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

R-LETTER

Vol. 6, No. 5, October 2015

IEEE COMMUNICATIONS SOCIETY

TABLE OF CONTENTS

Message from the Review Board Directors 2
Video Playback Time Maximization for Smartphones 3
A short review for "EQ-Video: Energy and Quota-Aware Video Playback Time Maximization for Smartphones" (Edited by Koichi Adachi) 3
How to share GPU amongst multiple jobs in cloud-based multimedia services? 5
A short review for "VIGRIS: Virtualized GPU Resource Isolation and Scheduling in Cloud Gaming" (Edited by Gwendal Simon) 5
Approaching New Limits of Synchrony with Multi-Sensorial Media 7
A short review for "Perceived Synchronization of Mulsemedia Services" (Edited by Ragnhild Eg and Carsten Griwodz) 7
Improving QoS for Media Streaming in Distributed Software Defined Networks . 10
A review for "Distributed QoS Architectures for Multimedia Streaming Over Software Defined Networks" (Edited by Roger Zimmermann) 10
Content-Oriented Geo Features in Hybrid Video Search 12
A review for "On Generating Content-Oriented Geo Features for Sensor-Rich Outdoor Video Search" (Edited by Pradeep K. Atrey) 12
Paper Nomination Policy 14
MMTC R-Letter Editorial Board 15
Multimedia Communications Technical Committee Officers 15

Message from the Review Board Directors

Welcome to the October 2015 issue of the Review Letter (R-Letter) of the IEEE Communications Society Multimedia Communications Technical Committee (MMTC). This issue comprises **five R-Letters** in the area of video playback time optimization, GPU virtualization for cloud services, mulsemedia, software defined networks, and hybrid video search.

We hope that this issue **stimulates your research in the area of multimedia communication**.

An overview of all R-Letters is provided in the following:

The **first paper**, published in *IEEE Wireless Communications Letter* and *edited by Koichi Adachi*, explores video playback time maximization for smartphones.

The **second paper**, published in the *ACM Transactions on Architecture and Code Optimization (TACO)* and *edited by Gwendal Simon*, discusses how to share GPU amongst multiple jobs in cloud-based multimedia services.

The **third paper** is *edited by Ragnhild Eg and Carsten Griwodz* and has been published within the *Proceedings of IEEE Transactions on Multimedia*. It approaches new limits of synchrony with multi-sensorial media (mulsemedia).

The **forth paper**, published in the *IEEE Transactions on Multimedia* and *edited by Roger Zimmermann*, improves the QoS for media streaming in distributed software defined networks.

Finally, the **fifth paper**, published in the *IEEE Transactions on Multimedia* and *edited by Pradeep K. Atrey*, is about content-oriented geo features in hybrid video search!

We would like to thank all the authors, nominators, reviewers, editors, and others who contribute to the release of this issue. In particular, we would like to thank Weiyi Zhang for his contribution to the IEEE MMTC review board as he resigned as board co-director.

IEEE ComSoc MMTC R-Letter

Director: Christian Timmerer
Alpen-Adria-Universität Klagenfurt, Austria
Email: christian.timmerer@itec.aau.at

Co-Director: Yan Zhang
Simula Research Laboratory, Norway
Email: yanzhang@simula.no

Video Playback Time Maximization for Smartphones

A short review for "EQ-Video: Energy and Quota-Aware Video Playback Time Maximization for Smartphones"
(Edited by Koichi Adachi)

W. Lee, J. Koo, S. Jin, and S. Choi, "EQ-Video: Energy and Quota-Aware Video Playback Time Maximization for Smartphones", IEEE Wireless Communications Letter, vol.19, no.6, pp. 1045-1048, June. 2015.

During recent several years, *smartphone* has drastically gained its popularity in our society. One of the main uses of such smartphone is watching video streaming. Current wireless mobile networks such as 3rd generation partnership project (3GPP) long-term evolution (LTE) and Wi-Fi are capable of such video streaming services and fast enough to support high quality video services including high definition (HD) video. However, such high quality video services require battery energy and data quota reserved for users' data plan. Thus, so called typical data-hungry users cannot enjoy the playback time of multimedia services satisfactorily or energy-hungry smartphones may drain its battery to provide such a high quality video streaming service. To address this problem, video streaming services have been studied in consideration of data usage or energy consumption [1]–[3]. However, smartphone users are generally concerned about both remaining data quota and battery energy. Typical users may stop playing video when either available data quota or energy is about to deplete.

