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

R-LETTER

IEEE COMMUNICATIONS SOCIETY

Vol. 3, No. 4, November 2012

CONTENTS

Message from R-Letter Directors

Welcome to the IEEE MMTC Review-Letter (R-Letter) 2012 November issue. The R-Letter Editorial Board is committed to continue the excellent work of the previous Board led by Dr. Guan-Ming Su and Dr. Nabil J. Sarhan. With the support of the MMTC community, eleven issues have been published since the launch of the R-Letter platform in October 2010.

Research publication databases have been expanding in both size and dimension, making choice difficult without consuming considerable reading time. The MMTC R-Letter Editorial Board is dedicated to assist by providing a selection of high quality publications for your quick review, which can serve as a convenient link to a network of related work. In order to establish such effective mechanism, we solicit your support to nominate high quality publications to the R-Letter Editorial Board.

.

In this issue, we present nine papers, discussing a wide range of multimedia related issues which include database creation, data delivery and quality assessment. The first paper, published in the *IEEE Multimedia Magazine*, discusses the difficulty in collecting facial expressions and proposes capturing the data from movies. The second paper, in *IEEE ICME 2012*, exploits random network coding to support better delivery and understanding of network conditions. The third paper, from the *IEEE Transactions on Wireless Communications*, studies Energy Harvesting for solving the battery problem and limited lifetime of wireless networks. The fourth paper, from *IEEE ICME 2012*, investigates concealing errors in videos based on the visual attentiveness of a human observer. The fifth paper, published in the *IEEE Journal on Selected Topics in Signal Processing*, discusses the evaluation of 3D video quality. The sixth paper, published in *IEEE ICME 2012*, proposes a scene segmentation technique for electronic navigation systems to assist blind people to better understand the environment. The seventh paper, published in *IEEE Packet Video Workshop*, presents a pDASH system to support peer-assisted streaming. The eighth paper, published in *IEEE Communications Magazine*, examines the importance of offering high quality content with guaranteed Quality of Experience and the monitoring of IPTV networks. The last paper, from the *IEEE Journal on Selected Areas in Communications*, discusses the tradeoff between reliability and throughput for multi-user cooperative networks.

We would like to thank all the R-Letter authors and editors who contributed to the completion of this issue. Their timely efforts given the tight deadline is sincerely appreciated.

IEEE ComSoc MMTC R-Letter

Director,
Irene Cheng, University of Alberta, Canada
Email: locheng@ualberta.ca

Co-Director
Xianbin Wang, University of Western Ontario, Canada
Email: xianbin.wang@uwo.ca

A Benchmark for Experimental Validation of Facial Expression Analysis Methods

A short review for "Collecting Large, Richly Annotated Facial-Expression Databases from Movies"

A Dhall, R. Goecke, S. Lucey, and T. Gedeon, " Collecting Large, Richly Annotated Facial-Expression Databases from Movies," IEEE Multimedia Magazine, vol. 19, vo. 3, pages 34-41, July-Sept. 2012.

Facial expression is closely tied to the emotion and intention of individuals. It is an effective way of non-verbal social communication. Automatic facial expression analysis has been an active research topic in the multimedia, computer vision, and human-computer interactions communities. For example, the First Facial Expression Recognition and Analysis Challenge, which was held in conjunction with the 2011 IEEE International Conference on Face and Gesture Recognition, has attracted 16 submissions and more than 90 attendees [1].

Research in facial expression is heavily relied on experimental datasets. These are normally collected in the form of posed or spontaneous expressions, depending on whether human subjects are asked to generate artificial expressions or not. Psychological studies have shown that these two types of expressions differ substantially [2]. Analysis of spontaneous expressions is preferred in research due to its realistic characteristics. However, collection of spontaneous expression is a non-trivial task due to several reasons. Firstly, it is difficult to capture such expressions because when subjects are aware that they are being observed, their facial expressions may deviate from the genuine representation [3]. Secondly, the collection of large database is time-consuming and expensive. It is difficult to gather subjects in various ages and genders, and capture data in various illumination and environmental conditions. Therefore, most existing databases are manually collected in controlled lab environment, with limited poses and expressions from tens or a few hundred subjects. Thirdly, labeling of facial expressions is not straightforward. This is not only caused by the occlusions in the scene, but also due to common muscle movement shared by some expressions, such as anger and disgust [2].

To address these difficulties, in this paper, the authors propose a semi-automatic method to collect and annotate facial expressions from movies. The data collection process starts from video subtitle extraction. This allows information about emotion, actors, scene, and time stamps be retrieved. A recommender system is then used to search the subtitle, and recommend to labelers only those clips with high probability of containing meaningful expressions from a subject, so that the labeler can input dense information related to the expression and subject.

The authors have generated two databases: Acted Facial Expressions in the Wild (AFEW) and Static Facial Expressions in the Wild (SFEW). The former consists of 1,486 short video clips with visible presence of subjects and their faces, while the latter contains a subset of static facial expressions with 1476 images. Both databases cover expressions in seven categories: anger, disgust, fear, happiness, sadness, surprise, and neutral.

As pointed out by the authors, there are several advantages of generating databases using this method [4]. Most importantly, there are plenty of movie data available, which contain large amount of facial expressions and natural head pose movements in settings that are close to real-world environments. Furthermore, data can be collected from actors in various race, gender, and ages. The professional training of these actors allows them to mimic the real-world human behavior, which looks more like spontaneous type other than intentional posing. The recommender system can greatly facilitate the annotation step by only suggesting clips with expression related keywords.

Contribution of this paper also comes from definition of six experimentation protocols based on the level of person dependency present in the datasets. These include Strictly Person Specific (SPS), Partial Person Independent (PPI), and Strictly Person Independent (SPI). The authors have computed baselines in classifying seven expression categories on the three protocols over AFEW and SFEW databases. In these experiments, PHOG and LPQ features were extracted from cropped faces, with PCA being used to reduce the dimensionality of concatenated feature vector. Then a nonlinear SVM was used to learn and predict facial expressions. This baseline method has achieved only 26.3% average classification accuracy for the PPI protocol on the AFEW database. The results also suggest that AFEW database is much more difficult than the CK+ database [5].

In conclusion, facial expression research has long been hindered by the lack of access to databases with real-world settings. The construction of AFEW and SFEW databases has partly solved this problem. This work has the potential of pushing forward the facial expression research by providing a benchmark for experimental validation of various methods. Further information of these two databases is available at: http://cs.anu.edu.au/few.

References:

[1] http://sspnet.eu/fera2011/

[2] P. Ekman and E. Rosenberg, *What the face reveals: basic and applied studies of spontaneous expression using the facial action coding system (FACS)*, Oxford University Press, 1998.

[3] V. Bettadapura, "Face expression recognition and analysis: the state of the art", *Technical Report, College of Computing, Georgia Institute of Technology*, arXiv:1203.6722, April 2012.

[4] A Dhall, R. Goecke, S. Lucey, and T. Gedeon, "Collecting large, richly annotated facial-expression databases from movies," *IEEE Multimedia Magazine*, vol. 19, vo. 3, pages 34-41, July-Sept. 2012.

[5] P. Lucey, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," In *CVPR Workshop'10*, pp.94-101, 2010.

R-Letter Editor:

Jun Zhou received the B.S. degree in computer science and the B.E. degree in international business from Nanjing University of Science and Technology, China, in 1996 and 1998, respectively. He received the M.S. degree in computer science from Concordia University, Canada, in 2002, and the Ph.D. degree in computing science from University of Alberta, Canada, in 2006.

He joined the School of Information and Communication Technology in Griffith University as a lecturer in June 2012. Prior to this appointment, he had been a research fellow in the Australian National University, and a researcher at NICTA. His research interests are in statistical pattern recognition, interactive computer vision, and their applications to hyperspectral imaging and environmental informatics.

Applying Random Network Coding for Improved Multimedia Delivery in LTE-Advanced Networks

A short review for "Random Network Coding for Multimedia Delivery over LTE-Advanced"

D. Vukobratovic, C. Khirallah, V. Stankovic, and J. Thompson, "Random Network Coding for Multimedia Delivery over LTE-Advanced," in Proceedings of the IEEE International Conference on Multimedia and Expo, July 2012.

