DECLARATION OF JUNE ANN MUNFORD

1. My name is June Ann Munford. I am over the age of 18, have personal knowledge of the facts set forth herein, and am competent to testify to the same.

2. I earned a Master of Library and Information Science (MLIS) from the University of Wisconsin-Milwaukee in 2009. I have over ten years of experience in the library/information science field. Beginning in 2004, I have served in various positions in the public library sector including Assistant Librarian, Youth Services Librarian and Library Director. I have attached my Curriculum Vitae as Appendix CV.

3. During my career in the library profession, I have been responsible for materials acquisition for multiple libraries. In that position, I have cataloged, purchased and processed incoming library works. That includes purchasing materials directly from vendors, recording publishing data from the material in question, creating detailed material records for library catalogs and physically preparing that material for circulation. In addition to my experience in acquisitions, I was also responsible for analyzing large collections of library materials, tailoring library records for optimal catalog

search performance and creating lending agreements between libraries during my time as a Library Director.

4. I am fully familiar with the catalog record creation process in the library sector. In preparing a material for public availability, a library catalog record describing that material would be created. These records are typically written in Machine Readable Catalog (herein referred to as "MARC") code and contain information such as a physical description of the material, metadata from the material's publisher, and date of library acquisition. In particular, the 008 field of the MARC record is reserved for denoting the date of creation of the library record itself. As this typically occurs during the process of preparing materials for public access, it is my experience that an item's MARC record indicates the date of an item's public availability.

5. Typically, in creating a MARC record, a librarian would gather various bits of metadata such as book title, publisher and subject headings among others and assign each value to a relevant numerical field. For example, a book's physical description is tracked in field 300 while title/attribution is tracked in field 245. The 008 field of the MARC record is reserved for denoting the creation of the library record itself. As this is the only date reflecting the inclusion of said materials within the library's collection, it is my experience

that an item's 008 field accurately indicates the date of an item's public availability.

6. I am familiar with the Internet Archive, a digital library accessible to the general public. Among other services that the Internet Archive makes available to the general public is the Wayback Machine, an online archive of the web. The Internet Archive's Wayback Machine service archives webpages as of a certain capture date to track changes in the web over time. The Internet Archive has been in operation as a nonprofit library since 1996 and has hosted the Wayback Machine service since its inception in 2001. During my time as a librarian, I frequently used the Internet Archive's Wayback Machine for research and instruction purposes. This includes teaching instructional classes on using the Wayback Machine to library patrons and using the Wayback Machine to research reference inquiries that require hard-to-find online resources. I consider the Internet Archive's recordskeeping to be as rigorous and detailed as other formal library recordskeeping practices such as MARC records, OCLC records and Dublin Core.

7. I have reviewed Exhibit 1009, "BilVideo-7: An MPEG-7 Compatible Video Indexing and Retrieval System" by Muhammet Bastan, et al as published in IEEE MultiMedia September 2010.

8. Attached hereto as Appendix BASTAN01 is a true and correct copy of the MARC record for IEEE MultiMedia as held by the Penn State University Library. I secured this record myself from the library's public catalog. The MARC record contained within Appendix BASTAN01 accurately describes the title, publisher and ISSN number of IEEE MultiMedia. The Showing Results field present on page 3 of this record indicates the library's collection includes the September 2010 issue of IEEE MultiMedia containing "BilVideo-7: An MPEG-7 Compatible Video Indexing and Retrieval System".

9. Attached hereto as Appendix BASTAN02 is a true and correct copy of "BilVideo-7: An MPEG-7 Compatible Video Indexing and Retrieval System" as published in IEEE MultiMedia as held by the Penn State University Library. These scans were secured from the library's collection. In comparing Exhibit 1009 to Appendix BASTAN02, it is my determination that Exhibit 1009 is a true and correct copy of "BilVideo-7: An MPEG-7

Compatible Video Indexing and Retrieval System" by Muhammet Bastan, et al as published in IEEE MultiMedia.

10. The 008 field of the MARC record in Appendix BASTAN01 indicates the date of record creation. The 008 field of Appendix BASTAN01 indicates the Penn State University Library first acquired IEEE MultiMedia as of July 13, 2000 and held the journal in perpetuity until the publication of "BilVideo-7: An MPEG-7 Compatible Video Indexing and Retrieval System". Considering this information, it is my determination that IEEE MultiMedia September 2010 and therefore "BilVideo-7: An MPEG-7 Compatible Video Indexing and Retrieval System" was made available to the public shortly after its initial publication in September 2010.

11. I have reviewed Exhibit 1005, "A Web-Enabled Video Indexing System" by Jian Zhou, et al as published in Proceedings of MIR '04.

12. Attached hereto as Appendix ZHOU01 is a true and correct copy of the MARC record for Proceedings of MIR '04 as held by the Penn State University Library. I secured this record myself from the library's public catalog. The MARC record contained within Appendix ZHOU01 accurately describes the title, publisher and ISBN number of Proceedings of MIR '04.

13. Attached hereto as Appendix ZHOU02 is a true and correct copy of "A Web-Enabled Video Indexing System" as published in Proceedings of MIR '04 as held by the Penn State University Library. These scans were secured from the library's collection. In comparing Exhibit 1005 to Appendix ZHOU02, it is my determination that Exhibit 1005 is a true and correct copy of "A Web-Enabled Video Indexing System" by Jian Zhou, et al as published in Proceedings of MIR '04.

14. The 008 field of the MARC record in Appendix ZHOU01 indicates the date of record creation. The 008 field of Appendix ZHOU01 indicates the Penn State University Library first acquired Proceedings of MIR '04 as of June 15, 2005. Considering this information, it is my determination that Proceedings of MIR '04 and therefore "A Web-Enabled Video Indexing System" was made available to the public at least as of June 15, 2005.

15. I have been retained on behalf of the Petitioner to provide assistance in the above-illustrated matter in establishing the authenticity and public availability of the documents discussed in this declaration. I am being compensated for my services in this matter at the rate of $200.00 per hour

plus reasonable expenses. My statements are objective, and my compensation does not depend on the outcome of this matter.

16. I declare under penalty of perjury that the foregoing is true and correct. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

Dated: 8/8/2025

June Ann Munford

Appendix CV

June A. Munford
Curriculum Vitae

Education

University of Wisconsin-Milwaukee - MS, Library & Information Science, 2009
Milwaukee, WI

- Coursework included cataloging, metadata, data analysis, library systems, management strategies and collection development.
- Specialized in library advocacy, cataloging and public administration.

Grand Valley State University - BA, English Language & Literature, 2008
Allendale, MI

- Coursework included linguistics, documentation and literary analysis.
- Minor in political science with a focus in local-level economics and government.

Professional Experience

Researcher / Expert Witness, October 2017 – present
Freelance ● Pittsburgh, Pennsylvania & Grand Rapids, Michigan

- Material authentication and public accessibility determination. Declarations of authenticity and/or public accessibility provided upon research completion. Experienced with appeals and deposition process.

- Research provided on topics of public library operations, material publication history, digital database services and legacy web resources.

- Past clients include Alston & Bird, Arnold & Porter, Baker Botts, Fish & Richardson, Erise IP, Irell & Manella, O'Melveny & Myers, Perkins-Coie, Pillsbury Winthrop Shaw Pittman and Slayden Grubert Beard.

Library Director, February 2013 - March 2015
Dowagiac District Library ● Dowagiac, Michigan

- Executive administrator of the Dowagiac District Library. Located in Southwest Michigan, this library has a service area of 13,000, an annual operating budget of over $400,000 and total assets of approximately $1,300,000.

● Developed careful budgeting guidelines to produce a 15% surplus during the 2013-2014 & 2014-2015 fiscal years while being audited.

● Using this budget surplus, oversaw significant library investments including the purchase of property for a future building site, demolition of existing buildings and building renovation projects on the current facility.

● Led the organization and digitization of the library's archival records.

● Served as the public representative for the library, developing business relationships with local school, museum and tribal government entities.

● Developed an objective-based analysis system for measuring library services - including a full collection analysis of the library's 50,000+ circulating items and their records.

November 2010 - January 2013
Librarian & Branch Manager, Anchorage Public Library ● Anchorage, Alaska

● Headed the 2013 Anchorage Reads community reading campaign including event planning, staging public performances and creating marketing materials for mass distribution.

● Co-led the social media department of the library's marketing team, drafting social media guidelines, creating original content and instituting long-term planning via content calendars.

● Developed business relationships with The Boys & Girls Club, Anchorage School District and the US Army to establish summer reading programs for children.

June 2004 - September 2005, September 2006 - October 2013
Library Assistant, Hart Area Public Library
Hart, MI

● Responsible for verifying imported MARC records and original MARC cataloging for the local-level collection as well as the Michigan Electronic Library.

● Handled OCLC Worldcat interlibrary loan requests & fulfillment via ongoing communication with lending libraries.

Professional Involvement

Alaska Library Association - Anchorage Chapter

● Treasurer, 2012

Library Of Michigan
● Level VII Certification, 2008
● Level II Certification, 2013

Michigan Library Association Annual Conference 2014
● New Directors Conference Panel Member

Southwest Michigan Library Cooperative
● Represented the Dowagiac District Library, 2013-2015

Professional Development

Library Of Michigan Beginning Workshop, May 2008
Petoskey, MI
● Received training in cataloging, local history, collection management, children's literacy and reference service.

Public Library Association Intensive Library Management Training, October 2011
Nashville, TN
● Attended a five-day workshop focused on strategic planning, staff management, statistical analysis, collections and cataloging theory.

Alaska Library Association Annual Conference 2012 - Fairbanks, February 2012
Fairbanks, AK
● Attended seminars on EBSCO advanced search methods, budgeting, cataloging, database usage and marketing.

Depositions

2019 ● Fish & Richardson

Apple v. Qualcomm (Case No. IPR2018-001281, 39521-00421IP, IPR2018-01282 and 39521-00421IP2)

2019 ● Erise IP

Implicit, LLC v. Netscout Systems, Inc (Case No. 2:18-cv-53-JRG)

2019 ● Perkins-Coie

Adobe Inc. v. RAH Color Technologies LLC (Case No. IPR2019-00627, IPR2019-00628, IPR2019-00629 and IPR2019-00646)

2020 ● O'Melveny & Myers

Maxell, Ltd. v. Apple Inc. (Case No. 5:19-cv-00036-RWS)

2021 ● Pillsbury Winthrop Shaw Pittman LLP

Intel v. SRC (Case No. IPR2020-1449)

2022 ● Perkins-Coie

Realtek v. Future Link (Case No. IPR2021-01182)

2023 ● Fish & Richardson

Neuroderm Ltd. v. Abbvie, Inc (Case No. PGR2022-00040)

2023 ● Fish & Richardson

Nearmap US Inc. v. Pictometry International Corp. (Case No. IPR2022-00735)

2023 ● Fish & Richardson

Samsung Electronics v. MemoryWeb LLC (Case No. 39843-0136PS1)

2023 ● Pillsbury Winthrop Shaw Whitman LLP

Gravel Rating Systems v. Costco Wholesale Corp. (Case No. 4:21-cv-149-ALM)

2024 ● Willkie-Farr

Netflix, Inc. v. VideoLabs. Inc. (Case No. IPR2023-00628)

2024 ● Quinn Emanuel Urquhart & Sullivan

Purdue University v. Wolfspeed (Case No. 1:21-CV-840)

2024 ● Fish & Richardson

Dish Network v. Entropic Communications (Case No. IPR2024-00393)

2025 ● Perkins Coie

Amazon.com, Inc. v. Nokia Technologies Oy (Case No. IPR2024-00572)

2025 ● Fish & Richardson

Dish Network v. Entropic Communications (Case No. IPR2024-00373)

Limited Case and Clientele History

Alston & Bird

● Ericsson

v. Collision Communications (Case No. IPR2022-01233)

● Nokia

v. Neptune Subsea, Xtera (Case No. 1:17-cv-01876)

● Universal Electronics Inc

v. Roku Inc (Case No. IPR2022-00818)

Alavi & Anaipakos PLLC

● Stingray Group Inc. and Stingray Music USA, Inc.

v. Edwin A. Hernandez-Mondragon and Egla Corp. (Case No. 1:24-cv-21226-RAR)

Arnold & Porter

● Ivantis

v. Glaukos (Case No. 8:18-cv-00620)

● Samsung

v. Jawbone (Case No. 2:21-cv-00186)

Benesch Friedlander Coplan & Aronoff

- Voyis

  v. Cathx (Case No. 5:21-cv-00077-RWS)

Buchanan, Ingersoll & Rooney PC

- Google LLC, AT&T Service Inc., T Mobile USA Inc., Cellco Partnership, Ericsson Inc. & Nokia of America Corp.

  v. KT Corp & Pegasus Wireless Innovation LLC (Case No. IPR2025-00293)

Deschert LLP

- Smaxtec Animal Care

  v. ST Reproductive Technologies, LLC (Case No. IPR2024-00885)

Duane Morris

- Kangxi Communication Technologies

  v. Skyworks Solutions Inc. (Case No. IPR2025-00373)

- Microsoft

  v. Edge Networking Systems LLC (Case No. 1:24-cv-00215)

Erise I.P.

- Apple

  v. Ericsson Inc. (Case No. IPR2022-00715)

  v. Future Link Systems (Case No. IPRs 6317804, 6622108, 6807505, and 7917680)

  v. INVT (Case No. 20-1881)

  v. Navblazer LLC (Case No. IPR2020-01253)

  v. Qualcomm (Case No. IPR2018-001281, 39521-00421IP, IPR2018-01282, 39521-00421IP2)

v. Quest Nettech Corp (Case No. 2:19-cv-00118-JRG)

v. Telefonaktiebolaget LM Ericsson (Case No. IPR2022-00275)

v. Theta IP, LLC (Case No. IPR2024-00818)

v. THL Holding Company LLC (Case No. 1:23-cv-00548)

- Fanduel

  v. CGT (Case No. 19-1393)

- Garmin

  v. Phillips North America LLC (Case No. 2:19-cv-6301-AB-KS)

- Netscout

  v. Longhorn HD LLC (Case No. 2:20-cv-00349)

  v. Implicit, LLC (Case No. 2:18-cv-53-JRG)

- Sony Interactive Entertainment LLC

  v. Bot M8 LLC (Case No. IPR2020-01288)

  v. Infernal Technology LLC (Case No. 2:19-CV-00248-JRG)

- Tesla

  v. Charge Fusion Technologies LLC (Case No. IPR2025-00032)

- Unified Patents

  v. GE Video Compression (Case No. 2:19-cv-248)

Fish & Richardson

- Apple

  v. AliveCor (Case No. 3:21-cv-03958)

  v. LBS Innovations (Case No. 2:19-cv-00119-JRG-RSP)

  v. Koss Corporation (Case No. IPR2021-00305)

v. Masimo (Case No. IPR 50095-0012IP1, 50095-0012IP2, 50095-0013IP1, 50095-0013IP2, 50095-0006IP1, 50095-0135IP1)

v. Neonode (Case No. 21-cv-08872-EMC)

v. Qualcomm (Case No. IPR2018-001281, 39521-00421IP, IPR2018-01282, 39521-00421IP2)

v. Resonant Sys. (Case No. 7:23-cv-00077-ADA)

- AsusTek Computer Inc.
  v. Videolabs, Inc. (Case No. 22-CV-00720-ADA)

- Dell
  v. Neo Wireless (Case No. IPR2022-00616)

- Dish Network

  v. Entropic Communications, LLC (Case No. 2:2023-CV-01043)

  v. Entropic Communications LLC (Case No. IPR2024-00393)

  v. Realtime Adaptive Streaming (Case No. 1:17-CV-02097-RBJ)

  v. TQ Delta LLC (Case No. 18-1798)

