MULTIMEDIA COMMUNICATIONS TECHNICAL COMMITTEE
IEEE COMMUNICATIONS SOCIETY
http://www.comsoc.org/~mmc

E-LETTER

Vol. 6, No. 1, January 2011

IEEE COMMUNICATIONS SOCIETY

CONTENTS

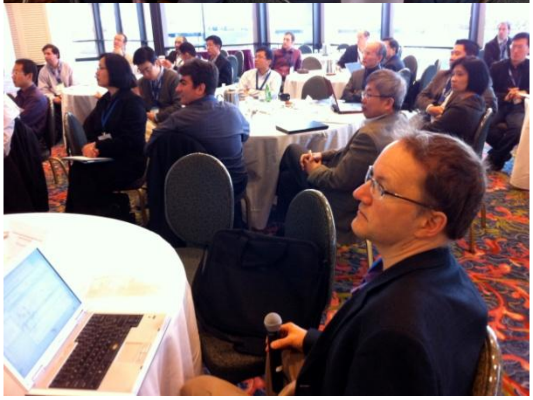
MESSAGE FROM MMTC CHAIR

Dear MMTC fellow members,

At the beginning of year 2011, I wish all of you a healthy, prosperous, and happy new year! I also wish 2011 become another successful year for MMTC community with strong growth and many achievements in both research efforts and professional activities.

MMTC had its last meeting 3 weeks ago at Miami during IEEE GLOBECOM 2010. As many TC members had attended and observed, this year's GLOBECOM turned out to be the most successful meeting in the conference's 53-year history, as this annual flagship event of the IEEE Communications Society (ComSoc) set numerous milestones including number of attendees (2,500+); symposia and workshop paper submissions (4,614); lowest tutorial acceptance ratio (8%); highest number of tutorial registrations with over 300 attendees per session; largest team of volunteers (8,500+) contributing to the conference's success; and the largest number of high-profile invited speakers (100+) delivering keynote and plenary sessions, technical symposia, workshops, panels and tutorials (469). The pictures below demonstrate how the tutorials and talks were attended.

Around fourty TC members attended our 2-hour TC meeting on Dec. 9 (Thursday) at the Intercontinental Hotel, Miami, which covered presentation and discussions on around 25 topics including conference activities reports from GLOBECOM 2010, ICC 2011, ICME 2011 and GLOBECOM 2011, IG reports from all 13 IGs, Board reports from membership, award, and service boards, activity reports from E-Letter and R-Letter, and status reports from ICME and TMM steering committee. Please see below 2 snapshots of the meeting room, and the full attendee list can be found in next page.



In order to create a better home event for MMTC members to meet and gather, TC decides to found an annual workshop called MMCom (International Workshop on Multimedia Communications), to be held in conjunction with the future GLOBECOMs in December every year. I would like to thank the following members who take the responsibility to co-chair the first version of MMCom (MMCom'11) at Houston, Texas.

- Prof. Dr. Thomas Magedanz, Fraunhofer FOKUS/ T. U. Berlin, Germany
- Prof. Jiangtao (Gene) Wen, Tsinghua University, China
- Dr. Xiaoli Chu, King's College, UK
- Prof. Yung-Hisang Lu, Purdue University, USA

More details of MMCom'11 will be announced shortly, please support this new event by submitting your papers.

On the other hand, I would like to call for nominations for the following 2 opportunities:

- TPC Chair of IEEE ICME 2012 at Melbourne, Australia, representing MMTC
- Steering Committee member of IEEE CCNC, representing MMTC

If you are interested at any of the opportunities above, please send your self-nomination to me at haohongwang@gmail.com by **Feb. 1, 2011** with your bio and short description of your background and experiences in conference organizing. The election will be conducted by our 40 IG Chairs and they will select the final winners of these two positions, based on the same procedure we used before for the ICME Steering Committee voting member election.

At last but not the least, I would like to encourage our members to submit papers to IEEE ICME 2011 workshops at Barcelona, Spain (http://www.icme2011.org/). The paper submission deadline is **Feb. 20, 2011**.

Thank you very much!

Haohong Wang
Chair of Multimedia Communication TC of IEEE ComSoc

Appendix
Attendee List of MMTC meeting @ Miami

Bin Wei	AT&T Labs - Research
John Buford	Avaya Labs
ChangWen Chen	SUNY-Buffalo
Wanjiun	National Taiwan University
C. C. Jay Kuo	University of Southern California
Dan Keun Sung	KAIST
Lei Cao	University of Mississippi
Jianwei Huang	The Chinese University of Hong Kong
LiSong Xu	University of Nebraska-Lincoln
Vincent Wong	University of British Columbia
Lin Cai	University of Victoia
Andres Kwashski	Rochester Insitute of Technology
Martin Reisslein	Arizona State Universtiy
Khaled El-Maleh	Qualcom
ChongGang Wang	InterDigital Communications
Yung-Hsisang Lu	Purdue University
Wendi Heinzelman	University of Rochester
Leonardo Badia	IMT Lucca
Amdrea Zamecca	University of Padova
Philip Chou	Microsoft Research
XianBin Wang	University of Western Ontario
Hsiao-Chun Wu	Louisiana State University
Shiwen Mao	Auburn University
Zhu Liu	AT&T Labs - Research
Jaim Lloret	Polytechnic University of Vaeencia
Pascal Lorenz	University of Kaule Alsace
Madjid Merabti	Liverpool John Moore University UK
DaPeng Oliver Wu	University of Florida
Sanjeer Mebrotra	Microsoft Research
Jin Li	Microsoft Research
Tsungnan Lin	National Taiwan University
Luigi Atzori	University of Cagliari (IT)
Tansu Alpcar	Tech. Univ. Berlin (Ger)
LingFen Sun	University of Plymouth (UK)
Vince Poor	Princeton University
Rob Fish	NETovations Group
Xi Zhang	Texas A&M University

SPECIAL ISSUE ON HUMAN-CENTRIC MULTIMEDIA COMMUNICATIONS

Advances in Human-Centric Multimedia Communications

Guest Editor: Zhenzhong Chen, Nanyang Technological University, Singapore
zzchen@ieee.org

Recent advances in broadband networking and multimedia technologies have led to the explosion of multimedia services. With the rapid growth of demand for multimedia content production, access, and distribution, the evolution has changed from the static system-centric communications to the dynamic human-centric communications including producing, sharing, and interaction. Human-to-computer and human-to-human technologies have led to many innovations by prompting standard resolution digital video broadcasting to immersive multimedia entertainment, low-quality video conferencing to high-end telepresence, traditional audio-video study to advanced haptic-audio-visual framework. The advanced applications not only require overcoming challenges of system-level services but also enhancing quality of experience (QoE) for end users. In this special issue, the editorial team has the great honor to invite some pioneer researchers for eight invited papers to present their state-of-the-art accomplishments, share their latest experiences, and outline future directions in human-centric multimedia communications.

The first paper, titled “Perceptual Haptic Data Reduction in Telepresence and Teleaction Systems”, addresses several key issues of the efficient communication of haptic signals. The authors introduce typical telepresence and teleaction (TPTA) systems and discuss the importance of haptic compression/data reduction scheme in such a communication system. This article provides some useful guidelines for the system design and implementation for human-machine interaction, virtual reality, etc.

In the second article, “Gaze Awareness and Eye Contact in Multimedia Communications”, the authors discuss the gaze awareness and eye contact in immersive multimedia communications which enrich the user experience in the collaboration meeting. Different solutions of the video cross-talk reduction are introduced. The developed system shows the promising natural interaction for remote users.

Using multimedia technologies to assist the online education plays an important role in today’s education. The authors of the third paper, titled “Multimedia Technology for Next Generation Online Lecture Video”, introduces the advances in multimedia technologies to enhance the user’s view experience. They present their online lecture platform, Stanford ClassX, in which some user friendly features, such as instructor tracking and interactive pan/tilt/zoom, are integrated.

In the paper “The Context Does Matter: Beyond the Data Pipes of Today”, the authors not only introduce their work on the context-aware techniques for online social networks, but also share their views on how context-driven applications will impact user collaborations and interactions.

The authors of the paper “Computational Audio-Visual Scene Analysis” present a real-time computational audio-visual scene analysis (CAVSA) system. It consists of microphone sensors, multicore processers, cameras, as well as various software modules. It will have long-term impacts on multimedia applications such as surveillance and tele-presence.

Interactive multimedia streaming is a key issue in human-centric multimedia communications. Two papers in this special issue present novel solutions in this area. The paper, titled “Autonomous infrastructures for networked interactive and personalized media experience”, introduces an interactive video streaming system as well as a personalized video summarization framework. The paper “High-dimensional Media Compression for Interactive Streaming” considers a challenging issue in interactive streaming, i.e., how to compress high-dimensional data. Several specific applications are discussed. The authors also point out the importance of integrating coding and streaming techniques.

The paper “Semantic Image Adaptation for User-centric Mobile Display Devices” addresses the human factors in mobile applications and presents a semantic image adaptation scheme. A Bayesian fusion approach is developed for low level features and high level semantics. This novel system brings mobile users better visual experience.

I hope you will enjoy this special issue dedicated to human-centric multimedia communications.

Zhenzhong Chen received the B.Eng. degree from Huazhong University of Science and Technology (HUST) and the Ph.D. degree from Chinese University of Hong Kong (CUHK), both in electrical engineering. His current research interests include video signal processing, visual perception, and multimedia communications. He is currently a Lee Kuan Yew research fellow at Nanyang Technological University (NTU), Singapore. Before joining NTU, he was an ERCIM fellow at National Institute for Research in Computer Science and Control (INRIA), France. He held visiting positions at Universite Catholique de Louvain (UCL), Belgium, and Microsoft Research Asia, Beijing.

He serves as voting member of IEEE Multimedia Communications Technical Committee (MMTC), invited member of IEEE MMTC Interest Group of Quality of Experience for Multimedia Communications (QoEIG) (2010-2012), technical program committee member of IEEE ICC, GLOBECOM, CCNC, and ICME. He has co-organized several special sessions at international conferences, including IEEE ICIP 2010, IEEE ICME 2010, and Packet Video 2010. He serves as a guest editor of Journal of Visual Communication and Image Representation. He received CUHK Faculty Outstanding Ph.D. Thesis Award, Microsoft Fellowship, and ERCIM Alain Bensoussan Fellowship. He is a member of IEEE.

Perceptual Haptic Data Reduction in Telepresence and Teleaction Systems

Fernanda Brandi, Rahul Chaudhari, Sandra Hirche, Julius Kammerl, Eckehard Steinbach, and Iason Vittorias, Technische Universität München, Munich, Germany
{fernanda.brandi, rahul.chaudhari, hirche, kammerl, eckehard.steinbach, vittorias}@tum.de

1. Introduction

Vision and hearing play a significant role in the perception of our surroundings. This fact has aptly justified and reinforced our inclination of focusing research in man-machine interaction traditionally on these modalities. Inspired by the recent progress in human-machine interaction, robotics, and augmented reality, contemporary scientists and engineers are concentrating efforts towards seamlessly integrating the haptic modality with the well established ones of audio and video. This realization is rapidly gaining the field of haptics (from the Greek *haptikos*, pertaining to the sense of touch), the attention that it has rightfully deserved.

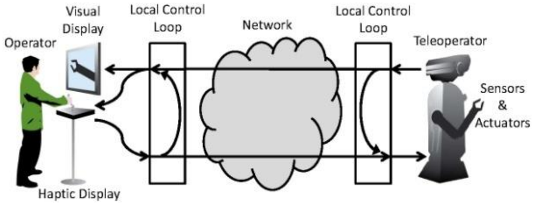


Figure 1: Schematic overview of a visual-haptic telepresence and telemanipulation system (adapted from [1]).

In particular, typical telepresence and teleaction (TPTA) systems rely on haptics. In these systems a human operator controls a remote teleoperator through a human-system interface device. As soon as the teleoperator encounters contact with its surroundings, corresponding feedback is transmitted to be displayed to the human operator. The communication system therefore closes a global control loop. In this way, the TPTA system enables the perception of objects including physical manipulation as well sensing their material properties. An overview of a TPTA system is illustrated in Fig. 1.

Communication unreliabilities, such as time delay and packet loss, can jeopardize the system's stability resulting in dangerous unbounded oscillations of the devices. Besides that, reduction of the transmitted haptic data requires proper reconstruction on the receiver side to guarantee a stable system.

For reasons of stability, haptic data samples are typically transmitted immediately upon generation on a 1 packet/sample basis for packet-based networks like the Internet. This leads to a packet being triggered for transmission at every millisecond (corresponding to the stringently required haptic update rates of 1 kHz). Essentially, too many packets are generated too fast with relatively less worth of information conveyed. Previous studies have established the fact that such high packet-rates are difficult to maintain over general-purpose networks like the Internet [2]. A good haptic compression/data reduction scheme should solve this predicament.

2. Perceptual coding of haptic signals

Perceptual deadband coding (PDC) schemes are developed to cope with these challenges [3]. The PDC approach exploits the limitations of human haptic perception for lossy data reduction. It is summarized by a simple mathematical relationship between the physical intensity of a haptic stimulus and its phenomenologically perceived intensity (known as the Weber's law). When trying to perceive the difference between physical quantities (e.g. weights) in succession, it is not the difference itself that makes an impression upon us, rather the ratio of this difference to the magnitude of the first quantity. This ratio is constant and is denoted by *k*, a percentile value. We translate this observation that Weber made to our field of interest, namely perceptual haptic data reduction. By defining perceptual thresholds – the so called deadband – we can distinguish perceivable changes from unperceivable changes in haptic signals. By transmitting only those haptic samples which lead to a perceivable change, we can significantly decrease the packet rate on the network without impairing the user experience. The samples skipped from transmission are approximated at the other side by simple interpolation schemes like the "hold last sample"-approach or via linear prediction. The size of such a deadband is controlled by the parameter *k*. The greater the *k* value is, the larger the deadband and hence the applied perceptual thresholds. Substantial average haptic data reduction of up to 85-90% of the otherwise bulky data is obtained using this approach.

3. Multiple-Degree-of-Freedom Extension

Real-world TPTA systems deploy typically more than one degree-of-freedom (DoF). In order to enable perceptual data reduction in multi-DoF TPTA scenarios, [4] proposes the construction of an isotropic deadzone. In two dimensions, the deadzone can be described by a circular; in three dimensions by a spherical region which is centered at the tip of the currently applied haptic sample vectors. Furthermore, its radius is defined to be a fraction of the haptic sample magnitude. However, when extending the haptic data reduction schemes from a single-DoF to multi-DoFs, the spatial orientation of haptic sample vectors acts as an additional perceptual domain. Therefore, its influence on haptic perception thresholds is to be investigated. In this context, psychophysical experiments in [5] reveal that haptic force feedback perception is a function of the spatial orientation of the force feedback itself. In order to adopt the perceptual deadband scheme to these findings, [6] proposes the construction of a novel deadzone shape that takes the form of a frustum of a cone. In this way, the perceptual data reduction approach can reflect the dependencies of the spatial direction of force feedback onto the perceptual thresholds which allows for a significant improvement in data reduction performance.

4. Control Issues

Several control architectures have been proposed to enable stable TPTA sessions in the presence of communication unreliabilities. To guarantee stability when there is an arbitrarily large constant time delay in the network the scattering transformation is proposed in [7]. Instead of the power conjugated variables, i.e. force and velocity, a linear combination of them is transmitted. The time-varying delay and packet loss challenge are addressed in [8] and [9] respectively.

By requiring each subsystem of the TPTA system to dissipate energy, and therefore be more conservative, stability can be guaranteed. Using the same rationale for the data reconstruction strategies of a haptic data reduction algorithm, stability is shown for the PDC approach in [10]. For robotic systems with more than one degree-of-freedom, the corresponding data reduction algorithm and an optimization-based reconstruction strategy are presented in [11].

The selected control architecture determines apart from stability, the robustness and transparency of the TPTA system. A detailed discussion exceeds the scope of this article and the reader is referred to the survey article [12].

5. Error-Resilient Haptic Data Reduction

Internet-based TPTA systems are subject to packet delays, jitter and packet losses. Particularly, when packet losses occur in haptic communication while using the PDC scheme in combination with predictive coding, several artifacts can be observed on the reconstructed signal such as bouncing, increased roughness and a “glue effect” [13].

Due to strict delay constraints traditional packet loss compensation strategies such as retransmissions based on time-outs are not feasible for haptic communication. Therefore, in order to achieve error-resilient haptic data reduction, [13] proposes the construction of a Markov tree similar to [14] that enables the estimation of the most likely state of the receiver. This allows us to adaptively add redundancy to the haptic channel if the estimated state at the receiver significantly deviates from the desired signal trajectory and if this deviation becomes perceivable. Combined with the state-of-the-art haptic data reduction approaches we can achieve perceptual error-resilient haptic communication. A more extensive discussion of the challenges of haptic communication can be found in [15].

6. Conclusion and Future Work

With recent advances in haptic technology, the efficient communication of haptic signals is gaining relevance. Integrating more degrees-of-freedom leads to an increased amount of data and strict delay constraints result in high update packet rates in the network. We address this challenge by deploying a mathematical model of human perception for multiple degrees-of-freedom which allows the reduction of the packet rate by up to 90%. Moreover, the stability issues due to packet losses and delays have been successfully addressed from a control engineering perspective. An error-resilient perceptual coding for networked haptic interaction has been developed allowing the haptic communication to operate seamlessly while operating in the presence of adverse communication conditions.

Future work will address the extension of the model of the human haptic perception by integrating additional findings from psychophysics, such as multimodal dependencies and dynamic perception thresholds. A comprehensive psychophysical model furthermore enables the development of novel methods for objective quality evaluation. The complexity of the error-resilient haptic communication approach is also to be decreased and its efficiency further improved.

Acknowledgment

This work has been supported, in part, by the German Research Foundation (DFG) within the Collaborative Research Center SFB 453 on high-fidelity telepresence and teleaction and project STE 1093/4-1. The financial support for Fernanda Brandi by the Erasmus Mundus ECW EUBRANEX is gratefully acknowledged.

References

[1] W. R. Ferrell and T. B. Sheridan. “Supervisory control of remote manipulation.” in IEEE Spectrum, 4(10):81–88, Oct. 1967.
[2] C. Mahlo, C. Hoene, A. Rosami, and A. Wolisz, “Adaptive coding and packet rates for TCP-friendly voip flows,” in Proc. 3rd Int. Symp. Telecommun., Shiraz, Iran, Sep. 2005.
[3] P. Hinterseer, S. Hirche, S. Chaudhuri, E. Steinbach, and M. Buss, “Perception-based Data Reduction and Transmission of Haptic Data in Telepresence and Teleaction Systems,” in IEEE Trans. on Signal Processing, vol. 56, no. 2, 2008.
[4] P. Hinterseer and E. Steinbach, “A psychophysically motivated compression approach for 3d haptic data,” in Proc. if the 14th Symp. on Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2006, Arlington, VA, USA, Mar. 2006, pp. 35–41.
[5] H. Pongrac, B. Färber, P. Hinterseer, J. Kammerl, and E. Steinbach, “Limitations of human 3d force discrimination,” in Human-Centered Robotics Systems 2006, Munich, Germany, Oct. 2006.
[6] J. Kammerl, R. Chaudhari, E. Steinbach, “Exploiting Directional Dependencies of Force Perception for Lossy Haptic Data Reduction,” in Proc. of the Int. Symp. on Haptic Audio Visual Environments and Games (HAVE), Phoenix, Arizona, USA, Oct. 2010.
[7] R. J. Anderson and M. W. Spong, “Bilateral control of teleoperators with time delay,” IEEE Trans. on Automatic Control, vol. 34, no. 5, pp. 494–501, May 1989.
[8] N. Chopra, M. Spong, S. Hirche, and M. Buss, “Bilateral Teleoperation over Internet: the Time Varying Delay Problem,” in Proc. of the American Control Conf., Denver (CO), US, 2003, pp. 155–160.
[9] S. Hirche and M. Buss, “Packet Loss Effects in Passive Telepresence Systems,” in 43rd IEEE Conf. on Decision and Control, Paradise Island, Bahamas, 2004, pp. 4010–4015.
[10] S. Hirche, P. Hinterseer, E. Steinbach, and M. Buss , “Transparent Data Reduction in Networked Telepresence and Teleaction Systems Part I: Communication without Time Delay,” in PRESENCE: Teleoperators and Virtual Environments 16(5), 2007.
[11] I. Vittorias, H. Ben Rached, and S. Hirche, “Haptic Data Reduction in Multi-DoF Teleoperation Systems,” in Proc. of the Int. Symp. on Haptics Audio-Visual Environments and Games (HAVE), Phoenix, AZ, USA , 2010.
[12] P. Hokayem and M. Spong, “Bilateral teleoperation: An historical survey,” Automatica, vol. 49, no. 12, pp. 2035–2057, Dec. 2006.
[13] F. Brandi, J. Kammerl, and E. Steinbach. "Error-Resilient Perceptual Coding for Networked Haptic Interaction”. In ACM Multimedia 2010, Firenze, Italy, Oct. 2010.
[14] P. A. Chou and Z. Miao. Rate-distortion optimized streaming of packetized media. IEEE Trans. on Multimedia, 8(2):390–404, April 2006.
[15] E. Steinbach, S. Hirche, J. Kammerl, I. Vittorias, and R. Chaudhari, "Haptic Data Compression and Communication for Telepresence and Teleaction," IEEE Signal Processing Magazine, vol. 28, no. 1, pp. 87-96, January 2011.

Fernanda Brandi was born in Brazil in 1981. She received the Bachelor degree in Electrical Engineering (2006) and the Master degree (2009) from the Universidade de Brasilia, Brasilia, Brazil. She is currently working towards the Ph.D. degree at the Institute of Media Technology at the Technische Universität München, Munich, Germany. Her research interests include video and haptic communication with focus on optimizing the signal compression in the presence of communication unreliabilities.

Rahul Chaudhari received the M.Sc. degree in communication systems from the Technische Universität München, Munich, Germany in 2009, focusing on signal processing and compression/reduction of data for haptic communication. He received an undergraduate degree (bachelor of engineering) in electronics and telecommunications from the University of Pune, India, graduating in 2006 as the top engineering student in his class. He joined the Institute for Media Technology at the Technische Universität München in 2010, where he is currently working as a member of the research and teaching staff. His research interests are in the field of haptic communication with a focus on compression/data reduction of haptic data and objective quality evaluation of compression schemes.

Sandra Hirche received the diploma engineer degree in Mechanical Engineering and Transport Systems in 2002 from the Technical University Berlin, Germany, and the Doctor of Engineering degree in Electrical Engineering and Computer Science in 2005 from the Technische Universität München, Munich, Germany. From 2005-2007 she has been a PostDoc at the Tokyo Institute of Technology, Tokyo, Japan. Since 2008 she is associate professor heading the Associate Institute for Information-oriented Control in the Department of Electrical Engineering and Information Technology, Technische Universität München. Her research interests include networked control systems, cooperative control, human-in-the-loop control, and haptics. Since 2009 she serves as Chair for Student Activities in the IEEE Control System Society.

Julius Kammerl studied computer science at the Technische Universität München in Munich, Germany. He received the degree “Dipl.-Inf. (Univ)” in January 2005. After working at the Audio and Multimedia Group at Fraunhofer Institute for Integrated Circuits IIS in Erlangen, Germany, he joined the Institute for Media Technology at the Technische Universität München in 2006, where he is currently working as a member of the research and teaching staff. His research interests are in the field of haptic communication with a focus on perceptual coding of haptic data streams. He is a member of the interdisciplinary research cluster on high-fidelity telepresence and teleaction, which is funded by the German Research Foundation, DFG. He is a Member of the IEEE.

