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

# MMTC Communications – Review

**Vol. 12, No. 3, June 2021**

IEEE COMMUNICATIONS SOCIETY

## TABLE OF CONTENTS

# Message from the Review Board Directors

Welcome to the June 2021 issue of the IEEE ComSoc MMTC Communications – Review.

This issue comprises four reviews that cover multiple facets of multimedia communication research including mobile video broadcasting, sound localization, UAV resource allocation, and 360 video streaming. These reviews are briefly introduced below.

The first paper, to be published in IEEE Transactions on Mobile Computing and edited by Dr. Takuya Fujihashi, proposes a spatial scalability enabled robust video broadcast system to accommodate diverse users with both heterogeneous resolutions and channel conditions.

The second paper, to be published in IEEE Transactions on Pattern Analysis and Machine Intelligence and edited by Dr. Debashis Sen. It designs a novel unsupervised algorithm to address the problem of localizing sound sources in visual scenes through a two-stream network structure handling each modality.

The third paper, published in IEEE Transactions on Wireless Communications and edited by Dr. Qichao Xu, proposes a resource pricing and trading scheme based on Stackelberg dynamic game to optimally allocate edge computing resources between edge devices and UAVs, as well as a blockchain-based scheme to record the entire resources trading process to protect the security and privacy.

The fourth paper, published in IEEE Transactions on Multimedia and edited by Dr. Yong Luo, explores optimal wireless streaming of a multi-quality tiled 360 virtual reality (VR) video from a server to multiple users through a multicast wireless channel.

All the authors, reviewers, editors, and others who contribute to the release of this issue deserve appreciation with thanks.

IEEE ComSoc MMTC Communications – Review Directors

Zhisheng Yan
Georgia State University, USA
Email: zyan@gsu.edu

Yao Liu
Binghamton University, USA
Email: yaoliu@binghamton.edu

Wenming Cao
Shenzhen University, China
Email: wmcao@szu.edu.cn

Phoenix Fang
California Polytechnic State University, USA
Email: dofang@calpoly.edu

# Robust Video Broadcast considering Resolution Heterogeneity of Mobile Devices

*A short review for "Robust Video Broadcast for Users with Heterogeneous Resolution in Mobile Networks"*

Edited by Takuya Fujihashi

*Y. Gui. H. Lu, F. Wu, C. W. Chen, "Robust Video Broadcast for Users with Heterogeneous Resolution in Mobile Networks," IEEE Transactions on Mobile Computing (Early Access).*

Demand for video traffic has been exponentially growing in recent years because of the development of mobile devices and video applications. Video broadcast is an effective way to reduce the traffic required to deliver the same video content for multiple mobile users. However, the heterogeneous channel conditions in mobile networks, in terms of Signal-to-Noise Ratio (SNR) have posed great challenges to existing digital video broadcast systems. Specifically, the digital video systems suffer from the cliff effect and staircase effect [1] owing to the vulnerability of communication errors. To eliminate such cliff effect and staircase effect, robust video transmission schemes [2] have aroused great interest in recent years. Since the robust video transmission schemes only consist of linear operations, the robust video transmission system intrinsically has the scalability to instantaneous channel conditions, which can provide each user with the video quality commensurate with its channel quality.

Although the channel heterogeneity of mobile devices can be solved by robust video transmission schemes, the heterogeneity of other aspects still impairs the user experience. In this paper, the authors deal with the resolution heterogeneity of mobile devices in video broadcast systems. Due to the diversity of mobile devices, such as smartphones, tablets, and laptops, the receivers in video broadcast systems have different screen sizes. One of the simplest solutions to deal with the resolution heterogeneity of mobile devices is to broadcast a single representation of a video sequence to the user devices. In this case, a device with low display resolution decodes and downsamples the received high-resolution videos. However, such a modification increases the cost and power consumption of the device. In addition, sending the details that are not shown on the display owing to the display resolution is a waste of its receiving channel bandwidth. Therefore, how to make robust video transmission adaptive to the resolution heterogeneity with one single video representation, is a very important issue in the robust video broadcast system.

Based on the above-mentioned issues, the authors propose a robust video broadcast framework called Spatial Scalability enabled Robust Video Broadcast (SSRVB), which can accommodate diverse mobile devices with heterogeneous resolutions, as well as maintaining the scalability of robust video transmission with respect to channel conditions.