- Evapco Dry Cooling

  v. SPG Dry Cooling (Case No. IPR2021-00688)

- Genetec

  v. Sensormatic Electronics (Case No. 1:20-CV-00760)

- Huawei

  v. Bell Northern Research LLC (Case No. IPR2019-01174)

- Kianxis

  v. Blue Yonder (Case No. 3:20-cv-03636)

- LG Electronics

  v. Bell Northern Research LLC (Case No. 3:18-cv-2864-CAB-BLM)

- MediaTek

  v. MOSAID (Case No. IPR2024-00718)

- Metaswitch

  v. Sonus Networks (Case No. IPR2018-01719)

- Microsoft

  v. Throughputer Inc (Case No. IPR2022-00757)

- Mom Enterprises

  v. Ddrops Company (Case No. 1:22-cv-00332)

- MLC Intellectual Property

  v. MicronTech (Case No. 3:14-cv-03657-SI)

- Nearmap Inc

  v. EagleView Technologies (Case No. IPR2022-01009)

- Neuroderm Ltd.

  v. Abbvie, Inc (Case No. PGR2022-00040)

- Posco Co Ltd.

  v. Pascal Drillet et al. (Case No. IPR2025-00370)

- Realtek Semiconductor

  v. Future Link (Case No. IPR2021-01182)

- Quectel
  - v. Koninklijke Philips (Case No. 1:20-cv-01710)

- Samsung
  - v. Aire Technology (Case No. IPR2022-00877)
  - v. Bell Northern Research (Case No. 2:19-cv-00286-JRG)
  - v. Communication Technologies Inc (Case No. IPR2022-01221)
  - v. Jawbone Innovations (Case No. IPR2022-00865)
  - v. MemoryWeb LLC (Case No. IPR2022-00885)
  - v. Dodots Licensing Solutions (Case No. IPR2023-00701)
  - v. Telefonaktiebolaget LM Ericsson (Case No. IPR2021-00615)

- Texas Instruments
  - v. Vantage Micro LLC (Case No. IPR2020-01261)

- Xilinx
  - v. Sentient Sensors LCC (Case No. 1:22-cv-00173)

Goodwin Proctor

- Intel
  - v. Proxense, LLC (Case No. 6:24-cv-00283)

- Samsung Electronics
  - v. Mobile Data Technologies (Case No. 2:24-cv-00435)

- Taiwan Semiconductor Manufacturing Company Ltd.
  - v. Advanced Integrated Circuit Process LLC (Case No. 2:24-cv-00623)

Groombridge Wu

- Nearmap

  v. Eagle View Technologies Inc. and Pictometry International Corp. (Case No. 2:21-cv-00283-TS-DAO)

Irell & Manella

- Curium

Latham & Watkins LLP

- Apple

  v. Advanced Coding Research LLC

Leydig, Voit & Mayer, Ltd.

- Public Availability Consultancy

O'Melveny & Myers

- Apple

  v. Maxell (Case No. 5:19-cv-00036-RWS)

- Disney Media & Entertainment Distribution LLC

  v. Digital Media Technology Holdings, LLC (Case No. 2:22-cv-01642)

- Micron Technology Inc.

  v. Besang Inc (Case No. IPR2023-00900)

- Samsung

  v. Daedalus Prime LLC (1335 Investigation)

Perkins-Coie

- Amazon
  - v. Nokia (Case No. IPR2024-00847 and IPR2024-00848)
  - v. Nokia Technologies Oy (Case No. IPR2024-00572)
- Heru Industries
  - v. The UAB Research Foundation (Case No. IPR2022-01148)
- Intel Corporation
  - v. BeSang Inc. (Case No. IPR2023-00991)
  - v. BitMICRO (Case No. 5:23-cv-00625)
  - v. Telefonaktiebolaget L M Ericsson (Case No. IPR2024-00610)
- Realtek Semiconductor
  - v. Future Link (Case No. IPR2021-01182)
- r-pac
  - v. Adasa, Inc. (Case No. 1:2024cv06102)
- Twitter Inc
  - v. VOIP-Pal.com (Case No. 3:20-cv-02397-JD)
- TCL Industries
  - v. Koninklijke Philips NV (Case No. IPR2021-00495, IPR2021-00496 and IPR2021-00497)
- VusionGroup
  - v. Hanshow (Case No. IPR2024-00857)

Pillsbury Winthrop Shaw Pittman

- Intel

  v. FG SRC LLC (Case No. 6:20-cv-00315)

- Gravel Rating Systems

  v. Costco (Case No. 4:21-cv-149)
  v. Lowe's Home Centers (Case No. 4:21-cv-150)
  v. T-Mobile USA (Case No. 4:21-cv-152)
  v. Kohl's Inc. (Case No. 4:21-cv-258)
  v. Under Armor (Case No. 4:21-cv-356)

Quinn Emanuel

- Exact Sciences Corporation

  v. Geneoscopy, Inc. (Case No. IPR2024-00459 and IPR2024-1330)

- ServiceNow, Inc.

  v. InQuisient, Inc. (Case No. 22-900-CJB)

- Wolfspeed

  v. The Trustees of Purdue University (Case No. 1:2021-cv-00840)

Sheppard, Mullin, Richter & Hampton LLP

- Advanced Micro Devices, Inc and Pensando Systems, Inc.

  v. Concurrent Ventures LLC and Xtreamedge, Inc. (Case No. IPR2025-00478)

- Cadence Design Systems Inc.

  v. Semiconductor Design Technologies (Case No. 3:23-cv-01001)

Shook, Hardy & Bacon LLP

- HP Inc.

  v. Universal Connectivity Technologies (Case No. 1:23-cv-01177)

Stern, Kessler, Goldstein & Fox

- Preliminary Research

Wilmer, Cutler, Pickering, Hale and Dorr

- Apple

  v. Shunock (Case No. 1:23-cv-08598)

- Quest Diagnostics Inc., Haystack Oncology Inc. and Haystack Oncology GmbH

  v. Natera, Inc. (Case No. 1:23-cv-08598)

- Roche Diabetes, Inc.

  v. Trividia Health, Inc. (Case No. 1:24-cv-00668)

Willkie, Farr & Gallagher LLP

- Lenovo, Dell, HP

  v. Universal Connectivity Technologies Inc. (Case No. 2:2023cv00449)

- Neurent

  v. The Foundry, LLC (Case No. IPR2024-00669)

- Netflix, Inc.

  v. VideoLabs, Inc. (Case No. IPR2023-00628)

Appendix BASTAN01

PennState University Libraries | Catalog

Bookmarks 0 Course Reserves My Account

Keyword Search... Search Advanced search Start Over

Share Bookmark Report an Issue

MARC View

LEADER 01938nas a22004817a 4500

001				48266976
003				SIRSI
005				20250714092815.0
006				m o d
007				cr n
008				000713c19949999nyuqr pss 0 a0eng c
010				a| 2008212915
016	7			a| 012433248 2| Uk
019				a| DO NOT EDIT this 360 MARC RECORD
022				a| 1941-0166 l| 1070-986X y| 1070-986X
035				a| (WaSeSS)ssj0003697
037				b| IEEE Service Center, 445 Hoes Ln., POB 1331, Piscataway, NJ 08855-1331
040				a| NTD b| eng c| NTD d| OCL d| OCLCQ d| NSD d| U9S d| OCLCQ d| OCLCE d| OCLCQ d| OCLCF d| OCLCO d| OCLCQ d| OCLCA d| UKMGB d| VT2 d| OCLCO d| OCLCQ d| UAB d| HNC d| OCLCL d| WaSeSS
042				a| pcc a| nsdp
050			4	a| QA76.575 b| .I358
060			4	a| INTERNET
082	1	0		a| 006 2| 14 1| https://id.oclc.org/worldcat/ddc/E379xBCBchbdmrqYKbR6BJFPFX
130	0			a| IEEE multimedia (Online)
210	0			a| IEEE multimed. b| (Online)
222		0		a| IEEE multimedia b| (Online)
245	1	0		a| IEEE multimedia h| [electronic resource].
246	3	0		a| Multimedia
260				a| Los Alamitos, CA : b| IEEE Computer Society, c| ©1994-
310				a| Quarterly
362	0			a| Vol. 1, no. 1 (spring 1994)-
500				a| Title from IEEExplore homepage (viewed.
538				a| Mode of access: World Wide Web.
650		0		a| Multimedia systems v| Periodicals.
650		0		a| Computer networks v| Periodicals.
650		0		a| Multimedia systems x| Periodicals.
650		0		a| Computer networks x| Periodicals.
655		0		a| Electronic journals.
710	2			a| Institute of Electrical and Electronics Engineers.
710	2			a| IEEE Computer Society.
776	0	8		i| Print version: t| IEEE multimedia x| 1070-986X w| (DLC) 94648260 w| (OCoLC)28524907
856	4	0		u| http://SK8ES4MC2L.search.serialssolutions.com/?V=1.0&L=SK8ES4MC2L&S=JCs&C=IEEEMUL&T=marc
910				a| CONSER online
949				a| Electronic resource w| ASIS c| 1 i| 48266976-1001 l| ONLINE m| ONLINE r| Y s| Y t| ONLINE u| 7/19/2025 o| .STAFF. 360 MARC

ASK A LIBRARIAN

PennState

Copyright ©2019 The Pennsylvania State University. All rights reserved. Except where otherwise noted, this work is subject to a Creative Commons Attribution 4.0 license. Details and exceptions.

Legal Statements | PSU Hotlines

PENN STATE UNIVERSITY LIBRARIES

Libraries Home
Libraries Intranet (Staff Only)
Accessibility Help
Website Feedback
Policies and Guidelines
Staff Directory

(814) 865-6368

CONNECT WITH PENN STATE UNIVERSITY LIBRARIES

Facebook
Twitter
Instagram

IEEE multimedia [electronic resource].

Uniform Title	IEEE multimedia (Online)
Additional Titles	IEEE multimed. (Online) and Multimedia
Published	Los Alamitos, CA : IEEE Computer Society, ©1994-
Additional Creators	Institute of Electrical and Electronics Engineers and IEEE Computer Society

Access Online
serialssolutions.com

Dates of Publication and/or Sequential Designation
Vol. 1, no. 1 (spring 1994)-

Subject(s)
Multimedia systems—Periodicals
Computer networks—Periodicals

Genre(s)
Electronic journals

ISSN
1941-0166

Publication Frequency
Quarterly

Note
Title from IEEExplore homepage (viewed.

Technical Details
Mode of access: World Wide Web.

ASK A LIBRARIAN

View MARC record | catkey: 48266976

PennState

Copyright ©2019 The Pennsylvania State University. All rights reserved. Except where otherwise noted, this work is subject to a Creative Commons Attribution 4.0 license. Details and exceptions.

Legal Statements | PSU Hotlines

PENN STATE UNIVERSITY LIBRARIES
Libraries Home
Libraries Intranet (Staff Only)
Accessibility Help
Website Feedback
Policies and Guidelines
Staff Directory
(814) 865-6368

CONNECT WITH PENN STATE UNIVERSITY LIBRARIES
Facebook
Twitter
Instagram

Online Journals, Books and more

All Journals Only Books Only Videos Only

Title begins with Search

Show more search options ▼

Showing results 1 through 1 of 1

for the search: "IEEEMUL"

Refine Results: All | Journals Only | Books Only | Videos Only

Limit by: Peer Reviewed | Open Access

IEEE multimedia

ISSN: 1070-986X

Peer Reviewed

1994 to Present in IEEE Electronic Library (IEL)

Search inside this journal Search

Back to Top

Copyright ©2025 The Pennsylvania State University.
All rights reserved. Except where otherwise noted, this work is subject to a Creative Commons Attribution 4.0 license. Details and exceptions.

Last Updated on 07-31-2025 © Ex Libris, a ProQuest Company, 2000-2025

Appendix BASTAN02

BilVideo-7: An MPEG-7-Compatible Video Indexing and Retrieval System

Muhammet Baştan, Hayati Çam, Uğur Güdükbay, and Özgür Ulusoy
Bilkent University

BilVideo-7 is an MPEG-7-compatible, distributed, video indexing and retrieval system that supports complex multimodal queries in a unified framework.

Early prototype multimedia database management systems used the query-by-example paradigm to respond to user queries. Users needed to formulate their queries by providing examples or sketches. The query-by-keyword paradigm, on the other hand, has emerged due to the desire to search multimedia content in terms of semantic concepts using keywords or sentences rather than low-level multimedia descriptors. After all, it's much easier to formulate some queries by using keywords. However, some queries are still easier to formulate by examples or sketches—for example, the trajectory of a moving object.

Moreover, there is the so-called semantic gap problem—the disparity between low-level representation and high-level semantics. This gap makes it difficult to build multimedia systems that are capable of effectively supporting keyword-based semantic queries while also being able to interpret an acceptable number of semantic concepts. The consequence is the need to support both query paradigms in an integrated way so that users can formulate queries containing both high-level semantic concepts and low-level descriptors. Another important issue to be considered is interoperability, which is especially crucial for distributed architectures if the system is to be used by multiple heterogeneous clients. In the system described here, we use the MPEG-7 standard as a multimedia content-description interface to address this issue.[1]

The design of a retrieval system is directly affected by the type of queries to be supported, especially because the types of descriptors and the granularity of the representation determine the system's speed and accuracy. Some example video query types that might be attractive for most users, but which also are not supported by the existing systems in an MPEG-7-compatible framework, include the following:

- *Content-based queries by examples*. The user specifies an image, an image region, or a video segment and the system returns video segments that are similar to the input query.
- *Text-based semantic queries*. Queries are specified by a set of keywords corresponding to high-level semantic concepts and relations between them.
- *Spatiotemporal queries*. Queries are related to spatial and temporal locations of objects and video segments within the video.
- *Composite queries*. These queries contain any combination of other simple queries. The user composes the query by putting together image and video segments and specifying their properties, then asks the system to retrieve similar ones from the database.

We developed BilVideo-7 as a comprehensive, MPEG-7-compatible system to support such multimodal queries in a video indexing and retrieval framework. The name BilVideo-7 derives from BilVideo, which was a prototype video database system that supported keyword-based spatiotemporal queries using a knowledge base and a Prolog inference engine.[2]

To create BilVideo-7, we designed an MPEG-7 profile for video representation that enables detailed queries on videos, and used our

029

1070-986X/10/$26.00 © 2010 IEEE

Published by the IEEE Computer Society

MPEG-7-compatible video feature extraction and annotation tool to obtain MPEG-7-compatible video representations. The BilVideo-7's visual query interface, which is designed to help formulate complex multimodal queries, supports a comprehensive set of MPEG-7 descriptors. The queries for this system are processed on a multithreaded query-processing server with a multimodal query-processing architecture suitable for parallelization.

MPEG-7 foundation

MPEG-7 is an ISO/IEC standard developed by Moving Picture Experts Group (MPEG), the committee that also developed the standards MPEG-1, MPEG-2, and MPEG-4. Different from the previous MPEG standards, MPEG-7 is designed to describe the content of multimedia. It is formally called the Multimedia Content Description Interface.

MPEG-7 offers a comprehensive set of audiovisual description tools in the form of descriptors and description schemes that describe the multimedia data, forming a common basis for applications. The Description Definition Language is based on W3C XML with some MPEG-7-specific extensions, such as vectors and matrices. MPEG-7 documents are XML documents that conform to particular MPEG-7 schemas for describing multimedia content. Descriptors describe features, attributes, or groups of attributes of multimedia content. Description schemes describe entities or relationships pertaining to multimedia content. They specify the structure and semantics of their components, which may be description schemes, descriptors, or data types.

