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

R-LETTER

Vol. 3, No. 1, February 2012

IEEE COMMUNICATIONS SOCIETY

CONTENTS

Message from R-Letter Director

Happy Near Year! Welcome to the first issue of IEEE MMTC Review-Letter (R-Letter) in 2012. In the past one and half year, we have approached our goals step-by-step to introduce cutting-edge and promising new concepts and ideas in all multimedia communication related areas. We witness the growth of R-letter, the noteworthy increase of multimedia communication researches, and the significant expansion of mobile multimedia streaming applications. Continuing our efforts, we hope readers can still benefit a lot from every issue of R-letter.

In this issue, we are pleased to introduce ten high quality papers, spanning a wide range of important and hot topics, namely, video conferencing, video quality assessment, video compression technology, network scheduling policy, and wireless security. The first paper, published in the *IEEE Transactions on Multimedia*, proposes a new distributed algorithm for P2P multi-party video conferencing. The second paper, from *IEEE Transactions on Multimedia*, exploits the visual attention mechanism and proposes a new quality model. The third paper, from the *IEEE Transactions on Circuits and Systems for Video Technology*, studies the impact of frame rate and quantization during the quality assessment. The fourth paper, from *IEEE International Conference on Multimedia & Expo*, proposes a reduced-reference quality metric for video plus depth coding. The fifth paper, published in the *IEEE Transactions on Image Processing,* addresses the challenges in the distributed wireless camera networks and develops a systematic approach to improve coding efficiency and meet tight latency constraints. The sixth paper, published in the *IEEE Transactions on Multimedia*, investigates a depth image-based rending approach with advanced inpainting methods in the multiview-video-plus-depth format. The seventh paper, from *IEEE Transactions on Multimedia*, proposes an unsupervised segmentation algorithm to extract moving regions from compression video. The eighth paper, published in the *IEEE Transactions on Networking*, analyzes the delay of a multi-hop wireless network and develops a new queue grouping method. The ninth paper, from *IEEE International Conference on Communications*, studies the physical layer security in the multiple multi-antenna scenarios. The last paper, from *IEEE Transactions on Information Theory*, analyzes the theoretical limit of secure MANET communication and proposes practical schemes.

I would like thank all the editors of this issue for their great work: Carl James Debono, Christian Timmerer, Jong-Seok Lee, Tao Liu, Vladan Velisavljević, Cheng-Hsin Hsu, Hassan Mansour, Ai-Chun Pang, Walid Saad, and Simon Pun. I also would like to thank the R-Letter Co-Director Nabil J. Sarhan for all his great efforts.

Guan-Ming Su

Director, IEEE ComSoc MMTC R-Letter
E-mail: guanmingsu at ieee.org

Video Conferencing over Peer-to-Peer Networks

A short review for "Optimizing multi-rate peer-to-peer video conferencing applications"

Edited by Carl James Debono

M. Ponec, S. Sengupta, M. Chen, J. Li, and P.A. Chou, "Optimizing Multi-Rate Peer-to-Peer Video Conferencing Applications," IEEE Transactions on Multimedia, vol. 13, no. 5, pp. 856 – 868, October 2011.

To reduce traveling costs and its associated time, business and personal meetings can be organized online through video conferencing. This is leading towards a steady growth in multi-party video conferencing applications. To allow continuous growth of users, these systems must be able to guarantee a high user's Quality of Experience (QoE). Provision of such quality becomes challenging especially with the introduction of mobile technology that needs scalable video services, to present the content on different screen sizes, and have different network connections, each with its set of quality parameters. Furthermore, some of the network technologies present an asymmetric channel with the uplink bandwidth being much smaller than the downlink. Thus, providing multi-party video conferencing services on such channels may result in severe packet delays.

A multi-rate Peer-to-Peer (P2P) conferencing system offers a possible solution to address the very diverse needs and available resources of all conference participants. It allows different receivers in the same group to receive the media content at different encoding bit-rates, using, for example, scalable layered coding [1]. To maximize the aggregate quality of experience of all participating peers, the video conferencing system needs to properly allocate shared network resources, in particular peers' upload bandwidth, and needs to encode and route peers' streams in an efficient way. The quality of experience of each video stream received by each peer can be objectively assessed through the Peak-Signal-to-Noise-Ratio (PSNR) of the decoded video [2].

In literature, traditional approaches to multi-party conferencing either employ a central server that first receives a stream from all participants and then distributes it to all others, or demand that every participant sends its own stream to all the other participants directly using unicast links. The former presents a significant resource bottleneck, which is the server, while the latter might provide low quality video because each peer's upload bandwidth usage must be divided among the data streams going to all the other participants. An alternative to this is to utilize multiple multicast trees consisting of paths between some of the peers [2].

In their paper the authors derive a new multicast tree packing scheme for P2P multi-rate multicast and show that the maximum utility (representing the aggregate QoE of the video conference) under multi-rate setting can be achieved by routing along a set of depth-1 and depth-2 trees, whose number is quadratic in the number of peers, for each source in the conference. This means that the delivery of the streams only involves at most one intermediate node, which is important for maintaining the total end-to-end latency of video distribution small.

Rate control in multicast routing for utility maximization has been well tackled in literature, such as [3] and [4]. However, most of the work assumes a single source, a fixed rate, and given tree settings. Link rates or path rates are then used as variables to optimize the system. In their work, the authors provide a new multi-tree based formulation for P2P multi-rate multicast utility maximization problem, where the variables are bit-rates of individual multicast trees. This contrasts with the nonlinear constraints used in other techniques found in literature that use link rates or path rates as variables. Two distributed algorithms were presented to solve this problem; a packet marking based Primal and a queuing delay based Primal-dual. These were given with the proofs of their global asymptotic convergence to the optimal solutions. Unique challenges introduced by maximizing utility in the multi-rate setting as compared to the single-rate case [2] were studied and addressed.

The presented algorithms are easy to implement in a peer-to-peer overlay network developed over the current Internet infrastructure. The techniques provide a solution to decide about encoding bit-rates, flow control, routing and tree construction in a distributed fashion. The system was evaluated in Virtual Lab [5] using ten peers.

Results from the presented prototype video conferencing system show that the approach converges to bit-rates which improve the user's experience. These are similar to the results predicted by the theoretical analysis and offers automatic adaptation to the dynamic network conditions, conference characteristics, and user preferences. The strict delay requirements for video conferencing are satisfied and additional servers can be deployed to aid nodes in the video distribution as well.

The algorithms presented in this paper can be applied to any real-time video application that needs to be delivered to a limited number of users. Further work involves the study of different multicast tree solutions and how the trees can recover when a link fails. Quality of experience is also affected by channel errors and thus error resilience and error corrections need to be in place. Real-time video conferencing does not permit feedback channels and thus resilience needs to be guaranteed on the transmission channels which may lead to more bandwidth requirements, especially on the uplink. Moreover, peer-to-peer on mobile devices can lead to fast depletion of the devices' battery resources and thus coding techniques which transmit less data while still guaranteeing high QoE are needed.

References:

[1] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC Standard," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1130-1120, September 2007.

[2] M. Chen, M. Ponec, S. Sengupta, J. Li, and P.A. Chou, "Utility maximization in peer-to-peer systems," in *Proc. ACM SIGMETRICS*, June 2008.

[3] K. Kar, S. Sarkar, and L. Tassiulas, "Optimization based rate control for multirate multicast sessions," in *Proc. of IEEE INFOCON*, April 2001.

[4] S. Deb and R. Srikant, "Congestion Control for fair resource allocation in networks with multicast flows*," IEEE Trans. on Automatic Control*, vol. 49, no. 4, pp. 274 – 285, April 2004.

[5] V. Padman and N. Memon, "Design of a virtual laboratory for information assurance education and research," in *Proc. of the IEEE Workshop on Information Assurance and Security*, West Point, NY, June 2002..

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.

Perceived Video Quality Assessment based on Attended and Entire Visual Stimuli

A short review for "Balancing attended and global stimuli in perceived video quality assessment"

Edited by Christian Timmerer

J. You, J. Korhonen, A. Perkis, and T. Ebrahimi, "Balancing Attended and Global Stimuli in Perceived Video Quality Assessment," IEEE Transactions on Multimedia, vol.13, no.6, pp.1269-1285, December 2011.

Video quality assessment plays a crucial role within the delivery chain due to the fact that the quality may be degraded by different errors introduced [1], e.g., compression errors in lossy coding schemes and transmission errors through error-prone transmission. Although subjective assessment is considered to be the most reliable means to evaluate quality degradation by asking human subjects to judge distorted video quantitatively or qualitatively based on opinion, it is always time-consuming and cannot be applied in real-time systems. For this reason, a lot of research efforts have been made on development of advanced objective video quality metrics. Mean-squared-error (MSE) and its derivative, peak signal-to-noise ratio (PSNR) are two common indices widely used in quality prediction in video compression and transmission schemes. However, they have been found not to highly correlative with subjective quality judgment, because the characteristics of the human visual system (HVS) were not taken into account [2], as HVS is the final receiver of any video stimuli.