For video streaming service, HTTP-based video streaming technology is usually adopted. This technology prefetches video data to cope with various unfavorable obstacles experienced by users such as time-varying network bandwidth and wireless channel fluctuation. Therefore, if a user decides to stop playing an ongoing video streaming, the prefetched video data will be dropped and this results in the wasting data quota and also the battery energy to prefetch this video data.

In this letter, the authors propose an algorithm to maximize overall video playback time while suppressing the unnecessary battery consumption. To achieve this goal, a normalized cost function in terms of both data usage and energy consumption is designed.

The authors focus on the *chunk download cycle* (*CDC*). Once the amount of unplayed video data in the buffer reaches the predetermined threshold, the client-side video player requests another chunk, whose size is also predetermined, for the next CDC cycle. To maximize the playback time, the period of CDC needs to be adjusted based on the current data quota, battery energy, and user's leaving-in-midstream probability. Intuitively, if the battery energy is running out, the CDC period should be increased. However, this increment may result in more data and energy waste if the user stops playing the video prior to the completion of the entire playback. Therefore, there is a trade-off relationship between the battery saving and the possible data and energy waste.

The authors derive the normalized data cost and the energy cost that are obtained from the expected data and energy wastes for a single CDC. Based on the derived costs, the authors propose a policy for the video playback time maximization, EQ-video, which finds the optimal CDC period for each video playback considering current remaining data quota and battery level.

The performance of the proposed EQ-video is evaluated by computer simulation. The authors successfully show that the proposed EQ-video can balance the unnecessary data download and energy waste in order to increase the total playback time.

The proposed algorithm does not require any protocol change or introduction of new parameter. The proposed algorithm can be easily integrated into already-existing video players if the video players alternatively use an optimal CRC period value returned by the proposed algorithm instead of the predefined value for their video chunk deliveries.

Waste of the data quota means that the precious mobile network resources are used for waste. From the view point of the network utilization, it is quite important to improve the utilization efficiency. The reviewer thinks this proposed algorithm also contribute to the improvement of the network utilization efficiency.

References:

[1] X. Li, M. Dong, Z. Ma, and F. Fernandes, "GreenTube: Power optimization for mobile video streaming via dynamic cache management," in *Proc. ACM MM*, Oct. 2012, pp. 279–288.

[2] M. A. Hoque, M. Siekkinen, and J. K. Nurminen, "Using crowd-sourced viewing statistics to save energy in wireless video streaming," in Proc. *ACM Mobicom*, Oct. 2012, pp. 377–388.
[3] J. Chen, A. Ghosh, J. Magutt, and M. Chiang, "QAVA: Quota aware video adaptation," in *Proc. ACM CoNEXT*, Dec. 2012, pp. 121–132.

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. Since 2010, he has been with the Institute for Infocomm Research, A*STAR, in Singapore. His research interests include cooperative communications and energy efficient communication technologies. 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.

Dr. Adachi served as General Co-chair of the 10th and 11th IEEE Vehicular Technology Society Asia Pacific Wireless Communications Symposium (APWCS) and Track Co-chair of Transmission Technologies and Communication Theory of the 78th and 80th IEEE Vehicular Technology Conference in 2013 and 2014, respectively. He was recognized as the Exemplary Reviewer from IEEE COMMUNICATIONS LETTERS in 2012 and IEEE WIRELESS COMMUNICATIONS LETTERS in 2012, 2013, and 2014. He was awarded excellent editor award from IEEE ComSoc MMTC in 2013.

How to Share GPU Amongst Multiple Jobs in Cloud-based Multimedia Services?

A short review for "VIGRIS: Virtualized GPU Resource Isolation and Scheduling in Cloud Gaming"
(Edited by Gwendal Simon)

Z. Qi, J. Yao, C. Zhang, M. Yu, Z. Yang, and H. Guan, "VGRIS: Virtualized GPU Resource Isolation and Scheduling in Cloud Gaming", ACM Transactions on Architecture and Code Optimization (TACO), vol. 11, no. 2, June 2014.

The multimedia research community has always kept strong ties with the research communities that are related to software and hardware. In many cases, the object of multimedia proposals is to exploit at best the available resources so that the Quality of Experience can be maximized. This objective often requires a deep understanding of the underlying technical aspects regarding hardware (processors), networks (protocols) and software (libraries, operating systems).

The paper presenting VIGRIS has not been published in a multimedia venue, but it reflects well that kind of broad knowledge that is required to make a good multimedia experience possible with respect to today's resource availability. In this paper, the authors deal with Graphics Processing Unit (GPU) and virtualization. Their overall goal is to better schedule concurrent jobs sharing the same GPU with respect to various constraints.

