate layers from an input photograph, by rendering brushstrokes along contours which are generated by a multi-scale edge detector. Low frequency edges are rendered with wider brushstrokes and high frequencies are rendered with narrower ones.

CHALLENGES FACED BY NOVICE PAINTERS

Novice painters face a variety of challenges in executing a complex painting, including the difficulty of positioning and sizing the elements of the composition on the canvas, the challenge of executing the required brushstrokes, and the complexity of mixing paints to achieve the desired colors. For example, as described in [10], inexperienced painters often make the mistake of working on one small area of a painting until it is finished and then move on to the next area. This can result in a "confused and disjointed image because each area of tone and colour is separate and unrelated to its neighbors." Our adoption of a layer-based approach and the development of interaction modes which support the execution of individual layers addresses this concern.

The desired look of the painting is achieved through careful control of the brush, by regulating the pressure, movement, speed, and distance of each stroke. *Impasto*, for example, is a technique for producing a "fresh, vigorously textured look" [21] using thick, expressive brushstrokes. Smith [21] recognizes the difficulty of applying brushstrokes to generate the characteristic *impasto* texture in a uniform manner throughout a layer. This is an example of the larger class of problems novices may face at the brushstroke application level. Our system can provide guidance in the application of brushstrokes to create a desired effect.

Color mixing is one of the most difficult tasks faced by a novice painter. The proportions of paints which must be combined will vary dramatically from one shade of color to the next, and most beginning painters are unfamiliar with the subtleties of the subtractive color space that governs the paint mixing process [9]. In our observations, users can easily spend 20 minutes or longer in an attempt to reproduce a specified color for a particular painting. These long delays are a source of frustration, particularly when users must re-mix a given color after running out of it. Our system includes support for color mixing by automatically identifying the components that are required for a particular color and adaptively guiding the mixing process.

INTERACTION MODES

Our system is designed around a set of interaction modes that assist the artist in painting the ordered set of layers that comprise a finished painting. We support hands-free switching between interaction modes using a foot pedal (see the included video for example usage). Each interaction mode is realized by projecting light onto the surface of the painting itself and adapting the projection to existing paint on the canvas in creating the desired display. We describe our adaptive projected display in the following section. Our system provides five interaction modes to support the painting process:

- **Preview Mode** to display layers

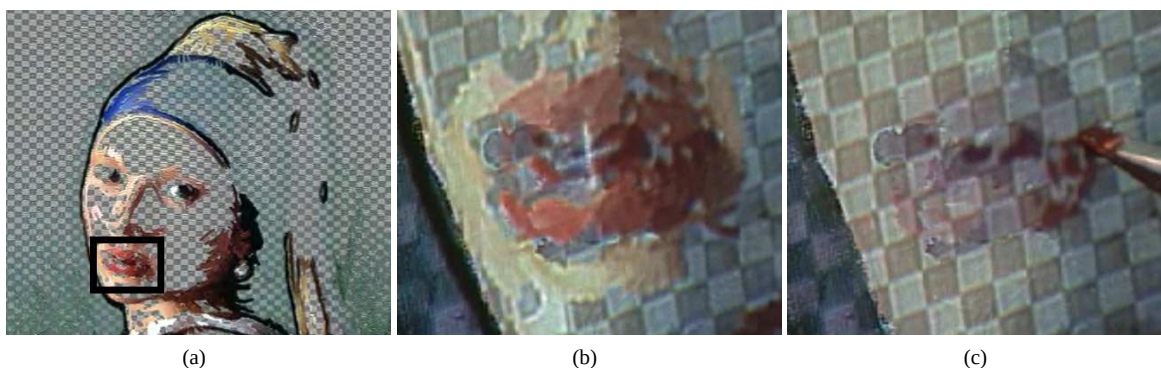


Figure 2: (a) Layer 4 out of 4 in preview mode, with highlighted region around lips. (b) Closeup of lip region. (c) Lip region following color selection. Note that all images are photographs of the actual canvas under compensation.

- **Color Selection Mode** to paint with specific colors
- **Blank Mode** to assess the state of the painting
- **Orientation Mode** to display brushstroke guidelines
- **Color Mixing Mode** to aid in creating the desired colors

Preview Mode: A visual guide for painting a given layer is provided by preview mode, which creates an overlay on the canvas that displays the current layer in progress while compensating for the underlying paint. The use of this mode is depicted in Figs. 1(b) and 2(a). In preview mode, areas of the canvas that are not to be painted in the current layer are hidden from the artist using a checkered pattern. This pattern is easily distinguished from the foreground elements of the layer. This reduces the chance that regions outside the current layer will distract the artist from a particular execution step.

Furthermore, the clear visibility of all layer regions and “in-visibility” of off-layer regions provided in preview mode helps the artist devise a strategy for selecting a color palette and set of brushes that are used to paint the layer. For example, it may be the case that each layer corresponds naturally to one specific brush size. It is easy to select a brush in preview mode by physically comparing brushes to their previewed strokes on the canvas.

Color selection mode: Another challenge in executing a particular layer is the efficient use of paint. An artist with a loaded brush would like to quickly identify regions of the canvas that need that particular color. To address this need, we designed a color selection mode that highlights all strokes of a chosen color in the current layer. The artist uses a Wacom tablet with a pen attached to the opposite end of the paintbrush to select a specific color in the layer. Figs. 1(c) and 2 illustrate the use of this mode. The white cross in Fig. 2(b) indicates the selection of a particular shade of red in the layer containing the girl’s lips. We employ a user-specified tolerance parameter to determine which parts of the layer match the selected color. The tolerance is adjusted interactively via the foot pedal, to expand or contract the color region of interest.

Following color selection, all other colors are hidden from view so that the artist can apply the color where it is needed. A comparison between Fig. 2(b) and 2(c) demonstrates that color selection mode effectively masks out regions in the

layer which do not correspond to the selected color. For areas like these where it is difficult to distinguish between various shades of red, the color selection tool is useful for clearly marking where paint should be added.

Blank mode: The artist can systematically complete each layer through the use of the preview and color selection modes. In order to assess the progress made in painting a layer, along with the current state of the painting, the artist may switch the projectors off at any time using blank mode. The included video shows the utilization of preview, color selection, and blank mode in completing a layer.

Orientation mode: While preview and color selection modes show the artist *where* to apply paint, orientation mode shows the artist *how* to apply it. A set of guidelines overlaid on the layer indicate the orientation, width, and extent of brushstrokes that should be applied to achieve a desired brushstroke texture. For example, long, linear guidelines would indicate the need for bold, expressive strokes, while a short hatching pattern would indicate the use of a dabbing technique. Guidelines can also be used to indicate directions for blending paints, as depicted in Figures 4(b) and 4(c). Orientation guidelines must be generated along with their associated layer images. We note that one benefit of using the painterly rendering method of Hays and Essa [12] to produce layers is that it automatically generates contours that can be used as orientation guidelines.

Color mixing mode: Support for the task of combining paints to achieve a desired color is provided by color mixing mode. Fig. 3 shows the interface, which depicts a standard color wheel along with a camera view of the artist’s palette. Using the Wacom tablet, the user identifies an area on the palette where paint is to be mixed. A desired color from the current layer can be specified via color selection mode. A precalibrated table maps the target color to a specific location on the color wheel, known as the target point. Labeled source points on the color wheel correspond to fully-saturated source pigments (labeled “18” and “24” in Figure 3).

The first step in mixing the target color is to identify the fully saturated colors that may be combined to make the target color. For any point on the color wheel, the system selects the two colors that lie on both the outer boundary of the color wheel and on a line connecting the center of the wheel to the

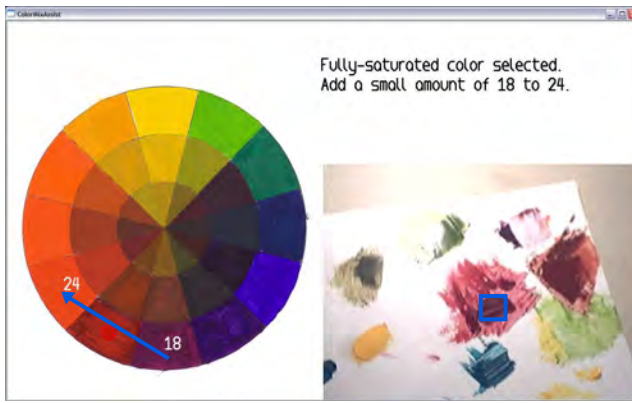


Figure 3: Interface for *color mixing mode*, which employs automated capture of and localization of palette colors on a color wheel to guide the paint mixing process. The blue arrow illustrates the movement of the current mixture color, which is the selected region in the camera image, to the desired target.

target color's location on the wheel. These two fully saturated components of the target color are complementary colors. Following identification of these two ingredient colors, the user may put a sample of the color on the palette directly from the tube or by means of another mixture process. The interface may be used to help mix these source ingredients if needed. After both ingredient colors have been placed on the palette, one color is gradually mixed with the other. Following each step of paint addition, the camera image is analyzed to identify the point on the model color wheel which corresponds to the mixture in progress. Estimation of the mixture point is accomplished using a pre-calibrated set of histograms obtained for a representative set of pigments. These points chart the geometric progress of moving the color along a gradient on the color wheel. The system guides the user to slowly add one ingredient color to the other to achieve the desired target point.

CREATING A DISPLAY ON A PAINTED CANVAS

We employ a system of two projectors and one camera to create an adaptive display directly on the canvas. The projector outputs and camera input can be automatically warped and aligned with the surface of the canvas using standard calibration methods [7].

Multi-Projector Adaptive Display: By placing the two projectors at an oblique angle to the canvas and overlapping and aligning their outputs, we gain the ability to illuminate each pixel on the canvas with two distinct sources of light. This has two benefits. First, by projecting simultaneously from both projectors we gain robustness to shadows and occlusions, as described in [7]. Second, by adapting with both projectors during compensation, we can increase the dynamic range of the display on the canvas. This is particularly useful in trying to adapt dark colors to make them lighter.

Photometric Adaptation: In order to control the appearance of the canvas, our system must account for the spatially varying reflectance of a painted canvas. Any image projected onto the canvas will be modulated by the reflectance of the paint and therefore create an undesired appearance as shown in Fig. 5(d). To achieve a desired appearance, we adapt a

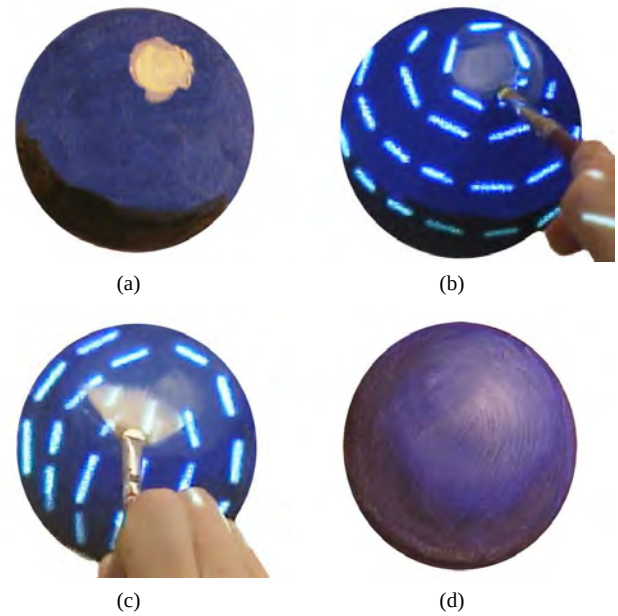


Figure 4: Painting a sphere using orientation mode. (a) Sphere with highlight and shadow; (b) Coarse blending with bristle brush in orientation mode; (c) Fine blending with fan brush in orientation mode; (d) Finished result.

feedback approach in which a captured camera image of the canvas is analyzed to create a compensation image for each projector to output.

Our compensation approach is based on the method of Nayar et. al. [16] for adapting the output of a single projector to a surface with varying reflectance. Our system employs multiple projectors to increase the dynamic range of the compensation method and supports the positioning of the projectors beside the artist instead of directly in front of the canvas. The algorithm works by projecting an initial input image and iteratively adjusting image intensities using visual feedback. Starting with all projectors projecting black, each projector iteratively reduces compensation error, the difference between desired and measured intensities for each pixel and channel independently, according to the proportional feedback law $P_{i+1} = k(I - C_i) + P_i$, where P_i is the compensation image for iteration i , I is the desired appearance image, C_i is the surface-aligned camera image for iteration i , and k is the gain. Once the error has dropped below a threshold tolerance, a completion mask is toggled for the finished pixel to avoid further adjustment of its compensation intensity. Adaptation halts after the change in error for all pixels has stabilized or a maximum number of iterations have occurred. Following adaptation, the compensation image for the current projector is fixed and adaptation begins for the next projector. Each subsequent projector reduces error for the remaining pixels that have not met the target intensities for each channel. Figure 5 illustrates the compensation process for 2 projectors positioned as in Figure 1(a). Clearly, the appearance of the canvas is closer to the desired image with two projectors than with one. These results establish the effectiveness of compensation in the case of oil paints.



Figure 6: 3 paintings made using our system by 3 painters who had never painted before.

RESULTS

We present the results of a formal user study of novice painters along with informal observations on the usage of our system by several subjects.

User Study

We conducted a formal user study to quantify the benefit that our system can provide to a novice painter, relative to a standard art projector. The study had two components. In the first component, a population of twenty subjects was asked to paint copies of the model painting illustrated in Fig. 8(b).¹ Each of the subjects was given a canvas containing an initial sketch of the desired shape (as produced by a standard art projector) and printouts of the model painting. Ten of the subjects served as a control group and produced their paintings without the use of our system. The other ten subjects had full access to our system.

Prior to their painting experience, each subject filled out a questionnaire in which they were asked to rate their confidence in their ability, on a scale of 1 to 5 where 1 means “not confident at all” and 5 means “very confident”, to paint a set of three paintings which were identified by an art expert to be of low, medium, and high difficulty.² The questionnaire also contained questions to assess whether the subject had any previous painting experience.³ (Subjects with prior experience were allowed to complete their paintings but one was dropped from the analysis due to experience.) Each subject was then given a brief orientation session and then asked to paint a short warm-up painting which took approximately 15 minutes.⁴ Each subject then completed the model painting, which took three hours on average.

Following their painting experience, the subjects filled out a second survey which again asked them to rate their ability to execute a different set of three paintings of equivalent levels of difficulty. In addition, subjects were asked to rate

their level of frustration, level of boredom, likelihood to paint again, assessment of the quality of their painting, and overall satisfaction with the experience, on a scale of 1 to 5.

We conducted an analysis of the self-assessed ability ratings for each subject population before and after their painting experience. The average prior confidence rating for the painting of Medium difficulty was 2.2 for the control group and 2.3 for the group that used our system. Following their painting experience, the average rating for the Medium difficulty painting dropped to 1.6 for the control group, and rose slightly to 2.5 for the group that used our system. We ran a t-test and Mann-Whitney-U nonparametric test for the null hypothesis that there was no significant difference in the dependent variables between the two groups. The difference in assessed ratings was found to be statistically significant (at level $p < 0.05$) using both tests. We conclude that our system has a quantifiable, positive benefit over what an art projector alone can provide in improving the confidence of novice users.⁵

In the second component of the study, a group of three experienced painting professors filled out two questionnaires to evaluate each of the twenty paintings completed by the novice users (see Fig. 8). The first questionnaire asked respondents to rate 6 attributes of painting quality on a scale of 1 to 7 and the second questionnaire asked the professors to rank their top ten paintings in terms of overall quality and overall similarity to the model painting.⁶ Fig. 7 reports the sum of the ranks of expert ratings for six survey attributes over the two populations of paintings, where 1 = “poor quality”, 4 = “medium quality” and 7 = “high quality.” The p-values for the Mann-Whitney-U test are shown for each attribute. A statistically significant difference (at $p < 0.05$) was observed for every attribute except “quality of color matching to original painting.” Furthermore, a majority of the top 10 paintings ranked by each professor were painted with the use of our system. These results serve to further validate our

¹An experienced artist created the barn image in two layers with associated orientation mode guidelines. The captured layers and guidelines were used by our system.

²The paintings used in our survey were as follows: Before painting, difficulty Low “Composition with Red, Blue, and Yellow” (Mondrian), Medium “The Lemon” (Manet), and High “Floor Scrapers” (Caillebotte). After painting, difficulty Low “Number 10” (Rothko), Medium “Dusk” (Monet), and High “The Music Lesson” (Vermeer).

³All questionnaires and raw data are available at <http://www.cc.gatech.edu/cpl/projects/painting>

⁴The warm-up session was conducted to help eliminate any negative effects of using basic painting tools for the first time.

⁵The average confidence ratings for the Low and High difficulty paintings did not show a statistically significant variation between the two groups. However the mean ratings for both attributes were higher for the group that used our system. Note that the model painting was designed to be of medium difficulty.

⁶In addition to ranking the paintings, the second questionnaire asked for open comments on standard attributes they have used in the classroom for comparing paintings and assessing their relative quality. These comments were found to validate our choice of survey attributes for the quality of the painting and its similarity to the model.

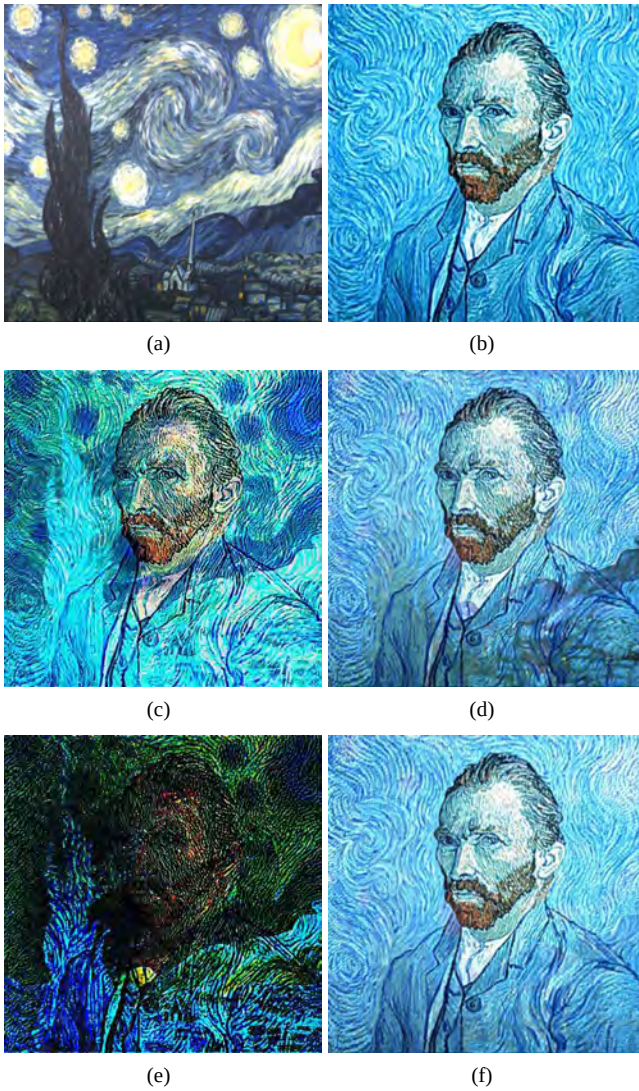


Figure 5: (a) Canvas before compensation; (b) Desired appearance; (c) Projector 1 compensation image; (d) Compensation result for projector 1 alone; (e) Projector 2 compensation image; (f) Final canvas appearance under the combination of projectors 1 and 2.

claim that our system can help novice painters to produce higher-quality artifacts than an art projector alone.

Usage Trends

Figure 6 shows three paintings created using our system by three novice artists who had never painted in oils before. Each painting took an average of 5 hours to create using our system. We observed the painting process for these artists and identified the following trends:

First, two of the subjects exhibited initial hesitation in applying the brushstrokes directly on top of the displayed virtual brushstrokes. For the two impressionist style paintings in Figure 6, for example, both artists applied the paint in very small and slow brushstrokes despite the implicit instruction to paint with quicker and more sweeping brushstrokes as visually suggested in preview, color selection and orientation modes.

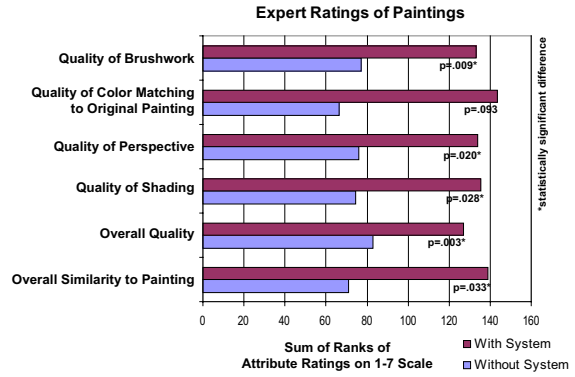


Figure 7: Aggregate quality ratings reported by three painting professors, for the two sets of paintings shown in Fig. 8.

Second, as the novices' hesitations faded and they switched between interaction modes more frequently, task execution speed increased. This indicates that they learned how to apply regions of color in a layer in a more effective manner and, as a result, became more confident in their ability to accomplish painting goals.

Third, as the paintings progressed, all three users switched to color selection mode more frequently. In one case, the artist repeatedly re-adjusted the color similarity tolerance for each color selection before applying a new color to the canvas. This enabled him to grow or shrink the layer segmentation until it appeared to conform to contiguous brushstroke areas as opposed to pieces of brushstrokes resulting from color variation across the brushstroke. This behavior suggests the need for a more sophisticated means of detecting and segmenting brushstrokes for color selection mode.

We were initially concerned about the potential negative effects of glare and shadows on the subjects in our experiment. Glare caused by projecting light onto wet oil paint could both hinder the ability to control the appearance of the canvas using photometric compensation and annoy users of our system. Shadowing on the display due to projector occlusions by the artist could prevent the effective use of interaction modes for painting guidance at the color application and manipulation level. We found that by positioning our projectors at oblique angles, we were able to reduce both of these effects in practice. None of the artists reported any discomfort caused by glare or shadows.

CONCLUSION AND FUTURE WORK

We have described a novel interactive system for guiding novice artists to paint using traditional methods and tools. We have demonstrated how a multi-projector adaptive display can support artistic painting by guiding users according to traditional painting techniques taught in artist's manuals. Furthermore, we presented a set of five interaction modes for a canvas-centric interface which supports a traditional layer-based approach to painting. We evaluated the effectiveness of our system in meeting the goals of improved quality of experience and high quality of output. The results of our user study indicate that these goals were met.



(a)



(b)



(c)

Figure 8: **(a)**: *Top two rows*: Paintings made with the system. *Lower two rows*: Corresponding zoomed-in regions for the paintings in the top rows. **(b)**: Model painting all subjects in the user study were told to copy **(c)**: *Top two rows*: Paintings made without the system. *Lower two rows*: Corresponding zoomed-in regions for the paintings in the top rows. Notice the larger deviation in shading from the model for the zoomed in portion between the two groups.

Our system has the flexibility to handle a wide range of painting styles and approaches. Portraiture, still life, and landscape paintings are examples of domains that are supported by our process model. However, Jackson Pollock's paintings provide an example of a less traditional means of applying paint that does not fall within the range of techniques that we currently support.

In future work, we are considering extending the painting interaction techniques presented in this paper to other areas of the fine arts. Multiple projectors could guide a sculptor in subtracting material from a block of marble, for example. Just as layers were used to build up the composition of a painting in the current system, a sculpting system could guide an artist in a hierarchical broad-to-fine chiseling approach.

ACKNOWLEDGEMENTS

We would like to thank Roxanne Flagg for providing expert painting demonstrations and making helpful comments that were instrumental to improving the painting interface. Special thanks to Lan Guo for her assistance in conducting the user study and thanks to our experts Philip Carpenter, Daryl White, and Tina Simonton. We also thank professors Jessica Hodgins, Irfan Essa and Michael Cohen for their helpful comments on the paper. This material is based on work supported in part by the National Science Foundation under NSF Grant IIS-0205507.

REFERENCES

1. M. Ashdown, M. Flagg, R. Sukthankar, and J. Rehg. A flexible projector-camera system for multi-planar displays. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 165–172, 2004.
2. Barron. *All About Techniques in Oil*. Barron's Educational Series, Inc., 1997.
3. W. Baxter, V. Scheib, M.C. Lin, and D. Manocha. dab: Interactive haptic painting with 3d virtual brushes. In *Proceedings of Siggraph*, pages 461–468, New York, 2001.
4. W. Baxter, J. Wendt, and M. Lin. Impasto: A realistic, interactive model for paint. In *Proceedings of the 3rd International Symposium on Non-Photorealistic Animation and Rendering*, pages 45–56, New York, June 2004.
5. O. Bimber, F. Coriand, A. Kleppe, E. Bruns, S. Zollmann, and T. Langlotz. Superimposing pictorial artwork with projected imagery. *IEEE MultiMedia*, pages 16–26, January-March 2005.
6. O. Bimber, G. Wetzstein, A. Emmerling, and C. Nitschke. Enabling view-dependent stereoscopic projection in real environments. In *In proceedings of International Symposium on Mixed and Augmented Reality (ISMAR'05)*, pages 14–23, 2005.
7. T.-J. Cham, J.M. Rehg, R. Sukthankar, and G. Sukthankar. Shadow elimination and occluder light suppression for multi-projector displays. In *Proceedings of Computer Vision and Pattern Recognition*, pages 513–520, 2003.
8. C.J. Curtis, S.E. Anderson, J.E. Seims, K.W. Fleischer, and D.H. Salesin. Computer-generated watercolor. In *Proceedings of Siggraph*, pages 421–430, 1997.
9. B. Edwards. *Color: A Course in Mastering the Art of Mixing Colors*. Penguin Group (USA) Inc., 2004.
10. A. Gair. *The Artist's Handbook: A Step-by-Step Guide to Drawing, Watercolor and Oil Painting*. MetroBook, 1998.
11. P.E. Haeberli. Paint by numbers: Abstract image representations. In *Proceedings of Siggraph*, pages 207–214, 1990.
12. J. Hays and I. Essa. Image and video based painterly animation. In *Proceedings of the 3rd Symposium on Non-Photorealistic Animation and Rendering*, pages 113–120, New York, 2004.
13. A. Hertzmann. Painterly rendering with curved brush strokes of multiple sizes. In *Proceedings of Siggraph*, pages 453–460, 1998.
14. P. McCorduck. *Aaron's Code: Meta-Art, Artificial Intelligence and the Work of Harold Cohen*. W H Freeman and Co., 1990.
15. B.J. Meier. Painterly rendering for animation. In *Proceedings of Siggraph*, pages 477–484, 1996.
16. S. Nayar, H. Peri, M. Grossberg, and P. Belhumeur. A projection system with radiometric compensation for screen imperfections. In *First IEEE International Workshop on Projector-Camera Systems (PROCAMS-2003)*, 2003.
17. C. Pinhanez, R. Kjeldsen, T. Levas, G. Pingali, M. Podlaseck, and P. Chou. Ubiquitous Interactive Graphics. Technical Report RC22495 (W0205-143), IBM Research Report, May 2002.
18. R. Raskar, G. Welch, and H. Fuchs. Spatially augmented reality. In *First IEEE Workshop on Augmented Reality (IWAR'98)*, 1998.
19. R. Raskar, G. Welch, and K.-L. Low. Shader Lamps: Animating real objects with image-based illumination. In *Proceedings of Eurographics Workshop on Rendering*, 2001.
20. M.P. Salisbury, M.T. Wong, J.F. Hughes, and D.H. Salesin. Orientable textures for image-based pen-and-ink illustration. In *Proceedings of Siggraph*, 1997.
21. R. Smith, M. Wright, and J. Horton. *An Introduction to Art Techniques*. DK Publishing, Inc., 1995.
22. Philip Steadman. *Vermeer's Camera: Uncovering the Truth behind the Masterpieces*. Oxford University Press, 2002.
23. J. Summet, M. Flagg, T.-J. Cham, J.M. Rehg, and R. Sukthankar. Shadow elimination and blinding light suppression for interactive projected displays. *IEEE Transactions on Visualization and Computer Graphics*, Accepted for publication in 2006.
24. T. Yoshida, C. Horii, and K. Sato. A virtual color reconstruction system for real heritage with light projection. In *Proceedings of International Conference on Virtual Systems and MultiMedia*, 2003.

Appendix B-7

Registration Techniques for Using Imperfect and Partially Calibrated Devices in Planar Multi-Projector Displays

Ezekiel Bhasker, Ray Juang, *Student, IEEE*, and Aditi Majumder, *Member, IEEE*

Abstract— Multi-projector displays today are automatically registered, both geometrically and photometrically, using cameras. Existing registration techniques assume pre-calibrated projectors and cameras that are devoid of imperfections such as lens distortion. In practice, however, these devices are usually imperfect and uncalibrated. Registration of each of these devices is often more challenging than the multi-projector display registration itself. To make tiled projection-based displays accessible to a layman user we should allow the use of uncalibrated inexpensive devices that are prone to imperfections. In this paper, we make two important advances in this direction. First, we present a new geometric registration technique that can achieve geometric alignment *in the presence of severe projector lens distortion* using a relatively inexpensive low-resolution camera. This is achieved via a closed-form model that relates the projectors to cameras, in planar multi-projector displays, using rational Bezier patches. This enables us to geometrically calibrate a 3000×2500 resolution planar multi-projector display made of 3×3 array of nine severely distorted projectors using a low resolution (640×480) VGA camera. Second, we present a photometric self-calibration technique for a projector-camera pair. This allows us to photometrically calibrate the same display made of nine projectors using a *photometrically uncalibrated* camera. To the best of our knowledge, this is the first work that allows geometrically imperfect projectors and photometrically uncalibrated cameras in calibrating multi-projector displays.

Index Terms—Geometric calibration, photometric calibration, tiled displays.

1 INTRODUCTION

Multi-projector displays are popular today for many applications such as visualization, entertainment, training, and simulation. Contemporary researchers envision future workspaces with *ubiquitous pixels* rendered by multi-projector displays of various scales and forms aiding users in collaboration, interface, and visualization [9, 12, 16, 17, 24, 25, 30, 33, 36].

To enable such readily accessible multi-projector displays, automated camera-based registration techniques that register the imagery coming from multiple projectors, both geometrically and photometrically, have been proposed [5, 6, 13, 19, 20, 21, 23, 26, 27, 28, 29, 34, 37, 38]. More recently, Bhasker, et al. [2], proposed a distributed network of compact projector-camera units that can self-register without any user input, relieving the user of installation and maintenance responsibilities. However, all these methods do not handle imperfections, such as lens distortions, that are common in inexpensive commodity devices. They assume pre-calibrated devices (*i.e.*, devices whose geometric and photometric parameters are known) leaving the responsibility of device calibration to the user. Device calibration is often more challenging than multi-projector display registration and demands a very educated user. This difficulty in deployment has quarantined tiled projection-based displays to national laboratories and universities despite huge advances in automated camera-based registration techniques.

The work in this paper makes initial advances by developing geometric and photometric registration techniques that allow imperfect uncalibrated devices to be used for the most common cases of planar displays. The ultimate goal would be to allow the use of projectors with *severe* geometric and photometric distortion and still achieve acceptable geometric and photometric registration. We achieve this goal

using cameras that are neither geometrically nor photometrically registered. The paper by Majumder and Stevens [21] is the only work to date in this direction and presents initial methods to handle large spatial variation in intensity within and across projectors. In this paper, we make further advances to these unexplored domains. These are: (a) achieving geometric registration while using projectors that have considerable geometric imperfections (*e.g.*, lens distortions); (b) achieving photometric registration with a photometrically uncalibrated camera.

1.1 Using Geometrically Imperfect Projectors

We present a new geometric registration technique that can correct for severe lens distortion using a relatively low-resolution camera for planar multi-projector displays. This method relies on a new parametric geometric model relating the projector to the camera used for registering the display. The relationship is modeled in a compact and efficient manner using a rational Bezier patch.

Projection technology is advancing at a tremendous pace. Pocket projectors that can fit on one's palm are no longer a myth but a reality [1]. Industry initiatives are finding ways to build short-throw, unbreakable projectors that use LEDs as light sources so that they can be embedded in cell phones. It is evident that lens distortion will be significant in such low-cost projectors. Our method will enable the use of such low-cost projectors to build commodity tiled displays.

Further, in designing a compact multi-projector display or changing an existing setup to increase space utility, a short-throw lens is required to decrease the throw distance of the projectors. These lenses currently require several optical elements to reverse the high lens distortion created by short focal lengths. Hence, they are cost-prohibitive (around \$2000 - \$6000) and at least 5-6 times the cost of regular projectors (around \$800 - \$1500). Our work handles severe lens distortions, thus paving the way for using inexpensive lenses on projectors, just as they are used on cameras today. This allows for even more compact multi-projector display designs.

1.2 Using Photometrically Uncalibrated Cameras

Existing methods for photometric calibration of a multi-projector display often use a camera that is first calibrated using high dynamic range images [8, 26]. This involves taking images in an outdoor environment where one does not have control of motion, illumination, etc. This calibrated camera is then used to recover the photometric proper-

• Ezekiel Bhasker is with University of California, Irvine, E-mail: ebhasker@ics.uci.edu.

• Ray Juang is with University of California, Irvine, E-mail: rjuang@ics.uci.edu.

• Aditi Majumder is with University of California, Irvine, E-mail: majumder@ics.uci.edu.

Manuscript received 31 March 2007; accepted 1 August 2007; posted online 27 October 2007. Published 14 September 2007.

For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

ties of the projectors which are then modified appropriately to achieve a photometrically seamless multi-projector display [21, 26].

In this paper, we present a photometric self-calibration technique for a projector-camera pair that allows us to register a multi-projector display photometrically using an uncalibrated camera. To the best of our knowledge, this is the first work that estimates both the intensity transfer function and the spatial variation in intensity of projectors simultaneously using a photometrically uncalibrated camera.

2 RELATED WORK

In this section, we review existing geometric and photometric registration techniques and relate them to the techniques we present in this paper.

2.1 Geometric Registration

Geometric registration of planar multi-projector displays entails reconstructing two functions: (a) the function that relates the individual projector coordinates to the camera coordinates; and (b) the function that relates the camera coordinates to the global screen coordinates. Most geometric registration techniques devised for planar displays so far fall into the following categories.

Linear methods assume linear models for cameras and projectors. They relate the projector, camera, and the screen coordinates by linear matrices called homographies [5, 34]. However, linear models do not account for geometric non-linearities, such as lens distortion, and cannot achieve accurate alignment in their presence. To achieve acceptable alignment in such cases, projectors are manually adjusted to the ‘sweet-spot’ in zoom level where the non-linear distortions become negligible.

Piecewise linear methods address these geometric non-linearities by using a piecewise linear function, essentially a triangulation, to relate the projector, camera and screen coordinates to each other. Though this achieves reasonable geometric accuracy, a dense triangulation is required to sample the parameter space adequately [37]. Thus, a high-resolution (5-6 Megapixel) camera is required to register even relatively small displays made of 4-6 projectors.

Hereld, et. al. [13], use a *non-linear model* where a closed-form cubic function handles some of the geometric non-linearities. This method has two limitations. First, a simple cubic polynomial cannot capture the perspective projection between the projector and the camera adequately. As a result, this method assumes a close-to-rectangular array of projectors resulting from on-axis projection. Such an arrangement is relatively easy to achieve manually in a rear-projection system, but not in front-projection system where the projector is on the ceiling projecting on a screen placed on the ground. Second, this work is not motivated by accurate analysis of the kind of non-linearities seen in projectors and hence cannot handle higher order non-linearities.

Our method incorporates the non-linear distortions of the projector into the function that relates the projector to other entities, such as the camera and screen, using a single accurate closed-form function – the rational Bezier patch. This allows combination of distortions arising from completely different reasons (e.g., lens distortion and key-stoning) into a single unified representation. Further, it results in an efficient camera-based geometric registration for planar, multi-projector displays, that can handle significant lens distortion.

2.2 Photometric Registration

Photometric properties of a camera/projector involve its intensity transfer function, or ITF (more commonly called the gamma function), and the spatial intensity fall-off in its field-of-view (more commonly called the vignetting effect). *Photometric registration* of projection-based displays requires estimation of the photometric properties of the projectors using a camera. To do this, the camera is first photometrically registered. The high dynamic range imaging technique [8, 22] can be used to estimate the camera’s ITF by taking images outdoors. Currently, there is no simple automated way of estimating the camera’s vignetting effect. The registered camera is usually set to a narrow aperture when using it to estimate the photometric properties of the projectors in a tiled display. The techniques in the work by Raji,

et al. [26], are then applied to estimate the projectors’ ITF using the registered camera. Using the projectors with known ITF, the techniques described in the papers by Majumder and Stevens [20, 21] are then used to estimate the spatial intensity fall-off of the projector, and to register the multiple projectors photometrically using the estimated photometric properties. No method currently exists that can use a photometrically uncalibrated camera and recover both the ITF and the spatial intensity fall-off of the projectors simultaneously. Our method in this paper estimates all the different photometric parameters of both devices in a photometrically uncalibrated projector-camera pair. It is accomplished in a controlled indoor setting without using any other physical props. Photometric registration techniques that use these parameters [20, 21, 38] can then be utilized to achieve photometric seamlessness in the multi-projector display.

3 USING GEOMETRICALLY IMPERFECT PROJECTORS

In this section, we present our geometric registration technique that corrects for projector geometric imperfections, such as lens distortion. We use a camera that is corrected for lens distortion [3]. First, we develop a new model that relates a single projector to the camera in a planar multi-projector display using a rational Bezier patch. The projector lens distortion is modeled by an affine-invariant *non-rational* Bezier patch. For a projector-camera setup with a planar display, this function is followed by a homography that relates the projector to the camera and to the screen linearly. We use a perspective-invariant *rational* Bezier patch to model the combination of these non-linear and linear functions. Thus, all the different distortions in a planar projection-based display (e.g., key-stoning, radial, and tangential distortions) are combined into a single closed-form function. To the best of our knowledge, this is the first work that models both linear and non-linear distortions of a projector-camera system in a unified manner.

To geometrically register multiple distorted projectors, we sparsely sample the rational Bezier function relating the projector coordinates with the camera and screen coordinates, and estimate the function parameters. These parameters are then used to warp the input image to the projector appropriately resulting in a geometrically aligned image. The sparse sampling proves adequate for acceptable geometric registration since the rational Bezier offers a piecewise curve representation (which is more suited for lens distortions that transform lines to curves), rather than a piecewise linear representation. This sparse sampling allows us to use a low-resolution camera to achieve geometric registration for even medium sized displays made of 6-16 projectors.

3.1 Model for Single Projector

We first consider the simple case of a single projector projecting onto a planar screen and being observed by a single camera corrected for lens distortion. Our goal is to design an accurate parametric function $\mathcal{F}$ that defines the relationship between the projector pixels (x, y) and the camera pixels (u, v) , i.e., $(u, v) = \mathcal{F}(x, y)$. Ideally, the function $\mathcal{F}$ has two components. The first is the *lens distortion function* $\mathcal{R}$ that relates the undistorted projector coordinates (x, y) to the distorted coordinates on the display screen (x_d, y_d) . The second component is the well-known *homography* $\mathcal{H}$ that relates the distorted coordinates to camera coordinates (u, v) .

$$(u, v) = \mathcal{H}(x_d, y_d) = \mathcal{H}(\mathcal{R}(x, y)) = \mathcal{F}(x, y) \quad (1)$$

We would like to model $\mathcal{F}$ by a single compact parametric function that accurately captures the effect of the concatenation of these two functions, $\mathcal{H}$ and $\mathcal{R}$, and is amenable to efficient data fitting computation.

Modeling $\mathcal{R}$ with Non-rational Bezier Patches: The classic lens distortion model [4, 10] consists of two independent distortions: *radial distortion* (e.g., barrel or pin-cushion) and *tangential distortion*.

Radial distortion is usually modeled by

$$x_d = x + (x - x_p)(k_1 r^2 + k_2 r^4 + k_3 r^6) = x + \rho_x, \quad (2)$$

$$y_d = y + (y - y_p)(k_1 r^2 + k_2 r^4 + k_3 r^6) = y + \rho_y, \quad (3)$$

Case No.	Lens Distortion Parameters			% Error For Bezier of degree					
	Center (x_c, y_c)	Radial (k_1, k_2, k_3)	Tangential (p_1, p_2)	2	3	4	5	6	7
(a)	(0.5, 0.5)	(-0.35, 0, 0)	(0, 0)	5.78	ϵ				
(b)	(0.5, 0.5)	(0, 0, 0)	(0.1, 0.1)	ϵ					
(c)	(0.5, 0.5)	(-0.35, -0.35, 0)	(0, 0)	9.00	0.91	0.56	ϵ		
(d)	(0.6, 0.55)	(-0.35, -0.13, 0)	(0.05, 0.05)	1.9	0.14	0.03	ϵ		
(e)	(0.6, 0.55)	(-0.35, -0.13, -0.016)	(0.05, 0.05)	2.0	0.17	0.038	0.0012	0.00022	ϵ

Table 1. This table shows the accuracy of our fit for lens distortion functions using non-rational Bezier patches of different degree. Due to the large number of lower degree terms, note that in case (e) distortion of higher degrees (e.g., degree 7) can be fitted to reasonable accuracy by a lower degree Bezier (e.g., degree 4). ϵ represents the convergence threshold used which is strictly less than 10^{-11} .

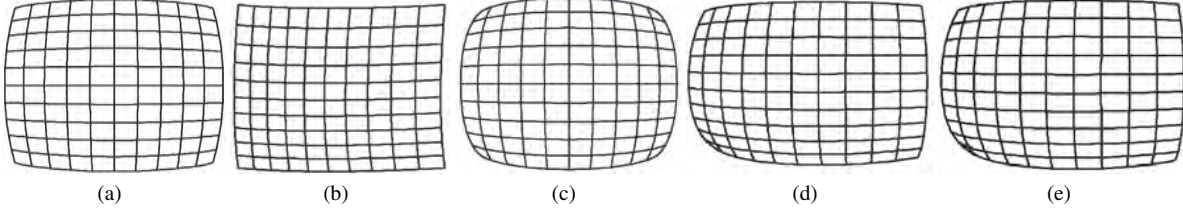


Fig. 1. Images showing the five cases of distortion mentioned in Table 1.

where (x_d, y_d) and (x, y) are the distorted and undistorted image coordinates respectively, $r = \sqrt{(x - x_p)^2 + (y - y_p)^2}$ is the radial distance from the principal center (x_p, y_p) in the undistorted image, and k_i , $1 \leq i \leq 3$ are the radial distortion coefficients. Negative values of k_i result in barrel distortion while positive values result in pin-cushioning. The principal center is a point in the image that is unchanged by radial distortion. In general, the principal center (x_p, y_p) need not be at the center of the image but is usually close to the center.

