Tutorial Proposal for ACM Multimedia 2004 on Understanding Media Semantics

Proposed by: Marc Davis, University of California at Berkeley
                  Chitra Dorai, IBM T.J. Watson Research Center
                  Frank Nack, CWI

I: Topic

The current goal of multimedia research is to make multimedia information pervasively accessible and useable. Achieving this goal requires bridging "the semantic gap," which describes the gulf between the meaningful descriptions that users expect systems to associate with media and the low-level features that systems actually compute. One promising approach to bridging the semantic gap and building high-level semantic descriptions of media content is founded on understanding the semantics of various media within a computationally informed and systematic study of media production and reception. The purpose of this tutorial is to provide an understanding of the role, technologies, and applicability of media semantics to the tasks of managing and reusing various media. Understanding media semantics is essential for progress in media content analysis and applications and will be of great interest to many attendees at ACM Multimedia 2004. In our tutorial we address a number of specific issues:

- The challenge of semantic gap and strategies for bridging it
- Basic communication, semiotic, and media theory underlying media semantics
- The particular semantics of audio-visual media
- The semantics of media combination (especially in video sequencing and editing)
- The applicability of media semantics to capturing, representing, processing, managing, repurposing, and personalizing media
- The future of media semantics

We propose the tutorial this year again for ACM Multimedia in response to the widespread interest and participation from the community in the tutorial we organized at the 2003 ACM MM conference centred on the same theme. Our 2003 session in Berkeley, California was successful and well-attended by a total of 18 participants, with 12 of them as registered attendees. We have also augmented this year's topics to cover semantics of audio.

II: Audience / Format / Requirements / Schedule

Full Day Tutorial. Level Introductory to Intermediate.

This tutorial is designed for researchers and practitioners who would like to learn how to understand the semantics of various media, how to describe them, and how to make use of such descriptions in the whole value chain of media creation, management, distribution, delivery, and reuse. While the tutorial is focused on Media Semantics, given Media Semantics' fundamental import to multimedia systems, the tutorial should also be of interest to people working in: Wireless Multimedia Systems, XML and Multimedia, Content-Based Indexing and Retrieval, Desktop Video Editing, Multimedia on the Web, Digital Asset Management, Media-Rich Homes / Digital Entertainment, Practical Digital Libraries, Multimedia User Interface Design, Multimedia in Collaboration, TREC Video Retrieval, and Multimedia Information Retrieval. The tutorial will include lectures, demonstrations, hands-on exercises, and audience participation. The exercises are introductory and fun, yet provide the direct experience needed to understand and apply Media Semantics concepts. The preliminary schedule is as follows:

8:30 – 9:00   **Welcome, Introductions, and Overview (30 minutes) Presenter: All**
Welcome participants, find out who they are and what they want, provide overview of tutorial goals, schedule, and ground rules
Sections:
Welcome Participants
Provision with tutorial material – printout of all slides + annotated bibliography (suggestions of books, journals, conferences, and other fields of interest) + additional information on media standards (annotated list of relevant industrial, W3C and ISO standards)
Introduce Tutorial Organizers
Self-Introductions of Participants (who they are, where they come from, their background, why they are taking the tutorial, what specific problem they are working on that is relevant)
Overview of Tutorial (goals, schedule, ground rules)

9:00 – 9:30	**Bridging the Semantic Gap (30 minutes) Presenter: Chitra Dorai**
	Explain the limits of signal-based media analysis and the need for media semantics (and syntax)
	Sections:
	Media Signal Analysis State of the Art, Promises and Problems
	The Semantic Gap
	Sensory Gap
	Bridging the Semantic Gap
	The Need for Media Semantics
	Computational Media Aesthetics Approach
9:30 – 10:30	**Fundamentals of Media Semantics (60 minutes) Presenter: Marc Davis**
	Explain the basic concepts of communication theory and semiotics as foundations for media semantics
	Sections:
	Communications Theory
	Introduction to Semiotics
	Semiotics for Media
10:30 – 10:45	**Break (15 minutes)**
10:45 – 11:30	**Semantics of Images (45 minutes) Presenter: Frank Nack**
	Introduce the main semantic concepts of static visuals and how these can be computationally described
	Examples are taken from existing applications
	Sections:
	Image Semantics
	Mise-en-scene Properties
	Photographic Properties
	Semantic Descriptions of Images (Media Streams, W3C and ISO standards as applied to images)
	Introduction to Metadata Structures (hierarchy, inheritance, facets)
11:30 – 12:00	**Exercise: Image and Metadata Creation (30 minutes) Presenter: Marc Davis**
	Provide participants hands-on experience in applying theory and technology of Image Semantics
	Sections:
	Participant teams will be supplied with a semantic rich photo to be described by the team
	Participants discuss Photo Descriptions
	Expected take-away lesson: variations in metadata lead to inter-indexer consistency issues; cross-domain issues; need for standardization of metadata
12:00 – 13:00	**Lunch (60 minutes)**
13:00 – 14:00	**Semantics of Audio (60 minutes) Presenter: Chitra Dorai**
	Provide an overview of audio semantics and representational structures and analysis for audio metadata
	Examples are taken from existing applications
	Sections:
	Audio Segmentation and Classification (speech, music, sound effects, and environmental sounds)
	Audio Content Analysis
	Semantics of Audio
	Standards for Audio Content Description
14:00 – 15:00	**Semantics of Video (60 minutes) Presenters: Marc Davis and Frank Nack**
	Introduce the main semantic concepts of video and how these can be computationally described
	Examples are taken from existing applications
	Sections:
	Relationship to Image Semantics
	Semantics of Time and Motion
	Montage, or the Semantics and Syntax of Editing and Sequencing
	Representing Space, Time, People, and Actions in Video
	Metadata Systems for Temporal Media (Media Streams, MPEG-7, SMIL)
15:00 – 15:15	**Exercise: Temporal Media Analysis and Repurposing (15 minutes) Presenter: Frank Nack**
	Provide participants hands-on experience in applying theory of the Semantics and Syntax of Video
	Sections:
	Introduction to Temporal Media Analysis and Repurposing Exercise
	Participant Teams Re-Edit a Simple Video Sequence
15:15 – 15:30	**Break**
15:30 – 15:45	**Exercise Discussion: Temporal Media Analysis and Repurposing (15 minutes) Presenters: All**
	Participants Discuss Results of Temporal Media Analysis and Repurposing Exercise

