

DECLARATION OF NATHANIEL E FRANK-WHITE

1. I am a Records Request Processor at the Internet Archive. I make this declaration of my own personal knowledge.
2. The Internet Archive is a website that provides access to a digital library of Internet sites and other cultural artifacts in digital form. Like a paper library, we provide free access to researchers, historians, scholars, and the general public. The Internet Archive has partnered with and receives support from various institutions, including the Library of Congress.
3. The Internet Archive has created a service known as the Wayback Machine. The Wayback Machine makes it possible to browse more than 450 billion pages stored in the Internet Archive's web archive. Visitors to the Wayback Machine can search archives by URL (i.e., a website address). If archived records for a URL are available, the visitor will be presented with a display of available dates. The visitor may select one of those dates, and begin browsing an archived version of the Web. Links on archived files in the Wayback Machine point to other archived files (whether HTML pages or other file types), if any are found for the URL indicated by a given link. For instance, the Wayback Machine is designed such that when a visitor clicks on a hyperlink on an archived page that points to another URL, the visitor will be served the archived file found for the hyperlink's URL with the closest available date to the initial file containing the hyperlink.
4. The archived data made viewable and browseable by the Wayback Machine is obtained by use of web archiving software that automatically stores copies of files available via the Internet, each file preserved as it existed at a particular point in time.
5. The Internet Archive assigns a URL on its site to the archived files in the format `http://web.archive.org/web/[Year in yyyy][Month in mm][Day in dd][Time code in hh:mm:ss]/[Archived URL]` aka an "extended URL". Thus, the extended URL `http://web.archive.org/web/19970126045828/http://www.archive.org/` would be the URL for the record of the Internet Archive home page HTML file (`http://www.archive.org/`) archived on January 26, 1997 at 4:58 a.m. and 28 seconds (1997/01/26 at 04:58:28). The date indicated by an extended URL applies to a preserved instance of a file for a given URL, but not necessarily to any other files linked therein. Thus, in the case of a page constituted by a primary HTML file and other separate files (e.g., files with images, audio, multimedia, design elements, or other embedded content) linked within that primary HTML file, the primary HTML file and the other files will each have their own respective extended URLs and may not have been archived on the same dates.
6. Attached hereto as Exhibit A are true and accurate copies of browser screenshots of the Internet Archive's records of the archived files for the URLs and the dates specified in the attached coversheet of each printout.

7. Attached hereto as Exhibit B are true and accurate copies of the Internet Archive's records of the archived files for the URLs and the dates specified in the following table.

Folder	URL
001	https://web.archive.org/web/20060504013848/http://www.research.ibm.com/ed/movies/SIGGRAPH01.avi
002	https://web.archive.org/web/20051218051417/http://www.research.ibm.com/ed/movies/siggraph%20bluespace.mov

8. Attached hereto as Exhibit C are true and accurate copies of the Internet Archive's record of the archived files for the URLs and the dates specified in the attached coversheet of each file
9. I declare under penalty of perjury under the laws of the United States of America that the foregoing is true and correct.

DATE: August 30, 2023

Nathaniel E Frank-White
Nathaniel E Frank-White

EXHIBIT A

<https://web.archive.org/web/20041013055749/http://www.time.com/time/covers/1101041011/gadgets/6.html>

PHOTO ESSAY | TECHNOLOGY

<< Previous | Next >>



TED THAI FOR TIME

Everywhere Display

WHO: IBM

WHY: Touch-screen kiosks have taken off, but it's not always easy to find space to install them. This invention removes the physical screen so that interactive images can be cast onto almost any surface.

HOW IT WORKS: A wall-mounted projector, paired with a camera, shows images on the floor or wall. The camera tracks people's interactions with the visuals, and the system interprets them the way it would a moving mouse or a tap on a screen.

APPLICATIONS: Retailers could lure shoppers with interactive window displays. Airports might offer visually aided instructions and directions throughout terminals. In a lab or office, collaborators might use it like a high-tech dry-erase board that displays Excel spreadsheets and other interactive material.

HOW MUCH: The display won't be sold in stores and will be priced according to its corporate customers' specifications.

<< Previous

NEXT >>

MORE PHOTOS



There's Music In My Glasses

Here's a peek at the coolest gadgets hitting the market in the coming months



Tomorrow's Sports Talent

Meet some of the young athletes who may have what it takes

MORE STORIES: [Forward Thinking](#) | [Next Chapter](#) | [Marketing](#)

https://web.archive.org/web/20040607145440/http://www.research.ibm.com/ed/ed_siggraph01.htm



IBM Research

Research home | IBM home | Order | Privacy | Legal | Contact IBM
Everywhere Displays Home | Technology | Applications | Publications | People | Events/Media

Everywhere Displays at SIGGRAPH'01

The exhibition

The [Emerging Technologies Exhibition](#) at SIGGRAPH is a juried, 6-day exhibition of the most innovative and advanced technologies in computer graphics and interaction technology.

The [Everywhere Displays project](#) was selected to be [presented](#) at the ET exhibition of SIGGRAPH'01, held in Los Angeles, California, August 12-17, 2001



The concept

Among the various [applications](#) envisioned for the ED-projector, we are particularly interested in augmented reality (AR) scenarios. Unlike traditional AR, the use of ED-projectors enable user visualization of information without encumbering goggles and interaction without contact with input devices. A typical AR application is the augmentation of an assembly task. In our demo at SIGGRAPH'01, assembly instructions are superimposed on the objects to be assembled. Each individual user contributes to the assembly of an object by executing specific parts of the assembly process, without knowledge of the complete structure of the assembled object.

Based on these premises, we conceived and implemented an augmented reality assembly environment to be experienced by the visitors of [SIGGRAPH'01](#). Given the characteristics and interests of our audience, we chose the object to be assembled to be a picture made of [M&M's](#) (multi-colored sugar-coated chocolates) where each M&M; is regarded as a "pixel" of the picture. Our experiment built pictures of 60x50 [M&M's](#); "pixels" based on portraits painted by Van Gogh. Each picture used a palette of 18 colors of the 21 colors available from [M&M ColorWorks](#). Since each visitor placed between 1 to 15 M&M's, in average 200 people were needed to complete a picture.



The setup

In the front of the 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](#). On the back walls of the space we shelved unlabeled buckets, each containing one of the 18 different colors of [M&M's](#) used. Two other surfaces are also used for projection: a painter's palette mounted on an easel and a wood board first covered with 1 inch-thick foam and then topped with white fabric. The ED-projector is positioned directly above the Plexiglas board. **No sensors of any kind are embedded on any surface of the environment. The interaction is always detected only by the camera.**



The 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". The goal of this image is simply to familiarize the user with the idea of touching



The ED-projector is then steered towards the foam-covered-board where images of different colored [M&M's](#) are projected along with the invitation "Pick a color". Here, the user can select which color to

a surface in order to interact with the system.

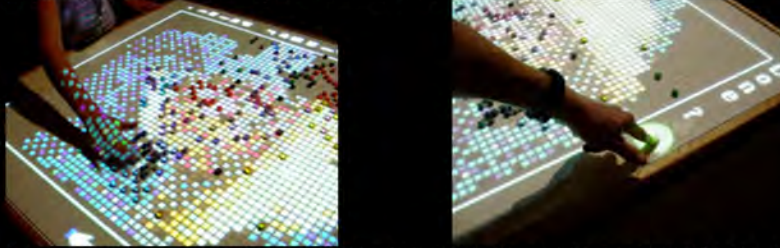
place in this iteration by simply making a "clicking" gesture on the M&M, of the desired color.



The ED-projector then rotates the mirror so it highlights exactly the bucket that contains the M&M's of that color. It displays the message "Get n", (where n is the number of M&M's to be picked up from the bucket) on the bucket's frontal face. In addition, a button with the word "done" 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. Immediately afterward an instruction to "Go to the table" is projected on the bucket.



The ED-projector then indicates the precise location where the M&M's should be placed by projecting an image of the target locations on the canvas.



At this point, the user is invited to "Finger paint", that is, to wave her hand over the board to reveal the completed picture. The camera system tracks the position of the user's hand and fills in circles of the appropriate color, interactively completing the picture. At any time the user can stop this activity by touching the "yes" button to the right of the image.

The video

Click [HERE](#) or on the picture to watch a video (19Mbytes, AVI), showing a run of the system, from the point of view of the ED-projector camera. This is exactly the image fed to the [computer vision system used to detect the interaction](#).

Copyright IBM Corporation, 2001.



The results



Publications

A publication describing this experience and the basic technology involved is [available](#). For general information, contact [Claudio Pinhanez](#)

M&M and ColorWorks are registered trademarks of [Mars, Inc.](#)
All pictures by Claudio Pinhanez, Copyright IBM Corporation 2001.

Last update: 10/10/02

[Research home](#) [IBM home](#) [Order](#) [Privacy](#) [Legal](#) [Contact IBM](#)

https://web.archive.org/web/20060524013342/http://domino.research.ibm.com:80/comm/research_people.nsf/pages/pinhanez.pubs.html



Publications

IBM Research

Publications by Claudio Pinhanez in chronological order

IS Interactive Spaces
CV Computer Vision
UI User Interface
CT Computer Theater
TL Temporal Logic
WE Web Entertainment

- IS** **To Frame or Not to Frame: The Role and Design of Frameless Displays in Ubiquitous Applications**
UI C. Pinhanez, M. Podlaseck
Proc. of Ubicomp'05. Tokyo, Japan, September 11-14. 2005
[PDF](#) [Request hardcopy](#)
- IS** **A Study on the Manipulation of 2D Objects in a Projector/Camera-Based Augmented Reality Environment**
UI S. Voids, M. Podlaseck, R. Kjeldsen, C. Pinhanez
In: Proc. of CHI'05. Portland, Oregon. 2005
[PDF](#) [Request hardcopy](#)
- IS** **Using Symbiotic Displays to View Sensitive Information in Public**
UI S. Berger, R. Kjeldsen, C. Pinhanez, M. Podlaseck, C. Narayanaswami, M. Raghunath
In: Proc. of IEEE International Conference on Pervasive Computing and Communications (PerCom'05). Kauai Island, Hawaii. 2005
[PDF](#) [Request hardcopy](#)
- CV** **Dynamically Reconfigurable Vision-Based User Interfaces**
UI R. Kjeldsen, A. Levas, C. Pinhanez
International Journal of Machine Vision and Applications, vol. 16(1): 6-12, 2004
[PDF](#) [Request hardcopy](#)
- CV** **A Portable System for Anywhere Interactions**
IS N. Sukaviriya, R. Kjeldsen, C. Pinhanez, L. Tang, A. Levas, G. Pingali, and M. Podlaseck
In Proc. of CHI'04 (Demonstrations). Vienna, Austria. 2004
[PDF](#) [Request hardcopy](#)
- IS** **Fostering a Symbiotic Handheld Environment**
UI M. Raghunath, C. Narayanaswami, C. Pinhanez
IEEE Computer, vol. 36(9): 56-65, 2003
[PDF](#) [Request hardcopy](#)
- UI** **An Architecture and Framework for Steerable Interface Systems**
IS A. Levas, C. Pinhanez, G. Pingali, R. Kjeldsen, M. Podlaseck, N. Sukaviriya
CV In: Proc. of Ubiquitous Computing 2003 (UbiComp'03). Seattle, Washington. October 2003
[PDF](#) [Request hardcopy](#)
- UI** **Applications of Steerable Projector-Camera Systems**
IS C. Pinhanez, R. Kjeldsen, A. Levas, G. Pingali, M. Podlaseck, N. Sukaviriya
CV In: Proc. of ICCV Workshop on Projector-Camera Systems (PROCAMS'03). Nice, France. October 2003
[PDF](#) [Request hardcopy](#)
- UI** **Embedding Interactions in a Retail Store Environment: The Design and Lessons Learned**
IS Noi Sukaviriya, Mark Podlaseck, Rick Kjeldsen, Anthony Levas, Gopal Pingali, Claudio Pinhanez
In: Proc. of the Ninth IFIP International Conference on Human-Computer Interaction (INTERACT'03). Zurich, Switzerland. September 2003
[PDF](#) [Request hardcopy](#)
- UI** **Ubiquitous Interactive Displays in a Retail Environment**
IS Claudio Pinhanez, Mark Podlaseck, Rick Kjeldsen, Anthony Levas, Gopal Pingali, Noi Sukaviriya
In: Proc. of SIGGRAPH'03 Sketches. San Diego, California. July 2003
[PDF](#) [Request hardcopy](#)
- TL** **Interval Scripts: a Programming Paradigm for Interactive Environments and Agents**
IS Claudio Pinhanez and Aaron Bobick
CT Pervasive and Ubiquitous Computing; vol. 7(1): 1-21, 2003
[PDF](#) [Request hardcopy](#)
- UI** **On Interfaces Projected onto Real-World Objects**
Mark Podlaseck, Claudio Pinhanez, Nancy Alvarado, Margaret Chan, Elisa Dejesus
In: Proc. of CHI'03 - Short Papers. Fort Lauderdale, Florida. April 2003
[PDF](#) [Request hardcopy](#)
- UI** **Augmenting a Retail Environment Using Steerable Interactive Displays**
IS Noi Sukaviriya, Mark Podlaseck, Rick Kjeldsen, Anthony Levas, Gopal Pingali, Claudio Pinhanez
In: Proc. of CHI'03 - Short Papers. Fort Lauderdale, Florida. April 2003
[PDF](#) [Request hardcopy](#)
- CV** **Dynamically Reconfigurable Vision-Based User Interfaces**
UI Rick Kjeldsen, Anthony Levas, Claudio Pinhanez
BEST PAPER AWARD
In: 3rd International Conference on Vision Systems (ICVS'03). Graz, Austria. April 2003
[PDF](#) [Request hardcopy](#)

