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

MMTC Communications – Review

IEEE COMMUNICATIONS SOCIETY

Vol. 12, No. 4, August 2021

TABLE OF CONTENTS

Message from the Review Board Directors

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

This issue comprises four reviews that cover multiple facets of multimedia communication research including UAV-aided multi-access edge computing, depth video recovery, federated learning, and artifacts removal in video-based point cloud compression. These reviews are briefly introduced below.

The first paper, published in IEEE Transactions on Vehicular Technology and edited by Dr. Qin Wang, proposes a solution for reducing the network latency in multi-access edge computing with the assistance of UAVs, where the UAV communicates with the edge node via a millimeter wave backhaul.

The second paper, edited by Dr. Ye Liu, was published in the 2020 IEEE International Conference on Multimedia and Expo (ICME). It proposes a two-stage depth video recovery method that can enhance depth videos while achieving long-range temporal consistency among video frames.

The third paper, edited by Dr. Qichao Xu, was published in IEEE Journal on Selected Areas in Communications. It proposes an asynchronous federated learning framework for UAV-enabled wireless networks, allowing machine learning model training without transmitting raw data to UAV servers.

The fourth paper, to be published in IEEE Transactions on Multimedia, and edited by Dr. Roberto G. de A. Azevedo, proposes a solution for removing artifacts introduced by quantization errors in video-based point cloud compression.

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

UAV-Aided Low Latency Multi-Access Edge Computing

A short review for "UAV-Aided Low Latency Multi-Access Edge Computing"

Edited by Qin Wang

Y. Yu, X. Bu, K. Yang, H. Yang, X. Gao, and Z. Han, "UAV-Aided Low Latency Multi-Access Edge Computing," IEEE Transactions on Vehicular Technology, vol. 70, no. 5, May 2021.

In recent years, unmanned aerial vehicle (UAV) is proposed as a promising technique in the future network. The communication network density is booming recently, which is driven by the explosive data traffic [1], [2]. UAV communication has many benefits especially for the flexibility and economic saving. The limited load capacity on the UAV is a significant weakness for the heavy tasks. Both UAV propulsion and signal transmission power consumptions are popular topics in the UAV communications for years [3]. However, industry and researchers have realized that latency has a stronger influence on the quality of services.

Multi-access edge computing (MEC) is proposed for the future networks. MEC is a promising technique to extend the cloud computing service closer to the user equipment. For example, popular video content can be stored in the edge server for downloading, which reduces the delay and cost of data transmission. Furthermore, the MEC server can also be adopted to implement intelligent image analysis, including capturing images in real-time and analyzing image content, without transferring data back to the data center. Although traditional cloud computing is compelling, the latency and inflexible architecture are intolerable for the new services. The latency of the network is becoming more and more essential for emerging techniques.

There are many challenges in developing a low latency communication network. First, the network architecture should be refactored, which is financial consuming [4]. Moreover, it is difficult to balance latency with other factors (such as data rate, reliability, energy consumption, etc.), since they contradict each other in many cases.

In this paper, a UAV-aided MEC network with millimeter wave (mmWave) backhaul is modeled to minimize the computing latency of all users. The users have two communication links to offload the tasks, namely UAV link and ad hoc link. They assume that the UAV cannot compute the offloaded data as an edge node. This is because the UAV often has a limited load, which cannot afford delay-sensitive tasks in many cases. Thus, the UAV can only collect the data and send back to the edge node for the further process through the mmWave backhaul. By adopting mmWave communication between the UAV and the edge node, the transmission delay can be significantly reduced.

In our proposed network, the users are labeled as UAV users or ad hoc users. A UAV user chooses to offload the task to the UAV, and the UAV sends the task to the edge node for computing. An ad hoc user will offload the task through the ad hoc link to the edge node. Each user chooses one route during the execution. The channel models between the users and the UAVs are formulated based on the assumption that the channels in the network, including the users-to-UAV task offloading channels and the mmWave backhaul, are LoS (Line of Sight)-dominated.

