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

R-LETTER

Vol. 5, No. 1, February 2014

IEEE COMMUNICATIONS SOCIETY

CONTENTS

Message from the Review Board

Introduction

In addition to recommend high quality articles to MMTC members, another goal of the Review Board is to promote MMTC and initiate helpful exchange between technical committees. Resulting from the hard work of Review Board members, we are glad to advise that, starting from this issue, the front page of our latest R-letter will be displayed on the SIGMM Records webpage. Names of R-article editors will also be included in the front page. This will certainly help establish a closer link between MMTC and SIGMM communities. R-letter is also becoming a regular contributor to IEEE Communications Society's Communications Technical News.

Distinguished Category

The recent advent of ubiquitous network access and affordable mobile devices has led to an unprecedented traffic growth in wireless networks. Such exponential growth has brought several technical challenges; one is the ever-increasing network energy consumption, especially on Base Stations (BSs). Another important challenge is how to manage/utilize such gigantic amount of data generated by mobile users.

The first paper, published in the *IEEE Transactions on Wireless Communications*, provided a new perspective on green cellular networks using hybrid energy, which consists of traditional grid and renewable green energy.

The second paper, published in the *IEEE Transactions on Multimedia*, studied a mobile cloud-based stream mining system, where mobile users send their captured data streams to the cloud for mining.

Regular Category

The topics of the regular category include video streaming, multiview video coding, natural language descriptions, shape similarity fusion, and 3D watermarking.

The **first paper**, published in the *IEEE Transactions on Vehicular Technology* and *edited by Koichi Adachi*, proposes a novel bandwidth management system for live video streaming in a public-shared network.

The **second paper**, published in the *IEEE Transactions on Multimedia* and *edited by Carsten Griwodz*, presents a joint bit allocation and rate control for the coding of multiview video plus depth based 3D video.

The **third paper** is *edited by Guillaume Lavoué* and has been published within the *Proceedings of IEEE International Conference on Computer Vision (ICCV)* and proposed means for the automatic natural Language descriptions of images.

The **forth paper**, published in the *IEEE Transactions on Multimedia* and *edited by Jun Zhou*, describes an effective shape similarity fusion method based on local data distribution.

Finally, we revisit a paper by incorporating the authors' view. The **fifth paper**, published in *IEEE Transactions on Image Processing* and *edited by Guillaume Lavoué and Adrian G. Bors*, proposes improvements for a 3D digital watermarking system with respect to perceptibility and robustness. Interestingly, it provides a different view on the same paper we reviewed in the December 2013 issue.

We would like to thank all the authors, nominators, reviewers, editors, and others who contribute to the release of this issue.

IEEE ComSoc MMTC R-Letter
Director:
Irene Cheng, University of Alberta, Canada
Email: locheng@ualberta.ca

Co-Director: Distinguished Category
Weiyi Zhang, AT&T Research, USA
Email: maxzhang@research.att.com

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

Mining Data Streams in Mobile Clouds

A short review for "Efficient Resource Provisioning and Rate Selection for Stream Mining in a Community Cloud"

Edited by Carl James Debono

S. Ren, and M. van der Schaar, "Efficient Resource Provisioning and Rate Selection for Stream Mining in a Community Cloud," IEEE Transactions on Multimedia, vol. 15, no. 4, pp. 723-734, June 2013.

Mining of multimedia data within large networks requires advanced search techniques that are computationally demanding. Implementing these solutions in mobile networks is not feasible due to the hardware and energy constraints of the mobile devices. Therefore, new techniques are needed to make multimedia data mining available in wireless systems.

The recent advent of ubiquitous network access and affordable mobile devices have led to an unprecedented growth of information flows and spurred a massive generation of widely-scattered data streams in wireless networks. Such a new wave of big data promises to revolutionize and bring fundamental changes to many sectors, including healthcare, energy and cloud computing, among others. The full potential of the enormous data streams in wireless networks can only be realized when the data is "mined" promptly. Mining the data streams by extracting valuable information out of the data in real time, however, is typically computation-intensive and exceeds the capabilities of many mobile devices, despite numerous technological innovations in the mobile industry during the last decade [1]. Resolving the conflicts between the soaring demand for mining big data, emerging in wireless networks, and the resource scarcity of mobile devices calls for a new computing paradigm, for which mobile cloud appears to be a promising candidate due to its capability of providing elastic and scalable computational resources to the mobile users [2, 3].

While mobile cloud is emerging as a critical component for reaping the full benefits of enormous data streams, it faces several challenges that involve various aspects of the system and need to be addressed. First, mobile devices, where data streams are generated, are highly energy-constrained, whereas sending data streams to the cloud side is power consuming [4]. Second, responsiveness of the data stream mining cannot be compromised, since otherwise the mining result may be of little value and/or even result in serious consequences [1]. Last but not least, on the cloud side, a large number of servers are indispensible for supporting the stream mining system, which thereby consumes a huge amount of energy [5].

The authors of the original paper study a mobile cloud-based stream mining system, where mobile users send their captured data streams to the cloud for mining. Three key metrics, energy consumption of the cloud, stream mining performance (e.g., accuracy of mining result) and responsiveness, are holistically combined into the study, while the energy consumption of battery-powered mobile users is also incorporated as a system constraint. Nonetheless, due to the inherent tradeoff, optimizing these three performance metrics simultaneously is not possible in practice. Thus, depending on the specific application environment, the cloud operator needs to have the ability to perform a desired tradeoff, which, however, is challenged by the fact that the environment in which the stream mining system operates is randomly changing over time and the distribution of the underlying stochastic process is often unknown a priori. Specifically, the wireless channels linking the cloud and mobile users are time-varying with possibly unknown channel gain distributions, and the data stream characteristics may also be constantly changing.