The topic of this paper is especially relevant considering the growing popularity of cloud-based resource-hungry multimedia services. The authors focus on cloud gaming but many other services, for example transcoding services, also need to both make intensive use of GPU and run in the cloud at minimum price. The motivation is that GPUs are now able to support virtualized tasks in a relatively efficient manner (which was definitely not true five years ago), but the sharing of the GPU among concurrent virtualized tasks is still not done properly, which calls for progresses.

The cloud providers talk about *consolidation* when it is to share hardware resources among concurrent jobs and to make a large number of virtual jobs run in a small number of physical machines. The consolidation of multimedia jobs has not been deeply explored so far. The lack of consolidation increases the costs since it requires the service provider to dedicate a large number of machines to serve a small number of end-users. Typically, the lack of consolidation is known as an Achilles' heel of cloud gaming: if one has to reserve one machine per end-user, the gain of running the service in the cloud from data-centers will be extremely limited. The goal is to make several game engines run on the same computer so that multiple end-users can be served by a small number of shared processing resources.

This paper addresses this problem. I found it especially interesting on several aspects.

First, this paper is a solid tutorial to state-of-the-art underlying concepts behind GPU and virtualization. The authors take time to motivate their studies and to explain carefully the background regarding the way multimedia jobs are usually designed and processed in modern computers. This part provides the necessary information for those in the multimedia community who would like to de-

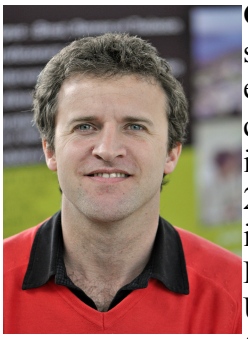
velop some proficiency on GPU programming and virtual jobs management.

Second, the authors propose a whole system, including the code, the Application Programming Interfaces (APIs) and other implementation details. This part may go too far into details for those who want to keep a high-level understanding of the underlying concepts, but it is nonetheless a fascinating journey into low-level commands. Such a level of details is required for journals like ACM Transactions on Architecture and Code Optimization.

Third, the authors provide a relatively simple solution as long as high-level aspects matter. It is not uncommon that such high-level vision does not really accommodate low-level issues, which makes the whole proposal impossible to implement in practice. In the current case, a long part of the paper addresses the most critical implementation details so the doubts about the implementability of the solution are non-existent. At the end, the simplicity of the high-level solution is an asset: the scheduling of tasks and the sharing of GPUs can be done in several ways that the authors explore with a pragmatic approach.

Finally, the authors provide a large set of real measurements to show the efficiency of their solution. The VIGRIS schedulers can guarantee good performances for several concurrent jobs running on the same cluster of servers. The consolidation of multimedia tasks (here video games) is better.

All in one, this paper provides a nice journey into the reality of implementing virtual jobs over GPU-equipped servers. With respect to the number of papers that deal with cloud-based multimedia services, such an accurate study of the underlying aspects is a great reading, which puts the efforts of researchers in a broader perspective.



Gwendal Simon is Associate Professor at Telecom Bretagne. He received his Master Degree in Computer Science in 2000 and his PhD degree in Computer Science in December 2004 from University of Rennes 1 (France). From 2001 to 2006 he was a researcher at Orange Labs, where he worked on peer-to-peer networks and social media innovations. Since 2006, he has been Associate Professor at Telecom Bretagne, a graduate engineering school within the Institut Mines-Telecom. He has been a visiting researcher at University of Waterloo from September 2011 to September 2012. His research interests include large scale networks, distributed systems, optimization problems and video delivery systems.

Approaching New Limits of Synchrony with Multi-Sensorial Media

A short review for "Perceived Synchronization of Mulsemedia Services"
(Edited by Ragnhild Eg and Carsten Griwodz)

Zhenhui Yuan; Ting Bi; Muntean, G.-M.; Ghinea, G.,"Perceived Synchronization of Mulsemedia Services,", IEEE Multimedia Transactions, vol. 17, no. 7, 2015, pp. 957-966.

The race to provide consumers with the most realistic and immersive multimedia experience is speeding up, picking up new sensations in the ongoing pursuit. An immersive multimedia presentation is one that engages the audience and captures their attention, and attention is directed using the senses. Presumably, the motivation behind multi-sensorial media, or mulsemedia, is to stimulate more than vision and hearing, incorporating additional sensory modalities in order to grab someone's full attention. With this bold idea follow new devices to connect, trigger and operate and new signals to encode, transmit and align. Decades of research lie behind the streaming technology that has become commonplace today, yet it is far from flawless. Available bandwidth comes with an upper limit, audio and video are encoded differently, and end-devices vary greatly from person to person. By including additional signal sources, product and service providers are adding a large degree of complexity to an already difficult situation.