The tangential distortion is modeled by

$$x_d = x + 2p_1xy + p_2r^2 + 2p_2x^2 = x + \tau_x, \quad (4)$$

$$y_d = y + 2p_2xy + p_1r^2 + 2p_1y^2 = y + \tau_y \quad (5)$$

where $r = \sqrt{(x - x_p)^2 + (y - y_p)^2}$ and p_1 and p_2 are the tangential distortion parameters.

Radial and tangential distortion are independent of each other and hence their effects can be combined into a comprehensive lens distortion equation, as

$$(x_d, y_d) = (x + \rho_x + \tau_x, y + \rho_y + \tau_y). \quad (6)$$

Estimating the unknown radial and tangential distortion coefficients involves solving complex non-linear optimization problems. Hence, several simplifying assumptions are usually made which are often not true in a real system. The most common simplifications are to assume that the principal point coincides with the center of the image and to ignore the higher degree terms [3, 7, 15, 32, 35, 39].

We model this lens distortion by *non-rational Bezier patches* instead. A non-rational Bezier patch of degree k and dimension d is defined by $(k+1)^d$ control points in a d -dimensional space. In our case, we consider a two-dimensional patch where $d = 2$. The Bezier patch is a tensor product patch where the control points, P_{ij} , are weighted by blending functions that are products of independent one-dimensional Bernstein polynomials. There are $k+1$ Bernstein polynomials of degree k given by

$$b_{nk}(t) = \frac{k!}{n!(k-n)!} t^n (1-t)^{k-n} \quad (7)$$

where $0 \leq n \leq k$ and $0 \leq t \leq 1$. The non-rational Bezier patch can then be defined from these polynomials as

$$\mathcal{NB}(x, y) = \sum_{i=0}^k \sum_{j=0}^k P_{ij} b_{ik}(x) b_{jk}(y) \quad (8)$$

$$= \sum_{i=0}^k \sum_{j=0}^k P_{ij} B_{ijk}(x, y), \quad (9)$$

where B_{ijk} is the tensor product of the Bernstein polynomials b_{ik} and b_{jk} and $0 \leq (x, y) \leq 1$. Hence, in the context of the single projector-camera pair for planar displays, the projector lens distortion function $\mathcal{R}(x, y)$ is represented by

$$(x_d, y_d) = \mathcal{R}(x, y) = \mathcal{NB}(x, y) \quad (10)$$

Our choice of using a non-rational Bezier patch is guided by the following rationale.

1. The non-rational Bezier patch provides a unified framework to represent a large range of lens distortions. More severe lens distortions can be included by increasing the *degree* of the Bezier patch. For example, a bicubic Bezier patch may be sufficient for standard narrow FOV lenses, but a Bezier patch of degree 4 may be required for the more severe distortions of wide FOV lenses. All of these are handled by the non-rational Bezier patch without affecting the underlying computational framework.
2. When Equation 10 is expanded and compared with Equation 6, one finds that the non-rational Bezier patch has many more higher-order cross terms, making it a more flexible model. For example, it is possible for short-throw lens projectors to have a mix of both barrel and pin-cushion distortions. Such cases are better handled by a more flexible model such as the Bezier patch.
3. The principal center creates a particular nuisance in traditional lens distortion estimation techniques, and forces most practical systems to assume it to be at the center of the image. When using a non-rational Bezier patch, the principle center is encoded within the control points themselves, and need not be estimated separately.
4. Like lens distortion, non-rational Bezier patches are also affine-invariant and hence provide an accurate model.

Table 1 and Figure 1 present results on the accuracy of our Bezier patch model. We consider a dense sampling of the (x, y) space, distort it using various traditional lens distortion parameters, and fit a non-rational Bezier patch to the distorted points. To evaluate the accuracy of the fitted Bezier, we find the average Euclidian distance error between fitted and the distorted data as a percentage of the distance from

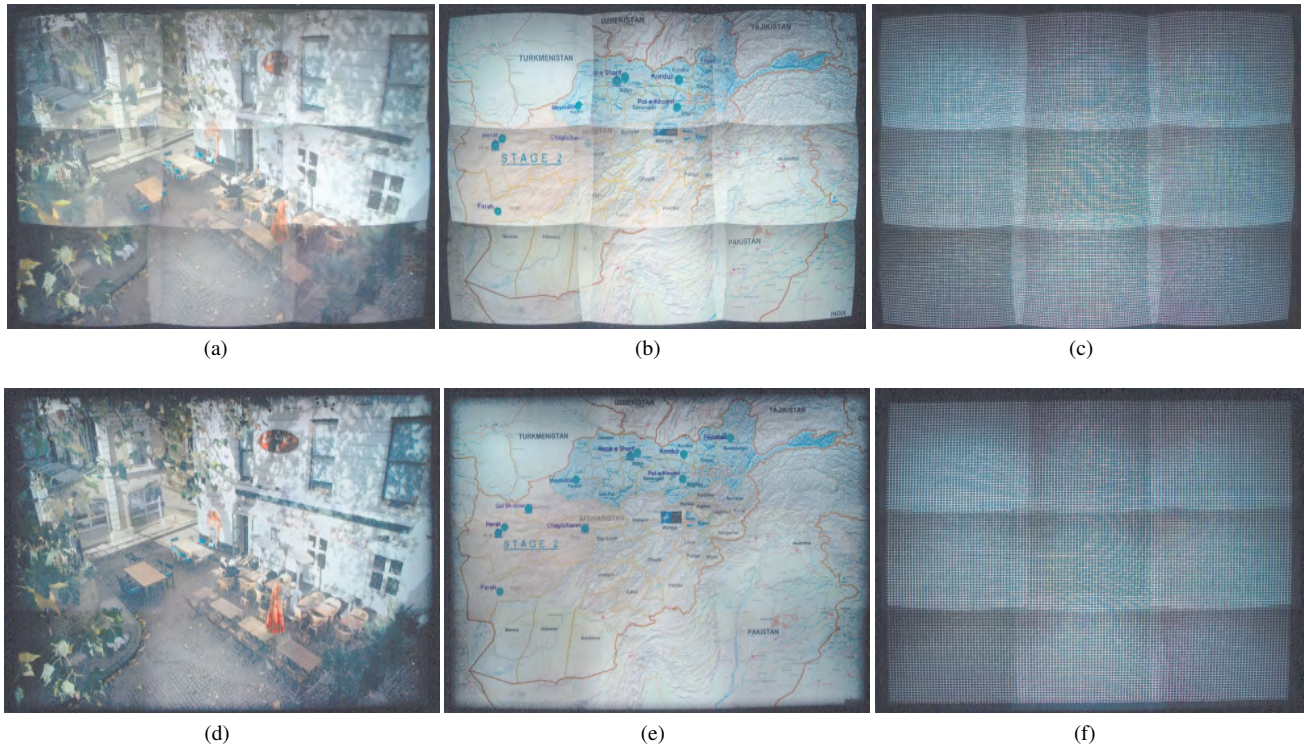


Fig. 2. Results of geometric registration shown on our display which is a 3×3 array of nine projectors. The geometric registration is achieved using bicubic rational Bezier patches with sparse correspondences (8×6 for a 1024×768 projector). (a, b, c) Before registration; (d, e, f) After registration. The grid image in the last column is made of lines that are one pixel wide. To retain clarity of the grids in the overlap region, photometric registration is not applied to the grid image in (c) when generating the registered image (f).

the principal center. Note that even when higher order distortions are present (e.g., x^7), lower degree Bezier patches (e.g., bicubic) can provide a good fit. This is due to the presence of a large number of lower-order cross-terms, which are absent in the traditional lens distortion model.

Modeling $\mathcal{H}\mathcal{R}$ with Rational Bezier Patches: Non-rational Bezier patches are affine-invariant but not perspective-invariant. Relating the projector to the camera not only involves a lens distortion but also a perspective projection. In the case of planar displays, this is a homography $\mathcal{H}$, a simpler 3×3 transformation induced by the planar display screen. So, to model the combined effects of $\mathcal{H}$ and $\mathcal{R}$ in Equation 1, we desire a function that retains all the properties of the Bezier and is also perspective-invariant. Retaining all the properties of Bezier patch assures retaining the accurate model of non-linearities due to distortion. Invariance under perspective projection assures the additional modeling of the perspective projection between the projector and the camera. To our advantage, a generalized form of Bezier patches called rational Bezier patches offer us a function that has all the properties of a non-rational Bezier and is also perspective-invariant. In addition to the 2D control points of Bezier patches, rational Bezier patches have a scalar weight associated with each control point, and the patch is defined as a weighted sum of the blended control points:

$$\mathcal{B}(x, y) = \frac{\sum_{i=0}^k \sum_{j=0}^k P_{ij} w_{ij} B_{ijk}(x, y)}{\sum_{i=0}^k \sum_{j=0}^k w_{ij} B_{ijk}(x, y)}. \quad (11)$$

The weights provide this function with the capability to sample the parameter space in a non-uniform (but controlled) manner. This makes it perspective-invariant. Hence, we propose the use of a rational Bezier patch to estimate the function $\mathcal{F} = \mathcal{H}\mathcal{R}$.

3.2 Geometric Registration of Multiple Projectors

Geometric registration of a multi-projector display requires estimating two functions. The first function maps the projector coordinates (x, y) to the camera coordinates (u, v) . We use a rational Bezier patch $\mathcal{B}$ for

this purpose. The second function maps the camera coordinates (u, v) to the normalized screen coordinates (p, q) . The presence of a planar screen allows us to use a homography $\mathcal{H}_S$ for this mapping. These functions are estimated as follows.

Parameter Estimation: To estimate $\mathcal{H}_S$, we require four correspondences between the camera and screen coordinates, (u, v) and (p, q) . The four correspondences result in a system of linear equations that can be solved to recover the homography matrix. To find the correspondences, we identify four points in the camera image that correspond to the four corners of the screen. This can be done automatically by attaching fiducials to the display. Since our display is nearly rectangular, we set all the projectors to display white light and use simple image processing to detect the extreme corners of the lighted regions to find the camera coordinates that correspond to the four corners of the normalized screen space.

To estimate the parameters of the rational Bezier patch, $\mathcal{B}$, we find correspondences between projector pixels (x, y) and camera pixels (u, v) . Let us assume N correspondences, i.e., (x_c, y_c) corresponding to (u_c, v_c) , $0 \leq c \leq N-1$. The unknowns are the control points in the camera's space, $P_{ij} = (u_{ij}, v_{ij})$, and their scalar weights w_{ij} . From the known correspondences, Equation 11 can be written for each correspondence as follows:

$$(u_c, v_c) \sum_{i=0}^k \sum_{j=0}^k w_{ij} B_{ijk}(x_c, y_c) - \sum_{i=0}^k \sum_{j=0}^k (u_{ij}, v_{ij}) w_{ij} B_{ijk}(x_c, y_c) = 0. \quad (12)$$

Assuming $U_{ij} = u_{ij} w_{ij}$ and $V_{ij} = v_{ij} w_{ij}$, Equation 12 yields two equations for each correspondence, as follows:

$$u_c \sum_{i=0}^k \sum_{j=0}^k w_{ij} B_{ijk}(x_c, y_c) - \sum_{i=0}^k \sum_{j=0}^k U_{ij} B_{ijk}(x_c, y_c) = 0. \quad (13)$$

$$v_c \sum_{i=0}^k \sum_{j=0}^k w_{ij} B_{ijk}(x_c, y_c) - \sum_{i=0}^k \sum_{j=0}^k V_{ij} B_{ijk}(x_c, y_c) = 0. \quad (14)$$

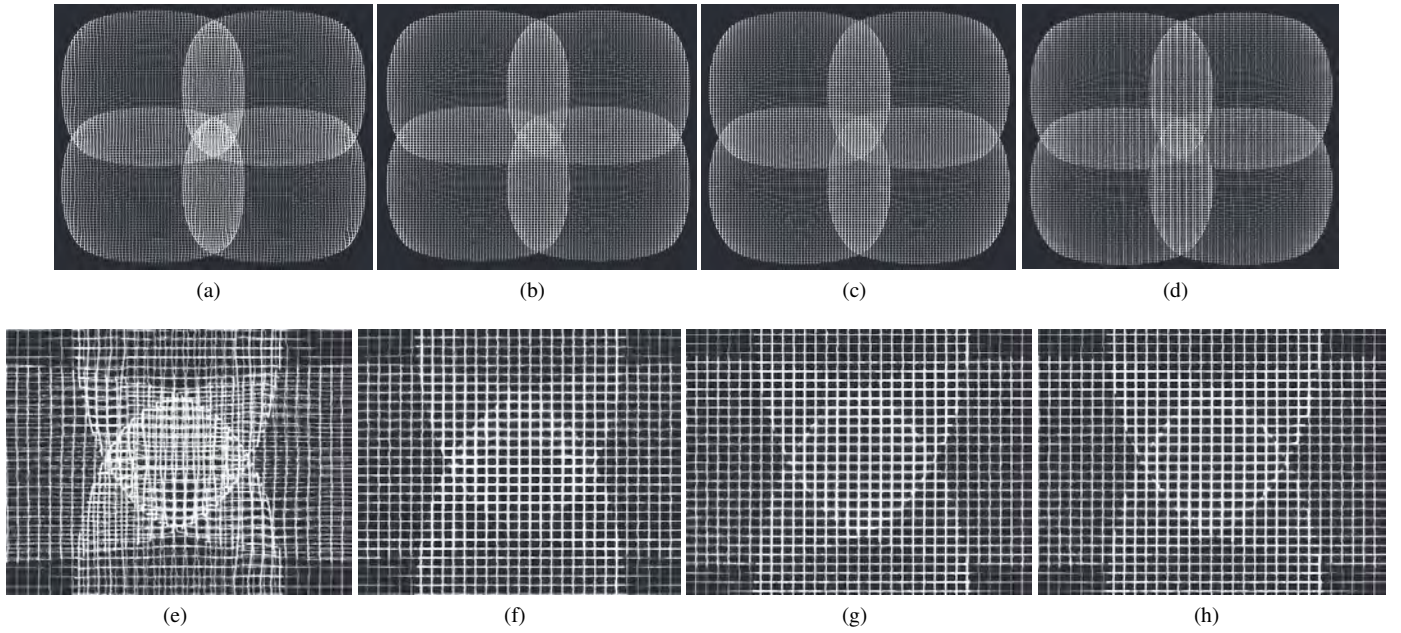


Fig. 3. Geometric registration of projectors with severe *barrel* distortion: using a linear piecewise method with sparse - 48 samples (a) and dense - 432 samples (b) correspondences; using rational *bicubic* Bezier patches with sparse - 48 samples (c) and dense - 432 samples (d) correspondences. (e,f,g,h) provides a zoomed in view of the projector overlap regions in (a,b,c,d) respectively. It is evident that piecewise linear methods with sparse correspondences lead to severe mismatches. The results using rational Bezier are similar, even when the sampling is reduced by an order of magnitude. In fact, at places, the results from using rational Bezier patches with sparse sampling is better than the results from using a dense sampling for the piecewise linear method. This is evident as the grid lines are thicker in the overlap region in (b) due to slight errors.

The unknowns for estimating a two-dimensional rational Bezier patch of degree k are $(k+1)^2$ control points and weights, which result in a total of $3(k+1)^2$ unknowns. The weights are variables that depend on the control points in a non-linear fashion. So, the above equations result in a non-linear least-squares fitting problem that can be solved efficiently using the Levenberg-Marquardt gradient descent optimization technique [31]. As an initial guess to this optimization, we fit the control polygon tightly around the set of correspondences (u_c, v_c) in the camera space and constrain the weights to be strictly greater than 0.

To find the correspondences between projector and camera coordinates, (x, y) and (u, v) , we display a rectangular grid of Gaussian blobs with known coordinates on each projector. These are captured by a calibrated camera in a series of non-overlapping images. We use a 2D-stepping procedure that depends on a user to identify the top-left blob and its immediate right and bottom neighbors in camera space. With this minimal three point user input, we can design a stepping procedure that (a) estimates the rough position of the next blob in scan-line order, and (b) searches for the correct blob position using the nearest windowed center-of-mass technique [11]. If this is not possible for more extreme projector distortions, one can binary-encode the blobs and project them in a time sequential manner to recover the exact ids of the detected blobs and find the correspondences [28, 37].

Image Correction: To create a geometrically seamless and undistorted display, we warp the image for each projector as follows: For each projector coordinate (x, y) , we first compute the corresponding camera coordinate (u, v) using the rational Bezier patch $\mathcal{B}$. Next, we compute the corresponding normalized screen coordinates (p, q) using the homography $\mathcal{H}_S$. The generated coordinates (p, q) are used as indices into the image being rendered on the tiled display. Bilinear interpolation is used to assign the final value of each projector pixel. Thus, the correction is achieved by the mapping

$$(p, q) = \mathcal{H}_S(u, v) = \mathcal{H}_S(\mathcal{B}(x, y)). \quad (15)$$

3.3 Implementation and Results

We have applied our method on our 8×6 -foot display made of 3×3 array of nine projectors. Our projectors have relatively large throw ra-

tios and hence do not reveal major lens distortions. Instead of choosing the cost prohibitive option of buying nine short-throw lenses, we chose to simulate the distortion digitally by distorting the input images to the projectors.

The registration is done offline and takes a couple of minutes. This generates the parameters of the rational Bezier for each projector, which are then stored to be used during image correction. We have implemented the image correction in real-time using modern GPUs through Chromium – an open-source distributed rendering engine for PC clusters [14]. A module for Chromium is written that first pre-computes the coordinate-mappings of all pixel coordinates using the stored parameters of the rational Bezier. This results in a per-pixel projector to screen lookup table. A fragment program for the pixel shader quickly maps pixels from the projector coordinate space to the screen coordinate space during rendering.

Figure 2 shows the results. The advantage of using an appropriate parametric function lies in an accurate alignment even with a sparse sampling of the function. We can achieve accurate alignment of our relatively high resolution display (3000×2100) by using a low-resolution VGA (640×480) camera. Notice that even with a low-resolution camera we get sub-pixel accuracy for fine grids (one pixel wide lines) and text. Figure 3 shows that we can achieve accuracy comparable to existing methods (that use much higher resolution cameras) even when our sampling is sparser by an order of magnitude. To achieve perceptual photometric seamlessness, we used the work by Majumder, et al. [21]. To make the quality of the geometric registration visible in the grid images, we did not register the display photometrically for Figure 2(c) and Figure 3. So, these images, though geometrically aligned, still show photometric seams. The real-time rendering is demonstrated in the submitted video. However the video, being in VGA resolution, cannot capture the high resolution details.

4 USING PHOTOMETRICALLY UNCALIBRATED CAMERAS

In this section, we present a method for photometric registration of multi-projector displays using a photometrically uncalibrated camera. Projector photometric parameters that include its ITF and spatial intensity variation, need to be estimated to instrument existing photometric registration techniques [21]. We use an uncalibrated camera (whose

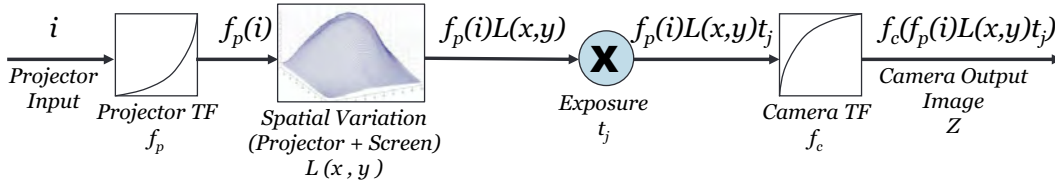


Fig. 4. The transformation process of an image as it passes through a projector-camera system.

ITF is unknown) to extract these projector photometric parameters. In the process, we also recover the camera ITF and hence calibrate the camera photometrically. To suppress the spatial intensity variation, such as vignetting effect, of the camera, like most previous work in scene reconstruction [8, 23], we operate the camera at a narrow aperture setting where it approaches the ideal pinhole model.

4.1 Model for a Single Projector

We assume a geometrically calibrated projector-camera system where a pixel (u, v) in the camera coordinate system is related to a pixel (x, y) in the projector coordinate system by a linear or non-linear warp $G(x, y) = (u, v)$. G can be determined by any geometric calibration method, including the one presented in the previous section [38].

Consider a spatially uniform grayscale input to the projector. Let the grayscale level be denoted by i . As per the model presented in the paper by Majumder and Gopi [18], the uniform image is first transformed by a spatially invariant input transfer function of the projector, f_p , to create a spatially uniform output, $f_p(i)$. Next, the projector optics introduce a spatially-dependent but input-independent intensity variation. This is further modulated by the screen reflectance/transmittance function. We denote this combined spatially dependent attenuation factor due to projector lens and screen as $L(x, y)$. Thus, the image produced after passing through the projector lens and screen is $f_p(i)L(x, y)$. The light from the screen then reaches the camera. The amount of light accepted by the camera is scaled by its exposure time t_j , where j indexes different exposure times. The different exposures are instrumented by changing the shutter speed of the camera. This produces an image $f_p(i)L(x, y)t_j$ that passes through the camera optics. Assuming negligible camera vignetting at narrow apertures, the image is transformed only by the input transfer function of the camera, f_c , to generate the grayscale value recorded by the camera Z . Thus, Z is a function of the input i , the exposure time index j , and the spatial coordinates (x, y) . This is illustrated in Figure 4. The final equation is

$$Z(i, j, x, y) = f_c(f_p(i)L(x, y)t_j). \quad (16)$$

For cameras, the intensity transfer function is monotonic [8], and hence invertible. Note that the same is not true for projectors [20]. Assuming invertible f_c , the above equation becomes

$$f_c^{-1}(Z(i, j, x, y)) = f_p(i)L(x, y)t_j. \quad (17)$$

Taking the natural logarithm of both sides we get,

$$\ln f_c^{-1}(Z(i, j, x, y)) = \ln f_p(i) + \ln(L(x, y)) + \ln(t_j). \quad (18)$$

To simplify the notation, we define $h_c = \ln f_c^{-1}$ and $h_p = \ln f_p$. The above equation then becomes

$$h_c(Z(i, j, x, y)) = h_p(i) + \ln(L(x, y)) + \ln(t_j) \quad (19)$$

where i ranges over the grayscale inputs, j ranges over the exposure times, and (x, y) ranges over the spatial coordinates of the projector. In this equation, Z and t_j are known while h_p , h_c and L are unknown. Varying i and t_j results in different values of Z for each pixel (x, y) . We can use this to setup a system of linear equations. We recover h_p , h_c and L that best satisfy Equation 19 in a least-squares sense. Note that recovering h_p and h_c involves solving the functions for a

finite number of samples in the complete range of input values. The recovered parameters are illustrated in Figure 5.

Improving Efficiency: The system of linear equations achieved by Equation 19 is very large. Let R_i and R_Z each be the cardinality of the set of input samples for h_p and h_c respectively, P be the total number of projector pixels, and T be the number of exposures. This results in $PT R_i$ equations and $P + R_i + R_Z$ unknown variables for the system of equations defined by Equation 19. Typically $R_i = R_Z = 256$ and $P = 1,000,000$ (assuming common XGA projector resolution). Since there are multiple exposures for each input P in the range R_i , the size of the linear system is on the order of at least a few million equations and cannot be solved efficiently.

To address this inefficiency we use a limited number of pixels in the projector space to solve for h_p and h_c . This is possible since h_p and h_c are both spatially constant. To ensure a sufficiently over-determined system, the criteria $PT R_i > P + R_i + R_Z$ should be satisfied. For $R_i = R_Z = 256$ and $T = 6$, a choice of 100 for P is more than adequate. So, we first subsample L to a resolution of 10×10 pixel and use this to setup a smaller linear system of equations to solve for h_p and h_c .

With the estimated h_p and h_c , we substitute these into Equation 19 and quickly solve for $L(x, y)$ at the various projector coordinates, by rewriting Equation 19 as

$$\ln(L(x, y)) = h_c(Z(i, j, x, y)) - h_p(i) - \ln(t_j) \quad (20)$$

Handling Noise: Our system does exhibit considerable random noise arising not only from the devices (camera and projector) but also from the screen. In particular, we use a rear projection screen with relatively high gain which has been shown to generate considerable random noise [20]. This has an adverse effect on the signal to noise ratio, especially for low projector input i or low camera output Z . Thus, estimating h_p and h_c directly from Equation 19 and L from Equation 20 results in noisy parameter estimation. We take several measures to reduce the effect of this random noise.

To assure smooth h_p and h_c while solving the system of equations, we minimize the error function

$$E = \sum_{j \in T} \sum_{(x, y) \in P} \sum_{i \in R_i} [(h_c(Z) - h_p(i) - \ln(L(x, y)) - \ln(t_j))]^2 \quad (21)$$

$$+ \lambda \left(\sum_{Z \in R_Z} h_c''(Z)^2 + \sum_{i \in R_i} h_p''(i)^2 \right). \quad (22)$$

The first term assures that the solution arises from the set of equations given by Equation 19 in a least-squares sense. The second term is a smoothing constraint on the curvature of h_p and h_c , given by their second derivatives. In the discrete domain, we use the Laplacian operator to find the curvature of h_p and h_c , i.e., $h_p''(i) = h_p(i-1) - 2h_p(i) + h_p(i+1)$. The scale factor λ weights the smoothness term relative to the data fitting term and should be chosen based on the amount of noise in Z . We choose $\lambda = 10$.

However, note that the first term in Equation 22 gives equal weights to all recorded camera values Z , and the second term gives equal weights to all i and Z . For the first term, we want to emphasize images with higher signal to noise ratio i.e. images for higher i with higher recorded values Z . For the second term, we want to emphasize smoothing of h_p and h_c in the lower ranges of i and Z respectively where low signal to noise ratio would result in a noisier estimate. To

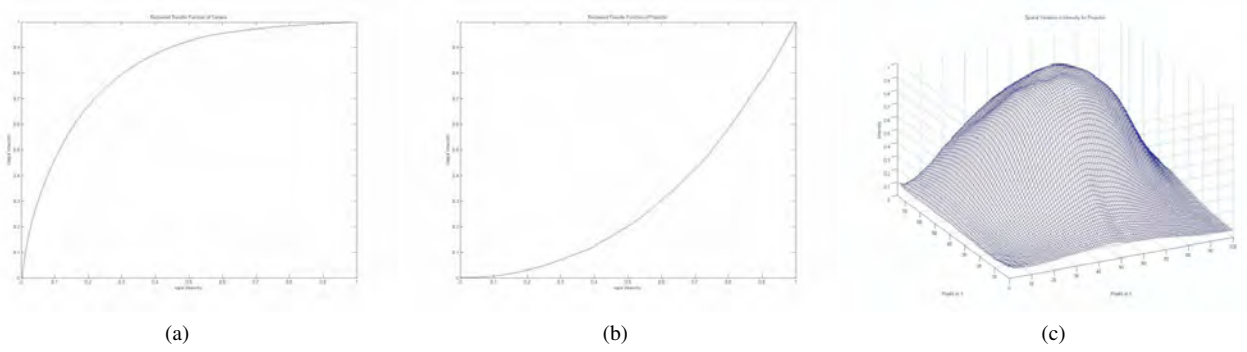


Fig. 5. (a) The estimated camera input transfer function f_c . (b) The estimated projector input transfer function f_p . (c) The estimated spatial intensity variation due to projector, screen, and camera L .

achieve this, we weight the first and second term of the error function in Equation 22 as follows,

$$E = \sum_{j \in T} \sum_{(x,y) \in P} \sum_{i \in R_i} [w_c(Z)(h_c(Z) - h_p(i) - \ln(L(x,y)) - \ln(t_j))]^2 \quad (23)$$

$$+ \lambda \left(\sum_{Z \in R_Z} [(Z_{max} - w_c(Z))h_c''(Z)]^2 + \sum_{i \in R_i} [(i_{max} - w_p(i))h_p''(i)]^2 \right). \quad (24)$$

where w_p and w_c are the weighting functions corresponding to the projector input and the recorded camera values, respectively, and Z_{max} and i_{max} are the maximum camera output and projector input respectively. To give the higher intensities greater confidence we use linear weighting functions, $w_c(Z) = Z$ and $w_p(i) = i$.

Similarly, noise can have adverse effects when estimating L using Equation 20 after h_p and h_c are recovered. Usually the image captured for input i at a particular exposure does not yield unsaturated outputs at all spatial pixel locations (x,y) . So, we need to use different images for reconstruction of L at different spatial locations. This will yield an L with different signal to noise ratio at different spatial regions depending on the image that was used to reconstruct L in any spatial region. Even if a rare situation occurs when one finds a single image (at a particular input and exposure) where all the pixels are unsaturated, L will still be noisy due to the inherent random noise of the projector-screen-camera system.

Averaging multiple observations is the standard way to reduce random noise significantly [11]. Using this fact, we reduce noise in L at any pixel (x,y) by solving it from multiple images and finding a weighted mean of the multiple solutions. The weighting function should emphasize higher i and Z to reduce the effect of random noise. We achieve this by a weighting function $w_L(Z,i) = w_c(Z)w_p(i)$. The solution can now be derived by modifying Equation 20 as

$$\ln(L(x,y)) = \frac{\sum_{j \in T} \sum_{i \in R_i} w_L(Z,i)[h_c(Z) - h_p(i) - \ln(t_j)]}{\sum_{j \in T} \sum_{i \in R_i} w_L(Z,i)}, \quad (25)$$

resulting in clean estimated parameters as illustrated in Figure 5.

4.2 Photometric Registration of Multiple Projectors

For photometric registration of a multi-projector display, we estimate the f_p and $L(x,y)$ of each projector independently using the same camera. Note that in order to recover the relative intensity levels of the projectors, we use the unnormalized $L(x,y)$ estimated independently from each projector. Using the extracted parameters from each projector, we use the photometric registration method proposed by Majumder and Stevens [21] to generate a perceptually seamless display.

4.3 Implementation and Results

We used a Kodak DCS ProSLR/n camera and projected 32 flat grayscale fields with intensity levels uniformly sampled from 0 to 255. For each intensity level, 15 exposures were taken. The camera was set to a narrow aperture, $f/32$. The data collection took 25 minutes. To

reduce the collection time, we tried reducing the number of exposures. Empirically, the minimum number of exposures needed before quantization effects became visible was eight. The exposures used, however, need to be well distributed amongst the range of available camera exposures. The program was written in C++ and utilizes Matlab and the OpenCV library. It takes about 11-15 minutes to register each projector on a Pentium 4, 2.8 GHz PC.

The photometric registration in Figure 2 (d) and (e) are achieved by using a photometrically uncalibrated camera. Additional results with projectors that do not exhibit lens distortion are shown in Figure 6. Here a linear geometric registration method is used [5]. These results show that a photometrically uncalibrated camera can be used irrespective of the kind of geometric registration method being used. When using the photometric registration technique of Majumder and Stevens [21] we use a low photometric uniformity parameter to retain the high contrast of the display as much as possible. This results in some perceivable seams near the boundaries of the bottom left and right projectors, which are significantly dimmer than the other projectors in our display. However, these seams exist even when the photometric registration is carried out with a calibrated camera, showing that they are artifacts of the photometric registration method itself and are not due to imperfect projector parameter estimation.

5 CONCLUSION

This paper presents new registration techniques to handle imperfect and uncalibrated devices in planar multi-projector displays. The four relevant cases include allowing (a) photometrically imperfect projectors, (b) geometrically imperfect projectors, (c) a photometrically uncalibrated camera, and (d) a geometrically uncalibrated camera. Our work addresses (b) and (c) in particular, while (a) is addressed by existing literature [21]. Our work allows severely distorted projectors to be used which opens up the possibility of mounting inexpensive lenses on projectors, just as with cameras, leading to very short-throw projectors. Using a photometrically uncalibrated camera allows the photometric registration to be conducted within a controlled indoor environment. Further, we have proposed the first method that estimates all projector/camera photometric parameters simultaneously. However, case (d) above of achieving geometric registration in multi-projector displays using a geometrically uncalibrated camera is still a challenging problem.

Further, our methods are within the realm of centralized architecture. Current distributed architectures [2] are very amenable to easy deployment, simple maintenance and hence more affordable multi-projector displays. Our goal is to extend our developed methodologies in this paper to the distributed framework of a network of projector-camera systems [2]. This poses a significant challenge which, if resolved, can be instrumental in realizing commodity tiled projection-based displays. All of these methodologies are stepping stones towards the vision of *ubiquitous pixels* in future workspaces.

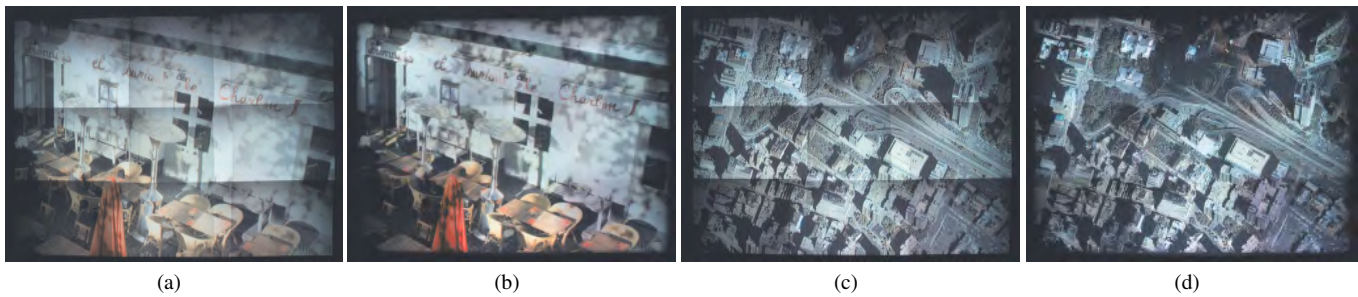


Fig. 6. Images of our 3×3 array of nine projectors. (a,c) The geometrically registered display without any photometric registration. (b,d) Photometric registration achieved by using an uncalibrated camera.

REFERENCES

- [1] Mitsubishi pocketprojector led dlp projector. Technical report, <http://www.mitsubishi-presentations.com/proj-pocket.asp>.
- [2] E. S. Bhasker, P. Sinha, and A. Majumder. Asynchronous distributed calibration for scalable reconfigurable multi-projector displays. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 12(5):1101–1108, 2006.
- [3] J. Y. Bouguet. Camera calibration toolbox for matlab. Technical report, <http://www.vision.caltech.edu/bouguetj/calib.doc/index.html>.
- [4] D. Brown. Decentering distortion of lenses. *Photometric Engineering*, 32(3):444 – 462, 1966.
- [5] H. Chen, R. Sukthankar, G. Wallace, and K. Li. Scalable alignment of large-format multi-projector displays using camera homography trees. *Proceedings of IEEE Visualization*, 2002.
- [6] Y. Chen, D. W. Clark, A. Finkelstein, T. Housel, and K. Li. Automatic alignment of high-resolution multi-projector displays using an uncalibrated camera. *Proceedings of IEEE Visualization*, 2000.
- [7] K. Cornelis, M. Pollefeys, and L. V. Gool. Lens distortion recovery for accurate sequential structure and motion recovery. *European Conference in Computer Vision*, pages 186 – 200, 2002.
- [8] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. *Proceedings of ACM Siggraph*, pages 369–378, 1997.
- [9] T. Disz, M. Papka, and R. Stevens. Ubiworld: An environment integrating virtual reality. *Heterogeneous Computing Workshop, Geneva, Switzerland*, 1997.
- [10] J. Fryer and D. Brown. Lens distortion for close-range photogrammetry. *Photogrammetric Engineering and Remote Sensing*, 52(1):51 – 58, 1986.
- [11] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Addison Wesley, 1992.
- [12] M. Hereld, I. Judson, and R. Stevens. Introduction to building projection-based tiled display systems. *IEEE Computer Graphics and Applications*, 2000.
- [13] M. Hereld, I. R. Judson, and R. Stevens. Dottyoto: A measurement engine for aligning multi-projector display systems. *Argonne National Laboratory preprint ANL/MCS-P958-0502*, 2002.
- [14] G. Humphreys, M. Houston, R. Ng, R. Frank, S. Ahem, P. Kirchner, and J. Klosowski. Chromium: A stream processing framework for interactive rendering on clusters. *ACM Transactions of Graphics*, 2002.
- [15] H. G. Jung, Y. H. Lee, P. J. Yoon, and J. Kim. Radial distortion refinement by inverse mapping-based extrapolation. *Proceedings to 18th International Conference on Pattern Recognition (ICPR)*, 2006.
- [16] J. Lai, A. Levas, P. Chou, C. Pinhanez, and M. Viveros. Bluespace: Personalizing workspace through awareness and adaptability. *International Journal of Human Computer Studies*, 57(5):415 – 428, 2002.
- [17] K. Li, H. Chen, Y. Chen, D. W. Clark, P. Cook, S. Damianakis, G. Essl, A. Finkelstein, T. Funkhouser, A. Klein, Z. Liu, E. Praun, R. Samanta, B. Shedd, J. P. Singh, G. Tzanetakis, and J. Zheng. Early experiences and challenges in building and using a scalable display wall system. *IEEE Computer Graphics and Applications*, 20(4):671–680, 2000.
- [18] A. Majumder and M. Gopi. Modeling color properties of tiled displays. *Computer Graphics Forum*, June 2005.
- [19] A. Majumder, Z. He, H. Towles, and G. Welch. Achieving color uniformity across multi-projector displays. *Proceedings of IEEE Visualization*, 2000.
- [20] A. Majumder and R. Stevens. Color nonuniformity in projection-based displays: Analysis and solutions. *IEEE Transactions on Visualization and Computer Graphics*, 10(2), March–April 2003.
- [21] A. Majumder and R. Stevens. Perceptual photometric seamlessness in tiled projection-based displays. *ACM Transactions on Graphics*, 24(1), January 2005.
- [22] T. Mitsunaga and S. Nayar. Radiometric self calibration. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1999.
- [23] S. K. Nayar, H. Peri, M. D. Grossberg, and P. N. Belhumeur. A projection system with radiometric compensation for screen imperfections. *Proceedings of IEEE International Workshop on Projector-Camera Systems*, 2003.
- [24] C. Pinhanez. The everywhere displays projector: A device to create ubiquitous graphical interfaces. *Proceedings of Ubiquitous Computing, Atlanta, Georgia*, 2001.
- [25] C. S. Pinhanez, F. C. Kjeldsen, A. Levas, G. S. Pingali, M. E. Podlaseck, and P. B. Chou. Ubiquitous interactive graphics. Technical Report RC22495 (W0205-143), IBM Research, 2002.
- [26] A. Raji, G. Gill, A. Majumder, H. Towles, and H. Fuchs. PixelFlex2: A Comprehensive, Automatic, Casually-Aligned Multi-Projector Display. *IEEE International Workshop on Projector-Camera Systems*, 2003.
- [27] R. Raskar. Immersive planar displays using roughly aligned projectors. In *Proceedings of IEEE Virtual Reality 2000*, 1999.
- [28] R. Raskar, M. Brown, R. Yang, W. Chen, H. Towles, B. Seales, and H. Fuchs. Multi projector displays using camera based registration. *Proceedings of IEEE Visualization*, 1999.
- [29] R. Raskar, J. van Baar, P. Beardsley, T. Willwacher, S. Rao, and C. Forlines. ilamps: Geometrically aware and self-configuring projectors. *ACM Transactions on Graphics*, 22(3), 2003.
- [30] R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, and H. Fuchs. The office of the future: A unified approach to image based modeling and spatially immersive display. In *Proceedings of ACM Siggraph*, pages 168–176, 1998.
- [31] S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice-Hall, 1995.
- [32] J. Salvi, X. Armangu, and J. Batlle. A comparative review of camera calibrating methods with accuracy evaluation. *Pattern Recognition*, 35:1617 – 1635, 2002.
- [33] P. Steurer and M. B. Srivastava. System design of smart table. *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications*, 2003.
- [34] R. Sukthankar, R. Stockton, and M. Mullin. Smarter presentations: Exploiting homography in cameraprojector systems. *Proceedings of International Conference on Computer Vision (ICCV)*, 2001.
- [35] R. Swaminathan and S. K. Nayar. Nonmetric calibration of wide-angle lenses and polycameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1172 – 1178, 2000.
- [36] J. Underkoffler, B. Ullmer, and H. Ishii. Emancipated pixels: Real-world graphics in the luminous room. *Proceedings of ACM Siggraph*, 1999.
- [37] R. Yang, D. Gotz, J. Hensley, H. Towles, and M. S. Brown. Pixelflex: A reconfigurable multi-projector display system. *Proceedings of IEEE Visualization*, 2001.
- [38] R. Yang, A. Majumder, and M. Brown. Camera based calibration techniques for seamless multi-projector displays. *IEEE Transactions on Visualization and Computer Graphics*, 11(2), March–April 2005.
- [39] Z. Zhang. Flexible camera calibration by viewing a plane from unknown orientation. *IEEE International Conference on Computer Vision*, pages 666 – 673, 1999.

Appendix B-8

User-Centric Design of a Vision System for Interactive Applications

Stanislaw Borkowski[†], Julien Letessier^{†§}, François Bérard[§], and James L. Crowley[†]

[†]INRIA Rhône-alpes

[§]CLIPS – IMAG

GRAVIR laboratory – PRIMA group

HCI group

655 avenue de l'Europe – Montbonnot

BP 53

38334 Saint Ismier Cedex – France

38041 Grenoble Cedex 9 – France

{stan.borkowski,julien.letessier,james.crowley}@inrialpes.fr

francois.berard@imag.fr

Abstract

Despite great promise of vision-based user interfaces, commercial employment of such systems remains marginal. Most vision-based interactive systems are one-time, “proof of concept” prototypes that demonstrate the interest of a particular image treatment applied to interaction. In general, vision systems require parameter tuning, both during setup and at runtime, and are thus difficult to handle by non-experts in computer vision. In this paper, we present a pragmatic, developer-centric, service-oriented framework for the construction of vision-based interactive systems. Our framework is designed to allow developers unfamiliar with vision to use computer vision as an interaction modality. To achieve this goal, we address specific developer- and interaction-centric requirements during the design of our system. We validate our approach with an implementation of standard GUI widgets (buttons and sliders) based on computer vision.

1. Introduction

Computer vision has the potential to provide a rich modality for user input in interactive systems. Numerous examples from the literature illustrate the diversity of application domains and interaction styles created using vision as input modality [7, 11, 9, 14, 2, 12, 10]. Yet, there are surprisingly few examples of commercially available vision-based interactive systems.