Eckehard Steinbach (M’96-SM’08) studied electrical engineering at the University of Karlsruhe, Karlsruhe, Germany, the University of Essex, Colchester, U.K., and ESIEE, Paris, France. He received the Engineering Doctorate from the University of Erlangen-Nuremberg, Germany, in 1999. From 1994 to 2000, he was a Member of the Research Staff of the Image Communication Group, University of Erlangen-Nuremberg. From February 2000 to December 2001, he was a Postdoctoral Fellow with the Information Systems Lab, Stanford University, Stanford, CA. In February 2002, he joined the Department of Electrical Engineering and Information Technology, Technische Universität München (TUM), Munich, Germany, as a Professor for Media Technology. Since 2009 he is heading the Institute for Media Technology at TUM. His current research interests are in the area of audio-visual-haptic information processing, image and video compression, error-resilient video communication, and networked multimedia systems.

Iason Vittorias was born in Rhodes, Greece in 1985. He received his diploma degree in Electrical & Computers Engineering in 2007 from the Aristotle University of Thessaloniki, Greece. Since 2008 he is a research assistant at the Institute of Automatic Control Engineering, Technische Universität München, Munich, Germany, pursuing his PhD degree. His research interests include teleoperation systems over networks, passivity-based control and haptic data reduction.

Gaze Awareness and Eye Contact in Multimedia Communications

Kar-Han Tan, Ramin Samadani, and John Apostolopoulos, HP Labs, Palo Alto, California, USA
{karhan.tan, ramin.samadani, john_apostolopoulos}@hp.com

1. The Human Connection

People do their best work in collaborative social networks. With advances in multimedia communications technologies it is now possible even for geographically distributed teams to collaborate productively. Modern teams work together across great distances and time zones without having to endure the inconveniences of frequent intercontinental travel. Fewer flights also mean reductions in expenses and lower environmental impact.

Early video conferencing systems suffered from technical issues like inadequate image quality and noticeable end-to-end latencies. As a result these systems often interfere with the users' natural social interactions because the experience delivered lacked *spontaneity* [5]. These issues were largely addressed in recent years by high end visual collaboration systems from HP, Cisco, Tandberg, PolyCom, and LifeSize. Using high quality cameras and displays to deliver low latency video, these systems create a realistic, immersive experience where meeting participants feel as if they are co-located in the same room, and can naturally interact with one another without having to worry about technology getting in the way.

2. Spatial Relationships

Although much progress has been made in improving remote collaboration experiences, there are still qualitative gaps between a remote meeting and a real, natural face-to-face meeting. One of the areas where the difference is significant is the sense of spatial relationships among meeting participants, an important aspect of truly immersive collaboration systems [2]. Ideally, if the sense of space were faithfully reproduced, then to address a meeting participant one would simply turn to face that person and start talking. In a real meeting, the person being spoken to would know it simply because the first participant would be facing him or her, and the two can make *eye contact* if they wish to. In today's collaboration systems, this simple but important interaction is in fact generally impossible. Except in limited 'sweet spot' configurations usually when two users are positioned in the center of their respective rooms, when user A turns to face a remote user B

(rendered on a local display), user B will not see a frontal image of user A with associated eye contact. This situation where two users are talking to each other but not appearing to be facing one another even when they are trying to make eye contact is clearly unnatural. For some users it can be disturbing - to the point where they *avoid* looking at the displays during a meeting.

3. Gaze Awareness

Another important aspect of a collaborative meeting is the ability to share documents and here reproducing the natural spatial relationship between users and documents is also a challenge. In a co-located meeting users can look at a shared document either printed in hardcopy or on a display, and it is easy to refer to specific portions of the shared document by pointing or simply by *looking*. In a remote collaboration system, often a shared document is displayed on a screen separating the users, and it is impossible to infer which part of the document a user is looking at. In other words, *gaze awareness* with respect to shared media is lacking in today's collaboration systems.

4. State of the Art

The twin problems of eye contact and gaze awareness has long been recognized as key issues standing in the way of truly natural remote collaboration experiences. Jones used a light field display in combination with a real-time 3D capture system to deliver one-to-many eye contact [8]. Gemmel used 3D graphics to modify images to improve facial expressions and gaze awareness [3]. In both cases, it is often easy for a user to see that the remote user's imagery is either not fully photorealistic or has been unnaturally manipulated, which takes away from how immersive the experience can be.

See-through displays offer a promising approach for improved eye contact and gaze awareness. Research in transparent displays [4] is driven by a number of new applications like collaboration [6, 11], gesture-based user interaction [7] and augmented reality [1]. For collaboration applications, these displays display information for each local participant and capture information to deliver to the remote participant through the same surface, as shown in Fig. 1. Aligning the system's camera with the participant display allows good eye contact and gaze awareness [11]. Collaboration systems based on see-through displays are often afflicted by *video cross-talk* which arises when the video signal to be displayed to the local user interferes with the local video signal that one desires to capture with the camera.

6. Video cross-talk reduction

A variety of techniques have been proposed for cross talk reduction. The most common approaches utilize hardware to multiplex, in different ways, the two signals. In [7] a liquid crystal switching diffuser provides temporal multiplexing of the signals by switching between diffusing display and transparent capture for gesture based interaction. In [6] a switching diffuser is discussed for collaboration applications. That paper, however, implements instead a prototype based on polarization multiplexed signals combined with a half-silvered mirror collaboration surface that needs to be at 45 degrees to vertical. A recent approach [11] uses a holographic diffusing screen together with light wavelength division multiplexing which allows a natural, vertical surface for collaboration with gaze awareness.

An alternative approach for cross-talk reduction uses software-only signal processing for cross-talk reduction [10]. Using this approach, no additional hardware is required, there is no light loss through the system, nor is synchronization between camera and projector required, since the method is able to reconstruct time-unsynchronized signals through careful photometric, geometric and optical characterization of the projector camera system.
In addition, software based approaches also apply to a broad range of collaboration applications. In [9] a software cross-talk classification method is applied to digital white board sharing.

7. Summary

New technologies and system prototypes are providing improved eye contact, gaze awareness and shared media capabilities, which in turn provide rich new collaboration experiences. See-through displays offer great support for these new experiences. Cross-talk reduction for these displays is a key technical challenge being actively addressed by recent research.

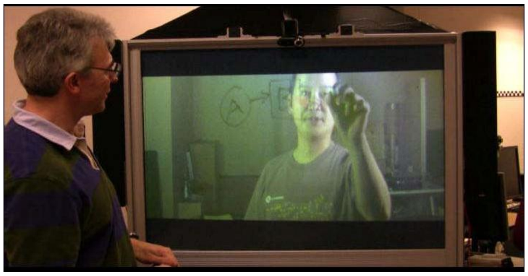


Figure 1. Two users collaborating on a ConnectBoard prototype with support for eye contact and gaze awareness.

References

[1] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier and B. MacIntyre, Recent advances in Augmented Reality, *IEEE Computer Graphics and Applications*, Nov/Dec. 2001, pp 34-47.
[2] J. Edwards. Understanding the Spectrum: Videoconferencing to Telepresence Solutions. Sponsored by Cisco. May 2010. http://www.cisco.com/en/US/prod/collateral/ps7060/idc_vc_to_tp_spectrum.pdf
[3] J. Gemmell, K. Toyama, C. L. Zitnick, T. Kang, and S. Seitz, "Gaze awareness for video-conferencing: A software approach," IEEE MultiMedia, vol. 7, pp. 26–35, 2000.
[4] P. Görrn, M. Sander, J. Meyer, M. Kröger, E. Becker, H.-H. Johannes, W. Kowalski and T. Riedl, Towards see-through displays: fully transparent thin-film transistors driving transparent organic light-emitting diodes, *Advanced Materials*, March, 2006.
[5] HP Halo Collaboration White Paper http://h20338.www2.hp.com/enterprise/downloads/Halo_Collaboration_White_Paper_3_21_06.pdf
[6] H. Ishii and M. Kobayashi, "Clearboard: a seamless medium for shared drawing and conversation with eye contact," in CHI 1992, pp. 525–532.
[7] S. Izadi, S. Hodges, S. Taylor, D. Rosenfeld, N. Villar, A. Butler, and J. Westhues, "Going beyond the display: a surface technology with an electronically switchable diffuser," UIST 2008.
[8] A. Jones, M. Lang, G. Fyffe, X. Yu, J. Busch, I. McDowall, M. Bolas, and P. Debevec, "Achieving eye contact in a one-to-many 3d video teleconferencing system," in ACM SIGGRAPH 2009.
[9] M. Liao, M. Sun, R. Yang, and Z. Zhang, "Robust and Accurate Visual Echo Cancelation in a Full-duplex Projector-camera System," *IEEE TPAMI*, vol. 30, no. 10, pp. 1831–1840, 2008.
[10] R. Samadani, J. Apostolopoulos, I. Robinson, and K.-H. Tan. Video Cross-Talk Reduction and Synchronization for Two-Way Collaboration. ICIP 2010.
[11] K.-H. Tan, I. Robinson, B. Culbertson, J. Apostolopoulos. Enabling Genuine Eye Contact and Accurate Gaze in Remote Collaboration. ICME 2010. HP Labs Technical Report HPL-2010-96

Kar-Han Tan is a Senior Research Scientist at the Mobile and Immersive Experience Lab (MIXL) at HP Labs, where he is working on 3D capture and display technologies as well as next generation remote collaboration systems. He received his

PhD in CS from the University of Illinois at Urbana-Champaign, where he was a Beckman Graduate Fellow, MS from UCLA, and B.Sc. from the National University of Singapore. Kar-Han contributes actively to the research community and has received several best paper awards. Prior to HP he was Manager of Algorithms Group at EPSON R&D, where he led the invention of View Projection, a technique that enables one-touch setup of light displays on arbitrary surfaces. He co-invented Multi-Flash Imaging at Mitsubishi Electric Research Lab (MERL), and the Virtual Structures algorithm at UCLA, widely recognized today as one of the fundamental techniques for mobile robot formation control.

Ramin Samadani is a Senior Research Scientist at HP labs Mobile & Immersive Experiences lab. His current research involves video processing algorithms to improve the presentation and quality of immersive video conferencing and collaboration systems. He serves as Associate Editor of IEEE Transactions of Image Processing and a member of the Multimedia Technical Committee. In the past he developed image processing algorithms applied to display browsing and printing of image collections, such as algorithms to reduce compression artifacts, to resize images while preserving image quality information, and to combine multimedia with continuous GPS location and time information. Prior to HP, at Electronics for Imaging (1994-2000), he worked on color technologies for high quality printers, in engineering and management positions, ending as Director of Imaging Technologies. At Stanford and NASA Ames (1987-1994), he worked on feature extraction and motion analysis algorithms applied to remote sensing applications, as well as on algorithms for color simulation of flat panel displays. He received the MS and PhD in EECS from Stanford University and a BS in Engineering Physics from UC Berkeley.

John Apostolopoulos is a Distinguished Technologist and the Director of the Mobile and Immersive Experience Lab (MIXL) at HP Labs. His research interests include immersive communication, and improving the reliability, fidelity, scalability and security of multimedia communications over wired and wireless packet networks. Apostolopoulos received his B.S., M.S., and Ph.D. degrees from MIT. In graduate school, he worked on the U.S. Digital TV standard and received an Emmy Award Certificate for his contributions. Apostolopoulos was named “one of the world’s top 100 young (under 35) innovators in science and technology” (TR100) by MIT Technology Review in 2003, and has received a number of best paper awards, and is an IEEE Fellow. He also teaches and conducts joint research at Stanford University, where he is a Consulting Associate Professor of Electrical Engineering, and he is a frequent visiting lecturer at MIT.

Multimedia Technology for Next Generation Online Lecture Video

Ngai-Man Cheung, Sherif Halawa, Derek Pang, Bernd Girod, Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA
{ncheung, halawa, dcypang, bgirod}@stanford.edu

1. Introduction

Recent years have seen a dramatic growth in online education. According to a recent Sloan Online Learning Survey, the number of US online students has increased by almost one million in 2009 [1]. This corresponds to a growth rate of 21%, which far exceeds the 2% growth in the overall population of higher education. Online education allows students to study materials at their own convenient time and location. Moreover, free online programs, such as MIT OpenCourseWare [2] or Stanford Engineering Everywhere [3], provide people from around the world with the opportunity to access high-quality education.

Central to online education is lecture video capturing and viewing. Nowadays, many lectures are being recorded and made available to students through the Internet. Quality of these lecture videos plays a crucial role in the overall online education experience. In spite of its importance, lecture capture remains costly and rather inefficient. Often, expensive camcorders and special equipments are needed to be installed for lecture recording. Dedicated staff is required for camera operation and post-processing. Videos are often being delivered in rather low resolutions with inflexible formats. Students have no control over the view regions, and there is no adaptation to individual need and pace. These hinder widespread deployment of online education and undermine its effectiveness.

In this article, we briefly discuss how advances in multimedia processing and computer vision could enable next generation online lecture videos, which are of lower cost, more engaging and effective. We also report a recent project employing advanced online lecture technology.

2. Multimedia technology to enable low-cost capturing and delivery

According to [4], a key to enable low-cost online lecture is to reduce the recurring labor cost in video capturing and processing. Thus, unmanned camcorders have been of research interest, e.g., [5, 6]. For instance, Microsoft *iCam2* [6] uses speaker tracking to control a pan/tilt/zoom (PTZ) camera for automated capturing. A digital/mechanical hybrid tracking scheme is developed to achieve smooth region following and wide area covering. More recent work, however, takes a different approach for automated capturing [7, 8, 13]. Thanks to advanced video coding algorithms such as H.264/AVC [9], it becomes possible to capture high-quality, high-resolution lectures into manageable sizes (e.g., a one-hour HD lecture takes only around 7GB). Therefore, recent work proposes to employ an off-the-shelf AVCHD camera mounted statically on a tripod to capture the entire field-of-view as well as fine details (text, figures, instructor's face, etc). To facilitate streaming to remote clients and viewing on non-HD displays or mobile terminals, [8] proposes to transcode the captured videos into multiple tiles and resolution layers, so that user-selected region-of-interests (RoI) can be readily extracted and streamed. Note that such coding algorithm can also improve viewing experience, as will be discussed. Alternatively, H.264/AVC Scalable Video Coding (SVC) extension [11] or automatic cropping [7, 8] can be employed to facilitate lecture delivery and display.

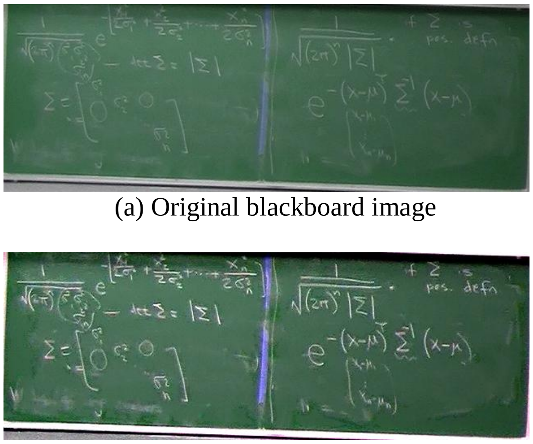
Unmanned video capture using off-the-shelf equipments may occasionally suffer from degradation. Automatic, low-cost restorations are important to deal with these issues, especially for institutions with cost-constraints. For example, [12] proposes an out-of-focus video restoration algorithm with reasonable computational complexity by leveraging slide images as side information in the process.

3. Multimedia technology to improve viewing experience

Often, different students would want to focus on different regions of a lecture scene. For instance, one may want to watch the instructor, while another may choose to focus on the blackboard content. To fulfill individual RoI needs, Stanford *ClassX* [13] uses spatial-random-access-enabled video compression [8, 10] along with an advanced client-server protocol to enable interactive pan/tilt/zoom lecture viewing. In particular, *ClassX* creates multiple resolution layers from the captured video. Each layer is subdivided into tiles, and tiles are compressed independently using H.264/AVC. Upon receiving a RoI request, the server determines the relevant tiles and streams them to the client for rendering. The result is a system providing student-centric viewing experience, where students can watch their own RoI according to their needs.

Viewing experience can also be improved by blackboard enhancement. Sometimes, the handwritings on blackboards can be difficult to recognize in videos due to background dirt, low contrast or poor illumination. Image enhancement can be applied to make blackboard contents easier to read (Figure 1). Blackboards can be automatically located in videos with edge detection and Hough transform followed by quadrangle detection [13, 14]. Handwritings and backgrounds can be separated and processed differently, using adaptive thresholding techniques such as Otsu algorithm on local windows [13, 15].

To better capture audience questions, microphone arrays and sound source localization algorithms can be employed [6].



(a) Original blackboard image

(b) Enhanced blackboard image

Figure 1. Blackboard enhancement result using adaptive contrast enhancement [13].

4. Multimedia technology to improve effectiveness of online lecture

There are various techniques to improve the effectiveness of online lectures. For example, electronic slides can be displayed alongside with the relevant sections of the videos for easy reference. Several approaches have been proposed for automatic synchronization between slide images and videos, including those based on text recognition [16] and local feature matching [17, 8]. To speed up processing of the entire lecture, low-complexity slide change detection can be applied to the video to identify the key-frames for slide synchronization [8].

In addition, blackboard images can be extracted from the videos for displaying alongside or for downloading as handouts. Quadrangle detection techniques previously discussed can be used to locate the boards. Geometric rectification and image enhancement can be applied to improve the readability of the board images [14].

Sometimes students may want to review only part of the lecture about a particular concept. It is advantageous to summarize and index the lecture archives to facilitate searching by the users. For example, videos can be segmented according to the content, and visualization tools such as concept maps [18] can be used to represent the relationships between video segments. Slide recognition [8] can be leveraged to facilitate this summarization process. In particular, textual information available in the synchronized slides can be used to annotate the videos and infer the relationships between video segments.

Effectiveness of online lectures can also be improved by analyzing the usage information such as viewer trajectories or video playback statistics. These analyses provide useful feedbacks to students and instructors. For example, if a certain section of the video is being watched multiple times by some students, this could indicate that the discussed concept could be complex. Besides, multimedia social network technology such as Internet forums, blogs, wikis or presentation sharing can facilitate and encourage collaboration among students, leading to improvement in learning effectiveness.

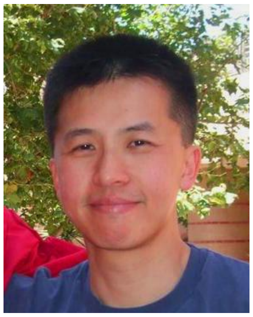
5. Example: Stanford *ClassX*

ClassX is an interactive online lecture capturing and viewing system developed at Stanford University [8, 13]. Unlike existing solutions that restrict the user to watch only a pre-defined view, *ClassX* allows interactive pan/tilt/zoom while watching the lectures. In addition, *ClassX* supports automatic tracking of instructors and automatic synchronization of slide images with videos. The system has been used in dozens of Stanford courses and colloquia during a pilot deployment period of near one year. Future plan is to release *ClassX* as open source software to promote research in next generation online lecture video that offers a cheaper and a more effective solution to both content distributors and viewers.

References

[1] I. E. Allen and J. Seaman, *Class Difference: Online*

Education in the United States 2010. Sloan Consortium, Nov. 2010.
[2] http://ocw.mit.edu/
[3] http://see.stanford.edu/
[4] L. A. Rowe, D. Harley, P. Pletcher, S. Lawrence, “BIBS: A Lecture Webcasting System,” Tech. rep. Berkeley Multimedia Research Center, U.C. Berkeley, 2001.
[5] Y. Rui, L. He, A. Gupta and Q. Liu, “Building an Intelligent Camera Management System,” in *Proc. ACM Multimedia 2001*, Sept. 2001.
[6] C. Zhang, Y. Rui, J. Crawford, and L.-W. He, “An Automated End-to-End Lecture Capturing and Broadcasting System,” *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, Jan. 2008.
[7] T. Nagai, “Automated Lecture Recording System with AVCHD Camcorder and Microserver,” in *Proc. ACM SIGUCCS fall conference on user services conference (SIGUCCS’09)*, St. Louis, Missouri, USA, 2009, pp. 47–54.
[8] A. Mavlankar, P. Agrawal, D. Pang, S. Halawa, N.-M. Cheung and B. Girod, “An Interactive Region-of-Interest Video Streaming System for Online Lecture Viewing,” in *Proc. International Packet Video Workshop (PV2010)*, Hong Kong, China, December 2010.
[9] *Advanced Video Coding for Generic Audiovisual Services*, ITU-T Rec. H.264 and ISO/IEC 14496-10 (MPEG-4 AVC), ITU-T and ISO/IEC JTC 1, May 2003.
[10] A. Mavlankar and B. Girod, “Video Streaming with Interactive Pan/Tilt/Zoom,” in M. Mrak, M. Grgic and M. Kunt (eds.), *High-Quality Visual Experience: Creation, Processing and Interactivity of High-Resolution and High-Dimensional Video Signals*, Springer. In print.
[11] H. Schwarz, D. Marpe, T. Wiegand, “Overview of the Scalable Video Coding Extension of the H.264/AVC Standard,” *IEEE Trans. Circuits and Systems for Video Technology*, vol. 17, no. 9, Sept. 2007.
[12] N.-M. Cheung, D. M. Chen, V. R. Chandrasekhar, S. Tsai, G. Takacs, S. Halawa, B. Girod, “Restoration of Out-of-focus Lecture Video by Automatic Slide Matching,” in *Proc. ACM Multimedia 2010*, Florence, Italy, October 2010.
[13] http://classx.stanford.edu/
[14] Z. Zhang and L.-W. He, “Notetaking with a Camera: Whiteboard Scanning and Image Enhancement,” in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2004)*, May 2004.
[15] M. Merler, J. R. Kender, “Semantic Keyword Extraction via Adaptive Text Binarization of Unstructured Unsourced Video,” in *Proc. IEEE International Conference on Image Processing (ICIP 2009)*, Nov. 2009.
[16] F. Wang, C. W. Ngo, and T. C. Pong, “Synchronization of Lecture Videos and Electronic Slides by Video Text Analysis,” in *Proc. ACM Multimedia 2003*, Nov. 2003.
[17] Q. Fan, K. Barnard, A. Amir, A. Efrat, and M. Lin, “Matching Slides to Presentation Videos Using SIFT and Scene Background Matching,” in *Proc. ACM International Workshop on Multimedia Information Retrieval (MIR 2006)*, New York, NY, USA, 2006.
[18] J. D. Novak, A. J. Canas, “The Theory Underlying Concept Maps and How to Construct and Use Them,” Tech. rep. Florida Institute for Human and Machine Cognition, 2006.



Ngai-Man Cheung received the Ph.D. degree from University of Southern California (USC), Los Angeles, in 2008. He is currently a postdoctoral researcher with the Information Systems Laboratory, Stanford University, Stanford, California. His research interests are multimedia signal processing and compression. He received paper awards from EURASIP Journal of Advances in Signal Processing, IEEE International Workshop on Multimedia Signal Processing (MMSP) 2007, IS&T/SPIE VCIP 2008, and from USC Department of Electrical Engineering in 2008.

Sherif Halawa received the B.Sc. (with First Class Honors) and M.Sc. degrees from the School of Engineering, Alexandria University, Egypt in 2005 and 2008 respectively. He is currently pursuing his PhD degree in Electrical Engineering at Stanford University. In 2009, he interned at Microsoft Research, Egypt. His major research interests are in the areas of multimedia encoding and encryption.

Derek Pang received the B.A.Sc. degree (with First Class Honors) from the School of Engineering Science, Simon Fraser University, Burnaby, BC, Canada, in 2009. He is currently pursuing the M.S. degree in Electrical Engineering at Stanford University, Stanford, CA and is a recipient of NSERC Postgraduate Scholarship. In 2007, he was an intern at NTT Media Information Laboratory. His research interests are in the areas of computer vision, novel interactive video applications and low- complexity video coding.