In their paper on mulsemedia synchronisation, Yuan, Bi, Muntean and Ghinea are tackling the timing of sensory signal transmission. The authors start out by describing the term mulsemedia as the combination of objects that target, for instance, the human sense of smell, taste or haptics with more traditional multimedia, such as text, image or video. Mulsemedia is therefore not defined far beyond the term multi-sensorial media, and its use is briefly mentioned as a means to provide immersive communication and enhance quality of experience. The authors go on to describe the challenge of synchronising the presentation of traditional multimedia content with the more novel sensory outputs. On the one hand, visual and auditory digital signals are typically transmitted across a network and conveyed locally through machine hardware. On the other hand, Yuan et al. remark the difference for olfactory and haptic signals, where the subjective experience must be produced by a local device that is triggered by a transmitted signal (although it should not be forgotten that the same is the case for a considerable amount of 3D content). As a result, the synchronisation of mulsemedia requires that the transmission, production and presentation of multiple signals converge in time at the end-point. To further complicate the matter, sensory signals tend not to reach the human recipient simultaneously. Sound and light travel at different speeds, and smells follow the airflow in a room or space. Moreover, quality of experience may not rely on absolute synchrony between sensorial stimulation [1]. Consequently, the authors argue that the optimal method for assessing the timing between sensory outputs is using subjective experiments.

Importantly, the resulting experiment is designed to study the effects of temporal asynchrony on subjective perceptual experiences during passive observation. Conversely, interactive applications typically demand a reaction from the recipient, for instance in response to a tactile or olfactory stimulus. This kind of interactivity could make temporal offsets more noticeable. Because this paper does not focus on interactive scenarios, the presented results may not be comparable to temporal thresholds derived in earlier works. For example, Lee and Spence [2] reported that vibrotactile feedback could reduce the time to perform a task by about 100ms, where the video-only solution would take up to 1s. Peon and Prattichizzo [3] report reaction time to haptic stimuli (through the Phantom) below 200ms, and considerably below the reaction time to visual stimuli.

Following their empirical standpoint, Yuan and colleagues conducted a user study where they asked participants to watch a short movie sequence. At one point during each movie presentation, participants would be exposed to either an air-flow from a desktop fan or to a haptic vibration from a gaming vest. The movie sequences contained one scene with an activity that was pre-matched to one of the external stimuli. Although both types of stimulation targeted the tactile modality, the experienced timing of the two was quite distinct. All over, participants considered the vibration to occur at the appropriate time when it happened at the same instance or no more than 1 second later than the corresponding movie activity. Compared to human sensitivity to temporal misalignments between audio and video [4], this can already be considered a

considerable tolerance to delayed haptic outputs. Yet, air-flow haptic devices get away with even greater temporal offsets; the wind stimulation could take place as long as 5 seconds before or 3 seconds after the movie activity and the timing would still be considered appropriate. The experimental setup and evaluation are described in great detail in the paper, allowing other researchers to validate the results. Unfortunately, the question of air-flow delay is described in less detail. Moreover, we wonder whether figures 3 and 4 may unintentionally be hiding outliers. We know from synchronisation studies that test subjects may be able to detect audiovisual asynchrony without being able to say which came first [4]. The same applies to distinct stimulations to the same sensory modality, such as the difficulty subjects experience when asked to determine which of two vibrotactile stimuli arrived first [5]. By leaving out variances in their chosen representations of score distributions, the authors may inadvertently be hiding contradictory results. Follow-up studies should confirm that the experiment is free from such uncertainties.

For every assessed appropriate haptic timing, participants also evaluated their subjective perceptual experience with the additional sensory stimulation. At this point, the experiment would have benefitted from some moderation. Yuan et al. introduced five separate terms to measure subjective experience, including perceived sense of relevance, reality, distraction, annoyance, and enjoyment. We can only assume that participants would eventually have become both fatigued and confused after considering and providing scores to so many terms. Nevertheless, figures 5 and 6 add some perspective to the results. It is clear that the authors set out to determine whether mulsemedia devices that convey touch impressions can give end-users an increased (or decreased) "sense of reality". This is an exciting term. It goes beyond the classical term Quality-of-Experience, which may satisfy itself with joy instead of realism. It also goes beyond the term immersion, which is mainly concerned with engaging attention. Sense of reality demands much more; it demands that sensations created by mulsemedia devices not only contribute to a user's experience, it requires a realistic sensation to arise from the stimulating touch or smell signal. In the same way, we would expect a perceived sense of reality for audiovisual content to resemble sights and sounds in the physical world. Thus, the authors evaluate the devices and the impact of their perceptual experiences with a far higher standard than seen in similar works.