In order to optimize total delay of the edge computing network, the best routes on the ground for the users should be provided in advance. Thus, the routing problem is proposed and solved firstly. A commonly used routing model [5] is adopted in the paper. For the ad hoc link, each user has different routes connecting to the edge node. All users are modeled as a directed graph. The edge computing users are assumed static and they offload tasks using orthogonal channels to avoid congestions. For each user, there are several potential paths to the edge node. The total delay is the summation of the transmission delay for all feasible paths for all users. The routing problem is formulated to minimize the delay in the network.

Then, the resource allocation optimization and UAV trajectory design are jointly considered to minimize the total delay in the network. This problem is designed to minimize the total delay of executing all users' tasks in the MEC network. For simplicity, they only involve the uplink delay problem in the formulation. There are several types of delay related to the problem, which are the transmission delay of users to the UAV, the UAV to the edge node, ad hoc link, and the computing delay at the edge node. The proposed utility function is formulated by considering these types of delay and the weights can be added to reflect the importance of delay types or the priority of user. The transmission delay and the computing delay are both considered to minimize the total delay of the proposed edge computing network.

In order to solve the proposed mixed integer nonconvex programming (MINCP) problem, the authors propose a designed algorithm framework. The algorithm framework can solve the problem iteratively and efficiently, which is composed of generalized Benders decomposition (GBD), alternating direction method of multipliers (ADMM), Dinkelbach algorithm, and successive convex approximation (SCA) algorithm. The GBD separates the integer variables and continuous variables in the outer loop. In the inner loop, a joint ADMM, Dinkelbach algorithm, and SCA algorithm is proposed to deal with the nonconvex fractional programming problem, which is the continuous primal problem.

Numerical results demonstrate that the UAV flies with a high speed when there are few users around the initial point. The UAV tends to stay closer and longer in the center of the area, where the user density is high. The users near the edge node are optimized to be ad hoc users and the UAV keeps away from the ad hoc users. In addition, the proposed trajectory has the lowest delay compared to the other schemes in which the UAV is stationary at the center point of the area, or the UAV can only move along the diagonal during the flight. It is also demonstrated that the proposed algorithm is more efficient than the BCD with relaxation algorithm.

In summary, the proposed low latency multi-access edge computing network has a better performance on reducing network latency with the assistance of UAVs.

References:

[1] J. An, K. Yang, J. Wu, N. Ye, S. Guo, and Z. Liao, "Achieving sustainable ultra-dense heterogeneous networks for 5G," IEEE Commun. Mag., vol. 55, no. 12, pp. 84–90, Dec. 2017.

[2] Z. Ni, J. A. Zhang, X. Huang, K. Yang, and J. Yuan, "Uplink sensing in perceptive mobile networks with asynchronous transceivers," IEEE Trans. Signal Process., vol. 69, pp. 1287–1300, 2021.

[3] L. Li, Q. Cheng, K. Xue, C. Yang, and Z. Han, "Downlink transmit power control in ultra-dense UAV network based on mean field game and deep reinforcement learning," IEEE Trans. Veh. Technol., vol. 69, no. 12, pp. 15 594-15 605, Dec. 2020.

[4] X. Gao, S. Feng, D. Niyato, P. Wang, K. Yang, and Y. Liang, "Dynamic access point and service selection in backscatter-assisted rf-powered cognitive networks," IEEE Internet Things J., vol. 6, no. 5, pp. 8270–8283, Oct. 2019.

[5] J. Chang and L. Tassiulas, "Energy conserving routing in wireless ad-hoc networks," in Proc. IEEE Int. Conf. Comput. Commun., 2000, pp. 22–31.

Qin Wang, Ph.D, is an Associate Professor at Nanjing University of Posts and Telecommunications (NJUPT), China. She received B.S. and Ph.D degrees from NJUPT, in 2011 and 2016. Prior to joining NJUPT, she was with the New York Institute of Technology (NYIT) between Feb. 2017 and Aug. 2020. From July 2018 to June 2020, she was a Postdoctoral Research Fellow at NJUPT. From 2015 to 2016, she was a visiting scholar at San Diego State University, USA. Her research interests include multimedia communications, multimedia pricing, resource allocation in 5G, and Internet of Things. She has published papers in prestigious journals such as IEEE Transactions on Vehicular Technology and IEEE Communications Magazine, in prestigious conferences such as IEEE INFOCOM SDP Workshop.