- UI Steerable Interfaces for Pervasive Computing Spaces**
IS Gopal Pingali, Claudio Pinhanez, Anthony Levas, Rick Kjeldsen, Mark Podlaseck, Han Chen, Noi Sukaviriya
CV **Mark Weiser BEST PAPER AWARD**
 In: IEEE International Conference on Pervasive Computing and Communications - PerCom'03. Dallas-Fort Worth, Texas. March 2003
[PDF](#) [Request harcopy](#)
- IS "It/I": A Theater Play Featuring an Autonomous Computer Character**
CT Claudio Pinhanez and Aaron Bobick
CV Presence: Teleoperators and Virtual Environments vol. 1(5): 536-548. 2002
[PDF](#) [Request harcopy](#)
- IS BlueSpace: Personalizing Workspace through Awareness and Adaptability**
UI Jennifer Lai, Anthony Levas, Paul Chou, Claudio Pinhanez, Marisa Viveros
 International Journal of Human Computer Studies vol. 57(5): 415-428. 2002
[PDF](#) [Request harcopy](#)
- UI User-Following Displays**
IS Gopal Pingali, Claudio Pinhanez, Tony Levas, Rick Kjeldsen, Mark Podlaseck
CV In: Proc. of the IEEE International Conference on Multimedia and Expo 2002 (ICME'02). Lausanne, Switzerland. 2002
[PDF](#) [Request harcopy](#)
- WE That's Entertainment! Designing Streaming, Multimedia Web Experiences**
 Clare-Marie Karat, John Karat, John Vergo, Claudio Pinhanez, Doug Riecken, Thomas Cofino
 International Journal of Human-Computer Interaction vol. 14(3&4): 369-384. 2002
[PDF](#) [Request harcopy](#)
- UI Interacting with Steerable Projected Displays**
IS Rick Kjeldsen, Claudio Pinhanez, Gopal Pingali, Jacob Hartman, Tony Levas, Mark Podlaseck
CV Proc. of the 5th International Conference on Automatic Face and Gesture Recognition (FG'02), Washington (DC), May 20-21 2002
[PDF](#) [Request harcopy](#)
- UI Ubiquitous Interactive Graphics**
IS Claudio Pinhanez, Rick Kjeldsen, Tony Levas, Gopal Pingali, Mark Podlaseck, Paul Chou
CV IBM Research Report RC22495 (W0205-143), May 17, 2002
[PDF](#) [Request harcopy](#)
- IS Hyper Mask - Talking Head Projected onto Real Object**
CV Tatsuo Yotsukura, Shigeo Morishima, Franck Nielsen, Kim Binsted, Claudio Pinhanez
CT The Visual Computer, vol. 18(2):111-120, 2002
[PDF](#) [Request harcopy](#)
- IS Creating Ubiquitous Interactive Games Using Everywhere Displays Projectors**
UI Claudio Pinhanez
 Proc. of the International Workshop on Entertainment Computing (IWEC'02), Makuhari, Japan May 14-17, 2002
[PDF](#) [Request harcopy](#)
- TL Interval Scripts: a Programming Paradigm for Interactive Environments and Agents**
IS Claudio Pinhanez, Aaron Bobick
CT Submitted to Personal and Ubiquitous Computing, May 2002
[PDF](#) [Request harcopy](#)
- CT "It/I: a Theater Play Featuring an Autonomous Computer Character**
IS Claudio Pinhanez, Aaron Bobick
CV To appear in Presence: Teleoperators and Virtual Environments, (final manuscript) May 2002
[PDF](#) [Request harcopy](#)
- IS BlueSpace: Creating a Personalized and Context-Aware Workspace**
UI Paul Chou, Marco Gruteser, Jennifer Lai, Tony Levas Scott McFaddin, Claudio Pinhanez, Marisa Viveros, Danny Wong, Sachiko Yoshihama
 IBM Research Report RC22281 (W0112-044), December 11, 2001
[PDF](#) [Request harcopy](#)
- UI Transforming Surfaces into Touch-Screens**
IS Claudio Pinhanez, Rick Kjeldsen, Tony Levas, Gopal Pingali, Jacob Hartman, Mark Podlaseck, Vivek Kwatra, Paul Chou
CV IBM Research Report RC22273 (W0112-016), December 4, 2001
[PDF](#) [Request harcopy](#)
- UI The Everywhere Displays Projector: A Device to Create Ubiquitous Graphical Interfaces**
IS Claudio Pinhanez
CV Proc. of Ubiquitous Computing 2001 (UbiComp'01), Atlanta, Georgia, September 2001
[PDF](#) [Request harcopy](#)
- IS Hyper Mask - Projecting Virtual Face onto a Moving Real Object**
CV S. Morishima, T. Yotsukura, F. Nielsen, K. Binsted, and C. Pinhanez.
CT Proc. of Eurographics'01, Manchester, England, September 2001
[PDF](#) [Request harcopy](#)
- WE Less Clicking, More Watching: Results of the Iterative Design and Evaluation of Entertaining Web Experiences**
 Clare-Marie Karat, Claudio Pinhanez, John Karat, Renee Arora, John Vergo
 Proc. of Interact'2001, Tokyo, July 2001
[PDF](#) [Request harcopy](#)
- UI Augmented Reality with Projected Interactive Displays**
IS Claudio Pinhanez
CV Proc. of International Symposium on Virtual and Augmented Architecture (VAA'01), Dublin, Ireland, 2001
[PDF](#) [Request harcopy](#)
- WE Can Web Entertainment Be Passive?**
 Claudio Pinhanez, Clare-Marie Karat, John Vergo, John Karat, Renee Arora, Doug Riecken, Thomas Cofino
 Proc. of International World Wide Web 2001 (IWWW'01), Hong Kong, May 2001
[PDF](#) [Request harcopy](#)
- UI Using a Steerable Projector and a Camera to Transform Surfaces into Interactive Displays**
IS Claudio Pinhanez
CV Proc. of CHI'01 - Short Talks, Seattle, April 2001

- [PDF](#) [Request harcopy](#)
- WE "Less Clicking, More Watching": An Option for Entertainment in the Web**
 Claudio Pinhanez, Clare-Marie Karat, John Karat, John Vergo, Renee Arora
 Proc. of CHI'01 - Short Talks, Seattle, April 2001
[PDF](#) [Request harcopy](#)
- IS Physically Interactive Story Environments**
CT Claudio Pinhanez, J. Davis, S. Intille, Michael Johnson, A. Wilson, A. Bobick, and B. Blumberg
 IBM Systems Journal, vol. 39 (3&4): 438-455, 2000
[PDF](#) [Request harcopy](#)
- IS HyperMask: Projecting Talking Head on Moving Surface (in Japanese)**
CV T. Yotsukura, K. Binsted, F. Nielsen, C. Pinhanez, N. Tetsutani., and S. Morishima
 Proc. of the 16th Japanese Conference of Computer Graphics (NICOGRAPH'00), Tokyo, Japan,
CT November 2000
[PDF](#) [Request harcopy](#)
- IS HyperMask: Talking Head Projected Onto Real Objects**
CV S. Morishima, T. Yotsukura, K. Binsted, F. Nielsen, and C. Pinhanez
 Proc. of Multimedia Modeling (MMM'00), pp. 403-412, 2000
CT [PDF](#) [Request harcopy](#)
- IS Scripting Interactive Environments with Interval Scripts**
CV Claudio Pinhanez, Aaron Bobick
 Proc. of VI Brazilian Symposium on Multimedia and Hypermedia Systems
CT Natal, Brazil, June 2000
[PDF](#) [Request harcopy](#)
- IS The KidsRoom**
CV A. Bobick, S. Intille, J. Davis, F. Baird, C. Pinhanez, L. Campbell, Y. Ivanov, A. Schutte, A. Wilson
 Communications of the ACM, 43(3), 2000
[Request harcopy](#)
- IS The SCD Architecture and its Use in the Design of Story-Driven Interactive Spaces**
 Claudio Pinhanez
 Proc. of 1st International Workshop on Managing Interactions in Smart Environments (MANSE'99)
 Dublin, Ireland, November 1999
[PDF](#) [Request harcopy](#)
- IS The KidsRoom: A Perceptually-Based Interactive Immersive Story Environment**
CV A. Bobick, S. Intille, J. Davis, F. Baird, C. Pinhanez, L. Campbell, Y. Ivanov, A. Schutte, A. Wilson
 PRESENCE: Teleoperators and Virtual Environments, 8(4), August 1999
[PDF](#) [Request harcopy](#)
- IS Representation and Recognition of Action in Interactive Spaces**
CV Claudio Pinhanez
 Ph.D. Thesis, M.I.T., May 1999
CT [PDF](#) [Request harcopy](#)
TL
- CT Using Computer Vision to Control a Reactive Computer Graphics Character in a Theater Play**
CV Claudio Pinhanez and Aaron Bobick
 Proc. of International Conference in Vision Systems '99 - Las Canarias, Spain, , pp. 66-82, January 13-15 1999
[PDF](#) [Request harcopy](#)
- CT "It/I": A Theater Play Featuring an Autonomous Computer Graphics Character**
CV Claudio S. Pinhanez and Aaron F. Bobick
 Proc. of the ACM Multimedia'98 Workshop on Technologies for Interactive Movies, Bristol, England, pp. 22-29, September 1998
[PDF](#) [Request harcopy](#)
- TL Human Action Detection using PNF Propagation of Temporal Constraints**
CV Claudio Pinhanez and Aaron Bobick
 Proc. of CVPR'98, Santa Barbara, California, pp. 898-904, June 1998
[PDF](#) [Request harcopy](#)
- CT Beyond the Current Use of Computers in Theater**
 Claudio Pinhanez
 Profile (edited by The New Zealand Association for Theatre Craft, Design, and Technology), 59, Autumn 1998
[Request harcopy](#)
- TL Fast Constraint Propagation on Specialized Allen Networks and its Application to Action Recognition and Control**
 Claudio Pinhanez and Aaron Bobick
 MIT Media Laboratory [TR-#456](#), January 1998
[PDF](#) [Request harcopy](#)
- IS Scripting Method Based on Temporal Intervals for Designing Interactive Systems**
TL K. Mase, C. Pinhanez, and A. Bobick
 (in Japanese) Transactions of the Information Processing Society of Japan, vol. 39 (5), pp. 1403-1413, 1998
[Request harcopy](#)
- IS Intelligent Studios: Modeling Space and Action to Control TV Cameras**
CV Claudio Pinhanez and A. Bobick
 Applications of Artificial Intelligence, vol. 11, pp. 285-305, 1997
[Request harcopy](#)
- CT Computer Theater**
 Claudio Pinhanez
 Proc. of the International Symposium of Electronic Arts (ISEA'97), Chicago, Illinois, September 1997
[PDF](#) [Request harcopy](#)
- CV Divide and Conquer: Using Approximate World Models to Control View-Based Algorithms**
 A. Bobick and C. Pinhanez
 Proc. of CVPR'97, Puerto Rico, USA, pp. 955-962, June 1997
[PDF](#) [Request harcopy](#)
- TL PNF Propagation and the Detection of Actions Described by Temporal Intervals**
CV Claudio Pinhanez and A. Bobick
 Proc. of the DARPA Image Understanding Workshop. New Orleans. Louisiana. May 1997

[Request harcopy](#)

IS Interval Scripts: a Design Paradigm for Story-Based Interactive Systems

Claudio Pinhanez, K. Mase, and A. Bobick
Proc. of CHI'97, Atlanta, Georgia, pp. 287-294, March 1997
[PDF](#) [Request harcopy](#)

CV Approximate World Models: Incorporating Qualitative and Linguistic Information into Vision Systems

Claudio Pinhanez and A. Bobick
Proc. of AAAI'96, Portland, Oregon, pp. 1116-1123, August 1996
[PDF](#) [Request harcopy](#)