Human-computer interface (HCI) developers are used to hardware-based input devices that provide data “out of the box” with little or no off-line or runtime tuning. In contrast, vision-based input systems can only be integrated by HCI developers with great difficulty: they require vision expertise for setup, as well as execution conditions incompatible with typical HCI requirements. In this sense, most current computer vision systems do not *just work*.

We believe that to popularize vision-based interfaces, computer vision systems must be made more developer-friendly as well as more user-friendly. To accomplish this, it is necessary that vision system developers adopt a different perspective. In this paper, we present an approach to vision system design that aims at minimizing the difficulties related to the deployment of vision-based interactive systems by: (a) encapsulating vision components in isolated services, (b) imposing these services to meet specific usability requirements, and (c) limiting communications between the services and the interactive applications to a minimum. We describe our approach in section 3.

In section 4, we design a system for vision-based interactive widgets for augmented surfaces, and propose a breakdown into atomic services. In this context, we identify the necessary information exchange between services and with the application. Meeting the requirements detailed in section 3 influences the choices of the image processing algorithms used; in particular, they require little or no setup and tuning and feature usability-grade latency. Our implementation is robust to light changes and is moderate in CPU usage. Finally, in section 5 we detail an implementation of a simple vision-based calculator built using the described vision system.

In this paper, we consider the requirements of developers and end-users. By “developers”, we refer to the developers of the interactive systems aimed at “end-users”. In particular, these are not the developers of the vision system and we do not expect them to have computer vision expertise. By “end-users”, we refer to the users of the final system, who are assumed to have neither developer nor computer vision expertise.

2. Related work

Much work has been done in vision-based gesture recognition and novel interaction styles research, but the problem

of integrating vision to standard development has only recently received attention. Kjeldsen et al.[5] present an architecture for dynamically reconfigurable vision-based user interfaces (VB-UIs). They propose to separate the application from the vision engine. The core application controls the vision engine by sending XML messages specifying the current interface configuration. The interface is composed of interaction widgets defined at the level of functional core by their function and position in the interface. As end-users actuate widgets, the image processing engine sends messages that correspond to interaction events to the application's functional core.

The modular decomposition of vision components, as well as the separation of image treatment from the functional core of applications are important steps toward "developer-usable" VB-UI components. However, as presented, the VB-UI approach requires its developer to define sets of parameters and calibration data for the vision treatment engine for each interface configuration and location. Setting parameters for particular conditions allows fine tuning of the vision system. On the other hand, it requires the user of the vision system (the developer) to understand the underlying image processing algorithms, which is contradictory with the idea of VB-UI deployment by developers unfamiliar with vision.

In [13] Ye et al. present a framework for vision-based interactive components design. Both Kjeldsen et al. and Ye et al. propose to analyze camera input only in regions of interest around interactive "zones" of the interface, thus making the vision engine more economic in CPU usage. However Ye et al. go a step further, and instead of applying one vision algorithm per widget type, they propose to use a set of different detection techniques to obtain visual events called Visual Interface Cues (VIC). Each VIC detector called also *selector* examines a small part of the camera image for visual events that can be of different nature, like color, texture, motion or object geometry. Selectors are structured into hierarchies, and each selector can trigger one or more other selectors. Interaction events are detected based on sequences of selectors output. For instance, a touch sensitive button uses only motion detection at first. Once motion is detected, a VIC-based button switches to color-based image segmentation and to segmentation-based gesture recognition.

While the VIC paradigm presented in [13] is appealing for widget-oriented interface design, it lacks the analysis of VIC selectors suitability for interactive systems. Indeed, the VICs framework is presented from the perspective of an expert in computer vision. It does not consider setup and maintenance issues related to each selector such as threshold setting or lighting requirements. In section 4 we present a VIC-based implementation of basic interactive widgets that was conceived to respect the requirements identified in

section 3.

Finally, the Papier-Mâché [6] toolkit features a framework for the creation of multimodal interactive applications, in particular using computer vision. The toolkit design is based on a detailed user study, and offers an abstract, event-based model to work with multiple modalities. The vision part allows object recognition and tracking to be used as application input. While the design seems to be well-thought from the HCI perspective, it has several architectural shortcomings; in particular typical problems linked to object recognition (such as the aspect variations due to lighting changes) are not dealt with, and several thresholds must be set manually, on-line, by the developer using the toolkit.

3. Service-oriented design

Vision-based user interfaces rely on vision processes that extract high-level, abstract information from streams of images. This information is what is relevant for the interaction task. Except for "proof of concept" demonstrations created by computer vision researchers, this information is to be used by interactive application developers who have little to no expertise in the computer vision. This implies that a vision system that aims to be used outside the laboratory must be designed from a developer-centric point of view. In addition, all interactive systems, including vision based ones, must consider end-user requirements. In this section, we describe how user-centric and developer-centric requirements influence both the structure of a VB system and its application programming interface (API).

Our goal is to provide a set of computer vision services to developers of interactive systems. We aim at offering these services at different levels of abstraction. At the highest level, a service is an autonomous "black-box" application that, once launched, provides abstract information through some communication channel (a network socket, or a shared memory for examples). Developers should not need any knowledge of the vision processes. They simply connect to the black-box application and receive abstract events such as the position of the fingers in the case of a finger tracker service, or activation of triplets in the case of the SPOD service that we describe later. We facilitate the communication between the vision application and the developer's application by using a very simple communication protocol. This approach is developer-centric in the sense that we only expose and make easy to access what matters to the developer: the high-level events.

3.1. Non-functional requirements

Typical user-centric criteria against which an interactive system is evaluated include overall latency, reliability and

autonomy [7]. Here, we shortly describe these criteria and how they constrain the architecture of vision systems.

Latency. Interactive systems generally require a constraint on latency. For user input systems latency is measured as the time between the user action and the notification of the application. Typically, when using a vision-based finger tracker or a mouse to drag projected objects on a surface, the latency must be under 50 *ms* to optimize usability [8]. On the other hand, for a system that monitors the number of persons present in a room, a latency of 1 second may be acceptable. In consequence, a general architecture for vision services must be able to respect strict latency limits in communicating information to the application.

Autonomy. Vision-based systems usually require operator intervention for initial setup and maintenance. While acceptable in a laboratory setting, such intervention is poorly suited for deployment. Our “black-box” approach reduces the possibility of operator maintenance by reducing communication between the vision service and the developer’s application to a minimum.

Ideally, vision system developers should devise automatic maintenance solutions without offering a maintenance API. However, we acknowledge that designing fully autonomous vision system is a difficult goal that will not be reached in the short run for many services. When operator intervention is mandatory, one must take into account that the calibration task will be performed by a non-expert. Therefore, the maintenance procedure must be designed appropriately to minimize disrupting the end-user from her/his primary tasks.

Reliability. In most cases, information needs to be reliably conveyed between the input system and the application. For instance, an application that receives events when a person enters or exits an augmented room cannot afford to “miss” an event because its state would become incoherent. Losing some events from a tracking system, on the other hand, would not necessarily break the interaction, but would deteriorate the end-user experience. Therefore, our architecture must include reliable transport for interactive events.

3.2 Functional requirements

We propose a number of developer-centric requirements for our system in order to successfully address the target audience. These requirements include abstraction, isolation, and contract. Together they form the notion of a service.

Abstraction. To HCI developers, “abstracting the input [is] the most time consuming and challenging piece of application development” [6]. Since we assume the user of our vision system (the application developer) has no vision expertise, vision-specific information should not be made visible.

Even though coupling processing results with a confidence factor might be richer, this information is, in most cases, of no interest to developers. For instance, optical mice use normalized cross-correlation to determine the direction and speed of the movement. As long as the correlation coefficient is above a threshold they send positional data to the computer, otherwise they remain silent. The correlation coefficient is not sent to the system. The services internal data, such as the confidence factor, should not be considered as first-class output of services and should not be made visible by default. However, in cases where developers need to access this information, it should be available by explicit request from the developer.

End-user input should be abstracted to interaction events that are the same for a given class of input. For example, a finger tracker should produce “motion” events compatible with mouse “motion” events so that developers can, as a first step, quickly substitute vision-based input to existing event processing code. In other words, the abstracted data from the vision service should be in a form that is easily connectable to existing graphical user interface code. Consequently, the API of a VB input system for interaction should (a) render the vision aspects invisible, and (b) generalize the input used for a given task.

Isolation. Input subsystems like a VB-UI may need to be used by multiple applications, possibly running on different machines – for performance or geographic reasons. For instance, a video capture service might be used concurrently by a surveillance system located on a remote server, by a video-conference application, and by a motion capture service, both co-located to the camera. Since “a particular piece of input can be used for many different types of output” [6], information generated by a vision system must be shareable, and both remotely and locally accessible. Moreover, any input system for interaction should be easily extensible (for instance using output adapters, aggregators, or supervisor patterns) and embeddable in other services or in applications. Designing a VB input system as a federation of “black box” components therefore appears to be adequate.

Contract. Developers of toolkits or input devices usually establish a contract with the UI developer, in terms of HCI-centric, non-functional criteria. For instance, for positional or tracking input (mice, trackpads, laser trackers, etc.), the

relevant criteria are latency, precision, robustness, and autonomy.

- *latency* is implicitly under a usability threshold for standard tracking input devices (the textbook value [8] is around 50 *ms*);
- *precision* is also implicitly defined for a mouse/trackpad (under 1 display pixel, scaled with respect to the device gain)
- *robustness* is generally “absolute”. In other words the device or system either works or doesn’t. For instance an optical mouse “just works” as long as used on an adequate surface, and stops emitting positional events as soon as the contact stops.
- *autonomy*, or the lack of setup and maintenance, is also tacit for traditional devices.

In the case of VB input systems, these criteria are difficult to meet. Therefore, it is the role of the vision system designer to explicitly state in contract form the limitations with respect to these criteria. For instance, the system should notify the client application about robustness-related failures rather than provide incorrect or distorted information along with a (low) confidence factor. This corresponds to a binary quality of service evaluation and requires the system to perform introspection.

3.3. A pragmatic approach

Our approach is to generally isolate as much as possible the VB input system from the application that uses it, and minimize the communications between them. We propose to encapsulate the relevant services into independent, “black-box” processes that use only serialized communications.

To allow for low latency while preserving reliability, we propose to use traditional socket-based (TCP and UDP) communications. We use the TCP link to guarantee the connection between services or with the application, and to allow for reliable, high-latency communications, and the UDP link for low-latency communications. This constitutes the base of the “BIP/1.0” protocol we use for communications between services (draft specification at <http://www-prima.inrialpes.fr/prima/pub/Publications/index.php>).

System autonomy cannot be enforced at the architectural level. Nevertheless, not providing support for synchronous communications (thus making it difficult to implement transactions between the application and a service) limits the possibility of implementing setup mechanisms where automatic setup can be devised.

Finally, since the *contract* offered by a VB input system cannot be expressed simply as a part of an API, we propose to document it explicitly. This means that the developer of the system must evaluate it against the criteria presented above in order to circumscribe conditions of use for the vision system.

4. Basic services for vision-based UI design

To illustrate our service-oriented approach with an implementation, we choose to study vision systems applied to traditional interfaces. Typically, we investigate the use of graphical WIMP-like interfaces projected on surfaces of mundane objects. We assume that end-users are free to interact with projected images without wearing markers or actuating any hardware.

We break down our VB input system into three services: (1) the widget service, which the client will interoperate with; (2) the image capture service, which abstracts the camera, allowing it to be used by other systems eventually; and (3) the calibration service, which establishes the geometrical mapping between the camera view and the display.

4.1. Simple Pattern Occlusion Detectors

In [3] we presented an appearance-based implementation of touch sensitive projected buttons which we called “Sensitive widgets”. The presence of an object over a button on the interaction surface is detected by observing the change of perceived luminance over the button center area with respect to a reference area. By defining the reference area around the central-one, the button is made robust to complete occlusion, and sensitive to appearance changes made by elongated objects. Very simple image processing allows this system to maintain several dozens of sensitive widgets at camera frame rate (PAL-size images at 25 Hz) on a typical desktop computer. Moreover, this approach is robust to lighting changes, and thus suitable for configurations in which the interface is projected frontally from a video projector.

Implementation and evaluation of several interface prototypes based on sensitive widgets demonstrated that, from the end-user’s perspective, robustness to partial occlusions is also necessary. Indeed, a user pointing at a far-part of the interface would likely hover her/his arm over projected buttons, thus triggering partially occluded widgets. A partial, unsatisfactory solution was obtained by deactivating partly occluded widgets based on the input from widgets placed further away from the user. The idea of combining inputs from several sensitive widgets led us to re-think the touch detection approach.

We now choose to assemble atomic occlusion detectors, which are to be placed within and around widgets, in a way

that allows the system to distinguish some simple occlusion patterns. The geometry of the detectors (called “striplets”) is simplified to a rectangular strip. These detectors are assembled into a federation to provide the SPOD (Simple Pattern Occlusion Detectors) service. The SPOD service is internally divided into two separated layers: the image processing layer called the Striplot Engine (SE), in charge of image processing, and the Vision Events Interpretation Layer (VEIL), in charge of input abstraction. Both of these layers were designed to meet the requirements described in section 3.

Striplotes are defined as sensitive patches on the interface. Their response is calculated as the integral of the perceived luminance multiplied by a gain function over the surface of the striplot. The gain function has to be chosen so that the integral equals zero when the luminance over the whole striplot is constant. In current implementation the gain function is a symmetric step function with positive value over the central part of the striplot and negative value at both ends of the striplot (Figure 1).

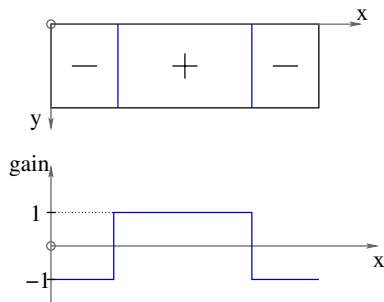


Figure 1. Striplot geometry and the gain function.

Striplotes are designed to detect occlusion by elongated objects. Each striplot is 40 millimeters long and 10 millimeters wide on the projected interface. These dimensions are chosen to ensure a maximal response to occlusions made by finger-sized objects occluding the striplot’s central area. Occlusions of any extremity of a striplot are intentionally ignored.

The camera coordinates of a striplot are calculated based on its position in the interface as given by the VEIL service and the camera-interface mapping. The event trigger threshold, on the other hand, is estimated individually for each striplot by the SE without any control from the client application. This threshold is automatically set to half of the maximal positive response of a striplot during interaction. The only assumption is that fingers contrast with the interface, which is true in most setups.

The VEIL is the “brain” of the SPOD service. The VEIL (a) translates widgets coordinates defined by the client application to a set of striplets coordinates, and (b) analyses the occlusion events generated by the SE and issues interaction events when appropriate.

Currently two types of interactive widgets are implemented: touch buttons and sliders. This allows a designer to build simple WIMP-like user interfaces. The button widget is composed of six striplets: two crossed in the center and four other surrounding the button center (Figure 2). Touch events are issued only if occlusion is detected by at least one of the center striplets and no more than one surrounding striplets. Sliders are simply obtained by assembling multiple partially overlapping buttons.

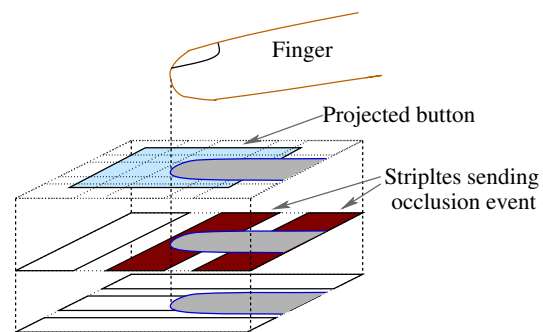


Figure 2. Button widget made of six striplets.

Since monocular vision systems cannot detect when a finger actually touches the interface, interaction events are generated only after detecting a short pause of the finger over a widget. The dwelling period, set to 250 ms, corresponds to the button-down event for mouse-based interaction. To move a slider, the end-user has to first “press” it and then drag it. This is coherent with existing WIMP interface behavior. In contrast, the dragging task requires from the end-user 2D movement coordination. If user’s finger exits the SPOD slider area, dragging is stopped.

By assembling striplot detectors in more sophisticated ways, it is also possible to develop different types of interactive widgets, for instance crossing-based menus [1]. An extreme case would cover the whole interface surface with SPOD-button-like structures, thus making a SPOD-based finger tracker.

Service API. The SPOD service requires the client application to specify the position of each interactive widget in a normalized coordinate frame of the interface. Additionally, the SPOD service needs to know the mapping between the camera view and the interface, as well as a rough estimate of the number of pixels per unit length of the interface in the scene. Both the mapping and the scale, are provided by a

calibration service (discussed below). The communication between SPOD service and calibration service is invisible to the client application. The SPOD service exclusively sends to the client application a stream of interaction events. All communication occurs via TCP/IP connections, using the BIP protocol described in section 3.3.

Inter-layer API. Both the SE and the VEIL are implemented as independent services, running in independent processes. Their communication also is asynchronous and event-driven, using BIP. For a given interface configuration, the VEIL sends to the SE coordinates of all triplets together with the interface-camera mapping matrix. The SE layer sends back to VEIL triplets state-change events that result from end-user interaction with the system.

Contract for the SPOD service. The initial SPOD service implementation can handle up to 300 triplets at camera frame rate (30 Hz) with images of 320x240 pixels size on a 2.8 GHz Pentium IV processor. In terms of widgets, this means the system can handle roughly 50 SPOD buttons simultaneously.

Because actuating buttons is not close coupled interaction, the latency is less of an issue. In fact, the VEIL makes the distinction between accidental occlusions and intentional actions based on a dwell time. On the other hand, the SE service is implemented to minimize latency.

Triplets only provide coarse resolution for finger positions. The resolution can be enhanced by averaging the position of several triplets detecting the same finger. Our implementation of a slider widget achieves a resolution of about 5 millimeters.

The SPOD service is made autonomous (i.e. except for the UI-camera mapping and scale there are no parameters to set), at the expense of robustness to certain condition changes. In particular, the SE layer would fail to detect occlusion from a finger on a dark background, it would also fail if the image contrast decreases due to a change in camera setting.

While the SPOD service was designed to respect the developer-centric requirements, it does not fully meet user-centric requirements. In particular, the simplistic automatic threshold estimation results in occasional false positive touch detections. However, the SPOD service can be described within the Visual Interface Cues (VICs) framework [13], with local luminance changes as the only visual cue. Using a similar approach for spacio-temporal gesture recognition like in [13], we can hope to alleviate the need of setting an occlusion threshold. Instead, triplets responses would be fed to a neural network based VEIL.

4.2. Support Services

Calibration consists in the geometrical coupling of the camera view (what the vision system perceives) and the displayed interface (what the application developer controls). The calibration service clients need to access the mapping information to transform vision information (e.g. positions in camera coordinates) into application-relevant data (e.g. positions in interface coordinates). This is achieved by providing the associated projective transformation in matrix form. Because the calibration service needs information from both the camera and the application, two approaches are possible: (a) If it controls the graphical output, it can work without interaction with the application. This is the case in the PDS example [3], where the interactive surface itself is tracked by the service. (b) In the general case, it must negotiate with the client application the display of a calibration grid [7].

Image acquisition service creates an abstraction of the camera. It allows concurrent access to the camera by multiple services. In our case, both the calibration service and the SE require access to the image stream. Low latency video sharing is implemented using shared memory buffers.

5. Application

Using the widget implementation described above, we have implemented a simple calculator application. The calculator interface can be projected and manipulated directly with fingers on the top of a desk. This interface allows a user to perform basic calculations such as addition, subtraction, division and multiplication. Numbers can be either typed on the calculator keyboard or chosen from the history buffer containing results of previous operations. The history buffer can be browsed using a slider on the left side of the calculator.

An informal evaluation of the calculator application, made by volunteers from our laboratory, showed that SPOD-widgets are easy to use and allow fast interface prototyping. A video showing the application working is available in the demo section of <http://www.prima.inrialpes.fr/>

6. Conclusions

This paper presents a developer oriented design approach for vision-based interactive systems. Inspired by [5], we decompose the vision-based applications to isolated processes of vision components and functional core of the application. The implementation of the vision-components draws on the VICs framework presented by Ye et. al in [13].

We believe that extending these two design approaches by an HCI-centric requirements analysis allows to build vision systems that can be used for interactive systems designed by developers unfamiliar with vision. Following the guidelines of the developer-centric and end-user-centric requirements analysis we implemented vision-based interactive widgets: buttons and sliders. We illustrate the feasibility of our approach with an implementation of a simple calculator for projection-augmented surfaces.

Acknowledgments

This work has been partially funded by the European project FAME (IST-2000-28323), the FGnet working group (IST-2000-26434), and the RNTL/Proact ContAct project.

References

- [1] G. Apitz and F. Guimbretière. Crossy: a crossing-based drawing application. In *UIST '04: Proceedings of the 17th annual ACM symposium on User interface software and technology*, pages 3–12, 2004.
- [2] F. Bérard. The magic table: Computer-vision based augmentation of a whiteboard for creative meetings. In *Proceedings of the ICCV Workshop on Projector-Camera Systems*. IEEE Computer Society Press, 2003.
- [3] S. Borkowski, J. Letessier, and J. L. Crowley. Spatial control of interactive surfaces in an augmented environment. In R. Bastide, P. A. Palanque, and J. Roth, editors, *EHCI/DS-VIS*, volume 3425 of *Lecture Notes in Computer Science*, pages 228–244. Springer, 2004.
- [4] J. L. Crowley, J. H. Piater, M. Vincze, and L. Paletta, editors. *Computer Vision Systems, Third International Conference, ICVS 2003, Graz, Austria, April 1-3, 2003, Proceedings*, volume 2626 of *Lecture Notes in Computer Science*. Springer, 2003.
- [5] R. Kjeldsen, A. Levas, and C. S. Pinhanez. Dynamically reconfigurable vision-based user interfaces. In Crowley et al. [4], pages 323–332.
- [6] S. R. Klemmer, J. Li, J. Lin, and J. A. Landay. Papier-mâché: toolkit support for tangible input. In *Proceedings of the 2004 conference on Human factors in computing systems (CHI'04)*, pages 399–406. ACM Press, 2004.
- [7] J. Letessier and F. Bérard. Visual tracking of bare fingers for interactive surfaces. In *UIST '04: Proceedings of the 17th annual ACM symposium on User interface software and technology*, pages 119–122. ACM Press, 2004.
- [8] I. S. MacKenzie and C. Ware. Lag as a determinant of human performance in interactive systems. In *Proceedings of the conference on Human factors in computing systems*, pages 488–493. Addison-Wesley Longman Publishing Co., Inc., 1993.
- [9] C. Pinhanez, R. Kjeldsen, T. Levas, G. Pingali, M. Podlaseck, and P. Chou. Ubiquitous interactive graphics. In *IBM Research Report RC22495 (W0205-143)*, May 2002.
- [10] R. S. Rao, K. Conn, S. H. Jung, J. Katupitiya, T. Kientz, V. Kumar, J. Ostrowski, S. Patel, and C. J. Taylor. Human robot interaction: Application to smart wheelchairs. In *Proceedings of the 2002 IEEE International Conference on Robotics and Automation (ICRA'02)*, pages 3583–3589, May 2002.
- [11] R. Raskar, J. van Baar, P. Beardsley, T. Willwacher, S. Rao, and C. Forlines. iLamps: geometrically aware and self-configuring projectors. *ACM Trans. Graph.*, 22(3):809–818, 2003.
- [12] P. Wellner. The digitaldesk calculator: Tactile manipulation on a desk top display. In *ACM Symposium on User Interface Software and Technology*, pages 27–33, 1991.
- [13] G. Ye, J. J. Corso, D. Burschka, and G. D. Hager. Vics: A modular vision-based hci framework. In Crowley et al. [4], pages 257–267.
- [14] Z. Zhang, Y. Wu, Y. Shan, and S. Shafer. Visual panel: virtual mouse, keyboard and 3d controller with an ordinary piece of paper. In *Proceedings of the 2001 workshop on Perceptive user interfaces*, pages 1–8. ACM Press, 2001.

Appendix C-1

Applications of Steerable Projector-Camera Systems

Claudio Pinhanez, Rick Kjeldsen, Anthony Levas, Gopal Pingali, Mark Podlaseck, Noi Sukaviriya

IBM Research, T.J. Watson

pinhanez,fcmk,levas,gpingali,podlasec,noi@us.ibm.com

Abstract

How can interactive computer interfaces be created anywhere in a space without wiring or modifying objects or people? We propose using steerable projector-camera systems employing computer vision to realize such "steerable interfaces." In this paper, we illustrate the potential of the new kinds of applications enabled by steerable interfaces and discuss the challenges imposed on computer vision through the presentation of four application prototypes: a collaborative assembly task coordinator; a multi-surface presentation viewer; a ubiquitous product finder for retail environments; and an interactive merchandise shelf.

1. Introduction

The main goal of our research is to create devices capable of realizing the *ubiquitous computing (ubicomp)* vision of everywhere computer access and seamless integration of computers with the real world [1]. However, unlike the most commonly proposed devices that have to be either worn or carried by the user (such as cell phones or PDAs), or embedded in the environment, we are investigating devices that can create interactive displays with minimal physical change to the user or to the environment.

In this context, using projectors to create interfaces becomes a natural candidate, especially when combined with cameras and computer vision systems able to directly detect user interaction with the projected surface. The combination projector-camera is thus able to create *projected interfaces* on most planar surfaces without requiring any modification on it.

However, if the goal is to provide multiple points of seamless computer access in an environment, it is not practical to install pairs of projector-cameras facing every possible surface. Even in the unlikely case that projectors become extremely cheap, installation of these devices will still be costly. Even worse, any change in the configuration of the environment will require re-installation and re-wiring of many devices.

To cope with those problems, we have been advocating the use of *steerable projected interfaces* [2, 3], referred in this paper simply as *steerable interfaces*. Steerable interfaces can be created, for example, by fitting a pan/tilt

mirror into a LCD projector and combining it with a pan/tilt/zoom camera.

We have described the basic structure of the vision systems used in our prototypes in previous publications [4, 5]. In this paper we focus on four different application prototypes that explore the potential of steerable interfaces. These applications employ vision systems that detect point-and-click interaction, toolbar scrolling, and simple user activity recognition. However, unlike traditional vision-based interfaces, these vision systems are required to work seamlessly on distinct surfaces under very different observation angles and lighting conditions. They also have to cope with the changes in the pattern of user interaction caused by moving the interface to a new surface, which alters the relative size and orientation of the gesture to the interface. In some ways, these applications require computer vision systems much more robust and adaptable to contextual changes than currently used vision systems.

This paper presents these new computer vision-based applications that are only possible due to use of steerable projectors and cameras. As we present these applications, we analyze and discuss the challenges created by steerable interfaces to computer vision and the ways we have addressed them. We conclude by suggesting possible ways to improve the performance of steerable vision systems.

2. Related Work

Deviceless interaction with projected images based on video processing was pioneered in the 1980s by Krueger [6] in systems that often used a backlit white background to simplify segmentation. Real-time static background segmentation became a reality in the 1990s, notably in the ALIVE system [7] and the technique has been a cornerstone of most segmentation work in vision-based gesture recognition.

However, only a little of that work has attended to the problems of segmentation in the context of a projected display. Because the projected image drastically changes the appearance of the hand, we tend to agree with Hardenberg and Berard [8] in that background subtraction does not give reliable results in the presence of strong projected images.

Different proposals have been made to use hand or body gestures to interact with the projected images. In

general, gesture-based interaction methods tend to fall into one of two categories: either variations of a point-and-click paradigm [8, 9] or application-specific pose or motion gestures [10]. Both these approaches have disadvantages. While the point-and-click approach is often not well suited to hand pointing due to limited resolution and the lack of a natural “click”, a large number of application-specific gestures makes interfaces complex: users have to learn and execute these gestures largely without feedback, as discussed in [11].

3. Computer Vision for Steerable Interfaces

The applications described in this paper use a device to create steerable interfaces called the *Everywhere Displays projector (ED-projector)* [2]. Figure 1 shows a prototype of this device, made of off-the-shelf components: a LCD projector, a pan/tilt mirror used in theatrical lighting, and a videoconference pan/tilt/zoom camera.

The system is currently run on two different computers, one responsible for the application software and the projector software; the other controlling the positioning of the camera and runs the computer-vision system responsible for detecting user interaction. The projector software stores and manages the calibration parameters to correct the distortion caused by oblique projection as well as the projector focus and zoom parameters. In the current implementation, these parameters are determined manually as described in [2].

Our methods to detect user activity involve the interpretation of the path of the user’s hand. Because the projection can radically change the appearance of objects moving through it, appearance-based techniques were ruled out. Instead motion is detected by looking at the differences between adjacent images in the video stream. Hands are tracked by looking for fingertips in the filtered motion data. Candidate fingertips are identified based on shape and evaluated using spatial heuristics such as orientation, distance from the user’s body, and the previous fingertip locations.

To build an interface, application designers assemble “widgets” that specify areas where users can interact. Each type of widget responds differently to the user’s motions. A touch widget looks for an in-pause-back motion of the hand. A continuous motion widget reports the relative location of a fingertip in one or two dimensions. The system can also track objects of distinct colours. As we see, our system employs traditional computer vision methods in the implementation of the basic widget set (see [4] for details).

Most real-life vision interface systems incorporate the vision system as a module that is hard-coded to operate under a fixed set of circumstances. Our system, on the other hand, requires the application to send the vision system a description of the user interface as a



Figure 1. Prototype of the ED-projector.

configuration of widgets (describing *what* the interface is). Based on these specifications, the vision system assembles a set of image processing components that implement the interface, sharing computational resources when possible. To change the interaction, a new interface description can be sent to the system at any time.

The system, described in detail in [5], also provides for the deployment of an interface onto different real-world planar surfaces. Like Zhang et al. [12], we dynamically map the interactive interface based on the four corners of the contour of the surface. However, in our system the parameters of the surfaces where the interface can be realized are represented and referenced independently of any particular interface. These include the size, location and perspective distortion within the image and characteristics of the physical environment around that surface, such as the user’s likely position while interacting with it. When the application requests an interface to be activated on a particular surface (that is, defines *where* the interaction should happen in the environment), the system retrieves the surface parameters and propagates them through the assembly of image processing components that implements that interface.

By explicitly decoupling the information describing the characteristics of *where* an interface happens in an environment, i.e., surface-specific information, we enable: (1) the porting an application to a new environment where the interaction surfaces could be different; (2) the use of one surface by multiple applications; and (3) the use of the same interface on multiple surfaces.

4. Applications Using Steerable Interfaces

Steerability considerably expand the potential of projected interface systems by making them devices much more adaptable to environmental changes, especially when compared to embedded displays and monitors. For example, if the furniture of an office is rearranged, it is



Figure 2. The M&M picture assembly task: a) choosing a color; b) “clicking” a bucket after getting the M&Ms; c) placing the M&Ms on the highlighted areas; and d) “finger painting” to reveal the complete picture.

necessary to rewire monitors and, many times, remove and reinstall embedded displays. Instead, if a steerable interface system is used, adaptation is simply a matter of recalibrating the position and parameters of the interactive surfaces by software.

However, the addition of steerability to a projected interface system goes beyond easy and flexible deployment of interfaces on different surfaces. It also enables a new set of applications that explore more direct connections to real-world objects. In particular, it allows the creation of augmented reality applications where information about objects is projected directly onto or around them, freeing the users from the inconvenience of wearing head displays and cumbersome input devices and avoiding the difficulty of registering the computer generated imagery to the real-world view.

The possibilities opened up by the use of steerable interfaces to create different and exciting applications come with a price. On the display side, there are issues such as reduction of perceived resolution, constraints on environment brightness, and occlusion, as discussed in [2].

4.1. Collaborative Assembly Task Coordinator

Steerable interfaces can augment reality without requiring users to wear goggles or other cumbersome devices. To demonstrate this premise, we developed for a the Emergent Technologies Exhibit of SIGGRAPH'01 a whimsical demonstration of a system that uses steerable interfaces to coordinate a collaborative assembly task. In this prototype application, the object to be assembled is a picture made of 3,000 M&M's (multi-colored sugar-coated chocolates) where each M&M is regarded as a pixel of the picture. Each user contributes to the task by placing 10 to 15 pixels. Here is a description of the main characteristics of this application:

Context: a collaborative assembly task in a factory.

Application Goal: to guide users to find the components to be assembled and to help placing them in the appropriate position.

Technology: the application uses an ED-projector overlooking the assembly area. The vision system implements widgets that detect touch events and that report continuous motion of the user's hand on a surface.

Interaction: as the user enters the environment, she is asked to touch a picture projected on a board to activate the system. The system responds by projecting an instruction asking the user to go to the selection area. The interface is then steered towards a table. Figure 2.a shows this table as it is transformed into an interactive menu for color selection. The user simply touches the color of his choice to select it. Then the user receives a projected instruction prompting him to look for a highlighted bucket. In Figure 2.b, the system highlights the bucket that contains the M&M's of the selected color and provides additional information on how many M&M's to extract from the bucket. The interface also includes a touchable button, where the user indicates that the retrieval of M&M's is completed. When this button is touched, the system projects on the bucket an instruction telling the user to walk to the picture board. In Figure 2.c, the system is pointing to the exact places on the picture board where the M&M's should be placed. After the M&M's are put in place, the user can signal the end of this task by touching a button projected on the side of the picture board. The user is then invited to wave her hand over the surface to progressively reveal the rest of the picture being built (see also the video, from the system's camera point of view, at <http://www.research.ibm.com/people/p/pinhanez/download/video1.avi>).

The computer vision system faced many issues in this application. The first difficulty involved the interactive menu for color selection. As shown in Figure 2.a, this menu presented a choice among 18 buttons, arranged in a

Table 1: Performance of the vision system (130 users)

	success	false positives	false negatives
Entrance	88%	9%	3%
select board	94%	2%	4%
buckets	55%	33%	12%
board (place)	91%	8%	2%
board (paint)	80%	2%	18%
total w/o buckets	89%	6%	5%
total	81%	12%	7%

grid configuration closely juxtaposed to each other. Correct selection was possible because the system looked for the complete gesture of touching and not simply by the presence of the hand in a particular spot. Therefore, the user could select a color in the top row without triggering the buttons on the rows below. As we see here, simple tracking without gesture recognition is often insufficient to implement even basic projected interfaces.

A second major challenge for the vision system was to realize the bucket interface dialog onto the many different buckets. At the time of this demonstration, the automatic mapping of configurations of widgets onto surfaces described in section 3 was not implemented, so each interface had to be painstakingly mapped to each surface that realized it. In fact, the difficulties experienced during this demonstration were a major factor to develop the description-based interface system described in section 3..

More than 650 novice users participated in the collaborative assembly task demonstration. Considering that the vision system and the projection system were integrated only weeks preceding the exhibition, the combined system worked remarkably well. As shown in Table 1, a sample with 130 consecutive users with 621 button touch events (touching gestures or false detections) yielded correct detection of touching gestures in 81% of the events, with 12% false negatives and 7% false positives. If the bucket events are excluded from the count, the performance exceeds 89%.

The buckets yielded a high number of errors for several reasons. The biggest problem was that while picking M&M's from the bucket, the user often partially or completely occluded the display with their head or back, accidentally triggering the button projected on the bucket. A possible way to handle this is to include in the vision system improved methods to distinguish between the shapes of hands and other body parts. Currently, the system only looks for fingertip-shaped patterns on the motion image. However, if view-based methods are used to identify body parts larger than a fingertip, the apparent shapes can vary dramatically from surface to surface due to the different relative positioning of camera.

Although we choose to use the M&M's placement task to demonstrate steerable interfaces in augmented reality, it

**Figure 3. Anywhere presentation viewer.**

should be evident the applicability of this technology in real-world assembly tasks and similar augmented reality scenarios.

4.2. Anywhere Presentation Viewer

Another important area of applications of steerable interfaces involves the access to desktop applications from different surfaces. A typical scenario occurs when people assemble in an office or meeting room to review electronic documents. According to the number of participants and their roles, documents are preferably viewed by individuals on monitor-like displays, by large groups of people on vertical screens, or by a small number of people in configurations that better facilitate collaborative work such as the one shown in Figure 3.

Trying to address these problems, we developed a presentation viewer that can move to different surfaces:

Context: office or meeting room.

Application Goal: to create a viewer for presentations that can be accessed on walls or tables to facilitate collaborative work.

Technology: an ED-projector overlooks the working area. The vision system utilizes only touch widgets; the interfaces are dynamically defined by the application and mapped automatically onto the calibrated surfaces.

Interaction: using a touch-screen LCD monitor, the user selects through an icon a surface in the environment to launch the application. The presentation is projected on the selected surface, topped by buttons allowing the users to select the next, the previous, or the first slides, or to dismiss the viewer. The viewer can be moved to other surface by selecting the corresponding icon on the LCD monitor.

This application clearly demonstrates the need of a vision system that can automatically map any application interface onto any given surface. In real offices and meeting rooms, tables move to different areas, walls are set up and removed, etc. Easily coping with these environmental changes is an essential feature of the vision system here.

A difficulty with this application, particularly in collaborative scenarios such as the one depicted in Figure 3, is that the context invites multi-user interaction. If that is the case, buttons have to respond to activation patterns from different directions corresponding to where each user is positioned relative to the interface. Such ability is not currently supported by our system: it assumes that all touch widgets in an interface are accessed from the same direction. Supporting multi-directional access requires either tracking and differentiating multiple fingertip paths simultaneously, which is currently very error prone, or monitoring the location of every user. In the second case, the vision system may use contextual information provided by an environmental user positioning tracking system. We are also evaluating more sophisticated hand path representation and recognition methods.

4.3. Ubiquitous Product Finder

Computerized access to information in public spaces such as malls and stores has been primarily deployed through kiosks. A major advantage of using virtual display systems such as the ED-projector in such spaces is that they can be installed on the ceiling, thereby preventing accidental damage and discouraging vandalism and theft.

We are exploring such ubicomp applications in public spaces in the context of retail environments. The first application developed in this framework is a system to help customers to find products in a large store. A major requirement for this application is ubiquity, i.e., the product finder has to be easily accessible from anywhere in the store. Another requirement is that it has to be simple to learn and use.

Figure 5 shows a conceptual view of the *ubiquitous product finder*. ED-projectors are installed in every aisle, and small white boards placed in strategic locations on the shelves. These boards can be easily moved to different positions to respond to marketing and customer needs. A large table is placed at the entrance to the store (on the right side of Figure 4). On it, an enlarged, intuitive-to-use version of the interface is projected aiming to make customers acquainted to the system. Signage boards are hung up from the ceiling or placed on top of the shelves, allowing the system to project directions to the product requested by the customer or to display advertisements when the system is idle.

A full-scale version of this application has not been deployed yet. Here, we describe a laboratory prototype of the system we have built and tested.

Context: a large retail environment.

Application Goal: to help customers everywhere in the store to find products by pointing directly to where they are located.

Technology: (laboratory prototype) an ED-projector is mounted on the ceiling. A table with a wooden red slider

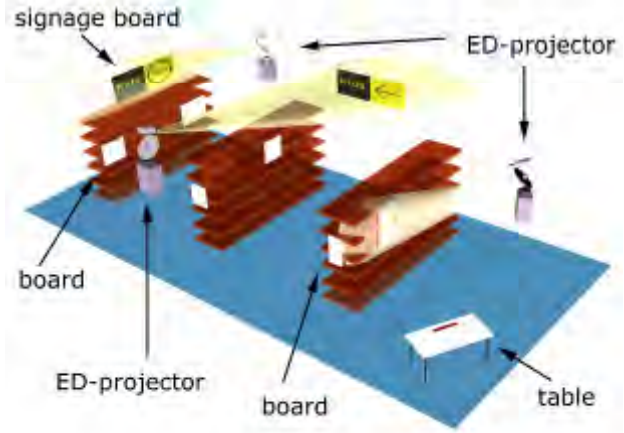


Figure 5 Concept of ubiquitous product finder.

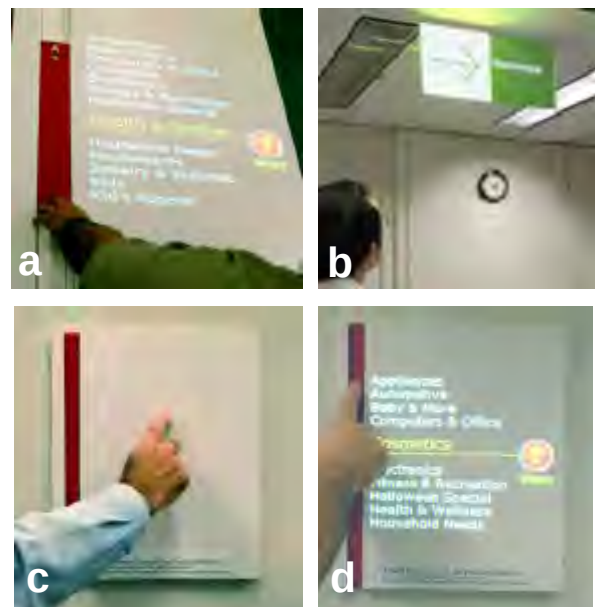


Figure 4. Prototype of ubiquitous product finder.

is used for projecting a larger introductory version of the product finder interface. Two smaller foam boards, visually similar to the table with the slider, are placed on the walls representing the smaller replicas of the same interface used in other areas of the store. The vision system is able to detect touch events, hand position and movement over the foam boards, and the position of the wooden slider relative to the table.

Interaction: the product list is projected on the product finder table located at the entrance of the store. The user can move the wooden red slider on the left of the table to find a product as illustrated in Figure 4.a. As he does this, the product list projected to the right of the slider scrolls up and down mirroring the motion of the slider. Once the user touches the “where” symbol on the right, arrows guiding the user to the location of the highlighted product are projected on signage boards hanging from the ceiling

as illustrated in Figure 4.b (see also video at <http://www.research.ibm.com/people/p/pinhanez/download/video2.avi>). Alternatively, the user may initiate the product finder from the foam boards at other locations of the store by touching the boards as shown in Figure 4.c. This action triggers the system to project the product finder interface onto these foam boards as shown in Figure 4.d. In the wall version, the user slides his finger up and down a red velvet strip to scroll the product list (see <http://www.research.ibm.com/people/p/pinhanez/download/video3.avi>).