The 3GPP Long-Term Evolution - Advanced (LTE-A) architecture aims at increasing network capacity and improving service quality by shifting the network design from the macro-cellular architecture to Heterogeneous Networks (HetNets) [1]. This is obtained by introducing closer-to-user small cells on top of the macro-cellular layout. In doing so it allows better prediction of the wireless channel conditions at the expense of increased cell density and inter-cell interference. To exploit this architecture, cooperation and coordination among small cells is necessary. This is well explored in the physical layer through a number of Coordinated Multi-Point (CoMP) techniques [2]. However, upper layer design is still largely similar to traditional setups.

The work in [3] has addressed this problem by introducing a Random Network Coding (RNC) sublayer within the Media Access Control (MAC) layer of the LTE/LTE-A Radio Access Network (RAN) protocol stack. This MAC-RNC protocol sits on top of the MAC layer indicated as a suitable position for RNC message processing. The target is to simplify delivery of upper layer messages within complex and dynamic topologies of evolved LTE-A RAN based on HetNets. Namely, even though the traditional MAC-HARQ (Hybrid Automatic Repeat Request) protocol efficiently handles message transmission over point-to-point wireless links, its capabilities to exploit multi-point and multi-hop RAN topologies are limited. On the other hand, by encapsulating fixed-length RNC-coded symbols derived from upper layer source messages into physical layer containers (transport blocks), and by exchanging these containers among HetNets nodes, more flexible and further efficient multi-point and multi-hop message delivery from the base station to any set of users within the cell becomes possible [3].

Introducing of MAC-RNC into the LTE-A RAN protocol stack does not only provide simpler and more flexible message delivery, but offers also a strong potential for efficient 3GPP multimedia delivery services deployment. Recent projections estimate a 18-fold increase in mobile data traffic between 2011-2016, with more than 70% accounting for mobile video services in 2016 [4]. Therefore the shift of 4G mobile networks towards extensive delivery of multimedia-based services demands significant efforts in adaptation and redesign of the system to multimedia traffic characteristics and requirements. A number of recent studies show feasibility of practical deployments of application-layer RNC-based multimedia streaming on latest generation smartphones [5]. In addition, recombining and sharing collaboratively the RNC-coded multimedia packets among mobile terminals and/or relay nodes can increase throughput by exploiting path diversity, cooperation and overhearing [6].

This suggests that combining multimedia streaming and RNC-based transmission improves its delivery. Based on [3], the authors of this reviewed paper explore the possible impact of the MAC-RNC solution on multimedia service delivery over LTE/LTE-A. They indicate the following benefits for using the protocol within the LTE-A RAN for multimedia traffic:

1) It introduces redundancy and protects data delivery only across the RAN, eliminating waste of resource due to forward error correction schemes in the application layer, which add end-to-end redundancy.
2) It exploits path, frequency (resource allocation) and time (TTI allocation) diversity within 4G HetNets to deliver upper layer messages to the desired user. This contrasts with current schemes that rely only on time diversity by exploiting long-length packet-level codes.
3) It performs as optimal short-length rateless codes over higher finite-fields. The amount of redundancy introduced by MAC-RNC is close to the minimum required for dynamic wireless channel conditions.

4) It reduces the number of feedback messages to a single ACK per upper layer message.
5) Coded packets may be further exchanged collaboratively by collocated User Equipment (UE) or re-encoded and forwarded by intermediate nodes in LTE-A multi-hop relaying or HetNets topologies to improve the media delivery process.
6) UE can send feedback related to the number of received linearly independent encoded packets. This can be exploited by MAC scheduler to better allocate resources on upcoming transmission time-intervals (TTIs) to match the number of remaining encoded packets UE needs to send to complete the reception of the upper layer message.
7) In MAC-RNC framework, the upper layer message is not segmented to match the size of the physical layer transport block. Instead, Random Linear Coding (RLC) messages can be produced directly from IP encapsulated video packets. Thus, the RLC layer may exploit content-awareness that could be enabled by minimal additional cross-layer interaction with the application layer video coding process.

The authors show that this MAC-RNC solution provides a simple and efficient RAN-wide rateless/network coding MAC sublayer scheme for reliable delivery of RLC encapsulated IP packets. Scheduling and resource allocation procedures have also been simplified while addressing the transmission process goals. Overall, the main impact on the video delivery is identified at the RLC layer through content awareness during IP packetized video encapsulation; at the MAC scheduler for explicit control over the number of encoded packets delivered; at the MAC-RNC mechanism itself through application of unequal error protection RNC solutions, and through collaborative network coded packet exchange in evolved LTE-A multi-hop RANs.

Random network coding provides a solution for better delivery of file sharing and multimedia applications. Further exploitation of RNC together with cross-layer solutions can help in providing the tools for the network to deal with the ever increasing traffic. Moreover, green networking design is needed to reduce the energy per bit required to transmit the huge multimedia generated traffic while maintaining low latency.

References:

[1] A. Khandekar, N. Bhushan, J. Tingfang, V. Vanghi, "LTE-Advanced: Heterogeneous Networks," European Wireless EW 2010, pp. 978-982, Lucca, Italy, April 2010.

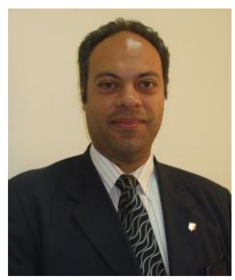
[2] M. Sawahashi, Y. Kishiyama, A. Morimoto, D. Nishikawa, M. Tanno, "Coordinated Multipoint Transmission/Reception Techniques for LTE-Advanced," IEEE Wireless Communications, vol.17, no.3, pp.26–34, June 2010.

[3] C. Khirallah, D. Vukobratovic and J. Thompson, "Performance Evaluation and Energy Efficiency of Random Network Coding in LTE-Advanced," IEEE ICC 2012, Ottawa, Canada, June 2012.

[4] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2011-2016, http://www.cisco.com/

[5] H. Shojania, B. Li, "Random Network Coding on the iPhone: Fact or Fiction?," ACM NOSSDAV 2009, Williamsburg, USA, June 2009.

[6] H. Seferoglu, L. Keller, B. Cici, A. Le and A. Markopoulou, "Cooperative Video Streaming on Smartphones," Annual Allerton Conference 2011, USA, September 2011.



R-Letter Editor:

Carl James Debono (S'97, M'01, SM'07) received his B.Eng. (Hons.) degree in Electrical Engineering from the University of Malta, Malta, in 1997 and the Ph.D. degree in Electronics and Computer Engineering from the University of Pavia, Italy, in 2000.

Between 1997 and 2001 he was employed as a Research Engineer in the area of Integrated Circuit Design with the Department of Microelectronics at the University of Malta. In 2000 he was also engaged as a Research Associate with Texas A&M University, Texas, USA. In 2001 he was appointed Lecturer with the Department of Communications and Computer Engineering at the University of Malta and is now a Senior Lecturer. He is currently the Deputy Dean of the Faculty of ICT at the University of Malta.

Dr. Debono is a senior member of the IEEE and served as chair of the IEEE Malta Section between 2007 and 2010. He is the IEEE Region 8 Conference Coordination sub-committee chair for 2012. He has served on various technical program committees of international conferences and as a reviewer in journals and conferences. His research interests are in wireless systems design and applications, multi-view video coding, resilient multimedia transmission and modeling of communication systems.

Energy Harvesting Relay Networks

A short review for "Voluntary Energy Harvesting Relays and Selection in Cooperative Wireless Networks"

B. Medepally and N. B. Mehta, "Voluntary Energy Harvesting Relays and Selection in Cooperative Wireless Networks," IEEE Trans. Wireless Commun., vol.9, no.11, Nov. 2010.

Battery life is an important issue for many wireless networks, e.g., wireless sensor network (WSN), where it is cumbersome to lay electricity cables to power the wireless nodes. As a result, the wireless nodes need to be equipped with pre-charged non-rechargeable batteries that provide the energy required for their sensing, computation, and communication tasks. Due to its limited capacity of battery, the nodes become dysfunctional because their batteries get drained out. Energy harvesting (EH) is a promising and green solution to solve the battery problem and limited lifetime of wireless networks [1]-[4]. An EH node can replenish its battery by harvesting solar energy, wind energy, radio frequency energy, piezoelectric energy, and several other renewable forms of energy. Therefore, the concept of EH overcomes the necessity of periodic battery replacements and it can contribute to the decrease of the network maintenance overhead. For the above advantage, EH has been gaining attention from various fields. However, the EH is totally dependent on the uncontrollable sources such as solar and wind. Thus, the amount of energy and the time instants at which it is available can be random. In general, this is a function of the *energy profile*, which mathematically models the energy harvesting random process [1].