The eXperimentation Model[3] is the software for all the reference code of the MPEG-7 standard. It implements the normative components of MPEG-7. MPEG-7 standardizes multimedia content descriptions but it doesn't specify how the description is produced. It's up to the developers of MPEG-7-compatible applications to determine how the descriptors are extracted from the multimedia, provided that the output conforms to the standard. MPEG-7 Visual Description Tools consist of basic structures and descriptors that cover the following basic visual features for multimedia content:

- Color descriptors, which include Color Structure Descriptor (CSD), Scalable Color Descriptor (SCD), Dominant Color Descriptor (DCD), Color Layout Descriptor (CLD), Face Recognition Descriptor (FRD), and Group-of-Frame (GoF) or Group-of-Picture (GoP) descriptor.
- Texture descriptors, which include Edge Histogram Descriptor (EHD), Homogeneous Texture Descriptor (HTD), and Texture Browsing Descriptor.
- Shape descriptors, which include Contour Shape Descriptor and Region Shape Descriptor (RSD).
- Motion descriptors, which include Motion Activity, Motion Trajectory (MTD), Camera Motion, and Parametric Motion Descriptors.
- Localization descriptors, which include Region Locator and Spatiotemporal Locator Descriptors.

In MPEG-7, the semantic content of multimedia (for example, objects, events, and concepts) can be described by text annotation (free text, keyword, and structured) and semantic-entity and semantic-relation tools. Free-text annotations describe the content using unstructured natural language text (for example, Barack Obama visits Turkey in April). Such annotations are easy for humans to understand but difficult for computers to process. Keyword annotations use a set of keywords (for example, Barack Obama, visit, Turkey, and April) and are easier to process by computers. Structured annotations strike a balance between simplicity (in terms of processing) and expressiveness. They consist of elements each answering one of the following questions: who, what object, what action, where, when, why, and how (for example, who: Barack Obama, what action: visit, where: Turkey, and when: April).

More detailed descriptions about semantic entities, such as objects, events, concepts, places and times, can be stored using semantic entity tools. The semantic relation tools describe the semantic relations between semantic entities using the normative semantic relations standardized by MPEG-7 (for example, agent, agentOf, patient, patientOf, result, resultOf, similar, opposite, user, userOf, location, locationOf, time, and timeOf) or by nonnormative relations.

MPEG-7's semantic tools provide methods to create brief or extensive semantic descriptions of multimedia content. Some of the descriptions can be obtained automatically, while most of them require manual labeling. For example, transcribed text obtained from automatic speech-recognition tools can be used as free-text annotations to describe video segments. Keyword and structured annotations can be obtained automatically to some extent using state-of-the-art auto-annotation techniques. Descriptions of semantic entities and relations between them, which cannot be obtained automatically with the current state-of-the-art technologies, require a considerable amount of manual work.

In 2007, MPEG adopted the MPEG Query Format[4] to provide a standard interface between clients and MPEG-7 databases for multimedia content-retrieval systems. The query format is based on XML and consists of three main parts:

- Input query format defines the syntax of query messages sent by a client to the server and supports different query by free text, query by description, query by XQuery, spatial query, temporal query, and so on.
- Output query format specifies the structure of the result set to be returned.
- Query management tools are used to search and choose the desired services for retrieval.

See the ''MPEG-7 Compatible Systems'' sidebar for related MPEG-7-compatible multimedia systems.

MPEG-7-compatible video representation

The first step in constructing an MPEG-7-compatible video-management system is to decide what query types to support and then to design an MPEG-7 profile. Representing video is crucial because it directly affects the system's performance. There is a trade-off between the accuracy of representation and the speed of access: more detailed representation will enable more detailed queries but will result in longer response time during retrieval. Keeping these factors in mind, we adapted our MPEG-7 profile from the one described in Bailer and Schallauer[5] to represent image, audio, and video collections.

Our profile corresponds to the video representation portion of the detailed audiovisual profile, with our own interpretation of what to represent with keyframes, still, and moving regions so our system can support a wide range of queries. First, audio and visual data are separated (using media-source decomposition[1]). Then, visual content is hierarchically decomposed into smaller structural and semantic units. Figure 1 (on page 66) shows an example of video decomposition according to this profile. First, the video is partitioned into non-overlapping video segments called *shots*, each having a start time and duration, with semantic annotation to describe the objects or events, and visual descriptors such as motion and group of frame or group of picture.

Next, the method performs temporal decomposition on the shots. The background content of the shots doesn't change much, especially if the camera is not moving. This static content can be represented by a single or a few keyframes. Therefore, each shot is decomposed into smaller, more homogeneous video segments (keysegments), which are represented by keyframes. Each keyframe is described by a temporal location, semantic annotation, and a set of visual descriptors. The visual descriptors are extracted from the frame as a whole. In addition, each keyframe is decomposed into a set of still regions (spatiotemporal decomposition) to keep more detailed region-based information in the form of spatial location by the minimum bounding rectangles (MBRs) of the region, semantic annotation, and region-based visual descriptors.

Each shot is also decomposed into a set of moving regions to represent the dynamic and more important content of the shots corresponding to the salient objects. Hence, more information can be stored for moving regions to enable more detailed queries about salient objects. We represent all salient objects with moving regions even if they are not moving. For example, faces are represented by moving regions, having an additional visual descriptor: the FRD.

Because the position, shape, motion and visual appearance of the salient objects might change throughout the shot, descriptors sampled at appropriate times should be stored. The trajectory of an object is represented by

MPEG-7 Compatible Systems

Although MPEG-7 was published in 2001, only a few MPEG-7-compatible multimedia systems have been developed so far. The comprehensiveness and flexibility of MPEG-7 allow its usage in a broad range of applications, but also increase its complexity and adversely affect interoperability. To overcome this problem, profiling has been proposed. An MPEG-7 profile is a subset of tools defined in MPEG-7, providing a particular set of functionalities for one or more classes of applications. One approach proposes an MPEG-7 profile for detailed description of audiovisual content that can be used in a broad range of applications.[1]

An MPEG-7-compatible database system extension to Oracle Database Management System is proposed in *MPEG-7 Multimedia Database System*.[2] The system is demonstrated by audio and image retrieval applications. Other research has proposed algorithms for the automatic generation of three MPEG-7 description schemes: video table of contents, for active video browsing; summary, to enable the direct use of metadata annotation of the producer; and still image, to allow interactive, content-based image retrieval.[3] Tseng et al. address the issues associated with designing a video personalization and summarization system in heterogeneous environments using MPEG-7 and MPEG-21.[4]

IBM's VideoAnnEx Annotation Tool (see http://www.research.ibm.com/VideoAnnEx) enables users to annotate video sequences with MPEG-7 metadata. Each shot is represented by a single keyframe and can be annotated with static scene descriptions, key object descriptions, event descriptions, and other custom lexicon sets that the user might provide. The tool is limited to concept annotation and cannot extract low-level MPEG-7 descriptors from the video. The M-OntoMat-Annotizer software aims at linking low-level MPEG-7 visual descriptions to conventional Semantic Web ontologies and annotations.[5] The visual descriptors are expressed in the Resource Description Framework. The IFinder system was developed to produce limited MPEG-7 representation from audio and video by speech processing, keyframe extraction, and face detection.[6] ERIC7 is a software testbed that implements content-based image retrieval using image-based MPEG-7 color, texture, and shape descriptors.[7] Caliph and Emir are MPEG-7 based Java prototypes for digital photo and image annotation and retrieval, supporting graph-like annotations for semantic metadata and content-based image retrieval using MPEG-7 descriptors.[8] Cao et al. describe a middleware solution to access a bundle of MPEG-7 based multimedia services from mobile devices.[9]

These MPEG-7 compatible systems have two major problems. Most of them use a coarse image or video representation, extracting low-level descriptors from whole images or video frames and annotating them, but ignoring region-level descriptors. This coarse representation limits the range of queries. Next, the user cannot perform complex multimodal queries by combining several video segments and descriptors in different modalities. BilVideo-7 addresses these two major problems by adopting an MPEG-7 profile with a more detailed video representation and using a multimodal query processing and bottom-up subquery result fusion architecture to support complex multimodal queries (such as composite queries) with a comprehensive set of MPEG-7 descriptors.

References

1. W. Bailer and P. Schallauer, ''Detailed Audiovisual Profile: Enabling Interoperability Between MPEG-7 Based Systems,'' *Proc. 12th Int'l Multi-Media Modelling Conf.,* IEEE Press, 2006, pp. 217-224.
2. M. Döller and H. Kosch, ''The MPEG-7 Multimedia Database System (MPEG-7 MMDB),'' *The J. Systems and Software,* vol. 81, no. 9, 2008, pp. 1559-1580.
3. Y. Rui, ''MPEG-7 Enhanced Ubi-Multimedia Access—Convergence of User Experience and Technology,'' *Proc. 1st IEEE Int'l Conf. Ubi-Media Computing,* IEEE Press, 2008, pp. 177-183.
4. B. Tseng, C.-Y. Lin, and J. Smith, ''Using MPEG-7 and MPEG-21 for Personalizing Video,'' *IEEE Multimedia,* vol. 11, no. 1, 2004, pp. 42-52.
5. K. Petridis et al., ''M-OntoMat-Annotizer: Image Annotation Linking Ontologies and Multimedia Low-Level Features,'' LNCS 4253, Springer, 2006, pp. 633-640.
6. J. Löffler et al., ''IFinder: An MPEG-7-Based Retrieval System for Distributed Multimedia Content,'' *Proc. 10th ACM Int'l Conf. Multimedia,* ACM Press, 2002, pp. 431-435.
7. L. Gagnon, S. Foucher, and V. Gouaillier, ''ERIC7: An Experimental Tool for Content-Based Image Encoding and Retrieval Under the MPEG-7 Standard,'' *Proc. Winter Int'l Symp. Information and Communication Technologies,* 2004, ACM Press, pp. 1-6.
8. M. Lux, J. Becker, and H. Krottmaier, ''Caliph&Emir: Semantic Annotation and Retrieval in Personal Digital Photo Libraries,'' *Proc. CAiSE Forum at 15th Conf. Advanced Information Systems Engineering,* Springer, 2003, pp. 85-89.
9. Y. Cao et al., ''Mobile Access to MPEG-7 Based Multimedia Services,'' *Proc. 10th Int'l Conf. Mobile Data Management,* ACM Press, 2009, pp. 102-111.

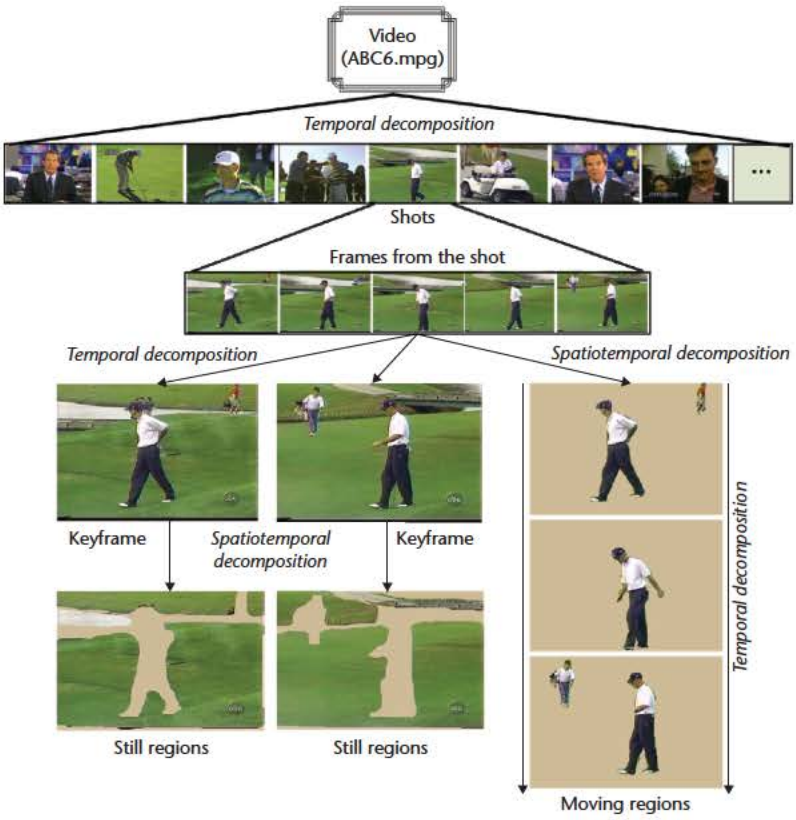


Figure 1. Decomposition of a video according to the MPEG-7 profile used in BilVideo-7. Low-level color, texture, and shape descriptors of the still and moving regions are extracted from the selected arbitrarily shaped regions, but the locations of the regions are represented by their minimum bounding rectangles.

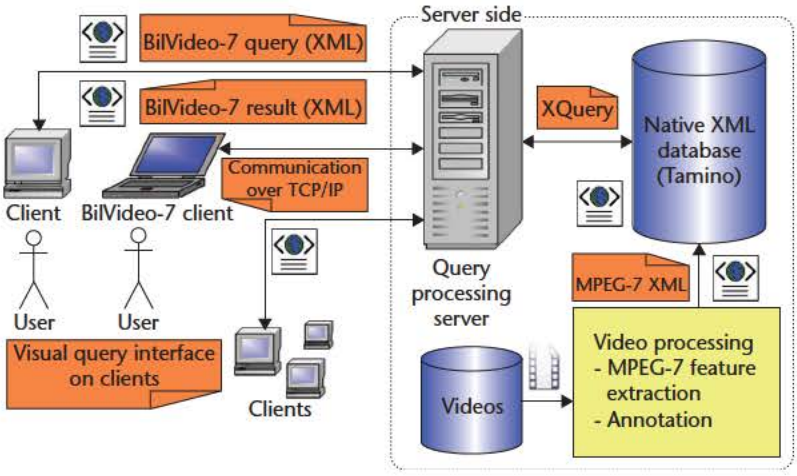


Figure 2. BilVideo-7's client–server architecture.

the MTD. The MBRs and visual descriptors of the object throughout the shot are stored by temporally decomposing the object into still regions. In this article, we refer to shots, keyframes, still regions, and moving regions as *video segments*.

System architecture

BilVideo-7 has a client–server architecture (see Figure 2). Users formulate queries on BilVideo-7 clients, which communicate with the BilVideo-7 query-processing server using an XML-based query language (see http://www.cs.bilkent.edu.tr/bilmdg/bilvideo-7) over TCP/IP. The query-processing server parses queries into subqueries, retrieves the required data from the XML database using XQuery for each subquery, executes subqueries, fuses the results of subqueries, and sends the results back to the clients.

Feature extraction and annotation

MPEG-7 video representations are obtained using the MPEG-7-compatible video feature extraction and annotation tool (see Figure 3a). In the figure, the current video frame is shown at the top left, the latest processed frame in the bottom left, the latest selected region in the top right, and the selected moving regions along with their trajectories in the bottom right. Selected video segments are shown on the right in a hierarchical tree view reflecting the structure of the video.

Videos, along with shot boundary information, are loaded and then processed on a shot-by-shot basis. Users can manually select keyframes, still regions, and moving regions and then annotate the video, shots, keyframes, still regions, and moving regions with free text, keyword, and structured annotations. The MPEG-7 visual descriptors (color, texture, shape, motion, and localization) for the selected video segments are computed by the tool, using an MPEG-7 feature-extraction library adapted from MPEG-7 eXperimentation Model (XM) Reference Software.[3] The semantic content is described by text annotations (free text, keyword, and structured annotation).

The output is saved as an MPEG-compatible XML file for each video. We use a native XML database, Tamino (see http://www.softwareag.com/Corporate/products/wm/tamino/default.asp), to store the MPEG-7 representations. Native XML databases use the XML data model directly. Thus, it's more convenient and natural to use a native XML database because no mapping and conversion to other data models is required and it's easy to set up the database using the publicly available MPEG-7 schema.