SSRVB designed a novel spatial decomposition method based on linear projection to provide differentiated resolution demands Specifically, the input videos are decomposed into multiple layers, i.e., one base layer and multiple refinement layers. A base layer guarantees the base quality of the video content, while the refinement layers progressively improve video quality. Since the total number of projection values is always equal to the number of original values in each projection process, SSRVB saves bandwidth requirement.

The authors then derived the expression of the decoding distortion of each user as well as the average distortion of multiple users in the broadcast scenarios. They defined the distortion minimization problem with the consideration of two types of resource allocation including subcarrier matching and power allocation. Since this problem is an NP-hard problem, they derived a closed-form optimal power allocation solution for any given subcarrier matching. With the optimal power allocation, they designed a near-optimal and low-complexity subcarrier matching scheme based on auction theory. Finally, an iterative algorithm is used to solve this joint subcarrier matching and power allocation problem.

All evaluations were carried out based on the six test video sequences downloaded from Xiph. They compared the performance of the proposed SSRVB with the existing robust video transmission schemes, including ECast [3], MCast [4], discrete Wavelet transform (DWT)-based, and H.264/SVC with convolutional codes and hierarchical modulation (SVC-HM) schemes. For comparison, they considered decomposing the source video into three layers, and the mobile user requested the low resolution, the middle resolution, and the high-resolution videos, respectively.

The evaluation results showed that SSRVB achieves the best average performance in terms of video quality. For example, SSRVB performs better for all users compared with the DWT-based scheme. It means the spatial decomposition in the proposed SSRVB is more suitable for robust video broadcast compared with the spatial decomposition in DWT. In addition, SSRVB outperforms the digital-based SVC-HM scheme owing to cliff effect prevention.

## References:

[1] S. Pudlewski, N. Cen, Z. Guan, and T. Melodia, "Video transmission over lossy wireless networks: A cross-layer perspective," IEEE Journal of Selected Topics in Signal Processing, vol. 9, no. 1, pp. 6–21, 2015.

[2] S. Jakubczak and D. Katabi, "SoftCast: One-size-fits-all wireless video," ACM SIGCOMM Computer Communication Review, vol. 41, no. 4, pp. 449-450, Oct. 2011.

[3] Z. Zhang, D. Liu, X. Ma, X. Wang, "ECast: An enhanced video transmission design for wireless multicast systems over fading channels," IEEE System Journal, vol. 11, no. 4, pp. 2566-2577, Dec. 2017.

[4] C. He, H. Wang, Y. Hu, Y. Chen, X. Fan, H. Li, and B. Zeng, "MCast: High-Quality Linear Video Transmission With Time and Frequency Diversities," IEEE Transactions on Image Process., vol. 27, no. 7, pp. 3599-3610, 2018.

**Takuya Fujihashi** received the B.E. degree in 2012 and the M.S. degree in 2013 from Shizuoka University, Japan. In 2016, he received Ph.D. degree from the Graduate School of Information Science and Technology, Osaka University, Japan. He is currently an assistant professor at the Graduate School of Information Science and Technology, Osaka University since April, 2019. He was an assistant professor at the Graduate School of Science and Engineering, Ehime University, Japan from Jan. 2017 to Mar. 2019. He was research fellow (PD) of Japan Society for the Promotion of Science in 2016. From 2014 to 2016, he was research fellow (DC1) of Japan Society for the Promotion of Science. From 2014 to 2015, he was an intern at Mitsubishi Electric Research Labs. (MERL) working with the Electronics and Communications group. He selected one of the Best Paper candidates in IEEE ICME (International Conference on Multimedia and Expo) 2012. His research interests are in the area of video compression and communications, with a focus on multi-view video coding and streaming.

.