Bernd Girod is Professor of Electrical Engineering and (by courtesy) Computer Science in the Information Systems Laboratory of Stanford University, California, since 1999. Previously, he was a Professor in the Electrical Engineering Department of the University of Erlangen-Nuremberg. His current research interests are in the areas of video compression, networked media systems, and image-based retrieval. He has published over 450 conference and journal papers, as well as 5 books, receiving the EURASIP Signal Processing Best Paper Award in 2002, the IEEE Multimedia Communication Best Paper Award in 2007, the EURASIP Image Communication Best Paper Award in 2008, as well as the EURASIP Technical Achievement Award in 2004. As an entrepreneur, Professor Girod has been involved with several startup ventures, among them Polycom, Vivo Software, 8x8, and RealNetworks. He received an Engineering Doctorate from University of Hannover, Germany, and an M.S. Degree from Georgia Institute of Technology. Prof. Girod is a Fellow of the IEEE, a EURASIP Fellow, and a member of the German National Academy of Sciences (Leopoldina).

The Context Does Matter: Beyond the Data Pipes of Today

Jacob Chakareski, EPFL, Lausanne, Switzerland
www.jakov.org

1. Introduction

We are witnessing an unparalleled integration of computer-communication technologies. People, devices, and computers interact in ways unforeseen before. At the heart of these technological advances lies our drive to interconnect and immerse into the environment. In particular, social networking has become a centerpiece of our online activities. From content sharing to live communication, we tend to carry out all these activities through community sites that we frequent on a daily basis. There is a growing understanding that engineering systems can benefit from the deluge of contextual information that is generated by our participation in online social networks. This article will highlight the major application areas of context-driven computing and communications.

2. From User-Centric to Context-Driven

Traditionally, information systems have been designed in a modular fashion. The seven layer Open Systems Interconnect (OSI) reference model for communication networks is a prime example of this design strategy [1]. Analogously, many computer-communications technologies employed today consider the actions of individual users as semantically independent of each other. This modular context-agnostic approach to systems design and operation has multiple advantages including lower complexity, easier inter-operability, and lower cost. However, at the same time it precludes exploiting prospective interdependencies across the different system layers and users that in turn can lead to inefficient performance and customer dissatisfaction.

In the rest of the paper, we will overview several application areas where context-aware operation can provide substantial benefits, in our opinion.

3. Networking

As shown recently, data networks can considerably boost their performance over multiple criteria by taking into account social data. In particular, efficient flow allocation, packet scheduling, and directory (look-up) services can be enabled by exploiting the social graph and the content preferences of the online community [2]. Similarly, in [3] the authors have shown that routing in mobile ad-hoc networks can be provided, at lower cost, by taking advantage of the reported social fabric between the participants. Finally, caching can also benefit from social network data, as shown in the study on short-video sharing in the context of peer-to-peer networks [4]. The context-aware techniques outlined here will become even more important in the future, due to the expected wave of network traffic caused by live media-rich communication in online communities [5].

For instructional reasons, we will describe further the framework of [2] in the remainder of the section. In particular, each member of the online community is characterized in [2] by a set of social contacts and a set of preferences for the content items shared in the social network. The underlying transport network serving the online community is characterized by uplink and downlink capacities affiliated with the nodes of the social graph, as well as by transmission costs of data communication between two social contacts in the graph. Via linear programming the authors optimize the information flow of the transport network, per unit cost of data transmission. In particular, by placing network resources on more popular content and less costly data links, substantial gains are achieved over network allocation serving the social graph that disregards such information in its operation. Note that the latter one is typical for present overlay networks created by social contacts for the purpose of content sharing.

In addition, the framework in [2] is employed to design a network directory service for efficient content discovery in peer-to-peer systems based on social network information. Specifically, each node in the overlay registers its content preferences with a management server that keeps track of the peer population. The tracker also registers the cost of data communication between nodes in the overlay using an ISP portal. Such portals are increasingly becoming common nowadays as ISPs seek to constructively address the issue of peer-to-peer traffic engineering. A low-complexity optimization algorithm is then designed that selects an appropriate set of prospective neighbour nodes in the overlay for an incoming peer. The set will maximize the likelihood of discovering a desired content item in the overlay for the incoming peer, per unit cost of data transmission. The optimization achieves that by correlating the content preferences of the new peer with those of the nodes already present in the overlay, while taking into account the underlying cost of data communication. Analogously, the algorithm detects (looks-up) the desired content item in the overlay in the shortest possible time. Significant gains in terms of likelihood and speed of content discovery, as well as savings in transmission cost are reported in [2], relative to conventional network directory services.

Finally, context-driven packet scheduling is also studied within the framework of [2]. In particular, the authors design an efficient algorithm for scheduling the exchange of data packets between social contacts interested in a content item. The context-aware scheduling can operate coherently with the previous optimization techniques of [2] described above, as each addresses a different data communication aspect in an online community. The algorithm takes into account the importance of each data unit, its availability in a neighbourhood of contacts, their dynamically varying networks resources., and any history of prior data transmissions. Thereby, the scheduling maximizes the utility of data exchange in a social graph, for the given available network resources. Substantial gains in terms of video quality of the delivered content are reported in [2], over context-agnostic packet scheduling schemes

4. Cyber Security

There is a range of applications in the field of cyber security where context-driven operation can make a difference. For example, colluding of copyrighted content replicas by malicious users in P2P networks can be effectively prevented by taking into account their social structure [6]. In addition, building trust relationships and avoiding free-riding in P2P systems can be achieved by inviting social contacts to join the same P2P network, as studied for example in [7,8]. Similarly, the authors in [9,10] have shown that user cooperation can be enhanced by employing social contacts as helpers in a P2P environment. Finally, network anomaly detection can profit by knowing the social structure between examined (existing) nodes in the network and nodes that need to be investigated for malicious activities via their traffic patterns.

5. Content Analysis

A number of diverse applications of content analysis and understanding can benefit from the contextual information that a social network provides. In particular, by taking into account the social engagement and interaction under which the data was created or consumed, the semantic gap in understanding exhibited by existing media analytics solutions can be substantially reduced. For example, the most popular content segments can be identified by studying the behavioral and interactivity pattern of the audience, as shown in [11]. Similarly, computing the aggregate behaviour of an online community over its individual data feeds can enable real-time event detection, as demonstrated by a recent study of Twitter data [12]. Finally, recommendation systems, targeted advertising [13,14], and viral marketing, are yet further application scenarios where semantic accuracy gains can be achieved via context-aware operation. Such gains will necessarily lead to higher customer satisfaction and profit margins.

6. Information Retrieval

Information retrieval and search are two other areas that can be greatly facilitated by employing contextual information. The key to such performance improvement techniques, as studied, e.g., in [2,13,14], is to take advantage of the correlation between the content preferences of social contacts, a phenomenon known as homophily [15]. Interestingly, recent studies [16,17] have shown that we tend to search for information via our online community contacts. For example, Twitter's search engine receives around 600 million queries per day [18]. The rationale behind this phenomenon is that we generally trust more the information obtained via our social network than that produced by online search engines.

7. Conclusions

The road to the context-aware future is not easy. The integration of the social layer into the operation of the various applications reviewed in this article requires a thorough revision of present system architectures. Simultaneously, the supporting algorithms and protocols that such applications will employ will necessarily become more complex, which is yet another disadvantage. Finally, a concerted effort on the part of industry with respect to standardization will be required if such solutions should receive a wide adoption.

It goes without saying though that every major shift in design principles and operation strategies in the history of modern information systems required time and perseverance. Context-driven applications have the potential to profoundly change the ways in which we collaborate and interact. At the same time, they can provide novel perspectives of ourselves as a community in our increasingly digital world. Therefore, we should give them all the required attention on their prospective road to success.

References

[1] A. S. Tanenbaum, Computer networks, 4th ed. Prentice Hall, Aug. 2002.

[2] J. Chakareski and P. Frossard, “Context-adaptive information flow allocation and media delivery in online social networks,” IEEE J. Selected Topics in Signal Processing, vol. 4, no. 4, pp. 732–745, Aug. 2010.

[3] G. Bigwood, D. Rehunathan, M. Bateman, T. Henderson, and S. Bhatti, “Exploiting self-reported social networks for routing in ubiquitous computing environments,” in Proc. Int’l Conf. Wireless and Mobile Computing, Networking and Communications. Avignon, France: IEEE, Oct. 2008, pp. 484–489.

[4] X. Cheng and J. Liu, “Nettube: Exploring social networks for peer-to-peer short video sharing,” in Proc. Conf. on Computer Communications (INFOCOM). Rio de Janeiro, Brazil: IEEE, Apr. 2009, pp. 1152–1160.

[5] “Approaching the zettabyte era,” in Cisco Visual Networking Index. Cisco Inc., Jun. 2008.

[6] H. Zhao, W. Lin, and K. Liu, “Behavior modeling and forensics for multimedia social networks: A case study in multimedia fingerprinting,” IEEE Signal Processing Magazine, vol. 26, no. 1, pp. 118–139, Jan. 2009.

[7] J. A. Pouwelse, P. Garbacki, J. Wang, A. Bakker, J. Yang, A. Iosup, D. H. J. Epema, M. Reinders, M. van Steen, and H. J. Sips, “Tribler: A social-based peer-to-peer system,” in Proc. 5th Int’l Workshop on Peer-to-Peer Systems, Santa Barbara, CA, USA, Feb. 2006.

[8] J. Wan, L. Lu, X. Xu, and X. Ren, “A peer-to-peer assisting scheme for live streaming services,” in Advances in Grid and Pervasive Computing, ser. Lecture Notes in Computer Science. Springer-Verlag Berlin / Heidelberg, 2008, vol. 5036, ch. 34, pp. 343–351.

[9] W. Wang, L. Zhao, and R. Yuan, “Improving cooperation in peer-to-peer systems using social networks,” in Proc. 3rd Workshop on Hot Topics in Peer-to-Peer Systems, Rhodes Island, Greece, Apr. 2006, pp. 50–57.

[10] J. Altmann and Z. B. Bedane, “A P2P file sharing network topology formation algorithm based on social network information,” in Proc. IEEE Int’l Workshop Network Science for Comm. Networks. Rio de Janeiro, Brazil, Apr. 2009.

[11] S. J. Davis, I. S. Burnett, and C. H. Ritz, “Using social networking and collections to enable videosemantics acquisition,” IEEE Multimedia, vol. 16, no. 4, pp. 52–61, Oct.-Dec. 2009.

[12] T. Sakaki, M. Okazaki, and Y.Matsuo, “Earthquake shakes twitter users: real-time event detection by social sensors,” in Proc. Int’l Conf. World Wide Web. Raleigh, NC, USA: ACM, Apr. 2010, pp. 851–860.

[13] A. Bagherjeiran and R. Parekh, “Combining behavioral and social network data for online advertising,” in Proc. Int’l Conf. Data Mining Workshops. Washington, D.C., USA: IEEE, Dec. 2008, pp. 837–846.

[14] P. Mitra and K. Baid, “Targeted advertising for online social networks,” in Proc. Int’l Conf. Networked Digital Technologies. Ostrava, The Czech Republic: IEEE, Jul. 2009, pp. 366–372.

[15] H. W. Lauw, J. C. Shafer, R. Agrawal, and A. Ntoulas, “Homophily in the digital world: A LiveJournal case study,” IEEE Internet Computing, vol. 14, no. 2, pp. 15–23, Mar.-Apr. 2010.

[16] W. Gao, Y. Tian, T. Huang, and Q. Yang, “Vlogging: A survey of videoblogging technology on the web,” ACM Computing Survey, vol. 42, no. 4, Jun. 2010.

[17] J. Gibs, “Social media: The next great gateway for content discovery?” in NielsenWire Blog (http://blog.nielsen.com), Oct. 2009.

[18] “Twitter Statistics,” released at Chirp: The official Twitter developer conference, April 2010.

Jacob Chakareski is an Associate Scientist at EPFL, where he conducts research, lectures, and supervises students. His present investigations include online social networks, multi-camera networks, and cognitive quality of experience in wireless multi-hop networks. He was granted the Ambizione fellowship for 2009-2012 from the Swiss NSF that recognizes research excellence among foreign nationals working at Swiss universities. Dr. Chakareski has held research positions with Microsoft and Hewlett-Packard. He has authored one monograph, three book chapters, and over 90 international publications, and has nine pending or approved patent applications. He actively participates in technical and organizing committees of several IEEE conferences and symposia on a yearly basis. He was a publicity chair for the Packet Video Workshop in 2007 and 2009 and for the Workshop on Emerging Technologies in Multimedia Communications and Networking at ICME 2009. He organized and chaired a special session on telemedicine at MMSP 2009. Chakareski won the best student paper award at the IS&T/SPIE VCIP 2004 conference. For further information, please visit www.jakov.org.

Computational Audio-Visual Scene Analysis[i]

Hao Tang, Hewlett-Packard Laboratories, Palo Alto, CA
Zicheng Liu, Microsoft Research, Redmond, WA
hao.tang@hp.com, zliu@microsoft.com

1. Overview

Human's ability to comprehend a complex scene instantaneously and effortlessly is remarkable. To date, how to program a computer to mimic this mysterious biological process is far from certain. Along with the dramatic increase of the computational power, memory, and bandwidth of modern computers, we can expect to perform certain simplified scene analysis tasks using automatic algorithms. The field of computational audio-visual scene analysis (CAVSA) concerns the use of a computer to automatically detect and locate objects (esp. people) in a scene, track their 3D movements, identify and isolate sounds in a complex mixture, and associate the individual sounds with their respective sources, based on audio and visual cues captured by camera and microphone sensors. Undoubtedly, this is an extremely challenging problem which has attracted continuous research efforts from the multimedia and computer vision communities. CAVSA has many potential applications in various areas such as smart surveillance, humanoid robotics, tele-presence, tele-immersion, and intelligent human-computer interaction, just to name a few.

At Microsoft Research, we developed a real-time CAVSA system using off-the-shelf web cameras and microphone sensors. This system is capable of answering the following questions regarding a room scene: 1. How many people are there in the room? 2. Where are their locations? 3. What are their look directions? 4. Who among them is speaking? Our design principles were rooted in the needs of practical applications: 1. **Real time**. The system must be able to generate scene analysis results at a fast rate of at least 24 frames per second. 2. **Easy configuration and setup**. The system must be easily configurable. The number and placement of sensors must be flexible. Sensor calibration must be sufficiently convenient. No special hardware is required. 3. **Good scalability**. The system must support the easy incorporation of an unlimited number of additional sensors (provided that there are sufficient computational power, memory, and bandwidth) to facilitate the analysis of larger scenes and to enhance the scene analysis results.

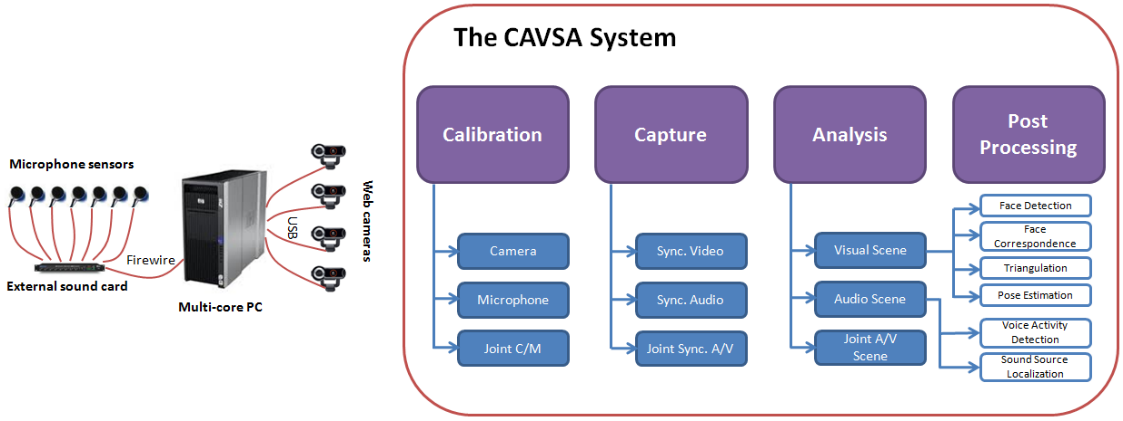


Figure 1: A schematic overview diagram of our CAVSA system.

Figure 1 schematically illustrates the hardware components and software architecture of our CAVSA system. The hardware components include a multi-core PC, 4 USB web cameras, 7 microphone sensors, and a firewire external sound card. The software architecture is composed of four essential parts, namely calibration, capture, analysis, and post processing. The system was built under Windows XP using Microsoft Visual Studio 2008, and is highly multithreaded. The system's work flow is as follows: The visual cues are multiple images of the scene, synchronously captured by the cameras from different angles. At each time instant, a multiview face detector [1] is first applied to identify the locations of all faces in all images. The detected faces are then corresponded across multiple calibrated cameras. After that, the 3D locations and orientations of the faces are recovered. The audio cues are multiple audio signals captured synchronously by the microphone sensors at a sample rate of 44.1 kHz. The time delays of arrival (TDOA) for the different signals are exploited to find the sound source location. Finally, the analysis results from the audio and visual modalities are fused to determine the speaker in the scene.

2. Hardware

One of our design principles demands that no special hardware devices should be required. Our hardware system consists of an 8-core PC, a MOTU 828pre firewire external sound card [2], 4 Logitech Quickcam Pro 9000 USB web cameras, and 7 tiny microphone sensors. These are all off-the-shelf devices that may be directly acquired from the market. The cameras are mounted at the four top corners of the room, and connected to the PC through USB cables. The microphone sensors are attached to the four walls, and connected to the PC via the firewire external sound card. Our design assures that the number of cameras and microphone sensors can be configured, and the placement of them is flexible, although certain optimized placement of the sensors may facilitate the analysis task better [3].

3. Software

In our software architecture, the calibration part is responsible for jointly calibrating the cameras and microphone sensors, which includes the determination of both intrinsic and extrinsic parameters of all cameras and the determination of the 3D locations of all microphone sensors. This is achieved through a combination of the direct linear transformation (DLT), nonlinear least squares (NLS), multidimensional scaling (MDS) and Procrustes analysis (PA) techniques [4-6]. The capture part is responsible for capturing multichannel audio and visual signals synchronously. The synchronization between the audio channels is based on the high-precision built-in hardware clock of the external sound card, and the synchronization between the visual channels as well as the synchronization between the audio and visual signals are based on the less accurate multimedia clock of the PC. The analysis part, which is the core component of the software system, is responsible for analyzing the audio and visual signals to infer the answers to the four aforementioned questions regarding the scene. There are four major functional modules operating on the visual scene, including a multithreaded face detector that concurrently locates all faces in all camera images, a multi-camera face corresponder that matches the same faces across different camera images, a stereo triangulator that recovers the 3D locations of the faces, and a pose estimator that determines the 3D poses of the faces. Likewise, there are two functional modules operating on the audio scene, namely a voice activity detector (VAD) [7] that separates human speech from ambient noise, and a sound source localizer (SSL) that finds the 3D location of the speaker using the steered response power (SRP) and peak picking methods [8]. The results from the audio and visual modalities are then combined to accurately determine the speaker. Finally, the post processing part is responsible for completing some extra work such as maintaining the person identities across multiple sessions.

4. Conclusion

We developed a real-time CAVSA system based on off-the-shelf web cameras and microphone sensors, which is used to analyze a room scene. The analysis results may be exploited to aid smart surveillance and tele-presence applications. The field of CAVSA is yet underway. We believe that our system can advance the state of the art and have a potential business impact.

References

[1] C. Zhang and Z. Zhang, Boosting-Based Face Detection and Adaptation, Synthesis Lectures on Computer Vision, Morgan and Claypool, 2010.

[2] http://www.motu.com.

[3] D. Rabinkin, R. Renomeron, J. French and J. Flanagan, "Optimum Microphone Placement for Array Sound Capture," In Advanced Signal Processing: Algorithms, Architectures, and Implementations VII, Franklin T. Luk, Editor, Proceedings of SPIE, Vol. 3162, pp. 227-239, 1997.

[4] Camera Calibration Toolbox for Matlab, http://www.vision.caltech.edu/bouguetj/calib_doc/.

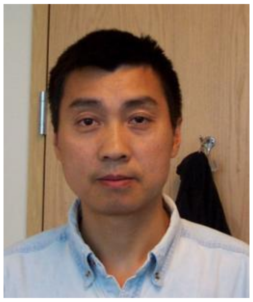
[5] Trevor F. Cox, Michael A. A. Cox, Trevor F. Cox, Multidimensional Scaling, 2nd Ed., Chapman and Hall/CRC, 2000.

[6] Gower, John C. and Dijksterhuis, Garmt B.: Procrustes Problems, Oxford University Press, 2004.

[7] S. Gökhun Tanyer and Hamza Özer, "Voice Activity Detection in Nonstationary Noise," IEEE Transactions on Speech and Audio Processing, Vol. 8, No. 4, July 2000.

[8] H. Do, H. F. Silverman, and Y. Yu, "A real-time SRP-PHAT source location implementation using stochastic region contraction (SRC) on a large-aperture microphone array," IEEE International Conference on Acoustics, Speech, and Signal Processing, Honolulu, HI, 2007, vol.1, pp.121 – 124.

Hao Tang is a researcher at Hewlett-Packard Laboratories, Palo Alto, CA. He received the Ph.D. degree in electrical engineering from the University of Illinois at Urbana-Champaign in 2010, the M.S. degree in electrical engineering from Rutgers University in 2005, the M.E. and B.E. degrees, both in electrical engineering, from the University of Science and Technology of China, in 2003 and 1998, respectively. His broad research interests include statistical pattern recognition, machine learning, computer vision, speech and multimedia signal processing. From 1998 to 2003, he was on the faculty in the Department of Electronic Engineering and Information Science at the University of Science and Technology of China, Hefei, Anhui, China, when he also served as a senor software engineer, research director, and research manager at Anhui USTC iFlyTEK Co., Ltd. He received the Scientific and Technological Progress Award, 1st Prize, from the Government of Anhui Province, China, in 2000, and the National Scientific and Technological Progress Award, 2nd Prize, from the The State Council of China in 2003. He was the recipient of the Microsoft Research Asia "Star of Tomorrow" Award in 2007, Best Student Paper Award at the International Conference on Pattern Recognition in 2008, and IBM T.J. Watson "Emerging Leader in Multimedia" Award in 2009.



Zicheng Liu is a senior researcher at Microsoft Research, Redmond. He has worked on a variety of topics including combinatorial optimization, linked figure animation, and microphone array signal processing. His current research interests include activity recognition, face modeling and animation, and multimedia collaboration. He received a Ph.D. in Computer Science from Princeton University, a M.S. in Operational Research from the Institute of Applied Mathematics, Chinese Academy of Science, and a B.S. in Mathematics from Huazhong Normal University, China. Before joining Microsoft Research, he worked at Silicon Graphics as a member of technical staff for two years where he developed a trimmed NURBS tessellator which was shipped in both OpenGL and OpenGL-Optimizer products. He has published over 70 papers in peer-reviewed international journals and conferences, and holds over 40 granted patents. He has served in the technical committees for many international conferences. He was the co-chair of the 2003 ICCV Workshop on Multimedia Technologies in E-Learning and Collaboration, the technical co-chair of 2006 IEEE International Workshop on Multimedia Signal Processing, and the technical co-chair of 2010 International Conference on Multimedia and Expo. He is an associate editor of Machine Vision and Applications journal, and a senior member of IEEE.

[1] This work was done during the internship of Hao Tang at Microsoft Research, Redmond, WA.

Autonomous Infrastructures for Networked Interactive and Personalized Media Experience

C. De Vleeschouwer, Université catholique de Louvain, ICTEAM
christophe.devleeschouwer@uclouvain.be

1. Introduction

Today's media consumption evolves towards increased user-centric adaptation of contents, to meet the requirements of users having different expectations in terms of story-telling, and heterogeneous constraints in terms of access devices. Individuals want to access contents through a personalized service that is able to provide what they are interested in, at the time when they want it, and through the distribution channel of their choice. In this letter, we explain how it is possible to address this challenge by merging computer vision tools and networking technologies to automate the collection or adaptation of contents, so as to personalize their distribution through interactive services.

