

1) T1: Full Day (6 November 2005)

Multimedia Collaboration and Human-Centered Multimedia Information Systems

Yong Rui (Microsoft Research, Remond)

Nicu Sebe (University of Amsterdam, Netherlands) and

Alejandro (Alex) Jaimes (FXPAL Japan, Fuji Xerox Co. Ltd)

Synopsis:

This tutorial will take a holistic view on the research issues and applications of *Multimedia Collaboration* and *Human-Centered Multimedia Information Systems* focusing on four main areas: (1) multimedia data: conceptual analysis at different levels (feature, cognitive, and affective); (2) indexing algorithms: context modeling, cultural issues, and machine learning for user-centric approaches; (3) multimodal interaction: visual (body, gaze, gesture), audio (emotion) analysis, body tracking, and sensor fusion; (4) multimedia collaboration: systems for automated lecturing, meetings, and real-time conferencing.

Motivation:

Multimedia lies at the crossroads of many research areas (psychology, artificial intelligence, HCI, etc.) and is used in a wide range of applications. In particular, there are many applications in which humans directly interact with multimedia data.

On one hand, the fact that computers are quickly becoming integrated into everyday objects (ubiquitous and pervasive computing) implies that effective natural human-computer interaction is becoming critical (in many applications, users need to be able to interact naturally with computers the way face-to-face human-human interaction takes place). On the other hand, the wide range of applications that use multimedia, and the amount of multimedia content currently available, imply that building successful multimedia applications requires a deep understanding of multimedia content. Another important aspect is that of multimedia communication and collaboration. It can facilitate students to attend classes from remote, and it can greatly increase information worker's productivity. The success of these systems, therefore, depends highly on two *joint* aspects: (1) the human factors that pertain to multimedia data (human subjectivity, levels of interpretation), and (2) the way humans interact naturally with such systems (using speech and body language) to express emotion, mood, attitude, and attention.

In this tutorial, we take a holistic approach to the multimedia collaboration aspects and to human-centered issues in multimedia information systems. We aim to identify the important research issues, and to ascertain potentially fruitful future research directions in relation to the two aspects above. In particular, we introduce key concepts, discuss technical approaches, and open issues in four areas: (1) multimedia data: conceptual analysis at different levels (feature, cognitive, and affective); (2) indexing algorithms: context modeling, cultural issues, and machine learning for user-centric approaches; (3) multimodal interaction: visual (body, gaze, gesture), audio (emotion) analysis, body tracking, and sensor fusion; (4) multimedia collaboration: systems for automated lecturing, meetings, and real-time conferencing.

The focus of the tutorial, therefore, is on two aspects. First we discuss the interaction techniques formulated from the perspective of key human factors in a user-centered approach to developing *Human-Centered-Multimedia Information Systems*. Second, we will cover examples of collaboration systems to motivate the work, and the underlying technologies to drill deep into fundamental research problems. Specifically, for systems, we will cover a) an automated lecture capture system, b) RingCam: a 360-degree meeting recording system, and c) a real-time room conferencing system with live whiteboard capture. For the technologies, we will cover microphone array sound source localization, real-time person tracking, and probabilistic sensor fusion for speaker tracking using particle filters.

Benefits & List of Topics

This tutorial will enable the participants to understand key concepts, state-of-the-art techniques, and open issues in the following areas:

- **Multimedia indexing:** an overview of how humans perceive, index, organize, and search multimedia content. Discussion of studies in art, psychology, library sciences, and the development of conceptual frameworks for computational frameworks.
- **Human issues:** the role of memory, subjectivity, culture, context, and examples of technical approaches to multimedia analysis and interaction that consider these factors.
- **Multimodal emotion recognition for affective retrieval and in affective interfaces:** approaches to multimedia content analysis and interaction that use speech and facial expression recognition.
- **Machine learning:** adaptive multimodal interfaces and learning of visual concepts from user input for automatic detection and recognition (detection of scenes, objects, or events of interest).
- **Vision for multimodal interaction:** overview of techniques and state of the art in body tracking, gaze detection, and gesture recognition.
- **Multimodal fusion:** technical approaches and issues in combining multiple media (e.g., audio-visual) for multimodal interaction and multimedia analysis.
- **Applications:** traditional and emerging application areas will be described with specific examples in smart conference room research, arts, interaction for people with disabilities, entertainment, and others.