Practical Depth Video Recovery from Low-Quality RGB-D Frames

A short review for "Two-Stage Depth Video Recovery with Spatiotemporal Coherence"

Edited by Ye Liu

Q. Li, J. Guo, Q. Tang, Y. Guo and J. Qian, "Two-Stage Depth Video Recovery with Spatiotemporal Coherence," in 2020 IEEE International Conference on Multimedia and Expo (ICME), 2020.

Depth maps [1], which are able to present the distances from a viewpoint to the surfaces of objects, are promising in computer vision and graphics. They can support a variety of applications including shallow depth simulating, 3D model reconstruction, scene rendering, and so forth. However, a challenge faced in current is that the commonly used low-cost depth sensors cannot capture high-quality depth images due to sensor noise and imperfect sampling such as the missing and invalid data. As a result, depth image enhancement [2] becomes a fundamental research direction and is paid great attention in recent years.

Previous research efforts on this topic mainly focus depth recovery either using Markov random field, joint bilateral filters, guided image filters, or deep learning-based approaches recently. However, most of them are designed for still image, and cannot well support depth video recovery because although they can achieve high quality of restoration in every frame, temporal consistency between the frames is a big problem in video situations. On the other hand, some depth video recovery methods based on weighted mode filtering and static structure were also proposed to improve temporal consistency for depth video recovery. But the problems about short-length depth sequence and failing issue with sliding camera need to be solved [3].

The above observations motive the authors to design an enhanced solution that can satisfy both restoration quality of every frame and temporal consistency for long-length depth video sequence. Thus, a two-stage depth video recovery with spatio-temporal coherence is proposed. Extensive experiments were also conducted, that show the proposed method outperforms the state-of-the-art solutions that based on weighted mode filtering (WMF) and static structure (SS) in synthetic images scenarios and real-world cases.

The proposed two-stage depth video recovery method consists of six steps, including (a) input of an RGB-D frames, (b) key frames selection, (c) Markov random field-regularized low-rank matrix completion, (d) key frames recovery, (e) confidence-guided Laplacian smoothing, and (f) output of the recovered frames, respectively. Here, the step (c) is designed to ensure high quality key frame recovery with RGB-D video streams, while the temporal consistency of video frames is improved through the step (e). The details of the two core mechanisms are presented in the follows.

At the first stage for recovering the key depth frames, the RGB-D frames are firstly divided into groups and one frame is selected as the key one in each group. Then, the well-known SIFT flow method [4] is adopted to warp the neighbors of each key frame for better spatial correspondence. It is worth noting that although the zero-rows/columns have little impact on the low rank property, the missing data is a big issue affecting matrix completion. To mitigate this issue., a new Markov random field-regularized low-rank matrix completion is designed, in which the new completion model and regularization term are formulated along with the discussion of chosen parameters.

At the second stage for guaranteeing temporal consistency, the authors formulate the in-between depth video frame recovering as a global optimization problem. First, associated tensor graph is established to map the spatiotemporal continuity of the video frames. Second, a novel confidence-guided Laplacian smoothing algorithm is designed for enhancing the in-between frames. The depth confidence map, the percentage of valid pixels in a region, pixel update, are well defined with in depth discussion, as well as the calculation of unnormalized correlation degree. Furthermore, convergence acceleration is conducted.

The proposed two-stage depth video recovery method was evaluated through experimental studies in either synthetic depth videos and many real-world situations in terms of still images, RGB-D videos, and ablation studies. The examining metrics are the mean absolute error (MAE), the peak signal-to-noise (PSNR), and the structural similarity index method (SSIM), respectively. Compared with the classical depth recovery approaches, the proposed one has better recovery results without edge blurring and ghosting, and its performance closes to the ground truths for still image. In the real RGB-D video cases, the results show that the proposed method has least error metrics, and thus can achieve the best pixel-wise and structural similarity. Finally, the ablation studies show that temporal continuity is maintained well with the help of the designed confidence-guided Laplacian smoothing algorithm.