# Nearly-Unsupervised Localization of Sound Sources in Videos

*A short review of "Learning to Localize Sound Sources in Visual Scenes: Analysis and Applications"*

Edited by Debashis Sen

*A. Senocak, T. -H. Oh, J. Kim, M. -H. Yang and I. S. Kweon, "Learning to Localize Sound Sources in Visual Scenes: Analysis and Applications," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 5, pp. 1605-1619, May 2021.*

Humans perceive events in scenes by sensing multi-modal information through different sensory organs. Just like visible content, sound in scenes contains rich information for understanding spatio-temporal cues of the objects /events that generate sound. Humans easily locate and interpret the visual appearance of a sound source by sensing the audio-visual data. Humans acquire this capability by learning to correlate audio and visual data over time without any supervision [1]. In this context, the authors investigate in the paper whether machines can also learn to perform human-like audio visual event perception in an unsupervised manner.

There has been a plethora of investigations in audio-visual learning, but sound source localization through audio-visual data without any handcrafted prior has hardly been attempted in an unsupervised manner. Learning such a mechanism is a challenging task. Moreover, it is very difficult to learn the localization of sound sources based on image sequences and mono-channel audio.

In the paper, the authors present a framework comprising of three networks, namely sound network, visual network, and localization network for sound source localization in videos. Sound network learns the semantic relationship between the mono-channel audio and single frame without any motion cues. The network emphasizes on extracting features related to context and concepts of sound signal. The visual network extracts visual context information while preserving the spatial information. Finally, the localization network perform the sound source localization by utilizing the extracted concepts from sound and visual networks. The network generates a confidence score map indicating the location. This is modelled by an attention mechanism similar to the human visual system.

The first few layers in the sound network are similar to SoundNet [2]. The rest of the layers are designed with ReLU [3] with fully connected (FC) layers [3]. The architecture of visual network is similar to the VGG-16 model [3]. To perform localization usin the localization network, the context representation vector similar to [4, 5] is used.

The learning model of the proposed framework determines whether the audio signals share local similarity with video frames. The framework generates representations from both sound and visual networks, which are projected to be similar or dissimilar to each other in a latent feature space. An unsupervised loss function is proposed by computing a distance in the common feature space. Empirical evidence suggests that such a framework with unsupervised learning works in a wide variety of scenarios. However, in some cases it is also found to arrive at wrong conclusions in terms of localization. This suggests that an unsupervised learning mechanism alone using a video frame and mono-channel audio in the proposed setting may not always be the best suited.

Further investigations show that in the early stages of the learning, the model can generate wrong conclusions. However, the conclusion can be improved with weak supervision. This leads the authors to device a mechanism to provide a little supervision in the form of prior-knowledge to the model. Interestingly, the problem of false conclusion is resolved using the semi-supervision setting that uses a supervised loss function as a prior. As an addition, the authors design a deep learning based framework which allows supervised, semi-supervised and fully supervised settings as choices, which can be selected according to the availability of the annotated data.

This paper also contributes a new evaluation measure, namely, localization consensus intersection over union (cIoU) similar to the consensus metric in the VQA task [6], but in the sound source localization context. Further, this paper contributes a new sound source dataset with

annotation for supervised training and performance evaluation.

Extensive evaluation shows that the proposed framework is able to localize the sound source very well with a semi-supervised setting. An analysis with non-object and ambient sounds shows that the model deals well with off-context sound. The proposed framework successfully generates the visual embedding that can be beneficial to analyze the effectiveness of the learned representation. The authors show the effectiveness of the proposed model in cross-modal searching, like audio query-based video retrieval and video query-based audio retrieval. Further, the authors show the utility of the proposed framework in sound saliency-based automatic camera view panning in 360° videos.

This paper tackles a relatively new and difficult problem of sound source localization in videos. Experimental results show that a completely unsupervised learning can perform well, but would sometimes draw wrong conclusions. However, it can be resolved with a small amount of supervision. The work shows its potential in different applications like multi-modal retrieval, sound-based saliency, and representation learning-based applications.

**Acknowledgement**
The editor would like to thank Mr. Sobhan Dhara of IIT Kharagpur, India for providing a preliminary draft of this review.

## References:

[1] W. W. Gaver, "What in the world do we hear? An ecological approach to auditory event perception," *Ecological Psychol.*, vol. 5, pp. 1–29, 1993.

[2] Y. Aytar, C. Vondrick, and A. Torralba, "SoundNet: Learning sound representations from unlabeled video," in *Proc. 30th Int. Conf. Neural Inf. Process. Syst.*, 2016, pp. 892–900.

[3] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Representations*, 2015, pp. 1–14.

