Guest Editorial Introduction to the Special Issue on Recent Advances in Point Cloud Processing and Compression

I. Introduction

A POINT cloud is a set of 3D points that can be used to represent a 3D surface. Each point has a spatial position (x, y, z) and a vector of attributes, such as colors, material reflection, or normal. As point clouds are capable of reconstructing 3D objects or scenes, they have the potential to be widely used in various applications such as auto-driving and 6-degree virtual reality. However, the following properties of point cloud make the point cloud compression and processing become rather challenging. 1) Unstructured. The point cloud is a series of non-uniform sampled points. On the one hand, it makes the correlations among various points difficult to be utilized for compression. On the other hand, the convolutional neural network that is widely used in image/video processing cannot be applied to the point cloud processing. 2) Unordered. Unlike images and videos, the point cloud is a set of points without a specific order. Therefore, both the point cloud processing and compression algorithms need to be invariant to any permutations of the input point clouds.

Recent years have witnessed considerable research efforts in point cloud processing and compression. Hence, the state-of-the-art in point cloud processing and compression is getting redefined. Yet, many research challenges still remain to be addressed. 1) Efficient framework to compress the point cloud especially the sparse point cloud is still an open problem. Both the traditional compression framework from MPEG and the end-to-end point cloud compression schemes are competing to be the state-of-the-art. 2) Simple yet efficient quality metrics to balance the trade-off between the geometry and the attribute that can reflect human perception are needed. Point clouds have two kinds of information: geometry and attribute. They have their own quality metrics individually. However, how do the qualities of geometry and attributes influence the overall qualities of point clouds are still unknown. A good quality metric is also important to guide the processing and compression algorithms. 3) Efficient deep-learning network structures for both point cloud low-level and high-level vision tasks are anticipated. The development of point cloud object detection and segmentation is still at an initial stage. More sophisticated and efficient algorithms are to be developed.

In this Special Issue, a total number of 14 articles that present state-of-the-art results are accepted. They cover mainly the following four topics: point cloud compression, point cloud quality metric, and point cloud low-level and high-level vision tasks. In the following, we will introduce these articles briefly with one paragraph for each article.

II. Point Cloud Compression

Zhang and Gao [A1] propose an adaptive geometry partition and coding scheme to improve the point cloud compression efficiency. They first introduce quad-tree (QT) and binary tree (BT) partitions as alternative geometry partition modes in addition to the commonly used octree (OT) partition. In addition, to address the complexity issue brought from searching for the optimal combination of OT, QT, and BT, they introduce two hyper-parameters to specify conditions when QT and BT partitions can be applied. Once they are determined, the partition mode can be derived according to the geometry shape of the current coding node. Finally, an adaptive parameter selection scheme is presented to optimize the coding gain adaptively. The proposed adaptive geometry partition scheme has been adopted in the state-of-the-art MPEG geometry-based PCC (G-PCC) standard because of its high coding efficiency.

Zhu *et al.* [A2] introduce region-wise processing to point cloud geometry compression leveraging the region similarity to exploit inter-region redundancies. First, the point cloud geometry is first segmented into numerous local regions each of which comprises a portion of point cloud surface, and can be represented by a surface vector that describes the geometry shape numerically in a projected principal space. Second, these regions are grouped into several discriminative clusters, assuring that inter-cluster similarity is minimized and intra-cluster similarity is maximized simultaneously, where the similarity is calculated using the regional surface vectors. Third, in each cluster, a reference region having the largest similarity score to the others is selected and compressed using the lossless mode of G-PCC. The other regions are predicted using the reference region with alignment transform. The effectiveness of the proposed method is demonstrated experimentally using a variety of common test sequences.

Zhao *et al.* [A3] present a novel image synthesis method for effective point cloud attribute compression. First, it segments a point cloud into a collection of fine-grained patches by

Digital Object Identifier 10.1109/TCSVT.2021.3129071

performing geometric structure analysis. Second, it transforms the patches from 3D to 2D using a low-dimensional embedding algorithm and then convert them into patch attribute images with patch rasterization and rectification. Finally, it assembles all the attribute images of patches by formulating it as a bin nesting problem and harvest an attribute image of the whole point cloud for image/video-based compression. The effectiveness of the proposed method in point cloud attribute compression and its superiority over state-of-the-art codecs are experimentally demonstrated.