In addition, the authors also provide supplemental materials to give a detail description of the implementation in the first stage, parameter settings in the experiments, further visual quality comparisons, and the evaluation on synthetic depth videos. These materials help the community to fast verify its authenticity, effectiveness, and develop new approach to promote this research direction.

In summary, this paper presents novel ideas, practical efforts, and solid scientific contributions for depth video recovery. This work promotes the development of depth map in computer vision and graphics. A broad of potential applications is thus expected [5].

References:

[1] R. Phan and D. Androutsos, "Robust Semi-Automatic Depth Map Generation in Unconstrained Images and Video Sequences for 2D to Stereoscopic 3D Conversion," in IEEE Transactions on Multimedia, vol. 16, no. 1, pp. 122-136, Jan. 2014.

[2] W. Liu, X. Chen, J. Yang and Q. Wu, "Robust Color Guided Depth Map Restoration," in IEEE Transactions on Image Processing, vol. 26, no. 1, pp. 315-327, Jan. 2017.

[3] Q. Li, J. Guo, Q. Tang, Y. Guo and J. Qian, "Two-Stage Depth Video Recovery with Spatiotemporal Coherence," 2020 IEEE International Conference on Multimedia and Expo (ICME), 2020.

[4] C. Liu, J. Yuen and A. Torralba, "SIFT Flow: Dense Correspondence across Scenes and Its Applications," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 5, pp. 978-994, May 2011.

[5] Y. Li et al., "Video Vectorization via Bipartite Diffusion Curves Propagation and Optimization," in IEEE Transactions on Visualization and Computer Graphics, Early Access, 2021.

Ye Liu, received the M.S. and Ph.D. degrees in electronic science and engineering from Southeast University, Nanjing, Nanjing, China, in 2013 and 2018, respectively. He was a Visiting Scholar with Montana State University, Bozeman, MT, USA from October 2014 to October 2015. He was a visiting Ph.D. Student from February 2017 to January 2018 with the Networked Embedded Systems Group, RISE Swedish Institute of Computer Science. He is currently a Researcher with Nanjing Agricultural University, Nanjing, China. He has authored or co-authored papers in several prestigious journals and conferences, such as the IEEE IEM, IEEE ComMag, IEEE Network, IEEE IoTJ, IEEE TII, ACM TECS, INFOCOM, ICNP, and EWSN. His current research interests include wireless sensor networks, energy harvesting systems, and smart agriculture. Dr. Liu was awarded the 1st place of the EWSN Dependability Competition in 2019, and the Macao Young Scholar in 2021.

An UAV-Enabled Federated Learning Joint Optimization Scheme

A short review for "Privacy-Preserving Federated Learning for UAV-Enabled Networks: Learning-Based Joint Scheduling and Resource Management"

Edited by Qichao Xu

H. Yang, J. Zhao, Z. Xiong, K. -Y. Lam, S. Sun and L. Xiao, "Privacy-Preserving Federated Learning for UAV-Enabled Networks: Learning-Based Joint Scheduling and Resource Management" in IEEE Journal on Selected Areas in Communications. June 2021.

With the increasing maturity of fifth-generation communication technology (5G) and UAV technology, UAVs have been widely used in various fields. UAVs can not only be used to monitor the road traffic environment and investigate hidden dangers, but also become air base stations in remote areas and disaster-stricken areas, providing communication services for the ground [1]-[3]. UAVs with strong computing capabilities can also become air servers to undertake computing tasks. Not only that, with the increasing maturity of machine learning (ML) and federated learning (FL) theories, the superior aerial perspective of UAVs enables them to easily complete tasks such as wireless communication, data collection, and large-scale coverage, which has become increasingly mature. ML and FL provide another paradigm for problem-solving. Deploying UAVs as air base stations can more conveniently complete the tasks of ground equipment data collection and data processing, that is, ML model training.