HVS is extremely complex and many latent mechanisms have not been understood adequately. An important mechanism of the HVS that overtly or covertly guides human visual behavior is the attention mechanism [3][4]. There is a debate that the selection of attended information occurs early or later in processing visual stimuli. According to the early selection theory, all stimuli that reach the sensory system are processed, until individual physical attributes are explicitly represented. In contrast, the alternative late selection theory assumes that the selection does not occur before categorization and semantic analysis of all the input [3]. Since visual behavior is often controlled by the visual attention mechanism, it should also be taken into account in video quality assessment. If we adopt the late attention selection theory in video quality assessment scenarios, not only the attended stimuli extracted from the field of vision, but also other unattended stimuli, will both make contributions to evaluation of quality of the entire visual stimuli.

In this paper, by adopting the late selection theory of attention mechanism, the authors proposed to divide the perceived quality of a distorted video into two components: global quality and local quality. The global quality is evaluated by observers allocating their attention to the entire video frame, which usually results into a coarse impression of the perceived quality. The local quality is a result derived from subjects allocating their attention to attended stimuli only. As visual perception is a combined result of both attended and unattended stimuli, the overall video quality assessment can be simulated based on a combination of the global quality and the local quality. Additionally, the authors proposed that this combination can be also influenced significantly by the content of individual video sequences.

In order to generate a local quality measure to evaluate the quality degradation on attended stimuli only, the authors first proposed an advanced video attention model. Several video features that are closely related to attentional behaviors have been modeled and extracted from video sequences. These features include skin, contrast and size of individual segmented regions, motion characteristics in terms of magnitude and coherence, and "surprising" events. As some psycho-visual experiments under quality assessment tasks have demonstrated that visual attentional behaviors can also be influenced by distortion information, specifically localized distortion, the perceived distortion has also been taken into account in the video attention model. Additionally, the smooth pursuit eye movement and its impact on attention region detection have also been applied in the attention model. Each video attention feature can generate an individual attention map. An overall attention map showing the conspicuity at every location of a visual scene, indicated by a scalar quantity, is naturally required for attention region extraction. Consequently, an advanced map fusion approach

based on the feature maps has been proposed using a 3-step fusion: (1) fusion of task-irrelevant attention maps; (2) motion tracking fusion; and (3) task-driven map fusion.

On the other hand, in order to evaluate the global quality and local quality, respectively, the authors carefully designed four image/video features that are related to quality perception. These features describe several important aspects of video quality, such as spatial and temporal video characteristics. Consequently, four quality indices between a reference video with perfect quality and its distorted version have been calculated, including their absolute difference, changes in motion smoothness and angular distribution or orientations, and distortion in color. Using a spatial averaging strategy considering the entire video frame and the detected attention regions, four global quality indices and four local quality indices corresponding to the four quality features have been derived. Subsequently, the authors proposed an assumption that the overall visual perception is a combined result of the entire visual stimuli and the attended stimuli, significantly tuned by the visual content types. Because an observer can receive a high amount of visual stimuli from the unattended regions in a complex scene, it can be reasonably assume that the attended stimuli have a stronger impact on human perception in a simple visual scene as opposed to a more complex scene. Consequently, the authors have proposed an accurate complexity index to measure the spatiotemporal activities, and accordingly, the fusion of global and local quality is tuned by this index, as well as the temporal variation of quality indices over different video frames. Finally, considering other proven conclusions, e.g., the frames in the beginning and the end of a video sequence have more significant impact on the perceived overall quality, and the tendency that increasing the quality of the frames in the end leads to a better perceived quality [5]. Another temporal pooling scheme has been proposed to combine quality values over different temporal video clips into a single quality value.

The authors conducted comprehensive experiments to validate the performance of the proposed video quality metric, comparing with state-of-the-art video quality models in different scenarios. This kind of study is very important, because it attempts to investigate some fundamental mechanisms of the HVS and to explore the mechanisms in practical multimedia applications. As Quality of Experience (QoE) is becoming a more and more important and popular concept representing user experience of multimedia services, investigation of user experience from different points of view and in an interdisciplinary manner, will be very critical.

References:

[1] J. You, U. Reiter, M. M. Hannuksela, M. Gabbouj, and A. Perkis, "Perceptual-based Objective Quality Metrics for Audio-visual Services - A Survey," *Signal Processing: Image Communication*, vol. 25, no. 7, pp. 482-501, 2010.

[2] B. Girod, "What's Wrong with Mean-square Error," in *Digital Images and Human Vision*, A. B. Watson, Ed. MIT Press, pp. 207-220, 1993.

[3] H. Pashler, *The Psychology of Attention*, MIT Press, Cambridge, MA, 1998.

[4] L. Itti, and C. Koch, "Computational Modeling of Visual Attention," *Nature Reviews Neuroscience*, vol. 2, no. 3, pp. 194-203, Mar. 2001.

[5] M. Zink, O. Künzel, J. Schmitt, and R. Steinmetz, "Subjective Impression of Variations in Layer Encoded Videos," *Lecture Notes in Comput. Sci*., vol. 2707, pp. 137-154, Jan. 2003.

Christian Timmerer is an assistant professor in the Institute of Information Technology (ITEC), Alpen-Adria-Universität Klagenfurt, Austria. His research interests include the transport of multimedia content, multimedia adaptation in constrained and streaming environments, distributed multimedia adaptation, and Quality of Service / Quality of Experience. He was the general chair of WIAMIS'08, ISWM'09, EUMOB'09, AVSTP2P'10, WoMAN'11 and has participated in several EC-funded projects, notably DANAE, ENTHRONE, P2P-Next, ALICANTE, QUALINET, 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.

Measuring Trade-off between Temporal Smoothness and Quality in Videos

A short review for "Perceptual quality assessment of video considering both frame rate and quantization artifacts"

Edited by Jong-Seok Lee

Y.-F. Ou, Z. Ma, T. Liu, and Y. Wang, "Perceptual Quality Assessment of Video Considering both Frame Rate and Quantization Artifacts", IEEE Transactions on Circuits and Systems for Video Technology, vol. 21, no. 3, pp. 286-298, March 2011

The demand of multimedia applications and services along with the state-of-the-art video compression techniques is becoming more and more prevalent, such as video telephony, mobile video broadcasting, high definition television (HDTV), and Internet Protocol television (IPTV). The success of emerging devices for video transmission and high-definition displaying over communication networks depends on a reliable method that can predict the video quality automatically and accurately to assure better quality of experience (QoE) of users.

In video streaming, given the bandwidth limitation and display resolution of a receiver, the encoder, a network transcoder, or adaptor has to decide at which spatial, temporal and amplitude resolutions to code, transcode, or adapt a video in order to achieve the best perceptual quality. Therefore, it is important to understand the impact of combinatorial variations along the three dimensions on the perceptual quality, so as to achieve the best trade-off between picture quality and temporal fluidity of the transmitted bit stream. The amplitude resolution is controlled by the quantization stepsize or equivalently quantization parameter (QP), whose impact on quality perception has been studied significantly. On the other hand, studying the joint impact of all three dimensions on the perceptual quality is a complex and challenging task.

This paper extends the authors' prior work [1], investigates the impact of frame rate (FR) and quantization on perceptual quality of a video for laptop devices, and accordingly develops a full-reference objective quality metric.

First, the authors conducted subjective quality assessment experiments with video sequences encoded by the scalable extension of H.264/AVC, scalable video coding (SVC). Seven sequences in CIF resolution (352x288) were used, and the test stimuli had frame rates of 30, 15, 7.5, and 3.75 Hz and QP values of 28, 36, 40, and 44 using coarse grain scalability. Each viewer watched a series of video in a random order, and asked to give overall rating of each video in the range of 0 to 100. In order to remove noisy ratings or outliers, two post screening methods were used in concatenation. The authors first applied the post-screening method specified in Recommendation ITU-R BT.500-11 [2] to remove all ratings by certain viewers whose ratings are outside the range of the majority of the viewers. Then, the viewers' inconsistency was checked, i.e., if a viewer tends to give lower (higher) ratings for videos having higher (lower) FRs than those having lower (higher) FRs for fixed QPs, the viewer is considered as an outlier.

The authors concluded from the subjective test results that the impact of FR and that of quantization is separable. Thus, they proposed to use the product of a spatial quality factor and a temporal correction factor to predict the overall video quality, which was named as Video Quality Metric considering Temporal resolution and Quantization (VQMTQ). The spatial quality factor, which assesses the quality of decoded frames with respect to the corresponding reference frames, was modeled by a sigmoid function with a Pearson correlation coefficient of 0.99. The temporal correction factor, which reduces the quality estimated by the first factor according to the actual FR, was approximated by an inverted exponential function with a Pearson correlation coefficient of 0.95. The proposed quality model contains three parameters. The authors suggest using a fixed value for a parameter for all cases, while the other two are dependent on the content characteristics.