CT Computer Theater: Stage for Action Understanding

Claudio Pinhanez and A. Bobick
Proc. of the AAAI'96 Workshop on Entertainment and AI/A-Life, Portland, Oregon, pp. 28-33, August 1996
[Request harcopy](#)

CV Scripts in Machine Understanding of Image Sequences

Claudio Pinhanez and A. Bobick
Proc. of the 1995 AAAI Fall Symposium on Computational Models for Integrating Language and Vision, Cambridge, Massachusetts, pp. 96-101, November 1995
[Request harcopy](#)

IS Intelligent Studios: Using Computer Vision to Control TV Cameras

Claudio Pinhanez and A. Bobick
Proc. of IJCAI'95 Workshop on Entertainment and AI/A-life, Montreal, Canada, pp. 69-76, August 1995
[PDF](#) [Request harcopy](#)

CV Using Approximate Models as Source of Contextual Information for Vision Processing

A. Bobick and C. Pinhanez
Proc. of ICCV'95 Workshop on Context-Based Vision, Cambridge, Massachusetts, pp. 13-21, June 1995
[PDF](#) [Request harcopy](#)

Internet in Developing Countries: The Case of Brazil

Claudio Pinhanez
Proc. of the Seminário Preparatório sobre Aspectos Sócio-Culturais da Internet no Brasil, LNCC, Rio de Janeiro, Brazil, August 1995
[HTML](#) [Request harcopy](#)

CV Behavior-Based Active Vision

Claudio Pinhanez
International Journal of Pattern Recognition and Artificial Intelligence, vol. 8 (6), pp. 1493-1526, 1994
[Request harcopy](#)

CV Controlling a Highly Reactive Camera Using a Subsumption Architecture

Claudio Pinhanez
Proc. of Applications of AI'93: Machine Vision and Robotics, SPIE, Orlando, Florida. pp. 100-111, 1993
[Request harcopy](#)

IBM publications

Claudio S. Pinhanez and Mark E. Podlaseck. To Frame or Not to Frame: The Role and Design of Frameless Displays in Ubiquitous Applications. *UbiComp'05 - The Seventh International Conference on Ubiquitous Computing*. May 2005.

Stefan Berger, Frederik C. Kjeldsen, Claudio S. Pinhanez, Mark E. Podlaseck, Chandra Narayanaswami and Mandayam Raghunath. Using Symbiotic Displays to View Sensitive Information in Public. *PerCom'05 - 3rd IEEE International Conference on Pervasive Computing and Communications*. IEEE, January 2005.

Claudio S. Pinhanez and Gopal S. Pingali. Projector-Camera Systems for Telepresence. *ETP'04 Effective Telepresence Workshop 2004 (part of ACM Multimedia 2004)*. January 2004.

Anthony Levas, Claudio S. Pinhanez, Gopal S. Pingali, Frederik C. Kjeldsen, Mark E. Podlaseck and Piyawadee Sukaviriya. [An Architecture and Framework for Steerable Interface Systems](#). *UBICOMP 03*. June 2003.

Claudio S. Pinhanez, Frederik C. Kjeldsen, Mark E. Podlaseck, Anthony Levas, Gopal S. Pingali and Piyawadee Sukaviriya. Ubiquitous Interactive Displays in a Retail Environment. *SIGGRAPH Sketches and Applications*. ACM, April 2003.

Gopal S. Pingali, Claudio S. Pinhanez, Anthony Levas, Frederik C. Kjeldsen, Mark E. Podlaseck, Han Chen and Piyawadee Sukaviriya. Steerable Interfaces for Pervasive Computing Spaces. *PerCom2003 -- IEEE Conference on Pervasive Computing and Communication*. IEEE, March 2003.

Frederik C. Kjeldsen, Claudio S. Pinhanez and Anthony Levas. Dynamically Reconfigurable Vision-based User Interfaces. *3rd International Conference on Computer Vision Systems*. ECVision European Network of Excellence on Cognitive Vision, December 2002.

Mandayam T. Raghunath, Chandrasekhar Narayanaswami and Claudio S. Pinhanez. Fostering a Symbiotic Handheld Environment. *IEEE Computer Magazine* 36(9):56-65, July 2003.

Jennifer C. Lai, Anthony Levas, Paul B. Chou, Claudio S. Pinhanez and Marisa Viveros. BlueSpace: Personalizing Workspace through Awareness and Adaptability. *International Journal of Human-Computer Studies* 57(5):415-28, November 2002.

<https://web.archive.org/web/20041013055749/http://www.time.com/time/covers/1101041011/gadgets/6.html>

PHOTO ESSAY | TECHNOLOGY

<< Previous | Next >>



TED THAI FOR TIME

Everywhere Display

WHO: IBM

WHY: Touch-screen kiosks have taken off, but it's not always easy to find space to install them. This invention removes the physical screen so that interactive images can be cast onto almost any surface.

HOW IT WORKS: A wall-mounted projector, paired with a camera, shows images on the floor or wall. The camera tracks people's interactions with the visuals, and the system interprets them the way it would a moving mouse or a tap on a screen.

APPLICATIONS: Retailers could lure shoppers with interactive window displays. Airports might offer visually aided instructions and directions throughout terminals. In a lab or office, collaborators might use it like a high-tech dry-erase board that displays Excel spreadsheets and other interactive material.

HOW MUCH: The display won't be sold in stores and will be priced according to its corporate customers' specifications.

<< Previous

NEXT >>

MORE PHOTOS



There's Music In My Glasses

Here's a peek at the coolest gadgets hitting the market in the coming months



Tomorrow's Sports Talent

Meet some of the young athletes who may have what it takes

MORE STORIES: [Forward Thinking](#) | [Next Chapter](#) | [Marketing](#)



Research home | IBM home | Order | Privacy | Legal | Contact IBM
Everywhere Displays Home | Technology | Applications | Publications | People | Events/Media

Everywhere Displays: Publications

Embedding Interactions in a Retail Store Environment: The Design and Lessons Learned

Noi Sukaviriya, Mark Podlaseck, Rick Kjeldsen, Anthony Levas, Gopal Pingali, Claudio Pinhanez

To appear in Proc. of the Ninth IFIP International Conference on Human-Computer Interaction (INTERACT'03). Zurich, Switzerland. September 2003

[Abstract](#) [PDF](#) [Request harcopy](#)

Ubiquitous Interactive Displays in a Retail Environment

Claudio Pinhanez, Mark Podlaseck, Rick Kjeldsen, Anthony Levas, Gopal Pingali, Noi Sukaviriya

To appear in Proc. of SIGGRAPH'03 Sketches. San Diego, California. July 2003

[Abstract](#) [PDF](#) [Request harcopy](#)

On Interfaces Projected onto Real-World Objects

Mark Podlaseck, Claudio Pinhanez, Nancy Alvarado, Margaret Chan, Elisa Dejesus

In: Proc. of CHI'03 - Short Papers. Fort Lauderdale, Florida. April 2003

[Abstract](#) [PDF](#) [Request harcopy](#)

Augmenting a Retail Environment Using Steerable Interactive Displays

Noi Sukaviriya, Mark Podlaseck, Rick Kjeldsen, Anthony Levas, Gopal Pingali, Claudio Pinhanez

In: Proc. of CHI'03 - Short Papers. Fort Lauderdale, Florida. April 2003

[Abstract](#) [PDF](#) [Request harcopy](#)

Dynamically Reconfigurable Vision-Based User Interfaces

Rick Kjeldsen, Anthony Levas, Claudio Pinhanez

BEST PAPER AWARD

In: 3rd International Conference on Vision Systems (ICVS'03). Graz, Austria. April 2003

[Abstract](#) [PDF](#) [Request harcopy](#)

Steerable Interfaces for Pervasive Computing Spaces

Gopal Pingali, Claudio Pinhanez, Anthony Levas, Rick Kjeldsen, Mark Podlaseck, Han Chen, Noi Sukaviriya

Mark Weiser BEST PAPER AWARD

In: IEEE International Conference on Pervasive Computing and Communications - PerCom'03. Dallas-Fort Worth, Texas. March 2003

[Abstract](#) [PDF](#) [Request harcopy](#)

BlueSpace: Personalizing Workspace through Awareness and Adaptability

Jennifer Lai, Anthony Levas, Paul Chou, Claudio Pinhanez, Marisa Viveros

International Journal of Human Computer Studies vol. 57(5): 415-428. 2002

[Abstract](#) [PDF](#) [Request harcopy](#)

User-Following Displays

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

Proc. of the IEEE International Conference on Multimedia and Expo 2002 (ICME'02), Lausanne, (Switzerland) , August 26-29 2002

[Abstract](#) [PDF](#) [Request harcopy](#)

Interacting with Steerable Projected Displays

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

Proc. of the 5th International Conference on Automatic Face and Gesture Recognition (FG'02), Washington (DC), May 20-21 2002

[Abstract](#) [PDF](#) [Request harcopy](#)

Ubiquitous Interactive Graphics

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

IBM Research Report RC22495 (W0205-143), May 17, 2002

[Abstract](#) [PDF](#) [Request harcopy](#)

Creating Ubiquitous Interactive Games Using Everywhere Displays Projectors

Claudio Pinhanez

Proc. of the International Workshop on Entertainment Computing (IWEC'02), Makuhari, Japan

May 14-17, 2002

[Abstract](#) [PDF](#) [Request harcopy](#)

BlueSpace: Creating a Personalized and Context-Aware Workspace

Paul Chou, Marco Gruteser, Jennifer Lai, Tony Levas, Scott McFaddin, Claudio Pinhanez, Marisa Viveros, Danny Wong, Sachiko Yoshihama

IBM Research Report RC22281 (W0112-044), December 11, 2001

[Abstract](#) [PDF](#) [Request harcopy](#)

Transforming Surfaces into Touch-Screens

Claudio Pinhanez, Rick Kjeldsen, Tony Levas, Gopal Pingali, Jacob Hartman, Mark Podlaseck, Vivek Kwatra, Paul Chou

IBM Research Report RC22273 (W0112-016), December 4, 2001

[Abstract](#) [PDF](#) [Request harcopy](#)

The Everywhere Displays Projector: A Device to Create Ubiquitous Graphical Interfaces

Claudio Pinhanez

Proc. of Ubiquitous Computing 2001 (UbiComp'01), Atlanta, Georgia, September 2001

[Abstract](#) [Postscript](#) [PDF](#) [Request harcopy](#)

Augmented Reality with Projected Interactive Displays

Claudio Pinhanez

Proc. of International Symposium on Virtual and Augmented Architecture (VAA'01), Dublin, Ireland, 2001

[Abstract](#) [Postscript](#) [PDF](#) [Request harcopy](#)

Using a Steerable Projector and a Camera to Transform Surfaces into Interactive Displays

Claudio Pinhanez

Proc. of CHI'01 - Short Talks, Seattle, April 2001

[Abstract](#) [Postscript](#) [PDF](#) [Request harcopy](#)

Last update: 06/03/03

[Research home](#)

[IBM home](#)

[Order](#)

[Privacy](#)

[Legal](#)

[Contact IBM](#)

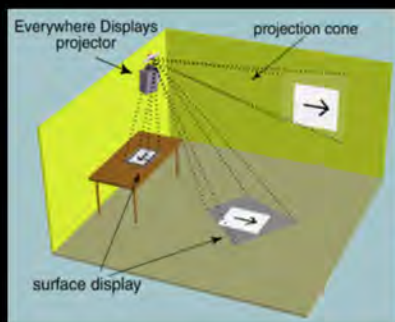


IBM Research

Research home | IBM home | Order | Privacy | Legal | Contact IBM
Everywhere Displays Home | Technology | Applications | Publications | People | Events/Media

The Everywhere Displays Project

The Everywhere Displays project aims to develop systems that allow the transformation of every surface in a space into a projected "touch screen". We are developing a prototype where we combine a LCD projector, a pan/tilt mirror, and a camera. The mirror is used to deflect the image of the projector to surfaces, walls, or the floor of a room. We process the projected image to compensate for the perspective distortion. We employ a pan/tilt video camera to detect hand/body activity on the projected area, so people can interact with the projected image by simply touching the surface.



No goggles, no wired surfaces

A major emphasis in this project is to avoid "wiring" surfaces with monitors or other sensing devices. Unlike traditional augmented reality applications, users do not need to wear graphics goggles or special input devices. Everything happens simply by projection and vision processing: light in, light out. Also, the Everywhere Displays projector can be easily steered to a new surface by a simple software command, thus avoiding fixed setups where projectors or cameras can only perform one specific task. The pan/tilt mechanisms provide dynamic adaptation to changes in the space usage and to different patterns of people activity.