To address these challenges, the authors of the paper leverage state-of-the-art control techniques [6] and develop an online computational resource provisioning and transmission rate selection algorithm to minimize the classification-energy cost. This formally quantifies both stream mining performance and energy consumption on the cloud side, subject to the stream mining responsiveness requirement. The algorithm specifies a sequence of online decisions without requiring the need of the future information. Specifically, at each time instant, each mobile user decides its transmission rate (which affects the user's transmission power and influences the stream mining performance), while the cloud operator decides its computational resource provisioning (e.g., how many servers, which CPU speed setting) for stream mining. The authors of the paper also show that at the expense of increasing the response time of stream mining, the classification-energy cost can be reduced and pushed arbitrarily close to the minimum cost achieved by the optimal

offline algorithm. The authors also extend the result to a profit-maximizing cloud operator that can dynamically price its computational resource for stream mining to maximize its long-term profit. Extensive simulations results were conducted and analyzed by the authors to validate the analysis.

In summary, the authors address four system issues in a mobile cloud mining: (1) minimizing energy consumption on the cloud side; (2) improving stream mining performance; (3) delivering timely mining results; and (4) reducing power consumption of mobile devices. The distinguishing feature of online execution makes the solution proposed by the authors an appealing candidate for future mobile cloud supporting interactive stream mining and for realizing the full potential of big data in wireless networks.

The mining of multimedia data in wireless networks will increase in demand and complexity as more users make use of multimedia capabilities of the devices to generate content and when more advanced content, such as 3D video content, is captured. To be successful, algorithms and solutions that are fast and consume less energy are needed. This needs to be coupled with improved accuracy, such as finding new descriptors that help the mining process. Future work needs to take into consideration these issues to make mobile cloud multimedia data mining viable.

Acknowledgement:
This paper is nominated by the MMTC Green Multimedia Communications (GMC) Interest Group.

References:

[1] R. Ducasse, D. Turaga, and M. van der Schaar, "Adaptive topologic optimization for large-scale stream mining," *IEEE Journal on Selected Topics in Signal Processing*, vol. 4, no. 3, pp. 620-636, Jun. 2010.
[2] B. Girod, V. Chandrasekhar, R. Grzeszczuk, and Y. A. Reznik, "Mobile visual search: Architectures, technologies, and the emerging MPEG standard," *IEEE MultiMedia*, vol. 18, no. 3, pp. 86-94, Jul.-Sep. 2011.
[3] K. Kumar, and Y.-H. Lu, "Cloud computing for mobile users: Can offloading computation save energy?," *IEEE Computer*, vol. 43, no. 4, pp. 51-56, Apr. 2010.
[4] N. Balasubramanian, A. Balasubramanian, and A. Venkataramani, "Energy consumption in mobile phones: A measurement study and implications for network applications," in *Proceedings of the 9th ACM SIGCOMM Internet Measurement Conference*, 2009.
[5] B. Guenter, N. Jain, and C. Williams, "Managing cost, performance and reliability tradeoffs for energy-aware server provisioning," in *Proceedings of IEEE INFOCOM*, 2011.
[6] M. J. Neely, *Stochastic Network Optimization with Application to Communication and Queueing Systems*, Morgan & Claypool, 2010.

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 an Associate Professor. He is currently the Deputy Dean of the Faculty of ICT at the University of Malta.

Prof. 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 Vice-Chair of Technical Activities for 2014. 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.

A New Perspective on Green Cellular Networks using Hybrid Energy

A short review for "On Optimizing Green Energy Utilization for Cellular Networks with Hybrid Energy Supplies"

Edited by Weiyi Zhang

Tao Han, and Nirwan Ansari (FIEEE),"On Optimizing Green Energy Utilization for Cellular Networks with Hybrid Energy Supplies'", IEEE Transactions on Wireless Communications, Vol. 12, No. 8, August 2013,pp. 3872-3882.

Green communications has received much attention recently. In wireless cellular networks, energy consumption is mainly drawn from BSs (base stations). According to the power consumption breakdown [1], BSs consume more than 50% of the power of a cellular network. In addition, the number of BSs is expected to be doubled in recent years [2]. Thus, reducing the power consumption of BSs is crucial to green cellular networks.

One of the popular techniques improving the energy efficiency of cellular networks is to design and optimize the power saving communication protocols that adjust the transmit power of the transceivers according to the traffic intensity. For example, one intuitive idea is to switch off the BSs when the traffic load is below a certain threshold for a certain time period [3]. When some BSs are switched off, radio coverage and service provisioning are taken care of by the devices that remain active. The BS switching problem can be formulated as an optimization problem that minimizes the number of active BSs while meeting the traffic load in the access network.

However, when the traffic demand is intense, few BSs can be switched off, and thus these "sleep mode" based algorithm may not be effective. When traffic demand is intense, heterogeneous radio access networks which utilize a diverse set of base stations achieves higher spectral and energy efficiency per unit area. The network deployment featuring high density deployments of small, low power BSs achieves higher network energy efficiency [4, 5].

Another proposed solution is to use green energy powered BSs to reduce the carbon footprint [6]. Green energy such as sustainable biofuels, solar and wind energy are promising options to save the on-grid energy consumed by BSs and reduce the CO2 footprint. However, owing to the dynamics of green energy generation and the limited capacity of energy storage, green energy may not guarantee sufficient power supplies for BSs.

In this paper, the authors proposed to envision future BSs to be powered by multiple types of energy sources, e.g., the grid, solar energy, and wind energy. In such cellular networks, BSs are powered by green energy if they have enough green energy stored in their batteries; otherwise, the BSs switch to on-grid energy to serve mobile users.