Through experiments using seven datasets, it was shown that the proposed metric outperforms two metric considering only FR effects [3,4] and one considering both FR and quantization effects [5].

The authors also presented ways to estimate optimal values of the two content-dependent

model parameters. Based on leave-one-out cross validation experiments, the parameters were modeled as linear combinations of content features such as motion direction activity, frame difference, and Gabor texture.

In the conclusion, the authors mention that the proposed model is expected to be applicable to other conditions different from those used in their experiments, such as videos having resolutions other than CIF or those encoded with other codecs. This needs to be further verified in the future. In addition, even if the same functional form can be used, the model parameters for the same video content may differ, depending on the encoder configurations. This also needs to be validated in future studies. Finally, it would be interesting to use the proposed quality models, together with a rate model as a function of spatial, temporal, and amplitude resolutions, to determine the optimal operation condition that maximizes the quality given a rate constraint, both for video encoding/transcoding and for scalable video adaptation.

References:

[1] Y.-F. Ou, T. Liu, Z. Zhao, Z. Ma, and Y. Wang, "Modeling the impact of frame rate on perceptual quality of video," in *Proc. Int. Conf. Image Processing* (ICIP), San Diego, Oct. 2008, pp. 689-692.

[2] Rec. ITU-T P.910, "Subjective video quality assessment methods for multimedia applications," 2001.

[3] Z. Lu, W. Lin, B. C. Seng, S. Kato, S. Yao, E. Ong, and X. K. Yang, "Measuring the negative impact of frame dropping on perceptual visual quality," in *Proc. SPIE Hum. Vision Electron. Imaging*, vol. 5666, Jan. 2005, pp. 554-562.

[4] H.-T. Quan and M. Ghanbari, "Temporal aspect of perceived quality of mobile video broadcasting," *IEEE Trans. Broadcast*., vol. 54, no. 3, pp. 641-651, Sep. 2008.

[5] R. Feghali, D. Wang, F. Speranza, and A. Vincent, "Video quality metric for bit rate control via join adjustment of quantization and frame rate," *IEEE Trans. Broadcast.*, vol. 53, no. 1, pp. 441-446, Mar. 2007.

Jong-Seok Lee is an assistant professor in the School of Integrated Technology at Yonsei University, Korea. He received the Ph.D. degree in electrical engineering from KAIST, Korea, in 2006. From 2006 to 2008, he was a postdoctoral researcher and an adjunct professor at KAIST. From 2008 to 2011, he was a research scientist in the Institute of Electrical Engineering, Swiss Federal Institute of Technology Lausanne (EPFL), Switzerland. He is author or coauthor of over 40 publications. His research interests include multimedia quality assessment, audio-visual signal processing, and multimodal human-computer interaction.

Reduced Reference Video Quality Assessment: Video Plus Depth

A short review for "Quality evaluation for real-time 3D video services"

Edited by Tao Liu

C. Hewage and M.G. Martini, "Quality Evaluation for Real-time 3D Video Services", In Proceedings of the IEEE International Conference on Multimedia & Expo (ICME), July 11-15, 2011.

Thanks to the added dimension of depth information to the traditional 2D video, 3D video provides viewing conditions close to the visual perception with Human Visual System (HVS). This enables more natural conditions for human interaction and enhances the Quality of Experience (QoE) of video consumption. Therefore, the demand for 3D video applications is increasing rapidly with the advance of 3D related technologies. However, 3D video communication applications may be impaired due to the limited bandwidth available in transmission channels and their time-variant quality. Artifacts caused by compression and transmission may be less tolerable in 3D viewing than 2D video due to the complex nature of 3D video formats and rendering processes. Therefore, in order to deliver good quality 3D video content to consumers over time varying channels, 3D video compression and transmission system parameters need to be adapted "on the fly". The measured 3D video quality at the receiver-side can be utilized as feedback information for such adjustment of compression and transmission system parameters. This paper proposes a real-time quality assessment method for depth maps associated with color plus depth map based 3D video.

The color plus depth based 3D video format is widely investigated and applied in research, standardization and industry due to its flexibility (e.g., backward compatibility), efficiency (high compression ratio) and ability to render novel views at low bitrates [1][2][3]. The depth map of a color plus depth 3D video sequence determines the position of the corresponding color image in the 3D space. This process is known as Depth Image-Based-Rendering (DIBR) [4]. With the DIBR method, the quality of depth maps is crucial to render good quality novel views and inaccurate predictions may cause discomfort to viewers. The studies carried out in [5][6][7] show the effect of degraded depth maps on 3D video quality. This paper investigates the real-time quality assessment of depth maps (associated with color plus depth map 3D video) which is essential to update the compression and system parameters for the depth map component of 3D video.

The effect of depth map artifacts that are introduced due to compression and transmission on 3D viewing can be measured subjectively as well as objectively. Even though subjective judgments remain the golden standard of 3D video quality evaluation due to their ability to get better judgment, their use in real-time 3D video applications is very limited. The effect of depth map impairments on image quality can be quantified using objective quality metrics as described in this paper. These could be either objective measures of the rendered left and right views or individual quality measures of color and depth map images. However, most of these objective image quality metrics, such as PSNR and SSIM, are Full-Reference quality metrics which need the original image sequence to obtain quality ratings with respect to it. Therefore, deployment of Full-Reference quality metrics in real-time applications is very limited. The alternative is to use Reduced-Reference (RR) and No-Reference (NR) quality metrics. The former extract features from the original and processed images (i.e., side-information) to evaluate quality in comparison and are more reliable than the latter methods, which solely depend on the received image information and system data (e.g., PLR). Due to limited amount of data required as side-information or no overhead for side-information (with No-Reference metrics) these metrics can be employed in real-time quality evaluations. Several No-Reference and Reduced-Reference quality metrics are reported in the literature for conventional 2D video [8][9], only a few studies are reported for 3D video [10][11].

The Reduced-Reference method proposed in the

paper utilizes specific characteristics of depth maps to quantify quality degradations of the depth map. The boundaries of depth map objects represent boundaries between different depth planes. Therefore, in the proposed method edge information (i.e., depth map boundaries) has been utilized to quantify the structural degradation of depth maps. The edge information generated from the original depth map will be sent to the receiver-side as the side-information required for Reduced-Reference quality evaluation. However, edge information based structural information alone will not be able to measure the impairments in the areas where there are no significant edges. Therefore in the proposed method, two other comparisons namely luminance and contrast comparisons have been introduced to tackle the pixel intensity and pixel variation changes in depth map images. These two comparisons (values based on statistical calculations) will be calculated at the sender-side and will be sent with the edge information generated for original depth map as side-information. The overhead is minimum in this case since edge information can be sent as 1 bit per pixel (and possibly further compressed) and side-information for luminance and contrast comparisons will cost only a few bytes per depth map frame.

The SSIM metric proposed by Wang, et al. [12] has been adapted to implement all three comparisons (structural, luminance and contrast) in the metric. In the proposed method, structural comparison is performed on the edge information generated from the original and received depth maps whereas in the original SSIM implementation this is calculated on the pixel domain of depth maps.

The performance of the proposed Reduced-Reference method is compared with its counterpart Full-Reference SSIM method, where original and received depth maps are used. The experiments are carried out with four different 3D depth map sequences from the Ballet, Orbi, Breakdance and Interview 3D video sequences. Both compression and packet loss impairments are considered for the evaluation. The results show a higher degree of accuracy with the proposed method in comparison to the Full-Reference SSIM method regardless of sequence type, compression level and PLR. The overall correlation coefficient (R-square) between the proposed and the Full-Reference methods is about 0.98. This shows the proposed method can achieve performances similar to that of the Full-Reference method with a lower overhead for side-information. This is also an advancement from the Reduced-Reference quality metric proposed by the authors in [11], where the quality is measured solely based on the edge information extracted from the original and received/processed depth maps. While [11] suggests to use lookup tables to predict the quality at different compression levels and for sequence categories, in the proposed Reduced-Reference method, the accuracy is high compared to its Full-reference counterpart regardless of the sequence type, compression level and packet loss rate being used.

In the near future, most of the conventional video communication applications will be updated into next generation multimedia content such as 3D video. In order to realize, good quality 3D video over unreliable and band-limited communication channels, it is essential to update the compression and transmission system parameters "on-the-fly". In order to measure 3D video quality at the receiver-side with minimum overhead for side-information from the sender-side, the quality evaluation methods as described in this paper will be immensely useful.

References:

[1] 3D Media Cluster, Three-Dimensional Community, www.3dmedia-cluster.eu/, May 2008.