Regardless of the high standards set by the evaluation terms, we note that the subjective ratings on perceived sense of reality are low. Much like the experience some might remember from watching a 5 fps video at 320x240, the devices are not yet fulfilling the ambition of realism. Perhaps devices need some years to improve, similar to the evolution of video encoders and decoders. Or perhaps the presented scores cannot truly represent a human's perceptual experience. The sensory influence from touch may be simply inaccessible using questions and rating scales, and it may be even more difficult to evaluate with multiple terms to relate to. Yuan et al. have assessed overt human responses that have served the multimedia community well for audiovisual synchronisation, and they have run a study that is an excellent example of thorough design, conduct and analysis. But maybe these new mulsemedia devices appeal to senses that are not equally available to human reflection? And maybe we should take this work as encouragement to design experiments and methodologies that can better tap into covert perceptual responses?

References:

[1] W. Fujisaki and S. Nishida, "Audio–tactile superiority over visuo–tactile and audio–visual combinations in the temporal resolution of synchrony perception," Exp. Brain Res., vol. 198, no. 2–3, pp. 245–259, Sep. 2009.

[2] J.-H. Lee and C. Spence, "Assessing the Benefits of Multimodal Feedback on Dual-task Performance Under Demanding Conditions," in Proceedings of the 22Nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction - Volume 1, 2008, pp. 185–192.

[3] A. R. Peon and D. Prattichizzo, "Reaction times to constraint violation in haptics: comparing vibration, visual and audio stimuli," in 2013 World Haptics Conference (WHC), 2013, pp. 657–661.

[4] M. Zampini, D. I. Shore, and C. Spence, "Audiovisual temporal order judgments," Exp. Brain Res., vol. 152, no. 2, pp. 198–210, Sep. 2003.

[5] C. Preuschhof, T. Schubert, A. Villringer, and H. R. Heekeren, "Prior Information Biases Stimulus Representations during Vibrotactile Decision Making," J. Cogn. Neurosci., vol. 22, no. 5, pp. 875–887, May 2010.

Ragnhild Eg is associate professor at Westerdals Oslo School of Arts, Communication and Technology, where she combines her background and interest in perceptual psychology with digital marketing. She completed a PhD in psychology at the University of Oslo while working at Simula Research Laboratory. At Simula, she was part of a multi-disciplinary project

that focused on the human perception of multimedia. Her research interests relate to the perceptual processing and integration of sensory information, and amongst other topics, her research has explored how the perceptual process is affected by the constraints imposed by technology.

Carsten Griwodz is chief research scientist in the Media Department at the Norwegian research company Simula Research Laboratory AS, Norway, and professor at the University of Oslo.
His research interest is the performance of multimedia systems. He is concerned with streaming media, which includes all kinds of media that are transported over the Internet with a temporal demands, including stored and live video as well as games and immersive systems. To achieve this, he wants to advance operating system and protocol support, parallel processing and the understanding of the human experience. He was area chair and demo chair of ACM MM 2014, and general chair of ACM MMSys and NOSSDAV (2013), co-chair of ACM/IEEE NetGames (2011), NOSSDAV (2008), SPIE/ACM MMCN (2007) and SPIE MMCN (2006), TPC chair ACM MMSys (2012), and systems track chair ACM MM (2008). He is currently editor-in-chief of the ACM SIGMM Records. More information can be found at http://mpg.ndlab.net

Improving QoS for Media Streaming in Distributed Software Defined Networks

A review for "Distributed QoS Architectures for Multimedia Streaming Over Software Defined Networks"
(Edited by Roger Zimmermann)

Hilmi E. Egilmez, A. Murat Tekalp, "Distributed QoS Architectures for Multimedia Streaming Over Software Defined Networks", IEEE Transactions on Multimedia, vol. 16, no.6, pp. 1597–1609, October 2014.