Visual query interface

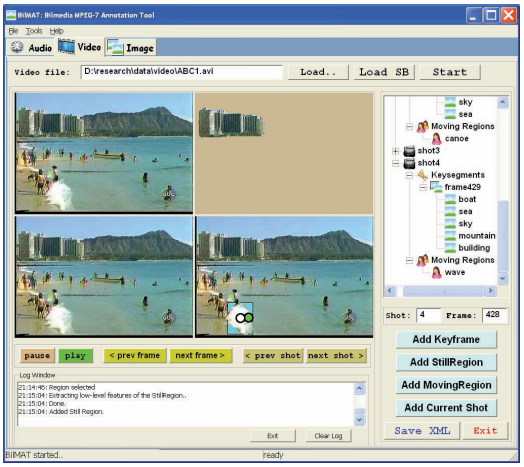
Users formulate queries on BilVideo-7 clients' visual-query interface, which consists of several tabs, each for a different type of query with a comprehensive set of descriptors and query options. As shown in Figure 3b, the query-formulation tabs are on the left, the query result list is displayed at the top right, the query results are at the bottom right, and the messages are at the bottom left. The user can select the media type, return type, and maximum number of results to be returned from the toolbar at the top. The queries are converted into the BilVideo-7 query format in XML and sent to the BilVideo-7 query-processing server.

The video table of contents is a useful facility to let the user browse through the video collection in the database, to see the contents of each video in a hierarchical tree reflecting the structure of the MPEG-7 representation of the video in XML format and to see the high-level semantic concepts in the collection and in each video separately. The user can browse through each video in the collection and see all the shots, keyframes, still regions, moving regions, temporal location, and the semantic concepts with which the the videos are annotated.

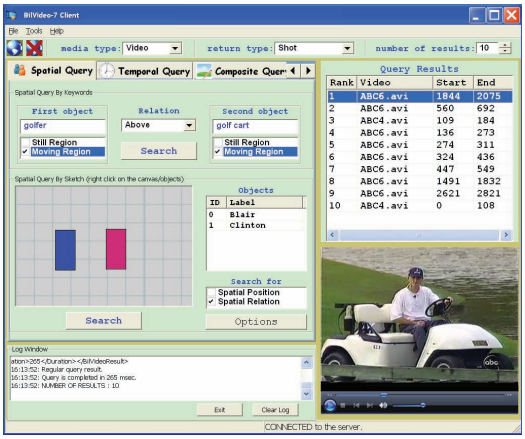
The textual query interface enables the user to formulate high-level semantic queries by entering keywords and specifying the type of video segment and annotation to be searched. The color, texture, and shape query interface is used for querying video segments by MPEG-7 color, texture, and shape descriptors. The input media can be a video segment, a whole image, or an image region. To provide this query functionality, the descriptors need to be extracted from the selected input media. Instead of uploading the input media to the server and extracting the descriptors there, we extract the descriptors on the client, form the XML-based query expression containing the descriptors, and send the query to the server. the BilVideo-7 clients include the MPEG-7 feature-extraction module to extract these descriptors. In addition to these query options, the user can specify the type of video segments to search, and also pick the weights and thresholds for each type of descriptor.

The motion-query interface is for the formulation of Motion Activity and Motion Trajectory queries. Trajectory points are entered using the mouse. The user can specify keywords for the moving region on which the trajectory query will be performed. The spatial-query interface enables the user to formulate spatial queries for still and moving regions using either keywords and a set of predefined spatial relations (left, right, above, below, east, west, and so on) or by sketching the MBR of objects using the mouse, and if desired, giving labels to them. It's possible to query objects on the basis of location, spatial relations, or both. The temporal-query interface is similar to the spatial query interface; this time, the user specifies temporal relations between video segments either by selecting from a predefined list (before, after, during, and so on) or by sketching the temporal positions of the segments using the mouse.

The composite-query interface is the most powerful query interface and enables the user to formulate complex queries. The query is composed by putting together any number of shots, keyframes, still regions, and moving



(a)



(b)

Figure 3. (a) MPEG-7-compatible video feature extraction and annotation tool. (b) BilVideo-7 client Visual Query Interface. This screenshot shows the spatial query interface.

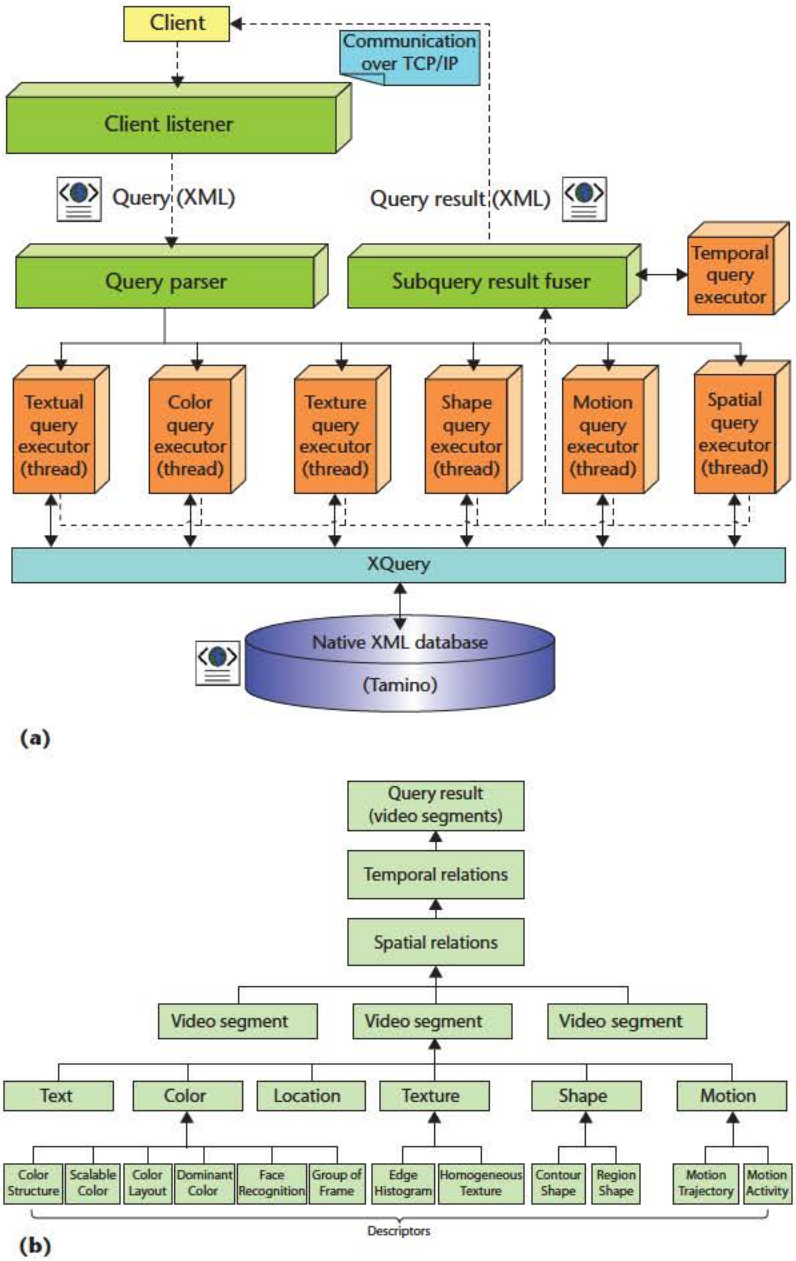


Figure 4. (a) The framework of the Query Processing Server. Each type of subquery is executed in a separate thread. (b) Subquery results are fused in a bottom-up manner.

regions, then specifying their properties as text-based semantic annotations, visual descriptors, locations, and spatial and temporal relations. Using this interface, the user can describe a video segment or a scene and ask the system to retrieve similar video segments.

The XQuery interface is best suited to experienced users who can write XQueries to search in the database. This is the most flexible interface in the system, and allows users to specify a wide range of queries.

Query processing

Query processing is done on the query-processing server, which is a multithreaded server-side component that listens to a configured TCP port and accepts incoming clients and processes their queries. Clients send queries in the XML-based BilVideo-7 query format and receive query results in XML-based BilVideo-7 result format, which contains a list of video segments (video name, start time, and end time) in ranked order. The current version of BilVideo-7 doesn't support MPEG Query Format query language because it's not possible to formulate some of the BilVideo-7 queries in MPQF (for example, spatial queries by location).

Multithreaded query execution

Each incoming query is parsed into subqueries and executed in a multithreaded fashion, with one thread for each type of subquery, as shown in Figure 4a. Queries with the same subquery type are accumulated in a queue and executed on a first-in-first-out basis. For example, subqueries for color descriptors (such as CSD, SCD, and DCD) are added to the end of queue of the color query executor thread and executed in this order. One XQuery is formed and executed for each subquery, consisting of a single video segment and a single descriptor (for example, keyframe with CSD). The XML database returns the XQuery results in XML format, which are parsed to extract the actual data (the descriptors).

Textual queries are the easiest to execute because the XML database can handle textual queries and no further processing is needed for the similarity computation. However, the database cannot handle the similarity queries for low-level descriptors. That is, the similarity between a query descriptor and a descriptor in the database cannot be computed by the database. Therefore, the corresponding query-execution thread retrieves the relevant descriptors from the database for the video segment in the subquery (for example, CSD for keyframes) and computes their distances to the query.

The distance measures suggested by MPEG-7 are implemented in MPEG-7 XM Reference Software but they are not normative. The evaluation of the distance measures for a set of MPEG-7 descriptors, presented elsewhere,[6] shows that although there are better distance measures (for example, pattern difference or Meehl index), the MPEG-7 recommendations

are among the best. Therefore, we adapted the distance measures from the XM software and briefly describe them here. In our computations, Q refers to a descriptor in the query, D to a descriptor in the database, and d is the computed distance between the descriptors.

We use L_1–norm to compute the distance between the CSD, SCD, GoF/GoP, and RSD.

$$d_{L_1}(Q,D) = \sum_i |Q(i) - D(i)|$$

The distance between two CLDs, $Q = \{QY, QCb, QCr\}$ and $D = \{DY, DCb, DCr\}$, is computed by

$$\begin{aligned} d(Q,D) = &\sqrt{\sum_i w_{yi}(QY_i - DY_i)^2} \\ &+ \sqrt{\sum_i w_{bi}(QCb_i - DCb_i)^2} \\ &+ \sqrt{\sum_i w_{ri}(QCr_i - DCr_i)^2} \end{aligned}$$

where i represents the zigzag-scanning order of the coefficients and the weights (w_{yi}, w_{bi}, w_{ri}) are used to give more importance to the lower-frequency components of the descriptor.

The distance between the EHDs Q and D is computed by adapting the L_1–norm as

$$\begin{aligned} d(Q,D) = &\sum_{i=0}^{79} \left|h_Q(i) - h_D(i)\right| \\ &+ 5\sum_{i=0}^{4} \left|h_Q^g(i) - h_D^g(i)\right| \\ &+ \sum_{i=0}^{64} \left|h_Q^s(i) - h_D^s(i)\right| \end{aligned}$$

where $h_Q(i)$ and $h_D(i)$ represent the histogram bin values of the descriptors Q and D, $h_Q^g(i)$ and $h_D^g(i)$ for global edge histograms, and $h_Q^s(i)$ and $h_D^s(i)$ for semiglobal edge histograms.

The distance between two HTDs Q and D (full layer, using both energy and energy deviation) is computed by

$$\begin{aligned} d(Q,D) = &w_{dc}|Q(0) - D(0)| + w_{std}|Q(1) - D(1)| \\ &+ \sum_{n=0}^{RD-1}\sum_{m=0}^{AD-1} w_e(n)|Q(n \cdot AD + m + 2) \\ &- D(n \cdot AD + m + 2)| \\ &+ w_{ed}(n)|Q(n \cdot AD + m + 32) \\ &- D(n \cdot AD + m + 32)| \end{aligned}$$

where w_{dc}, w_{std}, w_e, and w_{ed} are weights; the radial division, $RD = 5$, and angular division, $AD = 6$. Matching with this distance metric is not scale- and rotation-invariant.

The distance between two FRDs Q and D is computed as follows:

$$d(Q,D) = \sum_{i=0}^{47} w_i(Q(i) - D(i))^2$$

where w_i is the weight for the ith descriptor value.

The intensity of motion activity is a scalar value. Therefore, the distance is computed simply by taking the difference between two descriptor values. When the spatial localization of motion activity is given, we use Euclidean distance. For spatial-position queries, we use Euclidean distance between the center points of objects' MBRs. Due to space considerations, we omit the distance metrics for dominant color, contour shape, and motion trajectory.

If multiple instances of a descriptor are available for a moving region to account for the change in its appearance throughout the shot, the distance is computed for all the instances and the minimum is taken. If the computed distance for a video segment in the database is greater than the user-specified distance threshold for the query video segment and descriptor (for example, for a keyframe with CSD, if $d(Q,D)/d_{\max} > T_{Keyframe,CSD}$), that segment is discarded. Otherwise, the similarity, $s(Q, D)$, between two descriptors Q and D is computed as $s(Q, D) = 1 - d(Q, D)/d_{max}$, $0 \leq s(Q, D) \leq 1.0$; where $d(Q, D)$ is the distance between descriptors Q and D, and d_{max} is the maximum possible distance for the type of descriptor in the computation. We compute the maximum distance for each descriptor by taking the maximum distance from a large set of descriptors extracted from video segments.

Spatial locations of still and moving regions are stored in the database by their MBRs, without any preprocessing to extract and store the spatial relations between them. Therefore, we compute spatial similarity between regions at query-execution time, which is computationally more expensive but provides a more flexible and accurate matching for spatial-position and spatial-relation queries.

For each still and moving region in the query, queries related to the properties of the region are executed as described previously. Then the resulting video segments undergo spatial query processing to compute the spatial

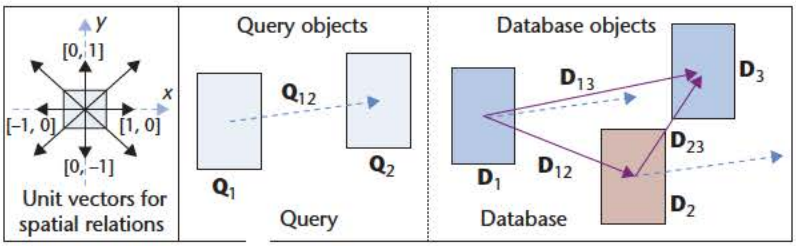


Figure 5. Spatial query processing by vector dot product between the vectors connecting centers of objects' minimum bounding rectangles.

similarities between them. We use the spatial similarity matching approach described elsewhere[7] because of its efficiency and robustness. First, we compute the vectors connecting the center points of objects' MBRs, $\mathbf{Q_{xy}}$ and $\mathbf{D_{ij}}$, as shown in Figure 5. Then, we compute the pairwise spatial similarity by the cosine of the angle between the vectors $\mathbf{Q_{xy}}$ and $\mathbf{D_{ij}}$, using vector dot product

$$d(\mathbf{Q_{xy}}, \mathbf{D_{ij}}) = \cos\theta = \frac{\mathbf{Q_{xy}} \cdot \mathbf{D_{ij}}}{|\mathbf{Q_{xy}}||\mathbf{D_{ij}}|}, \quad 0 \le \theta \le \pi,$$

$$-1 \le d(\mathbf{Q_{xy}}, \mathbf{D_{ij}}) \le +1$$

The output value is in the range [−1, +1], with +1 indicating identical spatial relation and −1 opposite spatial relation. In Figure 5, the spatial relation between the database objects D_1 and D_3 is the most similar to the spatial relation between query objects Q_1 and Q_2.

