Tutorial Proposal for ICME 2007

Human-Centered Multimedia

Beijing, July 2007

By Thomas S. Huang, University of Illinois at Urbana-Champaign (USA)
Alejandro (Alex) Jaimes, FXPAL Japan, Fuji Xerox Co. Ltd. (Japan)
Nicu Sebe, University of Amsterdam (The Netherlands)

I. Synopsis

This tutorial will take a holistic view on the research issues and applications of *Human-Centered Multimedia* focusing on three main areas: (1) multimodal interaction: visual (body, gaze, gesture) and audio (emotion) analysis; (2) image databases, indexing, and retrieval: context modeling, cultural issues, and machine learning for user-centric approaches; (3) multimedia data: conceptual analysis at different levels (feature, cognitive, and affective).

II. Motivation

Human-computer Interaction lies at the crossroads of many research areas (computer vision, multimedia, psychology, artificial intelligence, pattern recognition, etc.) and is used in a wide range of applications. In particular, we are aiming at developing human-centered information systems. The most important issue here is how to achieve synergism between man and machine. The term “human-centered” is used to emphasize the fact that although all existing vision systems were designed with human uses in mind, many of them are far from being user friendly. What can the scientific/engineering community do to effect a change for the better?

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 computer vision and multimedia applications requires a deep understanding of multimedia content.
The success of human-centered vision systems, therefore, depends highly on two *joint* aspects: (1) the way humans interact naturally with such systems (using speech and body language) to express emotion, mood, attitude, and attention, and (2) the human factors that pertain to multimedia data (human subjectivity, levels of interpretation).

In this short course, we take a holistic approach to the human-centered multimedia problem. 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 three areas: (1) multimodal interaction: visual (body, gaze, gesture) and audio (emotion) analysis; (2) image databases, indexing, and retrieval: context modeling, cultural issues, and machine learning for user-centric approaches; (3) multimedia data: conceptual analysis at different levels (feature, cognitive, and affective).

The focus of the short course, therefore, is on technical analysis and interaction techniques formulated from the perspective of key human factors in a user-centered approach to developing *Human-Centered Multimedia*.

III. Benefits & List of Topics

This short course will enable the participants to understand key concepts, state-of-the-art techniques, and open issues in the areas described below. In relation to ICME, the tutorial will cover parts of the following topic areas:

. • Active and real-time vision
. • Image databases, indexing and retrieval
. • Learning in vision
. • Model acquisition and validation
. • Tracking and surveillance
. • Object, event, and scene recognition
. • Segmentation and grouping
. • Applications

In particular, it will cover the following areas:

. • Vision for multimodal interaction: overview of techniques and state of the art in body tracking, gaze detection, and gesture recognition.

. • 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).

. • Multimodal fusion: technical approaches and issues in combining multiple media (e.g., audio-visual) for multimodal interaction and multimedia analysis.

. • 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.

. • 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.

IV. Intended Audience

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

V. Tentative Schedule & Format

Duration: half-day Format: the short course will consist of presentations by the organizers and will encourage discussion by the attendees.

VI. Materials

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

. • J. Flanagan and T.S. Huang, Special Issue on Human-computer Multimodal Interface, *Proceedings of the IEEE,* volume 91, number 9, 2003.

. • 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," U. Amsterdam Technical Report, March 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.

. • V. Pavlovic, A. Garg, and J. M. Rehg, “Boosted learning in dynamic Bayesian networks for multimodal speaker detection, “*Proceedings of the IEEE*, 91(9):1355-1369, 2003.

. • 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.

VII. Related Tutorials

This short course has been specifically designed with the audience of ICME in mind. Previous, but different tutorials given by the authors on related topics include the following:

• IEEE/ACM Pacific Rim Conference on Multimedia (PCM 2004), Tokyo, Japan (Sebe, Jaimes)

- **27 registered participants attended**

. **• IEEE International Conference on Multimedia and Expo 2005, Amsterdam, The Netherlands, July 2005 (Sebe, Jaimes)**

. - **40 registered participants attended**

. **• IEEE International Conference on Computer Vision 2005, Beijing, China, October 2005 (Sebe, Jaimes, Zhang)**

• IEEE International Conference on Computer Vision and Pattern Recognition 2006, New York, June 2006 (Huang, Jaimes, Sebe)

• IEEE International Conference on Pattern Recognition 2006, Hong Kong, August 2006 (Huang, Jaimes, Sebe)

The ICME tutorial differs from the CVPR and ICCV tutorials in that a stronger focus will be given to multimodal analysis and fusion, machine learning and multimodal interaction. Indexing and retrieval of image databases will also be covered.
Regarding the ICPR tutorial, the intention is to incorporate our new research results as well as to include a large section on human centered multimedia.