From the network perspective, our approach builds on an interactive streaming architecture that supports both user feedback interpretation, and temporal juxtaposition of multiple video bitstreams in a single streaming session. An instance of this architecture has been implemented by extending the liveMedia streaming library and using the H.264/AVC standard [1]. In this framework, the initial video content is split into segments that are encoded independently and potentially with distinct parameters. The server can then decide on the fly which segment and which version of it to send as a function of how it matches the preferences expressed by the user or the network constraints.

On the client side, two scenarios are envisioned to collect user preferences. The first scenario implements a set of dedicated RTSP commands that are exploited by the client to control on-the-fly the switching between the multiple versions of the pre-encoded video segments. In contrast, the second scenario queries the client off-line, and builds the summary to forward based on the preferences expressed by the user, either in terms of scene content or narrative style.

In Section 2, we explain how the first scenario can be exploited to facilitate the access to high-resolution video content through individual and bandwidth-constrained connections, as typically encountered on mobile networks. Section 3 considers the second scenario, and defines how to build personalized summaries as a resource allocation problem. Section 4 concludes.

2. Interactive video streaming

This first scenario fully exploits the interaction capabilities offered by the network infrastructure. As depicted in Fig.1, dedicated automatic video analysis tools are developed to split some initial content into non-overlapping segments, and to generate multiple cropped (for zoom-in) or sub-sampled (spatially or temporally) versions for each segment. Each version of each segment is encoded. The user then gets the opportunity to select interactively a preferred version among the multiple streams that are offered to render the scene at hand [2]. (S)he gets the opportunity to zoom in the video, to move forward or backward across time, or to request a replay of some actions. To demonstrate our system, automatic methods have been designed and implemented for segmenting and versioning the input video content in a semantically meaningful way, both in surveillance and soccer game contexts [3].

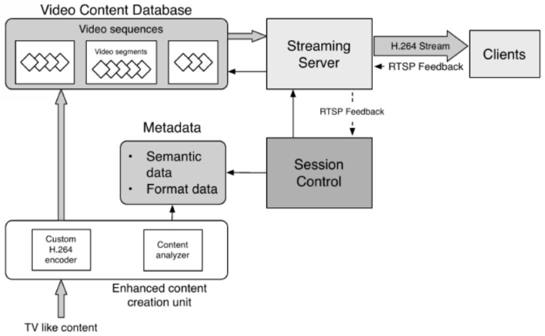


Figure 1. The content enhancement unit creates multiple versions of non-overlapping content segments, to be streamed through the interactive architecture.

3. Automatic and personalized summarization

The second scenario does not consider on-the-fly interactive feedback from the user. Instead, it collects the preferences of the user beforehand, to edit a personalized video summary, based on the concatenation of pre-encoded video clips. Our proposed system, depicted in Fig.2, adopts a divide and conquer paradigm [4]. Specifically, the video is split into semantically meaningful segments, i.e. into segments which are coherent with the actions of the game.

A generic resource allocation framework is then envisioned to adapt the selection of audiovisual segments to be included in the summary according to the needs and preferences of the user. Several contending local stories, also named sub-summaries, are considered to present each segment, so that not only the content, but also the narrative style of the summary can be adapted to user requirements. Hence, by tuning the benefit and the cost of the local stories, our framework, which searches for the combination of sub-summaries that maximizes the overall (user-dependent) benefit under a global duration constraint, becomes able to balance -in a natural and personal way- the semantic (what is included in the summary) and narrative (how it is presented to the user) aspects of the summary. This is a fundamental difference, compared to the approaches that are based on filtering or ranking mechanisms to extract the actions that best match the theme requested by the user, without taking care of fluent story-telling.

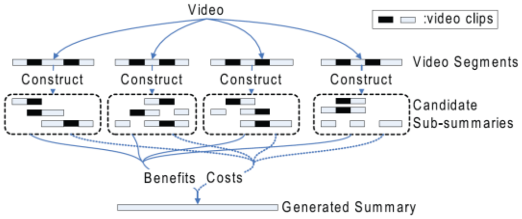


Figure 2. Summarization framework [4]

The proposed framework is definitely generic and flexible, since the cost and benefit functions can be tuned to the application needs. Recently, it has been deployed to summarize (1) the sport-event content broadcasted by common TV channels [4], and (2) the content produced through automatic production tools [5].

In the first case, the video is split into clips, based on conventional shot (or clip) boundary detection tools. The camera switching patterns and the view type structure of the clips are then investigated to divide the audio-visual feed into non-overlapping and semantically meaningful segments. Hot spot detection relies on audio analysis. Due to the lack of space, we refer the readers to an earlier publication for a detailed description of the system, and focus on the second scenario, which offers even more flexibility in terms of adaptation to user needs.

For this second scenario, considered by the FP7 European project APIDIS (www.apidis.org), the generation of concise video reports of the game involves numerous integrated technologies and methodologies, including multi-view capture of the scene, automatic scene analysis, camera viewpoint selection & control, and -ultimately- the generation of summaries through automatic organization of stories.

Specifically, salient video segments are identified based on game actions recognition, which in turns rely on player movement analysis and scoreboard monitoring. Player detection and tracking rely on the fusion of the foreground likelihood information computed in each camera view, which overcomes the traditional hurdles associated with single view analysis, such as occlusions and shadows [6].

Obviously, given the framework presented in Fig.2, the more action types are recognized, the more opportunities we have to personalize the video summaries by refining the benefit function definition. In order to produce semantically meaningful and perceptually comfortable video summaries, based on the extraction of sub-images from the raw content, we introduce three fundamental concepts, namely “completeness”, “smoothness” and “fineness” [6].

- Completeness stands for both the integrity of view rendering in camera/viewpoint selection, and that of story-telling in summary.
- Smoothness refers to the graceful displacement of the virtual camera viewpoint, and to the continuous story-telling resulting from the selection of contiguous temporal segments. Preserving smoothness is important to avoid distracting the viewer from the story with abrupt changes of viewpoint.
- Fineness refers to the amount of detail provided about the rendered action. Spatially, it favors close views. Temporally, it implies redundant story-telling, including replays. Increasing the fineness of a video does not only improve the viewing experience, but is also essential in guiding the emotional involvement of viewers through the use of close-up shots.

The integrated framework resulting from the above consideration is described in [6], and has been successfully demonstrated in the context of basketball games coverage.

4. Conclusions

This letter has surveyed a number of works aiming at personalizing the access to content, based either on interactive streaming mechanisms, or on user-driven edition of summary report. Through the letter, we observe that analysis directly impacts personalization mechanisms, in the sense that (1) richer semantic knowledge increases the number of opportunities for personalization, and (2) an increased accuracy of the analysis reduces the need for interactivity, since an autonomous system can then take the summary edition decisions by itself. The letter also reveals that multi-views acquisitions offers both effective analysis solutions and redundant media content, which in turns increases the flexibility when selecting samples to produce visually pleasant content. In final, multi-camera autonomous production can provide practical solutions to a wide range of applications, such as personalized access to local sport events through a web portal or a mobile hand-set, cost-effective and fully automated production of content dedicated to small-audience, or even automatic log in of annotations.

5. Acknowledgment

The author thanks European Commission and the Walloon Region for funding part of this work through the FP7 APIDIS and WIST2 WALCOMO projects, respectively.

References

[1] E. Bomcke and C. De Vleeschouwer, “An interactive video streaming architecture for H.264/AVC compliant players,” in IEEE Int. Conf. on Multimedia and Expo, New-York, USA, 2009.

[2] I. A. Fernandez, Fan Chen, F. Lavigne, X. Desurmont and C. De Vleeschouwer, Browsing Sport Content Through an Interactive H.264 Streaming Session, 2nd International Conference on Advances in Multimedia (MMEDIA), Athens, June 2010.

[3] I. A. Fernandez, C. De Vleeschouwer, F. Lavigne and X. Desurmont, Worthy visual content on mobile through interactive streaming, IEEE International Conference on Multimedia & Expo, Singapore, July 2010.

[4] Fan Chen, and C. De Vleeschouwer, A resource allocation framework for summarizing team sport videos, IEEE Int. Conf. on Image Processing, pp.4349-4352, Cairo, Egypt, 2009. Extended version accepted by IEEE Trans. On CSVT.

[5] Fan Chen and C. De Vleeschouwer, Personalized production of team sport videos from multi-sensored data under limited display resolution, Computer Vision and Image Understanding, Special Issue on Sensor Fusion, 114(6), 667-680, 2010.

[6] Fan Chen, D. Delannay, C. De Vleeschouwer, and P. Parisot, Multi-sensored Vision for Autonomous Production of Personalized Video Summary, in Book "Computer Vision for Multimedia Applications: Methods and Solutions" (Edited by Jinjun Wang, Jian Cheng, and Shuqiang Jiang), IGI Global, September 2010, ISBN-10: 160960024X.

Christophe De Vleeschouwer is a permanent Research Associate of the Belgian NSF and an Assistant Professor at UCL. He was a senior research engineer with the IMEC Multimedia Information Compression Systems group (1999-2000), and a post-doctoral Research Fellow at UC Berkeley (2001-2002) and EPFL (2004). His main interests concern video and image processing for analysis, communication and networking applications, including intelligent vision, content retieval, adaptive transmission, and media asset management. He is also enthusiastic about non-linear signal expansion techniques, and their use for signal compression and interpretation. He is the co-author of more than 20 journal papers or book chapters, and holds two patents. He serves as an Associate Editor for IEEE Transactions on Multimedia, has been a reviewer for most IEEE Transactions journals related to media and image processing, and has been a member of the (technical) program committee for several conferences, including ICIP, EUSIPCO, ICME, ICASSP, PacketVideo, ECCV, GLOBECOM, and ICC. He contributed to MPEG bodies, and several European projects. He now coordinates the FP7-216023 APIDIS European project (www.apidis.org), and several Walloon region projects, respectively dedicated to video analysis for autonomous content production, and to personalized and interactive mobile video streaming.

High-dimensional Media Compression for Interactive Streaming

Gene Cheung, National Institute of Informatics, Tokyo, Japan
Ngai-Man Cheung, Stanford University, Stanford, CA
cheung@nii.ac.jp, ncheung@stanford.edu

1. Interaction with High-dimensional Media

Due to the decreasing cost of media capturing devices and the proliferation of social networks, available media datasets are now in much higher dimension than traditional media like 2D images and single-view video. For example, *multiview videos* have been captured using up to 100 time-synchronized cameras [FM06]. Due to inherent physical limitations, however, display terminals typically show only a subset of the media to the viewers in lower dimension (e.g., only a single video view can be displayed on a conventional monitor at any point in time no matter how many views were captured). Typically then, a client browses the high-dimensional media by observing low-dimensional media subsets in succession across time. *Interactive streaming* captures this media interaction between server and client: a client continuously requests successive low-dimensional media subsets of her choosing, and in response the server transmits appropriate data for the client to reconstruct requested media subsets for display. By transmitting only data corresponding to the requested media subsets instead of the entire high-dimensional media, interactive streaming can potentially reap tremendous bandwidth saving over non-interactive streaming.

2. High-dimensional Data Compression

Compression of high-dimensional media has been intensively studied in the last few years, with the aim of improving the overall coding efficiency of the entire data set. For example, *multiview video coding* [MVC08] compresses jointly all captured images of all views across all time instants for optimal rate-distortion (RD) performance. For interactive streaming in a store-and-playback scenario (hence real-time encoding is not available), however, the compression problem is more challenging, because traditional differential coding techniques used for coding of correlated data is now difficult to employ. Recall that differential coding, typical in coding of single-view video, assumes a previous frame F_{i-1} of time instant *i-1* is available at decoder for prediction of target image F_i of instant *i*, so that only differential $F_i - F_{i-1}$ needs to be encoded. If media subset selection by client at stream time is not known at coding time, then no subset can be assumed to be available at decoder with certainty for prediction of the target subset, and traditional differential coding cannot be applied as is. A simple alternative strategy is to forego differential coding all together and encode every media subset independently. However, this will result in a large server transmission rate, since no correlation among subsets is exploited for coding gain.

3. Compression for Interactive Streaming

Over the past few years, researchers have devised novel coding structures and techniques to achieve different tradeoffs between media interactivity and coding efficiency. We provide a brief sampling of compression techniques for different interactive streaming applications in this paper. They include: reversible video playback, interactive light field streaming, interactive multiview video streaming, and interactive region-of-interest (RoI) video streaming.

Reversible video playback such as backward-play, backward-step or fast-backward has garnered a fair amount of research interest [WV99, LZ01, FC06, CW06]. These functionalities are particularly useful for surveillance applications, where videos may be carefully inspected in both forward/backward directions for unusual events. Supporting reversible video playback in MPEG/H.26X coded videos, however, is a challenging task, as differential coding is usually employed to encode the video in the forward-play direction only. Simple solution such as encoding all frames independently using intra-coding is costly for storage and transmission. Therefore, more sophisticated approaches have been proposed. For example, [LZ01] presented a MPEG/H.26X compatible solution by generating a reverse-encoded bitstream in addition to the forward-encoded one. [CW06] proposed to encode the video using distributed source coding (DSC) to facilitate reversible playback. In particular, the DSC encoder generates parity information to represent the current frame. The amount of parity is chosen so that current frame is bi-directionally decodable, i.e., it can be reconstructed using either the previous or the subsequent frame as side information, for forward- and backward-play, respectively.

Light field [LH96] is a large set of correlated images of the same static scene taken from a 2D array of closely spaced cameras. Typically, a client browses the light field data by observing single images along a view trajectory (contiguous succession of adjacent views) of her own choosing across time. This assumption of only neighboring view switches implies that one out of a small subset of adjacent frames must be available at decoder for prediction of the target image during a view-switch. [RG04] proposed to differentially encode one SP-frame for each predictor frame, so that the server can transmit an SP-frame corresponding to the predictor frame residing in the decoder during stream time. The identical construction property of SP-frames ensures the same reconstruction of the target image no matter which SP-frame (corresponding to the predictor frame in the decoder buffer) was actually transmitted. Alternatively, [AR04] proposed to use DSC instead, where the number of Least Significant Bits (LSB) bit-planes that need to be transmitted depends on the quality of the side information; i.e., the maximum difference between the predictor frame at decoder and the target image.

Interactive multiview video streaming is an application where a streaming client observes a single view a time, but can interactively switch views as video is played back in time uninterrupted. To provide view switching capability for the observer while maintaining good compression efficiency, [CO09] developed a *redundant P-frame* representation where multiple P-frames are encoded for the same original picture and stored at the encoder, each using a different previous frame as predictor that was on a possible observer's navigation trajectory. Thus, several P-frames are available to reconstruct a given frame when different decoding paths are followed. This reduces the need for retracing and leads to overall lower bandwidth. However, because multiple redundant P-frames are stored, overall storage is increased at the sender. Note that an indiscriminate use of redundant P-frame representation will lead to exponential expenditure in storage. To avoid such a problem without resorting to bandwidth-expensive I-frame, [CO09b] developed novel DSC implementations to merge multiple decoding paths into a single frame representation. Recent work [CC09] discussed preliminary results of using I-, P- and DSC frames in an optimized structure for interactive multiview video streaming.

Interactive RoI streaming [MB07, MA10] stems from the challenge that while videos are being captured at increasingly high resolution (e.g., Ultra HD), in many cases display terminals could be relatively small (e.g., on smart-phones), or the communication bandwidth could be limited. Interactive RoI streaming addresses the problem by encoding the high-resolution videos in a way that user-selected RoI can be readily extracted and streamed. This leads to video streaming with interactive pan/tilt/zoom functionality, where only those regions of the videos that are desired at the client's end are delivered. To support interactive RoI streaming, [MB07] proposed a spatial-random-access-enabled video compression algorithm, where differential coding is employed but the prediction structure is carefully designed to allow RoI extraction. In particular, input video is encoded into multiple resolution layers. Each resolution layer is subdivided into different tiles. Tiles in high resolution layers are predicted by collocated tiles of the lowest resolution layer (base layer) of the same time instant. To facilitate spatial random access, no temporal prediction is used for high resolution layers. During the streaming session, base layer is always delivered. Upon receiving a RoI request from the client, high resolution tiles of the corresponding spatial locations are streamed. Tile size could be optimized by considering the tradeoff between coding efficiency and pixel overhead.

4. Future Work

While coding techniques for different interactive streaming applications have been proposed to trade off compression efficiency with media interactivity, several problems remain unsolved towards a complete end-to-end networked system. First, if the transmission network is packet-loss prone, then an irrecoverably lost packet can lead to error propagation in differentially coded frames. Due to different navigation trajectories possible in interactive streaming, anticipating this error propagation a priori at coding time and devising an appropriate coding strategy to contain the damage is one open problem. Second, typical transmission networks exhibit non-negligible round-trip-time (RTT) delay between server and client, which greatly affects the reaction time of each client's media subset request. How intelligent coding structures and streaming protocols can be co-designed to overcome this RTT interactive delay is another open problem. We invite researchers in the media communication research field to tackle these challenging problems in interactive streaming.

References

[FM06] T. Fujii, T. Mori, K. Takeda, K. Mase, M.

Tanimoto, Y. Suenaga, “Multipoint measuring system for video and sound—100 camera and microphone system,” in *IEEE International Conference on Multimedia and Expo*, Toronto, Canada, July 2006.

[MVC08] “MPEG—technologies—introduction to multview video coding,” January 2008, ISO/IEC JTC 1/SC 29/WG 11 N9580.

[WV99] S. J. Wee and B. Vasudev, “Compressed-domain reverse play of MPEG video streams,” in *Proc. Multimedia Systems and Applications*, 1999.

[LZ01] C. W. Lin, J. Zhou, J. Youn, and M. T. Sun, “MPEG video streaming with VCR functionality,” *IEEE Trans. Circuits and Systems for Video Technology*, vol. 11, no. 3, Mar. 2001.

[FC06] C.-H. Fu, Y.-L. Chan, and W.-C. Siu, “Efficient reverse-play algorithms for MPEG video with VCR support,” *IEEE Trans. Circuits and Systems for Video Technology*, vol. 16, no. 1, Jan. 2006.

[CW06] N.-M. Cheung, H. Wang, and A. Ortega, “Video compression with flexible playback order based on distributed source coding,” in *Proc. Visual Communications and Image Processing (VCIP)*, 2006.

[LH96] M. Levoy, P. Hanrahan, “Light field rendering,” in *Proc. SIGGRAPH’96*, August 1996.

[RG04] P. Ramanathan, B. Girod, “Random access for compressed ligh fields using multiple representations,” in *IEEE International Workshop on Multimedia Signal Processing*, Siena, Italy, September 2004.

[AR04] A. Aaron, P. Ramanathan, B. Girod, “Wyner-Ziv coding of light fields for random access,” in *IEEE International Workshop on Multimedia Signal Processing*, Siena, Italy, September 2004.

[CO09] Gene Cheung, Antonio Ortega, Ngai-Man Cheung, “Generation of Redundant Frame Structure for Interactive Multiview Streaming,” *17th International Packet Video Workshop*, Seattle, WA, May 2009.

[CO09b] Ngai-Man Cheung, Antonio Ortega, Gene Cheung, “Distributed Source Coding Techniques for Interactive Multiview Video Streaming,” *27th Picture Coding Symposium*, Chicago, IL, May 2009

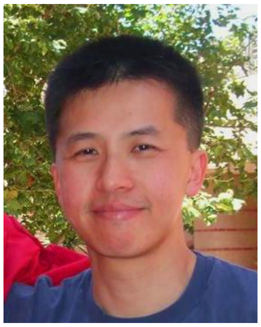
[CC09] Gene Cheung, Ngai-Man Cheung, Antonio Ortega, “Optimized Frame Structure using Distributed Source Coding for Interactive Multiview Video Streaming,” *IEEE International Conference on Image Processing*, Cairo, Egypt, November 2009.

[MB07] A. Mavlankar, P. Baccichet, D. Varodayan, and B. Girod, “Optimal Slice Size for Streaming Regions of High Resolution Video with Virtual Pan/Tilt/Zoom Functionality,” in *Proc. European Signal Processing Conference (EUSIPCO-07)*, Poznan, Poland, September 2007.

[MA10] A. Mavlankar, P. Agrawal, D. Pang, S. Halawa, N.-M. Cheung and B. Girod, “An Interactive Region-of-Interest Video Streaming System for Online Lecture Viewing,” in *Proc. International Packet Video Workshop (PV2010)*, Hong Kong, China, December 2010.

Gene Cheung received the B.S. degree in electrical engineering from Cornell University in 1995, and the M.S. and Ph.D. degrees in electrical engineering and computer science from the University of California, Berkeley, in 1998 and 2000, respectively. He was a senior researcher in Hewlett-Packard Laboratories Japan, Tokyo, from 2000 till 2009. He is currently an assistant professor in National Institute of Informatics in Tokyo, Japan.

His research interests include media representation & network transport, single- / multiple-view video coding & streaming, and immersive communication & interaction. He has served as associate editor of IEEE Transactions on Multimedia since 2007, served as area chair in IEEE International Conference on Image Processing (ICIP) 2010, and serves as technical program co-chair of International Packet Video Workshop (PV) 2010. He will serve as track co-chair for Multimedia Signal Processing track in IEEE International Conference on Multimedia and Expo (ICME) 2011. He is a co-recipient of the Top 10% Paper Award in IEEE International Workshop on Multimedia Signal Processing (MMSP) 2009.



Ngai-Man Cheung received the Ph.D. degree from University of Southern California (USC), Los Angeles, in 2008. He is currently a postdoctoral researcher with the Information Systems Laboratory, Stanford University, Stanford, California. His research interests are multimedia signal processing and compression. He received paper awards from EURASIP Journal of Advances in Signal Processing, IEEE International Workshop on Multimedia Signal Processing (MMSP) 2007, IS&T/SPIE VCIP 2008, and from USC Department of Electrical Engineering in 2008.

Semantic Image Adaptation for User-centric Mobile Display Devices

Wenyuan Yin, SUNY at Buffalo, Buffalo, NY 14260 USA
Jiebo Luo, Kodak Research Laboratories, Rochester, NY 14650 USA
Chang Wen Chen, SUNY at Buffalo, Buffalo, NY 14260 USA
wyin4@buffalo.edu; jiebo.luo@gmail.com; chencw@buffalo.edu

1. Introduction

It has becoming more and more popular for users to generate and share the media content on different types of terminals and networks through diverse portable mobile devices such as PDA, handheld PC and cellular phones. The popularity of the mobile devices triggers the need for mobile users to access and interact with the multimedia content anywhere at any time. To facilitate mobile usage, mobile devices are usually designed to be light weight and to serve the short viewing distance. On one hand, such a design benefits mobility and allows ubiquitous access of multimedia content; on the other hand, it has been the main reason for the small display size of mobile device and the main limiting obstacle to affect user's viewing experience.

Recent developments towards ubiquitous multimedia and the pressing need of improving the user's media experience have invoked the paradigm shifting from traditional device-centric multimedia to contemporary user-centric multimedia. In this new paradigm, users are placed in the center of ubiquitous multimedia environment and are provided with access to personalized multimedia content intelligently. Regardless of mobile device type and capabilities, seamless access is desired to maximize the quality of experience for mobile users, based on the media content and user's intention. To achieve such a goal, proper adaptation on high resolution media content is necessary to fit the various limited display of mobile devices, the heterogeneity of network links, and distinct customer intention of mobile users.

To obtain adaptation results which are consistent with the media contents and mobile user intentions, the semantic gap and user intention gap are two critical challenges that need to be properly addressed. According to 0, human tends to view and comprehend the images based on *semantics* which refers to the creatures and objects comprising the scene and the relations among the entities. Semantic based image adaptation aims at providing mobile users better perceptual experiences. To carry out semantic based image adaptation, semantic analysis to extract the key persons or objects and associate them with high level concepts 0 is an indispensible step. The well known 'semantic gap' 0 between low level visual features and high level concepts is one critical challenge we have to address for image adaptation.

The second critical challenge in image adaptation is how to narrow down the mobile user intention gap. The mobile user intention gap is caused by the following: 1) The mobile device interface poses great difficulty in obtaining users' real intentions because it often does not allow complicated inputs and frequent interactions. 2) The intention gap is further broadened on mobile devices by the small displays of mobile devices. When high definition images are presented on small mobile devices, direct down sampling to meet mobile display capacity may lead to unacceptable quality or even unrecognizable images.