[2] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, "Multi-View Video Plus Depth Representation and Coding, *IEEE International Conference on Image Processing* (ICIP'07), pp. 201 – 204, San Antonio, TX, Oct 2007.

[3] ISO/IEC JTC1/SC29/WG11, "Committee Draft of ISO/IEC 23002-3 Auxiliary Video Data Representations," Doc. N8038, 2006.

[4] C. Fehn, "Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV", in *Proc. SPIE Conf. Stereoscopic Displays and Virtual Reality Systems XI*, vol. 5291, pp. 93 – 104, CA, U.S.A., Jan. 2004.

[5] C.T.E.R. Hewage, S.T. Worrall, S. Dogan, S. Villette, and A.M. Kondoz, "Quality Evaluation of Color Plus Depth Map-Based Stereoscopic Video", *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 2, pp. 304-318, 2009.

[6] C.T.E.R. Hewage, S. Worrall, S. Dogan, H. Kodikara Arachchi, and A.M. Kondoz,

"Stereoscopic TV over IP", *Proceedings of the 4th IET European Conference on Visual Media Production* (CVMP 2007), pp. 1 - 7, London, UK, Nov. 2007.

[7] A. Tikanmäki, A. Gotchev, A. Smolic, and K. Müller, "Quality Assessment of 3D Video in Rate Allocation Experiments", *Proc. IEEE International Symposium on Consumer Electronics* (ISCE'08), pp. 1 - 4, Algarve, Portugal, April 2008.

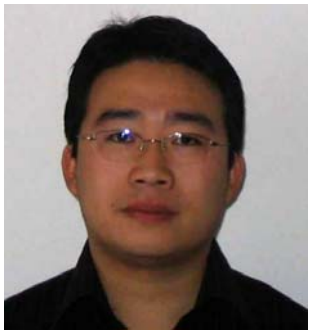
[8] Z. Wang, H. R. Sheikh and A. C. Bovik, "No-Reference Perceptual Quality Assessment of JPEG Compressed Images", *Proc. IEEE Int. Conference on Image Processing*, pp. 477 - 480, Rochester, New York, Sept. 22-25, 2002.

[9] S. Wolf, and M. H. Pinson, "Low Bandwidth Reduced Reference Video Quality Monitoring System", *First International Workshop on Video Processing and Quality Metrics for Consumer Electronics, Scottsdale*, Arizona, Jan. 2005.

[10] M. Solh and G. AlRegib, "A No-reference Quality Measure for DIBR-Based 3D Videos" *IEEE International Workshop on Hot Topics in 3D*, Barcelona, Spain, July 11-15, 2011.

[11] C. Hewage and M.G. Martini, "Reduced-reference quality metric for 3D depth map transmission", in *3DTV-Conference*: The True Vision - Capture, Transmission and Display of 3D Video, pp. 1 – 4, Tampere, Finland, June 2010.

[12] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image Quality Assessment: From Error Measurement to Structural Similarity", *IEEE Transactions on Image Processing*, vol. 13, no.4, pp. 600 - 613, April 2004.



Tao Liu received his B.S. degree in Electrical Engineering from Beijing Institute of Technology in 2004, and received his M.S. and Ph.D. degrees in Electrical Engineering from Polytechnic Institute of New York University in 2008 and 2010, respectively. He has conducted research in the fields of image and video quality assessment, pattern recognition, and video analysis and retrieval for organizations including Technicolor, AT&T, and Dialogic. The content-based video copy detection system he jointly developed at AT&T research lab was ranked among the top submissions at 2009 TREC Video Retrieval Evaluation (TRECVID). Currently, he is a research engineer at Dialogic Media Lab, NJ. He is an active participant in the Video Quality Experts Group (VQEG) and is a member of the IEEE. He is a member of MMTC and MMTC Review Board. He has been a TPC member or referee for various international conferences and journals in the field of image/video processing. He is a co-chair of 2011 IEEE Workshop on Multimedia Quality of Experience: Modeling, Evaluation, and Directions.

A Decentralized Robust Transmission Solution for Multiple Cameras

A short review for "Robust distributed multi-view video compression for wireless camera networks"

Edited by Vladan Velisavljević

C. Yeo and K. Ramchandran, "Robust Distributed Multi-view Video Compression for Wireless Camera Networks," IEEE Transactions on Image Processing, Vol. 19, No. 4, pp. 995-1008, April 2010.

The recent development and deployment of wireless sensor networks [1] and the availability of small and cheap CMOS camera chips has enabled the vision of populating the world with networked wireless video camera sensors. Such systems can be used for a wide variety of applications ranging from surveillance to entertainment. For example, a multiple view camera system can improve tracking performance by being able to overcome the effects of occlusion [2]. Similarly, free viewpoint TV and 3-D TV [3] [4] applications can also benefit from the simple deployment of dense networks of wireless cameras.

However, such a setup faces numerous challenges. The wireless environment poses bandwidth constraints and channel loss, while the sensor mote platform limits processing capability and battery life [1]. In applications such as real-time surveillance, there are very stringent end-to-end delay requirements, which impose tight latency constraints on the system. Furthermore, the cameras may be unable to communicate freely with each other, making it difficult to utilize inter-view correlation between cameras with overlapping views.

To tackle the above challenges, this paper recognizes that cameras with overlapping views provide redundancy, and develops a systematic approach that can effectively harness this redundancy for robust video transmission with completely distributed encoders while meeting tight latency constraints. In developing this approach, the authors exploit several results from information theory and computer vision.

Two important features have a key influence to the system. First, the encoder at each camera does not have access to views observed from other cameras. For that reason, the novel proposed method is based on the PRISM framework [5] of distributed source coding (DSC). The method is able to make statistical use of side-information (predictors) from other camera views for decoding even if they are not physically available at the encoder. Further, such an approach does not require the knowledge of explicit correspondence information between camera views at the encoder.

Second, the proposed coding approach models both the statistical relationship and the geometrical constraints between multiple camera views. In the paper, the authors describe two such models. The first model uses disparity estimation and view interpolation to generate side-information for decoding and it requires two other camera views at the decoder. The second model, in turn, requires only one other camera view at the decoder, and it uses epipolar constraints to generate side-information for decoding.

The corresponding encoding and decoding systems use inter-view statistical models within the PRISM framework. During the coding process, the encoder does not need to know block correspondences or the locations of other cameras; instead, the decoder performs motion or correspondence search. Due to the use of DSC rather than differential coding, this approach is also robust to transmission errors since drift can be mitigated even if the encoder and decoder do not have exactly the same predictors. Unlike in some other distributed video coding schemes (e.g. [6]), the temporal frame interpolation is not used to generate side-information at the decoder and feedback-based rate control is not required. Therefore, this approach is able to meet tight latency constraints. Furthermore, the use of decoder motion search is crucial in enabling both view synthesis search and disparity search to exploit redundancy between overlapping camera views for robustness.

In simulations, the authors show that with these two models, the proposed approaches are able to effectively exploit the redundancy in overlapping views for robustness. Moreover, the proposed approaches still allow for low complexity and fully distributed encoding while meeting tight latency constraints of no more than one frame

delay. The presented results also exhibit behavior of the performance when the camera network parameters are changed (e.g. distance between the cameras, density of the cameras, etc.). In particular, as the number of available neighboring views increases, the system becomes more robust, but with a diminishing gain. Furthermore, as the distance between neighboring views increases, the performance of the system is expectedly affected. Here, the reconstruction quality is significantly reduced for the view synthesis-based approach.

Interesting directions for future work include exploring "smarter" encoders that are able to estimate inter-view correlation based on intra-view properties such as edge strength. The regime of low frame rate video also promises to be an interesting area of research, since inter-view correlation could possibly dominate intra-view temporal correlation.

References:

[1] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, "Wireless sensor networks for habitat monitoring," in *Proc. ACM International Workshop on Wireless sensor networks and applications*, 2002, pp. 88–97.

[2] S. L. Dockstader and A. M. Tekalp, "Multiple camera tracking of interacting and occluded human motion," *Proc. of the IEEE*, vol. 89, no. 10, pp. 1441–1455, Oct 2001.

[3] W. Matusik and H. Pfister, "3D TV: a scalable system for realtime acquisition, transmission, and autostereoscopic display of dynamic scenes," *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 814–824, Aug 2004.

[4] A. Kubota, A. Smolic, M. Magnor, M. Tanimoto, T. Chen, and C. Zhang, "Multiview imaging and 3DTV," *IEEE Signal Processing Magazine*, vol. 24, no. 6, pp. 10–21, Nov. 2007.

[5] R. Puri, A. Majumdar, and K. Ramchandran, "PRISM: A Video Coding Paradigm With Motion Estimation at the Decoder," *IEEE Transactions on Image Processing*, vol. 16, no. 10, pp. 2436–2448, 2007.

[6] B. Girod, A. M. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proc. of the IEEE*, vol. 93, no. 1, pp. 71–83, Jan 2005..