[4] K. Xu et al., "Show, attend and tell: Neural image caption generation with visual attention," in *Proc. Int. Conf. Mach. Learn*., 2015, pp. 2048–2057.

[5] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *Proc. Int. Conf. Learn. Representations*, 2015, pp. 1–15

[6] K. Kafle and C. Kanan, "Visual question answering: Datasets, algorithms, and future challenges," *Comput. Vis. Image Understanding*, vol. 163, pp. 3–20, 2017

**Debashis Sen** is an Assistant Professor in the Department of Electronics and Electrical Communication Engineering and a faculty in the Centre of Excellence in Advanced Manufacturing Technology of Indian Institute of Technology - Kharagpur. He received his Ph.D. in Image Processing from Jadavpur University, Kolkata, India and his M.A.Sc. in Electrical Engineering from Concordia University, Montreal, Canada. He was a postdoctoral researcher at the Multimedia Analysis and Synthesis Laboratory, National University of Singapore and at the Center for Soft Computing Research, Indian Statistical Institute. He currently heads the Vision, Image and Perception research group and the ArtEye Lab in his department, which are funded by multiple agencies of Government of India and prominent industries in India. His current research interests are in Vision, Image and Video Processing, Uncertainty Handling, Eye Movement Analysis, Machine Vision and Deep Learning. He has authored/co-authored more than 50 research articles in high impact journals and conferences. Dr. Sen is on the editorial board of IET Image Processing, and Springer's Circuits, Systems and Signal Processing. He has received a young scientist award from The Institution of Engineers (India), a Qualcomm Innovation Fellowship, an ERCIM Alain Bensoussan Fellowship, a Ministry of Manpower (Singapore) Research Fellowship and a couple of best paper awards from IET.

# A UAV Assisted Secure Edge Computing Resource Allocation Scheme

*A short review for "Edge Computing Resource Allocation for Unmanned Aerial Vehicle Assisted Mobile Network With Blockchain Applications"*

Edited by Qichao Xu

*H. Xu, W. Huang, Y. Zhou, D. Yang, M. Li and Z. Han, "Edge Computing Resource Allocation for Unmanned Aerial Vehicle Assisted Mobile Network With Blockchain Applications," IEEE Transactions on Wireless Communications, vol. 20, no. 5, pp. 3107-3121, May 2021.*

With the rapid development of the Internet of Things(IoT), the number of various types of mobile terminals has grown rapidly, resulting in a large amount of data and computing network service requirements. Mobile Edge Computing (MEC), as a potential paradigm, extends the cache and computing capabilities of cloud computing to the edge of the network, bringing a better network service experience for mobile users[1],[2]. In particular, in some scenarios where mobile users are sparsely distributed or in extreme cases, traditional mobile edge computing cannot be directly applied [3],[4]. Unmanned aerial vehicles(UAVs) can be used to assist edge computing because of its flexible deployment, low cost, and high mobility. The edge computing stations (ECSs) provide services such as communication and computing for mobile users. UAVs obtain edge computing resources from ECSs to complete computing tasks of mobile users, which further enhances the quality of service(QoS) for mobile users.

However, due to the open and wireless characteristics of UAV communication, some security and privacy issues will be brought about in the process of resource allocation between ECSs and UAVs. For example, the ECS may refuse to acknowledge the receipt resource request of the UAV, the UAV may pretend that it has not received the resources of the ECS, and the privacy information leakage that may be caused by the resource transaction process. These obstacles will hinder the resource trading between UAVs and ECSs. Therefore, it is necessary to design a security mechanism to solve the security and privacy issues in the resource transaction process.

This paper proposes a blockchain-based dynamic resource allocation scheme in a UAV-assisted mobile edge computing network, including the ECS layer as the resource provider, the UAV layer as the resource purchaser and the mobile user layer as the service requester. Meanwhile, in the blockchain, the ECS also serves as the publisher and miner of the mining task in the execution process of the blockchain. This resource allocation plan uses the UAV as a bridge to connect the ECSs and the mobile users. The mobile user uploads a task request to the nearest UAV, after the UAV receives the task request, it purchases computing resources from the edge computing station to complete the user's request service. Blockchain is used to ensure safe resource transactions between the UAV and the ECS. After the UAV completes the resource transaction with the ECS, the ECS publishes a mining task, and the miner pack the transaction records into a block and add to the blockchain after verification among nodes of ECSs. To simulate the dynamic changes of the mobile user's demand received by the UAV, the dynamic evolution process is modeled by a differential equation. To attain the control price decision of the ECS and the resource selection decision of the UAV in the resource allocation process between the ECSs and the UAVs, the Stackelberg game is used to model this process to obtain the optimal resource allocation.