The text-based spatial queries (right, left, above, below, and so on) are executed in the same way, by converting each spatial-relation query to a unit vector (see Figure 5). For instance, Q_x *right* Q_y (Q_x is to the right of Q_y) query is converted to a query vector $\mathbf{Q_{xy}} = [-1, 0]$, from Q_x to Q_y. Multiple MBRs are stored in the database for moving regions to keep track of their locations. We compute the spatial similarity for all the MBRs and take the maximum similarity value as the final similarity. We execute temporal queries, if any, after spatial queries by checking if the list of video segments satisfies the temporal relations specified in the query. Spatial queries implicitly enforce a temporal relation between still and moving regions because they must co-appear on a scene for a certain time in the video to satisfy the spatial relations.

Fusion of subquery results

When multiple descriptors, possibly in different modalities, are specified for a query video segment, each is executed as a separate subquery, resulting in a list of video segments with similarity values. These subquery results must be fused to come up with the final query result. This fusion is done in a bottom-up manner, as shown in Figure 4b. Each node in the tree has an associated weight and threshold, which can be specified by the user during query formulation. We compute the similarity at each node as the weighted average of the similarities of its children. The fusion process continues upward in the tree until we obtain the final query result.

To illustrate the fusion process, consider a composite query consisting of a keyframe with color (CSD and DCD), texture (EHD and HTD), and text-based semantic (keyword *golf green*) descriptors. The query processor parses this query into five subqueries (CSD, DCD, EHD, HTD, and semantic), executes each, and produces five lists of keyframes from the database with similarity values. Then it fuses color (CSD, DCD) and texture (EHD, HTD) subquery results to obtain the color and texture similarities of each keyframe:

$$s_{i,color} = \frac{w_{keyframe,CSD} s_{i,CSD} + w_{keyframe,DCD} s_{i,DCD}}{w_{keyframe,CSD} + w_{keyframe,DCD}}$$

$$s_{i,texture} = \frac{w_{keyframe,EHD} s_{i,EHD} + w_{keyframe,HTD} s_{i,HTD}}{w_{keyframe,EHD} + w_{keyframe,HTD}}$$

where $s_{i,color}$ is the color similarity for the ith keyframe, $w_{keyframe,CSD}$ is the weight for CSD, and so on. If the similarity of keyframe i is less than the threshold specified by the user, it is discarded. At this point, we have three lists of keyframes having similarity values for color, texture, and semantics (text). We fuse these three lists to obtain the final list of keyframes:

$$s_i = \frac{w_{keyframe,color} s_{i,color} + w_{keyframe,texture} s_{i,texture} + w_{keyframe,text} s_{i,text}}{w_{keyframe,color} + w_{keyframe,texture} + w_{keyframe,text}}$$

If there are also spatial or temporal relation subqueries, they are executed and similarity values of the video segments are updated in the same way. Finally, we obtain N_{vs} lists of video segments, where N_{vs} is the number of video segments in the query. The final query result is obtained by fusing these lists using the same weighted average approach and by sorting the list in descending order of similarity.

Experimental results

The system is implemented in C++ and uses the Open Source Computer Vision Library (see http://opencvlibrary.sourceforge.net) to handle

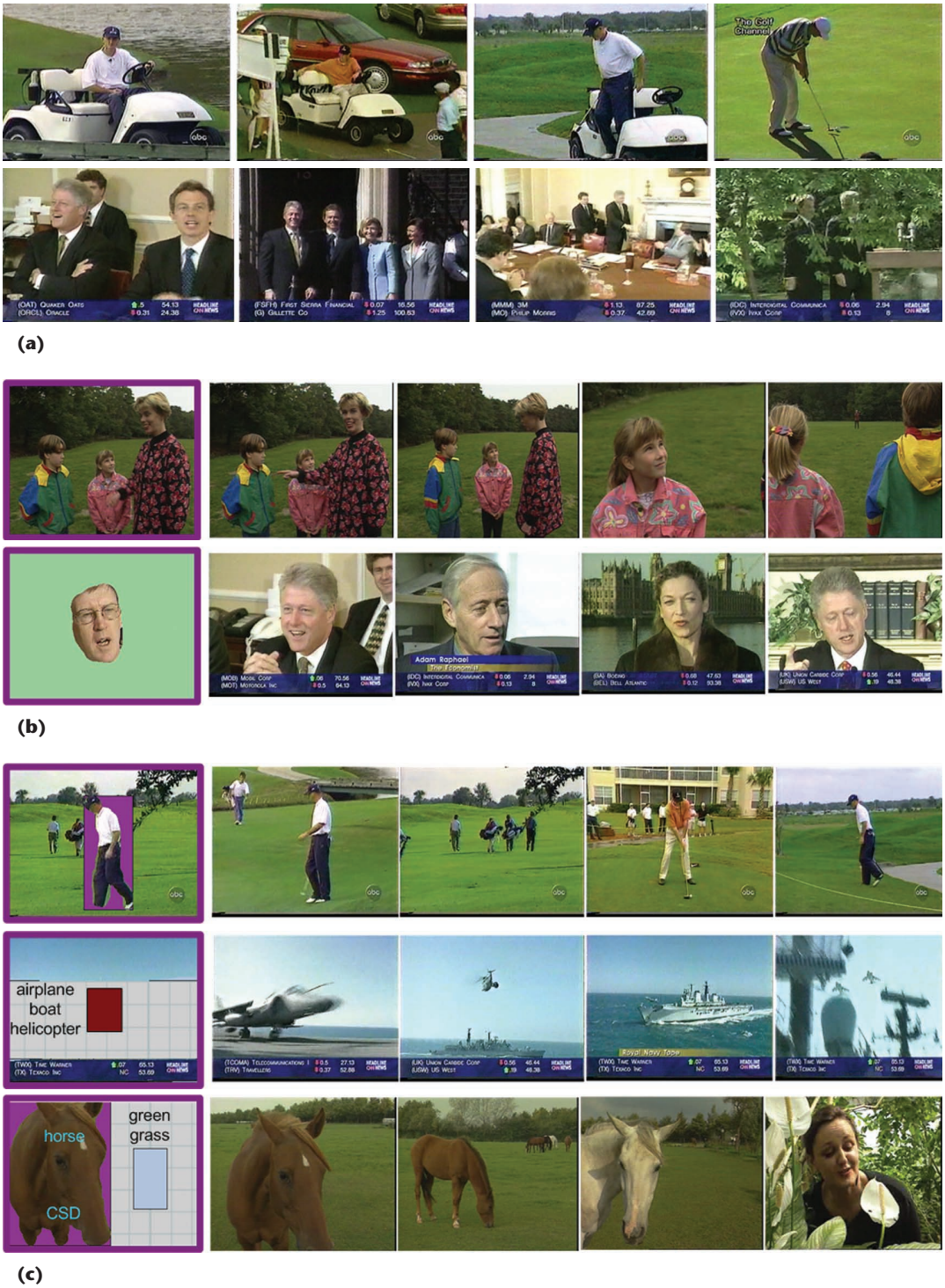


Figure 6. (a) Two spatial queries: top row shows a text-based spatia-relation query, "golfer above golf cart." The bottom row shows a sketch-based spatial-relation query, "Clinton left Blair," formulated by drawing two rectangles and labeling them as "Clinton" and "Blair." (b) Image-based (top row) and region-based (bottom row) low-level queries (queries are on the left). (c) Composite queries.

the image and video data, and wxWidgets (see http://www.wxwidgets.org) for the graphical user interfaces. We performed queries on a video data set consisting of 14 video sequences with 25,000 frames from the Trecvid 2004 and 2008 data sets,[8] which include news, documentary, educational, and archival videos. The query result is a list of shots in ranked order, each shown as a representative keyframe in Figure 6.

Figure 6a shows two spatial queries. The first query at the top searches for the video segments in which a golfer is above a golf cart, formulated as a text-based spatial-relation query, "golfer above golf cart." The system successfully returns three relevant video segments

Table 1. Query-execution times. The query-processing server and Tamino XML database are installed on a notebook PC with Intel Core 2 dual-core 2-GHz processors and 2 Gbytes of RAM, running Windows XP.

Query type	Description (Segments and descriptors)	Execution time (seconds)
Text-based semantic query	Keyframe (keyword)	0.125
Text-based semantic query	Moving region (keyword)	0.125
Text-based semantic query	Keyframe (keyword), Moving region (keyword)	0.188
Color query	Keyframe (CSD)	0.141
Texture query	Keyframe (HTD)	0.125
Color and texture query	Keyframe (CSD and HTD)	0.172
Shape query	Moving region (RSD)	0.156
Spatial query	Text-based, two still regions	0.172
Spatial query	Text-based, two moving regions	0.187
Spatial query	Sketch-based, two moving regions	0.187
Composite query in Figure 6c, top row	Keyframe (DCD and keyword), moving region (CSD, RSD, and keyword)	0.438
Composite query in Figure 6c, bottom row	Two still regions (CSD and EHD), moving region (keyword)	0.391

that exactly match the spatial query condition. The fourth result contains a golfer but no golf cart; the spatial condition is not satisfied, so its rank is lower than the first three.

The second query on the bottom row of Figure 6a, "Clinton left Blair," is sketch-based. The spatial-query condition is satisfied exactly in the first two video segments returned, while it's not satisfied in the last two, even though Clinton and Blair appear together. This is a desirable result of our bottom-up fusion algorithm; as the number of satisfied query conditions for a video segment decreases, the video segment's similarity also decreases, ranking lower in the query result. As a result, the ranking approach is effective and it produces query results that are close to our perception.

Figure 6b shows two low-level queries. In the image-based query (top row), the query image is represented by CSDs and DCDs and searched in keyframes. In the region-based query (bottom row), the query region is represented by CSDs and RSDs, and searched in moving regions. Both query results are satisfactory considering the types of descriptors used.

The three queries shown in Figure 6c are composite queries, in which high-level semantics in the form of keyword annotations and low-level descriptors (DCD, CSD, EHD, and RSD) are used together to describe the query video segments. In the first composite query, the keyframe is represented by the DCD and golf green. The moving region is represented by the CSD, RSD, and golfer. In the second query, two still regions at the top and at the bottom of Figure 6c are represented by the CSD and EHD. The moving region in the middle is represented by semantic concepts airplane or boat or helicopter.

Using such queries, the user can access video segments having any specific composition described in the query. The number and type of video segments in the query, as well as the descriptors used to describe them, are not limited. This makes the composite queries flexible and powerful, enabling the user to formulate complex queries easily. To our knowledge, our system is unique in supporting such complex queries.

Table 1 presents query-execution times for several queries. The query-execution time is proportional to the number of subqueries (number of video segments and descriptors in the query), database size (number of video segments in the database), sizes of the descriptors, and complexity of the matching algorithm (distance metric). Queries involving low-level descriptors take longer to execute in comparison to text-based queries because the distance computations between the low-level descriptors are computationally more expensive. The multithreaded query-processing architecture provides some degree of parallelism and shortens the query-execution times when the subqueries are executed in separate threads.

Conclusions and future work

To our knowledge, BilVideo-7 is the most comprehensive MPEG-7-compatible video database system currently available, in terms of the wide range of MPEG-7 descriptors and manifold query options supported. However, the system's retrieval performance depends very much on the MPEG-7 descriptors and the distance measures used. As for our future work, we plan to investigate distance measures other than the ones recommended by MPEG-7.[6]

Currently, the major bottleneck for the system is the generation of the MPEG-7 representations of videos by manual processing, which is time-consuming, error-prone, and suffers from human subjectivity. This processing hinders the construction of a video database of realistic size. Therefore, our current focus is on equipping the MPEG-7-compatible video feature extraction and annotation tool with as much automatic processing capabilities as possible to reduce manual processing time, errors, and human subjectivity during region selection and annotation.

In future versions of BilVideo-7, we plan to add support for representation and querying of audio and image data. The BilVideo-7 multimodal query-processing architecture makes it easy to add new descriptors for new modalities such as audio descriptors. Images could be considered to be a special case of keyframes that are decomposed into still regions, and hence can be supported easily in BilVideo-7. **MM**

Acknowledgments

We thank Rana Nelson for proofreading this manuscript.

References

1. B.S. Manjunath, P. Salembier, and T. Sikora, eds., *Introduction to MPEG-7: Multimedia Content Description Interface,* Wiley, 2002.
2. M.E. Dönderler et al., ''BilVideo: Design and Implementation of a Video Database Management System,'' *Multimedia Tools and Applications,* vol. 27, no. 1, 2005, pp. 79-104.
3. *Information Technology—Multimedia Content Description Interface—Part 6: Reference Software,* ISO/IEC 15938-6:2003, 2009; http://standards.iso.org/ittf/PubliclyAvailableStandards/index.html.
4. *Information Technology—Multimedia Content Description Interface—Part 12: Query Format,* ISO/IEC FCD 15938-12:2007, 2007; http://www.chiariglione.org/mpeg/working documents/mpeg-07/mp7qf/mp7qf-fcd.zip.
5. W. Bailer and P. Schallauer, ''Detailed Audiovisual Profile: Enabling Interoperability Between MPEG-7 Based Systems,'' *Proc. 12th Int'l Multi-Media Modelling Conf.,* IEEE Press, 2006, pp. 217-224.
6. H. Eidenberger, ''Distance Measures for MPEG-7-Based Retrieval,'' *Proce 5th ACM SIGMM Int'l Workshop Multimedia Information Retrieval,* ACM Press, 2003, pp. 130-137.
7. J.Z. Li and M.T. Ozsu, ''Stars: A Spatial Attributes Retrieval System for Images and Videos,'' *Proc. 4th Int'l Conf. Multimedia Modeling,* Springer, 1997, pp. 69-84.
8. A.F. Smeaton, P. Over, and W. Kraaij, ''Evaluation Campaigns and Trecvid,'' *Proc. 8th ACM Int'l Workshop Multimedia Information Retrieval,* ACM Press, 2006, pp. 321-330.

Muhammet Baştan is a PhD candidate in the department of computer engineering at Bilkent University, Ankara, Turkey. His research interests include computer vision, pattern recognition, multimedia retrieval, MPEG-7, image and video processing, saliency, segmentation, and annotation. Baştan has an MS in electronics engineering and computer science from Sabanci University, Istanbul, Turkey. Contact him at bastan@cs.bilkent.edu.tr.

Hayati Çam (1983-2009) was a software engineer at Siemens, Ankara, Turkey. His research interests included multimedia databases and computer vision. Çam had an MS from the Department Computer Engineering at Bilkent University.

Uğur Güdükbay is an associate professor in the department of computer engineering, Bilkent University, Ankara, Turkey. His research interests include multimedia databases, computer vision, and computer graphics. Güdükbay has a PhD in computer engineering and information science from Bilkent University. He is a senior member of IEEE and ACM. Contact him at gudukbay@cs.bilkent.edu.tr.

Özgür Ulusoy is a professor in the department of computer engineering, Bilkent University, Ankara, Turkey. His research interests include multimedia databases, web databases, peer-to-peer systems, data management for mobile systems, and real-time and active database systems. Ulusoy has a PhD in computer science from University of Illinois at Urbana-Champaign. He is a member of IEEE and ACM. Contact him at oulusoy@cs.bilkent.edu.tr.