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 ICIP-2011 on compression of high-dimensional media data for interactive navigation. His research interests include image, video and multiview video compression and processing, wavelet theory, multi-resolution signal processing and distributed image/video processing

An Automatic View Synthesis Algorithm for Multi-View-Plus-Depth Videos

A short review for "Depth image-based rendering with advanced texture synthesis for 3-D video"

Edited by Cheng-Hsin Hsu

P. Ndjiki-Nya, M. Koppel, D. Doshkov, H. Lakshman, P. Merkle, K. Muller, and T. Wiegand, "Depth Image-Based Rendering With Advanced Texture Synthesis for 3-D Video," IEEE Transactions on Multimedia, vol.13, no.3, pp. 453-465, June 2011.

The current 3-D cinema produces 2-way stereo videos on large displays that require the use of eyeglasses. In contrast, autostereoscopic multiview screens concurrently show several views from slightly different angles, and each viewer sees the two most suitable views depending on his/her position relative to the display. Displays with many views, e.g., 50-view, will soon be available [1]. However, capturing, encoding, storing, and transporting all views for such displays consumes tremendous amount of resources, and thus incur high costs, which render commercial services less viable.

A more cost-effective approach to support multiview displays is to dynamically synthesize additional views from a few original views using projective geometry. Depth Image-Based Rendering (DIBR) is a technique that uses image texture and depth map, carried by multiview-video-plus-depth videos, to synthesize additional views. While DIBR may largely reduce the number of the original views, and cut down the cost, the synthesized views may suffer from degraded visual quality. This is because certain background regions may be occluded by foreground objects in all the considered original views. These occluded regions form holes in the resulting views, which are visually unappealing. Hence, concealment algorithm, better known as hole-filling algorithms are required for good user experience.

This paper presents a new hole-filling approach for autostereoscopic multiview displays. The proposed view synthesis approach maintains spatial and temporal consistency in the synthesized views. The spatial consistency is achieved by filling holes with neighboring pixels, and the temporal consistency is achieved using a background overlay, which stores the background regions and texture images in previous frames. The synthesis approach consists of three steps.

First, the holes of the current depth map are filled with depth values inferred from neighboring, known depth values. The inference is done by a k-means clustering algorithm, which can reduce the chance to produce artifacts in the depth map. The filled depth map is used to update the background overlay in the next step.

Second, the background overlay is updated using a content-adaptive threshold based on the current frame; pixels with depth values smaller than this threshold are considered parts of the background. By keeping the background overlay across frames, the proposed approach achieves temporal consistency. Then, the texture image stored in the background overlay is used to fill the holes in the current frame. Note that depth maps are often unreliable, and thus even after this step; there may be still holes in the current frame, which will be filled in the next step.

Third, the remaining holes in the current frame are filled using neighboring pixels for spatial consistency. In particular, the holes are first filled using simple filter for initial pixel interference, which is followed by a patch-based advanced texture synthesis for refinements. The third step leads to the hole-filled current frame, which is used to update the background overlay. The proposed synthesis approach then iterates to the next frame and starts from step 1.

The proposed view synthesis approach is implemented and evaluated in Matlab. Four multiview videos are considered with a resolution up to 1024x768 pixels. In each scenario, a new view is synthesized using one or two original views, which are randomly chosen and may not be adjacent. The proposed approach is compared against View Synthesis Reference Software (VSRS) [2] and Fehn [3]. Two video quality metrics are considered: PSNR and SSIM. PSNR is computed within holes to avoid underestimating the video quality caused by geometric distortions that often occur in synthesized videos [4]. While SSIM is more resilient to geometric distortions, it cannot be applied to arbitrary shapes such as holes filled by

the algorithms. Hence, the SSIM is computed across whole synthesized frames.

The authors first empirically derive the optimum system parameters, such as search area and patch size, using experiments. In most of the experiments, the implications of varying system parameters are insignificant in terms of PSNR and SSIM values, and subjective tests are used to pick the optimum system parameters. The authors then compare their synthesis algorithm against the other two algorithms: VSRS [2] and Fehn [3]. The proposed approach outperforms Fehn in all test scenarios, and VSRS in 75% of the test scenarios. Some subjective comparisons indicate that VSRS leads to blurred edges, compared to the proposed approach. Hence, the authors conclude that the view synthesis approach proposed in this work achieves better performance than other state-of-the-art algorithms. Last, the authors report the running time of their approach on an Intel Xeon PC. With the optimum system parameters, synthesizing a frame takes at least 4 minutes to complete. In some scenarios, the per-frame running time could be as high as 67 minutes. The experimental results reveal two open problems in the video synthesis algorithms: (i) the existing objective quality metrics are not suitable to synthesized videos, and (ii) the excessive running time may prevent the proposed approach from being adopted in real systems.

Signal-based video quality metrics, such as PSNR, are too sensitive to the geometric distortions and could underestimate the video quality; while structure-based video quality metrics, such as SSIM, are not sensitive enough to small artifacts such as holes, and could overestimate the video quality. Hence, the experimental studies in this paper have to resort to subjective tests, which are costly and time-consuming. One possible solution is to design a hybrid quality metric with a weighted sum of PSNR and SSIM. However, determining the weight itself becomes a tough problem. The challenge of designing a good quality metric for synthesis videos is even more severe if we take the network transport into considerations; e.g., Shu et. al [5] recently propose a hybrid streaming quality metric combining video quality and rendering time.

While the projective geometry used in DIBR runs much faster than 3-D rendering, often in the order of ms per frame [6], the hole-filling algorithms, such as the one proposed in this paper could lead to high time complexity, probably due to the huge search space allowed by the filling techniques. Clearly, there exists a tradeoff between video quality and time complexity. How to systematically derive the best tradeoff between these two important performance metrics is a challenging issue.

Last, this paper does not consider the stream transport problem for multiview videos. Multiview videos result in a huge amount of data traffic, and could easily overload the existing and next-generation networks. Recently, we started to see more systems papers optimizing encoding [7] and streaming [8] of multiview videos. Given that autostereoscopic multiview displays allow viewers to watch 3-D videos in a more nature setup, e.g., without eyeglasses, we certainly hope to see more works along this direction.

References:

[1] K. Muller, P. Merkle, and T. Wiegand, "3-D Video Representation Using Depth Maps," *Proceedings of the IEEE*, vol. 99, no. 4, pp. 643–656, Apr. 2011.

[2] M. Tanimoto, T. Fujii, and K. Suzuki, "View Synthesis Algorithm in View Synthesis Reference Software 2.0 (VSRS 2.0)," ISO/IEC JTC1/SC29/WG11 M16090, Feb. 2008.

[3] C. Fehn, "Depth-Image-Based Rendering (DIBR), Compression, and Transmission for A New Approach on 3D-TV," In *Proc. of SPIE Stereoscopic Display Virtual Reality System*, San Jose, CA, Jan. 2004, pp. 93-104.

[4] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image Quality Assessment: From Error Visibility to Structure Similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[5] S. Shi, K. Nahrstedt, R. Campbell, "Distortion over Latency: Novel Metric for Measuring Interactive Performance in Remote Rendering Systems," In *Proc. of IEEE International Conference on Multimedia and Expo* (ICME'11), Barcelona, Spain, Jul. 2011, pp. 1-6.

[6] W. Yoo, S. Shi, W. Jeon, K. Nahrstedt, and R. Campbell, "Real-time Parallel Remote Rendering for Mobile Devices Using Graphics Processing Units," In *Proc. of IEEE International Conference on Multimedia and Expo* (ICME'10), Singapore, Jul. 2010, pp. 902-907.

[7] W. Feng, F. Liu, Y. Niu, and S. Price, "Systems Support for Stereoscopic Video Compression," In *Proc. of International Workshop on Network and Operating Systems Support for Digital Audio and*

Video (NOSSDAV'11), Vancouver, Canada, Jun. 2010, pp. 99-104.

[8] A. Hamza and M. Hefeeda, "Multicasting of Multiview 3D Videos over Wireless Networks," In *Proc. of ACM Workshop on Mobile Video* (MoVid'12), Chapel Hill, NC, Feb. 2012.

Cheng-Hsin Hsu received the Ph.D. degree from Simon Fraser University, Canada in 2009, the M.Eng. degree from University of Maryland, College Park in 2003, and the M.Sc. and B.Sc. degrees from National Chung-Cheng University, Taiwan in 2000 and 1996, respectively. He is an Assistant Professor in Department of Computer Science at National Tsing Hua University, Taiwan. He was a Senior Research Scientist at Deutsche Telekom R&D Lab USA, Los Altos, CA between 2009 and 2011. His research interests are in the area of multimedia networking and distributed systems. He has published more than 45 papers in leading journals, conferences, and workshops. He and his colleagues at Simon Fraser University developed a mobile TV testbed, which won the Best Technical Demo Award in the ACM Multimedia 2008 Conference. He is on the Review Board Committee of IEEE Technical Committee on Multimedia Communications (MMTC) and the Preservation Committee of ACM Special Interest Group on Multimedia (SIGMM). He served as the TPC Co-chair of the ACM Mobile Video Delivery Workshop (MoViD'11) and the Proceeding and Web Chair of the ACM International Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV 2010). He was on the technical program committees of several well-known conferences in his research areas, including ACM Multimedia Conference (Multimedia), ACM International Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV), and IEEE International Conference on Multimedia and Expo (ICME).