However, this program still has many problems. The current privacy protection issue is extremely important to users, which means that UAVs will face many difficulties in the process of collecting raw data. Firstly, users are not necessarily willing to upload private data, and data leakage and tampering may also occur during transmission. In addition, due to the limited computing power of UAVs, the speed and accuracy of training models also face challenges. Therefore, it is necessary to propose an algorithm to solve the above-mentioned problems in the process of privacy protection and resource allocation.

This paper proposes an enhanced federated learning algorithm based on UAV network and joint learning to solve privacy protection, scheduling, and resource allocation problems. In order to protect privacy to the greatest extent, consider that users and devices train the ML model locally, and then upload the locally trained local model parameters to the UAV. The device does not need to send any raw data to the UAV during the whole process. Among them, after the UAV server collects the local model parameters, it performs global aggregation, and then broadcasts the aggregated parameters to the relevant equipment for the next round of local model training until the FL model meets the requirements. In order to carry out reasonable scheduling and resource allocation and improve the training speed and accuracy of the FL model, a multi-agent reinforcement learning (RL) problem model combining equipment selection, UAV deployment and resource allocation is proposed. In this model, each UAV can make decisions based on its own instantaneous observation intelligence. Furthermore, in order to solve the joint problem, a multi-agent resource management algorithm based on asynchronous advantage actor-critic (A3C) is proposed. By implementing this algorithm, it is possible to further minimize the loss of training time and learning accuracy due to the large-scale dynamic environment.

Therefore, in this article, the author's main contribution is to propose an asynchronous federated learning framework to achieve privacy protection. In order to solve the multi-agent reinforcement learning problem of joint equipment selection, UAV deployment and resource allocation, A3C-based Multi-agent resource management algorithm to improve training speed and learning accuracy.

In the construction of the multi-agent reinforcement learning problem model of joint equipment selection, UAV layout and resource allocation, due to the difference in communication speed between the device and the UAV server, the device with good communication connection has a lower uploading of the local FL model parameters. When the UAV server broadcasts the global FL model parameters, the communication delay with each device is different. Therefore, selecting appropriate equipment instead of all equipment for FL model training can reduce communication delay and increase the training speed. In addition, the location of the UAV will also affect the overall communication delay. The devices for training the local FL model can be mobile phones, computers, vehicles, etc., so there are differences in computing power between devices. Assigning different computing tasks to different devices can also significantly improve the training speed of the FL model. Therefore, a multi-agent reinforcement learning problem model combining equipment selection, UAV deployment and resource allocation is proposed.

In order to solve the joint problem, RL is introduced to realize the FL aggregation process based on self-scheduling in the network supporting multiple UAVs. However, in this joint problem, changes in the communication and training environment due to differences in device mobility and computing capabilities increase the global learning delay and reduce the cycle efficiency [4]. At the same time, some actions of joint UAV deployment and resource allocation have continuous space. Therefore, an asynchronous advantage actor-critic-based asynchronous federated learning algorithm (A3C-AFL) is proposed to solve the above problems. The UAV server does not need to wait for all local FL model parameters to be uploaded before performing global FL aggregation. The proposed A3C-ALF framework includes three stages: equipment selection, UAV deployment, resource allocation; local training; global aggregation. In order to solve the combinatorial optimization problem, it is first transformed into a Markov Decision Process (MDP), and the multi-UAV network is regarded as multiple agents. Then use RL to solve the joint problem. Different from the traditional actor-critic (AC) algorithm, the proposed A3C enables asynchronous multiple agents (i.e., UAV servers) to interact with their environment in parallel and implement different exploration strategies. After using A3C to execute the joint problem, start the AFL program. Each UAV server aggregates the local model parameters trained by the selected equipment and broadcasts them after updating the complete local model parameters.

Extensive simulations have evaluated the performance of the proposed A3C-ALF scheme. The simulation results show that the proposed scheme perform better than AFL without device selection, Gradient-AFL, and A3C-synchronous federated learning(A3C-SFL) [5].