ED-projector in a Store

Three prototypes of application scenarios have been developed so far. The first examines the use of ED-projectors in a retail space. Steerable projected interactive displays are particularly suited to public spaces such as stores since they can be installed on the ceiling, preventing accidental damage and discouraging vandalism and theft. Three different applications have been developed for the retail environment including: interactive bins for women's pants (photo), projected product directories, and an augmented table of Halloween products.



ED-projector in BlueSpace

The second is a prototype of a context-aware, sensor-loaded future office environment called **BlueSpace**. In this space, the ED-projector is used to deliver silent notifications, to dynamically create collaborative displays on tables and walls, and to contribute to personalization.



ED-projector at SIGGRAPH'01

The third prototype is an augmented reality (AR) demonstration scenario of an assembly task. Unlike traditional AR, the use of ED-projectors enable user visualization of information without encumbering goggles and interaction without contact with input devices. In our demo at **SIGGRAPH'01**, assembly instructions are superimposed on the objects to be assembled. Each individual user contributes to the assembly of an object by executing specific parts of the

assembly process, without knowledge of the complete structure of the assembled object.

[Click HERE for a video of two of the prototype applications\(19MB, QuickTime\).](#)

Research Recognition

The research carried out in the Everywhere Displays project has recently been recognized by two major awards:

- **Best paper award** at the 3rd Int.l Conference on Vision Systems (ICVS'03). April 2003.
- **IEEE Mark Weiser Award** for the best paper at IEEE PerCom'03. March 2003.

More information

[Technology](#) | [Applications](#) | [Publications](#) | [People](#) | [Events/Media coverage](#)

Contact: [Claudio Pinhanez](#)

Last update: 06/02/03

[Research home](#)

[IBM home](#)

[Order](#)

[Privacy](#)

[Legal](#)

[Contact IBM](#)

EXHIBIT C

The Everywhere Displays Projector: A Device to Create Ubiquitous Graphical Interfaces

Claudio Pinhanez

IBM Thomas Watson Research Center
P.O. Box 218, Yorktown Heights, NY 10598 -USA
pinhanez@us.ibm.com

Abstract. This paper introduces the Everywhere Displays projector, a device that uses a rotating mirror to steer the light from an LCD/DLP projector onto different surfaces of an environment. Issues of brightness, oblique projection distortion, focus, obstruction, and display resolution are examined. Solutions to some of these problems are described, together with a plan to use a video camera to allow device-free interaction with the projected images. The ED-projector is a practical way to create ubiquitous graphical interfaces to access computational power and networked data. In particular, it is envisioned as an alternative to the carrying of laptops and to the installation of displays in furniture, objects, and walls. In addition, the use of ED-projectors to augment reality without the use of goggles is examined and illustrated with examples.

1 Introduction

Ubiquitous computing envisions a world where it is possible to have access to computer resources anywhere and anytime to the data and services available through the Internet [1]. Since most of current software and Internet data is designed to be accessed through a high-resolution graphical interface, to truly ubiquitously compute today users need devices with reasonable graphical capabilities. This means carrying laptops everywhere, wearing computer graphics goggles, or installing monitors and displays on the surfaces of spaces and objects, such as desks, fridges, and entrance doors. Or, simply, to resign to the low-resolution displays of mobile phones or PDAs.

In this paper we explore an alternative approach to create ubiquitous graphical interfaces. Our idea is to couple an LCD/DLP projector to a motorized rotating mirror and to a computer graphics system that can correct the distortion caused by oblique projection. As the mirror moves, different surfaces become available to be used as displays. Also, we plan to employ a video camera to detect hand interaction with the projected image using computer vision techniques.

Our target is to develop a projection-based system that creates interactive displays everywhere in an environment by transforming a surface into a projected “touch screen.” Such an *Everywhere Displays projector* can be installed, for example, on the ceiling of a space, to provide a generic computer interface to users in that environment (see Fig. 1).

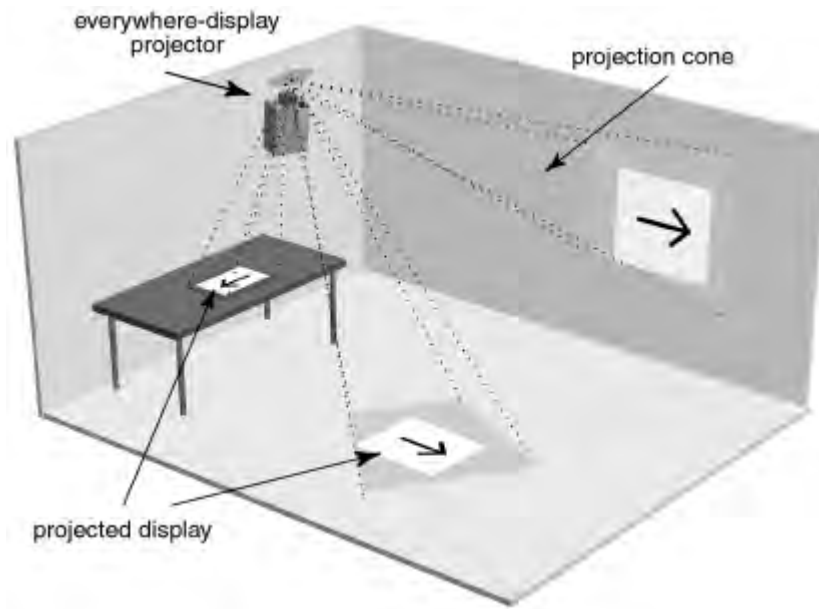


Fig. 1. Using the Everywhere Displays projector to create displays on different surfaces in an environment.

For example, an ED-projector can be installed in a meeting room and be used not only to project images on the walls but also to create individual displays for each of the participants of a meeting. Instead of today's meeting tables populated by bunkering personal laptops, a small set of ED-projectors can be shared by the participants to access their personal data, but easily reconfigured to allow teamwork. For instance, two people can be paired together to work on a sub-problem using a display projected in front of them while the other participants keep discussing the main themes using a display projected on the wall.

Moreover, ED-projectors have the ability to provide computer access in public environments without the risk of equipment being broken or stolen. Instead of carrying a computer, PDA, or phone everywhere, users can simply request a computer display, maybe by making a gesture to an overhead camera, and receive the projected image on a surface near their location. For example, an ED-projector in a store can transform pieces of white cardboard attached to shelves into interactive displays with product information. Unlike traditional kiosks, there is no need to bolt monitors and computers to the floor.

In other words, we are proposing a shift in the display paradigm, which ceases to be regarded as a device to be installed in an environment, or carried along by a user and becomes a service provided by a space, like electric power or phone lines. But like any innovation, ED-projectors not only solve a problem but also create a new set of applications. For instance, if information about location and identity of objects in an environment is known, an ED-projector can be used as a device to augment reality without requiring users to wear goggles. It can lead a visitor to its destination in a



Fig. 2. Current prototype of the Everywhere Displays projector (left) and the projector being used in an office-like environment (right).

building by directly projecting arrows on the floor; or project information directly onto the objects being assembled in an industrial plant. The ED-projector enables a computer system not only to augment a physical space with information but also to “act” in that environment and its users by projecting light patterns and symbols on objects and people.

This paper describes our current prototype of the Everywhere Displays projector and the technological solutions used in its implementation. We also demonstrate applications for different scenarios and examine new interaction paradigms for human-computer interaction that might emerge from the pervasive use of ED-projectors.

2 The Everywhere Displays Projector

The *Everywhere Displays projector*, or simply *ED-projector*, is composed of an LCD/DLP projector and a computer-controlled pan/tilt mirror. The projector is connected to the display output of a host computer that also controls the mirror. The left-side picture of Fig. 2 shows a prototype of an ED-projector built with an off-the-shelf rotating mirror used in theatrical/disco lighting, connected through a DMX network to the host computer.

In the configuration shown in Fig. 2-left, the projector’s light can be directed in any direction within the range of approximately 60 degrees in the vertical and 230



Fig. 3. Perception of contrast: global brightness (left); local brightness (middle); and a photomontage simulating the perceived contrast (right).

degrees in the horizontal. When positioned in the upper corner of a room (such as shown in Fig. 1), this prototype is able to project in most part of the two facing walls, half of the two adjacent walls, and almost everywhere on the floor.

Figure 2 also shows the current prototype of the ED-projector in use in an office-like environment. The top-right of Fig. 2 shows the ED-projector helping collaborative work. Notice the projector on the right upper corner of the picture and the angle of the rotating mirror used to direct the light onto the wall. The bottom-right picture of Fig. 2 shows the same surface being used to project the picture of an artwork as decoration for the room. Both photos were taken under normal office lighting conditions.

ED-projectors are feasible today due to the technical advances in two areas: video projectors and computer graphics engines. Current LCD/DLP projectors are able to create images that have enough contrast to be seen even when lights are turned on. Fast and cheap computer graphics engines are necessary to correct for the distortion caused by oblique projection. This and other implementation issues are discussed in the following sections.

2.1 Brightness and Contrast

Projecting images in a brightly lit room is possible because the human vision system perceives brightness and contrast locally. Consider a white wall in an environment with normal lighting: if no image is projected, subjects would describe the brightness of the wall as “white.” However, if a white and black pattern with sufficient brightness is projected on the same wall (typically 5 to 10 times brighter than the normal lighting), viewers perceive the white projected pattern as “white” and any neighboring area receiving only the ambient light as “black” [2].

Figure 3 exemplifies this mechanism in a situation where the ED-projector is used to create projected labels on white Styrofoam cups. The left picture of Fig. 3 shows the global brightness as “collected” by a photographic camera. In this picture, ambient light illuminates all the cups although the third cup of the top shelf also receives the projection of the pattern shown in the middle picture of Fig. 3. In the left image the

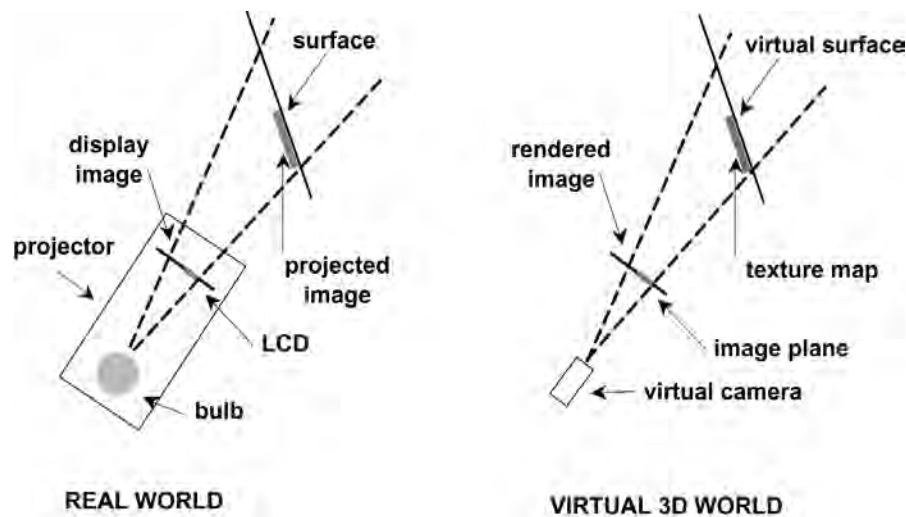


Fig. 4. Using a virtual computer graphics 3D world to correct the distortion caused by oblique projection by simulating the relationship in the real world between the projector and the projected surface.

projected pattern is barely visible because the photographic camera, unlike the human eye, equalizes the brightness globally. The camera takes in account the large black background of the shelf and compresses the range of brightness differences of the pattern projected on the cup into a few shades of white.

However, as shown in the middle of Fig. 3, there is enough local difference in brightness on the projected pattern to allow its clear perception if only local lighting is considered. In particular, notice that the black “OK” lettering corresponds to the white surface of the cup reflecting just the ambient light. Since human vision adjusts to local contrast, the resulting perceived brightness is more like the photomontage shown in Fig. 3-right where the projected pattern is clearly discernible.

Our first prototype employs a 1200 lumens LCD projector that has proved to have enough brightness and contrast to project images on the surfaces of an office room with the lights on. Although we have not conducted experiments to determine the perceived brightness and contrast, in typical home and office conditions a white pattern projected by our prototype is approximately 10 times brighter than its surroundings and, therefore, enough to create the illusion of contrast.

The second prototype we built employs a 3000 lumens LCD projector, enabling sharper contrast for most projected surfaces. In particular, the increase in brightness improved significantly the quality of images projected on horizontal surfaces such as tables and desks. Since such surfaces tend to be orthogonal to the sources of ambient light, the specular component of their reflected light is brighter than non-horizontal surfaces such as walls. However, with a 3000-lumen projector, even the extra brightness provided by the specular light is overshadowed by the projection light.