Moving Region Segmentation from Compressed Video

A short review for "Moving region segmentation from compressed video using global motion estimation and Markov random fields"

Edited by Hassan Mansour

Y.-M. Chen, I. V. Bajić, and P. Saeedi, "Moving Region Segmentation from Compressed Video Using Global Motion Estimation and Markov Random Fields," IEEE Transactions on Multimedia, vol. 13, no. 3, pp. 421-431, June 2011.

Moving object segmentation is a useful technique in a variety of applications, such as video surveillance, video database browsing, object-based video transcoding, video analysis and summarization, frame prediction, visual saliency estimation, etc. Segmentation is usually performed in the pixel domain, exploiting visual features such as shape, color and texture [1][2][3], or in the compressed domain, exploiting compression syntax and information to facilitate segmentation [4][5][6]. The former group of methods is characterized by generally higher accuracy and much higher complexity compared to the latter ones. Hybrid approaches, combining compressed-domain and pixel-domain operations in an effort to take the best of both worlds, have also been reported [7][8][9]. It should be noted, however, that almost all video content is available only in the compressed form, at least to the end users, so it is a matter of good engineering practice to make use of the already available compressed-domain information, even if the final goal is to produce pixel-wise segmentation boundaries in the decoded frames.

This paper presents a hybrid compressed- and pixel-domain segmentation framework for moving objects in compressed video. First, global motion estimation (GME) is performed directly on the motion vectors (MVs) from the compressed stream. Then the camera motion is removed from the MV field global motion compensation (GMC). Once the MV field is free of the influence of camera motion, MV vector quantization (VQ) based on local motion similarity is used to find the most likely number of moving regions. The statistics of the VQ clusters are used to initialize prior probabilities for subsequent Markov Random Field (MRF) classification, which produces a coarse segmentation map. In the final step, a coarse-to-fine strategy is utilized to refine region boundaries in the pixel domain.

Maintaining segmentation consistency, i.e., the number of moving regions and their motion statistics through the sequence, is the main difficulty in unsupervised segmentation, mostly caused by inaccurate MVs. This issue often manifests itself in the final result as over- or under-segmented moving regions. The proposed approach tackles this issue by removing camera motion and performing vector quantization (VQ) on decoded MVs. In order to remove camera motion, GME with 8-parameter perspective motion model is performed directly on the MV field using least squares regression with a M-estimator [10]. VQ is then carried out on the global motion compensated MV field using the generalized Lloyd algorithm. Multiple motion clusters are formed iteratively by minimizing the overall weighted distortion from all clusters. The MV statistics are finally computed from each cluster as initial MRF parameters.

The MRF motion model [3][6] is the core of MRF segmentation. In the adopted MRF motion model, MVs within a given moving region follow a bivariate Gaussian distribution, while region labels follow a 2-D Gibbs distribution based on a given neighborhood system. The goal of MRF segmentation is to create more compact clusters compared to the initial segmentation provided by VQ, by virtue of imposing the spatial constraints through the chosen neighborhood system. The segmentation that maximizes the posterior probability is found through the method of Iterated Conditional Modes (ICM).

Up to this point, all processing is done only on the MVs, and no texture decoding is needed. The resulting segmentation boundaries, however, follow block boundaries. In order to improve boundary localization, one must move to the pixel domain. Here, a coarse-to-fine strategy [9] is adopted, whereby the boundary location is refined starting from the block-wise boundaries obtained thus far.

Fine segmentation in the pixel domain utilizes visual features, such as edges and color, to obtain pixel-wise boundaries. The complexity is kept in check by performing boundary refinement only near predicted block-wise boundaries. First, each block is classified as either being a part of object interior or boundary. A Canny edge detector is used on the Y-component to identify edges within boundary blocks. Then, a region growing procedure is applied to grow interior regions towards each other via morphological erosion. This restricted erosion will move the interior region boundaries up to the nearest edge. If needed, further region growing based on color is applied until interior region boundaries finally meet.

In summary, this paper offers an efficient moving region segmentation framework, capitalizing on compressed-domain information available in any end-user application. The experimental results show improvements on mitigating under- and over-segmentation compared to state-of-the-art methods, and promising performance on maintaining moving region consistency in un-supervised MRF segmentation. The proposed method also delivers a good balance between accuracy and complexity, and could be applied to a broad range of video applications.

References:

[1] O. Brouard, F. Delannay, V. Ricordel, and D. Barba, "Spatio-temporal segmentation and regions tracking of high definition video sequences based on a Markov Random Field model," *Proc. IEEE ICIP*'08, pp. 1552-1555, Dec. 2008.

[2] Y. Deng, and B. S. Manjunath, "Unsupervised segmentation of color-texture regions in images and video," *IEEE Trans. on Pattern Anal. Mach. Intell*., vol. 23, issue 8, pp. 800-810, Aug. 2001.

[3] Z. Kato, T. C. Pong, and J. C. M. Lee. "Color Image Segmentation and Parameter Estimation in a Markovian Framework*," Pattern Recognition Letters*, 22(3-4):309--321, Mar. 2001.

[4] W. Zeng, J. Du, W. Gao, and Q. Huang, "Robust moving object segmentation on H.264/AVC compressed video using the block-based MRF model," *Real-Time Imaging*, vol. 11, pp. 290-299, Jun. 2005.

[5] M. Ritch and N. Canagarajah, "Motion-based video object tracking in the compressed domain," *Proc. IEEE ICIP*'07, vol. 6, pp. VI-301-VI-304, Oct. 2007.

[6] N. Vasconcelos and A. Lippman, "Empirical Bayesian motion segmentation*," IEEE Trans. on Pattern Anal. Mach. Intell*., vol. 2, issue 2, pp. 217-221, Feb. 2001.

[7] D. Zhong and S. F. Chang, "An integrated approach for content-based video object segmentation and retrieval," *IEEE Trans. Circuits Syst. Video Technol*., vol. 9, no. 8, pp. 1259-1268, Dec. 1999.

[8] X. Shi, Z. Zhang, L. Shen, "Multiresolution segmentation of video objects in the compression domain," *Optical Engineering*., vol. 46, no. 9, 097401, Sep. 2007.

[9] Y.-M. Chen, I. V. Bajić, and P. Saeedi, "Coarse-to-fine moving region segmentation in compressed video," *Proc. IEEE WIAMIS'09*, pp. 45-48, London, UK, May 2009.

[10] A. Smolić, M. Hoeynck, and J.-R. Ohm, "Low-complexity global motion estimation from P-frame motion vectors for MPEG-7 application," *Proc. IEEE ICIP*'00, pp. 271-274, Sep. 2000.

Hassan Mansour received his B.E. degree in Computer and Communications Engineering from the American University of Beirut in 2003, his M.A.Sc. and Ph.D. in Electrical and Computer Engineering from the University of British Columbia in 2005 and 2009, respectively. Since January 2010, he has been a postdoctoral research fellow in the Mathematics and Computer Science departments at the University of British Columbia.

His research interests include the theory and application of compressed sensing, sparse signal reconstruction algorithms utilizing prior information, seismic imaging, scalable video coding (SVC), stereo image processing, and high dynamic range (HDR) image and video compression.

Delay Analysis for Multi-Hop Wireless Networks

A short review for "Delay analysis and optimality of scheduling policies for multi-hop wireless networks"

Edited by Ai-Chun Pang

*G. R. Gupta, and N. Shroff, "*Delay Analysis and Optimality of Scheduling Policies for Multi-Hop Wireless Networks*," IEEE Transactions on Networking, vol. 19, no. 1, pp. 129-141, February 2011.*

There exist a considerable amount of research [1, 2] on multi-hop wireless networks that focuses on maintaining system stability and maximizing first-order metrics, such as throughput or utility. However, the arguably more important problem of delay optimal scheduling remains largely open, except for specialized scenarios and interference topologies. Even well-known throughput-optimal algorithms like back pressure, suffer from large delays [3]. Furthermore, for a large class of applications such as video or voice over IP, embedded network control and for system design; delay analysis is of prime importance. Due to the limited amount of literature on delay analysis and delay-optimal scheduling problem for general topology wireless networks, the results in the paper can be considered as a quantum advance in the state-of-the-art.