Intended Audience

The tutorial is intended for PhD students, scientists, engineers, application developers, computer vision specialists and others interested in the areas of multimedia information retrieval and human-computer interaction. A basic understanding of image processing and machine learning is a prerequisite.

Schedule & Format

The tutorial will consist of presentations by the organizers and will encourage discussion by the attendees.

Materials

Handouts will include presentation slides. In addition, the following papers (partial list) will be used as references:

- A.T. Duchowski, "A Breadth-First Survey of Eye Tracking Applications," *Behavior Research Methods, Instruments, and Computing*, 34(4):455-70, 2002.
- A. Hanjalic and L-Q. Xu, "Affective video content representation and modeling," *IEEE Trans. on Multimedia*, 7(1):143-154, 2005.
- A. Jaimes and N. Sebe, "Multimodal Human Computer Interaction: A Survey," *IEEE International Workshop on Human-Computer Interaction*, October, 2005.
- A. Jaimes and S.-F. Chang, "A Conceptual Framework for Indexing Visual Information at Multiple Levels", *Internet Imaging 2000*, IS&T/SPIE. January 2000.
- T. P. Minka and R. W. Picard, "Interactive Learning using a 'Society of Models'," *Pattern Recognition*, 30(4), 1997.
- M.R. Naphade and T.S. Huang, "Extracting semantics from audio-visual content: the final frontier in multimedia retrieval," *IEEE Trans. Neural Networks*, 13(4):793-810, 2002.
- S.L. Oviatt and P. Cohen, "Multimodal interfaces that process what comes naturally," *Communications of the ACM*, 43(3):45-48, 2000.
- V.I. Pavlovic, R. Sharma and T.S. Huang, "Visual interpretation of hand gestures for human-computer interaction: a review", *IEEE Trans. on PAMI*, 19(7):677-695, 1997.
- N. Sebe, I. Cohen, and T.S. Huang, "Multimodal emotion recognition," in *Handbook of Pattern Recognition and Computer Vision*, C.H. Chen and P.S.P. Wang eds, chapter 4.1, pp. 387-419, World Scientific, January 2005.
- L. Wang, W. Hu and T. Tan "Recent developments in human motion analysis," *Pattern Recognition*, 36, 585-601, 2003.
- Yong Rui, Anoop Gupta, Jonathan Grudin and Liwei He, Automating lecture capture and broadcast: technology and videography, *ACM Multimedia Systems Journal (Springer)*, 10:3-15 (2004)
- Yong Rui and Zicheng Liu, ARTiFACIAL: Automated Reverse Turing test using FACIAL features, *ACM Multimedia Systems Journal (Springer)*, May 2004
- Yunqiang Chen and Yong Rui Real-time Speaker Tracking Using Particle Filter Sensor Fusion, *Proceedings of the IEEE*, vol. 92, no. 3, pp. 485-494, Mar. 2004.
- Tao Wang, Yong Rui, Jia-guang Sun, Constraint Based Region Matching for Image Retrieval, *International Journal of Computer Vision (IJCV)*, vol.56(1/2), 2004, pp.37-45
- Xiang Sean Zhou, Yong Rui and Thomas Huang, *Exploration of Visual Data*, Kluwer Academic Publishers, ISBN 1-4020-7569-3 August 2003, 208 pp.

About the Organizers

Yong Rui is a Researcher in and the manager of the Multimedia Collaboration team at Microsoft Research Redmond. Dr. Rui is a Senior Member of IEEE and a Member of ACM. He is an Editor of *ACM/Springer Multimedia Systems Journal*, an Associate Editor of *IEEE Transaction on Multimedia*, and on the editorial board of *International Journal of Multimedia Tools and Applications*. He received his PhD from University of Illinois at Urbana-Champaign (UIUC).