Our laboratory prototype of this application currently does not address two important issues present in real stores. First, we use a camera-based user tracking system (based on [13]) to detect proximity to the foam boards. If a user is close to one of them, we aim the ED-projector camera to the board surface and, without projecting any information, wait for the user to “call” the system by touching the board. This is not a practical solution in real store where multiple customers may be close to different foam boards at the same time. We envision here the use of an array of steerable cameras that can be dedicated exclusively to monitor the foam boards. The second issue is the inability of the system to help more than one customer at a time. In particular, if many boards share an ED-projector, we may have situations where users will have to wait for their turn to use the product finder. The severity of this is yet to be tested in a real-world deployment, but we expect that customers may tolerate reasonable waiting times as the situation is similar to when they consult an information desk or a sales assistant.

The flexible structure of our computer vision system simplified significantly the implementation of the laboratory prototype. To detect the movements of the wooden slider of Figure 4.a, a widget to detect linear movements of blocks of solid color was implemented. In the case of the foam boards, fingertip tracking is used to determine its position on the velvet strip as shown in Figure 4.d. Switching between the two different widgets is accomplished by the application simply sending the vision system the description of the appropriate interface.

We initially had some difficulties with the interaction on the velvet strip. After scrolling through the list and finding the desired product, users move away to touch the “where” button on the right, sometimes causing unwanted scrolling before they exit the tracking region. We solved the problem heuristically, ignoring quick motions after a pause that immediately exit the tracking region

This problem is common to many gesture-based interfaces: the segmentation of continuous gestures into distinct “phrases”. In our case the problem is simplified somewhat by using the spatial adjacency of gestures to interaction widgets. As with the problem of continuous speech, there is some amount of “co-articulation” that

affects the ends of adjacent phrases, complicating their recognition. Gestural phrasing has been an issue for many years but, despite many different proposals (such as [14]), it is still largely unsolved.

4.4. Interactive Clothing Bins

Another application developed in the context of a retail environment uses a checkerboard arrangement of bins containing women’s pants and projectable surfaces as shown in Figure 6. It explores steerable interfaces as a way to augment the functionality of everyday objects and furniture. In this particular application, the system detects user interest with the product in a particular bin and automatically brings information designed to assist the customer.

Context: a retail environment.

Application Goal: to assist customers interested in a product by displaying information about it.

Technology: an ED-projector is aimed at the clothing bins. In our particular implementation, the system is unable to project an image large enough to cover all projectable surfaces on the bins. Instead, it selects groups of juxtaposed surfaces and steers the interface towards them according to the user hand’s position, i.e., the top parts of the bins, the bottom part of the bins, etc. The vision system can detect user activity in each of the clothing bins and touch events. No surface contains any sensor or wire. A camera-based user tracking system (similar to [13]) is used to determine the position of the user relative to the clothing bins.

Interaction: When the user is at a distance from the bins, a circulating series of advertisements for women’s clothing are projected on the panels, trying to attract customers to the area. When the user approaches the bins, the display is changed to advertise the store’s credit card and special promotions. Once the user starts examining pants in a bin, information pertaining to the pants in that bin is displayed in the proximity of the bin. Additional interactive panels are projected allowing the user to find information about similar pants styles, to check the size chart, or to check whether the desired size and color is available in stock (see demonstration video at <http://www.research.ibm.com/people/p/pinhanez/download/video4.avi>).

The major difference between this application and the ones previously examined is that this system proactively tries to engage and help the customer. By recognizing user activity in a particular bin (by simple hand tracking) and interpreting it as an interest in the products stocked there, the application takes the “liberty” of augmenting the environment with information. This example clearly demonstrates the potential of steerable interface systems to create new models of interaction by directly engaging

people while they execute everyday activities in the real world. Since application interfaces can be created on areas very close to where the activities are happening, user attention can be more easily grabbed. This kind of engaging application involving physical objects cannot be easily implemented with traditional embedded displays such as monitors or kiosks.

This example also shows how computer vision methods and algorithms for recognition of human activity and behavior can be used for the development of applications with steerable interfaces. Our system determines user interest in a particular bin by detecting long duration hand movements within the bin. The performance achieved with this simple measurement is adequate for a research prototype but we clearly see opportunity here for more sophisticated methods of action and behavior recognition.

5. Improving the Performance of the Vision System of a Steerable Interface

By decoupling the functional definition of the interface from the specification of its location in the physical environment and in the camera image, the ED-projector vision system supports a great deal more flexibility than most vision-based interaction systems. This flexibility requires significant architectural complexity to dynamically realize an efficient recognition process from the task and surface descriptions.

Up to this point, our efforts have mostly concentrated on developing this flexible internal architecture [5], while the interaction widgets themselves have been comparatively simple. We are now beginning to concentrate on developing widgets that are capable of supporting more interesting interactions. This work is proceeding along several dimensions.

People have difficulty seeing the exact shape of a moving object, so natural gestures most often consist of a sequence of motions and static poses, where a hand movement is followed by a static pose, possibly followed by another movement, and so on [9]. In order to recognize hand poses, it will be necessary to either incorporate techniques to recognize shape in the sparse and noisy motion data or to incorporate hand segmentation techniques that do not rely on motion. However, as mentioned above, to avoid problems caused by the change in appearance of the user's hands as they pass through the projected image, the current interactions have all been based on detection and analysis of motion in the video stream. One possible alternative is to use a segmentation method that uses knowledge about the image being projected, the physical surface where the interface is being projected on, and how they interact, so that it can estimate the appearance of the projection to the camera. By using techniques such as those presented in [15], it may be



Figure 6. Interactive clothing bins.

possible to estimate the appearance of the projection in the digitized video stream. This will also allow interactions to take place in the presence of a changing projection, such as a projected video clip.

Interestingly, the flexibility provided by computer vision allows interactions that are more abstract than the manipulative hand gestures implemented to date. One line of investigation concerns intentional control actions that the user takes with parts of the body other than their hands. These include interactions where whole body movements such as waves or jumps are used to control the interaction [16].

In addition to extending the gesture recognition capabilities of the system, we feel performance can be significantly improved by enabling the system to adapt to its current context. There are several areas where adaptation may be important. For instance, the current widget set requires information about the expected size and orientation of the user's hand. By making better use of user tracking information and observing user movements as they approach a widget, it is possible to determine this type of information automatically.

Every person moves with different rhythms and patterns. When users interact with a system using physical devices like a keyboard or mouse, the device standardizes and masks many of these differences, but because our vision system directly senses and interprets a user's movements, it can be sensitive to them. Techniques similar to [17] can allow adaptation to an individual user's movement patterns.

One constraint on any interaction system is that it be able to respond to the user smoothly and in a timely manner. For a vision-based system this implies it runs at a sufficiently high frame rate. Research with other interaction methods [18] has shown that tracking rates on the order of 10 Hz and response times less than 200ms are needed for a system to feel comfortable, except in the case of manipulation of virtual objects. If the frame rate falls below what is required to support these response times there are several trade-offs that can be adjusted to

compensate. For example the incoming resolution may be reduced, trading off accuracy for speed. We are exploring techniques to do this automatically.

6. Conclusion

We have shown in this paper how possible ways to use computer vision in applications based on steerable projected interfaces. Our basic approach involves using motion data to identify the user's hand and path analysis to identify touch events, hand movement, and simple manipulative actions.

A major factor enabling the development of three of the prototype systems was the application-driven dynamic definition of interfaces and their automatic mapping to any surface known by the vision system. With this architecture (defined in detail in [5]), surface and viewpoint-specific characteristics are abstracted from the development process of the application.

Some shortcomings of the current system and challenges for computer vision have also been identified as a result of our work developing these applications. In general, projected interfaces limit the use of appearance-based techniques. Also, steerability requires gesture recognition algorithms to be viewpoint independent or, at least, adaptable to new viewpoints if basic contextual information is provided. Gesture segmentation based on spatial location and greater use of contextual information such as user position contributed to the success of the system. As the complexity of the underlying widgets increases, a challenge will be to ensure that the additional complexity does not spill over to affect the application developer or the end user.

Finally, in this paper we demonstrate that ubicomp applications can be implemented based on the steerable interface paradigm. The extension that this paradigm will become a major part of the ubiquitous computing reality substantially depends, in our view, on the development of adequate and robust computer vision systems that satisfy the constraints imposed by projection and steerability.

References

- [1] M. Weiser, "The Computer for the Twenty-First Century," *Scientific American*, vol. 265, pp. 94-100, 1991.
- [2] C. Pinhanez, "The Everywhere Displays Projector: A Device to Create Ubiquitous Graphical Interfaces," in Proc. of Ubiquitous Computing 2001 (Ubicomp'01), Atlanta, Georgia, 2001.
- [3] G. Pingali, C. Pinhanez, A. Levas, R. Kjeldsen, M. Podlaseck, H. Chen, and N. Sukaviriya, "Steerable Interfaces for Pervasive Computing Spaces," in IEEE International Conference on Pervasive Computing and Communications - PerCom'03, Dallas-Fort Worth, Texas, 2003.
- [4] F. Kjeldsen, C. Pinhanez, G. Pingali, J. Hartman, A. Levas, and M. Podlaseck, "Interacting with Steerable Projected Displays," in Proc. of the 5th International Conference on Automatic Face and Gesture Recognition (FG'02), Washington, DC, 2002.
- [5] R. Kjeldsen, A. Levas, and C. Pinhanez, "Dynamically Reconfigurable Vision-Based User Interfaces," in 3rd International Conference on Vision Systems (ICVS'03), Graz, Austria, 2003.
- [6] M. W. Krueger, *Artificial Reality II*: Addison-Wesley, 1990.
- [7] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder: Real-Time Tracking of the Human Body," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, pp. 780-785, 1997.
- [8] C. v. Hardenberg and F. Berard, "Bare-hand human-computer interaction," in Proc. of Workshop on Perceptive User Interfaces, PUT'01, Orlando, Florida, 2001.
- [9] F. Quek, T. Mysliwiec, and M. Zhao, "Finger mouse: A freehand pointing interface," in Proc. of International Workshop on Automatic Face- and Gesture-Recognition, Zurich, Switzerland, 1995.
- [10] J. Segen, "GestureVR: Vision-Based 3D Hand Interface for Spatial Interaction," in Proc. of ACM Multimedia Conference, Bristol, England, 1998.
- [11] F. Kjeldsen, "Visual Recognition of Hand Gesture as a Practical Interface Modality." New York, New York: Columbia University, 1997.
- [12] Z. Zhang, Y. Wu, Y. Shan, and S. Shafer, "Visual Panel: Virtual Mouse, Keyboard, and 3D Controller with an Ordinary Piece of Paper," in Proc. ACM Perceptual/Perceptive User Interfaces Workshop (PUT'01), Florida, USA, 2001.
- [13] A. Senior, A. Hampapur, Y.-L. Tian, L. Brown, S. Pankanti, and R. Bolle, "Appearance Models for Occlusion Handling," in Proc. of Workshop on Performance Evaluation of Tracking and Surveillance (PETS2001), 2001.
- [14] J. W. Deng and H. T. Tsui, "An HMM-Based Approach for Gesture Segmentation and Recognition," in International Conference on Pattern Recognition (ICPR'00), Barcelona, Spain, 2000.
- [15] L. B. Wolff, S. K. Nayar, and M. Oren, "Improved Diffuse Reflection Models for Computer Vision," *International Journal of Computer Vision*, vol. 30, 1998.
- [16] W. Freeman, K. Tanaka, J. Ohta, and K. Kyuma, "Computer Vision for Computer Games," in Proc. Second Intl. Conf. on Automatic Face and Gesture Recognition (FG'96), Killington, Vermont, 1996.
- [17] M. A. Giese and T. Poggio, "Synthesis and Recognition of Biological Motion Patterns Based on Linear Superposition of Prototypical Motion Sequences," in Proceedings of the MVIEW 99 Symposium at CVPR'99, Fort Collins, Colorado, 1999.
- [18] S. Card, T. Moran, and A. Newell, "The Keystroke-Level Model for User Performance Time with Interactive Systems," in *Readings in Human-Computer Interaction: A Multidisciplinary Approach*, R. Baecker and W. Buxton, Eds.: Morgan Kaufmann, 1987.

Appendix C-2



Claudio Pinhanez

IBM Research - Brazil / C4AI - Univ. of São Paulo

Artificial Intelligence
 Conversational Systems
 HCI
 Ubiquitous Computing

GET MY OWN PROFILE

	All	Since 2018
Citations	4954	1096
h-index	37	14
i10-index	92	29

0 articles1 article

not availableavailable

Based on funding mandates

TITLE	CITED BY	YEAR
The everywhere displays projector: A device to create ubiquitous graphical interfaces C Pinhanez Ubicom 2001: Ubiquitous Computing: International Conference Atlanta Georgia ...	551	2001
The kidsroom: A perceptually-based interactive and immersive story environment SS Intille, JW Davis, CS Pinhanez, LW Campbell, YA Ivanov, A Wilson In PRESENCE: Teleoperators and Virtual Environments	505 *	1999
Multiple-surface display projector with interactive input capability CS Pinhanez US Patent 6,431,711	183	2002
Method and system for a user-following interface FCM Kjeldsen, A Levas, GS Pingali, CS Pinhanez US Patent 7,134,080	163	2006
Human action detection using pnf propagation of temporal constraints CS Pinhanez, AF Bobick Proceedings. 1998 IEEE Computer Society Conference on Computer Vision and ...	162	1998
Typefaces and the perception of humanness in natural language chatbots H Candello, C Pinhanez, F Figueiredo Proceedings of the 2017 chi conference on human factors in computing systems ...	107	2017
Interacting with steerable projected displays R Kjeldsen, C Pinhanez, G Pingali, J Hartman, T Levas, M Podlaseck Proceedings of Fifth IEEE International Conference on Automatic Face Gesture ...	105	2002
Steerable interfaces for pervasive computing spaces G Pingali, C Pinhanez, A Levas, R Kjeldsen, M Podlaseck, H Chen, ... Proceedings of the First IEEE International Conference on Pervasive ...	101	2003
Method and apparatus for image projection, and apparatus controlling image projection K Binsted, F Nielsen, C Pinhanez US Patent 6,554,431	93	2003

TITLE	CITED BY	YEAR
Interval scripts: A design paradigm for story-based interactive systems CS Pinhanez, K Mase, A Bobick Proceedings of the ACM SIGCHI Conference on Human factors in computing ...	82	1997
A study on the manipulation of 2D objects in a projector/camera-based augmented reality environment S Volda, M Podlaseck, R Kjeldsen, C Pinhanez Proceedings of the SIGCHI conference on Human factors in computing systems ...	81	2005
BlueSpace: personalizing workspace through awareness and adaptability J Lai, A Levas, P Chou, C Pinhanez, M Viveros International Journal of Human-Computer Studies 57 (5), 415-428	73	2002
Fostering a symbiotic handheld environment M Raghunath, C Narayanaswami, C Pinhanez Computer 36 (09), 56-65	70	2003
Perceptual user interfaces: the KidsRoom AF Bobick, SS Intille, JW Davis, F Baird, CS Pinhanez, LW Campbell, ... Communications of the ACM 43 (3), 60-61	67	2000
Representation and recognition of action in interactive spaces CS Pinhanez Massachusetts Institute of Technology	67	1999
Using a steerable projector and a camera to transform surfaces into interactive displays C Pinhanez CHI'01 Extended Abstracts on Human Factors in Computing Systems, 369-370	65	2001
Large-scale multi-agent-based modeling and simulation of microblogging-based online social network M Gatti, P Cavalin, SB Neto, C Pinhanez, C dos Santos, D Gribel, ... Multi-Agent-Based Simulation XIV: International Workshop, MABS 2013, Saint ...	60	2014
Kidsroom: Action recognition in an interactive story environment A Bobick, J Davis, S Intille, F Baird, L Cambell, Y Irinov, C Pinhanez, ... Mass. Inst. Technol., Cambridge, Perceptual Computing Tech. Rep 398	57	1996
Approximate world models: Incorporating qualitative and linguistic information into vision systems CS Pinhanez, AF Bobick PROCEEDINGS OF THE NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE, 1116-1123	57	1996
System and method for positioning projectors in space to steer projections and afford interaction G Pingali, C Pinhanez, M Podlaseck, A Levas, F Kjeldsen US Patent App. 10/735,053	56	2005

TITLE	CITED BY	YEAR
Dynamically reconfigurable vision-based user interfaces R Kjeldsen, A Levas, C Pinhanez Computer Vision Systems: Third International Conference, ICVS 2003, Graz ...	56	2003
Computer theater CS Pinhanez Proc. of the Eighth International Symposium on Electronic Arts (ISEA'97)	52	1996
To frame or not to frame: The role and design of frameless displays in ubiquitous applications C Pinhanez, M Podlaseck UbiComp 2005: Ubiquitous Computing: 7th International Conference, UbiComp ...	50	2005
Dynamically reconfigurable vision-based user interfaces R Kjeldsen, A Levas, C Pinhanez Machine Vision and Applications 16, 6-12	50	2004
Physically interactive story environments CS Pinhanez, JW Davis, S Intille, MP Johnson, AD Wilson, AF Bobick, ... IBM systems journal 39 (3.4), 438-455	49	2000
Centralized display for mobile devices C Narayanaswami, MT Raghunath, CS Pinhanez US Patent 7,512,402	43	2009
Services as customer-intensive systems C Pinhanez Design Issues 25 (2), 3-13	42	2009
Using symbiotic displays to view sensitive information in public S Berger, R Kjeldsen, C Narayanaswami, C Pinhanez, M Podlaseck, ... Third IEEE International Conference on Pervasive Computing and ...	42	2005
Intelligent studios: Using computer vision to control TV cameras CS Pinhanez, AP Pentland Perceptual Computing Section, Media Laboratory, Massachusetts Institute of ...	42	1995
Applications of steerable projector-camera systems C Pinhanez, R Kjeldsen, A Levas, G Pingali, M Podlaseck, N Sukaviriya Proceedings of the IEEE International Workshop on Projector-Camera Systems ...	41	2003
Embedding Interactions in a Retail Store Environment: The Design and Lessons Learned. PN Sukaviriya, M Podlaseck, R Kjeldsen, A Levas, G Pingali, ... INTERACT	41	2003
That's entertainment! Designing streaming, multimedia web experiences CM Karat, J Karat, J Vergo, C Pinhanez, D Riecken, T Cofino International Journal of Human-Computer Interaction 14 (3-4), 369-384	41	2002
User-following displays G Pingali, C Pinhanez, T Levas, R Kjeldsen, M Podlaseck	41	2002

TITLE	CITED BY	YEAR
Proceedings. IEEE International Conference on Multimedia and Expo 1, 845-848		
Less Clicking, More Watching: Results of the Iterative Design and Evaluation of Entertaining Web Experiences. CM Karat, CS Pinhanez, J Karat, R Arora, J Vergo INTERACT, 447-454	40	2001
"It/I" a theater play featuring an autonomous computer graphics character CS Pinhanez, AF Bobick Proceedings of the sixth ACM international conference on Multimedia ...	40	1998
" Less Clicking, More Watching": Results from the User-Centered Design of a Multi-Institutional Web Site for Art and Culture. J Vergo, CM Karat, J Karat, C Pinhanez, R Arora, T Cofino, D Riecken, ... Archives & Museum Informatics, 2008 Murray Ave., Suite D, Pittsburgh, PA 15217	38	2001
Augmenting reality with projected interactive displays B Fisher, K Dawson-Howe, C O'Sullivan, C Pinhanez Virtual and Augmented Architecture (VAA'01) Proceedings of the International ...	38	2001
User input apparatus, system, method and computer program for use with a screen having a translucent surface C Pinhanez, G Pingali, F Kjeldsen, A Levas, M Podlaseck US Patent App. 10/981,151	36	2006
Augmenting a retail environment using steerable interactive displays N Sukaviriya, M Podlaseck, R Kjeldsen, A Levas, G Pingali, C Pinhanez CHI'03 Extended Abstracts on Human Factors in Computing Systems, 978-979	36	2003
Hyper Mask-Talking Head Projected onto Real Object T Yotsukura, F Nielsen, K Binsted, S Morishima, CS Pinhanez The Visual Computer 18 (2), 111-120	36	2002
The effect of audiences on the user experience with conversational interfaces in physical spaces H Candello, C Pinhanez, M Pichiliani, P Cavalin, F Figueiredo, ... Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems ...	35	2019
Creating touch-screens anywhere with interactive projected displays C Pinhanez, R Kjeldsen, L Tang, A Levas, M Podlaseck, N Sukaviriya, ... Proceedings of the eleventh ACM international conference on Multimedia, 460-461	35	2003
Method and apparatus for indoor navigation FC Kjeldsen, A Levas, GS Pingali, CS Pinhanez, ME Podlaseck, ... US Patent 8,155,872	33	2012
Service systems as customer-intensive systems and its implications for service science and engineering C Pinhanez Proceedings of the 41st Annual Hawaii International Conference on System ...	33	2008

TITLE	CITED BY	YEAR
Ubiquitous interactive graphics C Pinhanez, R Kjeldsen, A Levas, G Pingali, M Podlaseck, P Chou IBM Research Report RC22495 (W0205-143)	33	2002
BlueSpace: Creating a personalized and context-aware workspace P Chou, M Gruteser, J Lai, A Levas, S McFaddin, C Pinhanez, M Viveros, ... Relatório técnico 1, 10-75	31	2001
Bottester: testing conversational systems with simulated users M Vasconcelos, H Candello, C Pinhanez, T dos Santos Proceedings of the XVI Brazilian Symposium on Human Factors in Computing ...	30	2017
A simulation-based approach to analyze the information diffusion in microblogging online social network MA de C Gatti, AP Appel, CN dos Santos, CS Pinhanez, PR Cavalin, ... 2013 Winter Simulations Conference (WSC), 1685-1696	30	2013
"It/I": a theater play featuring an autonomous computer character CS Pinhanez, AF Bobick Presence: Teleoperators & Virtual Environments 11 (5), 536-548	29	2002
Using approximate models as source of contextual information for vision processing A Bobick, C Pinhanez Proc. of the ICCV 95, 13-21	29	1995
Method and system for a user-following interface A Levas, GS Pingali, CS Pinhanez US Patent 8,589,796	25	2013
On interfaces projected onto real-world objects M Podlaseck, C Pinhanez, N Alvarado, M Chan, E Dejesus CHI'03 extended abstracts on Human factors in computing systems, 802-803	25	2003
Ubiquitous computing in the living room: Concept sketches and an implementation of a persistent user interface SS Intille, V Lee, C Pinhanez Proceedings of UBICOMP	24	2003
Humans inside" as the key characteristic of service systems C Pinhanez Proceedings of QUIS 11, 515-524	23	2009
Creating ubiquitous interactive games using everywhere displays projectors C Pinhanez Entertainment Computing: Technologies and Application, 149-156	23	2003
A hybrid architecture for multi-party conversational systems MG de Bayser, P Cavalin, R Souza, A Braz, H Candello, C Pinhanez, ... arXiv preprint arXiv:1705.01214	22	2017

TITLE	CITED BY	YEAR
Method and system for simulation of online social network AP Appel, MA de Cerqueira Gatti, SMB Neto, CS Pinhanez, ... US Patent App. 13/887,354	22	2014
Remote set-up and calibration of an interactive system S Borger, FCM Kjeldsen, GS Pingali, CS Pinhanez US Patent 8,162,486	22	2012
Information needs of system administrators in information technology service factories CRB de Souza, CS Pinhanez, VF Cavalcante Proceedings of the 5th ACM Symposium on Computer Human Interaction for ...	21	2011
Controlling view-based algorithms using approximate world models and action information AF Bobick, CS Pinhanez Proceedings of IEEE Computer Society Conference on Computer Vision and ...	21	1997
A service science perspective on human-computer interface issues of online service applications C Pinhanez International Journal of Information Systems in the Service Sector (IJISSS ...	20	2009
Intelligent studios modeling space and action to control tv cameras CS Pinhanez, AF Bobick Applied Artificial Intelligence 11 (4), 285-305	20	1997
Interval scripts: a programming paradigm for interactive environments and agents CS Pinhanez, AF Bobick Personal and Ubiquitous Computing 7, 1-21	19	2003
Fast constraint propagation on specialized allen networks and its application to action recognition and control C Pinhanez, A Bobick Submitted to AAAI 98	19	1998
Pnf propagation and the detection of actions described by temporal intervals CS Pinhanez, AF Bobick Proc. of the DARPA Image Understanding Workshop	19	1997
Hypermask talking head projected onto real object S Morishima, T Yotsukura, K Binsted, F Nielsen, C Pinhanez Multimedia Modeling: Modeling Multimedia Information and Systems, 403-412	18	2000
Role-based tracking and surveillance AP Appel, MA de Cerqueira Gatti, RA De Paula, CS Pinhanez US Patent 9,613,277	17	2017

TITLE	CITED BY	YEAR
Collaboration and social computing in emerging financial services DR Millen, C Pinhanez, J Kaye, SCS Bianchi, J Vines Proceedings of the 18th ACM Conference Companion on Computer Supported ...	16	2015
A proposal for a service science discipline classification system C Pinhanez, P Kontogiorgis IBM Corporation	16	2008
Can Web entertainment be passive C Pinhanez, CM Karat, J Vergo, J Karat, R Arora, D Riecken, T Cofino Proceedings of WWW01	16	2001
Improving out-of-scope detection in intent classification by using embeddings of the word graph space of the classes P Cavalin, VHA Ribeiro, A Appel, C Pinhanez Proceedings of the 2020 Conference on Empirical Methods in Natural Language ...	15	2020
Specifying and Implementing Multi-Party Conversation Rules with Finite-State-Automata. MG de Bayser, MA Guerra, PR Cavalin, CS Pinhanez AAAI Workshops, 713-719	14	2018
Having an Animated Coffee with a Group of Chatbots from the 19th Century H Candello, C Pinhanez, MC Pichiliani, MA Guerra, M Gatti de Bayser Extended abstracts of the 2018 CHI conference on human factors in computing ...	14	2018
Flood data collection and warning mechanism CS Pinhanez US Patent 8,614,631	14	2013
A services theory approach to online service applications C Pinhanez IEEE International Conference on Services Computing (SCC 2007), 395-402	14	2007
Teaching robots to act and converse in physical spaces: participatory design fictions with museum guides H Candello, M Pichiliani, M Wessel, C Pinhanez, M Muller Proceedings of the Halfway to the Future Symposium 2019, 1-4	13	2019
Ubiquitous interactive displays in a retail environment C Pinhanez, M Podlaseck, R Kjeldsen, A Levas, G Pingali, N Sukaviriya Proceedings of SIGGRAPH Sketches	13	2003
System and Method for Credit Score Based on Informal Financial Transactions Information SCS Bianchi, HCDSP Candello, DR Millen, CS Pinhanez US Patent App. 14/863,470	12	2017
Designing conversational interfaces H Candello, C Pinhanez Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais-IHC	12	2016

TITLE	CITED BY	YEAR
A ciência da inovação em serviços: estudo exploratório sobre os interesses e prioridades para uma agenda de pesquisa no Brasil PC Calabria, RC Bernardes, ER de Vargas, CS Pinhanez RAI Revista de Administração e Inovação 10 (4), 110-135	12	2013
Transforming Surfaces into Touch-Screens C Pinhanez, F Kjeldsen, A Levas, GS Pingali, J Hartman, A Levas, ... submitted to CHI	12	2001
Using computer vision to control a reactive computer graphics character in a theater play HI Christensen, C Pinhanez, A Bobick Computer Vision Systems: First International Conference, ICVS'99 Las Palmas ...	12	1999
Design Decisions for Interactive Environments: Evaluating the KidsRoom A Bobick, S Intille, J Davis, F Baird, C Pinhanez, L Campbell, Y Ivanov, ... Proceedings of	12	1998
System, method and computer program product for assessing the capabilities of a conversation agent via black box testing A Braz, HCDSP Candello, CS Pinhanez, MA Vasconcelos US Patent 10,102,846	11	2018
Different but equal: Comparing user collaboration with digital personal assistants vs. teams of expert agents CS Pinhanez, H Candello, MC Pichiliani, M Vasconcelos, M Guerra, ... arXiv preprint arXiv:1808.08157	11	2018
Recovering from dialogue failures using multiple agents in wealth management advice H Candello, C Pinhanez Studies in conversational UX design, 139-157	11	2018
An architecture and framework for steerable interface systems A Levas, C Pinhanez, G Pingali, R Kjeldsen, M Podlaseck, N Sukaviriya UbiComp 2003: Ubiquitous Computing: 5th International Conference, Seattle ...	11	2003
" Less clicking, more watching" an option for entertainment on the web? C Pinhanez, CM Karat, J Karat, J Vergo, R Arora CHI'01 Extended Abstracts on Human Factors in Computing Systems, 347-348	11	2001
Age density patterns in patients medical conditions: A clustering approach F Alhasoun, F Aleissa, M Alhazzani, LG Moyano, C Pinhanez, ... PLOS Computational Biology 14 (6), e1006115	10	2018
Shaping the experience of a cognitive investment adviser H Candello, C Pinhanez, D Millen, BD Andrade Design, User Experience, and Usability: Understanding Users and Contexts ...	10	2017

TITLE	CITED BY	YEAR
A service science perspective for interfaces of online service applications C Pinhanez Proceedings of the VIII Brazilian Symposium on Human Factors in Computing ...	10	2008
A portable system for anywhere interactions N Sukaviriya, R Kjeldsen, C Pinhanez, L Tang, A Levas, G Pingali, ... CHI'04 Extended Abstracts on Human Factors in Computing Systems, 789-790	10	2004
Segmenting social media users by means of life event detection and entity matching PR Cavalin, MG De Bayser, JPF De Melo, CS Pinhanez US Patent 10,108,714	9	2018
Ravel: a mas orchestration platform for human-chatbots conversations MG de Bayser, C Pinhanez, H Candello, M Affonso, MP Vasconcelos, ... The 6th International Workshop on Engineering Multi-Agent Systems (EMAS ...	9	2018
Evaluating the conversation flow and content quality of a multi-bot conversational system H Candello, M Vasconcelos, C Pinhanez Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors ...	9	2017
Method and system for automated service climate measurement based on social signals CS Pinhanez US Patent 8,121,890	9	2012
Computer theater: Stage for action understanding CS Pinhanez, AF Bobick Proc. of the AAAI'96 Workshop on Entertainment and AI/A-Life, 28-33	9	1996
Machine teaching by domain experts: towards more humane, inclusive, and intelligent machine learning systems C Pinhanez arXiv preprint arXiv:1908.08931	8	2019
A social network analysis framework for modeling health insurance claims data AP Appel, VF de Santana, LG Moyano, M Ito, CS Pinhanez arXiv preprint arXiv:1802.07116	8	2018
Real-time sentiment analysis in social media streams: The 2013 confederation cup case PR Cavalin, MAC Gatti, CN dos Santos, C Pinhanez Proceedings of BRACIS/ENIAC 2014	8	2014

Appendix C-3

Any time

Since 2023

Since 2022

Since 2019

Custom range...

2003 — 2005

Search

Sort by relevance

Sort by date

☑ Create alert

Applications of steerable projector-camera systems

☐ Search within citing articles

Searchlight-a lightweight search function for pervasive environments

[A Butz](#), [M Schneider](#), [M Spassova](#) - *Pervasive*, 2004 - Springer

We present a lightweight search function for physical objects in instrumented environments. Objects are tagged with optical markers which are scanned by a steerable camera and ...

☆ Save [Cite](#) Cited by 107 [Related articles](#) [All 24 versions](#)

Accommodating transient connectivity in ad hoc and mobile settings

[R Handorean](#), [C Gill](#), [GC Roman](#) - ... 2004, Linz/Vienna, Austria, April 21-23 ..., 2004 - Springer

Much of the work on networking and communications is based on the premise that components interact in one of two ways: either they are connected via a stable wired or ...

☆ Save [Cite](#) Cited by 103 [Related articles](#) [All 20 versions](#)

Active pursuit tracking in a projector-camera system with application to augmented reality

[S Gupta](#), [C Jaynes](#) - ... *Vision and Pattern Recognition (CVPR'05)* ..., 2005 - [ieeexplore.ieee.org](#)

We introduce an active object tracking algorithm that maintains relative calibration between a camera projector pair for augmented-reality applications. The approach is motivated by the ...

☆ Save [Cite](#) Cited by 17 [Related articles](#) [All 4 versions](#)

A demanufacturing projector-vision system for combined manual and automated processing of used electronics

[M Bailey-Van Kuren](#) - *Computers in industry*, 2005 - Elsevier

In order to process the growing quantities of used electronics, efficient demanufacturing methods are needed. The industry utilizes manual methods for dismantling products. A ...

☆ Save [Cite](#) Cited by 2 [Related articles](#) [All 7 versions](#)

スポットライトによる物探し支援システム

中田豊久, 金井秀明, 國藤進 - *情報処理学会研究報告グループ* ..., 2005 - [ipsj.ixsq.nii.ac.jp](#)

論文抄録 隠れていても頻りに移動する物でも, その物にスポットライトをあてるシステムを構築した. そのアプリケーションとして「部屋の中での紛失物探し」と「本棚での関連本探し」...

☆ Save [Cite](#) Cited by 5 [Related articles](#) [All 5 versions](#) [◇◇](#)

[PDF] UbiComp: Becoming Superhuman

[J Scott](#) - *Proceedings of the UbiComp 2005 Workshop on* ..., 2005 - [microsoft.com](#)

UbiComp: Becoming Superhuman Page 1 UbiComp: Becoming Superhuman James Scott Intel Research Cambridge [james.w.scott@intel.com](#) Abstract. UbiComp aims to offer ...

☆ Save [Cite](#) Cited by 1 [Related articles](#) [All 2 versions](#) [◇◇](#)

An intelligent disassembly assistant for man-machine demanufacturing

[MB Kuren](#), [J Soltani](#) - *Proceedings, 2005 IEEE/ASME* ..., 2005 - [ieeexplore.ieee.org](#)

In order to handle the complexity of demanufacturing operations, a combined man-machine system is proposed that incorporates projector-vision technology. As part of this system, an ...

☆ Save [Cite](#) [Related articles](#) [All 2 versions](#)

A Projector-Vision System for Automated Demanufacturing

[M Bailey-Van Kuren](#) - *IEEE/ASME International Conference on* ..., 2005 - [ieeexplore.ieee.org](#)

In order to process the growing quantities of used electronics, efficient demanufacturing methods are needed. The industry utilizes manual methods for dismantling products. A ...

☆ Save [Cite](#) [Related articles](#)

[CITATION] Projizierte Benutzerschnittstellen und steuerbare Projektoren

[M Schicker](#) - *Ausarbeitung zum gleichnamigen Vortrag*

☆ Save [Cite](#) Cited by 1 [Related articles](#)

Appendix C-4

Alois Ferscha
Friedemann Mattern (Eds.)

LNC3 3001

Pervasive Computing

Second International Conference, PERVASIVE 2004
Linz/Vienna, Austria, April 2004
Proceedings



Springer

Amazon Ex. 1008.0210

Alois Ferscha Friedemann Mattern (Eds.)

Pervasive Computing

Second International Conference, PERVASIVE 2004
Linz/Vienna, Austria, April 18-23, 2004
Proceedings



Springer

Amazon Ex. 1008.0211

Volume Editors

Alois Ferscha
Johannes Kepler Universität Linz
Institut für Pervasive Computing
Altenberger Str. 69, 4040 Linz, Austria
E-mail: ferscha@soft.uni-linz.ac.at

Friedemann Mattern
ETH Zürich (Swiss Federal Institute of Technology)
Department of Computer Science, Institute for Pervasive Computing
Haldeneggsteig 4, 8092 Zürich, Switzerland
E-mail: mattern@inf.ethz.ch

Library of Congress Control Number: 2004103931

CR Subject Classification (1998): C.2.4, C.3, C.5.3, D.4, H.3-5, K.4, K.6.5, J.7

ISSN 0302-9743

ISBN 3-540-21835-1 Springer-Verlag Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag. Violations are liable to prosecution under the German Copyright Law.

Springer-Verlag is a part of Springer Science+Business Media
springeronline.com

© Springer-Verlag Berlin Heidelberg 2004
Printed in Germany

Typesetting: Camera-ready by author, data conversion by PTP-Berlin, Protago-TeX-Production GmbH
Printed on acid-free paper SPIN: 10996911 06/3142 5 4 3 2 1 0

Amazon Ex. 1008.0212

Table of Contents

Activity Recognition

Activity Recognition from User-Annotated Acceleration Data.....	1
<i>Ling Bao, Stephen S. Intille</i>	
Recognizing Workshop Activity Using Body Worn Microphones and Accelerometers	18
<i>Paul Lukowicz, Jamie A Ward, Holger Junker, Mathias Stäger, Gerhard Tröster, Amin Atrash, Thad Starner</i>	
“Are You with Me?” – Using Accelerometers to Determine If Two Devices Are Carried by the Same Person	33
<i>Jonathan Lester, Blake Hannaford, Gaetano Borriello</i>	

Context Computing

Context Cube: Flexible and Effective Manipulation of Sensed Context Data.....	51
<i>Lonnie Harvel, Ling Liu, Gregory D. Abowd, Yu-Xi Lim, Chris Scheibe, Chris Chatham</i>	
A Context-Aware Communication Platform for Smart Objects.....	69
<i>Frank Siegemund</i>	
Siren: Context-Aware Computing for Firefighting	87
<i>Xiaodong Jiang, Nicholas Y. Chen, Jason I. Hong, Kevin Wang, Leila Takayama, James A. Landay</i>	

Near Body Interfaces

Spectacle-Based Design of Wearable See-Through Display for Accommodation-Free Viewing	106
<i>Marc von Waldkirch, Paul Lukowicz, Gerhard Tröster</i>	
A Compact Battery-Less Information Terminal for Real World Interaction	124
<i>Takuichi Nishimura, Hideo Itoh, Yoshiyuki Nakamura, Yoshinobu Yamamoto, Hideyuki Nakashima</i>	

Software

INCA: A Software Infrastructure to Facilitate the Construction and Evolution of Ubiquitous Capture and Access Applications	140
<i>Khai N. Truong, Gregory D. Abowd</i>	

Sensors

Activity Recognition in the Home Using Simple and Ubiquitous Sensors	158
<i>Emmanuel Munguia Tapia, Stephen S. Intille, Kent Larson</i>	
Automatic Calibration of Body Worn Acceleration Sensors	176
<i>Paul Lukowicz, Holger Junker, Gerhard Tröster</i>	
Reject-Optional LVQ-Based Two-Level Classifier to Improve Reliability in Footstep Identification	182
<i>Jaakko Suutala, Susanna Pirttikangas, Jukka Riekk, Juha Röning</i>	
Issues with RFID Usage in Ubiquitous Computing Applications	188
<i>Christian Floerkemeier, Matthias Lampe</i>	

Security

A Fault-Tolerant Key-Distribution Scheme for Securing Wireless Ad Hoc Networks	194
<i>Arno Wacker, Timo Heiber, Holger Cermann, Pedro José Marrón</i>	
ProxNet: Secure Dynamic Wireless Connection by Proximity Sensing	213
<i>Jun Rekimoto, Takashi Miyaki, Michimune Kohno</i>	
Tackling Security and Privacy Issues in Radio Frequency Identification Devices	219
<i>Dirk Henrici, Paul Müller</i>	

Architectures and Systems

Towards Wearable Autonomous Microsystems	225
<i>Nagendra B. Bharatula, Stijn Ossevoort, Mathias Stäger, Gerhard Tröster</i>	
Ubiquitous Chip: A Rule-Based I/O Control Device for Ubiquitous Computing	238
<i>Tsutomu Terada, Masahiko Tsukamoto, Keisuke Hayakawa, Tomoki Yoshihisa, Yasue Kishino, Atsushi Kashitani, Shojiro Nishio</i>	
eSeal – A System for Enhanced Electronic Assertion of Authenticity and Integrity	254
<i>Christian Decker, Michael Beigl, Albert Krohn, Philip Robinson, Uwe Kubach</i>	

Algorithms

A Distributed Precision Based Localization Algorithm for Ad-Hoc Networks	269
<i>Leon Evers, Stefan Dulman, Paul Havinga</i>	

Adaptive On-Device Location Recognition	287
<i>Kari Laasonen, Mika Raento, Hannu Toivonen</i>	
Accommodating Transient Connectivity in Ad Hoc and Mobile Settings	305
<i>Radu Handorean, Christopher Gill, Gruia-Catalin Roman</i>	
New Interfaces	
Microbiology Tray and Pipette Tracking as a Proactive Tangible User Interface	323
<i>Harlan Hile, Jiwon Kim, Gaetano Borriello</i>	
Augmenting Collections of Everyday Objects: A Case Study of Clothes Hangers as an Information Display	340
<i>Tara Matthews, Hans-W. Gellersen, Kristof Van Laerhoven, Anind K. Dey</i>	
MirrorSpace: Using Proximity as an Interface to Video-Mediated Communication	345
<i>Nicolas Roussel, Helen Evans, Heiko Hansen</i>	
SearchLight – A Lightweight Search Function for Pervasive Environments	351
<i>Andreas Butz, Michael Schneider, Mira Spassova</i>	
Author Index	357

SearchLight – A Lightweight Search Function for Pervasive Environments

Andreas Butz, Michael Schneider, and Mira Spassova

Department of Computer Science, Saarland University,
Stuhlsatzenhausweg, Bau 36.1, 66123 Saarbrücken, Germany
{butz,mschneid,mira}@cs.uni-sb.de

Abstract. We present a lightweight search function for physical objects in instrumented environments. Objects are tagged with optical markers which are scanned by a steerable camera and projector unit on the ceiling. The same projector can then highlight the objects when given the corresponding marker ID. The process is very robust regarding calibration, and no 3D model of the environment is needed. We discuss the scenario of finding books in a library or office environment and several extensions currently under development.

1 Introduction

Ubiquitous computing landscapes extend our physical surroundings by a computational layer providing new functionalities. One such functionality can be the capability of objects to make themselves known in order to be noticed or found by humans. This functionality was already proposed in the original Ubiquitous Computing vision [8]. A search function for physical environments would alleviate the need to keep track of all of the things in our environment.

One obvious application is keeping track of books in our office, in a library or a book store. Let's think of a library with ubiquitous display capabilities. Let's assume there will still be a conventional inquiry terminal to find out about books. The inquiry interface on the computer terminal in a library could then just provide an additional "show this book" button for the selected book, which prompts the environment to highlight its position on the shelf.

Let's think of an instrumented office in which there are many books and other things, either neatly sitting on the shelf where they belong, or left somewhere in the room on a pile. A physical search functionality will find and highlight missing objects for us, no matter where we left them. In this paper we describe such a search function for physical objects, and how it was implemented in our instrumented environment.