Appendix ZHOU01

PennState University Libraries | Catalog
Bookmarks 0 Course Reserves My Account
Keyword Search... Search Advanced search Start Over
Share Bookmark Report an Issue

MARC View

LEADER 02294nam a22004217a 4500

001				9177885
003				SIRSI
005				20121214090528.0
006				m d
007				cr n
008				050615s2004 nyuab sb 101 0 eng d
010			a|	2005278507
020			a|	1581139403 (pbk.)
040			a|	PVA c| PVA d| GSU d| IAY d| OCL d| XTL d| IXA d| C$A d| DLC d| WaSeSS
042			a|	lccopycat
050	0	0	a|	QA76.575 b| .A367 2004
082	0	0	a|	006.7 2| 22
111	2		a|	ACM SIGMM International Workshop on Multimedia Information Retrieval n| (6th : d| 2004 : c| New York, N.Y.)
245	1	0	a|	MIR '04 h| [electronic resource] : b| proceedings of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval, October 15-16, 2004, New York, NY, USA, in conjunction with ACM Multimedia 2004 / c| organizers, Michael S. Lew, Nicu Sebe, Chabane Djeraba ; sponsored by the ACM Special Interest Groups SIGMM & SIGGRAPH.
246	3	0	a|	Proceedings of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval
246	3	0	a|	6th ACM SIGMM International Workshop on Multimedia Information Retrieval
260			a|	New York, NY : b| ACM Press, c| c2004.
300			a|	x, 323 p. : b| ill., map ; c| 28 cm.
500			a|	"ACM order number: 433043"--P. ii.
504			a|	Includes bibliographical references and index.
506			a|	License restrictions may limit access.
650		0	a|	Multimedia systems v| Congresses.
650		0	a|	Information storage and retrieval systems v| Congresses.
655		0	a|	Electronic books.
700	1		a|	Lew, Michael S., d| 1965-
700	1		a|	Sebe, Nicu.
700	1		a|	Djeraba, Chabane.
710	2		a|	ACM Digital Library.
710	2		a|	ACM Special Interest Group on Multimedia.
710	2		a|	SIGGRAPH.
711	2		a|	ACM International Conference on Multimedia n| (12th : d| 2004 : c| New York, N.Y.)
856	4	0	u|	http://SK8ES4MC2L.search.serialssolutions.com/?sid=sersol&SS_jc=TC0000752847&title=MIR%20%2704%20%3A%20proceedings%20of%20the%206th%20ACM%20SIGMM%20International%20Workshop%20on%20Multimedia%20Information%20Retrieval%2C%20October%2015-16%2C%202004%2C%20New%20York%2C%20NY%2C%20USA%2C%20in%20conjunction%20with%20ACM%20Multimedia%202004
949			a|	Electronic resource w| ASIS c| 1 i| 9177885-1001 l| ONLINE m| ONLINE r| Y s| Y t| ONLINE u| 12/19/2012 o| .STAFF. 360 MARC

ASK A LIBRARIAN

PennState

Copyright ©2019 The Pennsylvania State University. All rights reserved. Except where otherwise noted, this work is subject to a Creative Commons Attribution 4.0 license. Details and exceptions.

Legal Statements | PSU Hotlines

PENN STATE UNIVERSITY LIBRARIES
Libraries Home (814) 865-6368
Libraries Intranet (Staff Only)
Accessibility Help
Website Feedback
Policies and Guidelines
Staff Directory

CONNECT WITH PENN STATE UNIVERSITY LIBRARIES
Facebook
Twitter
Instagram

MIR '04 [electronic resource] : proceedings of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval, October 15-16, 2004, New York, NY, USA, in conjunction with ACM Multimedia 2004 / organizers, Michael S. Lew, Nicu Sebe, Chabane Djeraba ; sponsored by the ACM Special Interest Groups SIGMM & SIGGRAPH

Conference Author	ACM SIGMM International Workshop on Multimedia Information Retrieval (6th : 2004 : New York, N.Y.)
Additional Titles	Proceedings of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval and 6th ACM SIGMM International Workshop on Multimedia Information Retrieval
Published	New York, NY : ACM Press, c2004.
Physical Description	x, 323 p. : ill., map ; 28 cm.
Additional Creators	Lew, Michael S., 1965-, Sebe, Nicu, Djeraba, Chabane, ACM Digital Library, ACM Special Interest Group on Multimedia, SIGGRAPH, and ACM International Conference on Multimedia (12th : 2004 : New York, N.Y.)

Access Online

serialssolutions.com

Restrictions on Access

License restrictions may limit access.

Subject(s)

Multimedia systems—Congresses

Information storage and retrieval systems—Congresses

Genre(s)

Electronic books

ISBN

1581139403 (pbk.)

Note

"ACM order number: 433043"--P. ii.

Bibliography Note

Includes bibliographical references and index.

ASK A LIBRARIAN

View MARC record | catkey: 9177885

PennState

Copyright ©2019 The Pennsylvania State University. All rights reserved. Except where otherwise noted, this work is subject to a Creative Commons Attribution 4.0 license. Details and exceptions.

Legal Statements | PSU Hotlines

PENN STATE UNIVERSITY LIBRARIES

Libraries Home (814) 865-6368

Libraries Intranet (Staff Only)

Accessibility Help

Website Feedback

Policies and Guidelines

CONNECT WITH PENN STATE UNIVERSITY LIBRARIES

Facebook

Twitter

Instagram

Appendix ZHOU02

A Web-Enabled Video Indexing System

Jian Zhou
Department of Electrical and Computer Engineering
Ryerson University
350 Victoria Street, Toronto, Ontario
Canada, M5B 2K3
jzhou@ee.ryerson.ca

Xiao-Ping Zhang
Department of Electrical and Computer Engineering
Ryerson University
350 Victoria Street, Toronto, Ontario
Canada, M5B 2K3
xzhang@ee.ryerson.ca

ABSTRACT
Video parsing and indexing is an important early stage of content-based video analysis. In this paper, we present a new web-enabled video indexing system that integrates Synchronized Multimedia Integration Language (SMIL) standard. New algorithms are proposed for video temporal segmentation. Sharp transition detection is achieved by an enhanced histogram-based method that is robust to illumination changes. For gradual transition detection, new features are introduced for dissolve detection. The proposed dissolve detector is based on a combined analysis of mean-variance-skewness of intensity. Compared with existing variance-based approaches, the introduced new features improve the discrimination ability on shot boundaries. We also describe methods for eliminating false positives. Experimental results show that the proposed algorithms can effectively detect shot boundaries. Detected scenes and other cinematic attributes are structured and organized by integrating HTML and SMIL. For each video file, the system generates a table-of-contents indexing file. The user-friendly interface provides web-based interaction, browsing and previewing of video content.

Categories and Subject Descriptors
H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems- *Video*; H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing.

General Terms
Algorithms, Design, Experimentation, Performance, Theory.

Keywords
Shot boundary detection, video temporal segmentation, video indexing, SMIL.

1. INTRODUCTION
Content based video analysis and retrieval has become an area of active research in recent years. The first step of video content analysis is to segment video into its constituent shots, and high-level scenes or episodes. As the building blocks of video projects, video shots need to be identified effectively and efficiently. The research of video parsing focuses on the detection of both sharp shot transition (cuts) and gradual shot transition (fade/dissolve). Many automatic techniques have been developed to detect video shot boundaries in both the compressed and the decompressed domains. Zhang *et al.* [1] proposed pair-wise pixel comparison, likelihood ratio and histogram comparison for sharp transition detection. Edge changes can also be used as a good feature for shot detection [2]. Yeo and Liu [3] detected scene changes by using pixel difference and luminance histograms based on DC-images in compressed domains. In [4], recently developed methods for shot detection were reviewed in detail, and a statistical detector was proposed based on motion compensation. Shot detection techniques can be categorized into feature-based [2], model-based [5] and statistical [6] methods. Most of the above techniques can achieve good performance on hard cut detection.

However, compared with sharp transition detection, gradual transition still remains a challenging problem. When dealing with gradual transition, Zhang *et al.* [7] used a twin threshold mechanism based on histogram difference metric. Frame differences were accumulated when inter-frame difference was above the lower threshold but smaller than the higher threshold. When the accumulated difference exceeded the higher threshold, a gradual transition was defined. In [2], edge and contour changes were used for gradual transition detection. Another feature that is commonly used for dissolve detection is intensity variance. During a dissolve transition, the intensity variance curve forms a downwards-parabolic shape. The variance-based approach was first introduced by Alattar [5], and many other researchers have used this feature to build their dissolve detectors [4] [6]. Alattar [5] proposed to take the second order difference of intensity variance, and then check two large negative spikes. However, such pattern might not be pronounced due to object/camera motion and noise. Truong *et al.* [6] proposed an improved version by adding more constraints. Lienhart [8] introduced a somewhat different approach that includes a transition synthesizer and a neural network classifier, and dissolves are detected by a multi-resolution search.

Most of the existing techniques require careful selection of thresholds to achieve good performance. Some key factors that affect the performance of shot detection include illumination changes and object/camera motion. Since histograms do not carry spatial information, they are expected to be robust to object and camera motion. However, illumination changes can cause serious problems. Some feature-based methods, for example, those

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
MIR'04, October 15–16, 2004, New York, New York, USA.
Copyright 2004 ACM 1-58113-940-3/04/0010…$5.00.

algorithms based on the appearance of intensity edges, are less sensitive to illumination changes, but they are not robust to the motions of large objects and extra computations are also introduced.

In this paper, new algorithms are proposed for shot detection. We present a cut detection algorithm that is robust to illumination changes. Dissolve detection is achieved by a combined analysis of intensity moments. In addition to the variance feature, the introducing of mean and skewness adds more constraints and improves the discrimination ability of frame distances on shot boundaries, and thus provides a more robust way for shot detection Experimental results show that the proposed mean-variance-skewness approach can capture those dissolves whose intensity patterns are not so obvious if using only variance feature.

To structure and describe video content, media description scripts or template need to be developed. In [9], MPEG-7 and MPEG-21 were used to design a video personalization and summarization system. While MPEG-7 defines several levels of abstraction and provides a standard set of tools for describing multimedia content, its scope dose not focus on the storage, delivery or presentation of digital video. Considering the growing amounts of online digital video with the success of the Internet, it is desirable to incorporate web-based technologies in content based video analysis projects. In this work, we describe the use of the Synchronized Multimedia Integration Language (SMIL) standard for building the web-enabled video indexing system. In the proposed system, SMIL bridges the gap between the structure of video content and its web-based presentations. User friendly web-enabled interaction, integration and synchronization of video segments are realized by hybrid documents combining HTML and SMIL.

2. SYSTEM OVERVIEW

As shown in Figure 1, the web-enabled video indexing system includes frame-level playback, feature extraction, cut detection, dissolve detection, a SMIL generator and Graphic User Interface (GUI).

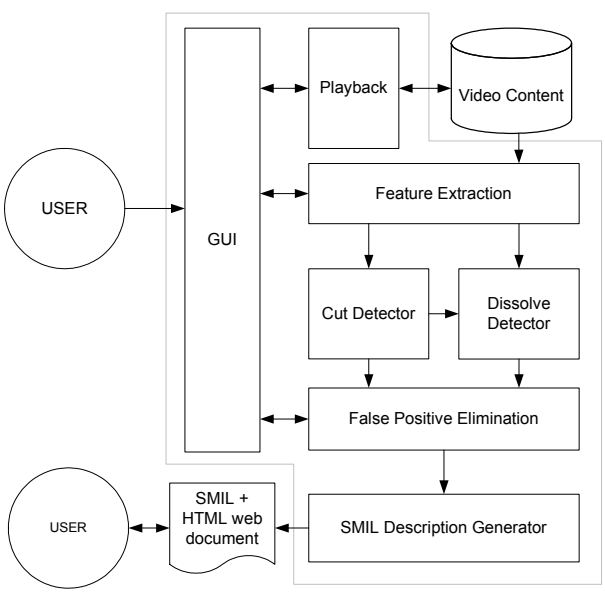


Figure 1: Architecture of a web-enabled video parsing system

After features are extracted from video content, cut detection is performed, followed by dissolve detection. Then, false positives are eliminated by a validation process. SMIL and HTML are used to describe the structures of video content including shot boundaries, video format, frame rate and other cinematic information. For each video clip, the system generates a web document as table-of-contents, which can be used for previewing and browsing video content online.

3. SHARP TRANSITION DETECTION

3.1 Improved Cut Detection

Histogram based methods are widely used for cut detection. Compared with other techniques, a histogram is robust to object motion, and it is simple and fast to calculate intensity or color histogram difference between two consecutive frames. Given a video sequence with N frames, denote the *n-th* frame as $f_n, n = 1, \ldots, N$. Let $H(f_n, k)$ denote the value of the *k-th* bin of the histogram for the frame f_n. Suppose there are total K bins in the histogram. Then the histogram difference at time n, denoted by $D_h(n)$, can be defined as follows:

$$D_h(n) = \frac{1}{A} \cdot \sum_{k=1}^{K} \left| H(f_{n+1}, k) - H(f_n, k) \right|, \; n = 1, 2, \ldots N, \qquad (1)$$

where A is the normalization factor. Histogram difference is a measure of dissimilarity between two consecutive frames. For cut transitions, visual characteristics are expected to change sharply at short boundaries, and thus the histogram difference can capture the visual discontinuities between shots. In practice, histogram based methods are the most common approach to shot detection, since they provide a good trade-off between accuracy and computational efficiency [1] [10] [11]. However, most histogram based methods are very sensitive to lighting changes. An example is shown in Figure 2. The luminance component is used to calculate the histogram difference. The first peak at frame 120 represents a real shot cut, and other peaks are caused by camera flashlight scenes. It can be seen that these illumination changes cause serious problems for cut detection.

In this paper, we propose a new histogram-based algorithm that is robust to lighting changes. The three channels in RGB color space are converted to the opponent color space. Only the red-green *(R-G)* component is used to compute the histogram difference. The opponent color representation of RGB color space is defined as [12]:

$$(R+G+B, R-G, 2 \cdot B - R - G), \qquad (2)$$

where R, G and B are red, green and blue channels respectively. By choosing the opponent color space, the proposed cut detection algorithm is less sensitive to lighting changes. As shown in Figure 3, the performance is significantly improved. Experimental results show that our method performs better, compared with working only with chromatic color components in YCbCr color space.

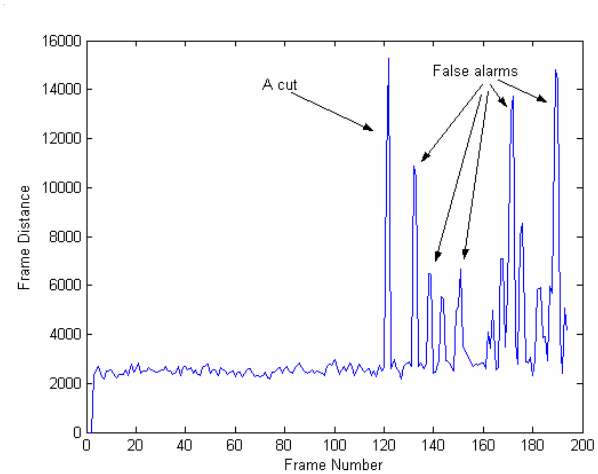


Figure 2: Histogram difference of luminance component in YCbCr color space.

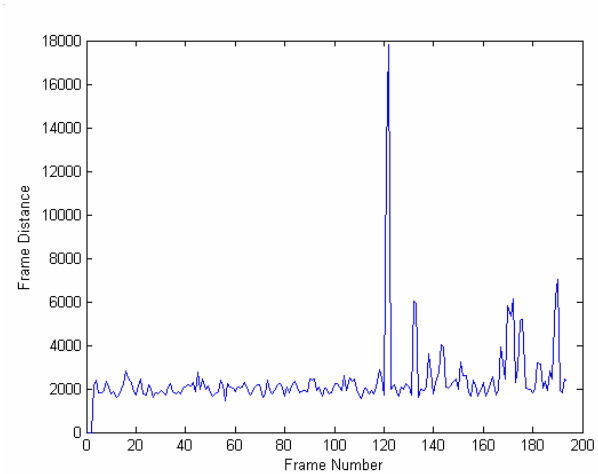


Figure 3: Histogram difference of (R-G) component in the opponent color space.