In order to reduce the on grid energy consumption of cellular networks during the peak traffic hours, the authors studied a green energy optimization (GEO) problem to balance the energy consumption among BSs. Since such problem involves the optimization in two dimensions: the time dimension and the spatial dimension, this paper decompose the GEO problem into two sub-problems: the multi-stage energy allocation (MEA) problem and the multi-BSs energy balancing (MEB) problem based on the characteristics of the green energy generation and the mobile traffic.

The MEA problem is to optimize the green energy allocation at individual BSs to accommodate the temporal dynamics of both the green energy generation and the mobile traffic. The authors proposed the MEA algorithm to solve the MEA problem. Taking advantages of the spatial diversity of the mobile traffic, the MEB problem is to balance the green energy consumption among BSs so as to reduce the on-grid energy consumption of the cellular network. The authors have proposed the MEA algorithm, the MEB algorithm and the EA algorithm to solve these sub-problems, and then address the GEO problem. In detail, the solution of the MEA problem estimates the amount of green energy allocated at individual BSs during each time slot. Based on this solution, optimal cell size adaptation is achieved by solving the MEB problem. Given the cell sizes, individual BSs apply the energy allocation (EA) algorithm to re-calculate their green energy usage, and determine

whether to consume on-grid energy at the current time slot. Therefore, the GEO algorithm consists of the MEA algorithm, the MEB algorithm, and the EA algorithm. In addition to theoretical analysis, extensive simulations have been carried out to evaluate the proposed solution and showed that the proposed work in this paper is able to save a significant amount of on-grid energy.

Acknowledgement:
This paper is nominated by the MMTC Multimedia Services and Applications over Emerging Networks (MEN) Interest Group.

References:

[1] C. Han, et al., "Green radio: radio techniques to enable energy-efficient wireless networks," IEEE Commun. Mag., vol. 49, no. 6, pp. 46–54, June 2011.
[2] "Community power using mobile to extend the grid." Available: http://www.gsmworld.com/documents/gpfm_communitypower11whitepaperlores.pdf
[3] "3gpp r3-100162: overview to LTE energy saving solutions to cell switch off/on," 3GPP RAN3 Meeting, Valencia, Spain, Jan. 2010.
[4] M. Etoh, T. Ohya, and Y. Nakayama, "Energy consumption issues on mobile network systems," in Proc. 2008 Int. Symp. Appl. Internet.
[5] K. Samdanis, D. Kutscher, and M. Brunner, "Dynamic energy-aware network re-configuration for cellular urban infrastructures," in Proc. 2010 IEEE GLOBECOM Workshops.
[6] "Sustainable energy use in mobile communications," Ericsson Inc., white paper, Aug. 2007.

Weiyi Zhang is currently a Research Staff Member of the Network Evolution Research Department at AT&T Labs Research, Middletown, NJ. Before join AT&T Labs Research, he was an Assistant Professor at the Computer Science Department, North Dakota State University, Fargo, North Dakota, from 2007 to 2010. His research interests include routing, scheduling, and cross-layer design in wireless networks, localization and coverage issues in wireless sensor networks, survivable design and quality-of-service provisioning of communication networks. He has published more than 80 refereed papers in his research areas, including papers in prestigious conferences and journals such as IEEE INFOCOM, ACM MobiHoc, ICDCS, IEEE/ACM Transactions on Networking, ACM Wireless Networks, IEEE Transactions on Vehicular Technology and IEEE Journal on Selected Areas in Communications. He received AT&T Labs Research Excellence Award in 2013, Best Paper Award in 2007 from IEEE Global Communications Conference (GLOBECOM'2007). He has been serving on the technical or executive committee of many internationally reputable conferences, such as IEEE INFOCOM. He was the Finance Chair of IEEE IWQoS'2009, and serves the Student Travel Grant Chair of IEEE INFOCOM'2011.

A Tree-Based Resource Management Architecture For Live Video Streaming Delivery

A short review for "A Novel Bandwidth Management System for Live video Streaming on a Public-Shared Network"

Edited by Koichi Adachi

H.-Y. Chang, N.-F. Huang, Y.-W. Lin, and Y.-J. Tzang, "A Novel Bandwidth Management System for Live Video Streaming on a Public-Shared Network", IEEE Transactions on Vehicular Technology, vol. 62, no. 8, pp. 3848-3862, Oct. 2013.

With the support of networking technologies such as wired networks, WiFi, and 3G/4G, smart handheld devices are able to support live video streaming more effectively than ever, which results in an increase of the popularity as a research topic.

Traditional multimedia systems, which are based on client-server architecture, cannot support the demands from a rapidly increasing number of clients. Clients impose a heavy burden on the bandwidth of the source server, and Internet protocol (IP) multicast **[7]**–**[9]** could be the most effective mean for resolving this problem because it is specifically designed to deliver content related to group-oriented applications efficiently. Nevertheless, IP multicast has certain limitations as indicated in **[10]**. Consequently, previous researchers have developed P2P schemes on application level overlay networks **[11]-[16]** for live media streaming applications.
However, the frequent changes in the P2P network topology make the live video streaming difficult. Therefore, an appropriate and stable video delivery platform for live video streaming in all types of online multimedia devices is necessary.

For more efficient use of Internet bandwidth, which may be different among clients, scalable video coding (SVC) **[17]** has been proposed. However, the heavy computational power is need for its encoding/decoding and clients may not have an ability to perform SVC. Instead the split-and-merge (S-M) method, which is based on asynchronous multisource streaming (AMSS) mode **[18][19]** developed for P2P, is adopted to satisfy the various bit rate requirements of clients in this study.