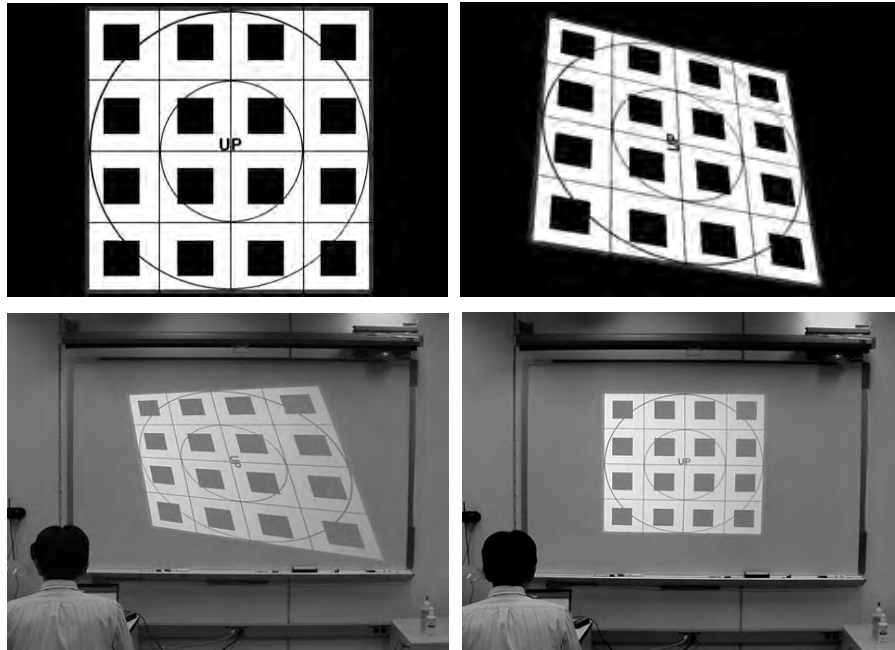


Fig. 5. Correction of oblique distortions: (left) the typical result of the oblique projection of a pattern on a surface; (right) the projection of the distorted pattern (top-right) creates a projected image free of distortion (bottom-right).

2.2 Correcting for Oblique Projection Distortion

When projection is not orthogonal to the projected surface, the projected image is distorted. In fact, as shown in Figs 1 and 2, in most cases the ED-projector is used to create displays on surfaces that are not orthogonal to the projection direction. To correct the distortions caused by oblique projection and by the shape of the projected surface (if not flat), the image to be projected must be inversely distorted prior to projection. In general, this distortion is non-linear and is computationally expensive to correct, involving the selective compression and the expansion of the original image.

We have developed a simple scheme that uses standard computer graphics hardware (present now in most computers) to speed up this process. Our method relies on the fact that, geometrically speaking, cameras and projectors with the same focal length are identical (as observed in [3, 4]). Therefore, to project an image obliquely without distortions it is sufficient to simulate the inverse process (i.e., viewing with a camera) in a virtual 3D computer graphics world.

As show in Fig. 4, we texture-map the image to be displayed onto a virtual computer graphics 3D surface identical (minus a scale factor) to the real surface. If the position and attitude of this surface in the 3D virtual space in relation to the 3D virtual camera is identical (minus a scale factor) to the relation between the real surface and the projector, and if the virtual camera has identical focal length to the

projector, then the view from the 3D virtual camera corresponds exactly to the “view” of the projector (if the projector was a camera). Since projectors do the inverse of viewing, i.e., they project light, the result is a projection free of distortions.

In practice we use a standard computer graphics board to render the virtual camera’s view of the virtual surface and send the computed view to the projector. If the position and attitude of the virtual surface are correct, the projection of this view compensates the distortion caused by oblique projection or by the shape of the surface. Of course, a different calibration of the virtual 3D surface must be used for each surface where images are projected in an environment.

An example of the results of the process is depicted in Fig. 5. In a typical situation of oblique projection, the pattern shown in the top-left is projected without any correction, resulting in the bottom-left image of Fig. 5. After calibration of the virtual 3D surface and camera parameters, the projection of the rendered image (top-right) creates a projection free of distortion (bottom-right).

So far we have experimented only with projecting on planar surfaces. The calibration parameters of the virtual 3D surface are determined manually by simply projecting the pattern shown in Fig. 5 and interactively adjusting the scale, rotation, and position of the virtual surface in the 3D world, and the “lens angle” of the 3D virtual camera. This process typically takes between 10 to 20 minutes but we are currently working on its automation using techniques similar to [5].

Another simple technique to correct for distortion on planar surfaces is simply to distort the texture to be projected by a homography [6]. In this case, calibration is obtained by interactively grabbing with the mouse each corner of the projected pattern and moving it to the desired location on the surface. Alternatively, the homography can be embedded to the graphics board projection matrix [3]. Unlike the previous approach, homographies work only for planar surfaces.

2.3 Focus

We currently use a LCD projector where focus and zoom parameters can be remotely controlled by computer commands issued through the serial port. However, another problem with oblique projection is that it is not possible to put all areas of the projected image simultaneously in focus. Fortunately, current commercial projectors have a reasonable depth of focus range, enough to maintain decent focus conditions in most cases. We have succeeded in projecting on surfaces with up to 30 degrees of inclination in relation to the projection axis without significant degradation of focus. However, the problem becomes more severe as the distance between the projected surface and the projector decreases.

2.4 Display Resolution

One problem with the techniques described above to correct oblique distortion is that it creates displays with resolutions that are smaller than the projector’s resolution. As can be seen in Fig. 5, the distortion correction process has to fit an irregular

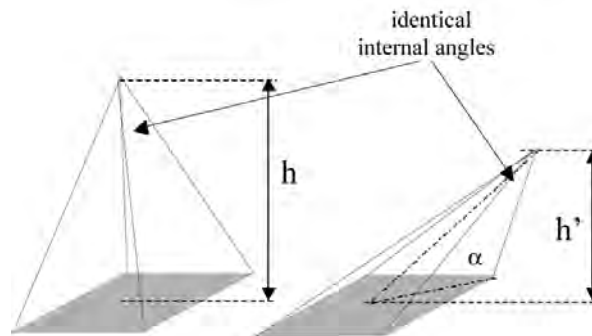


Figure 6. Change of volume for identical internal angles.

quadrangular into the 4:3 viewing area of typical displays. The result is that a considerable amount of display area is lost in the process.

In our prototypes we employ 1024x768 XVGA projectors. However, due to the loss of display area created by the distortion correction process, we have observed that the obtained resolution corresponds approximately to VGA, i.e., 640x480 pixels. This estimation takes in account that in the process of rendering the distorted image some pixels of the original image are compressed into single pixels of the projected image.

Other factors also influence the perceived resolution, among them the angle of projection and the texture of the projected surface. In the case of extreme angles of projection (<20 degrees), we have observed in some cases that while the center of the image is in focus, the area near the edges are somewhat blurred. This blurring is less visible if the projected image has less detail. Similarly, projecting onto textured surfaces such as carpets introduces a high frequency spatial component on the visual field, contributing to decrease the perceived resolution because of interference patterns. We are currently starting research aiming to determine how much perceived resolution is lost when projecting on surfaces of different colors, specular components, and textures.

2.5 Obstruction and Glare

Unlike LCD displays, the use of projectors as graphical displays face the problem of having the projected image being obstructed by people or moving objects in the environment. This was, in fact, one of our major concerns in the start of this project. However, our experience has shown that obstruction is far less common than we anticipated, particularly when the projector is positioned in the upper corner of a room (as shown in Fig. 1). Although we initially tried to position the projector on the center of the ceiling, the corner placement proved to be much more effective, mainly because in this situation the projection cone tends to be closer to the wall and therefore less prone to be intercepted by human beings.

Another reason for the relative lack of obstruction is the fact that oblique projection cones are smaller than orthogonal projection cones. To see this, consider

the two projection cones C and C' with identical bases and internal angle (i.e., lens configuration) and heights h and h' , as shown in Fig. 6. Since the loci of all vertices of pyramids of equal base and internal angle is a sphere¹, it's easy to see that the $h' = c.h.\sin\alpha$, where c is a constant smaller than 1 that depends on the geometry of the orthogonal pyramid C .

Since the volume of the pyramid is proportional to the product of the area of the base B by the height of the pyramid, $V = 1/3 B.h$, we obtain that the volume of the two pyramids are related by:

$$V' = V.c.\sin\alpha$$

For instance, if the angle of oblique projection α is 30 degrees, the volume of the oblique pyramid is less than half the volume of an equivalent orthogonal pyramid. Since the projection cones of oblique projections are smaller than those of orthogonal projections, we should expect a similar reduction in the likelihood of obstruction.

Similarly, the positioning of the projector on the ceiling, above human heads, contributes decisively to avoid glare and direct staring to the bulb of the projector. Again we see here a benefit of oblique projection that allows this placement of the projector on the ceiling. Although most commercial projectors today have some mechanical or electronic device to correct for keystone (typically less than 10%), this correction does not allow the positioning of the projector sufficiently high to avoid glare. In the office-like experimental setup of our environment, however, glare happens only in very unusual situations such as when the user is sitting on the floor.

3 Making the ED-Projector Interactive

The current prototypes of the ED-projector include the functionality to project on different surfaces a 512x512 portion of the interactive desktop display of the host computer. The user can interact in real-time with this projected desktop using mouse and keyboard.

We are now starting to explore the interaction with the projected surface without the need of users having to manipulate input devices. In other words, we would like to have the projected display behaving as if it was a "touch screen," making the ED-projector a system easily usable in public spaces or in hazardous environments. The goal is to have the user interact by moving her hand over the projected surface, as if the hand was a computer mouse (see Fig. 7); and by moving the hand rapidly towards the surface, to generate a "click" event. We are currently investigating the use of a pan/tilt video camera that is controlled by the computer so it has a complete view of the projected surface (depicted in Fig. 2, installed on the top of the ED-projector).

To track the position of the hand over the surface, we are considering the development of variations of the traditional background subtraction techniques used in computer vision [7]. However, unlike those cases, we have a situation where the

¹ This is the 3D equivalent of the known geometric property that for any two distinct points on a circumference, all triangles formed by the two points and one third point belonging to the arc between them have internal angles of the same size.

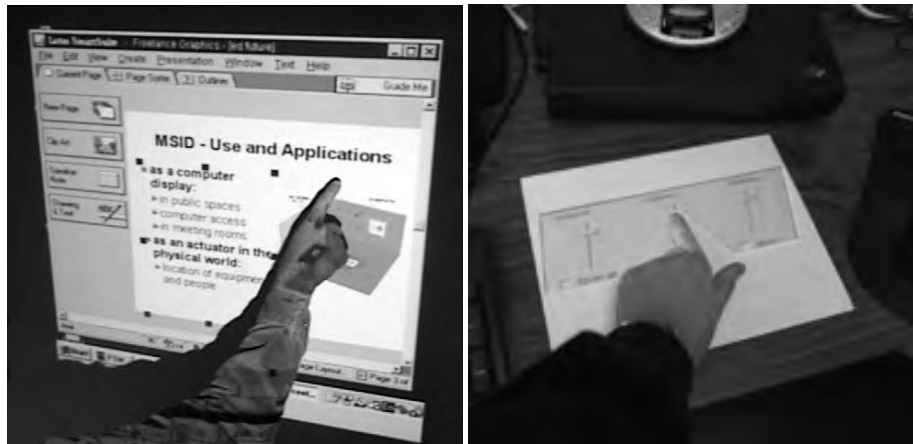


Fig. 7. Examples of interaction with the projected display by hand.

projected background changes significantly and abruptly over time. To overcome this problem, we are exploring two alternatives: the use of two synchronized cameras in a stereo configuration as in [8] and a method based on the estimation of the projected background of the image.

The second method explores the fact that although the projected background changes, what is being projected is known to the system. Therefore, if the patterns of light reflection are known for the projected surface, it is possible to estimate how the projected image will look like when seen by the camera. In other words, we are investigating whether we can calculate the expected background based on the surface reflectance characteristics (together with the camera influence on them) and the image being projected.

The detection of “clicking” events seems to be quite more complicated. To detect a fast movement of a hand or finger towards the projected surface, we need some estimation of the distance of the hand to the surface. Although we plan to experiment with stereo vision to solve this problem, we are considering a simpler scheme based on the measurement of the width of the shadow created by the hand. As seen in Fig. 7-left, the obstruction of the projected light by the user’s hand creates a black shadow, in this case to the left of the hand. As the hand gets closer to the surface the shadow decreases in width, especially in the area near the fingers. Our plan is to track the size of this shadow and use it to estimate the hand-surface distance.

A word of caution has to be made about using DLP projectors with standard video cameras. Since DLP projectors project each RGB separately, capturing the projected image with a camera normally results on images showing only one or two of the light components. The only way to overcome this problem, it is necessary to synchronize it with the projector and to increase the exposure time of the camera so it integrates all color components.