Therefore, the authors' main contribution in this article is to propose a secure resource pricing scheme based on blockchain to promote resource allocation between ECSs and UAVs and maximize the utility of ECSs and UAVs, and solve the security and privacy issues involved in resource transactions. The resource allocation problem between ECSs and UAVs is modeled as a Stackberg game model, which optimizes the resource pricing of ECSs and the resource selection of UAVs, improves the efficiency of resource utilization and provides satisfactory services for mobile users.

In order to solve the problem that mobile users cannot access edge nodes to obtain network services in special scenarios, this paper proposes to use UAV to provide mobile users with computing and communication capacities. Due to the limited resources, UAVs provide users with satisfactory services by purchasing resources from

ECSs. The service request from the users to the UAV affected by the service price and the resource price of the ECS changes over time. In order to obtain more benefits, the ECS first announces the resource price, and then the UAV selects the resource demand, and the resource allocation process is modeled as a two-stage Stackelberg game, the equilibrium solution of the ECS and the UAV is analyzed to obtain their optimal profit. Because resource transactions occur in an untrust environment, dishonest behaviors and even conflicts may occur in the process of resource transactions. In order to ensure that the security and privacy of resource transactions, blockchain technology is adopted to achieve secure resource transaction without the three-party trust organization. In the operation process of the blockchain, the mining reward is used to motivate the ECSs to participate in the mining task.

In addition, the problem of optimal resource requirements for UAVs is modeled as a differential game. Differential game originates from optimal control theory and game theory. It is a continuous dynamic random game which was first proposed by Isaacs in 1965[5],[6]. Different UAVs select their own resource requirements to obtain the optimal utility within a period of time. After a fixed period of negotiation and adjustment, the UAVs can obtain their own optimal resource choices. In this paper, Bellman dynamic programming is used to obtain the equilibrium solution of the UAV in the open-loop and feedback situations.

Extensive simulations have evaluated the performance of the proposed dynamic resource allocation scheme. The simulation results show that the proposed scheme analyzes the optimal resource allocation strategy of the ECS and the optimal dynamic resource selection strategy of the UAV in the open loop situation and the feedback situation respectively, and achieves the optimal objective of the ECS and the UAV.

In summary, the proposed dynamic resource allocation mechanism based on the Stackelberg dynamic game has been proved to achieve the dynamic optimal utility equilibrium of resource allocation between ECSs and UAVs. The edge computing station controls the price of resources to allocate resources to UAVs. Blockchain technology solves the security and privacy issues involved in the process of resource transactions.

## References:


[1]. Abbas, Y. Zhang, A. Taherkordi, and T. Skeie, "Mobile edge computing: A survey," *IEEE Internet Things J.*, vol. 5, no. 1, pp. 450–465, Feb. 2018.

[2]. S. Sardellitti, G. Scutari, and S. Barbarossa, "Joint optimization of radio and computational resources for multicell mobile-edge computing," *IEEE Trans. Signal Inf. Process. over Netw.*, vol. 1, no. 2, pp. 89–103, Jun. 2015.

[3]. F. J. Martinez, C.-K. Toh, J.-C. Cano, C. T. Calafate, and P. Manzoni, "Emergency services in future intelligent transportation systems based on vehicular communication networks," *IEEE Intell. Transp. Syst. Mag.*, vol. 2, no. 2, pp. 6–20, Oct. 2010.

[4]. E. Demir, T. J. Chaussalet, H. Xie, and P. H. Millard, "Emergency readmission criterion: A technique for determining the emergency readmission time window," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 5, pp. 644–649, Sep. 2008.

[5]. I. M. Mitchell, A. M. Bayen and C. J. Tomlin, "A time-dependent Hamilton-Jacobi formulation of reachable sets for continuous dynamic games," in *IEEE Transactions on Automatic Control*, vol. 50, no. 7, pp. 947-957, July 2005.