In this paper, a tree-based architecture for live video streaming delivery is proposed, where the smart handheld devices are leaves on delivery trees. A resource management system that exploits the sharable bandwidth of public-shared networks, such as FON, is designed and implemented to construct an efficient, robust, and high-availability video streaming delivery system (referred to as PSnet). In PSnet, the sharable bandwidth of an AP refers to the sharable uplink bandwidth of wired bandwidth. Therefore, streaming is delivered to the AP from the downlink, on which the AP uploads multiple copies to other APs. Thus, a sharable AP acts as an amplifier for the streaming video. Based on this concept of bandwidth expansion, the source of the video streaming system only requires a small amount of bandwidth to deliver the video and numerous clients are capable of receiving the video stream simultaneously. Two optimized algorithms are proposed to arrange the publicly shared bandwidth so that all clients are served and minimum resources are consumed. The PSnet system was implemented to demonstrate the overall feasibility of the concept.

The PSnet is composed of several parts; a streaming source to provide video streaming, a server (PSnet-S) to manage the entire system, the streaming delivery architecture (PSnet-G) comprising several groups of organized sharable APs (PSnet-Ns), the sharable AP pool for backup (PSnet-P), and clients who wish to receive the video stream. Based on requests from clients, the PSnet-S constructs the PSnet-G by organizing the sharable PSnet-Ns into groups. For efficient management of the resource, the PSnet-Ns in each group are organized as complete-binary-tree structures. Because PSnet-N is essentially an AP and usually remains online for a long time following initialization and powering up, the tree structure is usually stable. When a client requests video streaming from PSnet-S, it delivers streaming to the root node of the tree, where it is "duplicated" (uploaded) to the child nodes of the root. Thus, streaming is relayed to the internal nodes of the tree. Finally, the stream is delivered to the leaf nodes, which forward the stream to the clients. The degree to which each node contributes to the stream (PSnet-N) depends on the amount of sharable bandwidth of that node and the rate of video streaming.

How to arrange the nodes or how to establish the trees are among the most important issues in PSnet. Two optimization problems were formulated in this paper: Problem 1 finds a way to construct the tree(s) such that all requests are satisfied with the minimum number of constructed trees, and Problem 2 finds a way to construct the tree(s) such that all requests are

satisfied with the minimum number of nodes used. Two greedy algorithms were developed to solve the above two optimization problems. The authors showed the optimality of the developed algorithms and their worst-case overhead. Additionally, a resource management scheme was developed for recycling and reusing resources to improve the continuity of streaming experience and to reduce the overall system load on the devices involved.

The experimental results were presented to demonstrate the effectiveness of the proposed algorithms. The obtained results have shown that the proposed system is highly effective and appropriate for delivering live streaming compared to the Goalbit P2P live streaming system **[20]** in terms of video delay, redundancy ratio, and start-up latency. Furthermore, to enhance the system availability, a handoff scheme was designed for PSnet system which selects a new PSnet-N from PSnet-P for replacement if a PSnet-N crashes due to power failure, network failure, or even heavy wireless access. The results have indicated that a higher bit rate produced a shorter playout deadline, and vice versa.

References:

[7] T. Kim and M. H. Ammar, "A comparison of heterogeneous video multicast schemes: Layered encoding or stream replication," *IEEE Trans. Multimedia*, vol. 7, no. 6, pp. 1123–1130, Dec. 2005.

[8] D. Wu, Y. T. Hou, and Y.-Q. Zhang, "Scalable video coding and transport over broad-band wireless networks," *Proc. IEEE*, vol. 89, no. 1, pp. 6–20, Jan. 2001.

[9] Q. Zhang, Q. Guo, Q. Ni, W. Zhu, and Y.-Q. Zhang, "Sender-adaptive and receiver-driven layered multicast for scalable video over the Internet," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 4, pp. 482–495, Apr. 2005.

[10] E. Setton and B. Girod, *Peer-to-Peer Video Streaming*. New York, NY, USA: Springer-Verlag, 2007.

[11] N.-F. Huang, Y.-J. Tzang, H.-Y. Chang, and C.-W. Ho, "Enhancing P2P overlay network architecture for live multimedia streaming," *Inf. Sci.*, vol. 180, no. 17, pp. 3210–3231, Sep. 2010.

[12] N.-F. Huang, Y.-J. Tzang, H.-Y. Chang, and C.-S. Ma, "Construction of an efficient ring-tree-based peer-to-peer streaming platform," *in Proc. NCM*, Aug. 2010, pp. 75–80.

[13] C. Wu, B. Li, and S. Zhao, "On dynamic server provisioning in multichannel P2P live streaming," *IEEE/ACM Trans. Netw.*, vol. 19, no. 5, pp. 1317–1330, Oct. 2011.

[14] L. Zhou, Y. Zhang, K. Song, W. Jing, and A. V. Vasilakos, "Distributed media services in P2P-based vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 60, no. 2, pp. 692–703, Feb. 2011.

[15] B. Zhang, S. Jamin and, and L. Zhang, "Host multicast: A framework for delivering multicast to end users," *in Proc. IEEE INFOCOM*, New York, NY, USA, Jun. 23–27, 2002, pp. 1366–1375.

[16] D. Pendarakis, S. Shi, D. Verma, and M. Waldvogel, "ALMI: An application level multicast infrastructure," *in Proc. 3rd USENIX USITS*, San Francisco, CA, USA, Mar. 2001, pp. 49–60.

[17] H. Schwarz, D. Marpe, and T. Wieg, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE TCSVT*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.

[18] S. Itaya, N. Hayashibara, T. Enkido, and M. Takizawa, "Asynchronous multi-source streaming protocol to realize high-performance multimedia communication," *in Proc. DEXA Workshop*, Aug. 2005, pp. 116–120.

[19] S. Itaya, T. Enokido, and M. Takizawa, "A high-performance multimedia streaming model on multi-source streaming approach in peer-to-peer networks," *in Proc. IEEE AINA*, Mar. 2005, pp. 27–32.

[20] Official website. [Online]. Available: http://goalbit.sourceforge.net

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. 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. Currently he is with the Institute for Infocomm Research, A*STAR, in Singapore. His research interests include green communication and cooperative communications.