Expected take-away lesson: changes in video syntax change the semantics of video shots; need for video metadata that can handle fixed and variable semantics of video data

15:45 – 17:00   Applying Media Semantics (75 minutes) Presenters: All
Describe a few additional applications using media semantics to provide a better understanding of how media semantics help to restructure, represent, resequence, or repurpose media—for each application we provide a description of the application, in particular the relevant issues related to content description (e.g., linking, description structures, feature extraction, etc.), and references
Sections:
Computational Media Aesthetics Applications (Chitra Dorai)
Media Semantics for Active Capture and Adaptive Media Applications (Marc Davis)
MPEG-7 Cameras and Media-Aware Knowledge Spaces (Frank Nack)
Provide an outlook on the potential future of Media Semantics (All)

17:00 – 17:30   Final Discussion (30 minutes) Presenters: All
Provide participants the opportunity to reflect on the various concepts introduced in the Tutorial and allow for feedback, discussion, and questions and answers on the impact on their own research directions
Sections:
Tutorial Summary
Questions and Answers

The required equipment comprises: a video projector and audio speakers to be used with a laptop, a projection screen or surface. The laptops and software needed for the exercises are provided by the organizers. The presentation slides, audio-visual material used for examples and exercises, and a number of relevant theoretical texts will be made available to the participants.

III: Organizing Committee and Backgrounds

Marc Davis
University of California at Berkeley   Tel.: (510) 643-2253
School of Information Management and Systems (SIMS)   Fax: (510) 642-5814
314 South Hall   Email: marc@sims.berkeley.edu
Berkeley, CA 94720-4600   Web: http://garage.sims.berkeley.edu
USA

Marc Davis is an Assistant Professor at the School of Information Management and Systems (SIMS) at the University of California at Berkeley where he directs the Garage Cinema Research group. His work is focused on creating the technology and applications that will enable daily media consumers to become daily media producers. Prof. Davis' research and teaching encompass the theory, design, and development of digital media systems for creating and using media metadata to automate media production and reuse. As a pioneer in the repurposing of digital video, he has done groundbreaking work in video annotation, retrieval, and automatic movie assembly. As part of his doctoral dissertation at the MIT Media Laboratory, he developed Media Streams, an iconic visual language for annotating, retrieving, and repurposing digital video. From 1993 to 1998 at Interval Research Corporation, he led research and development teams in automatic media production technology for which a patent was awarded in 2001. From 1999 to 2002, Prof. Davis was Chairman and Chief Technology Officer of Amova, Inc., a developer of media automation and personalization technology. At the MIT Media Laboratory, Marc Davis co-founded the Narrative Intelligence Reading Group, which innovated interdisciplinary discourse at the intersection of literary and media theory, artificial intelligence, and media technology and design. At UC Berkeley, Prof. Davis is a Co-Founder and Executive Committee Member of the interdisciplinary Center for New Media (CNM), an Advisory Board Member of the Art, Technology, and Culture Colloquium (ATC), and an Affiliated Faculty Member of the Berkeley Institute of Design (BiD). Prof. Davis has numerous publications, presentations, and workshops, to his credit. He was one of the invited contributors to the 50th Anniversary Edition of the Communications of the ACM, for which he wrote a vision piece about the next 50 years of media technology. He authored an invited chapter on video representation for *Readings in Human-Computer Interaction: Toward the Year 2000*, which has become a standard textbook in the field. He was invited to address the MPEG committee that has developed the MPEG-7 standard for multimedia content description based in part on his dissertation work on Media Streams. This November, Prof. Davis participated in the ACM SIGMM Strategic Planning Retreat whose aim was to help set new directions for research in the field of multimedia. With Dr. Frank Nack and Dr. Chitra Dorai, Prof. Davis co-organized the well-attended and successful "Understanding Media Semantics" tutorial for ACM Multimedia 2003 in Berkeley, California, USA. Prof. Davis earned his B.A. in the College of Letters at Wesleyan University, his M.A. in Literary Theory and Philosophy at the University of Konstanz in Germany, and his Ph.D. in Media Arts and Sciences at the Massachusetts Institute of Technology Media Laboratory.