The reliability of the network can be significantly improved by cooperative communication, which utilizes the wireless node other than the source and the destination. In this paper, the authors propose the use of EH relays in a cooperative wireless network. They analyze the performance of a system consisting of a source node that communicates data to a destination node with the help of multiple intermediate EH amplify-and-forward (AF) relays. The choice of AF relay protocol is motivated by the fact that it is suitable for networks that require low complexity nodes, such as WSN. An AF relay is simple because it amplifies and forwards the signal it receives, without decoding it [5]. To avoid the tight synchronization among simultaneously transmitting relays that are at different locations while obtaining diversity gain, the best relay selection criterion has been widely considered [6]-[8]. For the best relay selection, the relay which improves the signal-to-noise ratio (SNR) the most is selected from the relay candidate set.

Different from the conventional relay system considered in literature, some of the relays are not available as it exploits EH for its power charge. The availability of the energy at each relay is subjected to the fundamental energy neutrality constraint. Thus, in this paper, the authors introduce the two categories of the relays, namely "*energy unconstrained relay*" and "*energy constrained relay*". The relay which harvested enough energy for forwarding a data packet from the source to the destination is considered as energy unconstrained relay. The relay which does not have enough energy for forwarding is considered as energy constrained relay. Among the energy unconstrained relays, the one that improves the SNR at the destination the most is selected. The selected EH relay consumes energy from its battery when it forwards data to the destination. In case all the relays are energy constrained, the source has to rely on its direct link to the destination for data transmission.

The authors evaluate the symbol error rate (SER) observed by the destination as a function of the channel parameters and the transmit power settings of the source and relays. First, the requisite intuition is developed by analyzing a system in which the source-to-relay channels are statistically identical and the relay-to-destination channels are also statistically identical. Thereafter, the general case is analyzed, where the different channels in the system are not identical is analyzed. Further insights are gained by considering asymptotic regimes in which the number of EH relays is large or when the mean channel gains are large.

A key outcome of the study is that using EH relays can significantly reduce the energy consumption at the source. Another design insight is the importance of characterizing the *energy unconstrained regime* of an EH relay, where the relay harvests more energy on average than it requires and the randomness in energy harvesting no longer affects its ability to serve as a relay. The optimal transmit power, which minimizes SER, is given at the point where all the relays are energy unconstrained. The analysis also quantifies how a relay becoming energy unconstrained depends on the rate at which it harvests energy, its transmit power, and also the other relays in the network.

The study brings out how the operation of wireless networks that utilize EH is different from the operation of conventional networks, where the nodes are equipped with non-rechargeable batteries. In the conventional relay network, the key design goal is the minimization of energy consumption in order to increase the lifetime of the network [9]. On the other hand, the focus changes to judiciously utilizing all the harvested energy in EH networks. For example, in the EH relay networks, the aggressive increase of the transmit power of each relay contributes to the lower SER performance. However, it also drains the relay's battery energy and may lead to it being unavailable later. This reduces the number of available relays. On the other hand, conservatively reducing the relay transmit power to save energy for later use increases results in the higher SER performance and may even lead to the relay not being able to utilize all the energy it harvests.

The general conclusion given in this paper, which is likely to impact the future designs of wireless systems, is that the physical and multiple access layers of an EH network need to be redesigned. Different from the conventional relay networks, the operation of an EH relay network is fundamentally governed by the energy neutrality constraint or the law of conservation of energy, which simply states that the energy utilized by an EH node cannot exceed the energy harvested by it. The important conclusion is that the system design is affected not only by the energy harvesting rate but also the energy utilization rate at which the energy is utilized by the sensing and communication protocols employed by the network.

References:

[1] A. Kansal, J. Hsu, S. Zahedi, and M. B. Srivastava, "Power management in energy harvesting sensor networks," *ACM Trans. Embedded Comput. Syst.*, vol. 7, pp. 1–38, Sep. 2007.

[2] J. A. Paradiso and T. Starner, "Energy scavenging for mobile and wireless electronics," *IEEE Trans. Pervasive Comput.*, pp. 18–27, Jan.–Mar. 2005.

[3] V. Raghunathan, S. Ganeriwal, and M. Srivastava, "Emerging techniques for long lived wireless sensor networks," *IEEE Commun. Mag.*, pp. 108–114, Apr. 2006.

[4] C. R. Murthy, "Power management and data rate maximization in wireless energy harvesting sensors," *Intl. J. Wireless Inf. Netw.*, July 2009.

[5] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, pp. 3062–3080, 2004.

[6] D. S. Michalopoulos and G. K. Karagiannidis, "Performance analysis of single relay selection in Rayleigh fading," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 3718–3724, Oct. 2008.

[7] Y. Zhao, R. Adve, and T. J. Lim, "Symbol error rate of selection amplify-and-forward relay systems," *IEEE Commun. Lett.*, vol. 10, pp. 757–759, Nov. 2006.

[8] A. Bletsas, A. Khisti, D. P. Reed, and A. Lippman, "A simple cooperative diversity method based on network path selection," *IEEE J. Sel. Areas Commun.*, vol. 24, pp. 659–672, Mar. 2006.

[9] I. Dietrich and F. Dressler, "On the lifetime of wireless sensor networks," *ACM Trans. Sens. Netw.*, vol. 5, pp. 1–39, Feb. 2009.

R-Letter Editor:

Koichi Adachi received the B.E., M.E., and Ph.D degrees in engineering from Keio University, Japan, in 2005, 2007, and 2009 respectively. From 2007 to 2010, he was a Japan Society for the Promotion of Science (JSPS) research fellow. Currently he is with the Institute for Infocomm Research, A*STAR, in Singapore. His research interests include cooperative communications. He was the visiting researcher at City University of Hong Kong in April 2009 and the visiting research fellow at University of Kent from June to Aug 2009.

Concealing Errors in Video without Drawing Attention

A short review for "Saliency -cognizant Error Concealment in Loss-corrupted Streaming Video"

Hadi Hadizadeh, Ivan V. Bajic, Gene Cheung, *"Saliency -cognizant Error Concealment in Loss-corrupted Streaming Video," IEEE International Conference on Multimedia and Expo 2012*, Melbourne, Australia, July 2012.

Recall a time when you were a little boy/girl and carelessly ruined a clothing article, *e.g.*, dirtied your socks by running in mud. When confronted by your mother, the best strategy is to divert your mother's visual attention away from the troubled spot to some other attention-grabbing spatial areas, perhaps to moving hand gestures, shaking head, etc. It turns out this simple lesson learnt at an early age—concealing errors without drawing attention—can be applied to error concealment in loss-corrupted streaming video as well.

The basic problem is the following. When streaming video over a loss-prone network like today's best-effort Internet or wireless networks that are susceptible to channel fading and interference, unavoidable packet losses will result in missing macroblocks (MB) in the received video, degrading visual quality. Because video has been efficiently compressed, there will remain little redundancy in the received data to recover the missing MBs perfectly. In order to limit the search space, one can design data fitting criteria, such as boundary matching [1], that the replacement MB must satisfy. Nonetheless, error concealment in loss-corrupted video is in general an *under-determined* problem, meaning that there are many candidate MBs *x* that will minimize a data fitting error function *fitErr(x)*:

$$\min_x \; fitErr(x) \qquad (1)$$

Instead of introducing other regularization terms to (1) to make the optimization well defined, Hadizadeh et al. introduced a *low-saliency prior* to the optimization objective:

$$\min_x \{ fitErr(x) + \lambda saliency(x) \} \qquad (2)$$

The idea is almost identical to the "muddy socks" analogy discussed earlier. Given the replacement MB *x* is very likely imperfect and contains errors due to the under-determined nature of the problem, let's find a MB *x* that has small data fitting error *and* induces *low visual saliency*—*i.e.*, draws little visual attention. Visual saliency—a measure of how different spatial regions in an image or a video frame will relatively draw viewer's visual attention—has been studied extensively in the computer vision literature in the past decade [2]. The basic idea is to detect low-level features such as color and luminance contrast, motion, flickers, etc. in the video frame, and sum up the effects locally into a saliency map that describes the visual attentiveness by a human observer per-pixel. The low-saliency prior in (2) essentially ensures that the replacement MB *x* will be less attention grabbing than the other spatial regions composed of correctly received MBs.