In recent years the concept of Software Defined Networking (SDN) has gained significant attention in the networking community. SDN separates the control-plane and the forwarding-plane that have traditionally co-existed in network hardware. In an SDN architecture, the control-plane is managed by a *controller* that is separate from the packet-level *forwarders*, with a standardized control protocol between the two planes [1] (very often the OpenFlow [2] protocol is used for this purpose). While the SDN abstractions provide significant new flexibilities, they do not account for all the realities of the practical world. In this manuscript the authors specifically investigate the case where an SDN paradigm is used to improve the QoS routing for media streaming, while acknowledging that the Internet is comprised of many independently managed sub-networks (so called domains or Autonomous Systems, AS) and there will likely not be one single, central SDN controller.

The authors make the realistic assumption that multi-domain networks exist, and while AS administrators may be willing to share some aggregate information about their networks, they likely would not want to expose all the details of their internal network architectures to the rest of the world. Starting from this premise, the authors propose three relevant components, namely methods for topology aggregation and link summarization, a general optimization framework for flow-based end-to-end QoS provision over multi-domain networks, and two distributed control plane designs for messaging between controllers.

The authors start by describing two multi-controller, SDN-based control plane architectures. In the more conventional hierarchical architecture, the controllers of multiple domains are coordinated at a higher level by a so called super-controller which has a global, but only aggregate, view of the network. On the other hand, the authors also propose a fully distributed architecture where each controller connects to its neighboring controllers in order to exchange topology aggregation and link summarization information about its own domain. To provide an aggregation of each AS domain, the authors propose that the internal, complex network be abstracted through virtual links between border nodes (i.e., forwarders). The border nodes represent the switches that connect each AS to other domains and the virtual links represent a full-mesh topology between these nodes. The problem that the authors aspire to solve is to establish an optimal QoS routing for audio and video streaming applications. For the optimization framework, the authors use the delay variation in the startup delay of media streaming. However, in their framework other metrics can be used as well. The dynamic QoS routing is posed as a Constrained Shortest Path (CSP) problem and it is solved through a Lagrangian relaxation based aggregated cost (LARAC) algorithm [3] that uses a heuristic since the CSP problem is NP-complete. Because in the proposed framework the overall LARAC solver does not have access to the internal parameters of each domain it is important to assign reasonable QoS parameters to the aggregate virtual links in each domain. For this purpose the authors propose three variations such that (i) the minimal cost and delay variation of the internal route between two border points are assigned to the virtual link, (2) the average cost and the average delay variation of routes calculated in a multi-step process are assigned to the virtual link, or (3) the maximum cost and the maximum delay variation of the *k*-disjoint minimum cost routes are assigned to the virtual link. The authors then introduce three algorithms that solve the overall optimization problem of assigning optimal routing paths based on the QoS parameters that each AS provides to the other domains. Two algorithms are designed for the hierarchical framework while the third works in a fully distributed manner.

The authors evaluate their proposed design and algorithms with a media streaming application that uses scalable video encoding MPEG-4 SVC [4] with a base layer and one enhancement layer. The base layer is assigned a QoS of 250 ms maximum tolerable delay variation, while the enhancement layer is transmitted on a best-effort basis. The resulting streaming quality is measured with the traditional PSNR metric.

The experimental evaluation reveals that the proposed framework can improve the end user quality by several dBs compared to a traditional BGP-like shortest path routing approach. The distributed approach using aggregation method (2) is shown to produce results that are very close to a globally optimum QoS routing (provided for comparison purposes). In the evaluation the authors vary a number of other parameters (for example the distances to the super-controller) and draw overall conclusions. They describe a QoS support vs. communication cost tradeoff between the fully distributed and the hierarchical approach, and they indicate that a hierarchically control plane can be a better option for in-nation service providers.

Overall the proposed framework is interesting, especially the realistic assumption that likely there will exist multiple SDN domains under the control of different service providers. With the proposed solution scalability can be achieved while confidentiality will be preserved as only aggregate information needs to be exchanged.

References:

[4] Open Networking Foundation (ONF), Palo Alto, CA, USA, "Software Defined Networking: The New Norm for Networks," 2012.

[5] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, "OpenFlow: Enabling Innovation in Campus Networks," *SIGCOMM Computer Communication Review*, vol. 38, no. 2, pp. 69–74, March 2008.

[6] F. Kuipers, P. Van Mieghem, T. Korkmaz, and M. Krunz, "An Overview of Constraint-based Path Selection Algorithms for QoS Routing," *IEEE Communications Magazine*, vol. 40, no. 12, pp. 50–55, Dec. 2002.

[7] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.