[6]. A. Isidori and A. Astolfi, "Disturbance attenuation and H/sub infinity /-control via measurement feedback in nonlinear systems," in *IEEE Transactions on Automatic Control*, vol. 37, no. 9, pp. 1283-1293, Sept. 1992.



**Qichao Xu**, Ph.D, is an Assistant Professor in the School of Mechatronic Engineering and Automation, Shanghai University. He received the Ph.D. degree from Shanghai University, Shanghai, China, in 2019. His research interests include Internet of Things, autonomous driving vehicles, and trust management. He has published more than 50 papers in prestigious journals such as IEEE TIFS, IEEE TMM, IEEE TITIS, IEEE TII, IEEE TVT, IEEE TBD and IEEE IoTs, in prestigious conferences such as IEEE ICC, IEEE INFOCOM.

# Optimization Frameworks for Adaptive Multicast Steaming of 360 VR Video

*A short review for "Optimal Wireless Streaming of Multi-Quality 360 VR Video by Exploiting Natural, Relative Smoothness-enabled and Transcoding-enabled Multicast Opportunities"*

Edited by Yong Luo

*K. Long, Y. Cui, C. Ye, and Z. Liu, "Optimal wireless streaming of multi-quality 360 VR video by exploiting natural, relative smoothness-enabled and transcoding-enabled multicast opportunities," in IEEE Trans. Multimedia, 2021 (Early Access).*

360 virtual reality (VR) video is typically recorded using a special rig of multiple cameras or a dedicated camera containing multiple camera lenses. 360-degree video is usually viewed via personal computers, mobile devices, or dedicated head-mounted displays. A user can view the scene of interest in any direction at any time, hence enjoying an immersive viewing experience.

Wireless streaming of 360 VR video lifts geographical or behavioral restrictions and has received growing interest. The tiling technique is widely adopted to improve the transmission efficiency of 360 VR video, which has a much larger encoding rate than traditional video. Specifically, 360 VR video is divided into tiles, and only the set of tiles covering a user's predicted field-of-view (FoV) is transmitted to him to reduce communications resource consumption. In practice, users may have heterogeneous conditions (e.g., channel conditions, display resolutions, etc.). Pre-encoding each tile into multiple representations with different quality levels and performing bitrate (quality) adaptation according to a user's condition can effectively alleviate rebuffering.

Recently, several works investigate adaptive wireless streaming of tiled 360 VR videos in the unicast scenario where multiple users request different 360 VR videos [1]. In some VR applications, such as VR gaming, VR military training, and VR sports, 360 VR video has to be transmitted to multiple users with overlapping FoVs simultaneously. When a tile is requested by multiple users concurrently, multicast opportunities can improve wireless transmission efficiency. Very few works use multicast opportunities in adaptive wireless streaming of tiled 360 VR video in the multicast scenario [2, 3, 4], which is a more challenging problem. Optimizing rate adaptions for tiles separately provides simple problem formulations but yields high computational complexity and inevitable quality variation in each FoV [2, 4]. Besides, for given FoVs, optimizing rate adaptions for tiles based on instantaneous channel conditions may not lead to practical solutions, as users' channel conditions change much faster than their FoVs [2, 4]. Last but not least, communication resources may not be sufficient for multicast steaming of 360 VR video to many users [2, 3, 4]. Therefore, it is highly desirable to have tractable and reasonable optimization frameworks for adaptive multicast streaming of 360 VR video.

This paper systematically investigates adaptive multicast streaming of 360 VR video from one server to multiple users in a wireless network. Specifically, two requirements for quality variation in one FoV, i.e., the absolute smoothness requirement (identical quality within one FoV) and the relative smoothness requirement (limited quality variation within one FoV), are considered. Two video playback modes, i.e., the direct-playback mode (without user transcoding) and transcode-playback mode (with user transcoding), are considered. It is worth noting that transcode-playback mode exploits not only conventional communications resources but also new computation resources available at the users' side. Furthermore, encoding rates of tiles are adapted to channel statistics rather than instantaneous channel conditions. The authors propose optimization frameworks for adaptive multicast streaming of 360 VR video in the four cases with different requirements for quality variation and video playback modes.