Hadizadeh et al. implemented this simple idea in the context of an existing error concealment scheme called RECAP [3]. The idea in RECAP is to transmit a low resolution (LR) thumbnail along with the high resolution (HR) video frame, so that in the event of packet losses, the LR thumbnail can be used as a template to identify good candidate MBs in previous correctly received HR frames for MB recovery in the current loss-corrupted frame. The loss recovery philosophy of RECAP is orthogonal to one in channel coding like *forward error correction* (FEC). FEC decreases the likelihood of packet losses (as observed by the application layer after FEC decoding) in the first place by judiciously increasing data redundancy; RECAP minimizes the adverse effects to video quality due to packet losses, given that losses are unavoidable.

In the RECAP context, where available thumbnail *y* readily constitutes a data fitting error function, optimization (2) of finding a replacement MB *x* becomes:

$$\min_x \|y - DLx\|_2 + \lambda S(x) \qquad (3)$$

Where *L* and *D* are respectively the low-pass filter and down-sampling matrices mapping a HR block to a LR thumbnail, and *S()* is the saliency term given replacement MB *x*. Hadizadeh et al. identified the replacement block *x* by first finding a short list of candidate MBs in previous correctly received HR frames, using available thumbnail *y*, that has small fitting error, then iteratively reducing the saliency of each candidate MB by changing the block's low-level features [4]. The chosen candidate MB is one that has the smallest objective value in (3). Experimental results show that the recovered video is much more appealing when using the low-saliency prior.

Finally, perhaps what is more surprising is that by inserting the low-saliency prior to solve (2) instead of (1), not only the perceptual subjective quality of the error-concealed video improves, but also the objective quality (in terms of Peak Signal-to-Noise Ratio (PSNR)) is also drastically increased. The reason is that if video is transmitted using Unequal Error Protection (UEP), where the more visually salient regions are protected more than the less salient regions, and thus the low-saliency prior in (2) is also a *true* prior. In other words, in an UEP transmission system, lost MBs are more likely to be of low visual saliency, so the low-saliency prior helps us identify the correct missing MBs with higher probability. Hadizadeh et al. showed that using the low-saliency prior, one can obtain PSNR improvement of up to 3.2dB over original RECAP.

References:

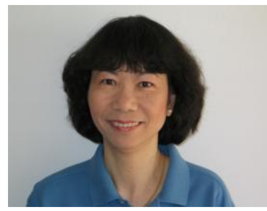
[1] Y. Chen, Y. Hu, O. Au, H. Li, and C. W. Chen, "Video error concealment using spatio-temporal boundary matching and partial differential equation," in *IEEE Transactions on Multimedia*, January 2008, vol. 10, no.1, pp. 2–15.

[2] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," in *IEEE Transactions on Pattern Analysis and Machine Intelligence,* November 1998, vol. 20, no.11, pp.1254–1259.

[3] C. Yeo, W. t. Tan, and D. Mukherjee, "Receiver error concealment using acknowledge preview (RECAP)–an approach to resilient video streaming," in IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei, Taiwan, April 2009.

[4] A. Hagiwara, A. Sugimoto, and K. Kawamoto, "Saliency-based image editing for guiding visual attention," in *1st International Workshop on Pervasive Eye Tracking and Mobile Eye-Based Interaction*, Beijing, China, September 2011.

R-Letter Editor:



Irene Cheng, SMIEEE, is the Scientific Director of the Multimedia Research Group, an Adjunct Faculty in the Faculty of Medicine & Dentistry and the Faculty of Science, University of Alberta, Canada. Her research interests, among others, include incorporating human perception, incorporating the concept of Just Noticeable Difference (JND) following psychophysical methodology, to improve multimedia, graphics and computer vision techniques. She completed her PhD at the University of Alberta and conducted postdoctoral research at the University of Pennsylvania. Before joining academia, she was a regional Information Technology executive in Lloyds Bank International, Asia. She received an Alumni Recognition Award in 2008 from the University of Alberta for her R&D contributions. Dr. Cheng was the Chair of the IEEE Northern Canada Section, EMBS Chapter (2009-2011), Board Member of the IEEE System, Man and Cybernetics (SMC) Society, Human Perception in Vision, Graphics and Multimedia TC, and was the Chair of the IEEE Communication Society, MMTC Interest Group on 3D rendering, processing and communications (2010-2012). She was the lead General Chair in IEEE ICME 2011 and was a visiting professor at Institut National des Sciences Appliquees (INSA) de Lyon, France in 2011. She has over 100 publications including two books.

Objective and subjective quality assessment of 3D synthesized views for 3DTV

A short review for "Towards a new quality metric for 3D synthesized view assessment"

E. Bosc, R. Pépion, P. Le Callet, M. Köppel, P. Ndjiki-Nya, M. Pressigout, L. Morin, " Towards a new quality metric for 3D synthesized view assessment", IEEE Journal on Selected Topics in Signal Processing, Volume: 5, Issue 7, pp. 1332 - 1343, Nov. 2011.

Emerging 3D video applications have encouraged investigations in various fields from video acquisition to display technologies. Most of these applications are under the scope of 3D television (3DTV) and free viewpoint video (FVV) [1]. 3DTV provides a depth feeling attributed from advanced 3D displays. FVV interactively allows the user to control the viewpoint in the scene. Considering the demand for high-quality visual content, the success of 3D video applications is closely related to its ability to provide viewers with a high quality level of visual experience.

The added value, compared to 2D conventional video, comes from the exploitation of multiple video sequences acquired at different viewpoints in the scene. These video sequences can be processed into different 3D representations [2]: image-based representations (conventional stereoscopic video, multi-view video, etc.), surface-based representations (polygonal meshes, etc.), point-based representations or depth image-based representations (2D+Z, multi-view video plus depth, layered depth video, etc.) among others. The study of this paper is in line with the depth image-based representation context, especially using multi-view video plus depth data, referred to as MVD.

MVD designates the association of multiple conventional color videos, referred to as *texture data* and their correspondent depth video sequences, referred to as *depth data*. Depth image-based rendering (DIBR) algorithms are then used to synthesize novel views of the scene, different from those captured by the cameras.

The problem of assessing 3D video quality first refers to the object under test. There is no standardized quality assessment framework for 3D video, so the object under test is chosen depending on the desired application, on the used 3D representation (mentioned above) and on the chosen display (i.e., stereoscopic, auto-stereoscopic or multi-autoscopic). This paper focuses on the quality of the synthesized views for different reasons. Firstly, 3DTV technology relies on the stereopsis phenomenon. This designates the fact that the human brain fuses two slightly different images, presented on each eye, and interprets the 3D content [3]. As a result, 3D displays should provide the appropriate stereoscopic images to ensure depth feeling. Yet, the captured views may not be stereo-compliant depending on the display characteristics. In that case, view synthesis is needed to create the correct stereoscopic pairs. Secondly, in the case of FVV, smooth navigation into the scene requires the generation of non-acquired views. Thirdly, for broadcast situations, constraints on bandwidth limit the amount of data to be transmitted. Generally, virtual views are synthesized from compressed texture and depth data. So compression performances can be evaluated by the rendered views quality. Fourthly, new compression methods using view synthesis prediction have recently been proposed [4][5].

3DTV technology has brought out new challenges regarding the question of synthesized view evaluation. Synthesized views are generated through a DIBR process. This process induces new types of artifacts whose impact on visual quality has to be identified considering various contexts of use. While visual quality assessment has been the subject of many studies in the last twenty years, there are still some unanswered questions regarding new technological improvement. DIBR brings new challenges mainly because it deals with *geometric distortions*.