In this research, we develop a novel semantic image adaptation scheme for heterogeneous mobile display devices. This scheme integrated the content semantic importance with user preferences under limited mobile display constraints. The main innovations of the proposed scheme are: 1) seamless integration of mobile user supplied query information with low level image features to identify semantically important image contents. 2) Integration of semantic importance and user feedback to dynamically update mobile user preferences. 3) Perceptually optimized adaptation for image display on mobile devices.

2. Mobile User Guided Adaptation System

2.1. System Components

As shown in Figure 1(a), the proposed system consists of (1) an adaptation proxy to process user request and feedback as well as to carry out semantic extraction, user preference learning and adaptation; (2) a server/database hosting original consumer photo content. We assume the annotation of the server side media content is processed off-line while the user request and feedback processing is carried out in real time.

2.2. System Workflow

The proposed system works as follows. As shown in Figure 1(b), a user first inputs the semantic request in the form of the keywords for the desired media content through the user interface at the mobile device. Such a semantic request is then sent to the query processing module of the adaptation proxy. As most people would like to input the activities or events as the keywords, we assume the system takes queries in the form of events. Upon preprocessing, a request containing the semantic event information is forwarded to the server or database, where the media contents are assumed already tagged with event concepts and other annotations. The database retrieves the photos best matching the request and sends the top thumbnail retrieval results to adaptation proxy.

Afterwards, as illustrated in Figure 1(c) and (d), user selects thumbnails containing his/her interested contents and then the server sends the selected full size image to the adaptation proxy. The original full size media contents stored on the server cannot be directly displayed on the user mobile devices due to the mismatch between these usually high resolution of original images and the small display size and resolution of the mobile devices. To provide optimal user perceptual experiences, media content adaptation is necessary to adapt the original media content into the personalized form that is semantically important and perceptually optimal to the current user's interests while still meeting the display size limitation.

To refine the adaptation for mobile users' interests, user feedback is designed to learn the mobile user preference as shown in Figure 1(c) and (d). Since the system initially has no idea about the individual user's true query intention, when retrieving the first batch of images for the given query of the current mobile user, he/she can select several thumbnails of interest as shown in (c). Then, the adaptation proxy will present four adaptation candidates for each selected image. Next, the mobile user can grade these adaptation candidates with preference values to them. Afterwards, the preference values are fed back to the adaptation proxy for user preference learning. Subsequently, as shown in (d), when the user select an image of interest, the system will present the optimal personalized adaptation to the user under the mobile display constraints and update the user preference continuously.

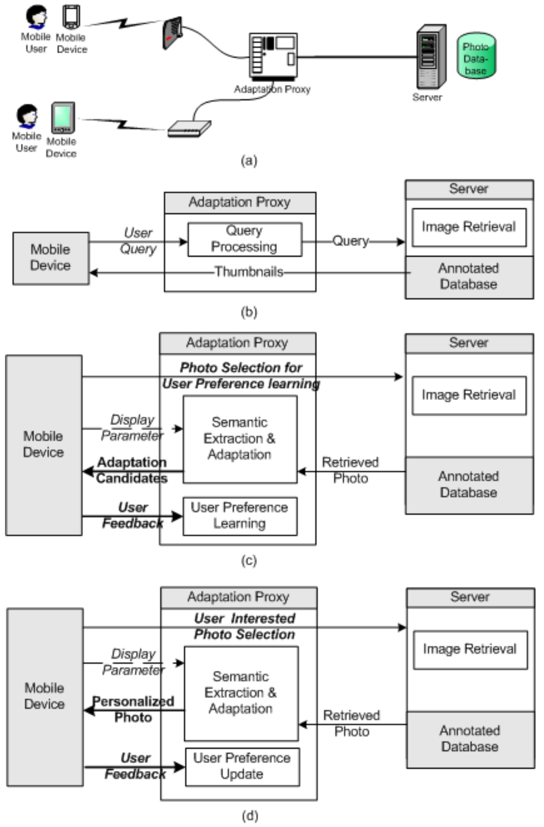


Figure 1. System Framework (a) General system (b) User query processing (c) Semantic image adaptation for mobile user preference learning (d) Mobile user guided semantic photo adaptation.

In the proposed system, mobile user supplied semantic information is seamlessly integrated with low level media features for important contents extraction. The user feedbacks are integrated to learn and update the mobile user preference. Based on the display capacity and the individual mobile user preference, the adaptation decision module determines the parameters for adaptation manipulation to select and resize contents of different significances. These parameters are fed into adaptation server to generate optimal adaptation results to send to the mobile users. This system intends to provide mobile user desired photo adaptation with personalized optimal perceptual experience under various limited display sizes. This paper mainly focuses on the semantic extraction, mobile user preference learning and adaptation modules, which will be introduced in details in the following sections.

3. Brief Description of Key Components

3.1. Mobile User Guided Adaptation System

To provide high quality semantic adaptation under display limitations, it is necessary to identify the location of the semantically important objects with different relevance to the corresponding events such as bride in wedding photos. In this research, to extract these key contents, we design a Bayesian fusion approach to properly integrate low level features with high level semantics. The innovation of the proposed extraction method resides in the seamless integration of user provided semantic information with the low level media features, by which the semantically important and probably user interested contents are extracted for media adaptation. Low level features are extracted in a bottom-up fashion while the high level semantic extraction is obtained in a top-down style. The proposed semantic object extraction is shown in Fig. 2. The details of the Bayesian fusion can be found in [4].

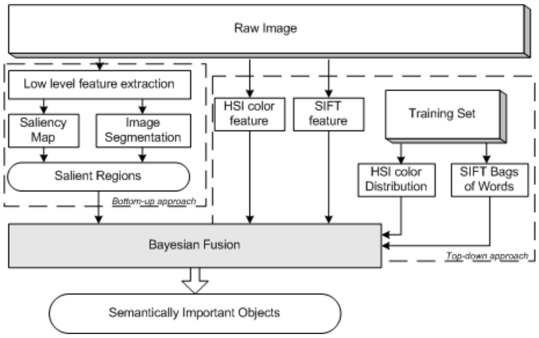
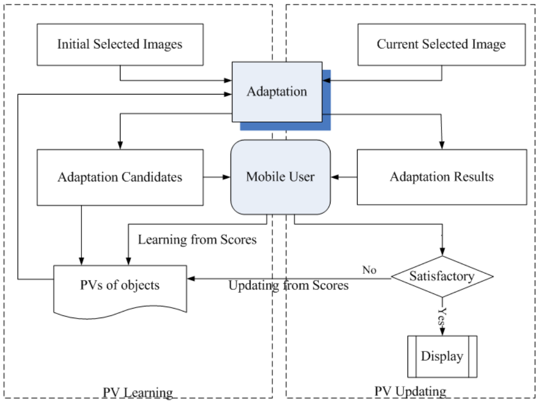


Figure 2. Semantic Extraction: Bayesian fusion of bottom-up low level features and top-down high level semantics.



3.2. *User Preference Learning for Adaptation*

To present adaptation results that are truly consistent with the mobile user's true interest, it is necessary to fine tune the importance of extracted objects and make them matched to the mobile users' preference. Since mobile user preferences are subjective measures varied among individuals, it is necessary to learn from users the subjective Preference Value (PV) on different objects. We adopt the well-known relevance feedback concept [5] to bridge the intention gap in retrieving more relevant images consistent with user's interests.

The feedback process for user PV learning and updating is illustrated in Figure 3. For each query, since the system has no idea initially about the mobile user's preference, it will allow the user to select several images to learn the user's interests through the user feedback scores upon those adaptation candidates for the selected images. In our implementation, four images are selected initially for user preference learning. For each image, four adaptation candidates are presented. Due to the small display, each time, only the four adaptation candidates for one image are displayed for grading. All graded scores of the mobile user for the sixteen adaptation candidates are utilized for user preference learning.

After learning from the first batch of adaptation candidates, reasonable PVs upon different objects have been obtained. Later on, whenever the mobile click an image, the system will automatically adapt the image based on the PVs of objects to display. Also, four personalized adaptation results will be shown for user selection. The user can decide whether to grade them or not depending on his satisfactory degree on them. If he selects one to display without grading, it means that the adaptation is satisfactory enough and the PVs need no more update. If he is not satisfactory on the adaptation result, he will grade the results and the PVs of objects will be updated by the new scores obtained.

Figure 3. Feedback process for PV learning and updating.

3.3. *User Centric Semantic Adaptation*

The goal of user centric adaptation is to simultaneously panelize the selection of contents not preferred by the user and reward the user preferred objects with high quality depending on the degree of their relevance to user, under the limited mobile display constraints. In the PV learning stage, before getting the user feedback, we can assign objective semantic importance (OSI) to meaningful objects. By the integration of OSI and feedback, we have already obtained such relevance of different objects to different mobile users which are denoted as PVs. In the following step, we utilize the PVs of objects to guide the adaptation to provide the mobile user the best possible perceptual experience. The optimal adaptation is performed and presented by formulating it into an information fidelity (IF) maximization problem as discussed below.

In this image adaptation model, each extracted semantically important object is assigned three attributes: region, Preference Value (PV) and minimum acceptable size (MAS). The regions of objects are obtained in the semantic adaptation process as described before. The PVs of objects are converged to the values consistent with the current user's preference by the integration of OSI and the user's feedbacks. The MAS of objects of interest is calculated by the MAS determination scheme that considers the information loss curve introduced by image region down scaling analyzed by Kullback–Leibler divergence [6]. Once the value of MAS is obtained, we measure the information loss caused by adaptation. The details of MAS determination can be found in [7].

4. Results

We build our consumer photo dataset of various images with different sizes obtained from popular Web sites, such as GOOGLE, FLICKR, and PICASA, etc. For each event concept, there are 100 photos. We conduct the experiments for the following five event concepts: wedding, graduation, baseball, football, and beach fun, on the consumer database using 500 images. For each concept, we use half of the photos as training set and the other half as test set.

We conduct experiments on semantic extraction for each event concept to compare our proposed fusion scheme with the bottom-up only and top-down only schemes. We use F-measure to evaluate the deviation of the detected bounding box. F-measure is the weighted harmonic mean of precision and recall with a non-negative β:

$$F_{\beta} = \frac{(1+\beta)\times Precision \times Recall}{\beta \times Precision + Recall} \qquad (1)$$

We set β =0.5 as in 0. The comparison result is demonstrated in Table I.

TABLE I
F-MEASURE COMPARISON OF DETECTION RESULTS

Event	Bottom-up	Top-down	Integration
Wedding	0.70	0.76	0.85
Graduation	*0.87*	*0.81*	*0.90*
Baseball	0.75	0.76	0.88
Football	0.88	0.79	0.88
BeachFun	0.89	0.79	0.90

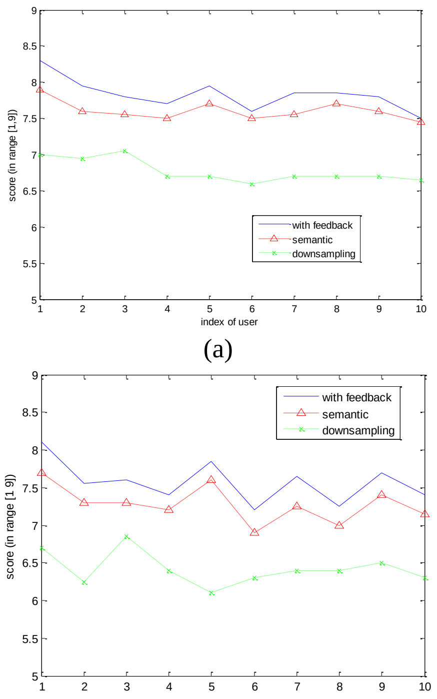
In the experiment, each test image is adapted given the target display size of 120 x 160 and 240 x 320, respectively. Our user guided semantic adaptation scheme is compared with results from direct down-sampling, attention-based adaptation approach [9] utilizing saliency and face detection, since most of the existing adaptation approaches such as in [10] are based on saliency and face detection. The results show that the proposed scheme can catch and highlight the objects that generally users are more interested in and can use the small screen of mobile device better than the direct resizing and attention-based adaptation schemes.

To validate that the proposed user preference learning and updating scheme can better catch individual user's interested objects and perform personalized adaptation to improve perceptual experience of users, we design experiments to quantify the improvement in adaptation results.

To compare the adaptation results based only on semantics, the results based on learned PVs of objects and the results from direct down sampling, 10 users are invited to provide their assessment to the three groups of adaptation results. Each user select 20 images from database and the three adaptation schemes are performed to the selected images. The subjective scores range from 1 to 9, in which scores 1-4 refers to non-relevant, scores 6-9 refers to relevant to the user interests, and 5 is no opinion. The larger the score of the result, the more satisfactory the user is. In Figure 7, the means of subjective scores of each user are drawn for adaptation results based on user preference, semantics and direct down sampling under mobile display of 320×240 and 176×144, respectively.

In Table II, the average scores of the adaptation results of all users using the three methods are shown. It demonstrates that the personalized adaptation results based on user feedback indeed provide more satisfactory results to the users.

To better illustrate the effectiveness of the proposed personalized adaptation, in Figure 8, we demonstrate the example adaptation results for two users with different preferences in wedding event given mobile display resolution of 176 ×144. One user is interested in bride only and the other is interested in both bride and groom. In this experiment, the same four images are used to learn their preferences respectively. In this figure, it shows the best adaptation results with learning in terms of the total information fidelity for the two users in (b) and (c), respectively, after user preference learning. It shows that the adaptation results are indeed personalized effectively according to their feedbacks.



(b)

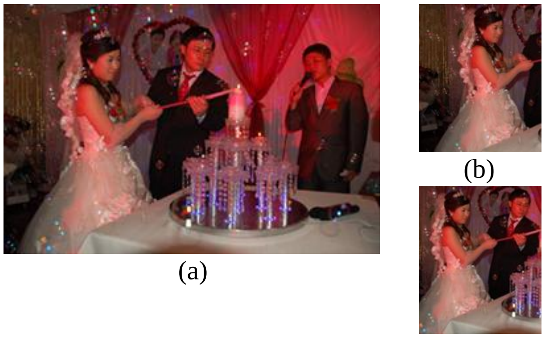
Figure 7. Means of subjective scores of each user for adaptation results using three methods: (a) Score distribution for 320×240 display size (b) score distribution for 176×144 display size

TABLE II
SUBJECTIVE SCORE COMPARISON OF ADAPTATION RESULTS

Resolution	Adaptation with preference	Semantic based adaptation	Direct down sampling
320×240	7.83	7.61	6.78
176×144	7.57	7.28	6.42

5. Conclusions

A novel semantic image adaptation scheme has been developed for mobile display devices. This scheme aims at providing mobile users with most desired image content by integrating the content semantic importance and user preferences under the limited mobile environment constraints. Furthermore, to bridge the semantic gap for adaptation, we have designed a Bayesian fusion approach to properly integrate low level features with high level semantics. To handle the preference variations among different mobile users, mobile user relevance feedback scheme has been developed to learn and update user preferences. Extensive experiments have been carried out to validate the proposed adaptation scheme. The experiments show that by adopting the proposed semantic adaptation scheme with integration of the semantics and mobile user preferences, perceptually highly relevant image adaptation can be effectively carried out to match the user intentions under the mobile environment constraints. We expect that the closing of semantic gap and user intention gap will continue to improve the personalized adaptation results to provide mobile users with improved perceptual experiences.



(a)

(b)

(c)

Figure 8. Example personalized adaptation results: (a) Original image (b) adaptation result (176×144) for user prefers bride (c) adaptation result (176×144) for user prefers bride and groom both

References

[1] I. Biederman, “On the Semantics of a Glance at a Scene”. In M. Kubovy and K. R. Pomerantz, editors, Perceptual Organization, pages 213-263. Lawrence Erlbaum Publisher, 1981.

[2] A.C. Loui, J. Luo, S.-F. Chang, et al. “Kodak video benchmark dataset: concept definition and annotation.” ACM MIR 2007.

[3] J. S. Hare, P. H. Lewis, P. G. B. Enser, and C. J. Sandom, “Mind the gap: Another look at the problem of the semantic gap in image retrieval,” Proc. SPIE, vol. 6073, 2006.

[4] W. Yin, J. Luo and C. W. Chen, “Semantic adaptation of consumer photos for mobile device access,” Proc. of IEEE International Symposium on Circuits and Systems, Paris, France, May 2010.

[5] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra, “Relevance Feedback: A Power Tool in Interactive Content-Based Image Retrieval”, IEEE Trans. on Circuits and Systems for Video Technology , Special Issue on Segmentation, Description, and Retrieval of Video Content, pp644-655, Vol 8, No. 5, Sept, 1998.

[6] X. Fan, X. Xie, W. Ma, "Detecting the Sufficient Display Resolution for Image Browsing", International Conference on Mobile Data Management, 2006.

[7] W. Yin, J. Luo and C. W. Chen, "User guided semantic image adaptation for mobile display devices," Invited paper at IEEE ICME Special Session on Human-Centric Multimedia Communications, Singapore, July 2010.

[8] D. R. Martin, C. C. Fowlkes, and J. Malik. "Learning to detect natural image boundaries using local brightness, color, and-texture cues". IEEE Trans. on PAMI, 26(5):530–549, 2004.

[9] L.Q. Chen, X. Xie, X. Fan, W.Y. Ma, H.J. Zhang, H.Q. Zhou, "A visual attention model for adapting image on small displays," ACM Multimedia Systems Journal, 2003.

[10] H. Liu, S. Jiang, Q. Huang, C. Xu, W. Gao: Region-based visual attention analysis with its application in image browsing on small displays. ACM Multimedia, pp. 305-308, 2007.

Wenyuan Yin received B.E. degree from Nanjing University of Science and technology in 2006. She is now pursuing the Ph.D. degree in the Department of Computer Science and Engineering, State University of New York at Buffalo. Her current research interests include image and video semantic understanding, media adaptation, video transcoding.

Jiebo Luo (F'08) received the B.S. degree from the University of Science and Technology of China in 1989 and the Ph.D. degree from the University of Rochester, Rochester, NY, in 1995, both in electrical engineering. He is a Senior Principal Scientist with the Kodak Research Laboratories, Rochester, NY. His research interests include signal and image processing, machine learning, computer vision, multimedia data mining, and computational photography. He has authored over 130 technical papers and holds nearly 50 U.S. patents.

Dr. Luo has been involved in organizing numerous leading technical conferences sponsored by IEEE, ACM, and SPIE. Currently, he is on the editorial boards of the IEEE T-PAMI and IEEE T-MM, Pattern Recognition (PR), and the Journal of Electronic Imaging. He is Editor-in-Chief of the Journal of Multimedia (Academy Publisher). He is a Kodak Distinguished Inventor, a winner of the 2004 Eastman Innovation Award (Kodak's highest technology prize), a member of ACM, and a Fellow of SPIE.

Chang Wen Chen (F'04) received the B.S. degree from the University of Science and Technology of China in 1983, the M.S.E.E. degree from the University of Southern California, Los Angeles, in 1986, and the Ph.D. degree from the University of Illinois at Urbana-Champaign in 1992.

He is a Professor of Computer Science and Engineering at the State University of New York at Buffalo. Previously, he has been Allen Henry Endow Chair Professor at the Florida Institute of Technology from 2003 to 2007, on the faculty at the University of Rochester from 1992 to 1996, on the faculty at the University of Missouri-Columbia from 1996 to 2003.

Prof. Chen served as the Editor-in-Chief for IEEE Trans. Circuits and Systems for Video Technology for two terms from January 2006 to December 2009. He has been an Editor for numerous IEEE Transactions and Journals, including Proceedings of IEEE, IEEE Journal of Selected Areas in Communications, IEEE Trans. Multimedia, and IEEE Multimedia Magazine. He has also served as Conference Chair for several major IEEE, ACM and SPIE conferences related to mobile wireless video communications and signal processing. He has received numerous research, service, and best paper awards, including the 2003 Sigma Xi Excellence in Graduate Research Mentoring Award and 2009 Alexander von Humboldt Research Award. He is also an SPIE Fellow.

TECHNOLOGY ADVANCES

Mobile Internet Television (IPTV)

Guest Editor: Kyungtae Kim, NEC Laboratories America, USA
kyungtae@nec-labs.com

Internet Protocol Television (IPTV) is a technology used to deliver high quality interactive traffic and other entertainment contents including television signal, digital video, audio/voice, text, and data, through over IP-based networks with support for guaranteed quality of service. This service is now expanded to mobile and wireless networks including 3G/4G wireless networks. According to ITU-T's definition (ITU-T is a front-head of Mobile IPTV standards), IPTV is access agnostic and thus various wireless networks can exist as an access network. However, much of the mobile/wireless enabled IPTV technology is still immature and has yet to be properly stressed. In this special issue, the editorial team has the great honor to invite some pioneer researchers from industry to academic area for six papers to present their state-of-the-art accomplishments, share their latest experiences, and outline future directions in mobile IPTV.

The first paper, titled "Mobile IPTV: Standards and Technical Challenges toward its Full Potential" by Soohong Park and Choong Seon Hong, addresses current standard activities and technical challenges of mobile IPTV service. The authors introduce ITU-T's work on a high-level architecture of Mobile IPTV system, network interface as well as service definition. It also discusses a number of technical challenges to be solved to enable Mobile IPTV service. This paper provides some useful guidance for the future IPTV service.

In the second article, "IPTV Convergence" by John Cosmas and Qiang Ni, the authors discuss digital video broadcast technologies over the fixed, mobile, and fixed/mobile converged environment. For fixed networks, the authors describe the necessity of a secured wireless LAN that operates in licensed spectrum and accommodates to the need of a broadband home user. For wireless networks, the paper proposes a convergence solution for collaborative cellular, terrestrial, and satellite broadcast networks.

Providing high resolution IPTV over broadband wireless access networks will open new horizons for surveillance, for news gathering and for users in underserved areas. The authors of the third paper, titled "High Resolution IPTV over Broadband Wireless Access Networks" by Omneya Issa, Wei Li and Hong Liu, discuss a potential of providing a good video quality, satisfying IPTV/video QoE requirements, for last-mile indoor and outdoor deployment scenarios. The authors also identify interesting issues which should be considered in optimizing Mobile IPTV service, such as, the tradeoff between network capacity and video quality, the adaptive versus fixed modulation modes, QoS classification and traffic prioritization, and the effect of line-of-sight on the quality of service.

The paper, titled "Efficient IPTV Services Delivery using SVC Adaptation and Cooperative Prefetching" by Toufik Ahmed, Ubaid Abbasi and Samir Medjiah discusses the use of scalable video coding (SVC) to support heterogeneous terminals while overcoming the bandwidth fluctuations which is very common in Internet. The authors introduce a novel system architecture using SVC adaptation with multicast groups for fast channel switching which enables an efficient delivery of mobile IPTV services. The cooperative prefetching scheme is presented to provide smooth seek operations by minimizing the seek operations which minimizes the seek delay in VoD streaming as well as provides continuous service delivery.

The fifth article, Cross-Layer Coding Optimization for Mobile IPTV Delivery" by James Ho and En-hui Yang, introduces a framework for mobile IPTV delivery based on the joint coding optimization of a number of cross-layer coding techniques. The proposed framework addresses the issues of multi-user diversity and transmission loss in wireless multicast multimedia applications.

P2P streaming has attracted much attention recently with promises for higher revenues and better load distribution. The paper, titled QoE-Aware P2P Video-on-Demand using Scalable Video Coding and Adaptive Server Allocation" by Osama Abboud, Konstantin Pussep, Ralf

Steinmetz, and Thomas Zinner, presents a streaming system that uses SVC to adapt to different user requirements and resources. The paper discusses a novel QoE-aware layer selection algorithm that maximizes flexibility through SVC.

I hope you will enjoy reading the interesting papers and how they approach to solve the challenges issues.