4 Using ED-Projectors in Ubiquitous Computing

The ED-projector is a generic input/output device that has been designed for use in multiple applications. These applications can be basically classified in two classes. The first class corresponds to the creation of interactive displays that provide computer access from surfaces of objects, furniture, walls, floor, etc. The second class of applications deals with typical augmented reality applications: the ED-projector can be used to point to physical objects, show connections among them, attach information to objects, and to project dynamic patterns to indicate movement or change in the real world. This section covers the applications in ubiquitous computing; the next section explores the uses of ED-projector as a tool to augment and affect reality.

4.1 Ubiquitous Access to Computational Resources and Information

The ED-projector is a tool that can create a high-resolution graphical interactive display to access computational resources, personal information, and the web. In this regard, it can be seen as a device that realizes the aspirations of ubiquitous computing [1] without the encumbering carrying of laptops or wearing of video-goggles. Since it can be installed on the ceiling of environments, it does not require the use of batteries or wireless links, eliminating two key problems of current laptops. And, unlike PDAs and phones, it does not require a change in most of today's interaction paradigms since it can provide access through a high-resolution (VGA, at least) display.

Fig. 8 depicts an example where a desktop application is moved around a room, being available in different surfaces that correspond to different uses of computer access. First, the display is laid on a desk, then moved onto a whiteboard, and finally projected on the wall besides the whiteboard. Notice that the display, in this last position, can be easily consulted while the users are scribbling on the whiteboard.

Similar applications can be created for professional environments such as hospitals where space is tightly constrained and there are restrictions on what can be carried by people. For example, a single ED-projector in an infirmary can provide on-demand computer access to nurses and doctors, freeing them from carrying laptops around. In another scenario, ED-projectors can substitute TV sets in patient rooms: although they can still be used to project TV images on the wall, they also enable computer access to data and communications from the patient's bed, without any need to move and connect equipment in the vicinity of the patient.

4.2 Collaborative Work

Video monitors, LCD screens, and PDAs are designed for individual use of computers. The only tool currently easily available for collaborative work with computer applications are LCD projectors, normally mounted on rooms to project on one of the walls, creating a "stage" effect for presentations. This configuration is



Fig. 8. The ED-projector moving a desktop application to the top of a table, to a whiteboard, and to a wall on the side of the whiteboard.

awkward for collaborative work since it creates a single point of attention and distracts people from looking at each other.

There has been a significant amount of HCI and design work to provide computer access to tables (for some examples, see [9, 10]). The ED-projector is clearly a device that enables this kind of interactive work, with the advantage that the resource can be easily moved to walls, adjacent tables, etc., allowing easy reconfiguration of a meeting space for different functions and teamwork styles.

Similarly, ED-projectors can be used in school classrooms. Instead of having a row of workstations with video monitors, the teacher can use ED-projectors to reconfigure the classroom for individual, group, or whole class activities. In day-care centers and kindergartens the device enables the use of walls and floors for education and entertainment, with the advantage of no infant contact with heavy or breakable objects.

There are many other similar applications in professional activities where teamwork is needed. Typical examples are surgery rooms, today crowded by arrays and arrays of bulk monitors displaying information and video imagery, and monitoring vital signs. A set of ED-projectors seems to substitute such configuration with the advantage that vital information can be easily brought in close proximity to nurses and physicians for detailed inspection or interaction.

4.3 Computer Access in Public Spaces and Hazardous Environments

The ED-projector can also be used to provide computer and information access in spaces where traditional displays can be broken or stolen, or create hazardous conditions, such as in public spaces or areas subject to harsh environmental conditions. Examples of spaces in the first category are waiting areas of transportation facilities such as subway stations and airports. The ceiling mounting of the ED-projector protects it against vandalizing, while the mirror allows easy change on where information is projected. This last feature is particularly important for this kind of environment where the cost of installing or reconfiguring equipment tends to be very high.

We also envision the use of ED-projectors in factory areas where normal electric devices are not allowed due to the potential risk of sparks, or simply because the environment itself contains hazardous elements to computers (such as water). The ED-projector can be encased in glass or other transparent material and still be able to create a useful display on such areas.

Another example of a “hazardous” environment is a home kitchen. By installing an ED-projector, and without the risk of spills, falls, or short-circuits, the kitchen user can access information, watch TV or surf the web, follow recipes, or simply set and control cooking time.

4.4 Bringing Computer Access to Disabled People

The ED-projector also permits an interactive display to 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 the sheet of a hospital bed without creating the risk of patient contact with any device. The patient, in this case, can interact with the display by simply using his hand and use it to search for information, call doctors and nurses, or to obtain access to entertainment.

5 Augmenting Reality with ED-Projectors

Ubiquitous computing research tends to focus on the issue of how computer resources can be spread through physical environments so computer and network access become seamless. However, an important consequence of embedding computers to physical environments is that it becomes easier to connect the computer to functions and knowledge that are specific of that space. A computer installed on the wall of a library is a natural candidate to manage the information about the books in that space, where they are located, and even how people are using them.

All this information about the environment can be provided to users. For instance, if the computer in the library knows where the books are located, it can help a user to find them. So far such applications have been explored in the context of augmented reality, mostly through the use of semi-transparent goggles [11, 12]. However,



Fig. 9. Augmented reality applications of Everywhere Displays: (1) an electronic phone directory is projected close to the phone; (2) the database corresponding to a file cabinet is accessed directly from the top of the cabinet; (3) a localization system points to the position of an object; (4) a notification of the arrival of an urgent e-mail is silently projected on the wall; (5) an emergency sign projected on the floor directly points to the nearest exit.

goggles not only are heavy and cumbersome, but they also require the fast and accurate tracking of the user's head to be effective.

We are exploring the use of the ED-projector in such environments. Unlike in previous works with static projectors [9, 10, 13], the rotating mirror expands the reach of projector to almost the entire room. It also enables the augmented reality image to follow an object around an environment (as long as there is a mechanism to track it). In the rest of this section, we examine augmented reality scenarios using the ED-projector.

5.1 Bringing Information to a Physical Location

Most applications of augmented reality are concerned with the virtual attachment of information to places and objects in the real world. Among such applications, we have been experimenting with the ED-projector to bring information to the physical location where the information is used or needed.

Figure 9.1 shows a simple example where a phone directory is projected close to a phone device, so it becomes easy to search and inspect information about the person a user needs to call. In an ideal situation, the simple act of picking up the phone could trigger the display of the directory on that surface. Figure 9.2 exemplifies another situation where a database application accessing a list of files is projected on the top of the file cabinet that contains those files. In this situation, the user can search the computer database using any kind of complex query, obtain and refine answers, and use the results to find the corresponding files in the cabinet.

5.2 Navigation, Tagging, and Localization of Resources

ED-projectors are ideal devices to support user navigation through an environment. For instance, by installing an ED-projector in a corridor it is possible to project information on any wall, door, or area in the floor. Moreover, if the intent is to provide directions for visitors, an ED-projector-based system has the advantage that no device needs to be given or worn by the visitor. Fig. 9.5 shows an example application where the projector is used to signal the direction of an emergency exit on the floor.

In a physical environment where the position of objects or components is known to the computer, it is possible to use an ED-projector system to visually point to an object's position or to tag it with relevant information, without requiring the user to wear or carry any kind of device. Fig. 9.3 shows an example where the ED-projector responds a verbal request for the position of an object (a digital camera) by directly pointing to the location of the object in the room. The system could follow by displaying a checkout list near the object that would allow the user to update the information about the item just by touching the projected checklist. Similarly, information about the digital camera or instructions for its use could have been displayed on the same area.

Unlike systems based on goggles, it is possible to have similar applications in public spaces. As shown in Fig. 10, a ED-projector can be used in a store to provide information about specific items, help people to find products, point to special promotions and sales events, and even provide entertainment for small children. All that is required is a set of white or light-colored surfaces where the information is to be projected.

5.3 Affecting People

The ED-projector can also be used as an actuator device to affect the people occupying a physical environment. If the system has information about the position of people, it can project light patterns that indicate possibilities or constraints to their movements and actions. For example, the projector can create an "electronic doormat" that helps the control of a line of people. It can indicate, directly on the floor, that the user should not move ahead by projecting, for instance, a red line. If the user keeps moving, the red line can follow him, and maybe increase in size or blink to stress the required human response.

Another application is the delivery of notification of important events to people, anywhere, and without the disruptive use of sound. Figure 9.4 shows a simple example of an e-mail notification being displayed on a wall close to where a user is working. Of course this scenario is only feasible if the identity, position, and head attitude of the user is known.

In many ways, the ED-projector is one of the first devices that allow a computer system to seamlessly act on the physical world we inhabit. It creates a harmless "robotic arm" of light that can affect people in multiple ways. Although sound and speech have been explored in the past as computer actuators in physical spaces, the

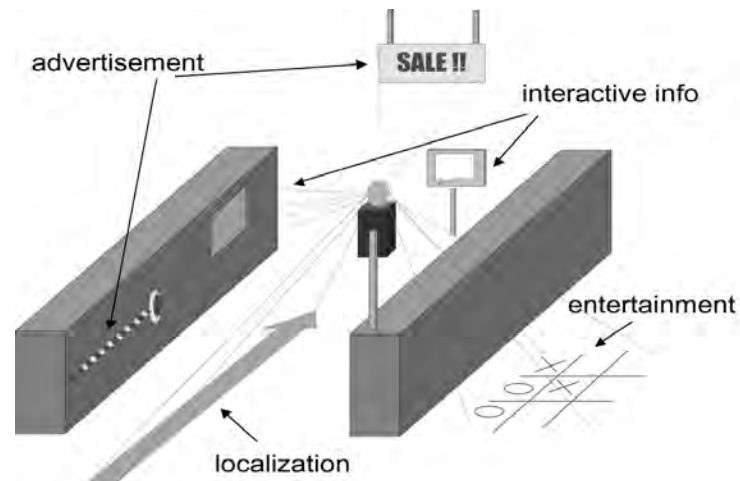


Fig. 10. Example of use of the ED-projector in a store.

ED-projector is unique in its possibilities of creating local, pinpointed ways for a computer to act on people.

6 Conclusion and Future Work

The main purpose of this paper is to present the ED-projector, a device consisting of a rotating mirror deflecting the light of a LCD projector. Although building on the previous research on interactive projected surfaces [9, 10, 14, 15], we believe that the introduction of the rotating mirror produces a very unique device. First, it allows a single I/O device to have multiple uses in the same environment and to dynamically create new displays as needed. An ED-projector can provide computer access to a user in one moment, project a notification of e-mail arrival in the next, and, in the case of an emergency, be used to direct people to exits. Second, an ED-projector creates a simple way for a computer to act on the objects and people in a physical environment, truly integrating computer actions to the real world.

The basic components for building an ED-projector are easily available. We have shown in this paper that the technology for correcting for oblique projection problems is simple and easy to implement. More research is needed, however, to make the calibration of the projector a simple process and certainly a lot more on the vision-based system for hand interaction with the projected image. Those problems, however, seem to be within the scope of currently known vision techniques.

The most exciting research is, in our view, on the new paradigms of human-computer interaction afforded by the concept of everywhere displays. Here we are proposing a scenario where ubiquitous computing happens without carrying displays around or installing them on furniture and walls. Unlike most of the current thinking

that looks into devices that can be carried by users, we are suggesting that graphical interactive displays can be a service provided by a space.

While initially we are expecting to support the current desktop paradigm, there is no reason to confine interaction to a rectangular frame like current monitors are forced to do. A possible avenue to explore is to consider the tangible interaction paradigm proposed by Ishii [16], but realize it without wiring or modifying the objects used for interaction. For example, a paper cup on a table can be easily transformed into a volume dial by simply projecting a scale around it with a “volume” label and responding to any rotation of the object by a user.

Finally, what kinds of collaboration can happen between human beings and computer in the moment that computers have the ability to point and affect objects and people in the real world? Kitchens and stores are particularly exciting scenarios to explore these novel concepts. We can also see situations, such as the control of a line of people mentioned above, where computers become devices able to act on people in the physical world. We are planning to explore further all these new paradigms for human-computer interaction that appear as a result of the introduction of ED-projectors in real environments.