The authors have conducted two subjective experiments to study the DIBR-based synthesized view evaluation problem. Key frames from 84 synthesized sequences coming from seven different view synthesis algorithms have been assessed by 43 observers according to two different methodologies: *Absolute Categorical Rating* and *Paired Comparisons* methods. Statistical analyses show that fewer observers were required for Paired comparisons tests to establish the algorithms distinctions, however both results highly correlate hence it suggests that these classical subjective protocols are well suited for synthesized view quality assessment.

The same material has also been evaluated using state of the art 2D objective quality metrics: PSNR, UQI, PSNR-HVS, SSIM, VSNR, etc. Surprisingly, while DIBR rendered virtual views are 2D images, these usual 2D metrics do not correctly render the human judgment. Indeed, synthesized views contain specific artifacts located around the disoccluded areas, but usual metrics seem to be unable to express the degree of annoyance perceived in the whole image.

These results provide hints for a new objective measure for synthesized view quality assessment. The authors propose two approaches: the first one is based on the analysis of the shifts of the contours of the synthesized view; the second one is based on the computation of a mean SSIM score of the disoccluded areas. They both provide encouraging preliminary results for this emerging quality assessment problem.

References:

[1] A. Smolic, K. M¨uller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, and T. Wiegand, “3-D video and free viewpoint Video-Technologies, applications and MPEG standards”, IEEE ICME 2006.

[2] A. Smolic, K. Mueller, N. Stefanoski, J. Ostermann, A. Gotchev, G. B Akar, G. Triantafyllidis, and A. Koz, “Coding algorithms for 3DTV-asurvey”, IEEE transactions on circuits and systems for video technology, vol. 17, no. 11, pp. 1606–1620, 2007.

[3] R. Patterson, “Human factors of 3-D displays,” Journal of the SID, vol. 15, no. 11, pp. 861–871, 2007.

[4] S. T Na, K. J Oh, and Y. S Ho, “Joint coding of multi-view video and corresponding depth map,” IEEE ICIP, 2008.

[5] Y. Morvan, D. Farin, and P.H.N. de With, “Joint depth/texture bit allocation for multi-view video compression”, Picture Coding Symposium, 2007.

R-Letter Editor:

Guillaume Lavoué received his engineering degree in signal processing and computer science from CPE Lyon (2002), his M.Sc. degree in image processing from the University Jean Monnet, St.-Etienne (2002), and his Ph.D. degree in computer science from the University Claude Bernard, Lyon, France (2005). Since September 2006 he is associate professor at the French engineering university INSA of Lyon, in the LIRIS Laboratory (UMR 5205 CNRS).
He is author or co-author of over 50 publications in international journals and conferences. His research interests include indexing, watermarking, compression and streaming of 3D models, and also perception and quality assessment for computer graphics.

Scene Segmentation and Pedestrian Classification from 3-D Range and Intensity Images

A short review for "Scene Segmentation and Pedestrian Classification from 3-D Range and Intensity Images"

X. Wei, S. L. Phung, and A. Bouzerdoum, "Scene Segmentation and Pedestrian Classification from 3-D Range and Intensity Images," in Proc. IEEE Int. Conf. on Multimedia and Expo (ICME), pp. 103-108, Melbourne, Australia, July 2012.

Electronic navigation systems for assisting blind people to sense the physical environment and detect obstacles have been in focus for a while. However, they are commonly based on two-dimensional color images having a limitation to obtain more exhaustive information on the detected obstacle, like type, velocity and distance, or to sense and truly represent the three-dimensional environment.

The authors propose a novel approach for detection of obstacles that uses time-of-flight three-dimensional (3-D) range cameras [1]. The proposed approach allows for efficient object segmentation and also provides estimates of the distance and speed of the objects in the scene. In addition, the method is based on distinguishing between pedestrian and non-pedestrian obstacles.

The input image captured by the range camera is segmented into depth layers. The segmentation algorithm consists of several steps. First, the image is pre-processed to reduce the noise and discard unreliable pixels. Next, to distinguish the objects in contact with the ground from the ground itself, segregation by normal surface vectors is applied. A 3-D point is considered a ground pixel if the ratio between the vertical and horizontal components of the average normal vector computed from all neighbor triangulation surfaces is larger than a selected threshold.

Then, the entire image is segmented into distinctive depth layers using multiple thresholds, which are determined adaptively as the local minima from the image histogram. The histogram-based segmentation is simple and efficient, but prone to under-segmentation because of thresholding only scalar depth values. To improve the performance, the authors deploy the mean-shift algorithm after the histogram processing to determine 3-D spatial relations in the under-segmented regions. The proposed mean-shift segmentation method is adapted to the standard deviation of each region, so that over-segmentation is avoided and the computational complexity is reduced. Two adjacent segmented regions are merged if the number of pixels along their common boundary is smaller than a given threshold. Finally, the average distance of each segment is calculated and the information is conveyed to the blind person together with the position, velocity, and type of the detected obstacle.

To classify the segmented regions, features are extracted from the range and intensity images captured by the time-of-flight camera so that each segmented region generates one feature vector for the classifier. The vector consists of the features related to the contours of the segmented regions and to the content of the range and intensity images.

First, the discrete Fourier transform (DFT) is used to capture the contours of the segmented regions. A Fourier descriptor invariant to scale, translation, and rotation of the image is formed from these DFT coefficients. Then, GIST features are extracted from the range and intensity images. These features are obtained by multi-scale oriented Gabor filtering that results in low-dimensional blurred representation of the image [2].

The final feature vector consists of 100 Fourier and 1,024 GIST features (512 from the range region and 512 from the intensity region). Finally, a support vector machine (SVM) classifier with the radial basis function (RBF) kernel is used to discriminate between pedestrian and non-pedestrian obstacles.

In the experiments, the dataset was acquired using a TOF camera with the resolution 144×176 pixels, at the frame rate of 30 frames per second and with the depth-of-field equal to 5m. The images were recorded from different indoor and outdoor scenes, under various daylight conditions. For each pixel, the camera produces five outputs: the x, y, and z coordinates, amplitude, and the confidence map. For segmentation evaluation, the ground-truth segmentation is generated manually. The segmentation performance is evaluated using the weighted Jaccard coefficient, which is related to the area of the segmented region. The proposed segmentation algorithm is compared with local variation [3], Markov random field [4], Graph-cut [5], Otsu and K-means. It achieves a segmentation rate of 73.1% and it outperforms that of the other tested methods. In the feature extraction stage, the Fourier and GIST features were evaluated on a set of 1000 range patterns and 1000 intensity patterns. The background is also varied to include both indoor and outdoor scenes. The classification rate was evaluated using ten-fold cross validation. For comparison purposes, three other image features, namely the SIFT [6], the HOG [7], and Fourier features, were evaluated on the same data set. Experimental results show that the proposed method achieves a classification rate of 99.5%, which is higher than those of SIFT, HOG, GIST, or Fourier features. A possible reason is that the GIST method extracts global texture information, from range and intensity images, and ignores details in the inner parts of the object. Furthermore, the Fourier descriptor enhances the boundary features of objects; for low-resolution range images, shape and contour are the most dominant features.

In summary, this paper presents a promising platform for improving the performance of obstacle detection using time-of-flight range cameras. Even though some parameters in the proposed method have been empirically chosen, the analysis and demonstrated efficiency in the experiments motivate further work along the same lines.

References:

[1] PolyWorks, "3-D metrology hardware review," 2010, http://www.innovmetric.com.

[2] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," International Journal of Computer Vision, vol. 42, no. 3, pp. 145–175, 2001.

[3] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," International Journal of Computer Vision, vol. 59, no. 2, pp. 167–181, 2004.

[4] O. Demirkaya, M. H. Asyali, and P. Sahoo, Image Processing with MATLAB: Applications in Medicine and Biology, CRC Press, 2008.

[5] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 9, pp. 1124–1137, 2004.

[6] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.

[7] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2005, vol. 1, pp. 886–893.

R-Letter Editor:

Vladan Velisavljević received the B.Sc. and M.Sc. (Magister) degree from the University of Belgrade, Serbia, in 1998 and 2000, respectively, and the Master and Ph.D. degree from EPFL, Lausanne, Switzerland, in 2001 and 2005.