2 The SUPIE Environment

The Saarland University Pervasive Instrumented Environment (SUPIE) is an instrumented room with various types of displays, sensors, cameras, and other

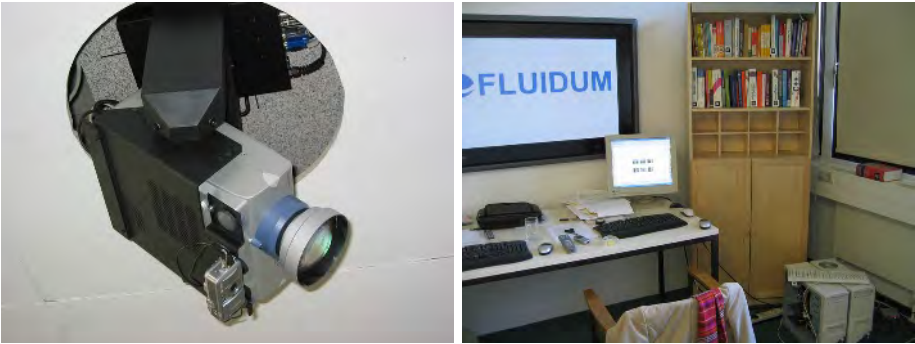


Fig. 1. The steerable projector mounted on the ceiling (left) and one corner of SUPIE with a book shelf (right)

devices. By means of a steerable projector, the room has continuous display capabilities which are limited in resolution and temporal availability (only one area at a time). The various other displays provide islands of higher resolution and interactivity within this display continuum.



Fig. 2. Images taken during the scanning process (left) and an example of an AR Toolkit marker (right)

The projector can project arbitrary light patterns onto all surfaces in the room with line of sight to its position in the center of the ceiling. An attached camera can take high resolution pictures from (almost) the same position and scan them for objects and markers. It also provides a low resolution video stream in real time for the recognition of movements.

3 Implementation of *SearchLight*

3.1 General Thoughts

All that is needed to implement *SearchLight* is the steerable projector with its camera and the capability to recognize optical markers. The only determining factors for finding objects and highlighting them, are the pan and tilt angles of the unit moving the projector. If a marker is found at given pan and tilt angles, these can be stored and used directly for projecting a highlight area around the corresponding object. Thus, not even a 3D model of the room is needed. Absolute registration errors of the camera and projector will cancel each other out since they are mounted in one unit and moved together.

3.2 Devices

The steerable projector is mounted on a moving yoke produced by a stage equipment manufacturer, controlled via a DMX interface, and mounted in the center of the ceiling of the room. The unit carries a 3.000 ANSI lumen projector with a lens of long focal length, in order to provide sufficient contrast and brightness in daylight. The camera attached to the projector (see figure 1 right) is a regular Digicam with an image resolution of four Megapixels. The high resolution pictures are triggered and read out over USB with the camera's freely available SDK. Each object that should be found by *SearchLight* is tagged by an AR Toolkit marker (see [1]). Each marker has a large black fringe and an individual small black symbol on a white background. (See figure 2) These markers are recognized by the Java version of AR Toolkit, the *JAR Toolkit*.

3.3 Implementation

SearchLight performs two main tasks: It *scans* the room for markers and memorizes the corresponding angles, which are used later to *show* searched objects.

Scan. The room is scanned by taking slightly overlapping pictures in all horizontal and vertical directions (see Fig. 2 left). Each picture is analyzed using JAR Toolkit. Marker IDs are stored in a list together with their pan and tilt angle, derived from their position and orientation in the picture and the orientation of the moving yoke when taking the picture.

Show. After the room has been scanned, the user can search for marked objects. The request to show a specific book in a library can, for example, be given from the library enquiry system through a "show this book" button. The object's Marker ID is then looked up and the projector unit moves to the position where this marker was detected during the last scan and projects a bright spot around the searched object. Due to the speed of the steerable projector, highlighting an arbitrary object takes less than a second, thus providing almost instant feedback to the search request.

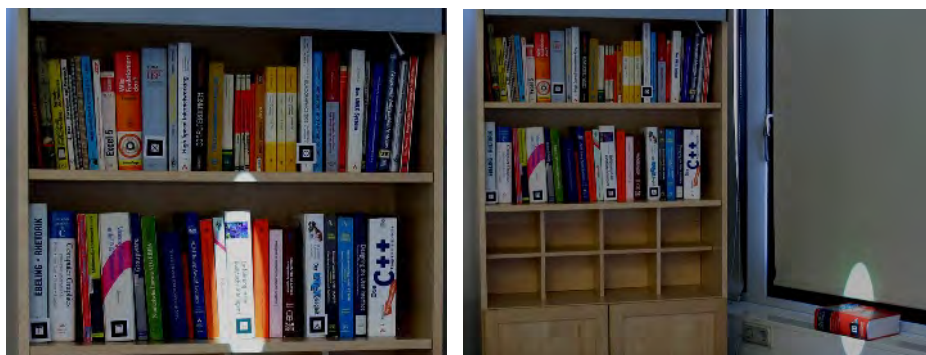


Fig. 3. A book was found on the shelf (left) and another one on the window sill (right)

4 Related Work

The use of a steerable projector to transform environments into continuous displays has first been proposed by Claudio Pinhanez [3] at IBM. In their Siggraph 2001 emerging technologies demo, where colored M&Ms were composed by visitors to form large pixel images, the Everywhere Display Projector was used to highlight trays with different colors on a shelf. The positions of these trays were, however, not acquired automatically.

Additional applications are presented in [4], including a ubiquitous product finder and an interactive shelf. The product finder guides people towards products in a store environment, using blank projection surfaces as individual direction signs and interaction spaces. The interactive shelf uses similar surfaces for the display of product-related information. While these prototypes are strongly related to the work presented here, both of them require dedicated projection space in the environment.

Raskar et al. present in [5] a method for simultaneous acquisition of room geometry and use of the acquired surfaces as an output medium for mobile projectors. They also use markers for object recognition and annotation. The FindIT Flashlight [2] can find objects instrumented with responsive electronic tags. Users scan the environment with an electronic flashlight emitting a digital code, and the corresponding tags respond when hit by the flashlight's beam. Both systems rely on the user pointing to the right direction, while in our demo, the environment actively controls the projector.

In [7] another augmented library environment is described, where books or their requested positions can be highlighted on the shelf through a head-mounted display. While this approach is more general in terms of scalability and topology of the environment, it requires an instrumentation of the user and an explicit 3D model of the environment, which we were able to avoid.

5 Current Results and Future Work

One immediate application scenario is a public library where each book has a marker on its front and back side and on the spine. In this way a book can be located even if it is not at its usual place in the shelf. When the library is closed, *Search Light* can scan all shelves and tables and take inventory. During opening hours, only certain areas must be re-scanned as described above. With our experimental setup in SUPIE (room size 5x6m, shelf on the wall, 4 Megapixel steerable camera with 3x optical zoom in the center of the ceiling) we were able to reliably recognize markers down to a size of 10mm in the whole room. Initial scanning of the room took roughly 1 hour, mostly due to the slow transmission of pictures over the USB 1.1 link to the camera. Currently we are improving *SearchLight* in various respects.

5.1 Continuous Model Update

In the current demo, scanning is done only once when *SearchLight* is started. In the future we will use idle times of the projector to systematically re-scan the room for changes. This process will also prioritize regions where changes are more likely by using additional sensors, such as RFID tags or motion detection with other cameras. On the down side, including external sensors will require the use of at least a simple 3D model of the room in order to relate locations of sensor events to pan and tilt angles of the projector. This model might, however, be acquired automatically by methods similar to those described in [6]. Upcoming versions of *SearchLight* may use the camera even when the projector is used for other tasks by just analyzing images as they appear rather than actively steering the camera to certain positions for scanning.

5.2 Extension to Other Markers

In theory, existing bar codes on many products could be used for the recognition process. In the case of books, OCR could even eliminate the need for markers altogether, since book spines and covers are designed to clearly identify books. Currently, we are working on integrating RFID tags by watching the region around the RFID antenna. When the antenna registers a change in its field, a new image is taken and the exact position of the new resp. removed object is identified from a difference image. As a side effect the exact silhouette of the object might be obtained for highlighting by the projector. To overcome the visibility problems inherent to the use of optical markers, radio tags might even fully replace them. This will, on the other hand, make a 3D model of the environment necessary, in order to relate object positions to pan and tilt angles of the projector.

5.3 Fuzzy Search Results

If the exact position or dimension of a searched object is unknown, a fuzzy search result can be visualized as a spot whose diameter reflects the amount of

uncertainty. If, in addition, we know the probabilistic distribution over possible locations, we can adjust the brightness of the spot to reflect this distribution. In the extreme case this means highlighting a whole area of the room, which would signify that the object must be "somewhere in that area".

6 Conclusions

We have discussed an approach to implementing a search functionality for physical environments by a steerable camera-projector unit, and we have shown that it is possible to implement this functionality without an explicit 3D model of the environment. One big advantage of our approach is, that no calibration between the projector-camera-unit and the environment is needed. One obvious limitation is that only objects within sight from the unit can be searched and found. This limits the topology of the environments in which our approach can be used.

Acknowledgements. The work described here was funded by "Deutsche Forschungsgemeinschaft" under a young investigator award for the project FLUIDUM. The SUPIE environment is shared with the REAL project. We thank our anonymous reviewers for helpful suggestions and pointers to additional related work.

References

1. M. Billinghurst, S. Weghorst, and T. Furness. Shared space: An augmented reality approach for computer supported collaborative work. *Virtual Reality*, 3(1):25–36, 1998.
2. H. Ma and J. A. Paradiso. The findit flashlight: Responsive tagging based on optically triggered microprocessor wakeup. In *Proceedings of UbiComp 2002, Gothenburg, Sweden*, 2002.
3. C. Pinhanez. Using a steerable projector and a camera to transform surfaces into interactive displays. In *Proc. ACM CHI 2001*, Seattle, Washington, USA, March 31 - April 5 2001. ACM Press.
4. C. Pinhanez, R. Kjeldsen, A. Levas, G. Pingali, M. Podlaseck, and N. Sukaviriya. Applications of steerable projector-camera systems. In *Proceedings of the IEEE International Workshop on Projector-Camera Systems at ICCV 2003*, Nice Acropolis, Nice, France, October 12 2003. IEEE Computer Society Press.
5. R. Raskar, J. van Baar, P. Beardsley, T. Willwacher, S. Rao, and C. Forlines. ilamps: Geometrically aware and self-configuring projectors. In *ACM SIGGRAPH 2003 Conference Proceedings*. ACM Press, 2003.
6. R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, and H. Fuchs. The office of the future: A unified approach to image-based modeling and spatially immersive displays. In *Proc. ACM SIGGRAPH '98*, pages 179–188, 1998.
7. G. Reitmayr and D. Schmalstieg. Location based applications for mobile augmented reality. In *Proceedings of the 4th Australasian User Interface Conference*, pages 65–73, Adelaide, Australia, Feb. 2003.
8. M. Weiser. The computer for the 21st century. *Scientific American*, 3(265):94–104, 1991.

Author Index

- Abowd, Gregory D. 51, 140
Atrash, Amin 18
- Bao, Ling 1
Beigl, Michael 254
Bharatula, Nagendra B. 225
Borriello, Gaetano 33, 323
Butz, Andreas 351
- Cermann, Holger 194
Chatham, Chris 51
Chen, Nicholas Y. 87
- Decker, Christian 254
Dey, Anind K. 340
Dulman, Stefan 269
- Evans, Helen 345
Evers, Leon 269
- Floerkemeier, Christian 188
- Gellersen, Hans-W. 340
Gill, Christopher 305
- Handorean, Radu 305
Hannaford, Blake 33
Hansen, Heiko 345
Harvel, Lonnie 51
Havinga, Paul 269
Hayakawa, Keisuke 238
Heiber, Timo 194
Henrici, Dirk 219
Hile, Harlan 323
Hong, Jason I. 87
- Intille, Stephen S. 1, 158
Itoh, Hideo 124
- Jiang, Xiaodong 87
Junker, Holger 18, 176
- Kashitani, Atsushi 238
Kim, Jiwon 323
Kishino, Yasue 238
Kohno, Michimune 213
- Krohn, Albert 254
Kubach, Uwe 254
- Laasonen, Kari 287
Laerhoven, Kristof Van 340
Lampe, Matthias 188
Landay, James A. 87
Larson, Kent 158
Lester, Jonathan 33
Lim, Yu-Xi 51
Liu, Ling 51
Lukowicz, Paul 18, 106, 176
- Marrón, Pedro José 194
Matthews, Tara 340
Miyaki, Takashi 213
Müller, Paul 219
Munguia Tapia, Emmanuel 158
- Nakamura, Yoshiyuki 124
Nakashima, Hideyuki 124
Nishimura, Takuichi 124
Nishio, Shojiro 238
- Ossevoort, Stijn 225
- Pirttikangas, Susanna 182
- Raento, Mika 287
Rekimoto, Jun 213
Riekk, Jukka 182
Robinson, Philip 254
Röning, Juha 182
Roman, Gruia-Catalin 305
Roussel, Nicolas 345
- Scheibe, Chris 51
Schneider, Michael 351
Siegemund, Frank 69
Spasova, Mira 351
Stäger, Mathias 18, 225
Starner, Thad 18
Suutala, Jaakko 182
- Takayama, Leila 87
Terada, Tsutomu 238
Toivonen, Hannu 287

Tröster, Gerhard 18, 106, 176, 225

Truong, Khai N. 140

Tsukamoto, Masahiko 238

Wacker, Arno 194

Waldkirch, Marc von 106

Wang, Kevin 87

Ward, Jamie A. 18

Yamamoto, Yoshinobu 124

Yoshihisa, Tomoki 238

Appendix C-5

スポットライトによる物探し支援システム

中田 豊久 金井 秀明 國藤 進

北陸先端科学技術大学院大学

あらまし 隠れていても頻繁に移動する物でも、その物にスポットライトをあてるシステムを構築した。そのアプリケーションとして「部屋の中での紛失物探し」と「本棚での関連本探し」を支援するシステムを提案する。前者のシステムでは、音による物探し支援ツールとの比較実験から視覚情報による物探しの有効性を示す。後者ではシステムを使って本棚の前で技術調査をする実行例を提示する。

A Supporting System for Finding Lost Objects by Spotlight

Toyohisa NAKADA, Hideaki KANAI and Susumu KUNIFUJI

Japan Advanced Institute of Science and Technology

Abstract We developed a system that illuminates an object even though the object is hidden by others or often moved. We propose two systems as its application: finding lost objects in indoor and finding related books in bookshelf. In former system, we indicate advantages of our system against a system using acoustic awareness. In latter system, we present an example of using our system.

1 はじめに

紛失した携帯電話やテレビのリモコンなどを部屋の中を探し回ったり、以前に見た記憶のある本を本棚から探したり、「物探し」という行為はよく行われるものである。一方、物の位置を測定する機器は、近年その性能を向上させてきているが、決定的な機器はまだない。そこで本稿では、複数の位置測定器を併用し、物の位置への気づきを向上させることで探し物を支援するシステムを提案する。

物探しという問題を本稿では、「必要としている物がどこにあるか分からない」と定義する。この問題に我々は、システムが物の位置を常に計測し続け、ユーザから要求があった時に視覚情報（スポットライト）によりその位置を提示する、というアプローチをとる。システムが物探しを支援する場合には、人に機器を付けるか、物に機器を付けるか、によって分類できる。前者の場合は、ユーザの行動履歴から

記憶を想起させ、探し物のある場所を特定するやり方となる。後者の場合は、物の位置が既に分かっているため、必要な時にその場所を提示する、という方法になる。我々は、何も特別な機器を身につける必要がない後者の方法による解決を目的としている。また、探し物の位置を提示する方法には、視覚情報の利用だけでなく、音（聴覚情報）による通知もある。しかし、我々の目的とする視覚情報による通知は、「一目瞭然」というように物の位置を伝えることに対して有効性が高い、と考える。

物探し支援システムは、次の機能要件を充たす必要があると考える。

透過性 箱や鞆の中などの隠れた位置にある物を探し出せる機能

追従性 頻繁に移動する物を探し出せる機能

本稿では、これらの機能を充たしたスポットライトによる物探し支援システムを構築し、そのアプリ

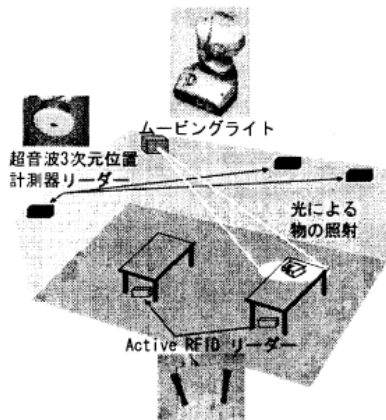


図 1: システム構成図

ケーションとして「部屋の中での紛失物探し (3 章)」と「書棚での関連本探し (4 章)」を提案する。

2 光による物探し支援システム

光による物探し支援システム [5] は、隠れた物でも頻繁に移動する物でもその位置を測定し、視覚情報によってユーザに知らせるシステムである。この章では、その構成、複数の位置計測器の統合方法、そして光の照射方法について記述する。

2.1 システム構成

図 1 に本システムのシステム構成を示す。物の場所を特定するための位置計測器として、2 種類の機器 (Active RFID: RF Code 社, 超音波位置計測器: 古河機械金属社) を併用する。それぞれの機器によって、位置情報の透過性 (隠れた状態にある物の位置を把握することが出来る機能) を実現し、測定精度の向上を図る。これらの 2 つの測定値の統合にはパーティクルフィルタを用いる。

探し物の位置への気付きを向上させるため、「スポットライト」を探し物のある付近に照射する。そのため、舞台などの照明機器として使用されているムービングライト (マーチン社) を用いる。

2.2 パーティクルフィルタによる物の位置推定

パーティクルフィルタとは、それぞれが状態と重みを持つ多数の粒子 (パーティクル) によって状態空間内全体の確率分布を近似する手法である。その特徴は、例えば位置推定に使用する場合、ある物がそこに存在する確率を検索領域のすべての箇所で計算せずに、局所的な箇所のみの計算で全体を近似できることである。Jeffrey Hightower らは、このパーティクルフィルタを物の位置推定に使用し、そのツールを公開している [2, 4]。本システムでは我々の使用する位置計測器についての部分をそのツールに付加し、使用している。パーティクルフィルタを位置推定に利用すると、精度の異なる位置情報の統合と隣接情報から位置情報への変換を統一的に処理することが可能となる。また、システムへ新たな位置計測機器を容易に導入することが出来るようになる。

本研究で使用した超音波位置計測器は、平均が 0 で 2 分散を測定誤差である 5cm とした正規分布として扱っている。Active RFID の場合は、事前実験から決定した誤差 20cm の幅と高さ (確率) 1/20 の一様分布としている。

2.3 ムービングライトによる光の照射

物の 3 次元位置情報が分かったら、そこに光が当たるようにムービングライトのパンとチルトを計算する。この計算は、パンとチルトを回転するアームと見ることにより、ロボットアームの手から腕の角度や伸縮を求める逆運動学問題 [9] と等価になる。

ムービングライトから見た座標系を T_m とし、その座標系での物の位置を P_m とする。一方、位置測定器から見た座標系を T_p とし、その座標系での物の位置を P_p とすると、物の位置の両者の関係は式 1 となる。

$$P_p = A_t A_{ry} A_{rz} A_{rp} A_{rt} P_m \quad (1)$$

各変換行列の意味は次のようである。座標系 T_p を平行移動 A_t , y 軸周りに回転 A_{ry} , z 軸周りに回転 A_{rz} するとムービングライトのパン、チルトが 0 の初期位置の座標系となる。その座標系からムービングライトのパン A_{rp} , チルト A_{rt} の変換をした座標系がムービングライト座標系 T_m となる。パーティクル

フィルタから得られた対象物の座標からムービングライトのパンとチルトを求めるには、機器の設置後のキャリブレーションとして求められた A_t, A_{ry}, A_{rz} と計測値から得られる P_m, P_p から A_{rp} と A_{rt} を求めればよい。

照射される光は、パーティクルフィルタによって計算される確率の分散に比例した大きさで照射する。色（無色）と形（円形）は変化しない。また、ムービングライトを制御するために、ムービングライトと計算機とのインターフェースには USITT DMX512¹ という標準のプロトコルを使用する。

3 部屋の中での紛失物探し

3.1 概要

部屋の中で見失った物を探すことを支援するシステムを光による物探し支援システムの1つ目のアプリケーションとして構築した。システムの実行例を図2に示す。Passive型RFIDタグを裏に貼り付けた紙に、タグを貼り付けた物の写真や、その物をイメージする絵を印刷したものを入力インターフェースとして使用している。RFIDリーダを絵の上に置くと、その物の付近をスポットライトが照射する。

3.2 音による紛失物探し支援ツールとの比較実験

表 1: 隠し場所タイプと、その時の光による物探し支援システムの位置測定精度の状態

Type	説明	位置精度状態
Type1	完全に目視できる。	高精度
Type2	一部が目視できる。	高精度
Type3	遮蔽物の下にある。遮蔽物を退かせば探し物を目視できる。	低精度
Type4	靴や箱の中。探し物を目視するまでにある程度の作業が必要。	低精度

¹ <http://www.usitt.org/standards/DMX512.html>

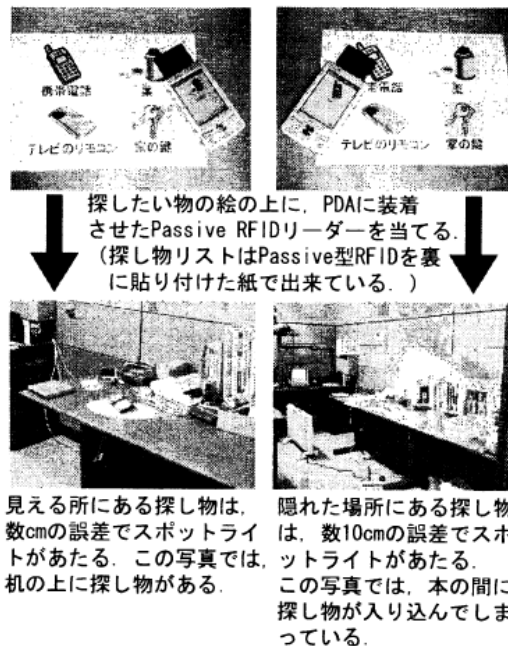


図 2: 紛失物探しシステムの実行例

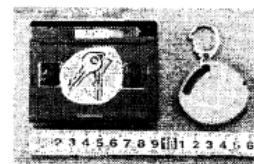


図 3: 光による探し物支援システム（左）と、音による紛失物探し支援ツール（右）のタグ

物の位置情報の光による気づきは、音によるものに比べて有効であるかを調べるために、コントローラのボタンを押すとタグから音が発生するワイヤレスロケータ（The Sharper Image Inc.）という製品との比較実験を行った。

実験は、机や椅子などがある一般的なオフィス環境の中の約 4m 四方の領域に両システムのタグを隠し、被験者に探してもらいその時間を計測する。探してもらったタグは、図3のようである。タグは、表1で示す4つのタイプの隠し場所に隠しておく。被験者は実験環境に対して予備知識の無い大学院学生10人であり、隠し場所タイプ毎に3回、計12回の測定を両システムで行った。隠し場所毎に分類し

た結果を図4に示す。図4のエラーバー（棒の先端の工の形をしたもの）は平均値の95%信頼区間を示している。結果は、Type2のみに両側1%水準で有意差があると確認された。この結果から光による物探し支援システムの優位性は、手などを使わずに視覚情報だけで探し物が出来る条件で、探し物が分かり難い所にある場合であると確認できた（図5）。

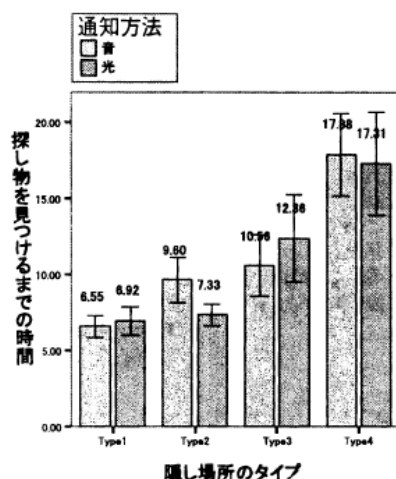


図4: 隠し場所タイプ毎の物を見つけるまでの時間(秒)

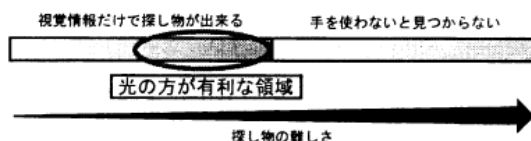


図5: 音に対しての光による物探し支援システムの優位性

4 本棚での関連本探し

4.1 概要

書店や図書館の前で、例えば何かについて調査をしている時に、いくつもの本を手にとって見る、ということをよく行うものである。この時には、Amazon.comなどで提供されている「この本を購入した人は、よくこういう本も購入しています」という情

報が役に立つ。ここでは、このような情報を光による物探し支援システムを使って、計算機を意識せずにユーザーに伝える「本棚での関連本探し支援システム」について記述する。

システムは次節以降で説明する「関連本検索」「学習による関連本検索」「部分ネットワークによるランキング表示」の3つの機能を提供する。ユーザーは、実物の本を指定位置にかざすことでシステムへ入力情報を与え、スポットライトで実物の本を照らし詳細情報をディスプレイに表示することでシステムからの出力情報を得る。

4.2 関連本検索

Amazon.comから提供されているWeb Service²を使用して、「この本を買った人はこんな本も買っています」と紹介されている³5つの本(XMLのタグは「SimilarProducts」)を関連本元データとして取得する。本棚にあるすべての本について2つ先まで関連本元データを取得し、ノードが「本」でエッジが「本の関係」を表す有向グラフを作成する（この有向グラフは以降のアルゴリズムでも使用する）。

入力データとして1つの本をシステムに与えると、それに関連する本を1つ出力する。ここでは関連本の事を、予め作られた関連本元データの有向グラフ上で最も少ないリンク数で結ばれる2つの本である、と定義する。ディスプレイに関連本とそこまでのリンク情報が表示され、実物の関連本にはスポットライトがあたる。

4.3 学習による関連本検索

学習による関連本検索は、2冊の本から関係を学習し、その関係を使って関連本を検索することである。この学習は次のようにして行われる。

STEP1 関連本元データの有向グラフを表す隣接行列を固有値分解し、最大と第二最大固有値を持つ固有ベクトルを使い、全ての本を平面配置する。

STEP2 2つの本の位置を示すベクトルの減算で得るベクトルをその2冊の関係とする。

² <http://www.amazon.com/gp/aws/sdk/>

³ 2005年2月現在の情報

STEP3 与えられた本にSTEP2のベクトルを足し、仮想的な関連本を作る。

STEP4 仮想的な関連本に最も近い本を推薦する。

出力は、ディスプレイに仮想的な関連本に類似する順番で本が表示される。その中の1冊を選択すると、実物にスポットライトが当たる。

4.4 部分ネットワークによるランキング表示

指定された本の中での重要度は、関連本元データの有向グラフを表現する隣接行列の最大固有値を持つ固有ベクトルの値の大きさに決定する。この方法は、Googleでの検索結果をランキングするアルゴリズム (PageRank [6]) として知られている。

指定された本から2リンク先までの関係ネットワークにPageRankを適用することで、特定領域内でのランキングを推定することが出来る。例えば、Amazon.comには売り上げランキングが表示されているのであるが、これはすべての本とすべてのユーザによる指標で、それがそのまま自分にとって重要な本であることを示しているわけではない。一方、自分で本を指定してPageRankを適用することは、それらの本とその周辺ではどの本が重要であるかが分かり、前者の方法よりは自分に適した重要度を教えてくれると思われる。

システムへの入力、指定位置にランキングしたい本を1つづつかざし、最後にディスプレイ上のランキングボタンをクリックする。ランキングがディスプレイ上に現れ、その中の1冊を選択すると実物にスポットライトが当たる。

4.5 実行例：ネットワーク分析関連の本を探す

図6に本棚での関連本探し支援システムを使用した例を示す。閲覧のスタートは、「Social Network Analysis: Methods and Applications (Structural Analysis in the Social Sciences, 8)(ISBN0521387078)」からで、「実践ネットワーク分析—関係を解く理論と技法 (ISBN4788507811)」を関連本検索で推薦される。学習による関連本検索機能に「自然科学の統計学 (ISBN4130420674)」, 「統

計学入門 (ISBN4130420658)」と与え「実践ネットワーク分析—関係を解く理論と技法」の入門編にあたる「ネットワーク分析—何が行為を決定するか (ISBN4788505843)」を見つける。最後にすべての本をランキング表示し、「ネットワーク分析—何が行為を決定するか」が最も重要であるという結果を得る。

5 関連研究

透過性と追従性を持った「聴覚情報」による物探し支援は、[10]で提案されている。視覚情報を使った物探し支援システムとしては、店舗内での商品探しを可動式プロジェクタを使って支援するシステムがある[7]。しかしこのシステムでは物の位置情報を静的に扱っているため追従性の機能を持っていない。

計算機上に仮想図書館を作り関連する図書などを推薦する研究は、例えば[3]などがある。一方、我々のシステムと同様に実世界志向の支援システムとしては、図書館でヘッドマウントディスプレイを使って現実を拡張するシステムが提案されている[8]。我々のシステムとの相違点は、我々のシステムは何も機器を身に着ける必要が無いことである。図書館で本探しを支援する[1]のシステムは、現実世界の活性化を目的とし何も身に着ける必要がないのであるが、システムが透過性と追従性の機能を持っていない。

6 おわりに

視覚情報を用いた「透過性」「追従性」を持つ光による物探し支援システムを提案し、そのアプリケーションとして「部屋の中での紛失物探し」と「本棚での関連本探し」の2つを構築した。今後は、さらに改良し、評価をしていきたいと考える。

謝辞 本研究の一部は文部科学省知的クラスター創成事業石川ハイテク・センシング・クラスターにおける「アウェアホーム実現のためのアウェア技術の開発研究」プロジェクトの一環として行われたものである。

参考文献

- [1] A. Butz, M. Schneider, M. Spassova, "Search-Light - A Lightweight Search Function for Per-



図 6: 本棚での関連本探し支援システムの実行例

- vasive Environments”, Pervasive Computing 2004, pp. 351-356, 2004
- [2] Jeffrey Hightower, Barry Brumitt, Gaetano Borriello, “The Location Stack: A Layered Model for Location in Ubiquitous Computing”, 4th IEEE Workshop on Mobile Computing Systems & Applications (WMCSA 2002), pp. 22-28, 2002
- [3] Saranya Maneeroj, Hideaki Kanai, Katsuya Hakozaiki, “An Improved Recommendation Method for Better Filtering Information out of Database”, 情報処理学会論文誌 データベース, Vol. 43, No. SIG5(TOD14), pp.66-74, 2002
- [4] The Location Stack, <http://portolano.cs.washington.edu/projects/location/>
- [5] 中田豊久, 金井秀明, 國藤進, “光による物探し支援システム”, 情報処理学会 インタラクシオン 2005 インタラクティブ発表, 2005
- [6] Lawrence Page, Sergey Brin, Rajeev Motwani and Terry Winograd, “The PageRank Citation Ranking: Bringing Order to the Web”, Stanford Digital Library Technologies Project, 1998
- [7] C. Pinhanez, R. Kjeldsen, A. Levas, G. Pingali, M. Podlaseck, N. Sukaviriya, “Applications of steerable projector-camera systems”, IEEE International Workshop on Projector-Camera Systems at ICCV 2003, 2003
- [8] G. Reitmayr, D. Schmalstieg, “Location based applications for mobile augmented reality”, 4th Australasian User Interface Conference, pp.65-73, 2003
- [9] 吉川恒夫, “ロボット制御基礎論”, 2.4 逆運動学問題, コロナ社, ISBN4-339-04130-1, 1988
- [10] 新西 誠人, 伊賀 聡一郎, 樋口 文人, 安村 通見, “Hide and Seek: アクティブに反応する ID タグの提案”, 第7回インタラクティブシステムとソフトウェアに関するワークショップ (WISS99)

Appendix C-6

An Intelligent Disassembly Assistant for Man-Machine Demanufacturing

Michael Bailey-Van Kuren, *Member, IEEE* and Justin Soltani

Abstract—In order to handle the complexity of demanufacturing operations, a combined man-machine system is proposed that incorporates projector-vision technology. As part of this system, an intelligent disassembly assistant has been developed to provide a flexible and efficient method for facilitating demanufacturing. The assistant is comprised of a set of software modules that are processed in a distributed computing environment. Hardware and software components of the intelligent assistant are identified. The assistant determines the base elements of the structure, checks for stability, and then determines a disassembly sequence. The projector-vision system provides information to a human operator and detects operations performed on the product. The system can adapt to unexpected human operator behavior.

I. INTRODUCTION

THE dismantlers of end of life electronic equipment process a wide variety of product models even for a single product type. Each product may contain unique components to be removed either due to hazardous material or reuse. Jovane et. al. noted a trend towards increased product take back and subsequent automation, with the emphasis on automated disassembly and separation for product recycling and reuse [1].

Currently, most dismantlers of electronic waste in the United States rely upon manual labor to disassemble used products. Operators must recognize the model under consideration and determine the appropriate level of processing. These operator decisions are based on varying product conditions, dynamic scrap lists, and dynamic material costs. However, operator information is often provided by static printouts and visual displays. Projector-vision systems such as the system investigated by Pinhanetz et. al. [2] have been proven as an effective interface between humans and computer systems. In demanufacturing, a projector-vision system can provide real time interface between the dynamic information in the dismantler databases and the disassembly operator.

Bailey-Van Kuren [3] investigated a demanufacturing approach that processes used products with a combination of

a human operator and the application of robotic machining methods. A combined man-machine demanufacturing system provides the advantages of automated methods on known tasks and human flexibility to deal with complex tasks.

Schmult [4] developed a system for automated disassembly of block structures for a fully automated system. This system addresses the general problem of acquiring desirable or undesirable material from a product, or segregation of dissimilar materials. This same problem must be addressed by the combined man-machine demanufacturing system.

Therefore, an intelligent system for product identification and disassembly assistance is proposed. Using projector-vision technology, product models can be identified on-line and relevant product information can be actively displayed at the work station. Projector-vision systems are being studied in a variety of fields as a flexible method for user-system interaction. For dismantling operations, a high speed inspection camera monitors the work station while human operators process the used electronic equipment and identifies models from a product database. The product database monitors scrap list information, material component cost information and refurbishment requirements. Thus, dynamic data, such as a disassembly sequence or a scrap list change, is broadcast to the operator when necessary. System efficiency is improved due to proper processing of the specified product model.

The distributed computing hardware implemented in the demanufacturing work cell is presented in section II. Section III presents the software components. Conclusions and future work are discussed in section IV.

II. DEMANUFACTURING WORK CELL HARDWARE

The hardware elements integrated in the intelligent demanufacturing work cell assistant include a robot controller, two autonomous machine vision systems, a projection system, tooling, and a supervisory controller, see Fig. 1. The robot controller, vision systems, and the supervisory controller contain processors thus establishing a distributed computing platform that can concurrently process independent software modules.

Manuscript received March 1, 2005.

M. Bailey-Van Kuren is with the Department of Manufacturing and Mechanical Engineering, Miami University, Oxford, OH 45056 USA (phone: 513-529-2647; fax: 513-529-1454; e-mail: baileym@muohio.edu).

J. Soltani is a student researcher with the Department of Manufacturing and Mechanical Engineering, Miami University, Oxford, OH 45056 USA. (e-mail: soltanj@muohio.edu).

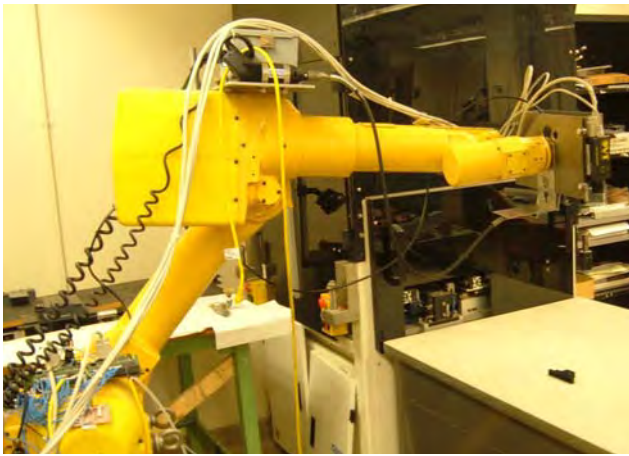


Fig 1. Automated disassembly work cell with projector-vision analyzing a block structure.

A. Robot / Robot Controller

The work cell utilizes a Fanuc S-10 robot which is a six-axis articulated arm robot powered by brushless AC servomotors. The S-10 has a maximum work load at the wrist of 10kg. In the demanufacturing application, the robot hand carries a lightweight projector system, two machine vision cameras, a high speed cutting tool, and a vacuum gripper. Thus, the robot controls the position and orientation of the projector-vision system as well as tools and manipulators for interacting with the used product. The S-10 has a large spherical workspace which enables the system to interact with the product on two separate work table surfaces. One work table is dedicated to automated processing and the second work table supports manual disassembly and projector-vision interaction with the human operator. The positioning repeatability of the S-10 robot is $\pm 0.2\text{mm}$. In general, a combined manual and automated demanufacturing station does not require high accuracy since the automation processes the used product with flexible cutting operations and the human operator is relied upon for accurate tasks. The S-10 Karel Controller communicates with the other system components using Ethernet based RS-232 communication or fixed I/O with 32 digital inputs and 24 digital outputs. The Ethernet based communications facilitate communication between the robot controller, the supervisory controller, and the vision systems.

B. Projection - Machine Vision System

In order to obtain information on the structure presented for disassembly and provide feedback to a human disassembly operator, the automated work cell utilizes a projector-vision system. The projection-vision system utilizes a Dell 3200MP projector and two DVT Series 600 Smart Image Sensors. The projector and cameras are mounted on the hand of the robot as shown in Fig. 2.



Fig. 2 Projection-vision system mounted to hand of the robot.

The Dell 3200Mp projector was selected for low weight and high brightness. The projector weight reduces the remaining effective capacity of the robot since it is mounted to the hand. The Dell projector weighs 1.6 kg and provides a brightness of 1300 lumens. The projector also provides digital zoom and keystone correction which can be adjusted from the supervisory control computer.

The image sensor is 4.8mm x 3.6mm CCD with a 640 x 480 pixel resolution, and 7.4um x 7.4 um pixels. The overlap in the cameras field of view provides depth perception for the work cell automation. The Series 600 has a programmable range of 10 μs – 1s exposure times, allowing a flexible platform to account for the varying amount of light intensity needed for different automated tasks. The cameras communicate with built in external ports for RJ-45 (10/100 megabit Ethernet communications, TCP/IP protocol) communications. A digital I/O board also accompanies each camera to communicate data. Each board has 12 configurable inputs and outputs that are 24Volt DC regulated.

The cameras facilitate the distributed computing environment with built in processors. Each camera is programmed with a logic structure that reacts to various camera inputs and generates various camera outputs. In this manner, the cameras can dynamically switch between different operating modes to identify product features or determine operator actions in the projector-vision interface. Camera outputs are directed toward the robot controller or the work cell supervisory controller.

C. Supervisory Controller

A desktop personal computer is used to communicate with the robot controller and vision systems, direct the projector – vision system and run high level logic. High speed computing is not required for supervisory control in this

distributed environment. The work cell supervisory controller that is currently employed is a Pentium II processor running at 448 Mhz. This processor speed is well below current available processing speeds in the gigahertz range. The supervisory controller communicates with the vision systems and robot controller via Ethernet based serial communication. Furthermore, the supervisory controller transforms the image coordinate data from the cameras to work space coordinates for the robot, providing surface data points of the used product surfaces. In the demonstration work cell, the supervisory controller maintains the product databases required for projecting real time data to the work cell operator. Figure 3 presents a schematic of the computing hardware set up for the work cell.

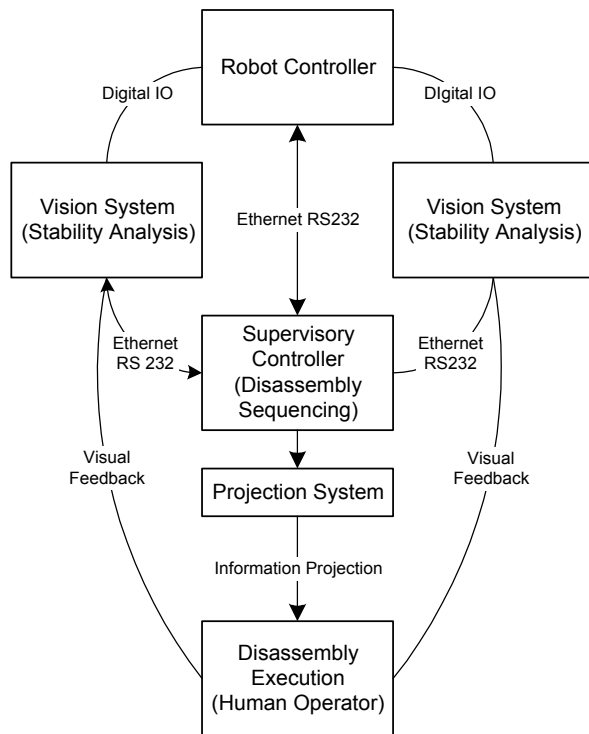


Fig.3. Relationship of disassembly assistant within the demanufacturing system

III. DEMANUFACTURING WORK CELL SOFTWARE

A. Logic Overview

The demanufacturing system relies upon three distributed software modules based at the robot controller, the machine vision systems, and the supervisory controller. The robot controller based module controls robot disassembly motion, reports robot positions, and interacts with the used product through peripheral software/hardware. The vision system based module provides image information regarding the phone surfaces and feature recognition requested by the robot controller. The supervisory controller coordinates the robot and machine vision information to coordinate and redirect robot motion based on the visual feedback from the used product. The tasks required to dismantle a used product are interrelated, therefore the programs on the vision system,

robot, and supervisory controller must run simultaneously.

The disassembly assistant software is triggered when the machine vision systems identify a used product on the work table. Schmult [4] implemented stability analysis in the automated system to prevent a disassembly operation that leaves the product unstable. A similar stability analysis is performed in the combined man-machine system. Once stable configurations are determined, a disassembly sequence is generated. Figure 3 shows the relationship of the disassembly assistant modules within the computing hardware structure. Note that "Execution" is carried out by the operator and visual feedback verifies task execution. Operator choices can change task planning and overall planning in the system.