References

1. Weiser, M.: The Computer for the Twenty-First Century. *Scientific American* (1991) 94-100
2. Hoffman, D.: *Visual Intelligence: How We Create What We See*. W. W. Norton. (1998).
3. Raskar, R.: Oblique Projector Rendering on Planar Surfaces for a Tracked User. In: *Proc. of SIGGRAPH'99*, Los Angeles, California (1999) 260
4. Pinhanez, C., Nielsen, F., and Binsted, K.: Projecting Computer Graphics on Moving Surfaces: A Simple Calibration and Tracking Method. In: *Proc. of SIGGRAPH'99*, Los Angeles, California (1999)
5. Yang, R. and Welch, G.: 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*. Plzen, Czech Republic (2001)
6. Faugeras, O.: *Three-Dimensional Computer Vision: A Geometric Viewpoint*. The MIT Press. Cambridge, Massachusetts. (1993).
7. Wren, C., Azarbayejani, A., Darrell, T., and Pentland, A.: Pfunder: Real-Time Tracking of the Human Body. *IEEE Trans. Pattern Analysis and Machine Intelligence* **19** (7) (1997) 780-785
8. Brumitt, B., Meyers, B., Krumm, J., Kern, A., and Shafer, S.: EasyLiving: Technologies for Intelligent Environments. In: *Proc. of 2nd International Symposium on Handheld and Ubiquitous Computing* (2000) 12-27
9. Rekimoto, J.: A Multiple Device Approach for Supporting Whiteboard-based Interactions. In: *Proc. of CHI'98*, Los Angeles, CA, April 18-23 (1998) 344-351
10. Underkoffler, J., Ullmer, B., and Ishii, H.: Emancipated Pixels: Real-World Graphics in the Luminous Room. In: *Proc. of SIGGRAPH'99*, Los Angeles, CA, August 8-13 (1999) 385-392
11. Krueger, M. W.: *Artificial Reality II*. Addison-Wesley. (1990).
12. Inami, M., Kawakami, N., Sekiguchi, D., Yanagida, Y., Maeda, T., Tachi, S., and Mabuchi, K.: Head-Mounted Projector for Projection of Virtual Environments on Ubiquitous Object-

- Oriented Retroreflective Screens in Real Environment. In: Proc. of SIGGRAPH'99, Los Angeles, California (1999) 245
13. Raskar, R., Welch, G., Cutts, M., Lake, A., and Stesin, L.: The Office of the Future: A Unified Approach to Image-Based Modeling and Spatially Immersive Displays. In: Proc. of SIGGRAPH'98, Orlando, Florida, July (1998) 179-188
 14. Crowley, J. L., Coutaz, J., and Berard, F.: Things that See. Communications of the ACM **43** (3) (2000) 54-64
 15. Binsted, K., Morishima, S., Nielsen, F., Pinhanez, C., and Yotsukura, T.: HyperMask: Virtual Reactive Faces for Storytelling. In: Proc. of SIGGRAPH'99, Los Angeles, California (1999) 186
 16. Ishii, H. and Ullmer, B.: Tangible Bits: Towards Seamless Interfaces between People, Bits, and Atoms. In: Proc. of CHI'97, Atlanta, Georgia, March (1997) 234-241

https://web.archive.org/web/20051014054021/http://lear.inrialpes.fr/people/triggs/events/iccv03/cdrom/procams03/pinhanez_kjeldsen_levas_pingali_podlaseck_sukaviriya.pdf

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

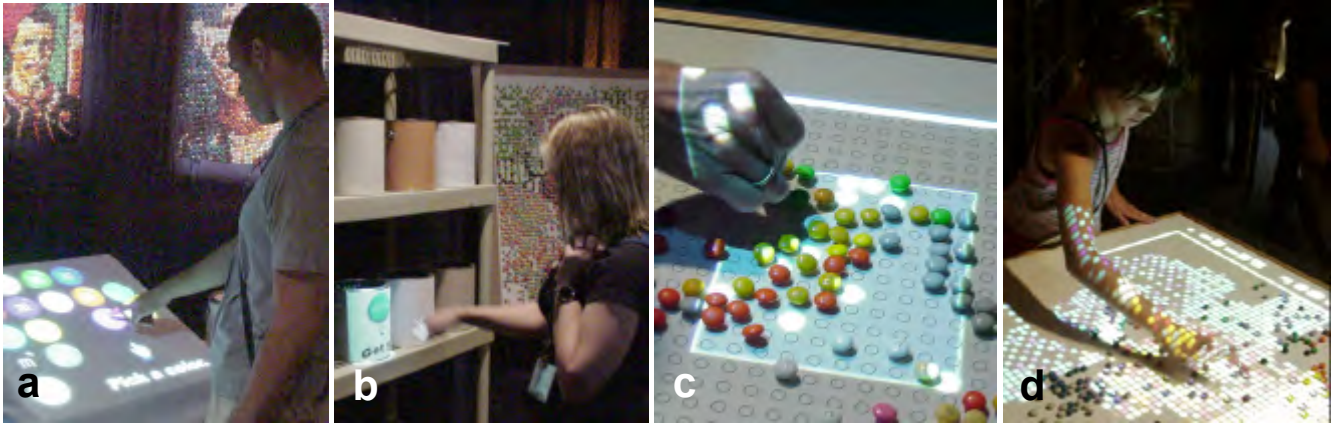


Figure 2. The M&M picture assembly task: a) choosing a color; b) “clicking” a bucket after getting the M&M’s; c) placing the M&M’s on the highlighted areas; and d) “finger painting” to reveal the complete picture.

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

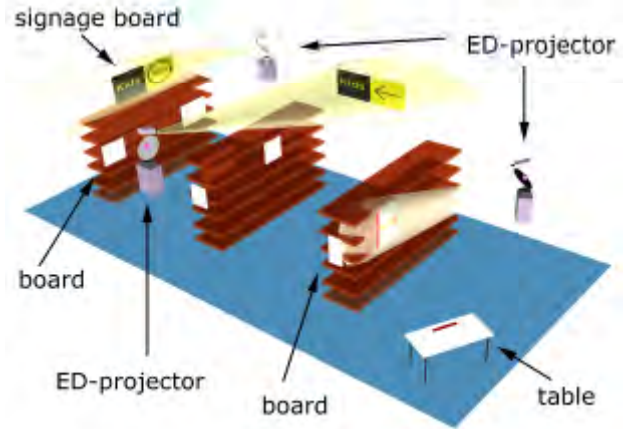


Figure 5 Concept of ubiquitous product finder.

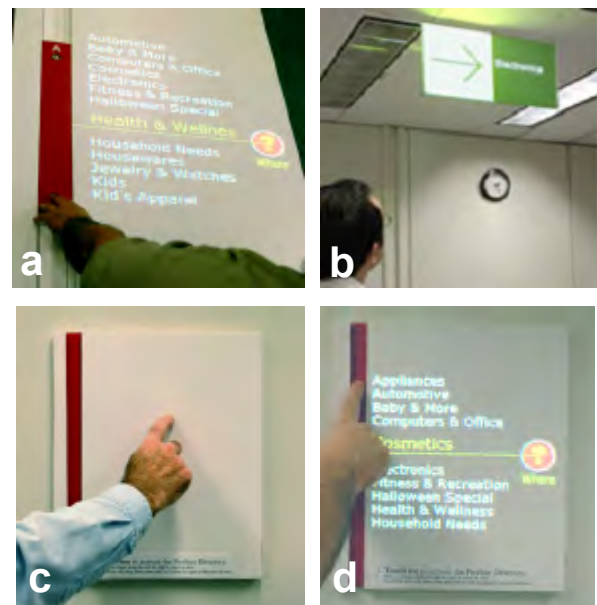


Figure 4. Prototype of ubiquitous product finder.

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, "Pffinder: 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, PUI'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 (PUI'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.

<https://web.archive.org/web/20040312224750/http://www.research.ibm.com:80/people/p/pinhanez/publications/rc22495.pdf>

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

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].

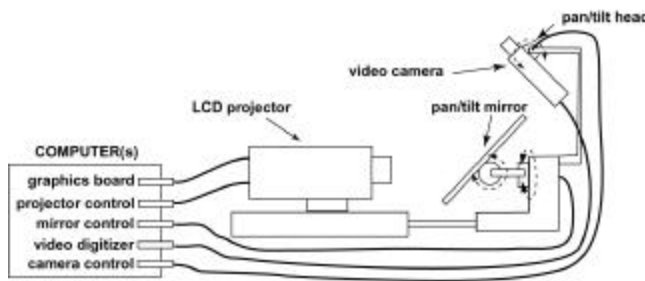


Figure 2. Diagram of the components of the ED-projector.



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

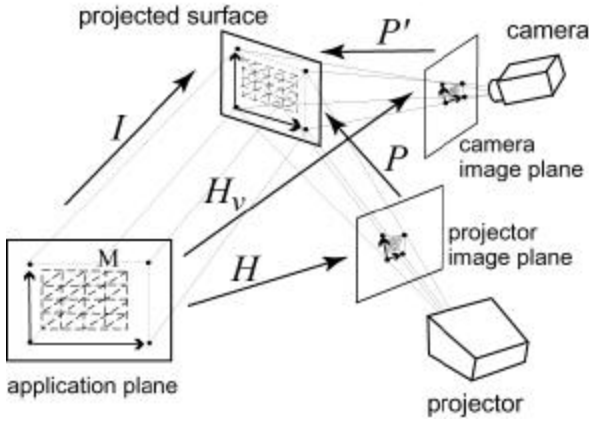


Figure 4. The transformations between the application plane, the projector's image plane, the camera's image plane and the projected surface.

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

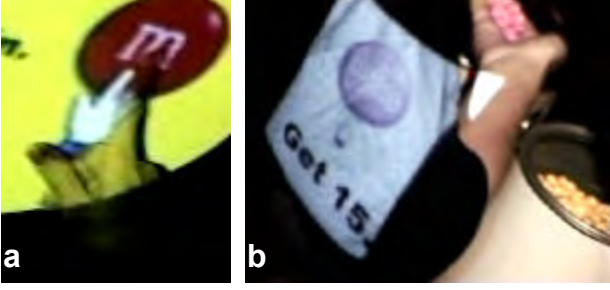


Figure 5. Difficulties faced by the vision system: a) the user's hand is partially hidden by the strong projection; b) the interactive surface is occluded by the user's shoulder as he moves to gather objects from inside a bucket.

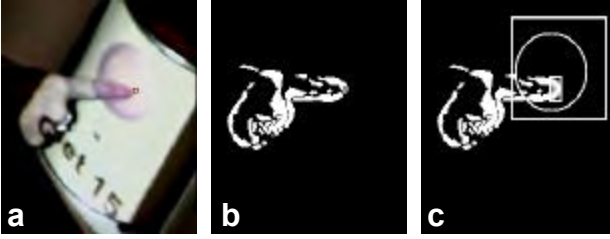


Figure 6. Detecting a touch: a) camera view of an interaction with a bucket; b) image difference data; c) overlay of search region (square), button active area (circle), and the fingertip template shown at the pointing location.

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 1/4 second and stop within the button, or start moving from the button and move in a consistent direction for 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.

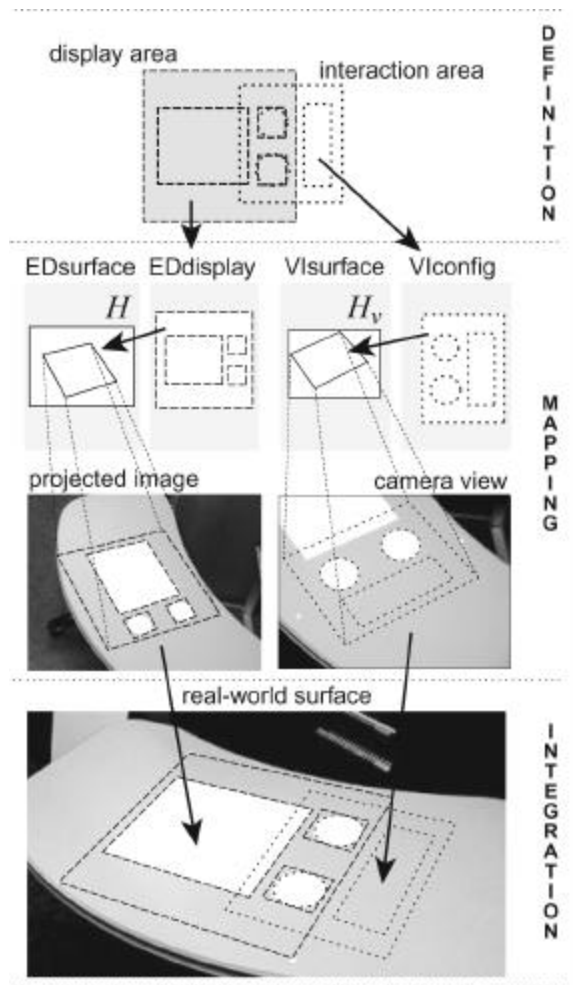


Figure 7. Process of creating an interactive surface on a table.

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.