Dr. Adachi served as General Co-chair of the Tenth IEEE Vehicular Technology Society Asia Pacific Wireless Communications Symposium (APWCS) and Track Co-chair of Transmission Technologies and Communication Theory of the 78th IEEE Vehicular Technology Conference in 2013. He was recognized as the Exemplary Reviewer from IEEE Communications Letters and IEEE Wireless Communications Letters in 2012.

Investigating the Trade-off in Allocating Bits to Texture or Depth Information in Rate-constrained Multiview Video Coding

A short review for
"Joint Bit Allocation and Rate Control for Coding Multi-View Video Plus Depth Based 3D Video"

Edited by Carsten Griwodz

F. Shao, G. Jiang, W. Lin, M. Yu, and Q. Dai, "Joint Bit Allocation and Rate Control for Coding Multi-View Video Plus Depth Based 3D Video", IEEE Transaction on Multimedia, vol. 15, no. 8, pp 1843-1854, Dec. 2013.

Multiview video with depth maps is a widely used approach for the encoding, transport and presentation of free viewpoint view. Based on several years of coding research, it has also been the basis for standardization in MPEG [1] and is implemented by 3D TVs. A challenge that has not attracted much attention so far is the interactive use of multiview-plus-depth video. In this use case, sending rates are restricted, and compression choices must be made in real-time. The trade-off between depth information and texture data has mostly been approached using the rule-of-thumb that depth information is more critical than texture information. This paper from December 2013 investigates the question thoroughly.

The paper addresses the rate-control of multiview-plus-depth video that can be performed during the compression step, and preferably in or close to real-time. The authors' goal is an increase in the quality of the virtual views, where the coding rate is restricted, and a choice between allocating bits either to maintain depth information or texture information must be made. At the same time, the authors aim at an encoding system that has low enough complexity for real-world use. They demonstrate both achievements by experimental evaluation, leading to a well-rounded study.

To start out, the paper has excellent state-of-the-art overview. The state-of-the-art overview demonstrates clearly the amount of effort that has recently gone into solving the special role of the depth map. Errors due to resolution, rounding or quantization can distort the reconstructed spatial placement of texture pixels and have adverse effects on reconstructed views. But the research contribution of this paper is not the introduction of new methods to reconstruct depth more accurately. This paper uses the standardized JMVC approach [2], but investigates the visual *accuracy* of reconstructed virtual views when the bandwidth tradeoff between depth plane and texture planes is shifted in favor of the one of the other.

The most valuable contribution of this paper is that this decision can be made solely at encoding time, and does not require the modification of standardized decoders. The rate distortion-based metric has the potential of executing real-time in an encoder, and is thus applicable in multiview-plus-depth real-time encoders.

The approach does not actually address the issue of human subjective quality of the resulting virtual view video, although this is implied in the text, but optimizes for picture fidelity. This means that the results of this paper may require a revision of the optimization step (and the evaluation) if the chosen metric is found to correlate badly with subjective studies. Considering the state-of-the-art in objective evaluation methods [3], the authors choose virtual view reconstruction from the original source data as reference for the reconstructed view. Their multiview-plus-depth stream that is compressed with a JMVC encoder was modified according to their rate-control method and rendered using VSRS [4] for comparison with the reference.

With this setup, they can demonstrate by example that achieving the highest visual fidelity in reconstructed views can require an increased allocation of bit-rate share to the texture just as well as to the depth map.

The results of this paper are based on an objective quality metric whose ability to reflect humans' subjective rating is very low. It should not be used carelessly, and it is most likely that the evaluation metric used in the rate-distortion model will require future revision. More about the deficiencies of current objective metrics can be read, for example, in a paper by Bosc et al. [5].

While this weakness of existing objective quality models is problematic, it does not invalidate the achievements of the authors. The method proposed by the authors is an important improvement to the state-of-the-art in encoding multiview-plus-depth video, one that increases the frame-by-frame picture fidelity compared to earlier work. While it is possible that the paper's technique fails on optimizing subjective quality (e.g., it does not preclude opposite bit

allocation decisions in subsequent frames, potentially leading to flicker or objective displacement over time in the virtual view), this may just as well be a non-problem because depth maps do not change intensely without major changes in the texture planes. Answering this question should be one of the most important goals of future papers that aim at advancing the authors' work. At the current point in time, the authors provide us with a major improvement of the state-of-the-art and can easily be extended to take such effect into account, certainly when a better objective quality metric becomes available.

References:

[1] A. Vetro, T. Wiegand, and G. Sullivan, "Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard," *Proceedings of the IEEE*, vol. 99, no. 4, pp. 626–642, Apr. 2011.

[2] ISO/IEC MPEG&ITU-T VCEG, Draft Reference Software for MVC, Doc. JVT-AE207. London, U.K., Jul. 2009.

[3] Patrick Le Callet, Sebastian Möller and Andrew Perkis, eds., "Qualinet White Paper on Definitions of Quality of Experience (2012)", European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), Lausanne, Switzerland, Version 1.2, March 2013.

[4] ISO/IEC JTC1/SC29/WG11, 3DV/FTV EE2: Report on VSRS Extrapolation, Doc. M18356. Guangzhou, China, Oct. 2010.

[5] E. Bosc, R. Pepion, P. Le Callet, M. Koppel, P. Ndjiki-Nya, M. Pressigout, & L. Morin. Towards a New Quality Metric for 3-D Synthesi"zed View Assessment.", *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 7, pp. 1332–1343, 2011.