B. Stability Analysis

Stability analysis is performed in two steps: identification of product feet and determination of stability. In order to maintain the structural stability of the product on the work surface, components in contact with the work surface that are supporting the structure or "feet" must be identified. Any product can be segregated into a body and limbs that protrude from a product body. In the simplest case, the product body is resting on the work surface and the identification of a "foot" is trivial. In order for stability analysis to be pertinent, a product must have the ability to rest in a variety of orientations based on some subset of protrusions contacting the work surface. Machine vision based segregation of the product body and limbs can be obtained through standard blob processing to segregate blobs. The identification of protrusions that are in contact with the work surface can be obtained by comparing blob location parameters within the image to the known work cell orientation of the camera relative to the work surface. Figure 4 presents the identification logic for product feet.

Each camera surveys objects in its field of view to identify feet. The supervisory controller resolves redundant feet by comparing transformed component coordinates. The result of this module is a set of feet components. After the feet of the structure are identified, the product structure undergoes a stability analysis to prevent product tipping during disassembly tasks.

The machine vision systems compute the visual center of mass of the structure in the image plane as well as the center of mass of each body or limb in the structure. In block structures, these individual center of mass coordinates are compared to the maximum X and Y extents of the feet structure to determine stability [4]. The block with its center of mass farthest from the feet is chosen for removal because it results in the most stable structure. This method assumes uniform density throughout the structure. This assumption would not be valid for various used products such as computers with dense power supplies and batteries. However, one principle product for the demanufacturing cell is used mobile phones after battery removal. Although component density varies for mobile phones, this variation is small.

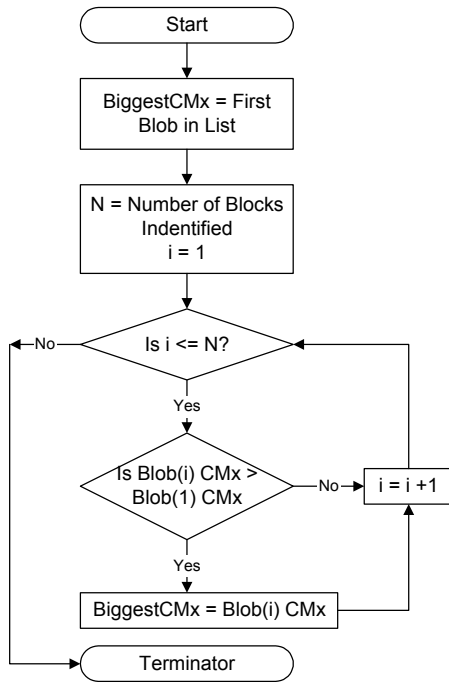


Fig. 4. Logical Structure of Feet Identification

The stability analysis was tested on basic block structures in order to demonstrate the identification scheme. Figure 5 shows a basic block structure with feet identified as well as the target, “most stable”, component. The target component is identified to a human operator via projected light on the target component. This stability analysis provides the basis for the disassembly of the product structure.

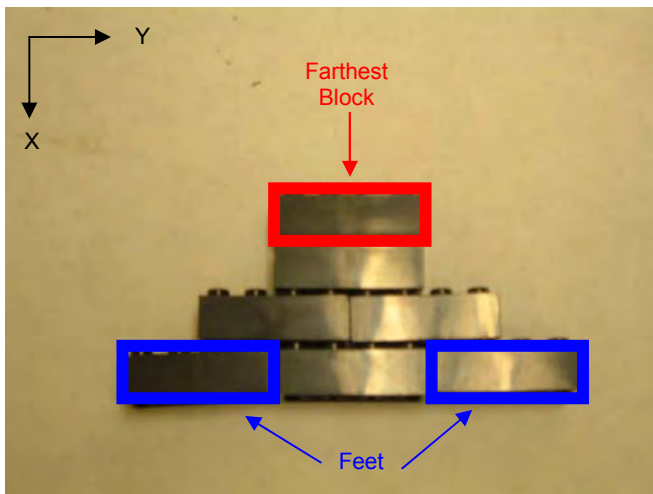


Fig. 5. Block structure with feet and target block as identified by the intelligent agent.

C. Real Time Disassembly Sequencing

The disassembly sequencing is not performed by a priori planning. In the previous work [4] that focused on

automated disassembly, the stability analysis can determine the entire disassembly sequence prior to disassembly because it is assumed that the robot will remove items in the appropriate sequence. In the case of human directed disassembly, the human operator may not perform the directed task. Therefore, a new analysis must be performed for the new state of the product.

The disassembly sequence is obtained one part at a time using the process defined in Fig. 6. The stability analysis is directly identifying the next part for removal. Each step in the disassembly sequence is presented to the operator via the projector. The projector visually “marks” on the structure which block is to be removed. The operator can then either correctly remove the piece identified or incorrectly remove a different block that was not specified by system to be removed at that time.

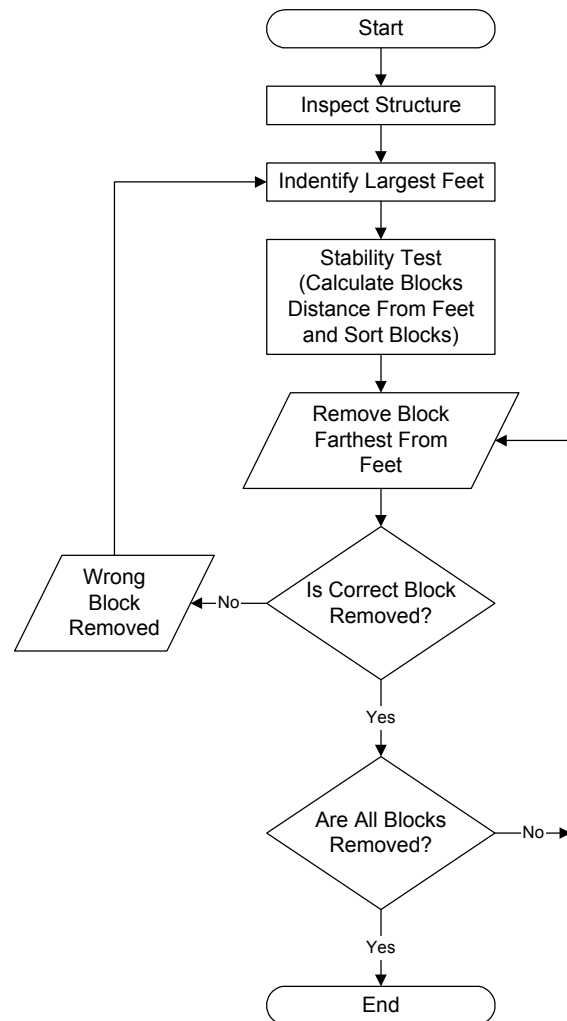


Fig. 6. Disassembly Sequence Logic

If the correct block was removed, the system does not need to recalculate the stability criteria and continues with the subsequent component identified in the disassembly sequence. However, if the user incorrectly removes a block

from the structure, the system visually detects the error and recalculates the feet identification and stability based on the new product structure. If there are no stability traps or other limiting factors for further disassembly in the structure, the system dynamically determines the disassembly sequence and provides the operator with a new component for disassembly. The inspection and verification process is iterative until the entire structure has been disassembled.

If an incorrectly block is removed and stability poses to be a limiting factor, then a visual message is presented to the operator via the projection system informing the operator of the mistake. This message only occurs when the previous piece removed must be put back for further disassembly to occur while maintaining stability of the structure. In this manner, the system accounts for the goal interaction problem between stability and disassembly.

IV. CONCLUSIONS AND FUTURE WORK

A disassembly agent has been proposed for combined automated and manual demanufacturing. The agent supports a distributed computing environment with camera based intelligence for determining structural stability. The system can react to operator changes in the disassembly sequence by dynamically determining the demanufacturing sequence. The existence of a projector-vision system facilitates the dynamic processing of the product through the capability of presenting dynamic information to the human operator.

Issues related to component density have been raised and will be a source for further study. Testing of the developed agent upon block structures and products with increasing complexity from that of block structures will be performed.

REFERENCES

- [1] F. Jovane et al., Key Issue in Product Life Cycle: Disassembly, CIRP Annals 42 (2) (1993) 651-658.
- [2] M. Bailey-Van Kuren, Automated Demanufacturing Studies in Detecting and Destroying Threaded Connections in Processing Electronic Waste, 2002 IEEE Symposium on Electronics and the Environment (2002) 295-298.
- [3] C. Pinhanez, R. Kjeldsen, A. Levas, G. Pingali, M. Podlaseck, N. Sukaviriya, Applications of Steerable Projector-Camera Systems, IEEE International Workshop on Projector-Camera Systems (2003).
- [4] B. Schmult, Autonomous Robotic Disassembly in the Blocks World, International Journal of Robotics Research 11 (5) (1992) 437-458.

Appendix D-1

FT MEDE
Gencoll

MLCM
2006/
41169
Copy 1

LIBRARY OF CONGRESS



0 013 263 676 A

Proceedings of the
ACM SIGMM 2004
Workshop on
Effective Telepresence
ETP 2004

October 15, 2004 • New York, New York, USA



Sponsored by
SIGMM, SIGCOMM, & SIGGRAPH

ETP'04: ACM SIGMM 2004 Workshop on Effective Telepresence • October 15 • New York, New York, USA

Proceedings of the
ACM SIGMM 2004
Workshop on
Effective Telepresence

ETP 2004

October 15, 2004 • New York, New York, USA



***Toward Seamless Remote
Interaction & Experience***

**Sponsored by
SIGMM, SIGCOMM, & SIGGRAPH**

Amazon Ex. 1008.0239

**The Association for Computing Machinery
1515 Broadway
New York, New York 10036**

Copyright © 2004 by the Association for Computing Machinery, Inc. (ACM). Permission to make digital or hard copies of portions of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyright for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permission to republish from: Publications Dept., ACM, Inc. Fax +1 (212) 869-0481 or <permissions@acm.org>.

For other copying of articles that carry a code at the bottom of the first or last page, copying is permitted provided that the per-copy fee indicated in the code is paid through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

Notice to Past Authors of ACM-Published Articles

ACM intends to create a complete electronic archive of all articles and/or other material previously published by ACM. If you have written a work that has been previously published by ACM in any journal or conference proceedings prior to 1978, or any SIG Newsletter at any time, and you do NOT want this work to appear in the ACM Digital Library, please inform permissions@acm.org, stating the title of the work, the author(s), and where and when published.

ISBN: 1-58113-933-0

Additional copies may be ordered prepaid from:

ACM Order Department
PO Box 11405
New York, NY 10286-1405

Phone: 1-800-342-6626
(US and Canada)
+1-212-626-0500
(all other countries)
Fax: +1-212-944-1318
E-mail: acmhelp@acm.org



ACM Order Number 105049
Printed in the USA



Foreword

It gives us great pleasure to welcome you to the *2nd ACM International Workshop on Effective Telepresence*. This workshop builds upon the ACM SIGMM 2003 Workshop on Experiential Telepresence (ETP03) held in conjunction with ACM Multimedia last year. Last year's workshop and panel involved a lively debate on what the notion of an experience implies. One of the conclusions was that we are in need of models for knowledge spaces that facilitate new forms of creativity, knowledge exploration and social relationships.

We would like this workshop to facilitate a more focused debate, namely about the means of experience in the context of seamless remote interaction. While sophisticated systems for distributed collaboration and immersive presence have emerged from time to time, their acceptance has been limited. The most commonly cited problems with such systems are that they are too difficult to set up and use, and that their value is not commensurate with cost.

The challenge, then, is to develop effective telepresence systems that offer appropriate interfaces and tools to enable seamless interaction with remote people and events. This has been recognized as one of the "grand challenges" for the multimedia community. The design and measures of performance of such systems should be centred on end-user experience. The goal of this workshop is to foster a community of researchers advancing the various technical and socio-technical aspects of telepresence. This workshop aims to bring together a good mix of people from academia, industry, and government, and spanning various disciplines including the sciences, engineering, and the arts.

An event of this kind is only as good as the quality of the people it can bring together. We have been fortunate to have a diverse program committee of research leaders of the highest expertise and reputation, who showed enthusiasm and support for this workshop. Our program this year includes six papers for oral presentation and nine papers for demo presentation. Together, they should provide an excellent snapshot of the state of research today in telepresence and we hope you like this mix of presentations. We also have three special sessions – two keynote talks and a panel session — based on invitations to some world-renowned people. The keynote talks by Jaron Lanier and Duffie White, and the panel on "Directions and frameworks for effective telepresence" should go a long way in stimulating and guiding future research in this area. This year, we invited workshop participants to post questions and discussions to the workshop website starting several weeks before the workshop. We hope participants find this leading to a more interesting debate in the panel session.

Producing a workshop requires the effort of many individuals. We wish to thank the authors for their hard work and for patiently meeting the demands placed on them for ACM-certified proceedings, and for the many demo participants for making the effort to produce live demos. We are grateful to the program committee members and additional anonymous reviewers for diligently reviewing the papers, and for their support and suggestions. Our special gratitude to the ACM Multimedia conference committee for accepting this workshop and helping in making it a reality, especially the following people: Nevenka Dimitrova and Henning Schulzrinne (General Co-Chairs), Chitra Dorai and Tat-Seng Chua (Workshops Chairs), and Lisa M. Tolles (Sheridan Printing).

We hope you enjoy *ETP 2004* and help in having similar stimulating and interactive workshops over the coming years!

Gopal Pingali, Frank Nack, & Hari Sundaram

Table of Contents

ETP '04 Symposium Organization	vii
---	------------

Oral Presentation Session 1

• A Framework for Understanding How Materials Science and Moore's Law are Changing the Landscape of Virtual Reality, Telepresence, and Tele-immersion Research	1
J. Lanier (<i>Advanced</i>)	
• Experiences of Using Wearable Computers for Ambient Telepresence and Remote Interaction.....	2
M. Drugge, M. Nilsson, R. Parviainen, P. Parnes (<i>Luleå University</i>)	
• Use of Telepresence in Informal Broadcasting over the Internet	12
N. Takamizawa, Y. Murayama (<i>Iwate Prefectural University</i>)	
• Teaching With an Intelligent Electronic Chalkboard (Demo)	16
G. Friedland, L. Knipping, R. Rojas, E. Tapia (<i>Freie Universität Berlin</i>)	

Oral Presentation Session 2

• Audio-based Head Motion Synthesis for Avatar-based Telepresence Systems	24
Z. Deng, C. Busso, S. Narayanan, U. Neumann (<i>University of Southern California at Los Angeles</i>)	
• Visual Contextualization and Activity Monitoring for Networked Telepresence.....	31
D. A. Fidaleo, R. E. Schumacher, M. M. Trivedi (<i>University of California at San Diego</i>)	
• FeelLight: A Communication Device for Distant Nonverbal Exchange.....	40
K. Suzuki, S. Hashimoto (<i>Waseda University</i>)	
• Telepresence in the Real World	45
D. White (<i>Duffie White and Company</i>)	

Demo Presentations

• 3D Visual Component Based Approach for Effective Telepresence Systems	46
Y. Okada (<i>Kyushu University</i>)	
• VITA: Visual Interaction Tool for Archaeology.....	48
H. Benko, E. W. Ishak, S. Feiner (<i>Columbia University</i>)	
• Interactive 3D Teleconferencing with User-adaptive Views	50
R. Yang (<i>University of Kentucky</i>), A. Nashel, H. Towles (<i>University of North Carolina at Chapel Hill</i>)	
• Heterogeneous Media Events Processing Systems.....	52
P. Kim, R. Jain (<i>Georgia Institute of Technology</i>)	
• Location-based Communication Services	55
X. Wu, R. Shacham, M. J. Mintz-Habib, K. Singh, H. Schulzrinne (<i>Columbia University</i>)	
• Asymmetric Collaboration through Tele-presence.....	57
A. Morde, C. Correa, J. Hou, S. K. Ganapathy, A. Krebs, I. Marsic, M. Bouzit, L. Rabiner (<i>Rutgers University</i>)	
• S3-R1: The IBM Smart Surveillance-Release-Release 1	59
A. Hampapur, L. Brown, J. Connell, N. Haas, M. Lu, H. Merkl, S. Pankanti, A. Senior, C.-F. Shu, Y. Tian (<i>IBM T.J. Watson Research Center</i>)	
• Projector-camera Systems for Telepresence	63
C. Pinhanez, G. Pingali (<i>IBM T.J. Watson Research Center</i>)	
• Visual End User Configuration of Hybrid User Interfaces	67
C. Sandor, B. Bell (<i>Columbia University</i>), A. Olwal (<i>The Royal Institute of Technology</i>), S. Temiyabutr, S. Feiner (<i>Columbia University</i>)	

Panel Session

- **Directions and Frameworks for Effective Telepresence** 69
S. Feiner (*Columbia University*), S. K. Ganapathy (*Rutgers University*), J. Lanier (*Advanced*),
G. Levin (*Carnegie Mellon University*), D. White (*Duffie White and Company*),
G. Pingali (*IBM T.J. Watson Research Center*)

Author Index..... 73

Projector-camera Systems for Telepresence

Claudio Pinhanez and Gopal Pingali
IBM T. J. Watson Research Center
19 Skyline Drive
Hawthorne, NY 10532
{pinhanez,gpingali}@us.ibm.com

ABSTRACT

Projector-camera systems provide tools for telepresence that enable novel experiences not possible with currently available schemes for remote interaction including video conferencing systems, virtual reality, and augmented reality systems. Projector-camera systems have the potential not only to allow remote individuals to view each other's environments but also to communicate in new ways by overlaying and embedding information and visualizations on *real* objects and surfaces in the remote environment, without requiring people in the remote environment to "wire" the space or wear special devices. We demonstrate some of this potential with an initial implementation.

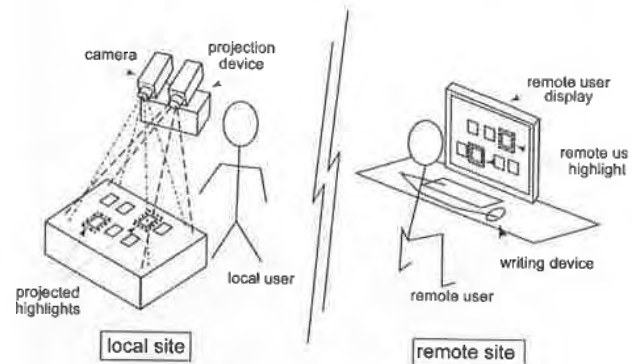


Figure 1. Overview of using a projector-camera system in a telepresence application.

Categories and Subject Descriptors

H.5.1 [Information Systems and Presentation]: Multimedia Information Systems – Artificial, augmented, and virtual realities; H.5.2: User Interfaces; H.5.3: Group and Organization Interfaces; H.4.3 [Information Systems Applications]: Communications Applications – Computer conferencing, teleconferencing, and videoconferencing.

General Terms

Algorithms, Design, Experimentation, Human Factors.

Keywords

Collaboration, remote interaction, interfaces, meetings, context modeling, action capture, presentation systems, ubiquitous computing, pervasive systems.

1. INTRODUCTION

Research in multimedia and telepresence is working towards the grand challenge of making “interactions with remote people and environments nearly the same as interactions with local people and environments” [10]. There have been many years of research in this area spanning collaborative virtual environments [4], video conferencing systems [11], and augmented reality systems [1]. For the most part, these systems have been centered on special devices – needing a special

computer or special conferencing system at either end or requiring people to sit in front of a special display or wear special devices to interact with remote people. With recent advances, it is increasingly becoming possible to look at controlling and manipulating an entire space rather than one special device in the space. This is leading to “smart” or “active” spaces and “room operating systems” [2] that treat the whole environment with its diverse set of computing, input and output devices, and sensors and actuators, as one collection of resources to be made available to users in a unified manner, much like a conventional operating system treats the resources on a desktop computer. This evolution of computing towards smart spaces is opening up new vistas for telepresence. It is now becoming possible to produce a different kind of immersive experience, remote interaction, and collaboration by augmenting the space as a whole [8]. This whole move towards smart spaces is also significant in the context of experiential computing as envisioned in [5].

In this context, we see a significant role for camera-projector systems [3]. These systems can add a new dimension to telepresence that has hitherto been uncommon. Projection and lighting can be used as an element of telepresence that allows overlaying and embedding information and visualizations on real objects, besides producing informative displays wherever they are needed. The Everywhere Display [9] and its dynamically reconfigurable interactive capabilities [6] is a good example of a projector-camera system that has significantly augmented spaces in this manner. When combined with people tracking and environment modeling, the Everywhere Display can be extended to produce steerable interfaces [7] that appear anywhere in a space and dynamically adapt to user activities in an environment and changes in the environment. While previous work on the camera-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ETP '04, October 15, 2004, New York, New York, USA.

Copyright 2004 ACM 1-58113-933-0/04/0010...\$5.00.



Figure 2. Example of control interface for a remote session chair using a projector-camera system to communicate with local conference participants.

projector systems and steerable interfaces has been primarily focused on local augmentation, we begin to explore its potential in telepresence or remote augmentation here.

Figure 1 shows an overview of our proposal for using camera-projector systems in telepresence. Here, an environment is equipped with steerable camera(s) and steerable projector(s). A user in a remote space can access/view and control both the projector and the camera through a display and writing device. A user at a remote site can view the local environment through video from the remote camera. The remote user can also control the projector in the local environment to project highlights, video, text, etc. in the local environment. As the projector is steerable, these virtual objects can be moved on to real objects and people in the local environment by the remote user. The remote user's interface can itself be an interactive steerable display although this is not a requirement. In this paper, we describe an initial demonstration we have put together to further illustrate this vision.

2. DEMONSTRATION OVERVIEW

This demonstration uses a steerable projector-camera system – the Everywhere Display – in a local environment and provides a control interface to a user in a remote environment. For illustrative purposes, we demonstrate this system in the context of a



Figure 3. The session chair projects live video showing his face on to the wall next to the speaker's podium in the remote environment.

conference/workshop taking place in a local environment for which a session chair of the workshop is in a remote location. The chair can view the workshop live from a remote location through on-site cameras and can provide instructions and messages to the local workshop attendees. We assume that the chair also has two-way audio communication to the conference.

Figure 2 shows a screenshot of the control interface available to the session chair. The upper window in this figure shows the interface to camera views. This window displays video from (one of) the remote camera(s) and enables the chair to manipulate the remote camera(s) through the pan, tilt, and zoom controls and also define camera views. In the example in Figure 2, pre-defined views include audience, podium, entrance, screen etc. The lower window in Figure 2 shows the interface for the chair to control the remote Everywhere Display projector. As discussed in [6], this projector has the notion of a "projection surface". Calibration of a projection surface ensures that projections can appear on these surfaces without keystone distortion as discussed in [9]. Different projection surfaces can be defined and calibrated for the environment in which the ED-projector is situated. The mid window in Figure 2 (labeled "Projection Surfaces") shows how these pre-defined projection surfaces are made available to the remote user, the session chair in this case. Several available projection surfaces are shown in Figure 2 such as podium, wall, screen, floor etc. In addition to pre-calibrated surfaces, there is also a "direct mapping" option which automatically associates the current camera view to an overlapping projection, as will be discussed shortly.

There are two basic ways determine what is going to be projected on the selected projection surface: grab and draw, corresponding to the two buttons on the upper right of Figure 2. The "grab" button on the upper right in Figure 2 allows the chair to grab any portion of his computer desktop and project it on the remote environment. Any computer application on the grab area is then projected onto the selected surface of the local environment. For example, the chair may choose to project his own image (as shown in Figure 3), a webpage running on a browser, etc.

Pressing the "draw" button on the upper right of the control interface shown in Figure 2 enables an embedded drawing application that allows the chair to easily "paint" the projected surface. Any drawing appearing on the lower window is immediately and efficiently transferred to the remote projector. The blank area in the lower window of Figure 2 is therefore a drawing canvas. Moreover, the real effect on the local



Figure 4. The remote session chair indicates to the speaker how much time is left for her talk.

environment can now be monitored through the live video in the upper window.

The icons on the left of the lower window show predefined drawing icons such as a highlight (which the session chair can use to highlight an object in the remote environment), a question mark (which the session chair can use to indicate it is time for questions), an arrow mark (to provide directions or point to objects in the remote environment), and a stop-watch icon (that can be used to show a speaker how much time is left for their talk). Besides these icons, the chair has a regular paint/drawing interface to draw on the canvas. The icons on the lower right allow the session chair to resize any element of the drawing canvas, and to switch between "live drawing" (the "play" button) where any modification on the canvas is immediately projected on the conference environment; and "hidden drawing" (the "pause" button), where the chair can draw, size, and compose all elements before displaying them.

The following three figures and descriptions provide some examples of how this interface is used by the session chair for telepresence in the remote conference session.



Figure 5. The remote session chair throws a spotlight on an audience member to give her the chance to speak.

In Figure 3, the remote session chair has selected the podium as the camera view, the wall as the projection surface, and used the grab option to provide the participants a video stream of his face. To do so, he simply runs his standard application to a local camera that displays his image on the computer desktop. By selecting the grab option, the chair can place the camera application exactly on the are where the desktop is being grabbed (shown by a yellow rectangle in Figure 3). As seen in the camera view on the right, the live video showing the chair's face is projected on the wall next to the podium in the conference room. Of course, the quality of the video depends on the speed of the connection between the local and the remote environments.

In Figure 4, the session chair has selected the same camera view and projection surface. In this case, however, he has used the "draw" option to transmit the drawing in the lower window, where he is showing the time left for the speaker. As seen in the live camera view, this information is projected on the wall next to the speaker in the conference.

In Figure 5, the chair has selected the camera view that shows the audience in the conference. In the lower window, he has chosen the highlight as the drawing option and moved this highlight onto

the face of one of the audience members who has raised her hand to ask a question. Using the direct mapping option instead of a pre-defined surface, he uses the highlight (without any keystone correction) to "point" to the audience member to give her a chance to speak. As seen in the live video in the upper window the audience member in the conference room is now highlighted with a projected spotlight. By using processes like this, the chair can effectively point to people and objects in the conference environment in a very intuitive manner.

3. APPLICATIONS

We see a number of other applications for such remote controlled projector-camera systems. In workspace collaboration, such systems allow people in remote offices to work with each other using the whiteboards, table etc. in each other's offices for collaboration. In a retail application, a remote expert can interact with a customer in the store who needs help and explain the features of a product by pointing to and drawing/writing on the product. Similarly, a remote expert on store design or advertising and signage can guide local store personnel in many stores equipped with such systems. Such systems can also help people with accessibility needs in assisted living homes. Here, remote nurses and healthcare professionals can view and help people with instructions and advice at the location where the person in need happens to be such as the bed or the floor. These systems also provide new capabilities for education and distance learning as illustrated by the example in this paper. Finally, new forms of entertainment and shared interactive games between remote people are enabled by such systems.

4. CONCLUSIONS AND FUTURE WORK

In this paper, we presented a simple demonstration to indicate the power of projector-camera systems in telepresence applications. The applications detailed in this work focus on providing mechanisms for a remote participant in a meeting to act on the local environment through projected imagery.

Future work will address several issues. One important problem is providing better means for remote people to calibrate new projection surfaces in a local environment. This can also be important simply as a way for a remote expert to set up such a system in an environment even in applications that are not focused on telepresence. Another interesting issue is exploring new forms of interaction when users at both ends have projector-camera systems as the primary means for interaction.

5. ACKNOWLEDGEMENTS

The authors would like to thank all the members of the Steerable Interfaces and Telepresence team for their valuable suggestions.

6. REFERENCES

- [1] R. Azuma. A Survey of Augmented Reality, In *Presence: Teleoperators and Virtual Environments*, Vol. 6, No. 4, August 1997, pp. 355-385.
- [2] Renato Cerqueira, Christopher K. Hess, Manuel Román, Roy H. Campbell. Gaia: A Development Infrastructure for Active Spaces. In *Workshop on Application Models and Programming Tools for Ubiquitous Computing* (held in conjunction with the UBICOMP 2001), September 2001, Atlanta, Georgia.
- [3] T.J. Cham, C. Pinhanez, J. M. Rehg, R.S. Sukthankar. IEEE International Workshop on Projector-Camera Systems, Nice, France, October 2003.
- [4] Churchill, E.F. and Snowdon, D. Collaborative Virtual Environment: An Introductory Review of Issues and Systems. *Virtual Reality*, 3, 1998, 3-15.
- [5] R. Jain. Experiential Computing. *Communications of the ACM*. Volume 46, Number 7, pp. 48-54, July 2003.
- [6] Kjeldsen, F.C., Levas, A. & Pinhanez, C.S. (2003), Dynamically Reconfigurable Vision-based User Interfaces, in *ICVS'03: Proceedings of the Third International Conference on Computer Vision Systems*.
- [7] Pingali, G.S., Pinhanez, C.S., Levas, A., Kjeldsen, F.C., Podlaseck, M.E., Chen, H. & Sukaviriya, N., (2003), Steerable Interfaces for Pervasive Computing Spaces, in *PerCom'03: Proceedings of the IEEE International Conference on Pervasive Computing and Communications*.
- [8] G. Pingali and N. Sukaviriya. Augmented Collaborative Spaces. ACM Workshop on Experiential Telepresence. December 2003.
- [9] Pinhanez, C. The Everywhere Displays Projector: A Device to Create Ubiquitous Graphical Interfaces. In *Proc. of Ubicomp'01*. 2001. Atlanta, Georgia.
- [10] Lawrence Rowe and Ramesh Jain. ACM SIGMM Retreat Report on Future Directions in Multimedia Research. http://www.acm.org/sigmm/main/events/sigmm_retreat/sigmm-retreat03-final.pdf. March 4, 2004
- [11] Y. Rui, L. He, A. Gupta, and Q. Liu, "Building an Intelligent Camera Management System," *Proceedings of ACM Multimedia Conference (MM'01)*, 2001.
- [12] V. Vinge. When Cyberspace leaks into the real world. *Proceedings of the 4th International Conference on Collaborative Virtual Environments*, Bonn, 2002.

Appendix D-2



Projector-camera Systems for Telepresence

Claudio Pinhanez and Gopal Pingali
IBM T. J. Watson Research Center
19 Skyline Drive
Hawthorne, NY 10532
{pinhanez,gpingali}@us.ibm.com

ABSTRACT

Projector-camera systems provide tools for telepresence that enable novel experiences not possible with currently available schemes for remote interaction including video conferencing systems, virtual reality, and augmented reality systems. Projector-camera systems have the potential not only to allow remote individuals to view each other's environments but also to communicate in new ways by overlaying and embedding information and visualizations on *real* objects and surfaces in the remote environment, without requiring people in the remote environment to “wire” the space or wear special devices. We demonstrate some of this potential with an initial implementation.

Categories and Subject Descriptors

H.5.1 [Information Systems and Presentation]: Multimedia Information Systems – Artificial, augmented, and virtual realities; H.5.2: User Interfaces; H.5.3: Group and Organization Interfaces; H.4.3 [Information Systems Applications]: Communications Applications – Computer conferencing, teleconferencing, and videoconferencing.

General Terms

Algorithms, Design, Experimentation, Human Factors.

Keywords

Collaboration, remote interaction, interfaces, meetings, context modeling, action capture, presentation systems, ubiquitous computing, pervasive systems.

1. INTRODUCTION

Research in multimedia and telepresence is working towards the grand challenge of making “*interactions with remote people and environments nearly the same as interactions with local people and environments*” [10]. There have been many years of research in this area spanning collaborative virtual environments [4], video conferencing systems [11], and augmented reality systems [1]. For the most part, these systems have been centered on special devices – needing a special

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ETP '04, October 15, 2004, New York, New York, USA.

Copyright 2004 ACM 1-58113-933-0/04/0010...\$5.00.

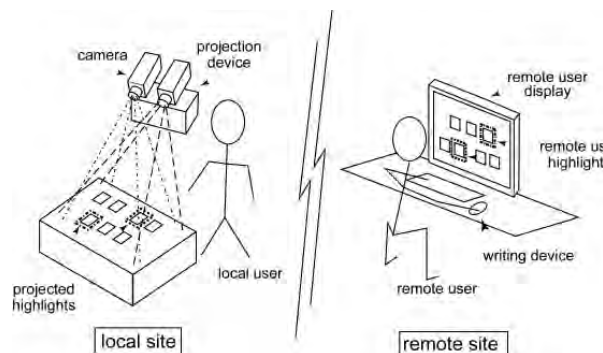


Figure 1. Overview of using a projector-camera system in a telepresence application.

computer or special conferencing system at either end or requiring people to sit in front of a special display or wear special devices to interact with remote people. With recent advances, it is increasingly becoming possible to look at controlling and manipulating an entire space rather than one special device in the space. This is leading to “smart” or “active” spaces and “room operating systems” [2] that treat the whole environment with its diverse set of computing, input and output devices, and sensors and actuators, as one collection of resources to be made available to users in a unified manner, much like a conventional operating system treats the resources on a desktop computer. This evolution of computing towards smart spaces is opening up new vistas for telepresence. It is now becoming possible to produce a different kind of immersive experience, remote interaction, and collaboration by augmenting the space as a whole [8]. This whole move towards smart spaces is also significant in the context of experiential computing as envisioned in [5].

In this context, we see a significant role for camera-projector systems [3]. These systems can add a new dimension to telepresence that has hitherto been uncommon. Projection and lighting can be used as an element of telepresence that allows overlaying and embedding information and visualizations on real objects, besides producing informative displays wherever they are needed. The Everywhere Display [9] and its dynamically reconfigurable interactive capabilities [6] is a good example of a projector-camera system that has significantly augmented spaces in this manner. When combined with people tracking and environment modeling, the Everywhere Display can be extended to produce steerable interfaces [7] that appear anywhere in a space and dynamically adapt to user activities in an environment and changes in the environment. While previous work on the camera-

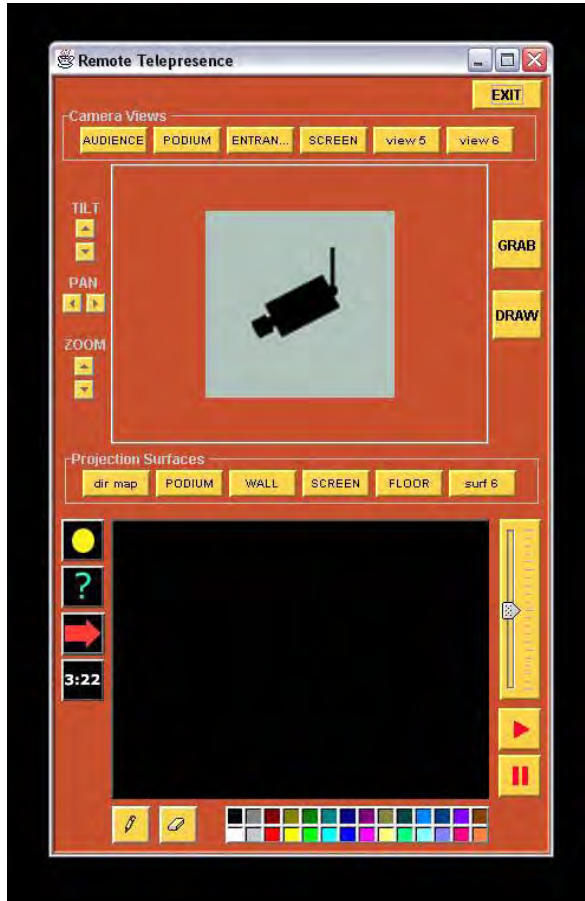


Figure 2. Example of control interface for a remote session chair using a projector-camera system to communicate with local conference participants.

projector systems and steerable interfaces has been primarily focused on local augmentation, we begin to explore its potential in telepresence or remote augmentation here.

Figure 1 shows an overview of our proposal for using camera-projector systems in telepresence. Here, an environment is equipped with steerable camera(s) and steerable projector(s). A user in a remote space can access/view and control both the projector and the camera through a display and writing device. A user at a remote site can view the local environment through video from the remote camera. The remote user can also control the projector in the local environment to project highlights, video, text, etc. in the local environment. As the projector is steerable, these virtual objects can be moved on to real objects and people in the local environment by the remote user. The remote user's interface can itself be an interactive steerable display although this is not a requirement. In this paper, we describe an initial demonstration we have put together to further illustrate this vision.

2. DEMONSTRATION OVERVIEW

This demonstration uses a steerable projector-camera system – the Everywhere Display – in a local environment and provides a control interface to a user in a remote environment. For illustrative purposes, we demonstrate this system in the context of a

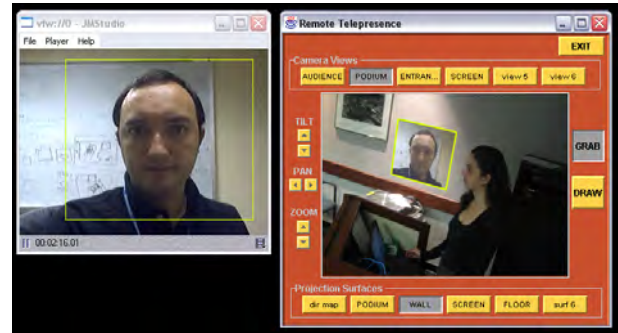


Figure 3. The session chair projects live video showing his face on to the wall next to the speaker's podium in the remote environment.

conference/workshop taking place in a local environment for which a session chair of the workshop is in a remote location. The chair can view the workshop live from a remote location through on-site cameras and can provide instructions and messages to the local workshop attendees. We assume that the chair also has two-way audio communication to the conference.

Figure 2 shows a screenshot of the control interface available to the session chair. The upper window in this figure shows the interface to camera views. This window displays video from (one of) the remote camera(s) and enables the chair to manipulate the remote camera(s) through the pan, tilt, and zoom controls and also define camera views. In the example in Figure 2, pre-defined views include audience, podium, entrance, screen etc. The lower window in Figure 2 shows the interface for the chair to control the remote Everywhere Display projector. As discussed in [6], this projector has the notion of a “projection surface”. Calibration of a projection surface ensures that projections can appear on these surfaces without keystone distortion as discussed in [9]. Different projection surfaces can be defined and calibrated for the environment in which the ED-projector is situated. The mid window in Figure 2 (labeled “Projection Surfaces”) shows how these pre-defined projection surfaces are made available to the remote user, the session chair in this case. Several available projection surfaces are shown in Figure 2 such as podium, wall, screen, floor etc. In addition to pre-calibrated surfaces, there is also a “direct mapping” option which automatically associates the current camera view to an overlapping projection, as will be discussed shortly.

There are two basic ways determine what is going to be projected on the selected projection surface: grab and draw, corresponding to the two buttons on the upper right of Figure 2. The “grab” button on the upper right in Figure 2 allows the chair to grab any portion of his computer desktop and project it on the remote environment. Any computer application on the grab area is then projected onto the selected surface of the local environment. For example, the chair may choose to project his own image (as shown in Figure 3), a webpage running on a browser, etc.

Pressing the “draw” button on the upper right of the control interface shown in Figure 2 enables an embedded drawing application that allows the chair to easily “paint” the projected surface. Any drawing appearing on the lower window is immediately and efficiently transferred to the remote projector. The blank area in the lower window of Figure 2 is therefore a drawing canvas. Moreover, the real effect on the local



Figure 4. The remote session chair indicates to the speaker how much time is left for her talk.

environment can now be monitored through the live video in the upper window.

The icons on the left of the lower window show predefined drawing icons such as a highlight (which the session chair can use to highlight an object in the remote environment), a question mark (which the session chair can use to indicate it is time for questions), an arrow mark (to provide directions or point to objects in the remote environment), and a stop-watch icon (that can be used to show a speaker how much time is left for their talk). Besides these icons, the chair has a regular paint/drawing interface to draw on the canvas. The icons on the lower right allow the session chair to resize any element of the drawing canvas, and to switch between “live drawing” (the “play” button) where any modification on the canvas is immediately projected on the conference environment; and “hidden drawing” (the “pause” button), where the chair can draw, size, and compose all elements before displaying them.

The following three figures and descriptions provide some examples of how this interface is used by the session chair for telepresence in the remote conference session.



Figure 5. The remote session chair throws a spotlight on an audience member to give her the chance to speak.

In Figure 3, the remote session chair has selected the podium as the camera view, the wall as the projection surface, and used the grab option to provide the participants a video stream of his face. To do so, he simply runs his standard application to a local camera that displays his image on the computer desktop. By selecting the grab option, the chair can place the camera application exactly on the are where the desktop is being grabbed (shown by a yellow rectangle in Figure 3). As seen in the camera view on the right, the live video showing the chair’s face is projected on the wall next to the podium in the conference room. Of course, the quality of the video depends on the speed of the connection between the local and the remote environments.

In Figure 4, the session chair has selected the same camera view and projection surface. In this case, however, he has used the “draw” option to transmit the drawing in the lower window, where he is showing the time left for the speaker. As seen in the live camera view, this information is projected on the wall next to the speaker in the conference.

In Figure 5, the chair has selected the camera view that shows the audience in the conference. In the lower window, he has chosen the highlight as the drawing option and moved this highlight onto

the face of one of the audience members who has raised her hand to ask a question. Using the direct mapping option instead of a pre-defined surface, he uses the highlight (without any keystone correction) to “point” to the audience member to give her a chance to speak. As seen in the live video in the upper window the audience member in the conference room is now highlighted with a projected spotlight. By using processes like this, the chair can effectively point to people and objects in the conference environment in a very intuitive manner.

3. APPLICATIONS

We see a number of other applications for such remote controlled projector-camera systems. In workspace collaboration, such systems allow people in remote offices to work with each other using the whiteboards, table etc. in each other's offices for collaboration. In a retail application, a remote expert can interact with a customer in the store who needs help and explain the features of a product by pointing to and drawing/writing on the product. Similarly, a remote expert on store design or advertising and signage can guide local store personnel in many stores equipped with such systems. Such systems can also help people with accessibility needs in assisted living homes. Here, remote nurses and healthcare professionals can view and help people with instructions and advice at the location where the person in need happens to be such as the bed or the floor. These systems also provide new capabilities for education and distance learning as illustrated by the example in this paper. Finally, new forms of entertainment and shared interactive games between remote people are enabled by such systems.

4. CONCLUSIONS AND FUTURE WORK

In this paper, we presented a simple demonstration to indicate the power of projector-camera systems in telepresence applications. The applications detailed in this work focus on providing mechanisms for a remote participant in a meeting to act on the local environment through projected imagery.

Future work will address several issues. One important problem is providing better means for remote people to calibrate new projection surfaces in a local environment. This can also be important simply as a way for a remote expert to set up such a system in an environment even in applications that are not focused on telepresence. Another interesting issue is exploring new forms of interaction when users at both ends have projector-camera systems as the primary means for interaction.