Many throughput-optimal algorithms make their scheduling decisions based on the backlog in the system, which in turn depends on past scheduling decisions and arrival rates. Such cross-dependency results in system dynamics that are difficult to analyze. Due to this reason, the behavior of throughput-optimal algorithms in terms of finer QoS performance metrics, such as mean delay or probability of buffer overflow, is difficult to quantify. Most results have been restricted to single-hop traffic using order optimal mean delay analysis [4], heavy traffic analysis [5] and large deviations [6]. Traditional heavy traffic results have focused on a single bottleneck in the system and proving a state-space collapse.

The authors devise novel analytical techniques to relax the interference constraints in the network to derive lower bounds. They develop a queue grouping technique to handle the complex correlations of the service process resulting from the multi-hop nature of the flows, thus circumventing the decades-long open problem of characterizing the departure process of a queue. They also introduce a novel concept of (K, X)-bottlenecks in the network and derive sample path bounds for a group of queues upstream of a bottleneck. They also derive a fundamental lower bound on the system-wide average queuing delay of a packet in multi-hop wireless network, regardless of the scheduling policy used. The lower bound can be used for analyzing a large class of arrival processes using known results in the queuing literature.

For a network under the primary interference model, their lower bound is tight in an asymptotic sense. For a tandem queueing network, the average delay of a delay optimal policy proposed by [7] numerically coincides with the lower bound provided in this paper. A clique network is a special graph where at most one link can be scheduled at any given time. Using existing results on work conserving queues, they design a delay optimal policy for a clique network and compare it to the lower bound. They also performed extensive experiments for the comparison of the lower bound derived in the paper and the average delay measured from the simulations. The lower bound is observed to be accurate via extensive numerical results. Thus, it provides a benchmark for designing delay optimal schedulers in general.

References:

[1] L. Tassiulas and A. Ephremides, "Stability Properties of Constrained Queueing Systems and Scheduling Policies for Maximum Throughput in Multihop Radio Networks," *IEEE Transactions on Automatic Control*, vol. 37, no. 12, pp. 1936–1948, December 1992.

[2] S. Shakkottai, R. Srikant, and A. Stolyar, "Pathwise Optimality of the Exponential Scheduling Rule for Wireless Channels," *Advances in Applied Probability*, pp. 1021–1045, December 2004.

[3] L. Ying, S. Shakkottai, and A. Reddy, "On Combining Shortest-Path and Back-Pressure Routing Over Multihop Wireless Networks," in *Proceedings of IEEE INFOCOM*, April 2009.

[4] Michael Neely, "Delay Analysis for Maximal Scheduling in Wireless Networks with Bursty

Traffic," in *Proceedings of IEEE INFOCOM*, April 2008.
[5] A. L. Stolyar, "MaxWeight Scheduling in a Generalized Switch: State Space Collapse and Workload Minimization in Heavy Traffic," *Annals of Applied Probability*, vol. 14, no. 1, pp. 1–53, 2004.
[6] V. J. Venkataramanan and X. Lin, "On Wireless Scheduling Algorithms for Minimizing the Queue-Overflow Probability," *IEEE/ACM Trans. On Networking*, vol. 18, no. 3, June 2010.
[7] L. Tassiulas and A. Ephremides. Dynamic scheduling for minimum delay in tandem and parallel constrained queueing models. *Annals of Operation Research*, Vol. 48, 333-355, 1993.

Ai-Chun Pang received the B.S., M.S. and Ph.D. degrees in Computer Science and Information Engineering from National Chiao Tung University, Taiwan, in 1996, 1998 and 2002, respectively. She joined the Department of Computer Science and Information Engineering (CSIE), National Taiwan University (NTU), Taipei, Taiwan, in 2002. Currently, she is a Professor in CSIE and Graduate Institute of Networking and Multimedia of NTU, Taipei, Taiwan. Her research interests include wireless networking, mobile computing, and performance modeling. She is a Senior Member of IEEE.

Wireless Physical Layer Security: The Multicast Transmit Beamforming Case

A short review for "Multicast secrecy rate maximization for MISO channels with multiple multi-antenna eavesdroppers"

Edited by Walid Saad

Q. Li and W.-K. Ma, "Multicast Secrecy Rate Maximization for MISO Channels with Multiple Multi-antenna Eavesdroppers" in Proceedings of 2011 IEEE International Conference on Communications (ICC'11), June 2011.

Physical-layer multicasting refers to downlink transmission techniques for broadcasting common information to a specific group of users. For example, in a multi-antenna base station setup, a common multicasting technique is the so-called "multicast transmit beamforming", in which the users are simultaneously beamformed and served [1]. Such multicast techniques provide resource-efficient means to deliver multimedia applications, such as mobile TV services or multicast streaming services. An overview of the services associated with multicasting has been considered, and implemented, in several wireless standards such as LTE and WiMax [2-3].

The ability to multicast innovative services is also accompanied with several challenges, notably from a security perspective. For instance, preventing unauthorized users from receiving the multicast information for free is an important problem that wireless operators need to deal with. With the ongoing decentralization of wireless networks, classical encryption techniques can become difficult to implement. For instance, secret key generation and management are becoming more and more difficult, especially for ad-hoc wireless networks which exhibit dynamic topologies. In this respect, recently, there has been an increased interest in investigating information security from an alternative point of view, particularly, physical-layer secrecy or information theoretic security. Physical-layer secrecy focuses on physical-layer-based signaling or coding techniques, wherein the rationale is to judiciously encode the confidential message into random symbols such that the legitimate users can correctly decode the message, while malicious eavesdroppers retrieve almost no information from their observations [4]. Two key notions in physical-layer secrecy are the secrecy capacity and the best achievable secrecy rate, which essentially relate to the maximum data rate at which the confidential information can be securely transmitted. While the concept of physical-layer secrecy can be traced back to the 1970's in Wyner's seminal work [4], the degree of freedom in modern wireless systems, such as multi-antenna systems, has provided physical-layer secrecy with many new and exciting opportunities in which the secrecy rate or capacity can be substantially increased via appropriate transmit designs.

In the reviewed paper, the authors consider the physical-layer secrecy problem in the context of multicasting. Essentially, the studied setting is that in which a multi-antenna base station needs to service a number of legitimate single-antenna users in the presence of a number of eavedroppers having multiple receive antennas. The authors' contribution lies in transmit design optimization, for which the maximum achievable multicast secrecy rate is sought. The key objectives are to derive fundamental information-theoretic limit of the multicast secrecy rate and to provide novel insights into the secure multicast transmit design problem. The multicast secrecy rate maximization (SRM) problem that is addressed is challenging since it is a non-convex problem with no known closed-form solution. To overcome this difficulty, the authors find a suboptimal, yet tractable, alternative to multicast SRM. In particular, using a convex relaxation of the SRM problem; a semidefinite program (SDP) approximation is derived which provides upper and lower bounds on the maximum achievable multicast secrecy rate. The proposed approach is evaluated using numerical simulations which show that the gaps between the upper and lower bounds are quite small, say, under a scenario of 3-5 legitimate multicast users. This implies that the SDP approach is a good approximation to the optimal achievable multicast secrecy rate.

Another interesting result reported in this paper is that the authors identified instances under which the SDP approximation is tight. They showed by analysis that for a reasonable number of eavesdroppers and legitimate users, the SDP

relaxation is tight, solving the SRM problem optimally. Interestingly enough, the proof also reveals that transmit beamforming is an optimal transmit design for those problem instances.

Hence, the significance of this paper is to provide fundamental insights on physical layer security in transmit multicast scenarios. The formulation is based on the assumption that the base station has perfect knowledge of eavesdroppers' channel state information (CSI). While this assumption is hard to maintain in practice, this work provides researchers with a way to evaluate the best achievable multicast secrecy rate, which is useful for benchmarking and examining performance gaps between perfect and imperfect CSI cases. Also, like many research endeavors in MIMO communications, the perfect CSI-based study by the authors should help pave the way for future research in the imperfect CSI case. Moreover, in cellular network applications where the eavesdroppers are unauthorized users in the same network, it is reasonable to assume that the base station has CSI of such eavesdroppers.

In a nutshell, with the on-going deployment of novel wireless services, it is envisioned that lightweight security techniques such as physical layer security will become more and more abundant. In this respect, studies such as the one provided in this paper are of particular interest since they provide key insights on the fundamental challenges and expectations that researchers must take into account when studying the implementation of physical layer security in next-generation wireless networks.

References:

[1] N. D. Sidiropoulos , T. N. Davidson and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting ," *IEEE Transactions on Signal Processing*, Vol. 54, No. 6, June 2006.

[2] G. Kambourakis, E. Konstantinou and S. Gritzalis, "Revisiting WiMAX MBS security," *Computers and Math. with Applications* 60 (2010) 217-223.

[3] 3GPP TS 33.246: Security of Multimedia Broadcast/Multicast Service (MBMS).

[4] A. D. Wyner, "The wiretap channel," in *The Bell System Technical Journal*, vol. 54, October 1975, pp.1355-1387.