Carsten Griwodz leads the Media Department at the Norwegian research company Simula Research Laboratory AS, Norway, and is 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 is area chair of ACM MM 2014, and was 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

Automatic Natural Language Descriptions of Images

A short review for "From large scale image categorization to entry-level categories"

Edited by Guillaume Lavoué

Vicente Ordonez, Jia Deng, Yejin Choi, Alexander Berg and Tamara Berg, "From large scale image categorization to entry-level categories", in Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 2768-2775, 2013.

Object recognition in images, as well as image classification have always interested researchers in computer science and have been very active since the very beginning of computer science. Today's techniques can afford complex computations and begin to work quite efficiently; they can classify images into predefined categories and/or recognize objects with a very good accuracy.

A very relevant application of these techniques is the automatic text description of images (e.g. for automatic indexing of personal or public /video databases). However many different labels can be associated to a given picture; e.g. a horse can be classified into a horse, an equine, a placental mammal, a mammal, and so on…depending on the categories learned by the classification system.

As raised by the authors, what really matters for automatic picture description is the entry level words [1], i.e. the words people use to describe image content (e.g. "horse" in the example above). This question is relevant with some recent works which have tried to make connections between computer vision outputs and natural language descriptions of images [2, 3].

In this context, the authors propose a system that (1) translates encyclopedic concepts to entry-level concept and (2) predicts entry-level concepts from Images. The proposed models are learnt on several existing huge databases of information as well as crowd-sourced experiments.

First, the authors ran an experiment to gather entry-level category labels directly from people. They use Amazon Mechanical Turk to crowdsource translations of images sampled from ImageNet synsets into entry-level categories.

Then the authors propose a system able to learn translations between encyclopedic concepts (represented by ImageNet leaf categories, e.g. Chlorophyllum molybdites) and entry-level concepts (e.g. mushroom). For this task they propose two methods: the first one is based on a language-only translation using the n-gram counts of the Google1T corpus [4] and the second one considers visual information.

Finally, the authors propose a system that predicts entry-level concepts directly from images. They propose three approaches: the first one is based on a variation on the hedging approach from [5]. They train a SVM on sift features for ImageNet leaf node categories and propagate predictions for internal nodes of the hierarchy. The most relevant internal nodes are then selected as entry-level categories using n-gram counts presented above. The second method is based on supervised learning from data where people have directly provided entry level labels. For this task the authors consider the 1 million images dataset from [6] with associated captions. Finally, the third method combines the two previous approaches.

The authors evaluate both their translations and image entry-level predictions approaches. Translations approaches work well; for instance "Gordon setter" is translated to "dog", "cactus wren" is translated to "bird". Image entry-level predictions are evaluated by studying how well they can predict nouns freely associated with images by users on Mechanical Turk. That's a pretty hard task. The hybrid method is the most efficient and provides very interesting results.

In conclusion, the proposed methods work very well and the raised topic is particularly relevant with the proliferation of social image databases. This approach should open the way to a whole new research area.

References

[1] P. Jolicoeur, M. A. Gluck, and S. M. Kosslyn. Pictures and names: making the connection. *Cognitive Psychology*, 16:243–275, 1984.

[2] G. Kulkarni, V. Premraj, S. Dhar, S. Li, Y. Choi, A. C. Berg,and T. L. Berg. Babytalk: Understanding and generating simple image descriptions. In *CVPR*, 2011.

[3] A. Farhadi, M. Hejrati, M. A. Sadeghi, P. Young, C. Rashtchian, J. Hockenmaier, and D. Forsyth. Every picture tells a story: generating sentences for images. In ECCV, 2010.

[4] T. Brants and A. Franz. Web 1t 5-gram version 1. In Linguistic Data Consortium, 2006.

[5] J. Deng, J. Krause, A. C. Berg, and L. Fei-Fei. Hedging your bets: Optimizing accuracy-specificity trade-offs in large-scale visual recognition. In CVPR, 2012.

[6] V. Ordonez, G. Kulkarni, and T. L. Berg. Im2text: Describing images using 1 million captioned photographs. In NIPS, 2011.

Guillaume Lavoué received his Ph.D. degree in computer science from the University Claude Bernard, Lyon, France (2005). Since September 2006 he is associate professor at the French engineering university INSA of Lyon, in the LIRIS Laboratory. Since 2013 he is the co-chair of IEEE Technical Committee on Human Perception and Multimedia Computing (SMC society). He is author or co-author of over 60 publications in international journals and conferences. His research interests include 3D mesh analysis and retrieval, 3D data transmission and streaming (including compression and watermarking), Web 3D, Perception and Human factors for computer graphics.

An Effective Shape Similarity Fusion Method Based on Local Data Distribution

A short review for "Shape similarity analysis by self-tuning locally constrained mixed-diffusion"

Edited by Jun Zhou

Lei Luo, Chunhua Shen, Chunyuan Zhang, and Anton van den Hengel, "Shape similarity analysis by self-tuning locally constrained mixed-diffusion", IEEE Transactions on Multimedia, Vol. 15, No. 5, 2013.

Shape retrieval or matching has been an important research topic in multimedia and computer vision. Almost all retrieval methods are based on pair-wise similarity measure. During the past years, many effective measures have been proposed, for example, shape contexts [1], inner-distance shape contexts [2], clique histogram [3], and tensor-based descriptor [4].

One of the research questions is how to combine different shape measures so as to improve the shape matching accuracy. An intuitive solution is that if a shape sample has high ranking by all measures, it should be retrieved. On the contrary, if all measures give low ranking to a sample, it should be excluded. However, the real practice is not always so easy, as different measures behave differently such that unanimous decision can seldom be reached. Furthermore, noise or outliers, which are difficult to be distinguished from high ranking by only one measure, often deteriorates the retrieval performance.

To address this problem, in this paper, Lei et al. proposed a locally constrained mixed-diffusion method, so that only the nearest neighbors given by different measures contribute the to the diffusion outcome. The resulting shape similarity measure is further improved by a self-adaptive neighborhoods strategy that automatically determines the size of the neighborhood for each sample. This method has generated exceptional performance on the MPEG-7 shape dataset with 100% in the bulls eye test, which, according to the authors, is the best result that has been reported by the date of paper submission.