5. ACKNOWLEDGEMENTS

The authors would like to thank all the members of the Steerable Interfaces and Telepresence team for their valuable suggestions.

6. REFERENCES

- [1] R. Azuma. A Survey of Augmented Reality, In Presence: Teleoperators and Virtual Environments , Vol. 6, No. 4, August 1997, pp. 355 -385.
- [2] Renato Cerqueira, Christopher K. Hess, Manuel Román, Roy H. Campbell. Gaia: A Development Infrastructure for Active Spaces. In Workshop on Application Models and Programming Tools for Ubiquitous Computing (held in conjunction with the UBICOMP 2001), September 2001, Atlanta, Georgia.
- [3] T.J. Cham, C. Pinhanez, J. M. Rehg, R.S. Sukthankar. IEEE International Workshop on Projector-Camera Systems, Nice, France, October 2003.
- [4] Churchill, E.F. and Snowdon, D. Collaborative Virtual Environment: An Introductory Review of Issues and Systems. *Virtual Reality*, 3, 1998, 3—15.
- [5] R. Jain. Experiential Computing. *Communications of the ACM*. Volume 46, Number 7, pp. 48-54, July 2003.
- [6] Kjeldsen, F.C., Levas, A. & Pinhanez, C.S. (2003), Dynamically Reconfigurable Vision-based User Interfaces, in *ICVS'03: Proceedings of the Third International Conference on Computer Vision Systems*.
- [7] Pingali, G.S., Pinhanez, C.S., Levas, A., Kjeldsen, F.C., Podlaseck, M.E., Chen, H. & Sukaviriya, N., (2003), Steerable Interfaces for Pervasive Computing Spaces, in *PerCom'03: Proceedings of the IEEE International Conference on Pervasive Computing and Communications*.
- [8] G. Pingali and N. Sukaviriya. Augmented Collaborative Spaces. ACM Workshop on Experiential Telepresence. December 2003.
- [9] Pinhanez, C. The Everywhere Displays Projector: A Device to Create Ubiquitous Graphical Interfaces. In *Proc. of UbiComp'01*. 2001. Atlanta, Georgia.
- [10] Lawrence Rowe and Ramesh Jain. ACM SIGMM Retreat Report on Future Directions in Multimedia Research. http://www.acm.org/sigmm/main/events/sigmm_retreat/sigmm-retreat03-final.pdf. March 4, 2004
- [11] Y. Rui, L. He, A. Gupta, and Q. Liu, "Building an Intelligent Camera Management System," *Proceedings of ACM Multimedia Conference (MM'01)*, 2001.
- [12] V. Vinge. When Cyberspace leaks into the real world. *Proceedings of the 4th International Conference on Collaborative Virtual Environments*, Bonn, 2002.

Appendix D-3



**Proceedings of the ACM SIGMM 2004 Workshop on Effective Telepresence :
ETP 2004 : October 15, 2004, New York, New York, USA**

Authors: [ACM SIGMM Workshop on Effective Telepresence](#), [Gopal Pingali](#), [Frank Nack](#), [Hari Sundaram](#), [ACM Special Interest Group on Multimedia](#), [ACM Digital Library](#)

 **Print Book**, English, 2004

Edition: [View all formats and editions](#)

Publisher: ACM Press, New York, NY, 2004

Genre:	Conference papers and proceedings
Physical Description:	vii, 75 pages : illustrations (some color) ; 28 cm
ISBN:	9781581139334, 1581139330
OCLC Number / Unique Identifier:	62896832
Subjects:	Computer Science Conference papers and proceedings Multimedia systems Multimedia systems Congresses Multimédia Congrès Video telephone Video telephone Congresses Visiophone Congrès

 3 editions in 304 libraries

Featured libraries

All libraries

Showing 237 libraries near [Washington, District of Columbia](#)

☐ Only public libraries



Academic

George Washington University



[1.2 miles](#) from your current location.

Melvin Gelman Lib, 2130 H St NW, Washington, DC, 20052-2521,
United States

Borrow 



Favorite
Library



Regional or
National

Library of Congress



[2.2 miles](#) from your current location.

101 Independence Avenue, SE, Washington, DC, 20540, United States

Borrow 



Favorite
Library

Appendix D-4

LIBRARY

CATALOG



1 of 1



BOOK

Proceedings of the ACM SIGMM 2004 Workshop on Effective Telepresence : ...

Full Record

[MARC Tags](#)

Meeting name

[ACM SIGMM Workshop on Effective Telepresence \(2004 : New York, NY, USA\)](#)

Main title

Proceedings of the ACM SIGMM 2004 Workshop on Effective Telepresence : ETP 2004 : October 15, 2004, New York, New York, USA / [general chairs: Gopal Pingali, Frank Nack, Hari Sundaram].

Published/Created

New York, NY : ACM Press, 2004.

[Request this Item](#)

[LC Find It](#)

Links



Links

<http://portal.acm.org/toc.cfm?id=1026776&coll=portal&dl=ACM&type=proceeding&idx=SERIES316&part=Proceedings&WantType=Proceedings&title=International%20Multimedia%20Conference>

More Information



LCCN Permalink	https://lccn.loc.gov/2005272817
Description	vii, 75 p. : ill. (some col.) ; 28 cm.
ISBN	1581139330
LC classification	MLCM 2006/41169 (T)
Portion of title	ACM SIGMM 2004 Workshop on Effective Telepresence ETP 2004 Effective Telepresence
Related names	Pingali, Gopal. Nack, Frank. Sundaram, Hari, 1971- ACM Special Interest Group on Multimedia. ACM Digital Library.
LC Subjects	Video telephone--Congresses. Multimedia systems--Congresses.
Browse by shelf order	MLCM 2006/41169 (T)
Notes	"ACM order number 105049"--T.p. verso. Includes bibliographical references and author index.
Additional formats	Also issued online.
LCCN	2005272817
Other system no.	(OCoLC)ocm62896832
Type of material	Book

Item Availability



CALL NUMBER	MLCM 2006/41169 FT MEADE Copy 1
Request in	Jefferson or Adams Building Reading Rooms - STORED OFFSITE
Status	Not Charged

CALL NUMBER

MLCM 2006/41169 FT MEADE

Copy 2

Request in

Jefferson or Adams Building Reading Rooms - STORED OFFSITE

Status

Not Charged

Appendix D-5

LIBRARY

CATALOG



1 of 1



BOOK

Proceedings of the ACM SIGMM 2004 Workshop
on Effective Telepresence : ...[Full Record](#)[MARC Tags](#)

000	01874cam a22003973a 4500
001	13927657
005	20081112165732.0
008	050412s2004 nyua b 101 0 eng d
906	— a 7 b cbc c copycat d 2 e ncip f 20 g y-gencatlg
925	0_ a acquire b 2 shelf copies x policy default
955	— a pv21 2005-04-12 to ASCD a jx00 2006-01-27 a ja15 2006-08-03 z-processor a jf01 2006-08-25 to Ft. Meade UNASSESSED; a xh00 2008-11-12 SEND TO DISCARD COPY NUMBERS GREATER THAN 2 UNLESS REFERENCE OR OTHER REQUEST TO BE KEPT
010	— a 2005272817
020	— a 1581139330
035	— a (OCoLC)ocm62896832
040	— a C\$A c C\$A d OCLCA d IXA d DLC
050	00 a MLCM 2006/41169 (T)
097	14 a TK6505 b .A36 2004
111	2_ a ACM SIGMM Workshop on Effective Telepresence d (2004 : c New York, NY, USA)
245	10 a Proceedings of the ACM SIGMM 2004 Workshop on Effective Telepresence : b ETP 2004 : October 15, 2004, New York, New York, USA / c [general chairs: Gopal Pingali, Frank Nack, Hari Sundaram].
246	30 a ACM SIGMM 2004 Workshop on Effective Telepresence
246	30 a ETP 2004

246	30	a Effective Telepresence
260	—	a New York, NY : b ACM Press, c 2004.
300	—	a vii, 75 p. : b ill. (some col.) ; c 28 cm.
500	—	a "ACM order number 105049"--T.p. verso.
530	—	a Also issued online.
504	—	a Includes bibliographical references and author index.
650	_0	a Video telephone v Congresses.
650	_0	a Multimedia systems v Congresses.
700	1_	a Pingali, Gopal.
700	1_	a Nack, Frank.
700	1_	a Sundaram, Hari, d 1971-
710	2_	a ACM Special Interest Group on Multimedia.
710	2_	a ACM Digital Library.
856	40	u http://portal.acm.org/toc.cfm?id=1026776&coll=portal&dl=ACM&type=proceeding&idx=SERIES316&part=Proceedings&WantType=Proceedings&title=International%20Multimedia%20Conference
985	—	e Reorg T

[Request this Item](#)[LC Find It](#)

Item Availability

CALL NUMBER

[MLCM 2006/41169 FT MEADE](#)

Copy 1

Request in

Jefferson or Adams Building Reading Rooms - STORED OFFSITE

Status

Not Charged

CALL NUMBER

[MLCM 2006/41169 FT MEADE](#)

Copy 2

Request in

Jefferson or Adams Building Reading Rooms - STORED OFFSITE

Status

Not Charged

Appendix D-6

Projector-camera systems for telepresence

☐ Search within citing articles

Information processing system, information processing apparatus, and information processing method

K Hirata, T Iyoda - US Patent 7,527,383, 2009 - Google Patents

An information processing system includes a projection unit that projects at least image data for calibration, an image data capturing unit that captures projection image data projected by ...

☆ Save  Cite Cited by 21 Related articles All 4 versions 

Unobtrusive augmentation of physical environments: interaction techniques, spatial displays and ubiquitous sensing

A Olwal - 2009 - diva-portal.org

The fundamental idea of Augmented Reality (AR) is to improve and enhance our perception of the surroundings, through the use of sensing, computing and display systems that make it ...

☆ Save  Cite Cited by 11 Related articles All 10 versions 

An Introduction to Augmented Reality

A Olwal - An introduction to augmented reality, 2009 - researchgate.net

The goal of Augmented Reality (AR) is to improve and enhance our perception of the surroundings by combining sensing, computing and display technologies. Most AR research ...

☆ Save  Cite Cited by 12 Related articles 

Evaluating Techniques to Share Hand Gestures for Remote Collaboration using Top-Down Projection in a Virtual Environment

T Teo, K Sakudara, M Fukuoka, M Sugimoto - 2022 - diglib.eg.org

Sharing hand gestures in a remote collaboration offers natural and expressive communication between collaborators. Proposed techniques allow sharing dependent ...

☆ Save  Cite Related articles All 2 versions 

Towards ambient business: enabling open innovation in a world of ubiquitous computing

C Schmitt, D Schoder, K Fischbach... - Advances in Ubiquitous ..., 2008 - igi-global.com

Ubiquitous computing enables the development of new innovative applications and services. Particularly influential on future business services will be the connection of the real ...

☆ Save  Cite Cited by 8 Related articles All 2 versions

Mobile devices and a new understanding of presence

K Weber - Workshop paper from SISSI2010 at the 12th annual ..., 2010 - academia.edu

Mobile devices help people to stay connected across huge distances and to socially interact while in spatially separated environments. However, compared to face-to-face-interaction ...

☆ Save  Cite Cited by 3 Related articles All 4 versions 

Mobile devices, virtual presence, and surveillance: questions concerning epistemology and some new challenges for privacy and data protection

K Weber - ... : Questions Concerning Epistemology and Some New ..., 2011 - papers.ssrn.com

Mobile devices help people to stay connected across huge distances and to socially interact while in spatially separated environments. However, compared to face-to-face-interaction ...

☆ Save  Cite Cited by 1 Related articles All 5 versions 

Appendix D-7

Advances in Ubiquitous Computing: Future Paradigms and Directions

Soraya Kouadri Mostefaoui
Oxford Brookes University, UK

Zakaria Maamar
Zayed University, UAE

George M. Giaglis
Athens University of Economics and Business, Greece



IGI PUBLISHING

Hershey • New York

Acquisition Editor: Kristin Klinger
Senior Managing Editor: Jennifer Neidig
Managing Editor: Sara Reed
Development Editor: Kristin Roth
Copy Editor: Brenda Leach
Typesetter: Michael Brehm
Cover Design: Lisa Tosheff
Printed at: Yurchak Printing Inc.

Published in the United States of America by
IGI Publishing (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: <http://www.igi-global.com>

and in the United Kingdom by
IGI Publishing (an imprint of IGI Global)
3 Henrietta Street
Covent Garden
London WC2E 8LU
Tel: 44 20 7240 0856
Fax: 44 20 7379 0609
Web site: <http://www.eurospanonline.com>

Copyright © 2008 by IGI Global. All rights reserved. No part of this book may be reproduced in any form or by any means, electronic or mechanical, including photocopying, without written permission from the publisher.

Product or company names used in this book are for identification purposes only. Inclusion of the names of the products or companies does not indicate a claim of ownership by IGI Global of the trademark or registered trademark.

Library of Congress Cataloging-in-Publication Data

Advances in ubiquitous computing : future paradigms and directions / Soraya Kouadri Mostéfaoui, Zakaria Maamar, and George M. Giaglis, editors.

p. cm.

Summary: "This book investigates the technology of ubiquitous computing, emerging applications and services, and social issues vital for the successful deployment of a ubiquitous computing application. Providing high quality, authoritative content on such topics as device design, wireless communication, location sensing, privacy concerns, attention focus, multi-person interaction, and direct interaction, work patterns, it is a must-have in library collections"--Provided by publisher.

Includes bibliographical references and index.

ISBN 978-1-59904-840-6 (hardcover) -- ISBN 978-1-59904-842-0 (ebook)

1. Ubiquitous computing. I. Mostéfaoui, Soraya Kouadri. II. Maamar, Zakaria. III. Giaglis, George M.

QA76.5915.A395 2008

004--dc22

2007037380

British Cataloguing in Publication Data

A Cataloguing in Publication record for this book is available from the British Library.

All work contributed to this book is new, previously-unpublished material. The views expressed in this book are those of the authors, but not necessarily of the publisher.

Advances in Ubiquitous Computing: Future Paradigms and Directions

Table of Contents

Foreword.....	vi
Preface.....	viii
Acknowledgment.....	xiii
 Chapter I	
Mobile Phone and Visual Tags:	
Linking the Physical World to the Digital Domain.....	1
<i>Marco Avvenuti, Università di Pisa, Italy</i>	
<i>Alessio Vecchio, Università di Pisa, Italy</i>	
 Chapter II	
Context-Aware Mobile Learning on the Semantic Web	23
<i>Rachid Benlamri, Lakehead University, Canada</i>	
<i>Jawad Berri, Etisalat University College, United Arab Emirates</i>	
<i>Yacine Atif, Massey University, New Zealand</i>	
 Chapter III	
Model-Driven Development for Pervasive Information Systems	45
<i>José Eduardo Fernandes, Bragança Polytechnic Institute, Bragança, Portugal</i>	
<i>Ricardo J. Machado, University of Minho, Guimarães, Portugal</i>	
<i>João Álvaro Carvalho, University of Minho, Guimarães, Portugal</i>	

Chapter IV**Device Localization in Ubiquitous Computing Environments 83***Rui Huang, The University of Texas at Arlington, USA**Gergely V. Záruba, The University of Texas at Arlington, USA**Sajal Das, The University of Texas at Arlington, USA***Chapter V****Enabling Programmable Ubiquitous Computing Environments:****A Middleware Perspective..... 117***Christine Julien, The University of Texas at Austin, USA**Sanem Kabadayi, The University of Texas at Austin, USA***Chapter VI****Determinants of User Acceptance for RFID Ticketing Systems 150***Dimitrios C. Karaiskos, Athens University of Business and Economics, Greece**Panayiotis E. Kourouthanassis, Athens University of Business and Economics, Greece***Chapter VII****Designing for Tasks in Ubiquitous Computing:****Challenges and Considerations..... 171***Stephen Kimani, Jomo Kenyatta University of Agriculture and Technology, Kenya**Silvia Gabrielli, University of Rome "La Sapienza," Italy**Tiziana Catarci, University of Rome "La Sapienza," Italy**Alan Dix, Lancaster University, UK***Chapter VIII****Kinetic User Interfaces: Physical Embodied Interaction with****Mobile Ubiquitous Computing Systems..... 201***Vincenzo Pallotta, University of Fribourg, Switzerland**Pascal Bruegger, University of Fribourg, Switzerland**Béat Hirsbrunner, University of Fribourg, Switzerland***Chapter IX****M-Traffic: Mobile Traffic Information and Monitoring System 229***Teresa Romão, Universidade de Évora, Portugal**Luís Rato, Universidade de Évora, Portugal**Antão Almada, YDreams, Portugal**A. Eduardo Dias, Universidade de Évora, Portugal***Chapter X****Towards Ambient Business: Enabling Open Innovation in a****World of Ubiquitous Computing 251***Christian Schmitt, University of Cologne, Köln, Germany**Detlef Schoder, University of Cologne, Köln, Germany**Kai Fischbach, University of Cologne, Köln, Germany**Steffen Muhle, University of Cologne, Köln, Germany*

Chapter XI**Activity-Oriented Computing..... 280***João Pedro Sousa, George Mason University, USA**Bradley Schmerl, Carnegie Mellon University, USA**Peter Steenkiste, Carnegie Mellon University, USA**David Garlan, Carnegie Mellon University, USA***Chapter XII****Privacy Threats in Emerging Ubicomp Applications:****Analysis and Safeguarding..... 316***Elena Vildjiounaite, VTT Technical Research Centre of Finland, Finland**Tapani Rantakokko, Finwe LTD, Finland**Petteri Alahuhta, VTT Technical Research Centre of Finland, Finland**Pasi Ahonen, VTT Technical Research Centre of Finland, Finland**David Wright, Trilateral Research and Consulting, UK**Michael Friedewald, Fraunhofer Institute Systems and Innovation Research
Germany***About the Contributors 348****Index 358**

**IGI PUBLISHING**701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-<http://www.igi-global.com>**ITB15286**

This paper appears in the publication, **Advances in Ubiquitous Computing: Future Paradigms and Directions** edited by S. Mostefaoui, Z. Maamar, and G. Giaglis © 2008, IGI Global

Chapter X

Towards Ambient Business: Enabling Open Innovation in a World of Ubiquitous Computing

Christian Schmitt, University of Cologne, Germany

Detlef Schoder, University of Cologne, Germany

Kai Fischbach, University of Cologne, Germany

Steffen Muhle, University of Cologne, Germany

Abstract

Ubiquitous computing enables the development of new innovative applications and services. Particularly influential on future business services will be the connection of the real with the virtual world by embedding computers or smallest processors, memory chips, and sensors into the environment and into physical objects, as well as using natural, multimodal customer interaction. Bearing in mind that ubiquitous computing entails the interaction of mobile smart objects and local smart environments hosted by different service providers, we propose an open approach to encourage the development of new and innovative smart applications and services. Considering the Open Source community, we assert that such an open approach reveals innovation potentials that cannot be achieved in proprietary information

Copyright © 2008, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

system environments. Most research projects, as well as commercial initiatives, however, focus mainly on specific, proprietary applications like smart shopping environments, and do not incorporate further prospects of using an open approach. Therefore, this chapter discusses as a first step the impact of Open Innovation in a world of ubiquitous computing from an economic perspective. Then, the design of an Open Object Information Infrastructure (OOII) that enables Open Innovation in the context of ubiquitous computing is presented. Finally, an innovative smart service called the Federative Library (FedLib), which represents a first instantiation of the OOII, is introduced to demonstrate the feasibility of the design.

Introduction

The connection of the real and the virtual world encourages the development of new and innovative applications and services. Google is a prominent case which distinctly underlines the creative power of combining both worlds (Roush, 2005). In June of 2005, Google released an official API for Google Maps and Google Earth. Through this, consumers and external developers now have access to detailed aerial and satellite maps and advanced geographical visualization tools that can easily be used to display data atop the Google maps. Almost immediately, a community of “geotagging” map makers was formed that ties information on the Web to geographical coordinates in order to build geospatial applications (examples of so-called “map mash-ups” can be found in Gibson, 2006). Geotagging means amplifying physical places with information so that, for instance, users of mobile devices with location technologies can retrieve additional information related to their current locations. This includes location-based ads and listings for nearby shopping, dining, entertainment, and business outlets or even photographs, videos, stories, and other personal information uploaded to the Internet and pinned to specific latitudes and longitudes. As a result, the community participates in the innovation process and in building “a browser for the earth,” as John Hanke, general manager of Google’s Keyhole Group, describes Google Earth (Roush, 2005). This could enable a new and natural means of accessing location-based information. Wade Roush (2005) describes this vision by drawing a comparison: “Every page on the Web has a location, in the form of a URL. Now every location can have a Web page [...] it means that navigating both the Web and the real geography around us is about to become a much richer experience, rife with occasions for on-the-spot education and commerce. It means that we will be able to browse the Web-and the virtual earth encompassed within it-simply by walking around.” (p. 60)

As the Google case also illustrates, opening access to information systems and intellectual property (IP) to complementors, suppliers, users, and even to competitors encourages innovation, increases the attractiveness of platforms, and enriches

service provision through the incorporation of external information sources. With the world evolving from Electronic Business to ubiquitous computing (Weiser, 1991) (and pervasive computing (Hansmann, 2003) or ambient intelligence (Aarts, 2001) respectively), even more comprehensive opportunities for implementing innovation are likely to arise. Innovative information gathering and access is enabled by the connection between the real with the virtual world by embedding computers or the smallest processors, memory chips, and sensors into the environment and into physical objects (Gershenfeld, 1999), as well as using natural, multimodal user interaction (Dam, 2001).

However, existing approaches frequently rely on rather proprietary information systems that reveal their data and services only partially or under certain conditions. Bearing in mind that ubiquitous computing entails the interaction of mobile smart objects and local smart environments hosted by different service providers, we propose using an open approach to encourage the development of new and innovative smart applications and services. Considering the Open Source community, we assert that such an open approach reveals innovation potentials that cannot be achieved in proprietary information system environments. Various developers with different perceptions can then search for relevant smart services, as well as information gathered by smart objects or environments, combine them freely, and hence create their own innovative services.

The rest of the chapter is structured as follows. The following section presents related work. Then, we first introduce our framework for ambient business. We then discuss the impact and potential of Open Innovation in a world of ambient business from an economic perspective. Next, we present the design of an architecture called the open object information infrastructure (OOII) that enables open innovation in the context of ambient business. To demonstrate the feasibility of the design, we introduce an innovative smart service called the Federative Library (FedLib) that will be the first instantiation of the OOII. The chapter concludes with a summary and further research questions.

Background

Several authors have considered the impact on business that comes with ubiquitous computing. Most work starts with a discussion of technological progress and the business impact arising from that progress. Roussos (2006) adopts this perspective. In his view, contingencies for ubiquitous commerce arise from emerging technologies which again have an impact on business innovation, as well as on society. Similar to that view, Mattern (2005) discusses technological progress that allows deploying smart objects in an informatized world around us. Within this world, informational

counterparts of everyday objects and of people induce social and political challenges and pave the way to new applications. Fleisch and Tellkamp (2006) act on this assumption and provide a framework that helps to identify value-creating application areas and discuss challenges in implementing these applications. Whereas Fleisch and Tellkamp focus on processes within a company or between companies, Fano and Gershman (2002) consider the impact that smart services will have on relationships between customers and the service provider. These relationships will be profoundly affected by new classes of services emerging in smart environments. Based on the vision that a service provider will steadily be able to access customers, they make a set of assumptions about how services will be provided in a world of ubiquitous computing.

By discussing technology, smart environments and several business aspects, the present literature gives considerable insights into how ubiquitous computing will impact future business. However, current frameworks do not even consider the potentials for innovation. For example, the German Federal Office for Information Security provides a framework to discuss the socio-economic impact of ubiquitous computing (BSI, 2006). It considers “determining factors,” “enabling technologies,” “pervasive computing features,” and “Impacts” without addressing innovation. Therefore, we propose in the following section our framework for ambient business before discussing the impact of ambient business on innovation.

The Framework for Ambient Business

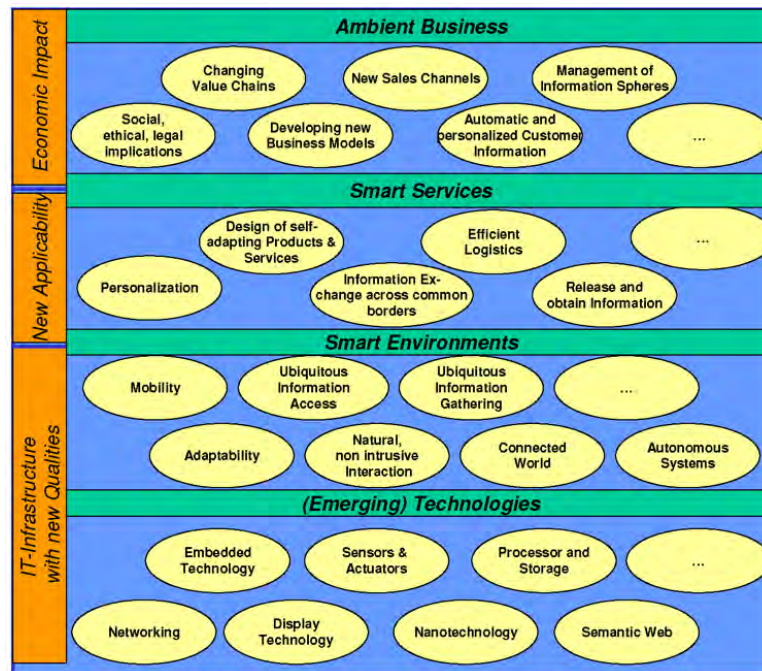
The framework for ambient business divides the analysis of ubiquitous computing in the context of entrepreneurial activities into four perspectives (see Figure 1).

First, the emerging technologies layer focuses on advances in technology during the last few decades that allow the development of new types of IT-infrastructures. These emerging technologies present new qualities that open up new forms of IT usage covered in the smart environments layer. Thus, new applications (smart services layer) can be implemented within smart environments. Finally, the ambient business layer deals with the impact of these innovative smart services on economy, business environment, and society.

Emerging Technologies

Advances in networking as well as embedded technology, display technology, and nanotechnology have been driving the research on ubiquitous computing during recent past decades (Hilty, 2005). Smaller and more powerful, but less power-con-

Figure 1. The Framework for Ambient Business



suming, information technology has facilitated the development not only of common IT-devices. Even objects such as coffee cups (Gellersen, 1999) and pens (Anoto, 2007) can be equipped with sensors, processors, storage capacity, and actuators. The objects can be connected with each other in order to “cooperate,” thus forming joint embedded IT environments. As a result, new opportunities for applications scenarios arise.

Current progress in the layer of emerging technologies is indeed quite remarkable. An in-depth discussion of technologies, therefore, is beyond the scope of this chapter. The remainder of this chapter assumes a portfolio of technologies which designers and developers of applications and services can use as a basis for presenting technical solutions for application demands in a world of ubiquitous computing. In this context, Semantic Web (Berners-Lee, 2001) technologies are of particularly great relevance.

Smart Environments

In the second layer, the qualities of emerging technologies are explored. The role of these qualities is twofold. First, they are the means for revealing new applica-

tions and services based on the emerging technologies. Second, when analyzing the shortcomings of current applications and services, the qualities can pave the way for satisfying people's innovation needs at the emerging technologies level.

An environment equipped with an IT-infrastructure as described above is able to gather information about people, things, and other surrounding conditions. The information can be used to recognize the context in which people and things are embedded, so that the environment becomes aware of the real world. Thus, the environment is able to estimate the needs arising from the real world and to adapt itself to (mobile) users and situations (Coutaz, 2005), or may even act autonomously. The environment becomes "smart."

Emerging technologies not only facilitate extensive information gathering, but also provide new ways to interact with users. Information access and services presented by the smart environment can be almost seamlessly integrated into the user's world. Interfaces become natural and interaction with them subliminal. The technology even disappears from the user's perception (Weiser, 1991; Norman, 1998; Want, 2002).

Although different qualities reveal different potential applications, the qualities of emerging technologies are not independent from each other. In fact, like the emerging technology layer, only by analyzing the interplay of different qualities and by transferring them to the application domain, can the full breadth of application potentials be revealed. Accordingly, these qualities point to smart services that may be developed on top of smart environments.

Smart Services

Merging the real and the virtual world, every relevant object in the real world acquires an informational counterpart in the form of a data object (Mattern, 2005). Thus, a service that uses this informational counterpart may know the location and condition of a physical object. Such services are already applied to ensure efficient logistic processes (Angeles, 2005) and whenever assets have to be managed. Due to the decreasing costs of corresponding technology, services will not only be offered to companies, but a vast range of services is also conceivable for individuals. Books, clothes, and CDs may be managed in a similar manner in the future. By analyzing the current and past context of an object, which comprises additional information from outside the environment, appropriate supportive services can be developed and offered through a so called smart product (Maass, 2006). Additional services may be presented for use, for example, at a medicine cabinet when the medicine runs short (Floerkemeier, 2003). Such services can also be individualized, so that they adapt to the preferences, habits, and current situation of the user (Coutaz, 2005). When services become fully aware of their context, they could be invoked automatically, without explicit manual interaction by the user (Tennenhouse, 2000).

Ambient Business

The ability of smart services to adapt themselves to situations and provide a huge amount of information instantly not only supports current business processes and distribution channels, but it also opens new ways of doing business.

Services in the application domain of supply chain and asset management reveal opportunities for monitoring tangible assets more closely and thus creating a higher degree of control. Yet, the costs of control decline at the same time. Not only high-value products, but also lower-value ones can be managed by efficiently smart services. This results in a very tight mesh of control throughout the company (Fleisch, 2005). Through means of such control, processes can be adjusted to incidents that are internal or external to the processes. This may result in more efficient and reliable processes within an organization.

Inter-organizational processes that utilize smart services may also cut costs in the supply chain. Through crossing internal company borders, information flows that are highly aligned with the flow of physical goods reduce losses, and control overhead that frequently occur at such borders (Strasser, 2005). Furthermore by, opening up their data base, companies are able to offer smart services to each other, for example, control and information services (Fleisch, 2005). Functions that were previously closely linked to one's own organization can now be outsourced. It becomes feasible for companies to offer their outstanding capabilities that were previously bound to the organization to suppliers, customers, and even to competitors.

In the future, products may be provided more as a bundle of the physical product itself and the services linked to it. It may be sold as a platform for generating additional revenues by means of services in the after-sale phase. In this case, the vendor comes closer to the customer through two means. First, he is able to address a customer not only through traditional sales channels like a store. He can also address the customer at the preferred location (Fano, 2002), which may have an enormous impact on business models and sales channels. Second, information may flow in the other direction as well. Users can provide usage information to the vendor and receive special services or a monetary reward in return. The vendor can use this information to align R&D-efforts to users' needs. Going one step further, smart services can be used to involve users in the R&D-process, whether directly or via mediators like Innocentive (Innocentive, 2007).

It can be assumed that ambient business entrepreneurs will find themselves in markets with short innovation cycles. Like the internet, smart environments follow the end-to-end principle. This principle states that the intelligence of an information system rests, that is, in the end-user-devices, not in the network (Saltzer, 1984). Someone who can exercise control as a central authority typically owns the network. In contrast, the "ends" of the network are typically in user's control. The user is able to replace or enhance the ends and to deploy innovative applications and content.

Thus, Lessig (2002) argues, systems that put the end-to-end principle into practice encourage innovation. The short innovation cycles in the internet can largely be explained by the fact that they implement this principle. Smart environments are largely considered to be built out of smart devices and, therefore, follow the end-to-end principle as well. Users of smart environments can decide what device they want to use and what smart services they should deliver. Therefore, the innovation cycles in ambient business are likely to be at least as short as in Electronic Business. Moreover, consideration must be given to the fact that smart services can be invented around all instances of our life that allow the creation of manifold potential applications. New entrants may attend to these potential applications, enter existing markets, and shift competition (Watson, 2002). Companies have to reconsider how to cope with new competitors in a high-velocity market as well as what strategies and business models to apply.

Smart environments and services that are deployed in our everyday life may have considerable social, ethical, and legal implications which also have to be considered in the business perspective. Different studies reveal corresponding potentials and perils. Positive effects can be expected throughout all activities of daily life. In particular, elderly people and handicapped persons may benefit from supportive applications and natural interfaces (Coroama, 2003). These interfaces may also close the digital divide between those with and those without access to computing facilities (BSI, 2006). Hilty et al. (2005) raises the question whether positive effects resulting from efficiency enhancements may be partially undone by side effects, that is, by increasing people's stress level, and whether smart products may lead to pollution and health hazards. However, most concerns relate to the amount of gathered data and resulting privacy issues. History suggests that balancing companies' benefits and customers' privacy is crucial for the success of commercial smart environments (McGinity, 2004). Bohn (2005), Hilty (2005), Bizer (2006), and BSI (2006) provide an in-depth analysis of possible perils and give hints how to balance benefits and user's privacy. Lahlou (2003) proposes a set of privacy guidelines to cope with this challenge.

In addition to these social, ethical, and legal implications, the aim of this chapter is primarily to discuss the impact of open innovation in a world of ambient business from an economic perspective.

Open Innovation from an Economic Perspective

The idea of encouraging innovation throughout openness with regard to IS was widely discussed in the early 1980s when mainframe vendors started to restrict the use of object program and source code of formerly unrestricted software. As a

countermove and in order to gain independence of the software industry, Richard Stallman (1999) propagated the idea of free software. This paradigm shift away from the conventional IP regime resulted in software that can be used, modified, and redistributed by everyone. Related movements like Open Source (Lerner, 2002; O'Reilly, 1999), Open Standards (Perens, 2006; West, 2004), Open Systems (Gabel, 1987), and Open Content (Ciffoilli, 2003; Wiley, 1999) all follow a similar idea: By modifying and recombining existing work, innovation can be collectively achieved. Thus, self-organized ecosystems of Open Innovation emerge when independent actors contribute to a product system (the product and its complementary assets). These Open Innovation ecosystems have proven that they are able to compete with companies that are reluctant to open up their systems and their IP. Software like Apache, Mozilla, and Linux has achieved a considerable market share, and Wikipedia comes close to the quality of the Encyclopaedia Britannica (Wales, 2005).

The idea of openness may contradict the common approach of value extraction of a kind of assets, that is, intellectual assets, in most societies. It is widely assumed that these assets can and should be owned by someone and hence are the IP of the creator or perhaps of the buyer. However, such IP shares the fate of most intangible assets. Once it is revealed and brought into the public domain, there is the risk that it becomes common. The creator and his competitors can use the asset to create value. The potential for the creator to differentiate his products from those of competitors is reduced. Hence, the use of the asset by competitors reduces the creator's return on the asset (Magee, 1977). In order to ensure compensation for the creators of innovative intellectual work, legislators enact laws to protect intellectual property rights (IPR). Creators and buyers of special rights are exclusively allowed to utilize property for a certain period of time. Through these laws, intellectual work should be rewarded and innovation stimulated (Chesbrough, 2006).

The management of intellectual assets often focuses on closing up companies' borders from the intellectual assets point of view in order to keep others from harvesting the fruit of their labor. However, legal instruments often cannot hinder third parties from gaining large shares of the profit arising from innovations: The protection mechanisms are often not enforceable. They are easy to bypass by using slight modifications of the IP in question, and this results in various forms of uncompensated adaptation (Teece, 1986). Under certain conditions, IPR can be ignored. The music and software industries are examples of this phenomenon. Thus, IPR are often not capable of protecting the owner from the exploitation of assets by others.

Closed companies' borders may not only frustrate the aim to protect assets; in fact, they may prevent companies from gaining profit from their assets at all. R&D departments regularly create intellectual assets that cannot be utilized within the company. Procter & Gamble, for example, estimated in 2002 that it only commercialized 10 % of its patents, whereas the remaining 90 % virtually gathered dust on shelves (Sakkab, 2002). Licensing may be a viable option to exploit unutilized patents (Rivette, 2000). However, potential licensees are often reluctant to adopt

external technology (Chesbrough, 2006), and the legal protection of an IP is often too weak to make licensing a viable option.

Companies such as Google and Amazon turn in part away from the closed way of managing intellectual assets and appear to profit from opening up their information systems and, therefore, revealing some of their intellectual assets. As another example, IBM donated 500 patents to the Open Source community in 2005 instead of generating revenue by licensing it. Moreover, IBM spends \$100 million each year in the development of Linux. IBM opens up these assets to regain a prominent role in the operating system market—a position it had lost beforehand with AIX (Chesbrough, 2006)—and is now able to use Linux as a platform to sell products and services around Linux (West, 2003).

Drivers for Open Innovation in Ambient Business

To reveal the drivers for making assets publicly available in a world of ubiquitous computing and the resulting potential benefits, the open approach can be discussed from a resource perspective (Barey, 1991; Penrose 1959; Wernerfelt 1984). This perspective reveals the possibilities arising from a resource exchange at the layers of the device infrastructure, the information infrastructure, and of the smart service.

Resources can be considered as tangible and intangible assets that are tied somehow semi-permanently to the firm (Wernerfelt, 1984). They compass all input factors that are needed to create value during a product life cycle. Companies can leverage the resource base to achieve a competitive advantage if these are valuable, rare, imperfectly imitable, and nonsubstitutable (Barney, 1991). As such, innovative related resources that fulfill these requirements can contribute to achieve an advantageous position in markets in which innovative products and services are valued by customers.

As argued before, in a world of ubiquitous computing, innovation life cycles may be short. From innovations in technologies and services often spring a disruptive force that leads to changes in the market environment: New innovators entering the market and unpredictable demand shifts may change the market structure. However, competitive advantage can only be achieved when the company's resources are in line with market conditions (Eisenhardt, 2000). Yet, in the changing markets of ubiquitous computing, frequent adaptation of the resource base is necessary. On the one hand, this is a major challenge for companies in UC markets. On the other hand, this should be seen as an outstanding opportunity to gain competitive advantage through superior capabilities in resource acquisition, namely adaptation, integration and reconfiguration of internal and external resources (Teece, 1997).

However, companies may not be able to alter their resource base on their own fast enough, and necessary resources may not be available for licensing. For example,

when Microsoft released Windows NT and Windows 95, Apple had to innovate in its operating system to cope with Microsoft. However, Apple was not able to alter its resource base accordingly. It, therefore, published its source code partly as an open source and allowed external programmers to contribute to the software development process (West, 2003). Following this example, opening up a company's border and letting resources and, thus, intellectual assets pass through may be a viable strategy for acquiring new resources.

Open Innovation in Ambient Business Settings

In ambient business, diverse resources can be incorporated to enhance the innovative potential that is associated with a company's innovation processes and the innovation potential of its products. These resources can be obtained from all stakeholders of the company's products, especially customers, suppliers and manufacturers of complementary products, and also from competitors as well. Accordingly, a transfer can be conducted at every level of the proposed framework.

The device infrastructure builds up a smart environment and, hence, provides the foundation for smart services by gathering data and interacting with the user. Access to a device infrastructure is an essential asset for service providers. A tightly restricted and exclusive access to the IT-infrastructure through proprietary standards allows the creation of a monopoly position. This facilitates the monopolist to set rules which rivals must obey (Adams, 1982). Furthermore, the proprietary standards of the infrastructure create a barrier for new entrants because a change to another infrastructure is linked with incompatibility costs. However, providers of complementary devices and services are likely to be discouraged from using a tightly controlled IT-infrastructure (Farrell, 1985; Garud, 2002). Therefore, it may be impossible for a company that is playing the proprietary game to build up the full range of different devices of various application domains. Hence, if a company opens up its device infrastructure and relies on open standards, access to foreign innovative components may be gained by allowing them to become integrated into the company's own infrastructure.

Garud and Kumaraswamy (1993) showed through the Sun Microsystems case how companies could improve their strategic position by implementing an open system strategy. By relying on open standards and by liberally licensing some of its own know-how, Sun attracted capable complementors that helped to improve the overall systems performance. Sun ensured its head start in competition by conducting shorter innovation cycles than its competitors.

Another strategy to ensure returns is to retreat to a certain layers of the product architecture. Layers that do not contribute to a distinctive resource base generally do not help companies to gain a competitive advantage. By opening up these layers, an unrestricted foundation for higher layers may be laid that will be further

developed by external innovators. Thus, the company may concentrate on extending its distinctive resource base on an upper layer, such as developing smart services. Apple, for example, published only source code of the operation system's lower levels. However, the Graphical User Interface and certain APIs, with which Apple can differentiate its operation system from its rivals, remained closed source (West, 2003).

As demonstrated earlier, the information infrastructure is the basis on which services are growing. It is, therefore, an important resource. The richer the information pool in terms of quality and quantity, the more services can be provided. These services will also be more reliable. For that infrastructure, the same principles apply as to the device infrastructure, since both are infrastructure layers. The information infrastructure that is available can be enhanced by opening it up and allowing other information infrastructures to interlink with it. Furthermore, companies may benefit from receding that layer and to commercialize the smart service layer. As distinct from the device infrastructure, when considering the information infrastructure, the user may play a more important role. Sensors deployed in the environment allow gathering data from a user's context without bothering the user, who may here fore willingly contribute this information to a company's information infrastructure assuming he trusts it. In this case, a services provider as well as a user may benefit from highly customized services. The user may play a more active role when it comes to semantically enriched data. Such enrichment can only be automatically conducted in exceptional cases.

Such tasks commonly require human judgment, for instance, to decide whether the Eiffel tower is ugly or beautiful. By opening up the base data, users can be encouraged to fulfill the tasks. Amazon's mechanical Turk is an example for a platform that brings together tasks that need human judgment to be solved and people that are willing to solve such tasks. Persons like students, housewives/ husbands, and pensioners may be attracted by being paid a small amount of money (Amazon, 2007). A voluntary participation, however, may be provoked whenever the user can benefit from an enriched information pool or consider his participation as an act of self-expression. Google Earth is a prime example of users being willing to contribute to a (partly) opened database (Roush, 2005). Other motivation mechanisms exist, such a mechanism, in which the work is disguised as a game (Ahn, 2006).

One of the major challenges of innovation—and therefore of creating smart services—is bringing together need information and solution information. These two types of information cannot usually be found at one and the same site (Hippel, 1994; Thomke, 2002). As a resource, solution information encompasses all information that provides a company with the technological competence and the potential that is necessary to supply customer needs and that resides traditionally in the company's site (Thomke, 2002; Thomke, 2003). Need information encompasses the user's preferences and requirements with respect to new products and services (Reichwald, 2006). It is generally in the user's domain. The innovator's lack of need information

is a critical drawback for creating smart services. Thackara (2001) states that in the current ubiquitous computing research, developers know how to create amazing things with ubiquitous computing technology, but they lack in need information. They do not know how to use their knowledge to create solutions that benefit the user. However, obtaining this information is far from easy. Need information is somewhat “sticky,” that is, it can only be conventionally revealed from potential users through costly and time consuming methods. Users often do not fully understand their needs or are not able to explicate them (Hippel, 1994). Smart environments can help to overcome this problem by studying contextual data gained from the user. Thus, it may be able to observe how a user behaves in using a smart service, which problems he encounters and what needs may remain unfulfilled.