The conversion from RGB color space to its opponent color representation is computationally efficient. The advantage of this representation is that the last two chromaticity axes are invariant to changes in illumination intensity and shadows. The same video sequence is used to calculate the histogram difference in Figure 2 and Figure 3. It can be readily seen that the effects caused by flashlight scenes are significantly reduced.

Because the distribution of frame distances varies from video to video, adaptive threshold is more reasonable than a globally fixed threshold. It is worth mentioning that the normalization factor *A* defined in equation (1) is crucial for robust adaptive threshold, since the video data may have different image sizes. A temporal sliding window of the length *2w+1* with *w*=8 and centered at current frame *n* is used to capture the local characteristics. A hard cut candidate is detected at frame *n* if the following conditions are satisfied:

1. $D_h(n)$ has the maximum value inside the sliding window, i.e., $D_h(n) \geq D_h(k), \forall k \in [n-w, n+w]$.
2. The difference between $D_h(n)$ and the median value of the sliding window is larger than a given threshold T_1.

An example based on the proposed cut detection algorithm is shown in Figure 4.

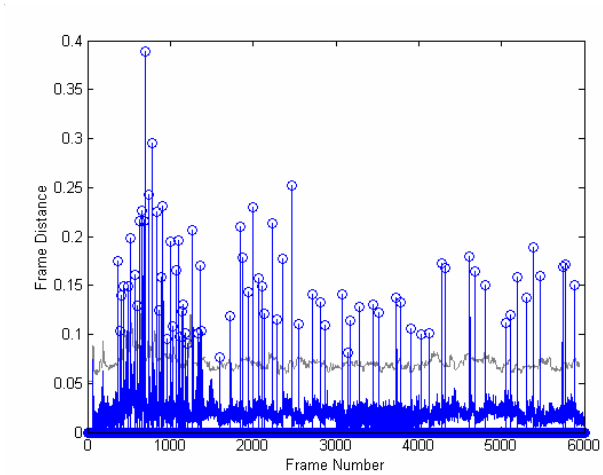


Figure 4: Cuts detected by adaptive threshold in the opponent color space (TV show "*Friends*").

3.2 False Positive Elimination

Due to camera/object motion and noise, some frames might be mistakenly identified as shot boundaries. Thus, a false positive elimination process is necessary. The validation process includes two criteria. First, a hard cut candidate at time *n* is declared as a false positive if frame *(n+2)* is similar as frame *(n-2)*, i.e.:

$$\sum_{k=1}^{K} \left| H(f_{n+2}, k) - H(f_{n-2}, k) \right| < T_2, \tag{3}$$

where T_2 is a given threshold. Second, excluding the maximum value $D_h(n)$ of the sliding window, we calculate the left half maximum denoted as M_L and the right half maximum M_R. Let M_V denote the average of the two maximums. If the difference between $D_h(n)$ and M_V is less than a given threshold T_3, the cut candidate $D_h(n)$ is deemed as a false positive; that is:

$$M_L = \max\{\forall\, D_h(i), i \in (n-1-w, n-1)\}, \tag{4}$$

$$M_R = \max\{\forall\, D_h(i), i \in (n+1, n+1+w)\}, \tag{5}$$

$$M_V = (M_L + M_R)/2, \tag{6}$$

$$\left| D_h(n) - M_V \right| < T_3. \tag{7}$$

The above criteria can effectively eliminate the false positives. The first criterion define in (3) is to measure the difference of visual characteristics between two shots. The second criterion compares the maximum value at the center with the second and third peaks, and such criterion can remove the false positives caused by dissolves.

4. GRADUAL TRANSITION DETECTION

4.1 Mean-Variance-Skewness

The detection of gradual transitions is less mature, compared with cut detection. One of the main reasons is that it is difficult to define and capture the visual discontinuities for gradual transitions. Most of the recent research work focuses on dissolve detection, since dissolve is the dominant editing style in gradual transitions. In this paper, we also focus on dissolves instead of other types of gradual transitions. During a dissolve, intensity variance has been proven to show a parabolic shape [5]. However, such pattern might not be so obvious due to motion and noise. In the proposed algorithm, in addition to intensity variance, mean and skewness are introduced as new features. The first order (mean - μ), the second order (variance - σ) and the third order (skewness - s) intensity moments for the *n-th* frame are defined as:

$$\mu(n) = \frac{1}{M \cdot N} \sum_{i=1}^{M} \sum_{j=1}^{N} f(i,j,n), \tag{8}$$

$$\sigma(n) = \left\{ \frac{1}{M \cdot N} \sum_{i=1}^{M} \sum_{j=1}^{N} [f(i,j,n) - \mu(n)]^2 \right\}^{1/2}, \tag{9}$$

$$s(n) = \left\{ \frac{1}{M \cdot N} \sum_{i=1}^{M} \sum_{j=1}^{N} [f(i,j,n) - \mu(n)]^3 \right\}^{1/3}. \tag{10}$$

Here $f(i,j,n)$ is the intensity value of the image pixel at location (i,j) for the *n-th* frame. *M* and *N* are image width and height respectively. The above moments are functions of time. Assuming a frame at time *t* is defined as $f(x,y,t)$, and *x, y, t* are continuous variables. A dissolve transition with duration of *T* can be considered as a mixture of two shots $f_1(x,y,t)$ and $f_2(x,y,t)$. During dissolve transition, the intensity of one shot decreases, and the intensity of the other increases. The dissolve editing style can be approximated by choosing two linear scaling functions $g_1(t)$ and $g_2(t)$ as:

$$g_1(t) = \frac{T-t}{T}, g_2(t) = \frac{t}{T}, \text{ where } t \in [0,T]. \tag{11}$$

The dissolve sequence $D(x,y,t)$ for $t \in [0,T]$ can be defined as [13]:

$$D(x,y,t) = g_1(t) \cdot f_1(x,y,t) + g_2(t) \cdot f_2(x,y,t). \tag{12}$$

Assuming two shots $f_1(\cdot)$ and $f_2(\cdot)$ are statistically independent and roughly ergodic random processes [5] [13]. The intensity variance is given by [5]:

$$\begin{aligned} \sigma^2(t) &\equiv Var(D(x,y,t)) \\ &= B_2 \cdot t^2 + B_1 \cdot t + B_0, \end{aligned} \tag{13}$$

where coefficients, B_2, B_1 and B_0, are independent of *t*. Thus, ideally, before and after a dissolve transition, the variance is roughly constant, and during the transition, the variance curve forms a parabolic shape. Based on the same assumptions, the mean can be calculated as:

$$\begin{aligned} \mu(t) &\equiv E[D(x,y,t)] \\ &= E[g_1(t) \cdot f_1(x,y,t) + g_2(t) \cdot f_2(x,y,t)] \\ &= g_1(t) \cdot \mu_1(t) + g_2(t) \cdot \mu_2(t) \\ &= \left(\frac{T-t}{T}\right) \cdot \mu_1 + \left(\frac{t}{T}\right) \cdot \mu_2 \\ &= A_1 \cdot t + A_0, \end{aligned} \tag{14}$$

where coefficients, A_1 and A_0, are independent of *t*. Equation (14) shows that mean curve forms a line during a dissolve. Similarly, the skewness is given by

$$\begin{aligned} & s^3(t) \\ &= Skew(D(x,y,t)) \\ &= E[(D(x,y,t) - \mu_D)^3] \\ &\equiv E\{[g_1(t) \cdot f_1(x,y,t) + g_2(t) \cdot f_2(x,y,t) - \mu_1(t) - \mu_2(t)]^3\} \\ &= C_3 \cdot t^3 + C_2 \cdot t^2 + C_1 \cdot t + C_0, \end{aligned} \tag{15}$$

where the coefficients, C_3, C_2, C_1 and C_0 are independent of *t*. It can be seen that the skewness forms a cubical curve during a dissolve. Skewness characterizes the degree of asymmetry of the distribution around its mean. Two shots with different visual characteristics are expected to have different skewnesses, which are connected by a cubical curve during the dissolve. Another interpretation is that we can easily make connections between moments and distance. The first moment actually defines the *1-norm*, and the *k-th* moment is related to the *k-norm*. Thus, if we only consider the absolute value, the skewness feature is nothing but a number that measures the difference between the distribution and its mean based on a real metric, i.e.:

$$s = d_3(f,\mu) = \| f - \mu \|_3 = [E(|f-\mu|^3)]^{1/3}. \tag{16}$$

The skewness that provides higher order information can be used as a feature for analyzing shot transitions. Figure 5 shows how variance and skewness are affected during a dissolve transition.

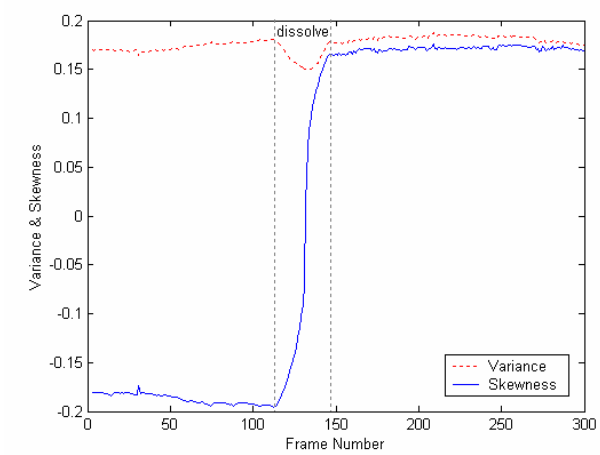


Figure 5: Variance and skewness curves during a typical dissolve.

It can be seen that intensity variance forms a parabola while skewness forms a cubical curve. Both features can capture the dissolve. But, numerically speaking, skewness provides higher

discrimination ability at the shot boundary since its value changes from -0.19 to 0.16 while variance only changes from 0.17 to 0.14.

Another example is shown in Figure 6. This transition contains some extreme factors such as fast camera motion and similar scenes between two shots. Variance and skewness features extracted from the video (see Figure 6) are plotted in Figure 7. In this case the parabola pattern of variance shown in Figure 7 is not obvious due to motion and noise. But skewness is still a good feature to identify the dissolve. When a dissolve joins two similar scenes, at the beginning of the dissolve, the intensity of one shot decreases, but at the same time, it is compensated by similar intensities from the other shot. Such situation can cause serious problems for variance-based approaches. But by exploiting higher order (such as the skewness curve) feature, it is still possible to capture such dissolve transitions.

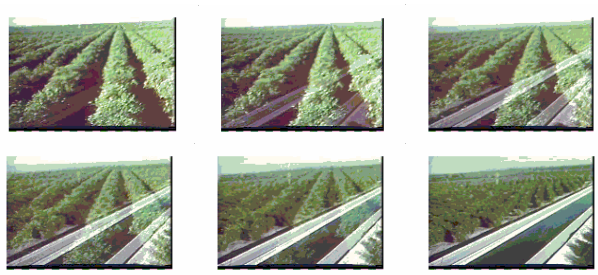


Figure 6: An example of dissolve transition (frames 995, 1000, 1005, 1010, 1015 and 1025 of "*The Miracle of Water*").

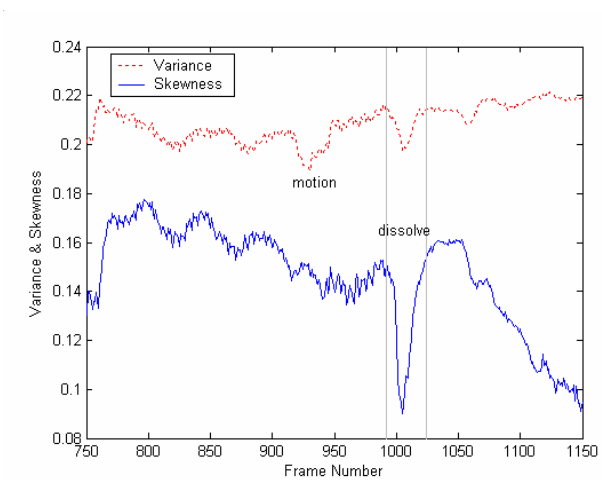


Figure 7: Variance and skewness curves during a dissolve.

In Figure 7, the cubical curve becomes a parabola-like curve when the cubical coefficient is close to zero.

4.2 Dissolve Detection

For dissolve detection, a new method based on a combined analysis of mean-variance-skewness is proposed in this paper. The dissolve detector takes the output of cut detector as input. Therefore, we assume there are no sharp transitions within each input video segment. A temporal sliding window of size *2w+1* with *w=16* and centered at current frame is chosen to adaptively detect the features. Dissolve detection is achieved by three parts consisting of a variance based detection, a mean-skewness detection and a validation phase that is used to remove false positives.

During variance-based detection, a median filter with a length of three is applied to variance curve to remove noise. And then a dissolve transition at time n is detected if all the following conditions are satisfied:

1. Variance $\sigma(n)$ has the minimum value in the sliding window; that is

$$\sigma(n) \le \sigma(k), \forall\, k \in [n-w, n+w]. \tag{17}$$

2. Let $\sigma_{mean}[i,j]$ denote the mean value of $\sigma(k)$ between the interval $[i,j]$, i.e.:

$$\sigma_{mean}[i,j] \equiv mean\{\forall\, \sigma(k), k \in [i,j]\}. \tag{18}$$

In order to match the downwards-parabolic shape of variance curve within the sliding window, the left half of the curve should be decreasing while the right half should be increasing. From this, we have the following two conditions:

$$\sigma_{mean}[n-w, n-w/2] > \sigma_{mean}[n-w/2, n], \tag{19}$$

$$\sigma_{mean}[n, n+w/2] < \sigma_{mean}[n+w/2, n+w]. \tag{20}$$

3. To make sure the curve has deep "valley" and strong "shoulders", the following two conditions should be satisfied:

$$\sigma_{mean}[n-w-w, n-w] - \sigma(n) > T_4, \tag{21}$$

$$\sigma_{mean}[n+w, n+w+w] - \sigma(n) > T_4, \tag{22}$$

where T_4 is a given threshold.

4. A curve fitting with a degree of two is used to further match the parabola pattern in the sliding window. The performance is evaluated by the estimated quadratic coefficient $\hat{B}_2$ (see (13)) and the fitting error. We have

$$\hat{B}_2 > T_5, \tag{23}$$

$$\sum_{i=n-w}^{n+w} \left| \sigma(i) - \hat{\sigma}(i) \right| < T_6, \tag{24}$$

where T_5 and T_6 are given thresholds, and $\hat{\sigma}(i)$, $i \in [n-w, n+w]$, is the value of the estimated curve evaluated at point *i*.

For variance-based detection, the above four conditions are used to match the parabolic shape. A dissolve is detected if all conditions are satisfied.