Besides natural multicast opportunity, two new types of multicast opportunities, namely, relative smoothness-enabled multicast opportunity and transcoding-enabled multicast opportunity, are introduced. The former allows a flexible tradeoff between viewing quality and communications resource consumption, whereas the latter enables a flexible tradeoff between computation and communications resource consumptions.

An elegant notation system is proposed to specify the relation between a set of tiles and their target user group. Furthermore, a novel mathematical model is proposed to characterize the impacts of multicast opportunities on the average transmission energy and transcoding energy under controllable quality variation for tiles in an FoV. The notation system and mathematical model greatly facilitate optimal exploitation of potential multicast opportunities in multicast streaming of 360 VR video.

The authors formulate the optimization of transmission resource allocation, playback quality level selection, and transmission quality level selection in the four cases with different requirements for quality variation and video playback modes. In particular, the minimization of the average transmission energy is considered in the two cases without user transcoding. In contrast, the minimization of the weighted sum of the average transmission energy and the transcoding energy is considered in the two cases with user transcoding. Optimization techniques are adopted to solve the challenging optimization problems in the four cases. By comparing the optimal values in the four cases, the authors prove that the energy consumption reduces when more multicast opportunities are utilized.

Numerical results show substantial gains of the proposed solutions over existing schemes in all four cases and demonstrate the importance of effective exploitation of the three types of multicast opportunities for adaptive multicast streaming of tiled 360 VR video.

In summary, this paper proposes novel ideas, fundamental optimization frameworks, and practical insights for adaptive multicast streaming of 360 VR video. The proposed solutions can be extended to wireless systems with advanced physical layer techniques [5].

## References:

[1] J. Chakareski, "Viewport-adaptive scalable multi-user virtual reality mobile-edge streaming," IEEE Trans. Image Process., vol. 29, pp. 6330- 6342, May 2020.

[2] H. Ahmadi, O. Eltobgy, and M. Hefeeda, "Adaptive multicast streaming of virtual reality content to mobile users," in Proc. of ACM Multimedia, Oct. 2017, pp. 170–178.

[3] C. Guo, Y. Cui, and Z. Liu, "Optimal multicast of tiled 360 VR video in OFDMA systems," IEEE Commun. Lett., vol. 22, no. 12, pp. 2563–2566, 2018.

[4] N. Kan, C. Liu, J. Zou, C. Li, and H. Xiong, "A server-side optimized hybrid multicast-unicast strategy for multi-user adaptive 360-degree video streaming," in Proc. of IEEE ICIP, Sep. 2019, pp. 141–145.

[5] C. Guo, L. Zhao, Y. Cui, Z. Liu, and D. W. K. Ng, "Power-efficient wireless streaming of multi-quality tiled 360 VR video in MIMO- OFDMA systems," IEEE Trans. Wireless Commun., 2021. (Early Access)

**Yong Luo**, Ph.D, is a Professor with the School of Computer Science, Wuhan University, China. He received B.E. and D.Sc. degrees in Computer Science from the Northwestern Polytechnical University and Peking University, China, in 2009 and 2014, respectively.

His research interests are primarily on machine learning and data mining with applications to visual information understanding and analysis. He has authored or co-authored over 40 papers in top journals and prestigious conferences including IEEE T-PAMI, IEEE T-NNLS, IEEE T-IP, IEEE T-KDE, IEEE T-MM, WWW, IJCAI, AAAI, CIKM, ICDM and ICME. He received the IEEE Globecom 2016 Best Paper Award.

## MMTC Communications – Review Editorial Board

## Multimedia Communications Technical Committee Officers

**Chair:** Jun Wu, Fudan University, China
**Steering Committee Chair:** Joel J. P. C. Rodrigues, Federal University of Piauí (UFPI), Brazil
**Vice Chair – America:** Shaoen Wu, Illinois State University, USA
**Vice Chair – Asia:** Liang Zhou, Nanjing University of Post and Telecommunications, China
**Vice Chair – Europe:** Abderrahim Benslimane, University of Avignon, France
**Letters & Member Communications:** Qing Yang, University of North Texas, USA
**Secretary:** Han Hu, Beijing Institute of Technology, China
**Standard Liaison:** Guosen Yue, Huawei, 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.