Kyungtae Kim received a M.S. degree in Department of Computer Science from the Columbia University in 2000, NY and his Ph.D. degree in the department of Electrical and Computer Engineering at Stony Brook University in 2006 and currently is an adjunct professor in the Department of Electrical and Computer Engineering at Stony Brook University as well as working for NEC Laboratories America Inc during 8 years until now in the areas of multimedia interactive communications over the wireless network, cognitive radio, wireless mesh networks, and mobility management over the heterogeneous networks including 802.16/802.11 networks.

Mobile IPTV: Standards and Technical Challenges toward its Full Potential

Soohong Park, Samsung Electronics, Korea
Choong Seon Hong, Kyung Hee University, Suwon, Korea
soohong.park@samsung.com, cshong@khu.ac.kr

1. Mobile IPTV

Mobile IPTV [1] lets mobile users transmit and receive multimedia traffic, such as TV signals, video, audio, text, and graphics, through IP-based networks with the support of quality of service (QoS) and quality of experience (QoE), security, mobility, and interactivity. In short, Mobile IPTV extends many IPTV services to mobile users. Mobile network has improved the quality of wireless interface but still has led to challenges in contents delivery and seamless mobility. Nevertheless, mobile network is being used for multimedia delivery, entertainment, information and communication.

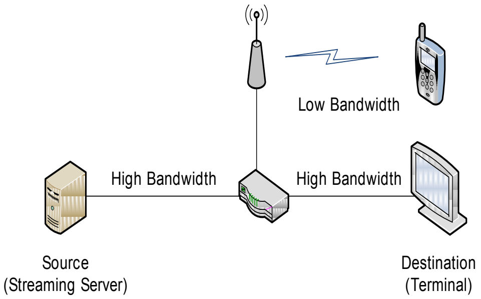


Figure 1. IPTV connectivity extension to mobile devices)

Figure 1 shows an overall IPTV architecture. In this figure, a wireless interface enables communication between the access network and the Mobile IPTV terminal. Because IPTV is access agnostic according to ITU-T's definition, various wireless access networks can exist. Each wireless technology has its own characteristics, and service providers should carefully consider them when deploying Mobile IPTV.

This paper describes the current status of Mobile IPTV in several outstanding standardization efforts, and the technical challenges.

2. Mobile IPTV Standards and Technical Challenges

Currently, ITU-T is a front-head of Mobile IPTV standards. In 2006, ITU-T formed a focus group called FG IPTV to coordinate and promote development of global IPTV standards, which took into account the existing work of IPTV and other standards development organizations. Then, in January 2008, the IPTV Global Standards Initiative (IPTV-GSI) took over the IPTV standardization role.

Until now, no Mobile IPTV architecture is clearly defined in the market, and several models are competing such as IPTV service over Wi-Fi [2], IPTV service over WiMAX [3], and IPTV service over Cellular [4]. For that, ITU-T is currently working on the Mobile IPTV high-level architecture. ITU-T Q.13/16 – Multimedia application platforms and end systems for IPTV is in progress (H.IPTV-TDES.4 in Table 1). The scope of ITU-T standardization is high-level design for Mobile IPTV architecture, network interface as well as service definition.

Table 1: Mobile IPTV Standardizations in ITU-T

Category	ITU-T Recommendations
IPTV Multimedia Application	-H.700(H.IPTV-MAP): Multimedia application platforms and end systems for IPTV -H.701(H.IPTV-CDER): Content delivery error recovery for IPTV services
IPTV Terminal Device	-H.720(TDES.0): Overview of IPTV TD and End Systems -H.721(TDES.2): IPTV Terminal Devices for Basic Services -H.IPTV-TDES.3: IPTV Terminal Device Full-Fledged Model -H.IPTV-TDES.4: IPTV Terminal Device Mobile Model
IPTV Middleware	-H.IPTV-WBTM: Web-based Terminal Middleware for IPTV -H.IPTV-DSMW: Distributed Service Middleware for IPTV

Mobile IPTV services should overcome several obstacles to a successful launch and wide use. Mobile IPTV implies at least one wireless link between the source, such as a streaming server, and the destination, such as a mobile terminal. Therefore, most of the technical challenges are related to the wireless link. The imperative is the operationalization of fast adaptive streaming for IPTV multimedia over wireless as well as unstable mobile networks. For that, HTTP based adaptive streaming is now booming up and being widely deployed in mobile devices. The best practices are

Apple’s HTTP Live Streaming and Microsoft Smooth Streaming. There are several standards activities such as MPEG DASH (Dynamic Adaptive Streaming over HTTP), 3GPP adaptive HTTP streaming [5], Open IPTV Forum (OIPF) extension of 3GPP TC (Technical Specification) [6].

Providing multicast in mobile environment is one of technical challenges for Mobile IPTV. Since 2009, the new IETF working group is chartered to work on the multicast mobility working group [7] in order to provide guidance for supporting multicast in mobile environment. The scope of work is limited to Proxy Mobile IPv6, IGMPv3/MLDv2 protocols and listener mobility. However, the current IPTV system is neither IPv6 address environment nor Mobile IP support. Therefore, multicast mobility is still a technical obstacle in Mobile IPTV service.

Contents adaptation for Mobile IPTV taking the mobile device’s characteristics into consideration is also technical challenges since almost IPTV contents is high quality resolution on Television for better user experience. MPEG developed the scalable video coding technology [7] that lets the IPTV system consider the mobile device types and available bandwidth. Although scalable video coding enables scalable representation of video content with high coding efficiency, it’s difficult to perform real-time encoding because of its encoders’ complexity. Additionally, how to best control the scalable video coding rate according to network resource availability is also technical obstacle.

The major business concern regarding Mobile IPTV is the possibility of low consumer demand for Mobile IPTV viewing on tiny screens. Wide adoption requires a business model for making Mobile IPTV services attractive to users.

User interface is another obstacle to a successful Mobile IPTV business. The small mobile device form hinders development of a fancy user interface. Mobile IPTV growth will require a highly creative and innovative human-machine interface suitable for the mobile device.

Watching live TV while mobile is one of mobile IPTV’s most attractive features. So, access to popular real-time TV programs and rich content should be provided. Content tailored for mobile environments, such as small screen size and random and short watching time, is key.

The last issue on Mobile IPTV in this paper is content. As we can expect, user is not in favor of seeing high-quality and long-length contents on Mobile IPTV device due to its limited wireless bandwidth and capabilities, particularly small screen size. Strictly speaking, 2 hours Hollywood movie does not fit for Mobile IPTV users. Instead, they would highly welcome contents that are suitable for short and occasional viewing. As studied in [8], Mobile TV users spend approximately 20 minutes a day watching contents, although more active users watched between 30 to 40 minutes per session. Mobile TV users also watched contents at different times than traditional TV peak hours. Therefore, how to fit for Mobile IPTV usage as well as user requirements with the current IPTV high-quality contents is also big challenge toward a Mobile IPTV success. Not only traditional broadcasting contents, but web-based video has improved the richness of the user experience today. As prices drop for consumer electronics devices that can create high-quality video, amateurs and professionals alike are producing increasing numbers of high-quality videos. These videos are good resources for Mobile IPTV in accordance with the IPTV commercial contents.

Publishing and interacting with video on the Web in a seamless manner with commonly available browsers is not possible with today’s Web standard. World Wide Web Consortium (W3C) tries to change that to support for improved access to contents and formed a new media fragments working group [9] that provides URI-based mechanisms for uniquely identifying temporal and spatial fragments of media objects in the web, such as video, audio, and images. Temporal addressing enables the referencing of a time or a segment of time in video and audio content, including a normal play time (or time offset), a frame-based time, or an absolute time. It allows the media player to jump to the specified time or frame or to only play a specific segment of the file. RFC 2326 (for the Real-Time Streaming Protocol) defines the notion of normal play time, which is the stream’s absolute position relative to the beginning of the video. On top of that, the Society of Motion Picture and Television Engineers’ time codes define the notion of frame-level accuracy.

3. Concluding Remarks

This paper has tried to describe the current standard activities and technical challenges based on the priority. However, the above items do not cover all of Mobile IPTV issues. Selection of wireless technologies and technical components relies on Mobile IPTV service model provided by a service provider. Nevertheless, the standards activities and technical challenges described in this paper should be deeply considered for enabling Mobile IPTV service, and this service will be expanding the value of the current IPTV service in near future.

References

[1] S. Park and S. Jeong, “Mobile IPTV Approaches, Challenges, Standards, and QoS Support”, *IEEE Internet Computing*, vol. 13, Issue 3, pp. 23-31, May-June 2009.

[2] I. Djama and T. Ahmed, “A Cross-Layer Interworking of DVB-T and WLAN for Mobile IPTV Service Delivery,” IEEE Trans. Broadcasting, vol. 53, no. 1, 2007, pp.382–390.

[3] F.E. Retnasothie, “Wireless IPTV over WiMAX: Challenges and Applications,” Proc. IEEE Annual Conf. Wireless and Microwave Technology (WAMICON 06), IEEE Press, 2006, pp. 1–5.

[4] F. Hartung, “Delivery of Broadcast Service in 3G Networks,” IEEE Trans. Broadcasting, vol. 53, no. 1, 2007, pp. 188–196.

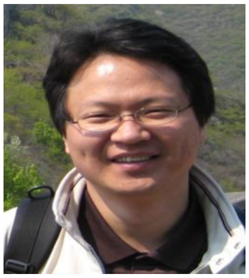
[5] 3GPP, "Transparent end-to-end Packet-switched Streaming Service (PSS) - Protocols and codecs (Release 9)", March 2010.

[6] OIPF, "HTTP Adaptive Streaming (Release 2)", September 2010.

[7] H. Schwarz, D. Marpe, and T. Wiegand, “Overview of the scalable video coding extension of the H.264/AVC standard,” IEEE Trans. Circuits Syst. Video Technol., vol. 17, no. 9, pp. 1103–1120, Sep. 2007.

[8] Carlsson, C. and Walden, P. “Mobile TV - To Live or Die by Content”, Proceedings of the 40th Hawaii International Conference on System Sciences.

[9] W3C Media Fragments Working Group, http://www.w3.org/2008/WebVideo/Fragments/



Soohong Park is is with Samsung Electronics, Digital Media & Communications R&D Center. His research interests include mobility, Internet applications, networking, and Web technologies. He chairs the following standards groups: IETF 16ng Working Group, W3C Media Annotation Working Group, and Mobile IPTV Working Group in Korea Standard Organization. He is currently Samsung representative in World Wide Web Consortium (W3C). He contributed to the IETF over ten years, especially coauthoring of RFC 4029, RFC 4039, RFC 4339, RFC 4968, RFC 5006, RFC 5154, RFC 5270, RFC 5948 and RFC 6160. He graduated from Dankook University, Electronics in 1999, and a Ph.D. candidate in the Department of Computer Engineering at Kyung Hee University. He is a member of IEEE.

Choong Seon Hong received his B.S. and M.S. degrees in electronic engineering from Kyung-Hee University, Seoul, Korea, in 1983, 1985, respectively. In 1988 he joined KT, where he worked on Broadband Networks as a member of the technical staff. From September 1993, he joined Keio University, Japan. He received the Ph.D. degree at Keio University in March 1997. He had worked for the Telecommunications Network Lab, KT as a Senior Member of technical staff and as a Director of the networking research team until August 1999. Since September 1999, he has been working as a Professor of the School of Electronics and Information, Kyung-Hee University. He has served as a Program Committee Member and an Organizing Committee Member for International conferences such as NOMS, IM, APNOMS, E2EMON, CCNC, ADSN, ICPP, DIM, WISA, BcN, and TINA. His research interests include ad hoc networks, network security and network management. He is a member of IEEE, IPSJ, KIPS, KICS, and KIISE.

IPTV Convergence

John Cosmas and Qiang Ni, Brunel University, Uxbridge, UK
{ John.Cosmas, Qiang.Ni }@brunel.ac.uk

1. Current Media Communication Landscape

For the last seventy years, first analog and then digital terrestrial, satellite and cable television (DTT, DST, DCT) in particular the ETSI's DVB-T, DVB-S, DVB-C standardized systems, have been phenomenally successful in delivering standard definition television (SDTV) media to the home. A particularly successful feature of DTT has been its efficient usage of radio spectrum through its configuration not only as multiple-frequency networks (MFN) but also as a single-frequency network (SFN), where a single frequency channel is transmitted over a network of transmitters with simultaneous transmissions from adjacent transmitters are simply treated as echoes at the receivers. Another most useful technology that is common to all DVB standards has been the transport stream since it provides a means for delivering high quality programs over any network because it contains useful time stamp, synchronization and programming data for the audio-visual data. The final most useful technology that is common to all DVB standards has been the conditional access system that requires the receiving terminal to receive continuously changing and unpredictable entitlement messages in order to continue descrambling the program signal thus making it very difficult for unauthorized viewing.

This success has not been repeated for digital mobile television (DMT), such as ETSI's DVB-H standardized system, since companies have been reluctant to invest in DVB-H (Handheld) because 3G cellular networks provide the same kind of service and still has spare capacity and because the only successful mobile television services in the world have been the 'free to air' broadcast to mobile services which have occurred in Japan and Korea.

The only 'free to air' television services in Europe have been DVB-T services but SISO DVB-T transmissions to mobile terminals experience a catastrophic increase in bit error rate unless MISO/SIMO/MIMO cyclic delay diversity (CDD) 00 is utilized. However there are two drawbacks with this approach. The first is that although MISO provides improved performance for mobile receivers that are non-Line of Sight (NLOS) with the transmitter antennas, it provides deteriorated performance for mobile and roof-top receivers that are Line of Sight (LOS) with the transmitter antennas and broadcast network operators will not accept any solution that deteriorates their transmissions to LOS roof-top receiving customers. The second is that MISO requires the investment of additional transmit antennas, power amplifier and CDD unit whilst not providing a significant performance improvement over DVB-H and that SIMO requires the investment of additional antennas and electronics at the receiver.

Recently, the emerging popularity of high definition television (HDTV) has spurred the development of a new generation of DTT, DST and DCT systems in particular ETSI's DVB-T2, DVB-S2 and DVB-C2 standardized systems, which promise to provide an increase in capacity of between 30% - 50%, thereby enabling more HDTV channels to be multiplexed in a single typical 8 MHz channel 0. Of particular interest is DVB-T2 incorporation of MIMO space-time block coding technology, which if implemented by broadcast network operators, may enable the successful mobile reception of 'free to air' television services to NLOS receivers without degrading the reception to LOS receivers because for this particular type of diversity technology the multiple transmitted signals are orthogonal to each other and thus do not interfere with each other.

The commercial requirements of DVB-NGH (Next Generation Handheld) stipulate the coexistence with other wireless communication systems such as cellular telecommunication networks, e.g. 2G/3G/LTE and wireless LAN/MAN e.g. IEE 802.11/802.16 00. The phenomenal success of the 2G cellular networks e.g. GSM/GPRS and the promising prospects of its successor 3G/4G cellular networks e.g. LTE, pose challenges for DVB-NGH particularly since LTE has the ability to configure its radio frequencies both as a cellular MFN and/or SFN and promises to provide up to 100M bits/second shared amongst users in a single coverage area. From the LTE standpoint, coexistence with DVB-NGH may be attractive as a complementary overlay network in areas and times of high congestion where LTE 0 will have limited capacity as its network resources are allocated for point-to-point active telecommunication users whilst coexistence with DVB-SH (Satellite

Handheld) as a complementary overlay network for remote areas since may be attractive as an overlay network with a coverage area that includes far more areas than any terrestrial network.

Wireless LAN/MAN technologies have experienced phenomenal success in the developed world for providing easy connectivity within individual homes and in the developing world for providing cheap connectivity of multiple homes and businesses to the core communication network 000. However the disadvantage of utilizing existing wireless LAN/MAN technologies for the last-hop are that they operate with unregulated spectrum, which renders them vulnerable to usage from others, thereby introducing unwanted congestion and interference.

In recent years, optical fibre network operators have been laying the infrastructure in towns and cities that will connect each of the millions of homes and businesses with 100M bit/second connections, boosting to 1G bit/second. However since it is economically unviable to connect 100% of homes with optical fibre connections, there is a need to use wireless LAN/MAN technologies for bi-directional broadband wireless connectivity to homes in remote locations. The emergence of optical fibre broadband networks will make the continued use of UHF radio spectrum to deliver to stationary applications to the home increasingly untenable, except for establishing wireless connectivity to homes in remote locations.

2. Main Media Consumer Trends

On the Internet, there has been an increasing dominance of streaming, communications and peer-to-peer video applications, and in the future this is projected to grow due to demand for increasingly higher resolution 2D and 3D video for both fixed and mobile end-user terminals 0 0. Furthermore, video clips and programs that are currently being viewed come from professionally generated content (e.g. past TV programs, newspapers, universities, government bodies companies etc.) or from user generated content (e.g. from smart devices capable of capturing digital photographs, video and audio) which is starting to dominate the Internet. *This requires bidirectional broadband communication.*

TV programs at scheduled times will still play an important role in a community's collective cultural experience e.g. News, Sports, Reality TV (X-Factor, Strictly Come Dancing, I'm a Celebrity Get Me out of Here, Big Brother, Unbreakable). *Thus there is still a need to transmit the same program to millions of viewers.*

The influence of Internet applications on society has developed the need for increasingly seamless integration of Internet applications (e.g. Facebook, MySpace, Twitter, YouTube) directly with TV monitor technologies in order to make the viewing experience more communal. *This has been possible because of TVs that are Ethernet enabled thereby providing an easy return channel.*

3D HDTV is an emerging and appealing new media genre which has been stimulated by the film industry with the success of films such as 'Avatar' and 'Alice in Wonderland' and the emerging "to the home" platforms have the potential to provide a popular, immersive entertainment experience that will bring a next generation viewing experience for 3D HDTV programs that go beyond the limited capabilities of the 2D TV systems that are currently available. ITU have defined three increasingly more sophisticated technologies: (1) Stereoscopic 3D, (2) Multiview 3D (3) Object Wave 3D (also known as Integral or Holoscopic 3D) 0, each of which *requires cameras and displays with increasingly larger spatial resolutions.*

3. Future Media Communication Landscape

In section 1, we argued that complementary wireless technologies are necessary both for fixed optical fibre networks due to the requirement for connectivity to homes in remote areas and for cellular networks due to the requirement of universal access at times of high congestion and in place of inaccessible terrestrial coverage. In this section we explore what possible additional technological developments that may be required to support main technological and media consumer trends 0.

3.1 Fixed Network

Increased Capacity and Universal Broadband Access: It may be timely to either introduce a more secure Wireless LAN system operating on licensed spectrum or for broadcasting standardization bodies to develop a secure wireless LAN type version of DVB-T2 that has an uplink bit rate of the same order of magnitude as the down link bit rate, for operation in licensed spectrum to complement optical fibre broadband networks. Strict access protocols are required to avoid congestion and interference and to detect unauthorized use of this spectrum.

3.2 Mobile Network

Increased Mobile Capacity: A new generation of converged mobile networks that Interwork with LTE with DVB-SH and DVB-NGH could provide improved mobile services and applications that takes advantage of these higher rates.

A converged LTE, DVB-SH and DVB-NGH: This should support user mobility where the user is allowed access to any network environment, terminal mobility where the terminal is able to operate in any network environment, network mobility where one (radio) network is able to be connected to any other network and service mobility where a service is accessible on any terminal over any network. This will provide businesses with the opportunity to provide different types of mobility services to people on the move.

Mobile Protocols: The Future Internet should support mobile protocols that support mobility (e.g. for seamless handover of AV services and for the support of mobile sensors and efficient transfer of small data units).

Scalable user Interface and scalable 2D/3D codecs: This should be provided by the future Internet that would permit mobile devices to access wired Internet sites.

Programmable/configurable mobile terminals: This would allow both complimentary and competing radio transceivers and different conditional access systems to coexist on a single phone. Scalable codec and media adaptation methods are required to support the huge variability of mobile phone types in the market.

Improved Addressing: This will result in efficient access to caches, proxies and the support for mobility.

3.3 Fixed and Mobile Network

Ability to Choose Price/Performance: Range of streamed media under the common IP umbrella (e.g. AV over IP, Transport Stream over IP) to facilitate providing the ability to choose price/performance that the customer wants based around Reliability, Resilience, Availability and QoS.

Media Awareness: Access networks at the interface between different parts of the network that sense the IP packets flowing through them and incorporate additional error correction measures on the type of media that is being transmitted and the QoS required by the user 0.

Network Awareness: At the home premises that use smart home appliance that detects appropriate type of network connectivity.

4. Conclusions

It is clear that DVB technologies could have an important role to play in the both fixed and mobile future IPTV services. For fixed networks we have identified the need of a secured Wireless LAN that operates in licensed spectrum and attuned to the need of a broadband home user. For wireless networks we have identified the need for cooperative cellular, terrestrial and satellite broadcast networks.

References

[1] R. Di Bari, M. Bard, Y. Zhang, K.M. Nasr, J. Cosmas, K.K. Loo, R. Nilavalan, and H. Shirazi, K. Krishnapillai, “Laboratory Measurement Campaign of DVB-T Signal with Transmit Delay Diversity” IEEE Transactions on Broadcasting 54(3, Part 2):532-541 2008

[2] R. Di Bari, M. Bard, A. Arrinda, P. Ditto, J. Cosmas, K.K. Loo, and R. Nilavalan, “Measurement Campaign on Transmit Delay Diversity for Mobile DVB-T/H Systems” IEEE Transactions on Broadcasting, Accepted for publication 2010/1

[3] “Understanding DVB-T2 - Key technical, business & regulatory implications”, DigiTAG, 2009.

[4] “Commercial Requirements for DVB-NGH” V1.01, 29 June 2009.

[5] “TM-H NGH Study mission report (Final)” v.1.1, 6 June 2008.

[6] H. Holma and A. Toskala, “LTE for UMTS OFDMA and SC-FDMA Based Radio Access, J. Wiley, 2009.

[7] Q. Ni, "Performance Analysis and Enhancements for IEEE 802.11e Wireless Networks". IEEE Network, Vol. 19, No. 4, July/August, 2005, pp. 21-27.

[8] Q. Ni, L. Romdhani, and T. Turletti, “A Survey of QoS Enhancements for IEEE 802.11 Wireless LAN”. Wiley Journal of Wireless Communications and Mobile Computing, Vol.4, Issue 5, Aug. 2004, pp. 547-566.

[9] Q. Ni, A. Vinel, Y. Xiao, A. Turlikov, and T. Jiang, "Investigation of Bandwidth Request Mechanisms under Point-to-Multipoint Mode of WiMAX Networks". IEEE Communications Magazine, Vol. 45, No. 5, May 2007, pp. 132-138.

[10] Michael Needham and John Harris, “Traffic and Network Modeling for Next Generation Applications”. IEEE International Symposium on Broadband Multimedia Systems and Broadcasting - Multiple Technologies for Multimedia, March 31 – April 2, 2008, Las Vegas, Nevada USA

[11] John Cosmas, Jonathan Loo, Amar Aggoun, and Emmanuel Tsekleves, “A Matlab Traffic and Network Flow Model for planning the impact of 3D Applications on Networks”. IEEE International Symposium on Broadband Multimedia Systems and Broadcasting 2010,

24- 26 March, 2010, Shanghai, China.

[12] Amar Aggoun, Emmanuel Tsekleves, John Cosmas, and Jonathan Loo, "Live Immerse Video-Audio Interactive Multimedia" IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, 2010, 24- 26 March, 2010, Shanghai, China.

[13] John Cosmas, Amar Aggoun, and Emmanuel Tsekleves et al., "Multimedia Delivery in the Future Internet: A Converged Network Perspective" Media Delivery Platforms Cluster White Paper Version: 1.0, October, 2008.

John Cosmas is a Professor in Multimedia System at Brunel University. He leads the Wireless Networks and Communications Research Centre and is the course director for the MSc in Advanced Multimedia Design and 3D Technologies. His research interests are concerned with Multimedia Systems applied to digital television and virtual reality and the synergies between the two. He has participated in eleven EU-IST and two EPSRC funded research projects since 1986 and he has led three of these. He is an associate editor of IEEE Transactions on Broadcasting and has published over 50 refereed papers in journals and over 100 technical papers in conferences around the world.