Another way to overcome the problem is suggested by von Hippel (Hippel, 2002, 2005). By outsourcing need-related tasks to users, the problem of how to directly acquire sticky need information can be avoided. Thus, the user handles a task, which the company formally conducted, and returns a solution he is satisfied with. This solution is commonly less sticky than the need information itself. However, to fulfill the task, the user needs relevant solution-information that sticks with the company. Here, the company partly opens up its information pool by providing a design tool, called toolkit, to the user. This toolkit implements the company’s solution information, which is provided in such a manner that the user can handle it easily. Smart Environment may broaden the facility of toolkits. Sophisticated toolkits may be able to be provided to the user at that point where his need generally occurs. Multimodal interfaces (Dam, 2001) may facilitate a richer interaction with the toolkit and, thus, allow solving complex tasks. Tangible interfaces can be used to let the user not only interact by “mediating” devices such as keyboard, mouse, and display, but to directly manipulate physical objects (Sharlin, 2004). An example of such a tangible interface is the “illuminating clay” from Piper et al. (2002). The computing system is able to gather the form of the clay the user works with and is able to augment it with additional information. Recent advancements in 3D-Printers may open up additional opportunities to integrate need information. The project Fab@Home, for example, provides documentation about how to build a 3D-Printer for private purposes (Fab@home, 2006).

Enabling Open Innovation in a World of Ambient Business

As mentioned above, the information infrastructure is an important resource in ambient business settings. Accordingly, opening up ubiquitous computing systems enables innovation in a world of ambient business. If smart devices and smart environments

are built on common standards and are opened up mutually, new possibilities for innovation arise through the combination of information pools, devices, and services. Furthermore, natural, multimodal, and tangible interfaces may simplify remote interaction within Open Innovation ecosystem (e. g., Pinhanez, 2004).

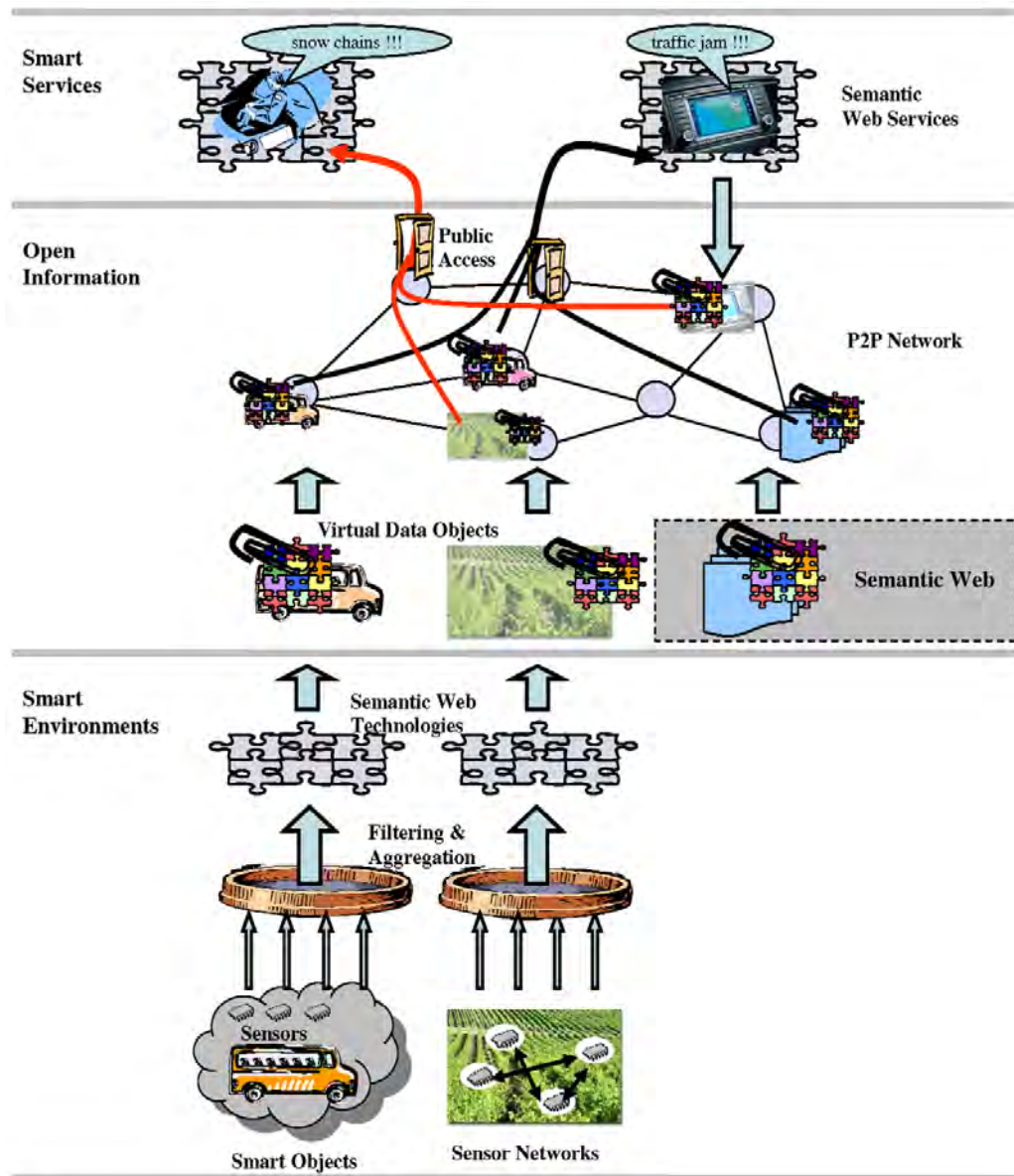
Xerox took the first steps to implementing smart environments in their research project Parctab (Want, 1995). Since then, many more research projects, as well as commercial initiatives, have appeared (for an overview of selected projects, see Endres, 2005). However, these initiatives focus mainly on specific, proprietary applications like smart shopping environments and do not incorporate further prospects of using an open approach (for an overview of smart environments in general, see Cook, 2005). Implementing an open approach as proposed in this chapter means, in essence, granting access to mobile smart objects and local smart environments hosted by different service providers, and to information gathered by them. To enable the reuse of smart environments and information gathered by smart objects, a common understanding of their semantics is essential to the recombination and to the creation of new smart applications and services. Accordingly, we propose the creation of an Open Object Information Infrastructure (OOII) as shown in Figure 2, which complies with the requirements to enable Open Innovation. The various components will be explained below.

Building up Smart Environments

The embedding of computers or smallest processors, memory chips, and sensors into physical objects, referred to as smart objects, constitutes a step toward connecting the real with the virtual world. Other examples are sensor networks—smallest sensors introduced into the environment, forming ad-hoc networks in order to recognize, for example, the whereabouts, speed, size, and form of objects and to observe them over a period of time (Culler, 2004). Sensors may perform an important role in the context of managing object information. This is particularly true for the use of sensors which permit a clear identification of smart objects by remote enquiry. With the assistance of Radio Frequency Identification (RFID), objects are identified automatically without line of sight, by transponders which are attached to them (Finkenzeller, 2003).

It has to be taken into account that the information content of sensor data is rather limited. Hence, the data must be gathered and formed into usable events by means of appropriate filter and aggregation procedures. Numerous projects for the development of suitable middleware have already commenced, but focus primarily on integration of sensor data in closed, business-related applications or communication between smart objects (for an overview of selected projects, see Henrickson, 2006 and Schoch, 2005). Most projects, however, ignore the further use of the information in other contexts or by external developers.

Figure 2. The Open Object Information Infrastructure (OOII)



Ensuring the utilization of gathered data for external developers requires extending existing middleware technologies with mechanisms for the semantic markup of sensor data. Semantic Web standards such as Topic Maps (Durusau, 2006; Garshol, 2006; Pepper, 2001;), or RDF(S)/OWL (Brickley, 2004; McGuinness, 2004) can be employed for this purpose. Several middleware technologies have already been extended with mechanisms for the semantic markup of sensor data. SOAM (Smart Object Awareness and Adaptation Model) (Vazquez, 2006) is one initial approach. SOAM is a smart environment reactivity model in which user preferences are for-

malized using the Standard Ontology for Ubiquitous and Pervasive Applications (SOUPA) (Chen, 2005). These preferences lead to adaptation on smart objects without explicit user intervention, because of the use of automatic, semantic reasoning with respect to the environmental conditions.

An extended version of this ontology is also used by the Context Broker Architecture (CoBrA) (Chen, 2005). This is a broker-centric agent architecture for supporting context-aware systems in smart meeting rooms. The project goal is to create a meeting room that can facilitate typical activities as setting up presentations (allowing users to control services via speech), or adjusting lighting and background music in a room in a manner that is tailored for the specific meeting.

Even if these projects incorporate semantic technologies, they do not consider the reuse of their smart environments within external smart services. Nevertheless, they could be suitable for application within the OOII, as they enable the combination of ubiquitously gathered information with additional semantically-enriched information provided, for example, by the Semantic Web.

Building up Open Information

In developing smart environments, the large number of smart objects must be taken into account. Additionally, the prevailing energy and capacity limitations of embedded computers and sensors make it necessary to save accompanying virtual data objects and to process information on additional network nodes which have sufficient memory capacity and computer power. These prerequisites advise the implementation of peer-to-peer (P2P) networks that offer better scalability and a self-organized and decentralized coordination of unused or limited resources, compared to the alternative client/server architecture (Barkai, 2001; Milojevic, 2002; Oram, 2001; Schoder 2003). Other characteristics, including a higher fault tolerance and a more effective support of spontaneous networking of entities, may be advantageous as well, because no continuous connection can be ensured for smart objects, due to their mobility.

In order to enable Open Innovation, arbitrary access to the stored semantically enriched object information must be ensured for external developers. As the Napster case revealed, central approaches tend to limit access to the information offered or even deny it completely. However, P2P projects like Freenet (Clarke, 2000) and Free Haven (Dingledine, 2000) have demonstrated how an anonymous and censor-free information access can be made possible. Using P2P networks for storing and retrieving semantically enriched information requires an extension of possibilities to pose semantic queries. On the one hand, scientific research is pursuing scheme-based approaches, and, on the other hand, ontologies or P2P Semantic Web services are applied (for an overview of selected projects, see Staab, 2006).

Using semantic query facilities, various developers with different perceptions can search for relevant (object) information, combine the information freely, and, hence, create innovative services (as indicated in Figure 2). For example, information about the speed of vehicles (low speed for a number of cars) combined with their actual location (they are all in a certain area) and additional information (a road map) might indicate that they are stuck in a traffic jam on a highway. This information could then be provided to others as a traffic-information service. In order to facilitate even the reuse of this new smart service, the service has to be specified semantically and published in the OOIL. Subsequently, other developers can reuse this service (and the connected smart environments and objects), combine services with information gathered by another local sensor network (the outdoor temperature in a certain area is below zero), and set up a warning system (you are traveling to a snowy area—do not forget your snow chains).

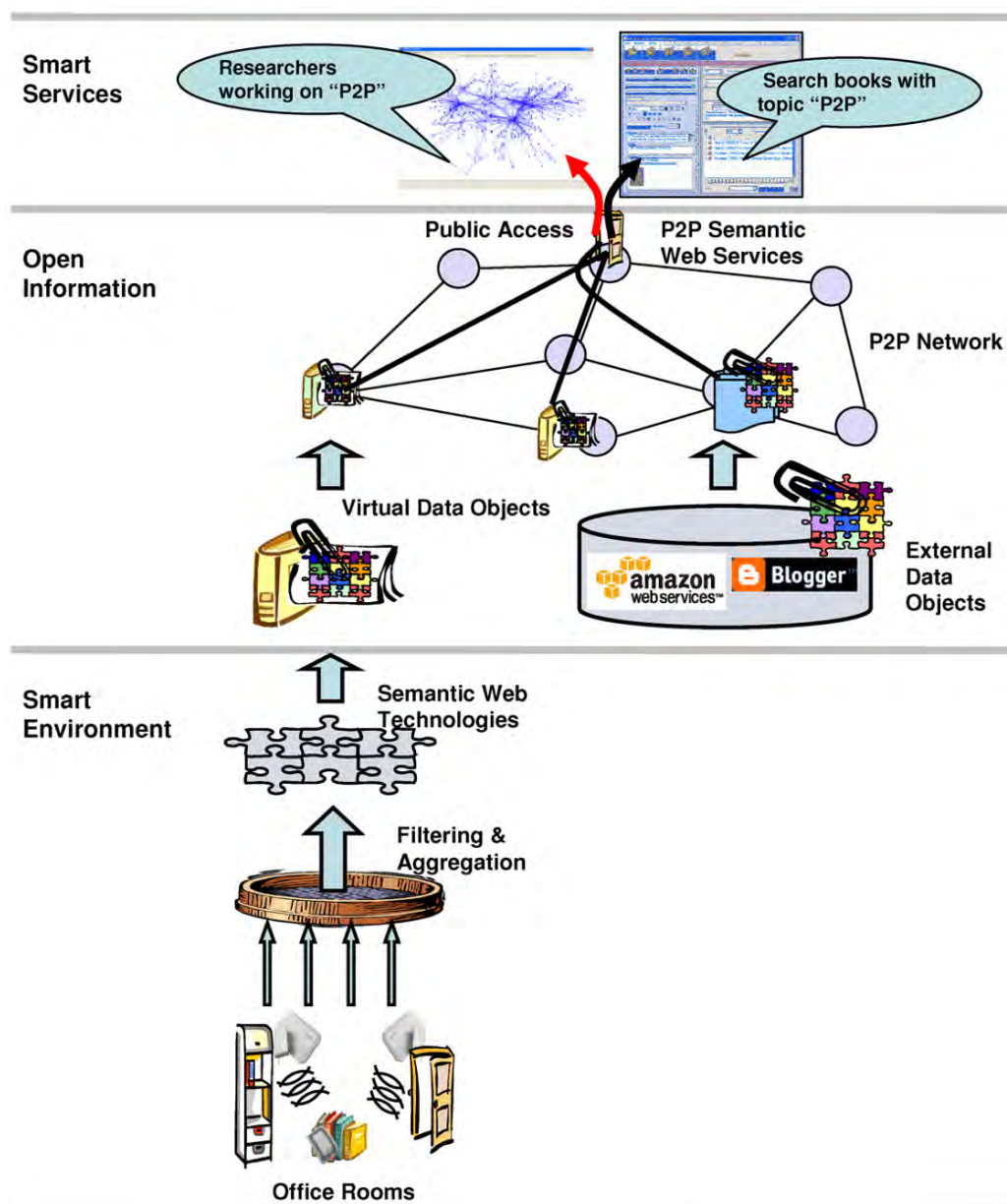
In order to discover and integrate semantically enriched information, a number of semantic web services (Alesso, 2004; Fensel, 2002) and smart agents (Konnerth, 2006; Lee, 2005;) have been developed. However, these projects mainly solve negotiation problems with respect to a common understanding of certain application contexts. The user-oriented framework of Task Computing (Masuoka, 2004) is a promising approach in the context of ubiquitous computing. It uses the Task Computing Environment Object Ontology (TCE, 2004) and is designed to operate dynamic ubiquitous environments in which a mobile computing user dynamically discovers the current set of available semantically defined services (Song, 2004). The end-user can easily and seamlessly integrate and manipulate services found on their own computer, the nearby devices and relevant remote Web services. However, Ben Mokhtar et al. (2006) claim that even this opportunity, or need to select the right composition among all the possible compositions suggested to the user, presents a major drawback of this framework. Accordingly, they developed a COntversation-based service COmposition middleware (COCO) that minimizes the necessary user interaction by enabling the automatic and transparent deployment and execution of user tasks.

Nevertheless, these approaches are limited to the discovery of smart services within smart environments in order to improve user interaction. They do not address global service discovery as needed within our OOIL. METEO-S WDSI (Verma, 2005) and GloServ (Arabshian, 2005) are approaches that perform global service discovery through the use of P2P networks. METEO-S WDSI presents a scalable environment for Web service publication and discovery among multiple registries in general. It uses an ontology-based approach to organize registries into domains and enable domain based classification of all Web services. Each of these registries supports semantic publication and the discovery of Web services. GloServ is a global service discovery architecture that can be used to build a distributed context-aware service discovery system related to pervasive computing scenarios (Arabshian, 2006). Currently, we examine which one of these approaches would be applicable within the OOIL.

Building Up Smart Services

Based on Open Information, innovative services can be developed. The Federative Library (FedLib) represents one first instantiation of our OOII (see Figure 3) and demonstrates the feasibility of the design. The FedLib is a library service for work groups that functions without the need for collecting books at a centralized location. In fact, the collection is held by the participants themselves. Book owners decide

Figure 3. The Federative Library (FedLib)



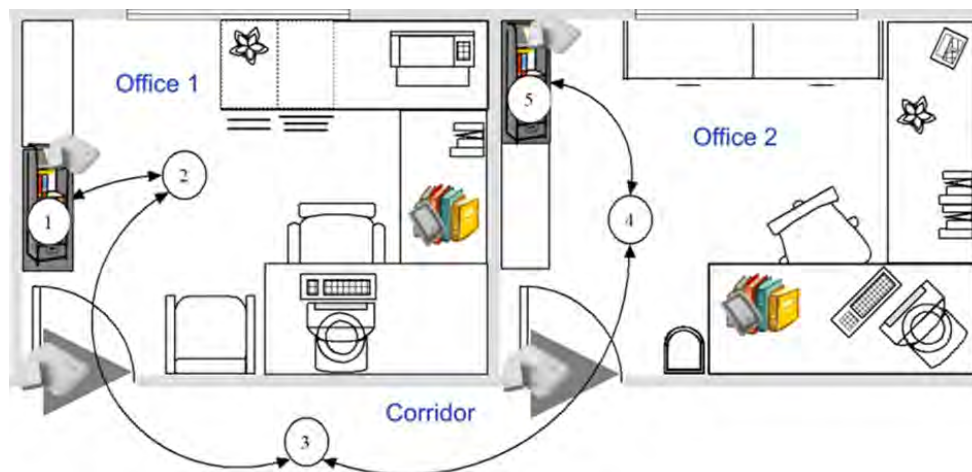
about making their books available for use by others. Thus, originally exclusive private collections of books that are common in organizations could be opened up to other would-be readers.

The pilot implementation of the FedLib will be realized within the Department of Information Systems and Information Management and includes our own office rooms and books, as well as books from the faculty library. We already built up the necessary smart environment and started to extend a reference information manager to the corresponding search-client in order to integrate the smart service into the daily workflow of a research assistant. Besides building up the smart environment, we focus on reviewing existing technology for further use within the FedLib and OOII.

Implementing the smart environment of the FedLib, bookshelves and doors of our office rooms are equipped with RFID readers (see Figure 4). Books labeled with RFID tags are distributed in the participants' offices. In order to enable ubiquitous information gathering about the presence of books, the detection of a book's current location must be combined with former "reads" to corresponding events. For instance, when the system detects a book at the door reader in Office 1, information must be available as to whether the book is leaving or entering the office. It is leaving if the book was previously detected on the bookshelf in Office 1. It is entering if the book was previously located at the door reader in Office 2.

By aggregating these semantically enriched events with additional information about the occupant of each office, the FedLib can track and allocate books automatically in terms of the current borrower. This minimizes administrative overhead, due to the ubiquitous integration of processes in the daily workflow of the research assistants. It therefore becomes possible to borrow books by collecting them directly from a colleague's shelf, taking them to one's own office, and placing them at the table

Figure 4. Smart Office Environment for the FedLib



Copyright © 2008, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

there. In order to return a book, it can simply be put back on one's own bookshelf. During that natural interaction with the Federative Library, the location and status of the book changes according to the events occurring (numbered 1 to 5 in Figure 4). In order to integrate even the management and the retrieval of books into the daily work of a research assistant, the corresponding search-client is integrated in a reference information manager. Furthermore, services like Amazon's Web Service (Levitt, 2005) and special semantically-annotated resources produced by Semblogging (Cayzer, 2004), are suitable for integration into the manager in order to provide additional useful information.

Conclusion

Previous research and the movement towards openness suggest that an opening of information systems may reveal new contingencies for innovation. In order to analyze these contingencies in a world of ubiquitous computing, we applied the framework of ambient business. We showed that opening up smart environments as well as smart services, reveals potentials for innovative services that can be created by external developers. Thus, the innovative strength of companies that act in open systems cannot be achieved by those that rely on rather proprietary information systems. However, most contemporary projects ignore the further use of the information in other contexts or by external developers.

By designing the Open Object Information Infrastructure, we showed how to open up ubiquitous computing systems. Within this infrastructure, information pools that emerge from smart devices are connected by existing Semantic Web standards. This allows developers to search for and combine information they need, create innovative services, and reuse and combine them in an innovative manner. By introducing the first instantiation of the OOII, we demonstrated the feasibility of our design. The Federated Library combines information from decentralized objects and external services to an innovative library services that cannot be realized in closed environments.

However, the open approach is not only applicable from the technological point of view. Economic rationality can argue similarly in favor of opening information systems and thus companies' intangible assets. In some instances, property rights cannot fulfill their aim of fostering innovation, but in fact encumber a further development that builds up on the protected assets. Therefore, in terms of economic and social welfare, the guarantee to freely use innovation as well as the incentive to commercialize innovations has to be perfectly balanced.

Additionally, a cross-border flow of intellectual resources can be beneficial. In high-velocity markets as can be expected in ambient business, the alignment of

company resources to the corresponding market becomes critical. In particular, innovation-related resources have to cope with market developments. Opening up the company's borders can be a strategy for acquiring external innovation-related resources and, therefore, a head start in competition with closed innovators.

Future Research Directions

Questions arise regarding the willingness of market players to adopt the idea of Open Innovation. First, open smart environments may infringe privacy. Consequently, we may ask what information a user considers private and in what situations. What information should, therefore, *not* be freely available to service providers? How should laws be formulated in a world in which information is freely exchanged? The first studies have already appeared which analyze the dangers associated with the emergence of ubiquitous computing (Hilty, 2005; Bizer, 2006; BSI, 2006), but they are only the first steps towards answering the core questions.

Likewise, companies must consider which resources and information should be allowed to cross their borders. Sophisticated strategies have to be implemented to ensure the appropriability and, therefore, the return on the innovations. Thus, a manager may ask what products and resources should be opened and to what extent. What are the company's distinct resources that have to be maintained inside the company? What kind of resources can be utilized more effectively outside? Control of the resource transfer must be established.

Furthermore, organizational structures and processes have to be adapted to facilitate the adaptation of external resources, and employees have to be trained to interact with external stakeholders. Moreover, external innovators have to be attracted. When increasingly more companies try to harness users' resources and creative power, how will the users react? Will their attention and time become a limited resource so that companies have to compete for it?

References

- Aarts, E., Harwig, R., & Schuurmans, M. (2001). Ambient intelligence. In P. Dennin (Ed.), *The invisible future-the seamless integration of technology in everyday life* (pp. 235-250). New York: McGraw-Hill.
- Adams, W., & Brock, J. W. (1982). Integrated monopoly and market power: System selling, compatibility standards, and market control. *Quarterly Review of Economics and Business*, 22(4), 29-42.

- Alesso, H., & Smith, C. (2004). *Developing Semantic Web Services*. Natick, MA.: Peters.
- Amazon (2007). *Mechanical Turk*. Retrieved on April 10, 2007 from <http://www.mturk.com/mturk/help?helpPage=whatis>, 2007
- Angeles, R. (2005). RFID Technologies: Supply-Chain Applications and Implementation Issues. *Information Systems Management*, 22 (1), 51-65.
- Anoto (2007). *Anoto Group AB, Lund, Sweden*. Retrieved on April 10, 2007 from <http://www.anoto.com/>
- Arabshian, K., Schulzrinne, H., Trossen, D., & Pavel, D. (2005, June). *Global Service Discovery using the OWL Web Ontology Language*. Paper presented at the International Workshop on Intelligent Environments, Colchester, England.
- Arabshian, K., & Schulzrinne, H. (2006). Distributed Context-aware Agent Architecture for Global Service Discovery. In *Proceedings of the 2nd International Workshop on the use of Semantic Web Technologies for Ubiquitous and Mobile Applications* (pp. 17-19).
- Barkai, D. (2001). *Peer-to-peer computing. technologies for sharing and collaboration on the net*. (Technical Rep.). Hillsboro: Intel Press.
- Barney, J. (1991). Firm Resources and Sustained Competitive Advantage. *Journal of Management*, 17 (3), 99-120.
- Berners-Lee, T., Hendler, J., & Lassila, O.: The Semantic Web. *Scientific American*, 284(5), 34-43.
- Bizer, J., Dingel, K., Fabian, B., Günther, O., Hansen, M., Klafft, M., Möller, J., & Spiekemann, S. (2006). *TAUCIS: Technikfolgen-Abschätzung Ubiquitäres Computing und Informationelle Selbstbestimmung*. Retrieved on April 10, 2007 from https://www.datenschutzzentrum.de/taucis/ita_taucis.pdf
- Bohn, J., Coroama, V., Langheinrich, M., Mattern, F., & Rohs, M. (2005). Social, Economic, and Ethical Implications of Ambient Intelligence and Ubiquitous Computing. In W. Weber, J. M. Rabaey, & E. Aarts (Ed.), *Ambient Intelligence* (pp. 5-29). Berlin, Heidelberg : Springer.
- Brickley, D., & Guha, R. (2004). RDF Vocabulary Description Language 1.0: RDF Schema, Retrieved on April 10, 2007 from <http://www.w3.org/TR/rdf-schema/> 2004
- BSI, Federal Office for Information Security (2006). Pervasive Computing: Trends and Impacts. Retrieved on April 10, 2007 from http://www.bsi.bund.de/literat/studien/percenta/Percenta_elay.pdf
- Cayzer, S. (2004). Semantic blogging and decentralized knowledge management. *Communications of the ACM*, 47(12), 47-52.
- Chen, H., Finin, T., & Joshi, A. (2005). The SOUPA Ontology for Pervasive Computing. In T. Valentina (Ed.), *Ontologies for Agents: Theory and Experiences*,

- Whitestein Series in Software Agent Technologies (pp. 233-258). Basel: Birkhäuser.
- Chesbrough, H. (2006). *Open Business Models: How to Thrive in the New Innovation Landscape*. Boston, MA: Harvard Business School.
- Ciffolilli, A. (2003). Phantom authority, self-selective recruitment and retention of members in virtual communities: The case of Wikipedia. *First Monday*, 8(12). Retrieved on April 10, 2007 from http://www.firstmonday.org/issues/issue8_12/ciffolilli/index.html, 2003
- Clarke, E., Sandberg, O., Wiley, B., & Hong, T. (2000). Freenet: A distributed anonymous information storage and retrieval system. In H. Federrath (Ed.), *Proceedings of the ICSI Workshop on Design Issues in Anonymity and Unobservability*. Berlin: Springer.
- Cook, D. J., & Das, S. K. (2005). *Smart Environments: Technology, Protocols, and Applications*. Hoboken, NJ: Wiley.
- Coutaz, J., Crowley, J. L., Dobson, S., & Garlan, D. (2005). Context is Key. *Communications of the ACM*, 48 (3), 49-53.
- Coroama, V. et al. (2003). Leben in einer smarten Umgebung: Ubiquitous-Computing-Szenarien und –Auswirkungen. (Technical Rep. No. 431). *ETH Zurich*. Retrieved on April 10, 2007 from <http://www.vs.inf.ethz.ch/publ/papers/szenarien-FINAL.pdf>
- Culler, D., Estrin, D., & Srivastava, M. (2004). Overview of sensor networks. *Computer*, 37(8), 41-49.
- Dingledine, R., Freedman, M., & Molnar, D. (2000). The free haven project: Distributed anonymous storage service. In H. Federrath (Ed.), *Proceedings of the ICSI Workshop on Design Issues in Anonymity and Unobservability*. Berlin: Springer.
- Durusau, P., Newcomb, S., & Barta, R. (2006). Topic Maps Reference Model, ISO 13250-5. Retrieved on April 10, 2007 from <http://www.isotopicmaps.org/tmrm/>
- Eisenhardt, K., & Jeffrey, M. A. (2000). Dynamic Capabilities: What are they? *Strategic Management Journal*, 21 (10/11), 1105-1121.
- Endres, C., Butz, A., & MacWilliams, A. (2005). A Survey of Software Infrastructures and Frameworks for Ubiquitous Computing. *Mobile Information Systems Journal*, 1(1), 41-80.
- Fab@Home (2007). *Fab@Home*. Retrieved on April 10, 2007 from <http://www.fabathome.org/>
- Fano, A. & Gersham, A. (2002). The Future of Business Services in the Age of Ubiquitous Computing. *Communications of the ACM*, 45 (12), 83-87.

- Farrell, J., & Saloner, G. (1985). Standardization, compatibility, and innovation. *RAND Journal of Economics*, 16 (1), 70-83.
- Fensel, D. (2002). Semantic web enabled web services. In I. Horrocks, J. A. Hendler (Ed.), *Proceedings of the First International Semantic Web Conference on The Semantic Web* (pp. 1-2). London: Springer.
- Finkenzeller, K. (2003). *RFID-Handbook*. Chichester, West Sussex : Wiley.
- Fleisch, E., & Mattern, F. (2005). *Das Internet der Dinge Ubiquitous Computing und RFID in der Praxis*. Berlin: Springer.
- Fleisch, E., & Tellkamp, C. (2006). The business value of ubiquitous computing technologies. In G. Roussos (Ed.), *Ubiquitous and Pervasive Computing New Frontiers for Electronic Business* (pp. 93-114). London: Springer.
- Floerkemeier, C., Lampe, M., & Schoch, T. (2003). The Smart Box Concept for Ubiquitous Computing Environments. In *Proceedings of sOc'2003*, Grenoble, (pp. 118-121),
- Gabel, H. L. (1987). Open Standards In Computers. In H. L. Gabel (Ed.), *Product Standardization And Competitive Strategy* (pp. 91-123). Amsterdam: Elsevier.
- Garshol, L., & Moore, G. (2006). Topic Maps-Data Model. Retrieved on April 10, 2007 from <http://www.isotopicmaps.org/sam/sam-model/>
- Garud, R., & Kumaraswamy, A. (1993). Changing Competitive Dynamics in Network Industries: An exploration of Sun Microsystems' Open Systems Strategy. *Strategic Management Journal*, 14 (5), 351-369.
- Garud, R., Jain, S., & Kumaraswamy, A. (2002). Institutional entrepreneurship in the sponsorship of common technological standards: The case of Sun Microsystems and Java. *Academy of Management Journal*, 45(1), 196-274.
- Gellersen, H., Beigl, M. & Krull, H. (1999). The MediaCup: Awareness Technology embedded in an Everyday Object. In H.-W. Gellersen (Ed.), *Proceedings of the Handheld and ubiquitous computing first international symposium, (HUC'99)* (pp. 308-310). Heidelberg: Springer.
- Gershenfeld, N. (1999). *When Things Start to Think*. London: Hodder & Stoughton.
- Gibson, R., & Schuyler, E. (2006). *Google Maps Hacks*. Sebastopol, CA: O'Reilly.
- Hansmann, U. (2003). *Pervasive Computing*. Berlin: Springer.
- Henrickson, K., Robinson, R. (in press). A survey of Middleware for Sensor Networks: State-of-the-art and future Directions. *Proceedings of the International Workshop on Middleware for Sensor Networks*.

- Hilty, L., Behrendt, S., Binswanger, M., Bruinink, A., Erdmann, L., Fröhlich, J., Köhler, A., Kuster, N., Som, C., & Würtenberger, F. (2005). *The Precautionary Principle in the Information Society*. Retrieved on April 10, 2007 from http://www.ta-swiss.ch/a/info_perv/2005_46e_pervasingcomputing_e.pdf
- Innocentive (2007). *Innocentive Inc., Andover, USA*. Retrieved on April 10, 2007 from <http://www.innocentive.com>
- Konnerth, T., Hirsch, B., & Albayrak, S. (2006). Jatl: An agent description language for smart agents. In *Proceedings of Declarative Agent Languages and Technologies (DALTO6)*.
- Lahlou, S., & Jegou, F. (2003). *European Disappearing Computer Privacy Design Guidelines V1.0. Ambient Agoras Report D15.4. Disappearing Computer Initiative*. Retrieved on April 10, 2007, from http://www.ambient-agoras.org/downloads/D15%5B1%5D.4_-_Privacy_Design_Guidelines.pdf
- Lee, W. (2005). An intelligent agent for rfid-based home network system. In Yang, L. (Ed.), *EUC 2005, LNCS 3824* (pp. 499-508), Berlin: Springer.
- Lerner, J., & Tirole, J. (2002). Some Simple Economics of Open Source. *The Journal of Industrial Economics*, 52 (2), 197-234.
- Lessig, L. (2002). The Architecture of Innovation. *Duke Law Journal*, 51(6), 1783-1801.
- Levitt, J. (2005). *The Web Developer's Guide to Amazon E-Commerce Service: Developing Web Applications Using Amazon Web Services*. Lulu Press.
- Maass, W., & Filler, A. (2006). Towards an Infrastructure for Semantically Annotated Physical Products. In C. Hochberger, R. Liskowsky (Ed.), *Proceedings Informatik 2006* (pp. 544). Bonn: Gesellschaft für Informatik.
- McGinity, M. (2004). RFID: Is This Game of Tag Fair Play? *Communications of the ACM*, 47(1), 15-18.
- Magee, S. P. (1977). Multinational Corporations, the Industry Technology Cycle and Development. *Journal of World Trade Law*, 11, 287-321.
- Masuoka, R., Labrou, Y., & Song, Z. (2004). Semantic Web and Ubiquitous Computing-Task Computing as an Example. *AIS SIGSEMIS Bulletin*, 1(3), 21-24.
- Mattern, F. (2005). Ubiquitous computing: Scenarios from an informatised world. In A. Axel Zerdick, A. Picot, K. Schrape, J. Burgelman, R. Silverstone, V. Feldmann, C. Wernick, & C. Wolff (Ed.), *E-Merging Media-Communication and the Media Economy of the Future* (pp. 145-163). Berlin: Springer.
- McGuinness, D., & van Harmelen, F. (Ed.) *OWL Web Ontology Language Overview. W3C Recommendation*. Retrieved on April 10, 2007 from <http://www.w3.org/TR/owl-features/>

- Milojicic, D., Kalogeraki, V., Lukose, R., Nagaraja, K., Pruyne, J., Richard, B., Rollins, S., & Xu, Z. (2002) *Peer-to-peer computing*. (Technical Rep), HP Laboratories Palo Alto.
- Mokhtar, S. B., Georgantas, N., & Issarny, V. (2006, June). *COCOA: COncussion-based Service COmposition in Pervasive Computing Environments*. Paper presented at the IEEE International Conference on Pervasive Services (ICPS'06), Lyon, France.
- Norman, A. D. (1998). *The Invisible Computer*. Cambridge, MA: MIT Press.
- Oram, A. (2001) *Peer-to-Peer: Harnessing the Benefits of a Disruptive Technology*. Sebastopol: O'Reilly.
- O'Reilly, T. (1999). Lessons from open-source software development. *Communications of the ACM*, 42 (4), 32-37.
- Penrose, E. T. (1959). *The Theory of the growth of the firm*. Oxford: Blackwell.
- Pepper, S., & Moore, G. (2001) *XML Topic Maps (XTM) 1.0*. Retrieved on April 10, 2007 from <http://www.topicmaps.org/xtm/>
- Perens, B. (2006). *Open Standards Principles and Practice*. Retrieved on April 10, 2007 from <http://perens.com/OpenStandards/Definition.html>
- Pinhanez, C., & Pingali, G. (2004). Projector-camera systems for telepresence. In G. Pingali, F. Nack, H. Sundaram (Ed.), *Proceedings of the 2004 ACM SIGMM workshop on Effective telepresence* (pp. 63-66). New York, NY: ACM Press.
- Piper, B., Ratti, C., & Ishii, H. (2002). Hands-On Interfaces: Illuminating Clay: a 3-D tangible interface for landscape analysis. In Wixon, D. (Ed.), *Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves* (pp. 355-363). New York, NY: ACM Press.
- Reichwald, R., & Piller, F. (2006). *Interaktive Wertschöpfung: Open Innovation, Individualisierung und neue Formen der Arbeitsteilung*. Wiesbaden: Gabler.
- Rivette, K. G., & Kline, D. (2000). *Rembrandts in the attic*. Boston, MA.: Harvard Business School Press.
- Roush, W. (2005). Killer Maps. *Technology Review*, 108 (10), 54-60.
- Roussos, G. (2006). Ubiquitous computing for electronic business. In G. Roussos (Ed.), *Ubiquitous and Pervasive Computing: New Frontiers for Electronic Business* (pp. 1-12). London, Springer, 2006.
- Saltzer, J. H., Reed, D. P., & Clark, D. D. (1984). End-to-end arguments in system design. *ACM Transactions on Computer Systems*, 2 (4), 277-288.
- Sakkab, N. Y. (2002). Connect & Develop Complements Research & Development at P&G. *Research-Technology Management*, 45(2), 38-45.

- Schoch, T. (2005). Middleware für Ubiquitous-Computing-Anwendungen. In E. Fleisch, F. Mattern (Ed.), *Das Internet der Dinge Ubiquitous Computing und RFID in der Praxis* (pp. 119-140). Berlin: Springer.
- Schoder, D., & Vollmann, C. (2003). Value Scope Management-Beherrschung der Wertschöpfungskette im M-Commerce am Beispiel "i-mode". *Mobile Commerce : Gewinnpotenziale einer stillen Revolution* (pp. 125-143). Berlin: Springer.
- Sharlin, E., Watson, B., Kitamura, Y., Kishino, F., & Itoh, Y (2004). On tangible user interfaces, humans and spatiality. *Personal and Ubiquitous Computing*, 8(5), 338-346.
- Song, Z., Labrou, Y., & Masuoka, R. (2004). Dynamic Service Discovery and Management in Task Computing. In *MobiQuitous 2004* (pp. 310-318). Boston: IEEE.
- Staab, S., & Stuckenschmidt, H. (2006). *Semantic Web and Peer-to-Peer*. Berlin: Springer.
- Stallman, R. (1999). The GNU Operating System and the Free Software Movement. In C. Di Bona (Ed.), *Open sources: voices from the open source revolution*. Sebastapol, CA: O'Reilly.
- Strasser, M., Plenge, C., & Stroh, S. (2005). Potenziale der RFID-Technologie für das Supply Chain Management in der Automobilindustrie. In E. Fleisch, F. Mattern (Ed.), *Das Internet der Dinge Ubiquitous Computing und RFID in der Praxis* (pp. 177-196). Berlin: Springer.
- Thackara, J. (2001). The design challenge of pervasive computing. *Interactions*, 8(3), 46-52.
- TCE, Task Computing Environment Object Ontology (2004). Retrieved on April 10, 2007 from <http://www.facp.fujitsulabs.com/tce/ontologies/2004/03/object.owl>
- Teece, D. J. (1986). Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Research Policy*, 15 (6), 285-305.
- Teece, D. J., Pisan, G., & Shuen, A. (1997). Dynamic Capabilities and Strategic Management. *Strategic Management Journal*, 18 (7), 509-533.
- Tennenhouse, D. (2000). Proactive computing. *Communications of the ACM* 43(5), 43-50.
- Thomke, S., & von Hippel, E. (2002). Customers as Innovators: A New Way to Create Value. *Harvard Business Review*, 80 (4), 74-81.
- Thomke, S. (2003). *Experimentation Matters: Unlocking the Potential of New Technologies for Innovation*. Boston, MA: Harvard Business School Press.

- Van Dam, A. (2001). User interfaces: Disappearing, dissolving, and evolving. *Communications of the ACM*, 44(3), 50-52.
- Vazquez, J. Delpiña, D. & Sedans, I. (2006). SoaM: An environment adaption model for pervasive Semantic Web. (LNCS 3983, 108-117.
- Verma, K., Sivashanmugam, K., Sheth, A., Patil, A., Oundhakar, S., & Miller, J. (2005). METEOR-S WSDI: A Scalable Infrastructure of Registries for Semantic Publication and Discovery of Web Services. *Journal of Information Technology and Management, Special Issue on Universal Global Integration*, 6(1), 17-39.
- Von Ahn, L. (2006). Games with a Purpose. *IEEE Computer*, 39(6), 92-94.
- Von Hippel, E. (1994). "Sticky Information" and the Locus of Problem Solving: Implications for Innovation. *Management Science*, 40(4), 429-439.
- Von Hippel, E., & Katz, R. (2002). Shifting Innovation to Users via Toolkits. *Management Science*, 48(7), 821-833.
- Von Hippel, E. (2005). *Democratizing Innovation*. Cambridge, MA: MIT Press.
- Wales, J. (2005). Internet encyclopaedias go head to head. *Nature* 438(15), 900-901.
- Want, R., Schilit, B., Adams, N., Gold, R., Petersen, K., Ellis, J., Goldberg, D., & Weiser, M. (1995). *The PARCTAB ubiquitous computing experiment* (Technical Rep. No. CSL-95-1). Palo Alto, CA: Xerox Palo Alto Research Center.
- Want, R., Borriello, G., Pering, T., & Farkas, K. I. (2002). Disappearing Hardware. *IEEE Pervasive Computing* 1(1), 36-47.
- Watson, R. T., Pitt, L. F., Berthon, P., & Zinkhan, G. M. (2002). U-Commerce: Expanding the Universe of Marketing. *Journal of the Academy of Marketing Science*, 30(4), 333-347.
- Weiser, M. (1991). The computer of the 21st century. *Scientific American*, 256(3), 94-104.
- Wernerfelt, B. (1984). A Resource-based View of the Firm. *Strategic Management Journal*, 5(2), 171-180.
- West, J. (2003). How open is open enough? Melding proprietary and open source platform strategies. *Research Policy*, 32(7), 1259-1285.
- West, J. (2004, May). *What are Open Standards? Implications for Adoption, Competition and Policy*. Paper presented at the Standards and Public Policy Conference, Chicago, Illinois.
- Wiley, D (1999). Open Content. Open Publication License. Retrieved on April 10, 2007 from <http://www.opencontent.org/openpub/>