From 1999 to 2000, he was a member of academic staff at the University of Belgrade. In 2000, he joined the Audiovisual Communications Laboratory (LCAV) at EPFL as teaching and research assistant, where he was working on his Ph.D. degree in the field of image processing. In 2003, he was a visiting student at Imperial College London. From 2006 to 2011, Dr. Velisavljević was a Senior Research Scientist at Deutsche Telekom Laboratories, Berlin, Germany. Since October 2011, he is Senior Lecturer (Associate Professor) at Bedfordshire University, Luton, UK.

He has co-authored more than 40 research papers published in peer-reviewed journals and conference proceedings and he has been awarded or filed 4 patents in the area of image and video processing. He co-organized a special session at IEEE ICIP-2011 on compression of high-dimensional media data for interactive navigation and he is a co-chair of the Multimedia Computing and Communications Symposium (MCC) at ICNC-2013. His research interests include image, video and multiview video compression and processing, wavelet theory, multi-resolution signal processing and distributed image/video processing.

Peer-assisted Dynamic Adaptive Streaming over HTTP

A short review for "Towards peer-assisted dynamic adaptive streaming over HTTP"

S. Lederer, C. Mueller, and C. Timmerer, "Towards peer-assisted dynamic adaptive streaming over HTTP," in Proc. IEEE Packet Video Workshop (PV), pp. 161-166, May 2012.

Nowadays, multimedia is omnipresent on the Internet, comprising 58% of the entire Internet traffic in North America's fixed access networks [1]. Given the amount of traffic, but also the inherent characteristics of the networks and the requirements for improved Quality of Service (QoS), considerable research and development was devoted on adaptive HTTP streaming methods. This has also resulted, in recent years, in the standardization of the Dynamic Adaptive Streaming over HTTP (DASH) [2][3] standard. The basic concept behind DASH is that one may encode the media content at a variety of formats, e.g. different resolutions, bitrates, frame rates, etc. as possible representations of the content. These representations are then also *chopped* into fixed-size segments. These segments can then be addressed via HTTP GET requests from the client and thus served using conventional HTTP Web servers. The entire logic is placed on the client side as to achieve the best streaming performance for the given user context. Therefore, the system could be described as a pull based system. Since the media retrieval is based on a per-segment basis, this provides the possibility to adapt the media stream during the session to changing network or playback system capabilities, such as bandwidth variations, battery life and so on. DASH also provides a few infrastructure advantages. In particular, no dedicated streaming servers are necessary, therefore allowing a conventional HTTP-infrastructure, such as Content Delivery Networks (CDN), to be used. This can have a significant impact on the infrastructure costs of, e.g., streaming providers

Point-to-Point (P2P) services and MPEG-DASH have several elements in common, such as, for example, the client-initiated pull approach as well as the segment-wise nature of the content. Given these commonalities, it makes sense to investigate the combination of both of these concepts, which was done by the authors of this paper.

P2P-based video streaming comes together with several restrictions. From a QoS perspective the most important limitation is the existence of asymmetric home Internet connections with significant lower upload bandwidth than download bandwidth. This leads to the problem where a client cannot serve the same amount of data as it receives. This has the implication that it is no more possible to consume media content at the best quality while, at the same time, maintaining smooth playback. Peer-assisted streaming is a promising compromise to achieve both high quality and smooth media consumption, while at the same time reducing the server bandwidth requirements significantly. The authors of the paper presented a combination between conventional client-server-based streaming via DASH and P2P traffic, which is referred as pDASH (peer-assisted DASH). In the proposed system the client has the ability to download files or parts thereof from other peers that have already consumed the desired content. Since these clients may not have suficcient upload capacities, the client can also always rely on the high bandwidth of the origin server, e.g. a CDN, to download missing parts of the content. This can help maintain smooth playback. Due to this approach, it is possible to achieve significant reductions in infrastructure and bandwidth needs, as well as cost, while maintaining the same QoS [4].

The presented pDASH approach was designed in a straightforward way, maintaining only DASH-compliant communication between the client and the server as well as with the other clients/peers. Therefore, information, which a peer has already downloaded from the currently needed DASH segments, is integrated in the Media Presentation Description (MPD), an XML document describing the DASH representations, segments, or other related information. This is done in a compatible way to the DASH standard, offering the client the possibility to download the segments or parts thereof from one or more clients, as well as directly from the CDN. Since the MPD can be used directly to identify the segment's location in the P2P network, this enables the system to work without the need for any overlay network or distributed hash tables [5]. This results in a simple but effective architecture, which limits the modifications on the client to mainly three changes: a) each DASH client has to run an HTTP server component to satisfy requests from other peers; b) the DASH adaptation and download logic has to be modified to handle the different content sources, i.e., multiple peers or the CDN; and c) a local cache for downloaded content needs to be maintained at the client. This is needed to serve segments or parts thereof to requesting clients. The size of this cache obviously influences the number of accommodatable peer-requests and thus the performance of the entire system.

On the other hand, on the server side, the authors proposed a central segment tracker, in addition to the HTTP Web server that serves the DASH content. This tracker monitors the received segment requests, together with each client's Internet Protocol (IP) address and timestamp, into a database or file. This information is used by a MPD generation component that serves MPDs containing the latest P2P network content distribution snapshot to clients requesting the content in the future. Although these components introduce logic on the server, they can run on today's Web Servers, e.g., using PHP, CGI, etc., and thus they are not preventing the usage of cheap and existing HTTP-based infrastructure

The performance of pDASH has been evaluated by the authors based on an OMNeT++ [6] simulation, comprising of different types of clients under different bandwidth conditions, such as clients supporting different bandwidth while utilizing asymmetric or symmetric Internet connections. Using a client arrival scheme for the simulation scenario, the server bandwidth reduction compared to a non-peer-assisted scenario, i.e., a classical DASH client-server system, was analyzed. The results show that the P2P traffic can definitely contribute to the traffic needed for downloading segments, whereas, in certain cases, more than 50% of the content can be served by other peers. This leads to a reduction of 15% to 25% of the needed server bandwidth in this first evaluation, which can be directly converted to infrastructure and traffic costs. The authors showed this by taking the pricing model of the Amazon CloudFront service into account. This can definitely influence relevant business areas and give providers a competitive advantage against others.

Commercial products, like the audio-streaming service Spotify [4], are already showing the potential of peer-assisted streaming systems. The presented pDASH system provides a standard-conformant solution to such services. Future work could further improve the system's performance, for example, by improving the user with peer-selection algorithms or by leveraging MPD-update mechanisms during the streaming session. In addition to this, the Content Centric Networking (CCN) [6] approach of the Future Internet (FI) movement may be an interesting overlay, which can help to further simplify the peer-assisted DASH architecture.

References:

[1] Sandvine 2012, *Global Internet Phenomena Report 1H 2012,* Sandvine Intelligent Broadband Networks.

[2] T. Stockhammer, Dynamic Adaptive Streaming over HTTP – Design Principles and Standards", In *Proceedings of the Second Annual ACM Conference on Multimedia Systems 2011* (MMSys11), ACM, New York, NY, USA, 133-144,.

[3] ISO/IEC DIS 23009-1.2, *Information technology — Dynamic adaptive streaming over HTTP (DASH) — Part 1: Media presentation description and segment formats*

[4] G. Kreitz, F. Niemelä, "Spotify – Large Scale, Low Latency, P2P Music-on-Demand Streaming", In *Proceedings of the IEEE International Conference on Peer-to-Peer Computing (P2P) 2010*, Delft, Netherlands, pp. 1-10, 2010.

[5] L. Keong Eng, J. Crowcroft, M. Pias, R. Sharma, S. Lim, "A survey and comparison of peer-to-peer overlay network schemes", *in IEEE Communications Surveys & Tutorials*, Vol.7, No.2, pp. 72- 93, Second Quarter 2005.

[6] OMNeT++, http://www.omnetpp.org/

[7] Jacobson, V., Smetters, D., Thornton, J., Plass, M., Briggs, N. and Braynard, R. . 2009. Networking named content. In *Proceedings of the 5th international conference on Emerging networking experiments and technologies* (CoNEXT '09). ACM, New York, NY, USA, 1-12.