The technical value of this paper can be summarized in two aspects. The first one is on how to control the contribution of different measures during the diffusion process. A simple method is performing a linear combination of multiple measures. However, the authors pointed out that useful information may be submerged into the background during the diffusion process if pairwise similarities between all samples are used. On the other hand, if those low similarity sample pairs are removed, the influence from background can be reduced. Therefore, the authors proposed locally constrained mixed-diffusion, such that only the k-nearest neighbors (k-NN) of each sample are adopted in the measurement.

The question is then how to determine the appropriate sizes of neighborhoods. Manual tuning is inconvenient, and often lead to one size for each dataset. The ideal case is that the sizes are adapted to each instance in a dataset. To achieve this goal, the authors proposed to combine dominant neighbourhood [5] with the ε-nearest-neighborhood (ε-NN), which is another main contribution of this paper. Given a data sample x, all other k-NN samples are weighted by their distances to x. This forces the dominant set to be the area surrounding x after an optimization process, and guarantees the near neighbors with large weights be selected by the dominant set. The farthest neighbor of a sample in the dominant neighborhood is then used as the radius of ε-NN, which allows size of neighbourhood be automatically determined.

This method was tested on several datasets, including MPEG-7, Swedish Leaf, Nister and Stewenius, and Caltech-101 datasets. The test data various in shape, object categories, sizes and views. The proposed method has shown clear advantage over several state-of-the-art methods, such as tensor product graph with dominant neighbourhood [5], in terms of retrieval accuracy. The authors also showed that the reported method can be used for cluster analysis.

Probably the only disadvantage of this method is the computational cost. Due to the multiplication of affinity matrix in the diffusion step, and the iterative optimization for obtaining the dominant

set, the time complexity is in the order of O(n3t) for the number of data points n and iteration number t. However, this does not prevent the adoption of this approach for large datasets, because the diffusion process only needs to use the top k-NN neighbors. Furthermore, similarity fusion is a generic problem. Therefore, the proposed method can be applied to other retrieval/matching problems, but not limited to shape application.

References:

[1] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 4, pp. 509–522, 2002.

[2] H. Ling and D. Jacobs, "Shape classification using the inner-distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, pp. 286–299, 2007.

[3] H. Zhao, J. Zhou, A. Robles-Kelly, J. Lu, and J. Yang. "Automatic Detection of Defective Zebrafish Embryos via Shape Analysis". *Proceedings of the Digital Image Computing: Techniques and Applications*, pages 431-438, 2009.

[4] F. Andalo, P. Miranda, R. Torres, and A. Falcao. "Shape feature extraction and description based on tensor scale". Pattern Recognition, Vol 43, No. 1, pages 26-3, 2010.

[5] X. Yang and L. Latecki, "Affinity learning on a tensor product graph with applications to shape and image retrieval". Proceedings of the IEEE Conference on Computer Vision Pattern Recognition, pages 2369-2376, 2011.

[6] A. Kumar and H. Daumé, III, "A co-training approach for multi-view spectral clustering". Proceedings of the International Conference on Machine Learning, pages 393-400, 2011.

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

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

Optimizing 3-D Watermarking for Minimal Surface Distortion

A short review for "Optimized 3D Watermarking for Minimal Surface Distortion"

Edited by Guillaume Lavoué and Adrian G. Bors

Adrian G. Bors and Ming Luo, "Optimized 3D Watermarking for Minimal Surface Distortion", in IEEE Transactions on Image Processing, vol. 22, no. 5, pp. 1822-1835, May 2013.

We revisit this paper, reviewed in the December issue, by incorporating the perspective of the authors. There is a renewed interest in 3-D watermarking due to the increase use of 3-D printing technologies for manufacturing and concern for the copyright for the 3-D models being widely used. Digital watermarking has a very challenging set of requirements requiring non-visibility of shape changes, large message length being embedded, robustness to a variety of changes and attacks and a high level of crypto-security [1]. The challenge is represented by the fact that when trying to enforce any of these requirements, it leads to worsening the results for all the others.

3-D watermarking approaches can be classified as being in the space or in the spectral domain. For embedding the watermark message directly in the shape's geometry the local moments [1], or statistics of geodesic distances [2] have been used. Other space-based watermarking methods consider modifying the distance from vertices to the object center [3], while in [4] the quadric error metric with respect to the surface is also minimized. The initial spectral approaches were non-blind, requiring information about the original 3-D shape for extracting the embedded message code. Various types of harmonics have been used for extracting the spectral domain including the Laplace-Beltrami operator [5], radial basis functions or the Dirichlet manifold harmonics [6]. Surface changes caused by watermarking are specifically masked in [7].

The paper discussed in this R-Letter describes a non-blind approach to 3-D watermarking in the spatial domain which specifically addresses the key requirement of low visibility of the watermark. The paper proposes an optimization approach for 3-D watermarking, where the optimizing function is provided by the surface error distortion function.

The 3-D surface distortion function is based on the sum of Euclidean distances from a displaced vertex (by watermark embedding) to the original object surface, the watermarked surface as well as its original location. This error measure represents a generalization of surface distortions used in [3] and [4]. A statistical approach is employed for the 3-D watermarking. The proposed methodology relies on applying Levenberg-Marquardt [8,9] optimization in spherical coordinates in order to minimize the distortion of the object surface. The convergence of the algorithm is very quick in just a few iterations.

An additional contribution of this paper is that it provides a crypto-security study according to the Kerckhoffs's principle, according to which the watermark system is entirely known to a crypto-attacker aiming to break the watermark code.