VIII. Organizers and Backgrounds

Thomas S. Huang, Beckman Institute, University of Illinois at Urbana-Champaign
email: huang@ifp.uiuc.edu
Web: http://www.beckman.uiuc.edu/profiles/faculty/t-huang1.html

Alejandro (Alex) Jaimes, 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

Nicu Sebe, 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

Bios:

Thomas S. Huang received the B. S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, R.O.C., and the M.S. and Sc.D. degrees in electrical engineering from the Massachusetts Institute of Technology (MI), Cambridge. He was with the Faculty of the Department of Electrical Engineering at MIT from 1963 to 1973 and with the Faculty of the School of Electrical Engineering and Signal Processing at Purdue University, West Lafayette, IN from 1973 to 1980. In 1980, he joined the University of Illinois at Urbana-Champaign, where he is now William L. Everitt Distinguished Professor of Electrical and Computer Engineering, Research Professor at the Coordinated Science Laboratory, and

Head of the Image Formation and Processing Group at the Beckman Institute for Advanced Science and Technology. He is also Co-Chair of the Institute's major research theme (Human Computer Intelligent Interaction). During his sabbatical leaves, he has been with the MIT Lincoln Laboratory, Lexington, MA; the IBM Thomas J. Watson Research Center, Yorktown Heights, NY; and the Rheinishes Landes Museum, Bonn, West Germany. He held visiting professor positions at the Swiss Federal Institutes of Technology, Zurich and Lausanne, Switzerland; University of Hannover, West Germany; INRS-Telecommunications, University if Quebec, Montreal, QC, Canada; and University of Tokyo, Tokyo, Japan. He has severed as a consultant to numerous industrial forms and government agencies both in the United States and abroad. His professional interests lie in the broad area of information technology, especially the transmission and processing of multidimensional signals. He has published 14 books and over 500 papers in network theory, digital filtering, image processing, and computer vision. He is a Founding Editor of the *International Journal Computer Vision*, *Graphics*, and *Image Processing* and Editor of the *Springer Series in Information Sciences*, published by Springer Verlag.

Dr. Huang is Member of the National Academy of Engineering; a Foreign Member of the Chinese Academies of Engineering and Sciences; and a Fellow of the International Association of Pattern Recognition and of the Optical Society of America. He has received a Guggenheim Fellowship, an A. V. Humboldt Foundation Senior U.S. Scientist Award, and a Fellowship from the Japan Association for the Promotion of Science. He received the IEEE Signal Processing Society's Technical Achievement Award in 1987 and the Society Award in 1991. He was awarded the IEEE Third Millennium Medal in 2000. In addition, in 2000, he received the Honda Lifetime Achievement Award for "contributions to motion analysis." In 2001, he received the IEEE Jack S. Kilby Medal. In 2002, he received the King-Sun Fu Prize from the International Association of Pattern Recognition and the Pan Wen-Yuan Outstanding Research Award.

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.

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 following books: Robust Computer Vision—Theory and Applications (Kluwer, April 2003) and Machine Learning in Computer Vision (Springer, May 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 Human Computer Interaction Workshops, HCI ’04, HCI '05, and HCI '06 (in conjunction with ECCV 2004, ICCV 2005, and ECCV 2006). He was also the co-chair of the ACM Workshop on Human-centered Multimedia (in conjunction with ACM Multimedia 2006).
He is the guest editor of three special issues on multimedia information retrieval and human computer interaction in ACM Transactions on Multimedia Computing, Communication, and Applications, ACM Multimedia Systems, and Image and Vision Computing. 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, UK (2003). He has published more than 60 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.