Qiang Ni is Reader in Wireless Communication and Networks at Wireless Networks and Communications Centre, Brunel University. Prior to joining Brunel University, he was a Senior Researcher with the Hamilton Institute, National University of Ireland Maynooth. He also previously worked with INRIA France as a Researcher (2001–2004). His research interests are in the wireless networking and mobile communications. He was an IEEE 802.11 wireless standard working group Voting Member and a contributor to the IEEE wireless standards. He is a Senior Member of the IEEE Communications Society.

High Resolution IPTV over Broadband Wireless Access Networks

Omneya Issa, Wei Li and Hong Liu, Communications Research Centre, Ottawa (ON),Canada
{omneya.issa, wei.li, hong.liu}@crc.gc.ca

1. Wireless/mobile IPTV

The recent release of broadband wireless access (BWA) technologies has encouraged the extension of IPTV services to new application scenarios involving wireless access and mobility.
The idea is to offer the users various IPTV services anywhere and even on the move.

ITU-T is at the front end of standardization on IPTV including mobile IPTV. As defined by ITU-T, IPTV is access agnostic. Among various wireless access technologies, such as WLAN, WiMAX or cellular networks, WiMAX, based on IEEE 802.16 family standards, emerges as a very promising access network for providing IPTV services. At a fraction of the costs of wired access and cellular networks, it is currently being deployed across the world as a main network infrastructure or as an alternative to cable or DSL lines in rural and underserved areas. While offering much larger coverage than WLAN, it also provides fixed and on-the-move users IP connectivity at a data rate of several megabits per second.

While this type of network is being increasingly deployed, the delivery of high resolution video over IP networks (i.e. SD/HD IPTV) is becoming a reality thanks to advanced video compression technologies and accelerated computing power. These favorable factors, in addition to the affordable broadband access and high data rates offered by 802.16 networks and the fact that it is an IP-based network, may extend high resolution IPTV services in the dimension of wireless and mobility.

Nevertheless, there might be obstacles for successful launch and wide use of wireless/mobile IPTV services. The wireless link still has less bandwidth than the wired line. Also, the wireless channels are exposed to a variety of disturbances due to shadowing and fading. Such channel quality variations are inevitable in wireless networks, thus leading to received video quality degradation. So, there is still work to be done on analyzing how this BWA technology can accommodate high-resolution commercial video and IPTV services.

2. Scenarios

The work done so far on IEEE 802.16-based network mainly concerned the physical and scheduling aspects [1]-[4]. Other studies investigated low resolution video applications over simulated wireless links [5]-[9].

The focus of our work is to study the feasibility of exploiting WiMAX in different scenarios involving high and standard definition TV applications. In particular two types of scenarios were investigated: HD/SD IPTV transmission over downlink, and HD/SD TV streaming on uplink as in camera surveillance and live TV news gathering applications. Fig. 1 shows an overview of these scenarios.

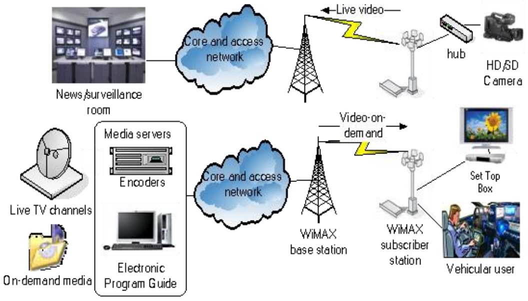


Figure 1. Scenarios of IPTV over BWA networks

In fact, high resolution imagery is important in surveillance and news gathering, where detailed images allow for better display and analysis. To our knowledge, no other work has yet covered the possibilities and limitations of providing high-resolution video and TV in indoor and outdoor environments to/from mobile users.

3. Possibilities and challenges

A testbed involving professional IPTV and video streaming equipment was developed for this purpose. Tests were done on a real 802.16d link. First, we analyzed the quality of high resolution video in fixed indoor conditions. Then, we investigated the outdoor performance in rural-like areas (e.g. university or military campuses).

In fact, video resolution and quality that can be offered to customers mainly depend on the bit rate allocated to the video. The bit rate per user is controlled by many factors including the modulation scheme, which depends on the link condition, and the number of simultaneous connections. We assessed standard and high definition video quality for different modulation schemes and bit rates for both uplink and downlink scenarios. HD (1080i) and SD (480i) broadcast material were H.264-encoded at 5-15 Mbps and 1-4 Mbps respectively. The 5 Mbps HD video quality remains acceptable for many scene types, according to the subjective study in [10]. The feasibility of providing a sustainable TV/video quality was analyzed in different link conditions. Hence, performance limits and guidelines were drawn in order to ensure a viable service in different environments.

In general, when the channel had acceptable conditions, the difference in video quality between test-cases was mainly due to the change of encoding bit rate and encoding parameters to suit the modulation scheme (RF condition). SD video was more tolerant than HD video to higher packet loss rates, experienced near a certain threshold of signal strength, below which packet loss rate started to dramatically increase. Another difference is that, even in low RF conditions where BPSK is the only choice for modulation, SD service may still be provided. In general, it is advised to operate the network a few dBs above the threshold of operation to guarantee a good video quality, especially, when QAM schemes are selected.

Our work revealed important aspects that can considerably influence the business case of delivering IPTV over BWA networks. Among them, is the tradeoff between quality and network capacity. If the operator objective is to maximize the number of subscribers, he would use QAM modulations if allowed by the RF condition. Considering the minimum encoding rate for SD (1 Mbps) and HD (5Mbps), a couple of HD channels or around 10 SD channels could be provided simultaneously. However, if his aim is to insure the service quality in average RF environment, the best option would be to operate in QPSK compromising the number of subscribers to almost half of the number QAM could support.

Most of broadband wireless access equipments offer an adaptive modulation feature. This feature enables the transmitter and receiver to negotiate the highest mutually sustainable data rate (modulation), then dynamically changes the bit rate to adapt to RF conditions. The adaptive modulation feature can help in sustaining IPTV delivery in most of link conditions. However, it must be coupled with H.264 encoders capable of dynamically changing the encoding bit rate to match the bit rate change of the channel. Scalable video coding may also be a good solution in case of software encoders. Video base layer can be transmitted in low bit rate cases while both base and enhanced layers can be received when higher bit rate channel can be afforded. However, our extensive experi-mentation showed that the adaptive modulation feature, when enabled, sometimes has difficulty coping with the rapid signal level change experienced by users in movement for both uplink and downlink communications.

It is also worth noting that H.264-encoded video rates above 8 Mbps are very challenging for on-the-move scenarios in outdoor environments. Although higher bit rates may be achieved, the throughput fluctuation greatly affects the video quality. Nevertheless, the encoding bit rate can go up to 7Mbps, offering good quality high resolution video, in most of the scenarios while maintaining a sustainable service.

Concerning the RF condition, IPTV over BWA performed better in LOS and near LOS conditions. In NLOS outdoor environment, especially on-the-move, users experienced high throughput variations; hence, proper video reception was not possible. However, other types of traffic such as low data rates (e.g. basic web surfing) and non real-time (e.g. TCP data) ones may be sustained under these conditions. When LOS is guaranteed, a video communication can be sustained, even, with occasional packet loss. Thus, the use of relays is needed in order to restore LOS conditions in shed areas. Other techniques such as diversity and smart antennas also become interesting elements to counteract reflected signals and, thus, offering a successful service.

A last aspect that needs to be considered is that the wireless access link is quite different from a classic broadcasting cable (ATSC or DVB). When a user changes the TV channel on his STB, it does not tune a channel like in a cable system, but it switches to another stream. This way, only channels that are currently being watched are actually sent from the local office to users. These interactive and flexible services are gaining more interest of new generations who like to watch their favorite programs when they want. Another difference is that 802.16 BWA equipments are implemented to provide unicast medium by default. A key advantage is individual modulation adaptation that matches user channel condition. However, multicast and broadcast services are recently added to the standard. This feature will be advantageous in areas where users can have more or less the same channel condition, so that no subscriber would be penalized because of high channel attenuation at another subscriber premise.

4. Conclusions and work in progress

Providing high resolution IPTV over BWA networks will open new horizons for surveillance, for news gathering and for users in underserved areas. Our studies have shown a potential of providing a good video quality, satisfying IPTV/video QoE requirements, for last-mile indoor and outdoor deployment scenarios. In particular, use-cases, such as the delivery of IPTV as well as remote surveillance and news gathering services, represented feasible scenarios. However, some issues need to be considered such as, the tradeoff between network capacity and video quality, the adaptive versus fixed modulation modes, and the effect of line-of-sight on the quality of service. Future research should be undertaken to assess quality of high resolution video on a BWA system supporting full vehicular mobility. Also, other interesting features are yet to be explored, such as, QoS classification and traffic prioritization in optimizing service delivery.

References

[1] B. S. Krongold and D. L. Jones, "PAR Reduction in OFDM via Active Constellation Extension," *IEEE Trans. Broadcasting*, vol. 3, pp. 258-268, Sept 2003.

[2] G. Armada, "Understanding the effects of phase noise in orthogonal frequency division multiplexing (OFDM)," *IEEE Trans. Broadcasting*, vol. 47, June 2001, pp. 153-159.

[3] Y. Ben-Shimol, I. Kitroser, and Y. Dinitz, "Two-dimensional mapping for wireless OFDMA systems," *IEEE Trans. Broadcasting*, vol. 52, Issue 3, Sept. 2006 pp. 388 – 396.

[4] M. Ergen, S. Coleri, and P. Varaiya, "QoS aware adaptive resource allocation techniques for fair scheduling in OFDMA based broadband wireless access systems," *IEEE Trans. Broadcasting*, vol. 49, Dec 2003.

[5] Jong Min Lee, H. Park, S. G. Choi, and Jun Kyun Choi, "Adaptive Hybrid Transmission Mechanism for On-Demand Mobile IPTV Over WiMAX," *IEEE Trans. Broadcasting*, vol 55, Issue 2, Part 2, June 2009, pp:468 – 477.

[6] O. Hillested, A. Perkis, V. Genc, S. Murphy, and J. Murphy, "Adaptive H.264/MPEG-4 SVC Video over IEEE 802.16 Broadband Wireless Networks," *Proc. Packet Video*, Nov. 2007, pp. 26-35.

[7] O. Hillested, A. Perkis, V. Genc, S. Murphy, and J. Murphy, "Delivery of On-Demand Video Services in Rural Areas via IEEE 802.16 Broadband Wireless Access Networks," Proc. the 2nd ACM international workshop on Wireless multimedia networking and performance modeling, Spain, 2006, pp. 43-52.

[8] C. Huang, J. Hwang, and D. Chang, "Congestion and Error Control for Layered Scalable Video Multicast over WiMAX, " Proc. IEEE Mobile WiMAX Symposium, 2007, pp. 114-119.

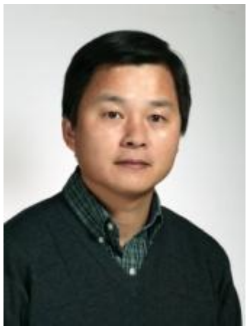
[9] J. She, X. Yu, F. Hou, P. Ho, and E. Yang, "A Framework of Cross-Layer Superposition Coded Multicast for Robust IPTV Services over WiMAX," IEEE Journal on Selected. Areas of Comm., no. 2, vol. 27, Feb. 2009, pp. 235-245.

[10] F. Speranza, A. Vincent, and R. Renaud, "Bit-Rate Efficiency of H.264 Encoders measured with Subjective Assessment Techniques," *IEEE Trans. Broadcasting*, vol. 55, Issue 4, Dec. 2009, pp. 776 – 780.

Omneya Issa received her BSc and MSc, both in Computer Engineering, from Cairo University, and the PhD degree in Telecommunications from the Institut National de la Recherche Scientifique, Montreal, Canada in 2004. From 2004 to 2008, she was a research scientist in International Institute of Telecommunications where she conducted R&D on multimedia applications for wireless and mobile technologies. Since 2008, she has been with the Communications Research Centre (CRC) Canada, where she is currently the team leader of multimedia communications. Her research interests include multimedia communication services, Quality of Service and video optimization for wireless and mobile telecommunications. She is also the BTS representative at the ITU-T IPTV-GSI.

Wei Li is currently a Research Scientist with the Communications Research Centre (CRC) Canada. He received the B.E. degree from Shandong University, the M.S. degree from the University of Science and Technology of China, and the Ph.D. degree from the Institut National des Sciences Appliquées of Rennes in France, all in electrical engineering. In October 2001, He joined the CRC where his major focus is broadband multimedia systems and digital television broadcasting. He was with Motorola Canada Software Centre, Montreal, Canada, from

2000 to 2001, where he conducted R&D in telecommunication networks. From 1998 to 1999, he was with Emulive Imaging in Montreal, Canada. Prior to that, he had worked as a researcher at Sherbrooke University, Canada from 1997 to 1998. He served as session chair for the IEEE International Symposium on Broadband Multimedia Systems and Broadcasting 2006 and 2007. He was the managing editor for the IEEE Transactions on Broadcasting, special issue on IPTV in broadcasting applications in 2009. He also served as reviewer for many renowned international journals and conferences in the area of broadcasting, multimedia communication and multimedia processing. He is the BTS IPTV representative at the ITU-T, and the vice-chair of Internet-enhanced TV planning team of ATSC.



Hong Liu is a research engineer at Communications Research Centre Canada, Ottawa, Canada. He received the B.Sc degree from Nanchang University, Nanchang, China in 1993 and the M.Sc degree from University of Ottawa, Ottawa, Canada in 2001. From 1993 to 1998, he worked as a lecturer in Electrical Engineering department of East China Jiao tong University, Nanchang, China. From 2000 to 2001, he worked in Nortel Networks at Ottawa as a software engineer and Manitoba Telecom Services Inc at Winnipeg as a network planner. He has been involved in the ITU-T IPTV standardization as a representative of the IEEE BTS since 2006. His research interests include video processing and communication, network communication, error control coding and DTV system.

Efficient IPTV Services Delivery using SVC Adaptation and Cooperative Prefetching

Toufik Ahmed, Ubaid Abbasi and Samir Medjiah, University of Bordeaux-1, 351, Cours de la libération 33405, Talence France
{tad, abbasi, medjiah}@labri.fr

1. Introduction

As the demand for Internet-based applications grows around the world, Internet Protocol Television (IPTV) has been becoming very popular. Mobile IPTV allows the users to receive multimedia contents over mobile IP networks using unicast, multicast or peer-to-peer (P2P) communications whenever they want and wherever they are.

Common Mobile IPTV services include Live TV and Video-on-Demand (VoD) streaming. The delivery of such services requires addressing the following two issues related to user Quality of Experience (QoE): (1) ensuring minimum start-up delay during channel switching for Live TV, and (2) guaranteeing minimum delay for seek operations in VoD streaming.

Solving the above-mentioned issues is crucial for the successful deployment of mobile IPTV services. In Live TV, users perform frequent channel switching to discover interesting programs. Similarly, in VoD streaming, the freedom of choosing content leads to a large number of seek operations 0. In both cases, delay reduction is more important than providing the full quality of the video.

In the other hand, to support clients with diverse capabilities and requirements, the MPEG is defining a scalable extension of H.264 AVC namely Scalable Video Coding (SVC) that is able to simultaneously support multiple spatial, temporal and SNR resolutions with minimum processing and transmission overhead 0. The SVC is being attracting great interests in the research community.

An SVC video stream is a collection of atomic data units namely NALU (Network Abstraction Layer Unit). It consists of a single base layer and one or multiple enhancement layers. The base layer is the most important layer and is essential for decoding the other layers while the exclusion of some or all of the enhancement layers still allows a reasonable quality.
The adoption of SVC in streaming services faces two key design challenges: *(1) layer selection*: how many layers each user should receive; and *(2) layer delivery*: how to efficiently deliver those layers.

2. Efficient IPTV Services Delivery

It has been noticed that mobile users nowadays have enough bandwidth to satisfy the reception of desired service as well as other services simultaneously. Based on this assumption, we propose, in this letter, the use of SVC adaptation to mitigate the problem of delay for the two IPTV services (Live and VoD). We propose an architecture that ensures fast channel switching (zap-delay) using SVC for live streaming and fast seek operations using cooperative prefetching 0 of SVC NALUs for VoD streaming.

2.1 Enhancing channel zap-delay for live services

In traditional architectures, the content of each live TV channel is sent through a single multicast group. This way, a user has to subscribe to this multicast group to view this channel. When switching to a different channel, the user has to wait a certain amount of time (ranging from hundreds of milliseconds to seconds) before he can start watching. This delay is required for IGMP *leave* and *join* operations but also needed to fill the buffer with the full quality content of the selected channel. However, in our architecture, a TV channel, as it is encoded into a base layer and several enhancement layers, is sent through at least two different multicast groups. The first group provides the transmission of the SVC base layer, while the second group will be used for all the enhancement layers. According to the remaining download bandwidth, we propose that users prefetch base layer content for fast channel switching while acquiring the full quality content once a channel is selected. Thus, in order to view a channel, the user terminal has to subscribe to the two multicast groups (base layer and enhancement layers). While viewing a channel, the remaining bandwidth is utilized to acquire the base layer NALUs of the popular channels. The popular channel information can be provided by the central entity (the streaming server) or discovered using gossip signaling. When a user performs a channel switching (i.e. zap), the user starts viewing the content with minimum delay due to the advance acquisition of the corresponding base layer (see

Figure 1). Once a channel is selected by a user, the maximum bandwidth is assigned for acquiring the NALUs of this channel (base layer as well as enhancement layers) to enhance the quality of the selected channel.

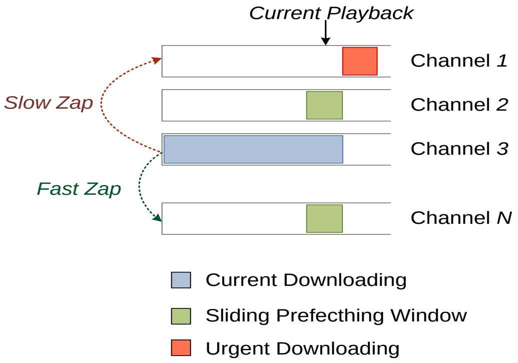


Figure 1. User behavior with respect to Buffer Window for Live TV Streaming

Due to the limited *"extra"* bandwidth used for prefetching the base layer of the other channels, all the channels cannot be prefetched simultaneously and the fast channel switching is only possible for a small number of channels. To enhance this mechanism, a history-based prediction scheme can be used in order to forecast the user zaps. This dynamic scheme aims to prefetch the appropriate channels to meet the dynamicity of the available *extra* bandwidth and the user behavior.

Consequently, this mechanism allows a fast channel switching trading video quality for reduced start-up delay while considering user behavior.

2.2 Enhancing seeks operation for VoD services

One of the existing scenarios for providing VoD streaming in mobile IPTV is based on P2P network. The users are organized into different sessions according to their preferences and their play head positions. Every *x* minutes the server starts a new session to broadcast the video from beginning, where *x* is the predefined session length. This service is also known as near VoD. Initially, when a new user selects a video, it joins a certain streaming session on the basis of its arrival time. A central entity called *tracker* estimates the user available bandwidth and decides whether this user can become the child of the server or a child of another user. Parent peers push the desired content to their child peers. To support VCR operations, the tracker provides the user two lists: *session neighbor list* and *shortcut neighbor list*. The former consists of peers within the same session. This list is used to support a continuous playback of video. The latter contains peers in different sessions and is used to support the playback leap, which occurs due to VCR operations. In order to support fast seek operations, we propose the use of cooperative prefetching 0 among the peers within a VoD session.

In the proposed cooperative prefetching, each peer in a session maintains a buffer map having the state information of available NALUs of different layers (base layer as well as enhancement layers). The peers within a session exchange this state information and create a table of available NALUs within that session. Each peer performs the necessary computation to remove redundancy and creates a list of available non redundant NALUs in the session. The peer then prefetches the desired base layer NALUs (which do not exist in the session) from other peers in other sessions. If it receives no response, these NALUs are requested from the original streaming server. The main advantage of this policy is to reduce the seeking delay by acquiring the maximum NALUs of the base layer. Whenever a peer performs a seek operation (see Figure 2), it can quickly acquire the requested NALUs of the base layer from peers within its session and starts requesting the NALUs of the enhancement layers. As a result, the seeking delay is reduced significantly.

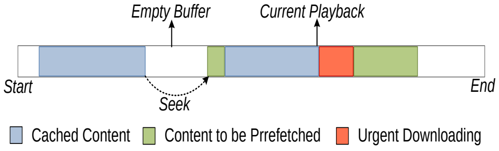


Figure 2. User behavior with respect to Buffer Window for VoD Streaming

3. Conclusion

The use of SVC adaptation allows the support for heterogeneous terminals while overcoming the bandwidth fluctuations which is very common in Internet. Based on SVC, the proposed architecture achieves efficient delivery of mobile IPTV services. In this letter, we proposed the use of SVC adaptation with multicast groups for fast channel switching in Live TV streaming. Moreover, we utilized the cooperative prefetching along with SVC to provide smooth seek operations by minimizing the seek delay in VoD streaming while providing continuous service delivery.

References

[1] H. Yu, D. Zheng, B. Zhao, and W. Zheng, *"Understanding User Behavior in Large-Scale Video-on-Demand Systems"*. In Proc of EuroSys2006, Leuven, Belgium, 2006.

[2] T. C. Thang, J.-G. Kim, J. W. Kang, and J.-J. Yoo, *"SVC adaptation: standard tools and supporting methods,"* EURASIP Signal Processing: Image Communication, vol. 24, no. 3, pp. 214–228, 2009.
[3] Abbasi, U, Ahmed, T*;"COOCHING: Cooperative prefetching strategy for P2P video-on-demand system"*. In Lecture Notes in Computer Sciences, Wired-Wireless Multimedia Networks and Services Management, vol. 5842, pp. 195-200. Springer, Berlin (2009).

Toufik AHMED is currently a professor at ENSEIRB-MATMECA school of engineers in Institut Polytechnique de Bordeaux (IPB) and performing research activities in CNRS-LaBRI Lab-UMR 5800 at University Bordeaux 1. T. Ahmed's main research activities concern Quality of Service (QoS) management and provisioning for multimedia wired and wireless networks, media streaming over P2P network, cross-layer optimization, and end-to-end QoS signaling protocols. T. Ahmed has also worked on a number of national and international projects. He is serving as TPC member for international conferences including IEEE ICC, IEEE GlobeCom, IEEE WCNC. He is currently leading CNRS-LaBRI in EU Project "ENVISION".

Ubaid ABBASI is currently a PhD Student under the supervision of Dr. Toufik Ahmed at the University of Bordeaux-1, France. His main research interests are video streaming in P2P networks and wireless mesh networks.

Samir MEDJIAH is currently a PhD Student under the supervision of Dr. Toufik Ahmed at the University of Bordeaux-1, France. His main research interests are routing, transport and congestion control of video streams in challenged networks including wireless multimedia sensor networks, delay tolerant networks and P2P networks.

Cross-Layer Coding Optimization for Mobile IPTV Delivery

James Ho and En-hui Yang, University of Waterloo, Canada
{james.ho, ehyang}@uwaterloo.ca

1. Introduction

The emergence of broadband wireless access (BWA) technologies such as IEEE 802.16 and Long Term Evolution (LTE) has accelerated the vision to realize mobile real-time multimedia content delivery such as IPTV over wireless medium. However, to enable reliable and efficient video multicasting services over BWA networks, a number of legacy issues need to be thoroughly investigated and addressed.

The first issue of interest is the impact of transmission loss on video quality degradation observed by end-users, which can be quantified as end-to-end distortion (EED), resulting from the combination of source quantization distortion and channel transmission loss. The distortion caused by transmission loss can contribute to significant portions of the total EED perceived at a recipient, especially given the time-varying nature of wireless channels and end-user mobility.