R-Letter Editor:

Alexis M. Tourapis (SM'07) received the Diploma degree in electrical and computer engineering from the National Technical University of Athens (NTUA), Greece, in 1995 and the Ph.D. degree in electrical engineering from the Hong Kong University of Science & Technology, HK, in 2001. During his Ph.D. years, Dr. Tourapis made several contributions to MPEG standards on the topic of Motion Estimation. He joined Microsoft Research Asia in 2002 as a Visiting Researcher, where he worked on next generation video coding technologies and was an active participant in the H.264/MPEG-4 AVC standardization process. From 2003 to 2004, he worked as a Senior Member of the Technical Staff for Thomson Corporate Research in Princeton, NJ, on a variety of video compression and processing topics. He later joined DoCoMo Labs USA, as a Visiting Researcher, where he continued working on next generation video coding technologies. From 2005 until 2011 Dr. Tourapis was with the Image Technology Research Group at Dolby Laboratories where he managed a team of engineers focused on multimedia signal processing and compression. He later joined Magnum Semiconductors as a Senior director, whereas since June 2012 he has been a Video Coding Architect with Apple Inc. In 2000, Dr. Tourapis received the IEEE HK section best postgraduate student paper award and in 2006 he was acknowledged as one of 10 most outstanding reviewers by the IEEE Transactions on Image Processing. Dr. Tourapis currently holds 21 US patents and has more than 80 US and international patents pending. He has made several contributions to several video coding standards ona variety of topics, such as motion estimation and compensation, rate distortion optimization, rate control and others. Dr. Tourapis currently serves as a co-chair of the development activity on the H.264 Joint Model (JM) reference software.

Cross-Layer IPTV Monitoring knocking at Quality of Experiences' Door

A short review for "Cross-layer monitoring in IPTV networks"

Gardikis, G.; Boula, L.; Xilouris, G.; Kourtis, A.; Pallis, E.; Sidibe, M.; Negru, D.; , "Cross-layer monitoring in IPTV networks," *Communications Magazine, IEEE*, vol.50, no.7, pp.76-84, July 2012.

Service assurance in IPTV services becomes imperative in a time when best-effort Internet media is gaining momentum. The content available on the Web is rising exponentially and more users are tending to use the Internet for unmanaged media consumption. Confronting this reality, IPTV providers must provide a clear benefit in order to continue to attract subscribers into their "fenced" networks. Offering high-quality content with guaranteed Quality of Experience (QoE) would be a clear benefit. In order to achieve this, an integrated service and network management architecture is essential, supported by a real-time monitoring system. Cross-layer monitoring is crucial for service quality assurance, fault detection, and system optimization by employing procedures spanning from physical to application layer and across all system segments, i.e., service provider, network provider, and customer domains.

The required functionalities of an IPTV monitoring system can be identified in two main categories: *reactive* and *proactive*. The former is referring to the response of the monitoring system to an abnormal situation whereas the latter is referring to the behavior of the monitoring system under normal operation.

Reactive functionalities include the detection of service outage or quality deterioration, the estimation of the magnitude of a problem, the localization of the failure point and determination of its impact, and the assessment of the impact on the QoE.

Proactive functionalities comprise failure/outage prevention by checking the resource utilization, the workload of system components, the detection of Service Level Agreement (SLA) status, and the monitoring the user behavior.

When it comes to monitoring metrics, they can be categorized according to the architectural layer (e.g., application, network) to which they correspond, specifically i) *user/QoE metrics* [1] (such as Mean Opinion Score), ii) *application/service metrics* (such as video frame loss or player buffer overflow/underflow), and iii) *transport/network metrics* (such as packet loss, jitter). For services relying on the MPEG-2 Transport Stream (M2TS) for delivery, a dedicated set of metrics measured at M2TS level has also been recommended [2].

With regard to the observation point, i.e., where the aforementioned cross-layer metrics are measured, the monitoring procedure may take place either within the distribution network (*in-network monitoring*) and/or at the customers' premises (*client-side monitoring*).

In-network monitoring mainly collects transport/network metrics such as packet loss, inter-arrival jitter, etc. These metrics are measured either by the network elements themselves or by monitoring devices which capture and analyze the traffic or a subset thereof.

Client-side monitoring is performed at the customer premises at three different locations: at the customer network gateway, at the decoder (set-top box), and at the presentation device (after decoding). A challenging issue in client-side monitoring is the derivation of user/QoE metrics, especially the video/audio quality expressed by the Mean Opinion Score (MOS). In the simplest approach, the MOS is directly calculated from network and application metrics using psychometric models. The latter takes into account also bitstream parameters, such as bitrate, resolution, and image complexity and tries to map the impairments introduced by the network to their actual impact on the QoE. That is, estimate how much the objective picture or sound quality is degraded [3]. A more complicated approach is the direct analysis of the decoded visual or audio information. Image-based quality assessment is a computationally intensive procedure which can, however, yield results quite close to user perception [4].

The EU-funded research project ALICANTE (FP7/ICT-248652) [5] involves the design and implementation of a cross-layer distributed monitoring system, tailored to networked media ecosystems, which aims at integrating most of the aforementioned approaches and techniques.

In particular, content stems from the servers/head-ends, traverses one or more associated "VCANs" (virtual networks), reaches the "Home-Boxes" (HB, a media-centric gateways) of the users who have subscribed to the service, and is finally presented in one or more user terminals. The ALICANTE monitoring system is based on distributed agents and deployed at the content server, within the VCAN, the Home-Box, and the terminal. Depending on their role, these agents collect various types of metrics within the delivery network such as:

- Host metrics including host status, CPU/memory/interface utilization, and number of services handled.
- VCAN metrics comprising the nominal and available capacity, average delay, and loss/jitter for each traffic aggregate within the VCAN.
- Session metrics such as per-session packet loss, jitter, and reordering measured at transport/session layer.
- Application/QoE metrics featuring video and audio MOS.

All the measured data are collected from the respective monitoring modules, formatted in XML structure, and communicated over SOAP interfaces providing basic support for interoperability.

Service monitoring in ALICANTE goes beyond the typical centralized paradigm. It also provides an increased level of awareness across the service delivery network in a decentralized manner and, thus, enabling real-time cross-layer and cross-domain interactions and optimizations respectively. Such interactions include:

- *Network-aware service management* via the controlled exposure of VCAN metrics to the service provider.
- Facilitation of *network-aware applications* via the provisioning of network metrics to media applications and *context-aware applications* via the provisioning of terminal monitoring parameters and information.
- Exploitation of network monitoring information for *in-network media adaptation* and *client-side media adaptation* within the home network using terminal and HB metrics.

As a consequence, the authors recommend that future research in multimedia service monitoring should follow a de-centralized paradigm. Instead of aggregating observed metrics to a single entity (for assessment and decision-taking), distributed architectures shall be promoted enabling the collaboration among actors and domains supporting the deployment of network- and context-aware services facilitating real-time service adaptation for improved resource utilization and optimized Quality of Experience.

Finally, interested parties are encouraged to actively participate in research networks working in the field such as the MMTC QoEIG [6] or the COST Action IC1003 Qualinet [7].

References:

[1] ITU-T Rec. G.1080, "Quality of experience requirements for IPTV services", Dec. 2008.
[2] ETSI TR 101 290 V1.2.1, "Digital Video Broadcasting (DVB); Measurement guidelines for DVB systems", Technical Report, May 2001.
[3] S. Winker, P. Mohandas, "The Evolution of Video Quality Measurement: From PSNR to Hybrid Metrics", *IEEE Trans. on Broadcasting*, vol. 54, no. 3, Sep. 2008.
[4] M. Pinson, S. Wolf, "A New Standardized Method for Objectively Measuring Video Quality", *IEEE Trans. on Broadcasting*, vol. 50, no. 3, Sep. 2004.
[5] ALICANTE EC FP7 project (MediA Ecosystem Deployment through Ubiquitous Content-Aware Network Environments), http://www.ict-alicante.eu/.
[6] MMTC Interest Group on QoE for Multimedia Communications (QoEIG), http://committees.comsoc.org/mmc/IG/QoEIG.asp.
[7] COST IC1003 Qualinet, European Network on Quality of Experience in Multimedia Systems and Services, http://www.qualinet.eu/.