In summary, this paper proposed an AFL framework to provide asynchronous distributed computing without transmitting raw sensitive data to UAV servers. Moreover, an A3C-based joint device selection, UAVs placement, and resource management algorithm is proposed to enhance the learning convergence speed and accuracy.

References:

[1]. X. Lu, L. Xiao, C. Dai, and H. Dai, "UA V -aided cellular communications with deep reinforcement learning against jamming," *IEEE Wireless Commun*, vol. 27, no. 4, pp. 48-53, Aug. 2020.

[2]. F. Qi, X. Zhu, G. Mang, M. Kadoch, and W. Li, "UA V network and IoT in the sky for future smart cities," *IEEE Network*, vol. 33, no. 2, pp. 96-101, Mar. 2019.

[3]. Y. Zeng, Q. Wu, and R. Zhang, "Accessing from the sky: A tutorial on UAV communications for 5G and beyond," *Proc. of IEEE*, vol. 107, no.12, pp. 2327-2375, Dec. 2019

[4]. V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley,D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," in *Proc. Int. Conf. Machine Learn*., 2016, pp.1928C1937.

[5]. Y. Chen, X. Sun, and Y. Jin, "Communication-efficient federated deep learning With layerwise asynchronous model update and temporally weighted aggregation," *IEEE Trans. Neural Netw. Learn. Syst*., vol. 31, no. 10, pp. 4229-4238, Oct. 2020.

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.

On the Removal of 3D Point Cloud Compression Artifacts

A short review for "Video-based Point Cloud Compression Artifact Removal"

Edited by Roberto G. de A. Azevedo

Anique Akhtar, Wen Gao, Li Li, Zhu Li, Wei Jia, and Shan Liu, "Video-based Point Cloud Compression Artifact Removal," IEEE Transactions on Multimedia, Early access, 2021. doi: 10.1109/TMM.2021.3090148

In recent years, 3D point clouds have emerged as one of the most important alternatives to represent static and dynamic 3D scenes/objects for immersive media applications (e.g., virtual/augmented reality streaming and teleconference). A 3D point cloud (PC) is composed of a set of discrete 3D points <x, y, z> (positions) associated with additional information (attributes) for such positions, e.g., color, normal, and reflectance. As one of the main advantages, PCs have real-time processing potential and provide a simpler alternative when compared to other 3D scene representations, e.g., meshes. As a drawback, point clouds commonly carry a huge amount of data; thus, requiring efficient compression techniques.

Currently, the two main approaches for 3D point clouds, developed by the MPEG standardization group, are Geometry-based (G-PCC) and Video-based (V-PCC) Point cloud compression [1]. G-PCC uses an octree-based approach to encode voxelated PCs, while V-PCC projects PCs into 2D surfaces and then employs state-of-the-art video encoding technologies on such projected content (e.g., HEVC). (See Fig. 1) Although many exciting recent works are being proposed regarding G-PCC-related approaches (including deep learning-based ones), V-PCC is still widely used as the current best rate-distortion trade-off for many types of 3D PC content.

In their paper [2], A. Akthar, W. Gao, L. Li, Z. Li, W. Jia, and S. Liu provide a novel approach for the artifacts' removal in V-PCC. Although a well-studied topic in the 2D-only image and video compression, overall, compression artifact removal for point cloud compression is still an underexplored research area. Thus, the paper provides one of the first steps in such a direction. The authors' proposal is based on: i) a projection-aware 3D sparse convolutional neural network and ii) a patch-based scalable sampling and aggregation scheme that finds the correspondence between the reconstructed and the original geometries.