Another issue to be addressed in wireless multicasting is multi-user diversity, which is the result of inevitable channel diversity between multicast recipients from varying distances, channel characteristics, and interference from the transmitter. Channel diversity varies the transmission rates each individual receiver can sustain, and hence, the selection of an effective transmission strategy at the base station becomes a challenge when all recipients are serviced by the same multicast transmission. One solution is to use a very conservative transmission policy or some retransmission or relay approach to enable service continuity. However, these solutions are not efficient since its realization requires additional wasteful resource consumption, impairing the system's economic scalability while increasing system cost and management overhead.

The issue of multi-user channel diversity for multicasting scenarios has been partially tackled by coupling *superposition coding* (SPC) with *scalable video coding* (SVC) to exploit the progressive refinement nature of SVC with the use of a multi-resolution SPC modulation scheme [1]. The proposed IPTV delivery framework in [1] features the multicast of SVC video content encoded into two quality layers. The two layers are superimposed into a two-level SPC multicast transmission to deliver IPTV content. The SPC multicast signal is designed such that most receivers could decode the lower quality layer, while recipients experiencing better channel conditions can obtain both quality layers. To further improve the perceived quality quantified using EED at the recipient, the issue of transmission loss can be partially tackled by applying *multiple description coding* (MDC) [2] at the application layer, which has yet to be considered in reference to SPC multicast.

Using the three aforementioned coding techniques, this letter serves to provide an overview of our research into the establishment of a complete optimization framework in which the integration of SVC, MDC, and SPC can be jointly designed and optimally configured to overcome both the multi-user diversity and transmission loss problems for minimized end-to-end distortion at intended receivers on a coded video broadcast system for mobile IPTV applications.

2. Framework Overview

The joint consideration of SVC, MDC, and SPC yields an overall framework illustrated in Fig. 1. The original video source is first encoded by a video codec into a scalable bitstream characterized by segments corresponding to multiple quality layers. Each quality layer is then encoded using MDC into protected units with added redundancy before modulated and superimposed for a single multicast transmission.

The framework designed in such a manner provides the transmitter with full flexibility in selecting the level of loss protection enabled through MDC, and the appropriate modulation scheme for each scalable video quality layer. The inclusion of multi-resolution modulation realized through SPC further addresses the multi-user channel diversity issue [1]. With each coding technique selected for a specific purpose, the interplay of all three coding schemes can be investigated to provide design insight into system configuration optimization to minimize EED. An expression for EED must thus be formulated to include all aspects of the proposed framework.

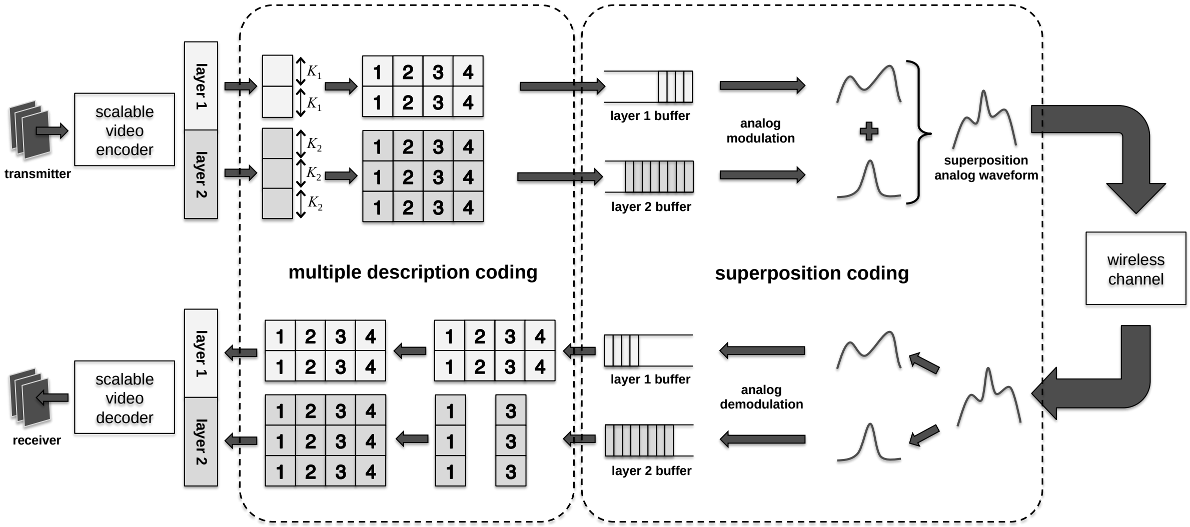


Figure 1. Overview of interplay between SVC, MDC, and SPC in an overall framework for two layers.

3. EED Formulation

Although previous research endeavors and simulations have demonstrated the potential of cross-layer coding design [3], [4] for minimizing EED, they remain focused on the asymptotic behavior in an idealistic system setting, where a Gaussian source is progressively encoded and transmitted with the compression rate at each layer perfectly matched with the capacity of the corresponding channel realization. Such results establish the potential advantages of such schemes, but fail to provide any design principles on practical systems where losses due to characteristics of the wireless channel frequently occur. To gain further design insights into the interplay between SVC, MDC, and SPC, the derivation of key formulations are necessary to study the effects of each coding scheme and provide design guidelines for joint coding optimization in realizing minimized EED.

Previous work in such an area derived a closed-form EED expression in a point-to-point tandem source-channel coding system [5], [6]:

$$\overline{D} = \left(1 - \frac{N}{N-1} p_{err}\right) D_Q + \frac{N}{N-1} \sigma^2 + \frac{N}{N-1} S_Q, \quad (1)$$

where σ^2 is the source variance, p_{err} is the average channel symbol error probability, D_Q is the conventional quantization distortion of the source into N codewords, and S_Q is the scatter factor, a newly discovered parameter representing the distances of the codewords from the source mean. Although the above equation is applicable for any source and any coded or uncoded channel given random source-channel index assignment, it remains limited to single-resolution source coding for point-to-point communications. Since multi-resolution source codes are more structured than single-resolution codes, there is less flexibility in the codeword selection of each resolution. Furthermore, transmission losses in each resolution layer strongly affect the end-user perceived quality due to the strong dependencies between resolution layers in scalable video coding. To address these shortcomings, Eq. (1) has been further extended in [7] to include multi-resolution source codes. However, further extensions are necessary to generalize the EED expression to include channel characteristics of the SPC multicast and also, the impact of MDC in combating transmission loss for each individual layer.

4. Framework System Configuration

With an EED formulation incorporating all aspects of the proposed framework for mobile IPTV multicasting, the interplay between SVC, MDC, and SPC can be studied to obtain results serving to deepen the comprehension into the joint design of each coding technique for minimized EED. Each coding scheme consists of numerous parameters that require both design and configuration.

The *scalable video coding* portion of the framework first requires the development of quantization techniques for these general multi-resolution sources codes, which are subject to inter-layer decoding dependencies that must be accounted for since each layer is subject to different channel symbol errors. The *multiple description coding* portion of the system, realized using Reed-Solomon error correcting codes to combat transmission losses, require configuration of the (N, K) parameters for each resolution layer to minimize transmission losses of the layered broadcast system. Finally, the employment of *superposition coding* (SPC) to generate multi-resolution broadcast signals requires appropriate modulation scheme selection and inter-layer power allocation.

Each of the design components outlined above can be formulated into an optimization problem for the minimization of the long-term end-to-end distortion (EED) for the typical receiver among a large group of receivers, with multi-resolution quantization parameters, MDC redundancy configuration, and inter-layer power allocation as variables. Furthermore, since live IPTV broadcasting is a real-time application, efficient and fast algorithms are essential to achieve and maintain the optimal joint configuration for all three design aspects to quickly adapt to the rapidly changing wireless channel conditions for mobile IPTV subscribers.

5. Conclusions

This letter briefly summarizes a framework for mobile IPTV delivery based on the joint coding optimization of a number of cross-layer coding techniques. The proposed framework addresses the issues of multi-user diversity and transmission loss in wireless multicast multimedia applications. Using the practical measure of end-to-end distortion to gauge the quality experienced by the end-user in such environments, a number of design parameters can be optimally configured to facilitate the efficient delivery of multimedia content over broadband wireless access networks. Furthermore, the behavior of the system can generate insight toward the convergence of future wireless video multicast technologies, fostering ubiquitous access to multimedia information regardless of location and mobility.

References

[1] J. She, F. Hou, P.-H. Ho, and L.-L. Xie, “IPTV over WiMAX: Key Success Factors, Challenges and Solutions,” *IEEE Commun. Mag.*, vol. 45, no 8, pp. 87-93, Aug. 2007.

[2] R. Venkataramani, G. Kramer, and V. K. Goyal, “Multiple description coding with many channels,” *IEEE Trans. Info. Theory*, vol. 49, pp. 2106-2114, Sept. 2002.

[3] J. She, X. Yu, P.–H. Ho, and E.–H. Yang, “A Cross-Layer Design Framework for Robust IPTV Services over IEEE 802.16 Networks,” *IEEE Journal of Selected Areas on Commun. (JSAC)*, vol. 27, no. 2, pp. 235-245, Feb. 2009.

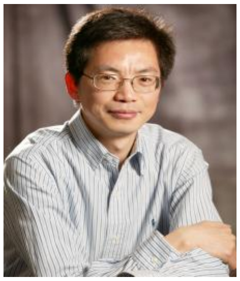
[4] C. T. K. Ng, D. Gunduz, A. J. Goldsmith, and E. Erkip, “Distortion Minimization in Gaussian Layered Broadcast Coding with Successive Refinement,” *IEEE Trans. Info. Theory*, vol. 55, no. 11, pp. 5074-5086, Nov. 2009.

[5] X. Yu, H. Wang, and E.-H. Yang, “Optimal quantization for noisy channels with random index assignment,” *Proc. IEEE Intern. Symp. Info. Theory*, Toronto, Canada, Jul. 2008. pp. 2727-2731.

[6] X. Yu, H. Wang, and E.-H. Yang, “Design and analysis of optimal noisy channel quantization with random index assignment," *IEEE Trans. Info Theory,* vol. 56, no. 11, pp. 5796-5804, Nov. 2010.

[7] F. Teng, E.-H. Yang, and X. Yu, “Optimal Multiresolution Quantization for Broadcast Channels with Random Index Assignment,” *Proc. IEEE Intern. Symp. Info. Theory*, Austin, TX, Jun. 2010. pp. 181-185.

James Ho received his BASc degree in Electrical Engineering with the award of Dean’s Honours List from the University of Waterloo, Ontario, Canada in 2008. In the following year, he completed his MASc degree and is currently pursuing PhD studies investigating wireless multimedia delivery from both coding theory and networking perspectives.



En-hui Yang has been with the Department of Electrical and Computer Engineering, University of Waterloo, Ontario, Canada since June 1997, where he is currently a Professor and Canada Research Chair in information theory and multimedia compression. He is a recipient of several research awards and a Fellow of IEEE, the Canadian Academy of Engineering, and the Royal Society of Canada. He currently serves as an Associate Editor for IEEE Transactions on Information Theory (IT) and is sitting on the Awards Committee for IT. Dr. Yang is also the founding Director of the Leitch-University of Waterloo Multimedia Communication Lab, and a Co-Founder of SlipStream Data Inc., which is now a subsidiary of Research In Motion.

QoE-Aware P2P Video-on-Demand using Scalable Video Coding and Adaptive Server Allocation

Osama Abboud, Konstantin Pussep, Ralf Steinmetz, Multimedia Communications Lab,
Technische Universität Darmstadt
{abboud,pussep,steinmetz}@kom.tu-darmstadt.de

Thomas Zinner, Chair of Communication Networks, University of Würzburg
zinner@informatik.uni-wuerzburg.de

1. Introduction

P2P techniques for streaming are becoming more popular since they allow for higher revenues for content providers, better load distribution, and scalability. Additionally, Internet devices that are joining typical streaming systems are becoming more heterogeneous. In addition, the extra level of mobility allows for many possible network connections that have different connection characteristics. All of this poses stringent requirements on any streaming system to provide support for many kinds of devices and connection characteristics.

A natural method to support these requirements is through the usage of Scalable Video Coding (SVC) [12]. SVC refers to the ability of having a flexible video bit stream in which the video can be adapted to system dynamics.

Our research aims at providing an insight on how SVC can be used in a P2P context to provide quality adaptation to network and device dynamics. Additionally, we use a layer decision controlled by a Quality-of-Experience (QoE)-aware algorithm, which assures the best possible user experience. Additionally, we consider the utilization of SVC in commercial scenarios, which results in a hybrid structure where peer-assistance is used to reduce the load on content servers. Our design and simulative studies have been conducted to proof the feasibility of our approach, cf. [3, 11, 6, 10].

2. Scalable Video Coding and QoE Control

The video codec H.264/SVC [12, 9] is based on H.264/AVC, a video codec widely used in the Internet The SVC extension enables the encoding of a video file at different quality levels within the same layered bit stream. This includes besides different resolutions also different frame-rates and different image qualities. These three dimensions are denoted to as spatial, temporal and quality scalability. Figure 1 gives an example of different possible scalabilities for a video file. The left bottom *subcube* is the base layer, which is necessary to play the video file, here with CIF resolution, 15 Hz frame-rate, and quality Q0. Based on this layer, different additional enhancement layers permit a better video experience with a higher resolution, better image quality or higher frame rate, respectively.

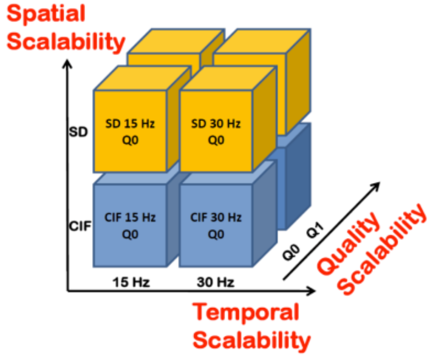


Figure 1. SVC cube, illustrating the possible scalability dimensions for a video file.

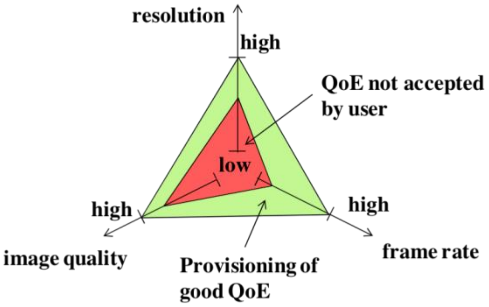


Figure 2. Acceptable area of QoE control settings.

Since SVC provides three different scalabilities, the question arises which layers should be downloaded if the overlay cannot provide enough capacity for the complete video file. At least a minimum resolution, quality and frame rate, which also depends on the user's context, has to be provided. Otherwise the user will not accept the video service. Higher layers will increase the QoE further, but also require a higher bandwidth.

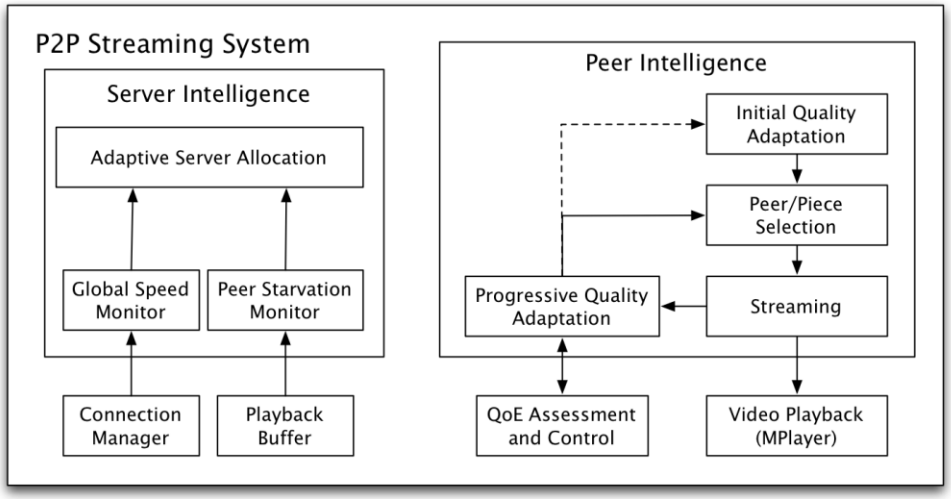


Figure 3. Quality adaptive P2P streaming system architecture with adaptive server allocation

The impact of the different scalabilities on the QoE is depicted in Figure 2. In general, spatial-/quality scalability enables bandwidth reduction by having a minor impact on the QoE, while temporal scalability enables only a minor bandwidth reduction as discussed in [13]. Insights achieved in the mentioned study were used to build a P2P streaming system that enables quality adaptation

3. Quality Adaptive P2P Streaming

The QoE-aware P2P streaming system is already detailed in [3, 6, 4, 11]. The basic architecture of our system, depicted in Figure 3, is divided into two parts: *Server Side Intelligence* and *Peer Side Intelligence*.

Within the *Peer Side Intelligence*, adaptation algorithms that run at peer were developed. These algorithms are based on the idea of dividing quality adaptation, or layer selection, into two stages. The first stage, called *Initial Quality Adaptation*, allows adapting to static resources at the peers, e.g. screening resolution, bandwidth, and processing power. After the initial layer has been selected, peers that can provide the selected layer are contacted and required pieces are requested. Our system further employs a closed loop adaptation algorithm, called *Progressive Quality Adaptation* that constantly monitors playback performance and throughput and will change the selected layer accordingly. Here this module relies heavily on the *QoE Assessment and Control* module that, based on investigations in [13], takes QoE into consideration. For example, this module gives indications on best layer combinations and layer switching frequency that do not deteriorate QoE. For the actual playback, we use a modified version of the MPlayer [2] that supports SVC [7].

Within the *Server Side Intelligence*, the main idea is to allocate servers in such a way that they provide the required resources to the system but avoid over-provisioning. One possible scenario is the utilization of servers on demand, e.g. by using Amazon Elastic Compute Cloud (EC2) [1] or similar services. Since the content provider pays for online time and allocated server bandwidth, minimizing these resources is a necessity. We encounter this issue by monitoring both the global download speed and the playback buffer states of single peers. After that, the system determines the required server resources and allocates them according to the current demand despite the continuous fluctuations of the throughput and overlay population. Note that the server side intelligence is used to address the global demand for content resulting from the heterogeneous peer resources, bandwidth asymmetry at the peer level, and fluctuating content popularity, while the client side intelligence tackles the issues with the throughput and performance of individual peers.

4. Conclusions

P2P streaming has attracted much attention recently with promises for higher revenues and better load distribution. Still, the majority of P2P video streaming systems today employ the one-size-fits-all concept where the same video bit-rate is offered to all users. To this end, the promising H.264/Scalable Video Coding (SVC) standard is seen as a necessity in not only supporting heterogeneous resources, but also in reducing the impact of P2P dynamics on the perceived Quality-of-Experience (QoE). In this paper we have presented our streaming system that uses SVC to adapt to different user requirements and resources. Our system employs a novel QoE-aware layer selection algorithm that maximizes flexibility through SVC while taking impact on QoE into consideration. Furthermore, the system embeds the support for commercial scenarios by enhanced usage of server resources. Such adaptation techniques are necessary for next generation streaming systems. We have already started implementing a prototype, which enables us to assess the impact on QoE while having real users [5]. The prototype is currently being tested within the German Lab testbed [8].

5. Acknowledgments

This work was funded by the Federal Ministry of Education and Research of the Federal Republic of Germany (support code 01 BK 0806, G-Lab).

References

[1]Amazon Elastic Compute Cloud (EC2). http://aws.amazon.com/ec2/.

[2] MPlayer : http://www.mplayerhq.hu/.

[3]Osama Abboud, Konstantin Pussep, Aleksandra Kovacevic, and Ralf Steinmetz. Quality Adaptive Peer-to-Peer Streaming using Scalable Video Coding. *Proceedings of the 12th IFIP/IEEE International Conference on Management of Multimedia and Mobile Networks and Services, MMNS09*, pages 41–54, Oct 2009.

[4]Osama Abboud, Konstantin Pussep, Markus Müller, Aleksandra Kovacevic, and Ralf Steinmetz. Advanced prefetching and upload strategies for p2p video-on-demand. *ACM Workshop on Advanced video streaming techniques for peer-to-peer networks and social networking*, Oct 2010.

[5]Osama Abboud, Thomas Zinner, Konstantin Pussep, Simon Oechsner, Ralf Steinmetz, and Phuoc Tran-Gia. A qoe-aware p2p streaming system using scalable video coding. pages 1–2. IEEE Computer Society Press, Aug 2010.

[6]Osama Abboud, Thomas Zinner, Konstantin Pussep, and Ralf Steinmetz. On the Impact of Quality Adaptation in SVC-based P2P Video-on-Demand. In *2nd ACM Multimedia Systems Conference 2011 - ACM MMSys 2011 (To Appear)*, San Jose, USA, February 2011.

[7]Mederic Blestel and Mickael Raulet. Open SVC Decoder. Available online: http://sourceforge.net/projects/opensvcdecoder/.

[8]G-Lab. German National Platform for Future Internet Studies. Available at http://www.german-lab.de//.

[9]ITU-T Rec. & ISO/IEC 14496-10. Advanced Video Coding for Generic Audiovisual Services.

[10]Simon Oechsner, Thomas Zinner, Jochen Prokopetz, and Tobias Hoßfeld. Supporting Scalable Video Codecs in a P2P Video-on-Demand Streaming System. In *21th ITC Specialist Seminar on Multimedia Applications - Traffic, Performance and QoE*, Miyazaki, Jap, March 2010.

[11]Konstantin Pussep, Osama Abboud, Florian Gerlach, Ralf Steinmetz, and Thorsten Strufe. Adaptive Server Allocation for Peer-assisted VoD. In *International Parallel and Distributed Processing Symposium (IPDPS)*. IEEE, 2010.

[12]Heiko Schwarz, Detlev Marpe, and Thomas Wiegand. Overview of the Scalable Video Coding Extension of the H.264/AVC Standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(9):1103–1129, 2007.

[13]Thomas Zinner, Oliver Hohlfeld, Osama Abboud, and Tobias Hoßfeld. Impact of Frame Rate and Resolution on Objective QoE Metrics. In *Quality of Multimedia Experience, 2009. QoMEx 2009. International Workshop on*, Trondheim, June 2010.

Osama Abboud is a final year PhD student at the Technische Universität Darmstadt, Germany. His work has focused on next generation p2p multimedia systems that harness scalable video coding and intelligent multimedia-aware networks. He has a Master's degree from Universität Ulm, Germany. For more details, see http://www.kom.tu-darmstadt.de/en/people/staff/osama-abboud

Konstantin Pussep is a final year PhD student at the Technische Universität Darmstadt. In his work he has dealt with peer-assisted content distribution, focusing on the video-on-demand streaming. His research interests include the optimized utilization of overlay resources, network-aware traffic management, and energy efficiency of end-user devices. He has a Diplom degree from Technische Universität Darmstadt. For more details, see http://www.kom.tu-darmstadt.de/en/people/staff/konstantin-pussep

Dr.-Ing. Ralf Steinmetz is professor of Multimedia Communications at the Technische Universität Darmstadt, in Germany. Together with more than 30 researchers, he has been working towards the v ision of real "seamless multimedia communications". He has contributed to over 400 refereed publications, become an ICCC Governor; and is a Fellow of both the IEEE and ACM. Professor Dr. Ralf Steinmetz is a member of the Scientific Council and president of the Board of Trustees of the international research institute: IMDEA Networks.

For more details, see http://www.kom.tu-darmstadt.de/en/people/staff/ralf-steinmetz

Thomas Zinner studied computer science and physics at the University of Wuerzburg, Germany. He received his diploma degree in computer science in 2007. Since then he has been a researcher at the Institute of Computer Science and pursuing his PhD. His current research focuses on Quality of Experience for Video Streaming - especially scalable video codecs - in combination with performance evaluation and network virtualization techniques.

E-Letter Editorial Board

MMTC Officers