The mean-skewness detection combines mean and skewness features for shot detection. First, a median filter with a length of three is applied to the skewness curve to remove noise. And then, the first order difference of skewness curve is calculated and its absolute value is used as input. Dissolve detection now becomes measuring the input and finding the large positive spikes. A sliding window with a length of *2w+1* with *w=16* is used to adaptively calculate the local properties. Mean curve is used to

validate the detected transitions. Let $s(k), \forall k \in [1, N]$, denote the skewness curve after being median-filtered. The first order difference $S(k), \forall\, k \in [2, N]$, is given by:

$$S(k) = \left| s(k) - s(k-1) \right|, \forall\, k \in [2, N]. \quad (25)$$

The frame at time *n* is declared as a dissolve boundary if all the following conditions are satisfied:

1. $S(n)$ has the maximum value in the sliding window; that is:

$$S(n) \geq S(k), \forall\, k \in [n-w, n+w]. \quad (26)$$

2. Denote $S_{median}(n)$ as the median value of the sliding window centered at current frame *n*; that is:

$$S_{median}(n) \equiv median\{\forall\, S(k), k \in [n-w, n+w]\}. \quad (27)$$

The difference between the median value and $S(n)$ should be greater than a given threshold T_7; that is:

$$S(n) - S_{median}(n) > T_7. \quad (28)$$

3. A regression line is used to fit the mean curve $\mu(k), \forall\, k \in [n-w, n+w]$, inside the sliding window. This condition requires that the fitting error should be less than a given threshold T_8. From this, we have:

$$\sum_{i=n-w}^{n+w} \left| \mu(i) - \hat{\mu}(i) \right| < T_8, \quad (29)$$

where $\hat{\mu}(i), \forall\, i \in [n-w, n+w]$ is the value of the estimated curve evaluated at point *i*.

A dissolve transition is detected if all three conditions are satisfied.

4.3 False Positive Elimination

In the false positive elimination phase, the shot lists obtained from variance-based method and mean-skewness method are merged into one list for further analysis. If the distance between two consecutive dissolves is less than the length of the sliding window *2w+1*, duplicate entries are defined. In that case, we merge the overlapped dissolves into one dissolve. For each dissolve in the shot list, the histograms from frame *(n+w)* and *(n-w)* are compared to validate the results. If their difference is less than a given threshold T_9, the dissolve is considered as a false positive.

$$\sum_{k=1}^{K} \left| H(f_{n+w}, k) - H(f_{n-w}, k) \right| < T_9. \quad (30)$$

The elimination criterion is based on the assumption that the visual characteristics from two shots are expected to be different.

5. EXPERIMENTAL RESULTS FOR SHOT DETECTIONS

Extensive experiments are performed to test the proposed shot detectors. Two TV shows, "Friends" and "Sex and the city" were selected, and documentary video data were collected from Carnegie Mellon University's The Informedia Project at "The Open Video Project" [14]. Experimental results are presented in Table 1. Precision and recall for hard cuts were obtained as 93.4% and 97.4% respectively. For dissolve detection precision and recall were 73.7% and 82.4%. In TV show "Friends", the three false alarms are caused by object motion, and fade in/out effects. In the show "Sex and the city", fast motion blur is used to connect two scenes. In fact, all three false alarms for hard cuts were caused by such special editing effects. Even though they were counted as errors in our tests, we could argue that they are actually shot boundaries. The documentary video data contain many water scenes and camera motions from close-up to establishing shot. Editing effects such as zoom-ins, zoom-outs, and camera panning are also used extensively. As it can be seen from the table, the general performance of documentary is not as good as TV shows, especially for cut detection. Part of the reason is that some transitions join similar outdoor scenes.

Table 1: Detection results for hard cuts (H) and dissolves (D)

Test Data	Total		Missed		False	
	(H)	(D)	(H)	(D)	(H)	(D)
Friends	73	4	0	1	3	0
Sex & the City	53	3	0	1	3	4
Documentary	64	44	5	7	7	11
Total	190	51	5	9	13	15

In one video clip from the documentary data, among the six dissolve transitions, only one dissolve can be detected if using only variance feature. But all six dissolves can be successfully identified if skewness feature is added. To compare the results with other works, we refer to [4] [6] [13]. For dissolve detection, a precision of 75.1% and recall of 82.2% is reported in [6], and Lienhar [8] obtained a precision of 82.4% and recall of 75% by using Neural Networks. Hanjalic [4] reached a precision of 79% and recall of 83% with a smaller test set containing only 23 dissolves. Best results for dissolve detection still use intensity variance feature [4] [6]. Even though video data collected in the above works were carefully selected to contain as many effects as possible, the performance evaluation from different researchers is still based on different materials. However, by introducing mean-variance-skewness and the combined analysis of these new features, we present new patterns and criteria for analyzing dissolve transitions. The experimental results show that the proposed algorithms are effective for shot boundary detection. Also, the methods are computationally efficient.

6. A WEB-ENABLED INTEGRATED SYSTEM

A system tool is developed to integrate the proposed shot detection algorithms. The graphic user interface of the system is shown in Figure 8. The system provides a frame-level playback. Both manual shot detection and automatic shot detection are supported. After shots are automatically detected, users can edit the shot list, for example, to merge or split shots.

In the proposed system, Synchronized Multimedia Integration Language (SMIL) standard is chosen as multimedia content descriptor. SMIL is a web multimedia format developed by the

World Web Consortium (W3C) and released in 1998. SMIL provides a cross-industry support for synchronized multimedia integration [15]. It is built on Extensible Markup Language (XML) and allows users to write and publish interactive multimedia online. SMIL syntax and semantics can also be incorporated into other XML-based languages for multimedia timing and synchronization. A simple example of hybrid document combing HTML and SMIL is shown below.

```
<html xmlns:t="urn:schemas-microsoft-com:time">
<head>
<?import namespace="t" implementation="#default#time2">
</head>
<body>
...
<input id="button1" type="button"
value="preview" fill="freeze" />
<t:video style="width:100; height:80px;"
   src="./aquarium1.mpeg" clipBegin="00:00:00.000"
   clipEnd="00:00:08.068" begin="button1.Click"
   type="mpeg" />
...
</body>
</html>
```

After shot boundaries are detected, the shot list and other cinematic attributes are managed by a SMIL-based web document. The table-of-contents web-enabled indexing file generated by the tool is shown in Figure 9. Web users can browse and preview the video segments, and jump to a specified location from frame-level. During the implementation, we found that SMIL standard is an effective media description for video structuring and indexing, and its close connection to web makes it very convenient to build and present structured web-enabled multimedia content. Also, keywords and conceptual attributes can be embedded in the SMIL-based indexing file that could be used by existing text-based search engines to realize video web search request.

7. CONCLUSIONS

We have presented new algorithms for shot boundary detection. Cut detection is achieved by choosing the opponent color space that is robust to illumination changes. Dissolve detection is based on a combined analysis of mean-variance-skewness. By introducing these new features and criteria, the proposed dissolve detector has provided a new way to identify and analyze dissolve transitions. Experimental results show that the proposed algorithms can effectively detect both sharp transitions and dissolve transitions, and are computationally efficient. We also presented a system tool to structure and organize the detected shots. Shots and video information are managed and indexed by integrating SMIL web multimedia standard. That makes the system interoperable with existing web-based techniques. The generated indexing file provides functionalities like web-based user interaction, browsing and previewing of video content.

8. REFERENCES

[1] H. Zhang, A. Kankanhalli, and S. W. Smoliar. Automatic Partitioning of full-motion Video. *Multimedia Systems*, 1(1): 10-28, 1993.

[2] R. Zabih, J. Miller, and K. Mai. A Feature-Based Algorithm for Detecting and Classifying Scene Breaks. *Proc. ACM Multimedia 95*, San Francisco, CA, pp. 189-200, Nov. 1995.

[3] B. Yeo and B. Liu. Rapid scene analysis on compressed video. *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, No. 6, pp. 533-544, Dec. 1995.

[4] A. Hanjalic. Shot-Boundary Detection: Unraveled and Resolved? *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, No.2, pp. 90-104, Feb. 2002.

[5] A. M. Alattar. Detecting and Compressing Dissolve Regions in Video Sequences with a DVI Multimedia Image Compression Algorithm. *IEEE International Symposium on Circuits and Systems (ISCAS)*, vol. 1, pp. 13-16, May 1993.

[6] B. T. Truong, C. Dorai and S. Venkatesh. New Enhancements to Cut, Fade, and Dissolve Detection Processes in Video Segmentation. *Proceedings of the 8th ACM International Conference on Multimedia*, pp. 219-227, Nov. 2000.

[7] H. J. Zhang, C. Y. Low, S. W. Smoliar, and J. H. Wu. Video parsing, retrieval and browsing: An integrated and content-based solution. *Proc. of ACM Multimedia'95*, Nov. 1995.

[8] R. Lienhart. Reliable dissolve detection. *SPIE Storage and Retrieval for Media Database 2001*, vol. 4315, pp.219-230, Jan. 2001.

[9] B.L. Tseng, C.Y. Lin and J.R. Smith. Using MPEG-7 and MPEG-21 for personalizing video. *IEEE Multimedia*, vol. 11, No. 1, pp. 42- 52, Jan.-Mar. 2004.

[10] J. Boreczky and L. Rowe. Comparison of video shot boundary detection techniques. *Proc. SPIE Conference on Storage and Retrieval for Image and Video Databases*, vol. 2670, pp. 170-179, 1996.

[11] A. Nagasaka and Y. Tanaka. Automatic video indexing and full-video search for object appearances. *Visual Database Systems*, vol. 2, pp. 113-127, 1992.

[12] A.B. Poirson and B.A. Wandell. Pattern-Color Separable Pathways Predict Sensitivity to Simple Colored Patterns. *Vision Research*, vol. 36, No. 4, pp. 515-526, 1996.

[13] R. Lienhart. Reliable transition detection in videos: A survey and practitioner's guide. *International Journal of Image and Graphics (IJIG)*, vol. 1, no. 3, pp. 469-486, Aug. 2001.

[14] The Open Video Project [Online]. Available: http://www.open-video.org/

[15] J. Ayers *et al*. Synchronized Multimedia Integration Language (SMIL) 2.0. World Wide Web Consortium Recommendation, Aug. 2001 [Online]. Available: http://www.w3.org/TR/smil20/

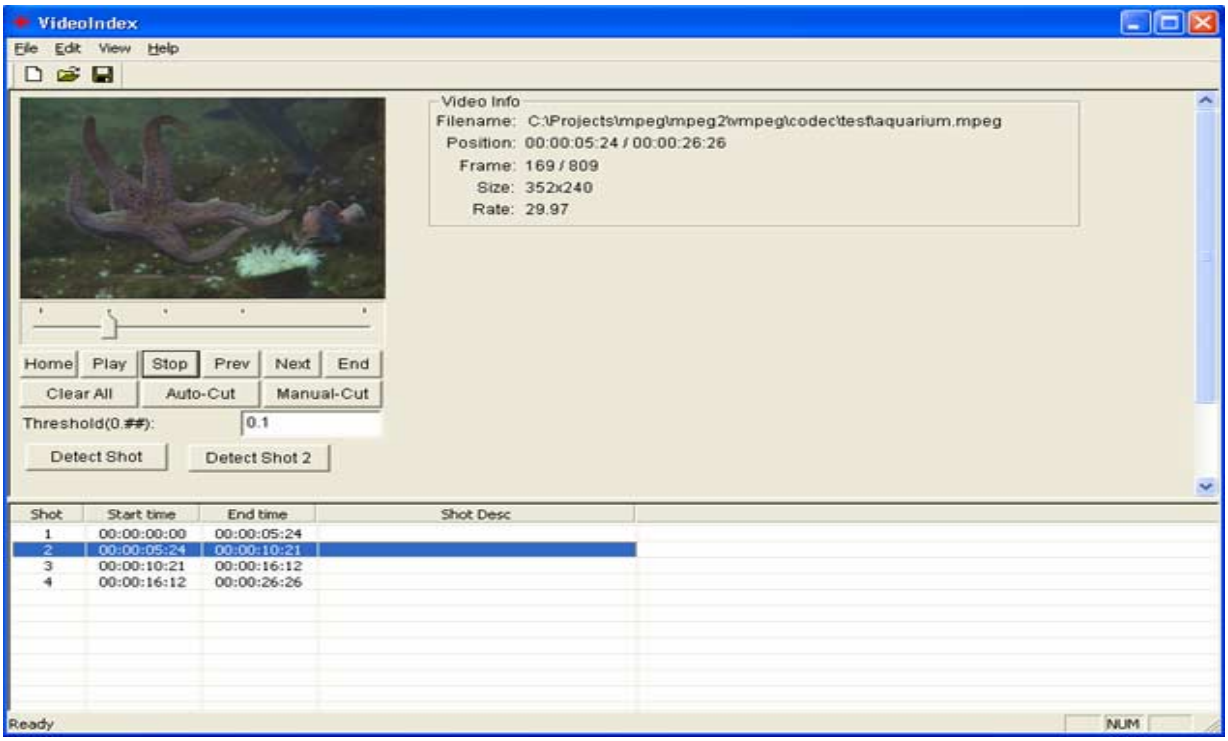


Figure 8: System Graphic User Interface.

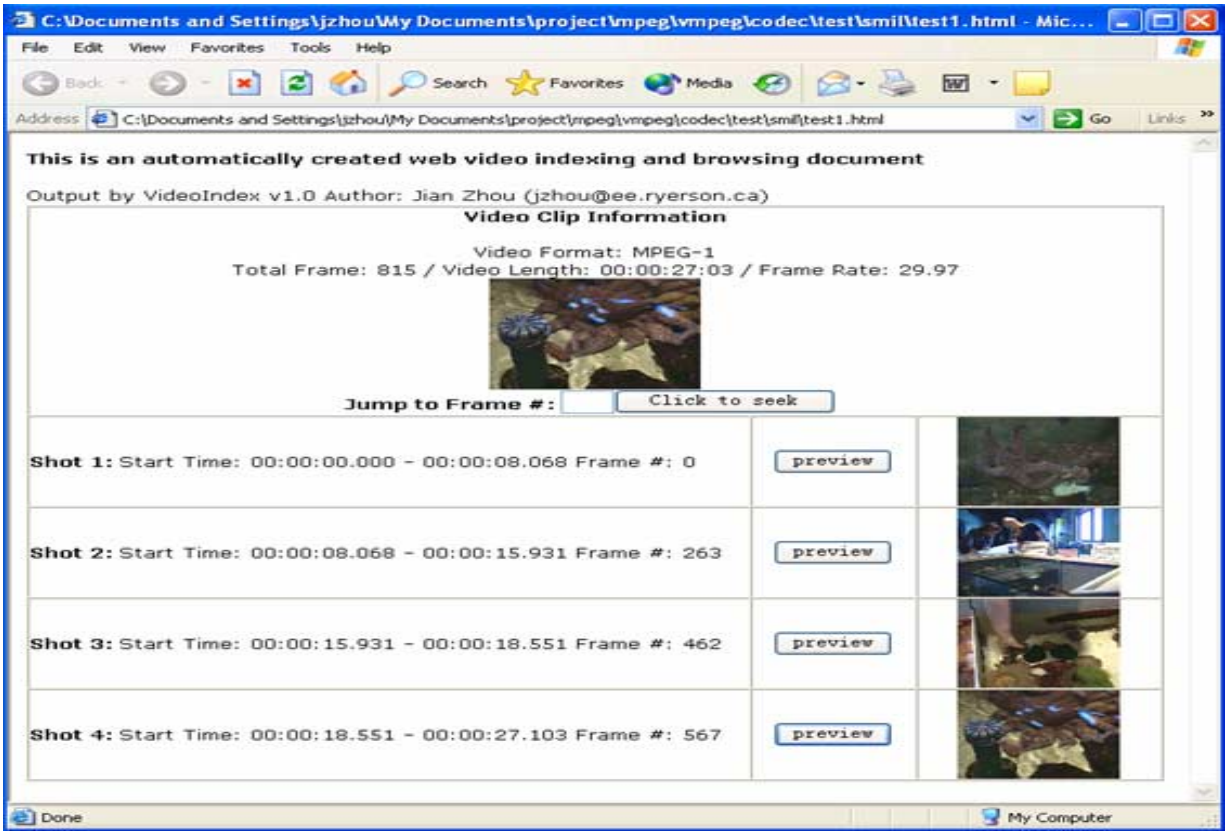


Figure 9: Generated HTML+SMIL indexing file.