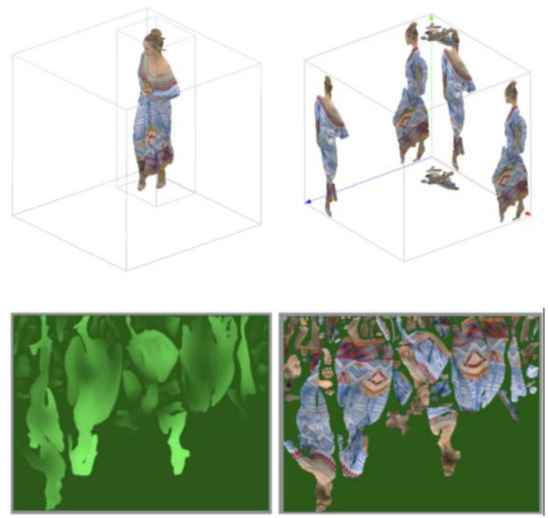

Fig.1. Example of V-PCC 3D-to-2D projection. Projection on the "bounded box" plane (top) [3] and V-PCC patch packing (bottom) [4].

The paper focuses on the geometry distortions of the V-PCC reconstructed point cloud. By observing that, in V-PCC, each point is projected into only one plane (see Fig. 1), and thus geometric distortions only occur in such a projection direction, the authors can limit the degree of freedom of the learned quantization noise in such a direction. The proposed approach, first, sample a large point cloud into smaller neighborhood patches and then, by using a SparseConv-based 3D U-Net, learn the per-point quantization noise. Chamfer distance is used as the loss function. Then, after having the artifacts removed per patch, the patches are aggregated back to the entire point cloud.

The experimental results show quality improvements when comparing the proposed approach to the noisy compressed PC, especially in the edges and the surface of the point cloud.

The numerical results also show that more improvement is achieved in the lower bitrate settings. Moreover, the authors perform some ablation studies providing insights on the problem and into their solution. In particular, they compare their sampling methods with an octree-based cube division method and study the sampling of the number of overlapping patches.

In sum, Akthar et al. propose an interesting approach to a still underexplored research topic. The proposal is backward-compatible with the current V-PCC standard since it is an out-of-the-loop artifact removal. Similarly, it will also be interesting to see how similar approaches can be applied in the context of G-PCC or into more recently deep learning-based PCC frameworks. Overall, PCC and auxiliary pre-/post-processing and streaming techniques are a very exciting research topics nowadays, and it is expected that we will see many new important breakthroughs in the coming years.

References:

[1] H. Liu, H. Yuan, Q. Liu, J. Hou and J. Liu, "A Comprehensive Study and Comparison of Core Technologies for MPEG 3-D Point Cloud Compression," in IEEE Transactions on Broadcasting, vol. 66, no. 3, pp. 701-717, Sept. 2020, doi: 10.1109/TBC.2019.2957652.

[2] A. Akhtar, W. Gao, L. Li, Z. Li, W. Jia, and S. Liu, "Video-based Point Cloud Compression Artifact Removal," in IEEE Transactions on Multimedia, Early access, 2021. doi: 10.1109/TMM.2021.3090148.

[3] "Coding of moving pictures and audio," document ISO/IEC JTC1/SC29/WG11 w18892

[4] S. Schwarz, M. Preda, V. Baroncini, M. Budagavi, P. Cesar, P. A. Chou, R. A. Cohen, M. Krivokuća, S. Lasserre, Z. Li et al., "Emerging MPEG standards for point cloud compression," IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol. 9, no. 1, pp. 133–148, 2018.

Roberto G. A. Azevedo, Ph.D., is an associate researcher at ETH Zurich, working on visual perception, quality of experience, compression, and streaming of immersive media technologies. Previously, he was a post doc at EPFL, Lausanne. He holds a Ph.D. (2015) and M.Sc. (2010) degrees in Informatics from PUC-Rio, and the degree of Computer Scientist from the Federal University of Maranhão (UFMA) (2008). Roberto also actively contributed to the specifications and reference implementation for the standards of the Brazilian Digital TV System and ITU-T Recommendations for IPTV middleware, currently adopted by more than 19 countries in Latin America, Africa, and Asia. Roberto's main research interests are focused on immersive and interactive media, with roots on the intersection of the broad areas of: multimedia systems, human-computer interaction, and computer graphics.

MMTC Communications – Review Editorial Board

Multimedia Communications Technical Committee Officers

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.