Roger Zimmermann is an associate professor with the Department of Computer Science at the School of Computing with the National University of Singapore (NUS) where he is also an investigator with the Interactive & Digital Media Institute (IDMI). His research interests are in both spatio-temporal and multimedia information management, for example distributed and peer-to-peer systems, spatio-temporal multimedia, streaming media architectures, georeferenced video management, mobile location-based services and geographic information systems (GIS). He has co-authored a book, six patents and more than two hundred conference publications, journal articles and book chapters in the areas of multimedia and databases. He has received the best paper award at IEEE ISM 2012 and was part of the team who won second place at the ACM SIGSPATIAL GIS Cup 2013. He has been involved in the organization of conferences in various positions, for example program co-chair of ACM Multimedia 2013. He co-directs the Centre of Social Media Innovations for Communities (COSMIC) at NUS and is an investigator with the NUS Research Institute (NUSRI) in Suzhou, China. Roger Zimmermann is an Associate Editor of the ACM Transactions on Multimedia journal (TOMM, formerly TOMCCAP) and the Multimedia Tools and Applications (MTAP) journal. He is a Senior Member of the IEEE and a member of ACM. For more details, see http://eiger.ddns.comp.nus.edu.sg

Content-Oriented Geo Features in Hybrid Video Search

A review for "On Generating Content-Oriented Geo Features for Sensor-Rich Outdoor Video Search"
(Edited by Pradeep K. Atrey)

Yifang Yin, Yi Yu, and Roger Zimmermann, "On Generating Content-Oriented Geo Features for Sensor-Rich Outdoor Video Search", IEEE Transactions on Multimedia, vol.17, no.10, pp.1760-1772, Oct. 2015.

The ubiquity of smartphones and tablets has enabled users to conveniently record, upload and view videos on a number of multimedia-sharing platforms, such as YouTube. This phenomenon results in a tremendous increase of user-generated videos on the Internet every year. Accordingly, the huge amount of videos challenges the traditional retrieval frameworks that only rely on the low-level visual features or the surrounding textual annotations. Such methods may suffer from low recall due to the changes in illumination and the noise in tags.

In this paper, the authors propose a hybrid video search framework by utilizing the geographic metadata associated with sensor-rich videos. Related work has been presented, for example, probably the most common procedure is to first filter the multimedia documents based on the location extracted from geo-metadata, and then perform text- or content-based ranking [1]. Alternatively, the ranking can be conducted based on a conjunctive function that jointly considers the geographic distance and the visual distance [2]. However, the authors argue that the existing fusion techniques usually leverage the location information directly, which is simply one of the camera properties and is inconsistent with the video content. Moreover, the use of camera direction information has not been thoroughly studied yet. To solve the above problems, the authors suggest generating a new content-oriented geo feature that can be fused with visual features more seamlessly. Based on this new feature, advanced techniques have been proposed to improve the robustness and diversity of the searching results.

In detail, the authors first present a hybrid model as video representation, which is a two-level hierarchical structure. At level one, a geo-histogram is generated describing the overall geographic coverage of a video. It is calculated based on the camera's field-of-view [3] and a pre-defined geo-codebook. In order to effectively encode the geo-coverage, a geo-codebook generation module is presented that is able to segment a map into a set of coherent regions with estimated saliency values. At level two, frames are matched to regions and representatives are selected in terms of visual appearances. Next the authors discuss how an effective similarity measure can be derived from the pro-posed representation. The spatial relevance between two videos is estimated by applying the cosine similarity between the geo-histograms. This measure is highly efficient and generally works well, but may face performance issues due to unpredictable occlusions, changes in viewpoints or illuminations, or errors in the geo-metadata. In order to solve the above issue and promote videos that are more visually similar to a query in the result list, the authors modify the spatial relevance measure by introducing a second factor that quantifies the local visual similarity on a region-by-region basis. By indexing frames geographically, this measure reduces potential mismatches while consuming less computational costs for query by example. Additionally, the authors also describe the use of the proposed representation in semantic annotation. Multiple clues including texts, visual content, as well as geographic priors collected from OpenStreetMap have been taken into consideration.

The authors implement a video search prototype and conduct analysis on each component. The dataset for evaluation is collected from the GeoVid website, which hosts sensor-rich user-generated videos from all over the world. They illustrate four examples of the generated geo-codebook in Singapore, Chicago, Japan, and Hong Kong, and briefly discuss the scalability of their approach. In terms of similarity search, the authors compare their method to two state-of-the-art visual approaches [1,2] and two popular fusion techniques [4,5]. The results indicate that the proposed innovative fusion of visual and geo features is able to achieve significant improvements in terms of computational complexity and retrieval accuracy.