Walid Saad received his B.E. degree in Computer and Communications Engineering from the Lebanese University, Faculty of Engineering, in 2004, his M.E. in Computer and Communications Engineering from the American University of Beirut (AUB) in 2007, and his Ph.D degree from the University of Oslo in 2010. From August 2008 till July 2009 he was a visiting scholar in the Coordinated Science Laboratory at the University of Illinois at Urbana Champaign. From January 2011 till August 2011, he was a Postdoctoral Research Associate at the Electrical and Computer Engineering Department at Princeton University.

Currently, he is an Assistant Professor at the Electrical and Computer Engineering Department at the University of Miami. His research interests include applications of game theory in wireless networks, small cell networks, cognitive radio, wireless communication systems (UMTS, WiMAX, LTE, etc), and smart grids. He was the first author of the papers that received the Best Paper Award at the 7th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt), in June 2009 and at the 5th International Conference on Internet Monitoring and Protection (ICIMP) in May 2010.

Achieving Secure MANET: Fundamental Limits and Practical Schemes

A short review for "Secrecy throughput of MANETs under passive & active attacks"

Edited by Simon Pun

Yingbin Liang, H. Vincent Poor and Lei Ying, " Secrecy throughput of MANETs under passive and active attacks", IEEE Transactions on Information Theory, vol. 57, no.10, pp. 6692-6702, October 2011.

Mobile ad hoc networks (MANETs) have recently drawn considerable research interests thanks to their numerous potential applications in military and emergency networks. In MANETs, autonomous mobile nodes can efficiently communicate with each other even *without* any centralized and fixed network infrastructure, which is particularly attractive for applications that require rapid network deployment. Despite many advantages, MANENTs are susceptible to malicious attacks due to the open nature of their infrastructure and wireless medium. To make security issue more challenging, most traditional cryptographic techniques commonly adopted in *wired* networks are not directly applicable to MANETs. This is because these traditional cryptographic techniques usually rely upon network infrastructure for security key distribution and management. Thus, it is difficult to apply such traditional techniques to infrastructure-less MANETs.

In this paper, the authors proposed a novel approach for designing secure communication strategies (including coding/routing/scheduling schemes) and characterizing the fundamental communication limits for MANETs. The key idea is to establish a connection between end-to-end communication under malicious attacks in MANETs and a wiretap channel [1] studied in information theory; and then apply physical layer coding techniques that achieve the security capacity of the wiretap model to secure MANETs. This approach also enables to characterize the optimal secrecy throughput (an extension of the notion of the throughput studied in [2]) of MANETs based on the information theoretic characterization of the rate limit of the wiretap channel. To the reviewer's best knowledge, this paper stands for the first theoretical results on the fundamental limits of secrecy throughput for MANETs. In the following sections, key innovations in this paper will be highlighted.

The first contribution of this paper is the mathematical identification of equivalent wiretap models developed in the information-theoretic security [1] for MANETs with malicious nodes. For each packet transmitted from a source node, its transmission can be modelled into *two* channels. With probability 1-α, the packet is successfully transmitted to its legitimate destination node. However, the package can be also overheard by malicious nodes with probability 1-β. Thus, the end-to-end packet delivery between each source-destination pair can be modelled by an erasure wiretap channel. Taking advantage of the well-established results on the secrecy capacity [1] and the coding scheme [3] for the erasure wiretap channel, the authors then derived the secure communication protocols and the optimal secrecy throughput for MANETs assuming two types of malicious attack models, namely the passive and active attack models. We will first review the passive attack model before extending our discussions to the active attack model.

In the passive attack model, malicious nodes simply eavesdrop the transmission channel. A transmission channel becomes insecure if one of the malicious nodes successfully intercepts any packet. Through meticulous mathematical derivations, the authors have shown that the secrecy throughput is characterized by the numbers of legitimate and malicious nodes as well as the maximum delay. Denote by n, m and D the number of legitimate and malicious mobile nodes and maximum delay, respectively. The secrecy throughput of a MANET is given by either $\Theta\left(\sqrt{D/n}\right)$ when m is comparable to $\sqrt{nD}$ or $\Theta(1/m)$ when m is on the order of $\sqrt{nD}\,poly(D)$ where $\Theta(\cdot)$ stands for the asymptotic tighter bound of the enclosed quantity or function.

This analytical result provides some very interesting insights about the secrecy throughput of MANETs as well as designing practical capacity-achieving algorithms. In particular, when the number of malicious nodes is small

(i.e. $m = o\left(\sqrt{nD}\right)$), the secrecy throughput is dictated by the delay constraint (i.e. D) and the number of legitimate nodes (i.e. n).

The authors subsequently proposed a two-hop transmission scheme to achieve the secrecy throughput. The scheme consists of three steps, namely (1) stochastic secure coding [1,3] and message interleavin; (2) cell scheduling and (3) two-hop transmission scheme. The three steps work as follows. After applying secure coding and interleaving on the message over a super-time-slot, the two-hop scheme schedules mobile nodes to transmit/relay the packet over the MANET until the destination node successfully receives the packet. This two-hop scheme is shown to achieve the secrecy throughput $\Theta\left(\sqrt{D/n}\right)$.

However, as the number of malicious nodes grows, the number of malicious nodes becomes the dominating factor in secrecy throughput. To cope with this obstacle, the authors proposed to replace the two-hop transmission scheme in Step (3) with a one-hop scheme. It is shown that the resulting secrecy throughput is given by $\Theta(1/m)$.

Finally, the authors extended their results on the passive attack model to the active attack model. In the active attack model, malicious nodes first eavesdrop a package before modifying and sending the modified packet to the destination node. In the paper, it is shown that a MANET under active attack can also be modelled as a wiretap channel. However, in contrast to the passive attack model where legitimate nodes see erasure channels, legitimate nodes under active attack experience binary symmetric erasure channels. Thus, secrecy throughput under the active attack model can be derived in a manner similar to the passive attack model. Interestingly, the theoretical results show that the secrecy throughput is the same for both the active and passive attack models when the number of malicious nodes is large. However, when the number of malicious nodes is small, it requires a more stringent condition on the number of malicious nodes for the active attack model in order to guarantee the same throughput as compared to the passive attack model.

In summary, this paper has developed a new approach for securing MANETs and established theoretical limits on the secrecy throughput of MANETs under both passive and active attacks. Furthermore, it proposed practical schemes to achieve the optimal secrecy throughput. Without any doubt, the results derived in this paper will exert a profound impact upon the future research on secure transmission in MANETs.

References:

[1] A. D. Wyner, "The wire-tap channel," Bell Syst. Tech., vol. 54, pp. 1355–1387, Oct. 1975.

[2] P. Gupta and P. Kumar, "The capacity of wireless networks," IEEE Trans. Inf. Theory, vol. 46, no. 2, pp. 388–404, Mar. 2000.

[3] A. Thangaraj, S. Dihidar, A. Calderbank, S. McLaughlin, and J.-M. Merolla, "Application of LDPC codes to the wiretap channel," IEEE Trans. Inf. Theory, vol. 53, no. 8, pp. 2933–2945, Aug. 2007.

Man-On Pun received the BEng (Hon.) in Electronic Engineering from the Chinese University of Hong Kong in 1996, the MEng. degree in Computer Science from University of Tsukuba, Japan in 1999 and the Ph.D. degree in Electrical Engineering from the University of Southern California (USC) in 2006, respectively. He has been with Huawei's Wireless Solution Department in Bridgewater, NJ since 2011. Previously, he held research positions at Mitsubishi Electric Research Labs (MERL), Cambridge, MA from 2008 to 2011, Princeton University from 2006 to 2008 and Sony Corporation, Tokyo from 1999 to 2001. Dr. Pun received the MERL president's award in 2009 and three best paper awards from Infocom 2009, ICC 2008 and VTC-Fall 2006. He serves as Associate Editor of the IEEE Transactions on Wireless Communications.

Paper Nomination Policy

IEEE MMTC R-letter welcomes review paper nomination. Any paper published in an IEEE ComSoc journal/magazine or in the MMTC sponsored proceedings: IEEE GLOBECOM, ICC and ICME, in the two years preceding the next award board's election, is eligible.

The paper nomination is always open. Paper nominations have to be sent to the IEEE MMTC Review Board Director by email. The nomination should include the complete reference of the paper, author information, a brief supporting statement (maximum one page), the nominator information, and an electronic copy of the paper when possible. Only papers published in the two years preceding the nomination will be considered.

Each nominated paper will be reviewed by two members of the IEEE MMTC Review Board, according to the area of expertise, and avoiding any potential conflict of interest. The reviewer names will be kept confidential. If both members agree that the paper is of award quality, they will recommend publishing the review of the paper (partially based on the nomination supporting document) in the IEEE MMTC Review Letter.

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

R-Letter Editorial Board

MMTC Officers