The performance of the proposed approach is evaluated on various 3-D objects such as those commonly used in computer graphics, displaying a variety of surfaces. Surface distortion is measured using the Metro algorithm proposed in [10]. An extensive study for the robustness to additive noise, Laplacian smoothing, mesh simplification, bit quantization, sampling and remeshing the object surface, is performed. The bit capacity embedding, the influence of various parameters and receiver operating curve (ROC) analysis, when considering noise and mesh simplification, is provided as well for the proposed method. The method is shown to embed robustly watermark codes in 3-D surfaces that can resist a large variety of attacks.

The proposed methodology can be adapted for being used on other 3D object representations such as voxel-based or on parametric models after using quantization. The security of the watermark can be improved if additional key-generated parameters would be added, as suggested in the paper. The proposed surface optimisation methodology can be adapted to be used with other surface distortion errors, such as recent perceptual quality metrics [11], which

were shown to be excellent predictors of the human perception of shapes.

To conclude, this paper proposes a 3-D watermarking method in the spatial domain which is optimal with respect to a given surface distortion measure. Levenberg-Marquardt algorithm is used in spherical coordinates in order to minimise the given surface distortion measure. 3-D watermarking has many potential applications, including for graphics animation and for protecting CAD designs in the context of 3-D printing.

References

[1] A. G. Bors, "Watermarking mesh-based representations of 3-D objects using local moments," *IEEE Trans. Image Process.*, vol. 15, no. 3, pp. 687–701, Mar. 2006.

[2] M. Luo and A. G. Bors,"Surface-preserving watermarking of 3-D shapes," IEEE Trans. Image Process., vol. 20, no. 10, pp. 2813–2826, Oct. 2011.

[3] J. W. Cho, R. Prost, and H. Y. Jung, "An oblivious watermarking for 3-D polygonal meshes using distribution of vertex norms," IEEE Trans. Signal Process., vol. 55, no. 1, pp. 142–155, Jan. 2007.

[4] M. Luo and A.G. Bors,"Shape watermarking based on minimizing the quadric error metric," in Proc. IEEE Int. Conf. Shape Model. Appl., Jun. 2009, pp. 103–110.

[5] K. Wang, M. Luo, A. G. Bors, F. Denis, "Blind and robust mesh watermarking using manifold harmonics," Proc. of IEEE Int. Conf. on Image Processing, Cairo, Egypt, 2009.

[6] Y. Liu, B. Prabhakaran, X. Guo,"Spectral watermarking for parametrized surfaces," IEEE Tran. on Information, Forensics and Security, vol. 7, no. 5, pp. 1459-1471, 2012.

[7] K. Kim, M. Barni, and H. Z. Tan, "Roughness-adaptive 3-D watermarking based on masking effect of surface roughness," IEEE Trans. Inform. Forensics Security, vol. 5, no. 4, pp.721–733, Dec. 2010.

[8] K. Levenberg,"An algorithm for least-squares estimation of nonlinear parameters," Quart. Appl. Math., vol. 11, no. 2, pp. 164–168, 1944.

[9] D. Marquardt, "A method for the solution of certain non-linear problems in least squares," SIAM J. Appl. Math., vol. 11, pp. 431–441, 1963.

[10] P. Cignoni, C. Rocchini, and R. Scopigno, "Metro: Measuring error on simplified surfaces," Comput. Graph. Forum, vol. 17, no. 2, pp. 167–174, 1998.

[11] M. Corsini, M. C. Larabi, G. Lavoué, O. Petrik, L. Vasa, K. Wang, "Perceptual metrics for static and dynamic triangle meshes," Computer Graphics Forum, vol. 32, no. 1, 2013, pp. 101-125.

Guillaume Lavoué received his Ph.D. degree in computer science from the University Claude Bernard, Lyon, France (2005). Since September 2006 he is associate professor at the French engineering university INSA of Lyon, in the LIRIS Laboratory.

Since 2013 he is the co-chair of IEEE Technical Committee on Human Perception and Multimedia Computing (SMC society).

He is author or co-author of over 60 publications in international journals and conferences. His research interests include 3D mesh analysis and retrieval, 3D data transmission and streaming (including compression and watermarking), Web 3D, Perception and Human factors for computer graphics.

Adrian G. Bors (M'00–SM'04) received the M.S. degree in electronics engineering from the Polytechnic University of Bucharest, Bucharest, Romania, and the Ph.D. degree in informatics from the University of Thessaloniki, Thessaloniki, Greece, in 1992 and 1999, respectively. In 1999, he joined the Department of Computer Science, University of York, U.K., where he is currently a Lecturer. He authored and co-authored more than 90 papers in international journals and conference proceedings. His current research interests include digital watermarking, computational intelligence, computer vision and image processing.

Dr. Bors has been an Associate Editor of the IEEE Trans. on Image Processing since 2010 and was an Associate Editor of the IEEE Trans on Neural Networks from 2001 to 2009. He was a member on the Organization Committee of CAIP 201

Paper Nomination Policy

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

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

Nomination Procedure

Paper nominations have to be emailed to R-Letter Editorial Board Directors:

Irene Cheng (locheng@ualberta.ca),
Weiyi Zhang (maxzhang@research.att.com), and
Christian Timmerer
(christian.timmerer@itec.aau.at)

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

Review Process

Each nominated paper will be reviewed by members of the IEEE MMTC Review Board. To avoid potential conflict of interest, nominated papers co-authored by a Review Board member will be reviewed by guest editors external to the Board. 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 (MMTC) Officers

Chair Jianwei Huang

Steering Committee Chair Pascal Frossard

Vice Chair – North America Chonggang Wang

Vice Chair – Asia Yonggang Wen

Vice Chair – Europe Luigi Atzori

Vice Chair – Letters & Member Communications Kai Yang

Secretary Liang Zhou

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.