To summarize, this paper presents a novel hybrid video description for similarity search and semantic annotation. The proposed content-orientated geo feature is parallel to the existing compact video representations that are mostly designed for efficient visual indexing [6]. The authors briefly discuss the possibility to integrate the approximate visual similarity measure on the region-level in their model after frames are geographically indexed. With the growing

of geo-referenced videos on the Internet, this could be an interesting future extension of this work.

Acknowledgement:
The R-Letter Editorial Board thanks the authors of the paper for providing a summary of its contributions.

References:

[1] N. O'Hare, C. Gurrin, G. Jones, and A. Smeaton, "Combination of Content Analysis and Context Features for Digital Photograph Retrieval," European Workshop on Integration of Knowledge, Semantics and Digital Media Technology, 2005, pp. 323–328.

[2] J. Kamahara, T. Nagamatsu, and N. Tanaka, "Conjunctive Ranking Function using Geographic Distance and Image Distance for Geotagged Image Retrieval", ACM GeoMM, 2012, pp. 9-14.

[3] S. Arslan Ay, R. Zimmermann, and S. H. Kim, "Viewable Scene Modeling for Geospatial Video Search," ACM Multimedia, 2008, pp. 309–318.

[4] L.-J. Li, H. Su, Y. Lim, and L. Fei-Fei, "Object Bank: An Object-Level Image Representation for High-Level Visual Recognition," International Journal of Computer Vision, pp. 20–39, 2014.

[5] O. A. B. Penatti, L. T. Li, J. Almeida, and R. da S. Torres, "A Visual Approach for Video Geocoding Using Bag-of-Scenes," ACM ICMR, 2012, pp. 53:1–53:8.

[6] H. T. Shen, B. C. Ooi, and X. Zhou, "Towards Effective Indexing for Very Large Video Sequence Database," ACM SIGMOD, 2005, pp. 730–741.

Pradeep K. Atrey is an Assistant Professor at the State University of New York, Albany, NY, USA. He is also an (on-leave) Associate Professor at the University of Winnipeg, Canada and an Adjunct Professor at University of Ottawa, Canada. He received his Ph.D. in Computer Science from the National University of Singapore, M.S. in Software Systems and B.Tech. in Computer Science and Engineering from India. He was a Postdoctoral Researcher at the Multimedia Communications Research Laboratory, University of Ottawa, Canada. His current research interests are in the area of Security and Privacy with a focus on multimedia surveillance and privacy, multimedia security, secure-domain cloud-based large-scale multimedia analytics, and social media. He has authored/co-authored over 90 research articles at reputed ACM, IEEE, and Springer journals and conferences. His research has been funded by Canadian Govt. agencies NSERC and DFAIT, and by Govt. of Saudi Arabia. Dr. Atrey is on the editorial board of several journals including ACM Trans. on Multimedia Computing, Communications and Applications, ETRI Journal and IEEE Communications Society Review Letters. He was also guest editor for Springer Multimedia Systems and Multimedia Tools and Applications journals. He has been associated with over 30 international conferences/workshops in various roles such as General Chair, Program Chair, Publicity Chair, Web Chair, Demo Chair and TPC Member. Dr. Atrey was a recipient of the ACM TOMM Associate Editor of the Year (2015), the Erica and Arnold Rogers Award for Excellence in Research and Scholarship (2014), ETRI Journal Best Editor Award (2012), ETRI Journal Best Reviewer Award (2009) and the three University of Winnipeg Merit Awards for Exceptional Performance (2010, 2012 and 2013). He was also recognized as "ICME 2011 - Quality Reviewer".

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 includes, but is 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:

Christian Timmerer (christian.timmerer@aau.at) and Yan Zhang (yanzhang@simula.no).

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

Members of the IEEE MMTC Review Board will review each nominated paper. In order to avoid potential conflict of interest, guest editors external to the Board will review nominated papers co-authored by a Review Board member. The reviewers' names will be kept confidential. If two reviewers agree that the paper is of R-letter quality, a board 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). Nominators external to the board 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 R-Letter Editorial Board

Multimedia Communications Technical Committee Officers

Chair: Yonggang Wen, Singapore
Steering Committee Chair: Luigi Atzori, Italy
Vice Chair – North America: Khaled El-Maleh, USA
Vice Chair – Asia: Liang Zhou, China
Vice Chair – Europe: Maria G. Martini, UK
Vice Chair – Letters: Shiwen Mao, USA
Secretary: Fen Hou, China
Standard Liaison: Zhu Li, USA

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.