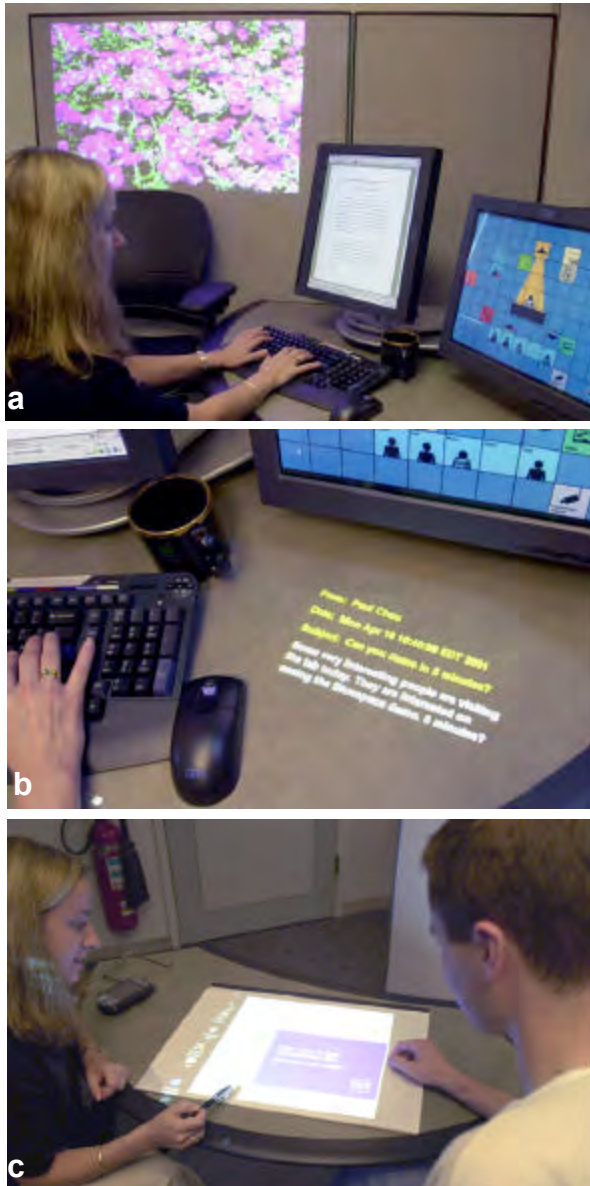


Figure 8. Uses of the ED-projector in an office application: a) to create dynamic wallpapers; b) to provide silent notification of e-mail; and c) to create a computer desktop on a working surface.

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 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

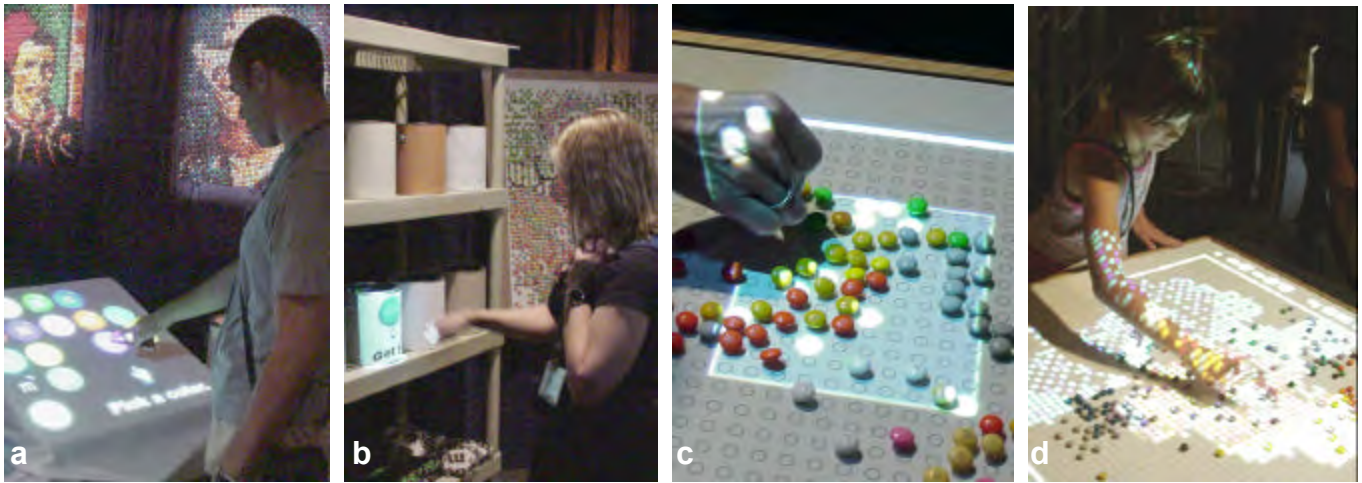


Figure 9. The M&M picture assembly task: a) choosing a color; b) “clicking” a bucket after getting the M&Ms; c) placing the M&Ms on the highlighted areas; and d) “finger painting” to reveal the complete picture.

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, PUI'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.

<https://web.archive.org/web/20080423153608/http://www.research.ibm.com/people/p/pinhanez/publications/procams03.pdf>

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

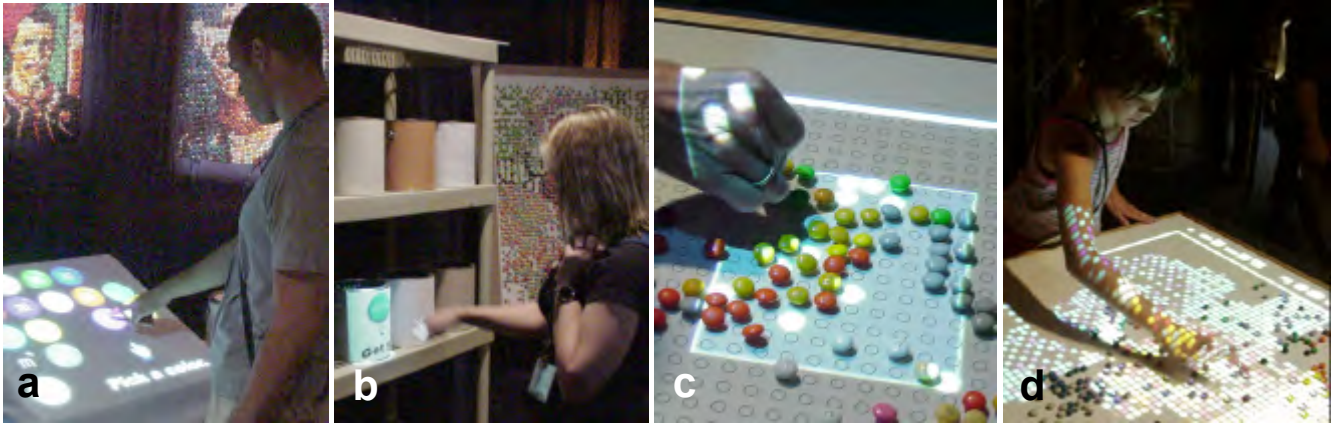


Figure 2. The M&M picture assembly task: a) choosing a color; b) “clicking” a bucket after getting the M&M’s; c) placing the M&M’s on the highlighted areas; and d) “finger painting” to reveal the complete picture.

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&Ms 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

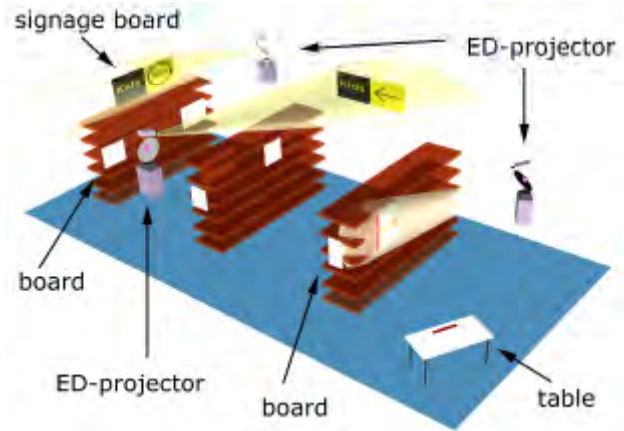


Figure 5 Concept of ubiquitous product finder.

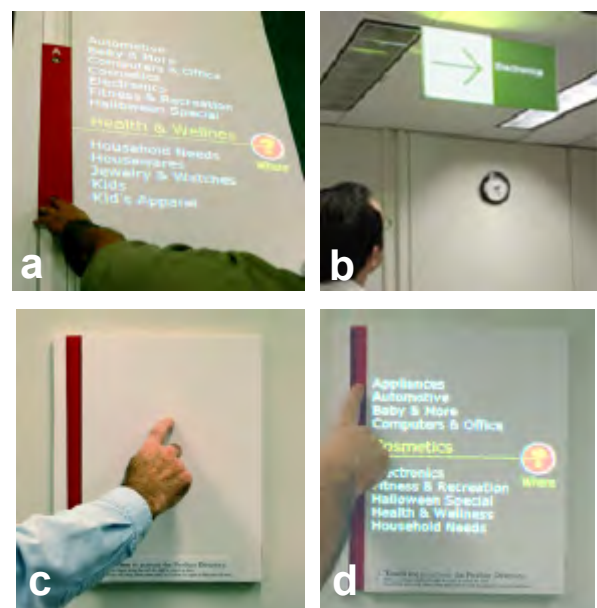


Figure 4. Prototype of ubiquitous product finder.

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, "Pfnder: 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, PUI'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 (PUI'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.

<https://web.archive.org/web/20050410205612/http://www.research.ibm.com/ed/publications/fg02.pdf>

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.

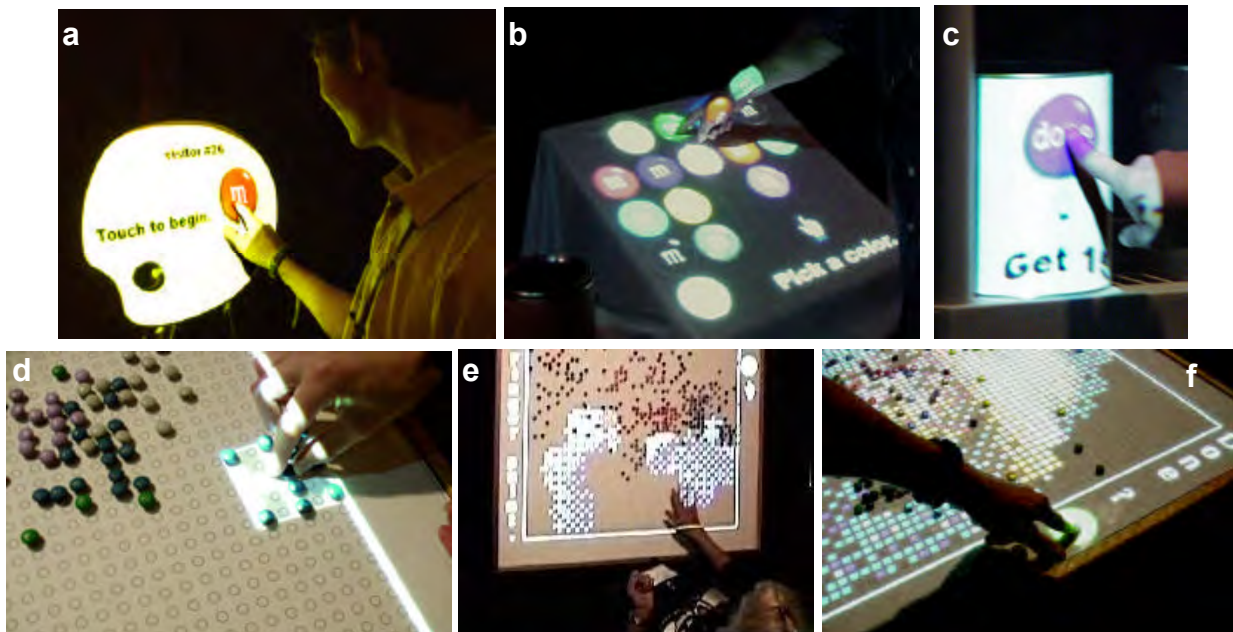


Fig. 2: Steps followed by each user when placing the M&M's in the SIGGRAPH'01 demo: a) the entrance panel is touched to start the process; b) a color is selected; c) the bucket containing the selected color is projected upon and the "done" button bucket is touched after the M&M's are picked up; e) the 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^{®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&M's of that color. It displays the message "Get n ", (where n is the number of M&M's 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.

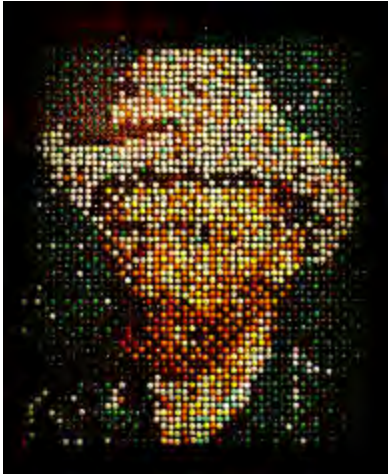


Fig. 3: A picture made of M&M's

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

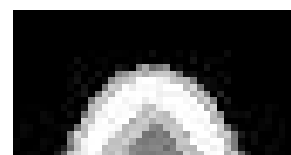


Fig. 5: Fingertip template.

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.