R-Letter Editor:

Christian Timmerer is an assistant professor in the Institute of Information Technology (ITEC), Alpen-Adria-Universität Klagenfurt, Austria. His research interests include immersive multimedia communication, streaming, adaptation, and Quality of Experience. He was the general chair of WIAMIS'08, ISWM'09, EUMOB'09, AVSTP2P'10, WoMAN'11, QoMEX'13 and has participated in several EC-funded projects, notably DANAE, ENTHRONE, P2P-Next, ALICANTE, and SocialSensor. He also participated in ISO/MPEG work for several years, notably in the area of MPEG-21, MPEG-M, MPEG-V, and DASH/MMT. He received his PhD in 2006 from the Alpen-Adria-Universität Klagenfurt. Publications and MPEG contributions can be found under research.timmerer.com, follow him on twitter.com/timse7, and subscribe to his blog blog.timmerer.com.

Towards a Better Understanding of Multi-User Cooperation: A Tradeoff between Transmission Reliability and Rate

A short review for "Capacity-Outage-Tradeoff for Cooperative Networks"

W. Guo and I. J. Wassell, "Capacity-Outage-Tradeoff for Cooperative Networks," *IEEE Journal on Selected Areas in Communications*, *vol. 30 (9)*, *Oct. 2012.*

Cooperative communications involve nodes that exchange information and transmit jointly to a common destination [1]. Whilst different users sharing contents is increasingly common on the network- and data-layers, it has not occurred on the physical wireless interface. Cooperative transmission has the potential to dynamically tradeoff data rate with reliability, depending on the multi-media content transmitted. The main weakness of this technology is a lack of understanding of the penalties and benefits in sharing signals on the physical layer.

The rationale for cooperative transmission is that by transmitting the same (or similar) data along different channels, the stochastic nature of multipath fading can be exploited. This has proven to be especially effective in quasi-static (slow) fading channels, where the information coding-length is smaller than the fading variation period. In such a channel, the achievable capacity at arbitrarily high reliability is zero, and reducing the outage probability becomes a challenge [2].

A key drawback with repetitively transmitting the same information along multiple channels is the inefficient use of the channel resource [2]. Therefore, given a fixed power and spectrum constraint, increased cooperation leads to a decreased amount of power and bandwidth per transmission in the cooperation process.

The aim of the paper [3] is to present a tradeoff between data throughput and transmission reliability for cooperative transmission. This tradeoff can assist in achieving content aware cooperation on the physical layer, whereby depending on the transmission reliability requirements of the multi-media content, maximum data rate can be achieved by selecting the optimal number of cooperation partners.

In the reviewed paper [3], the authors first present the novel relationship between achieving greater transmission reliability through cooperation and the associated reduction in transmission efficiency. The primary contribution of the paper is formalizing a tradeoff between transmission reliability (outage probability) and throughput (capacity). A key distinction between this work and existing literature is that signal transmission in this paper employs realistic modulation and forward-error-correction (FEC) codes. This offers a realistic insight compared to the commonly used Shannon expression (infinite code length), which has been shown to be over-optimistic and can lead to misleading results [4].

The analysis performed in the paper [3] uses theoretical expressions based on the bit-error-rate of transmitted information, which is reinforced by Monte-Carlo numerical simulation results. The specific cooperation protocol considered is Decode-and-Forward (DF), which has two key advantages: no noise amplification and no channel estimation at the relays.

The main conclusion from the tradeoff is that increased cooperation doesn't monotonically lead to increased transmission reliability. In fact, the relationship is convex, and for any given system setup (channel conditions and transmission scheme), there exists an optimal set of cooperation partners which maximizes the transmission reliability. Furthermore, maximizing the reliability doesn't lead to maximizing the throughput. Therefore, the system designer or the user needs to tradeoff between:

- Throughput and
- Reliability,

depending on the higher-layer multi-media content transmitted. For example, speech may require high transmission reliability, but a very low throughput rate. For a user with a poor quality channel, cooperating with a large number of partners is desirable. On the other hand, for downloading data, the throughput rate is more important than reliability.

The second contribution of the paper [3] is optimizing the system-level outage-capacity performance through partner selection, which draws on the authors' previous work [5]. The authors found theoretical expressions for partner selection, based on the channel conditions and the desirable performance metrics. For a topology where all nodes are roughly equal-distant to each other (symmetrical), it was found that the optimal number of partners is directly proportional to both the mutual channel strength and the transmission scheme's signal-to-noise ratio (SNR) threshold [3]. For a topology where all nodes are arbitrarily located (asymmetrical), the optimal number of partners can be found using a step-by-step numerical solution [3].

The third contribution of the paper is that given the selected partners, power can be optimally distributed amongst the cooperative transmission slots, maximizing the transmission reliability and data rate. The results show that this can actually lead to requiring fewer cooperation partners, thereby achieving a joint optimality between partner selection and power allocation. Future work can focus on joint optimality solutions, as well as how to combine media streams of different requirements into the same multi-user cooperation cycle.

In summary, the paper [3] has presented a tradeoff between transmission throughput and reliability for multi-user cooperative protocols. This tradeoff can achieve content aware cooperation on the physical layer, whereby depending on the transmission reliability requirements of multi-media contents, maximum data rate can be achieved by selecting the optimal number of partners. There remains significant work to be done on joint partner selection and power allocation strategies, as well as how to combine different multi-media contents into the same cooperation cycle.

References:

[1] J. Laneman, G. Wornell and D. Tse, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, pp. 3062-3080, Dec 2004.

[2] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, 2005.

[3] W. Guo and I. Wassell, "Capacity-outage-tradeoff for cooperative networks," *IEEE Journal on Selected Areas in Communications*, pp. 1641–1648, Oct 2012.

[4] D. Guo, S. Shamai and S. Verdu, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Transactions on Information Theory*, Apr 2005

[5] W. Guo, I. Chatzigeorgiou, I. Wassell and R. Carrasco, "Partner selection and power control for asymmetrical collaborative networks" *IEEE Vehicular Technology Conference*, May 2010.

R-Letter Editor:

Xiaoli Chu is a lecturer in the Dept. of Electronic and Electrical Engineering at the University of Sheffield.

She received the B.Eng. degree from Xi'an Jiao Tong University in 2001 and the Ph.D. degree from Hong Kong University of Science and Technology in 2005. From Sep 2005 to Apr 2012, she was with the Centre for Telecom Research at King's College London. Her research interests include heterogeneous network, interference management, cooperative communication, cognitive communication and green radio.

Paper Nomination Policy

Following the direction of MMTC, the R-Letter platform aims at providing research exchange, which includes examining systems, applications, services and techniques where multiple media are used to deliver results. Multimedia include, but are not restricted to, voice, video, image, music, data and executable code. The scope covers not only the underlying networking systems, but also visual, gesture, signal and other aspects of communication.

Any HIGH QUALITY paper published in Communications Society journals/magazine, MMTC sponsored conferences, IEEE proceedings or other distinguished journals/conferences, within the last two years is eligible for nomination.

Nomination Procedure

Paper nominations have to be emailed to R-Letter Editorial Board Directors:
locheng@ualberta.ca and
xianbin.wang@uwo.ca.

The nomination should include the complete reference of the paper, author information, a brief supporting statement (maximum one page) highlighting the contribution, the nominator information, and an electronic copy of the paper when possible.

Review Process

Each nominated paper will be reviewed by two members of the IEEE MMTC R-Letter Editorial Board. To avoid potential conflict of interest, nominated papers co-authored by an Editorial Board member will be reviewed by guest editors external to the Board. The reviewers' names will be kept confidential. If both members agree that the paper is of award quality, an editor will be assigned to complete the review letter (partially based on the nomination supporting document) for publication. The review result will be final (no multiple nomination of the same paper). The nominator will be acknowledged in the review letter.

R-Letter Best Paper Award

Accepted papers in the R-Letter are eligible for the Best Paper Award competition if they meet the election criteria (set by the MMTC Award Board).

For more details, please refer to http://committees.comsoc.org/mmc/rletters.asp

MMTC examines systems, applications, services and techniques in which two or more media are used in the same session. These media include, but are not restricted to, voice, video, image, music, data, and executable code. The scope of the committee includes conversational, presentational, and transactional applications and the underlying networking systems to support them.