Dr. Rui's research interests include computer vision, signal processing, machine learning, and their applications in communication, collaboration, and multimedia systems. He has published one book (*Exploration of Visual Data*, Kluwer Academic Publishers), six book chapters, and over sixty referred journal and conference papers in the above areas. Dr. Rui was on Organizing Committees and Program Committees of ACM Multimedia, IEEE CVPR, IEEE ECCV, IEEE ACCV, IEEE ICIP, IEEE ICASSP, IEEE ICME, SPIE ITCOM, ICPR, CIVR, among others. He is a Program Chair of Int. Conf. Image and Video Retrieval (CIVR) 2006, a Program Area Chair of ICME 2002 and ICME 2005, and Program Co-Chair of IEEE International Workshop on Multimedia Technologies in E-Learning and Collaboration (WOMTEC) 2003. He was on NSF review panel and National Academy of Engineering's Symposium on Frontiers of Engineering for outstanding researchers. Dr. Rui's gives many public speeches at conferences, tradeshow, and internal training sessions. His tutorial on "Multimedia Collaboration" at Pacific-Rim Multimedia (PCM) 2004 is one of the highest rated tutorials.

Yong Rui, PhD., Microsoft Research, Redmont, USA

email: yongrui@microsoft.com

Web: <http://www.research.microsoft.com/~yongrui/>

Nicu Sebe is an assistant professor in the Faculty of Science, University of Amsterdam, The Netherlands, where he is doing research in the areas of multimedia information retrieval and human-computer interaction in computer vision applications. He is the author of the book *Robust Computer Vision—Theory and Applications* (Kluwer, April 2003) and of the upcoming book *Machine Learning in Computer Vision* (Springer, Spring 2005). He was a guest editor of a CVIU special issue on video retrieval and summarization (December 2003) and was the co-chair of ACM Multimedia Information Retrieval Workshops, MIR'03 & MIR'04 (in conjunction with ACM Multimedia conferences). He also was the co-chair of the first Human Computer Interaction Workshop, HCI '04 (in conjunction with ECCV 2004) and is the co-chair of the upcoming IEEE Workshop on Human computer Interaction Workshop (in conjunction with ICCV 2005). He is the guest editor of two special issues on multimedia information retrieval and human computer interaction in ACM Multimedia Systems journal and Image and Vision Computing Journal. He was the technical program chair for the International Conference on Image and Video Retrieval, CIVR 2003. He was a visiting researcher in the Beckman Institute, University of Illinois at Urbana-Champaign (2002) and was a research fellow of the British Telecomm in Ipswich (2003). He has published more than 50 technical papers in the areas of computer vision, content-based retrieval, pattern recognition, and human-computer interaction and has served on the program committee of several conferences in these areas. He is a member of the IEEE and the ACM.

Nicu Sebe, PhD., Intelligent Sensory & Information Systems Group

Faculty of Science, University of Amsterdam

direct tel: +31-20-525-7552, fax: +31-20-525-7490

email: nicu@science.uva.nl

Web: <http://www.science.uva.nl/~nicu>

Alejandro Jaimes is an Advanced Multimedia Specialist at FXPAL Japan, Fuji Xerox, where he leads the efforts in Multimedia Analysis and Interaction. Dr. Jaimes received a Ph.D. in Electrical Engineering (2003) and a M.S. in Computer Science from Columbia University (1997) in New York City. He holds a Computing Systems Engineering degree from Universidad de los Andes (1994) in Bogota, Colombia. Prior to joining the Ph.D. program at Columbia he was a member of Columbia's Robotics and Computer Graphics groups, where he worked on projects related to computer vision and computer graphics. He has held summer research positions at AT&T Bell Laboratories, Siemens Corporate Research, and IBM (TJ Watson and Tokyo Research Laboratories). His recent professional activities include co-chairing the ACM Multimedia 2005 and 2004 Interactive Art program, and the PCM 2004 and ICME 2004 special sessions on "Immersive Conferencing: Novel Interfaces and Paradigms for Remote Collaboration," and "Novel Techniques for browsing in Large Multimedia Collections" respectively. He is co-founder and co-chair of the Workshop on Technology for Education in Developing Countries (TEDC '05, TEDC '04, TEDC '03), and serves as the TPC member for several international conferences (ICME, ICIP, CIVR, ICCV and ECCV Workshops on HCI, etc.), among others. His work has led to over 35 technical publications in international conferences and journals, and to numerous contributions to the MPEG-7 standard. He has 7 patents pending. He is a member of the IEEE and ACM.

Alejandro (Alex) Jaimes, Ph.D., FXPAL Japan, Fuji Xerox Co., Ltd.
direct tel: +81-465-80-2081, fax: +81-465-81-8951
email: alex.jaimes@fujixerox.co.jp
Web: <http://www.ee.columbia.edu/~ajaimes>