Graduate Courses taught on related subjects at the UC Berkeley School of Information Management and Systems:
IS202 Information Organization and Retrieval   http://www.sims.berkeley.edu/academics/courses/is202/f03
IS246 Multimedia Information   http://www.sims.berkeley.edu/academics/courses/is246/f03

Chitra Dorai
IBM T.J. Watson Research Center   Tel.: (914) 784-6049
P.O. Box 704   Fax: (914) 784-7455
Yorktown Heights, NY 10598   Email: dorai@watson.ibm.com
USA   Web: http://www.research.ibm.com/people/d/dorai

Dr. Chitra Dorai is a Research Staff Member at the IBM T.J. Watson Research Center, New York, where she leads the Media Semantics and e-Learning Media projects. Her research interests are in the areas of multimedia systems and digital video analysis, computer vision, pattern recognition, and machine learning. Her current research focuses on developing technologies for digital media analysis and content management in various domains, such as education and training media and motion pictures that are useful in content-based structuring, annotation and search and smart browsing. Recently, Dr. Dorai has created a new research approach, jointly with Svetha Venkatesh, called "Computational Media Aesthetics" to address the problem of the semantic gap in automatic content annotation and management systems. Computational Media Aesthetics is defined as the algorithmic study of a variety of image and aural elements in media, founded on their patterns of use in film grammar, and the computational analysis of the principles that have emerged underlying their manipulation, individually or jointly, in the creative art of clarifying, intensifying, and interpreting some event for the audience. This media production-guided semantic analysis approach has been well received in the research community with awards.   She recently served as a Guest-Editor of a special issue of IEEE Multimedia dedicated to this theme in spring 2003. She also edited a volume titled *Media Computing: Computational Media Aesthetics* published in June 2002 by Kluwer Academic Publishers. She was an author and contributor of the Videotext Multimedia Description Scheme which is now part of the MPEG-7 standard. With Dr. Frank Nack and Prof. Marc Davis, Dr. Dorai co-organized the well-attended and successful "Understanding Media Semantics" tutorial for ACM Multimedia 2003 in Berkeley, California, USA. Chitra received her Ph.D. from the Department of Computer Science at Michigan State University, where she was a recipient of the Distinguished Academic Achievement Award from the College of Engineering. She is a senior member of the IEEE and a member of the ACM.

Chitra has also organized and presented a tutorial on "3D Objects – Representation, Recognition, and Coding,' at the 14th International Conference on Pattern Recognition, 1998.

Frank Nack
Centrum voor Wiskunde en Informatica (CWI)   Tel.: +31 20 592 4223
Kruislaan 413   Fax: +31 20 592 4312
P.O. Box 94079   Email: Frank.Nack@cwi.nl
1090 GB Amsterdam   Web: http://homepages.cwi.nl/~nack
Netherlands

Dr. Frank Nack is a senior researcher at CWI, currently working in the Multimedia and Human-Computer Interaction group. He obtained his Ph.D. with a thesis on "The Application of Video Semantics and Theme Representation for Automated Film Editing," at Lancaster University, UK. The main thrust of his research is on the representation, retrieval and reuse of media in distributed hypermedia systems, educational hypermedia systems that enhance human communication and creativity, computational assistance for the development, maintenance and usage of hypermedia systems and distributed hypermedia systems, computational applications of media theory & semiotics, automated video editing, interactive storytelling, and computational humour theory. He is an active member of the MPEG-7 standardization group where he served as the editor of the Context and Objectives Document and the Requirements Document, and chaired the MPEG-7 DDL development group. With Prof. Marc Davis and Dr. Chitra Dorai, Dr. Nack co-organized the well-attended and successful "Understanding Media Semantics" tutorial for ACM Multimedia 2003 in Berkeley, California, USA. Frank is on the editorial board of IEEE Multimedia, where he serves as associated editor in chief and edits the Media Impact column.

Graduate Courses taught on related subjects at the Media Department, University of Applied Sciences, Darmstadt, Germany:
Multimedia System Design: Multimedia Techniques I (Media Types and Semantics)
Multimedia System Design: Multimedia Techniques II (Media Application Design)