Song *et al.* [A4] propose a novel layer-wise geometry aggregation (LGA) framework for LiDAR point cloud lossless geometry compression. Based on content properties, it adaptively partitions point clouds into ground, object, and noise three layers. The ground layer is fit with a Gaussian Mixture Model, which can represent ground points using much fewer model parameters than adopting the original 3D coordinates. The object layer is tightly packed to reduce the space between objects. For the noise layer, the difference between neighboring points is reduced by reordering using Morton Code, and the reduced residuals help save coding bits. It is demonstrated by experimental results that LGA significantly outperforms competitive methods without prior knowledge. Some other properties of LGA such as robustness, stability, and time complexity are also examined.

Nguyen *et al.* [A5] propose a lossless point cloud geometry compression method that uses neural networks to estimate the probability distribution of voxel occupancy. To take into account the point cloud sparsity, the proposed method first adaptively partitions a point cloud into multiple voxel block sizes. The partitions are signaled using an octree. Then, a deep auto-regressive generative model is employed to estimate the occupancy probability of each voxel given the previously encoded ones. A context-based arithmetic coder is used to code a block efficiently, wherein the context may have variable size and can expand beyond the current block to learn more accurate probabilities. The authors also consider using data augmentation techniques to increase the generalization capability of the learned probability models, in particular in the presence of noise and lower-density point clouds. The performance of the proposed method is evaluated on a variety of point clouds with diverse characteristics. The experimental results demonstrate that the proposed method can significantly reduce the rate for lossless coding compared to the state-of-the-art MPEG codec.

III. Point Cloud Quality Metric

Wu *et al.* [A6] consider the problem of subjective and objective quality assessment for point clouds. First, they introduce a publicly released subjective dataset, using a head-mounted display with six degrees of freedom. It includes 340 distorted point clouds using MPEG point cloud compression. The impact of compression on geometry and texture attributes is then investigated. Next, the authors propose two projection-based objective quality assessment techniques. The first one is based on a weighted view projection model, whereas the second one is based on a patch projection model. The dataset and the results of the article are proved to be useful for point cloud coding and transmission, as well as in the context of virtual reality applications.

Liu *et al.* [A7] propose a novel deep learning-based no reference point cloud quality assessment method, namely, PQA-Net. Specifically, the PQA-Net consists of a multi-view-based joint feature extraction and fusion (MVFEF) module, a distortion type identification (DTI) module, and a quality vector prediction (QVP) module. The DTI and QVP modules share the feature generated from the MVFEF module. Using the distortion type labels, the DTI and MVFEF modules are first pre-trained to initialize the network parameters, based on which the whole network is then jointly trained to finally evaluate the point cloud quality. The experimental results demonstrate the effectiveness of the proposed method.

IV. Point Cloud Low-Level Vision

Ding *et al.* [A8] introduce a new learning-based technique for upsampling sparse point clouds. More specifically, they introduce an efficient neural network framework that encompasses feature extraction, perturbation learning, and coordinate reconstruction. As a key element of the proposed approach, the authors propose to learn a 2D perturbation through multilayer perceptrons (MLPs) to estimate the coordinate shift from the input point to the upsampled one. This 2D perturbation is then combined with the extracted features to fine-tune the coordinate shift. A large-scale point cloud dataset has been produced for training purposes, which includes 36000 pairs. The experiments show that the proposed scheme outperforms state-of-the-art techniques in both qualitative and quantitative evaluations. Moreover, the network size remains small when compared to competing methods.

Zhang *et al.* [A9] introduce a novel progressive upsampling framework for point clouds. More specifically, the authors introduce an Up-UNet feature expansion module. It learns the local and global point features via down-feature and up-feature operators, in order to address the non-uniform distribution during the upsampling process and to remove outliers. In addition, the authors present a hybrid loss function considering both the multi-scale reconstruction loss and the rendering loss, generating a denser point cloud while preserving its structures. Qualitative and quantitative experimental results show that the proposed technique achieves state-of-the-art performances.

Wang *et al.* [A10] propose a new sequential point cloud upsampling method called SPU, which aims to upsample sparse, non-uniform, and orderless point cloud sequences by effectively exploiting rich and complementary temporal dependency from multiple inputs. Specifically, these inputs include a set of multi-scale short-term features from the 3D points in three consecutive frames (i.e., the previous, current, and subsequent frames) and a long-term latent representation accumulated throughout the point cloud sequence. Considering that these temporal clues are not well aligned in the coordinate space, they propose a new temporal alignment module (TAM) to transform each individual feature into the feature space of the current frame, which is analogous to the cross-attention mechanism. They also propose a new gating mechanism to learn the optimal weights for these transformed features, based on which the transformed features can be effectively aggregated as the final fused feature. The fused feature can be readily embedded into the existing single-frame-based point cloud upsampling methods (e.g., PU-Net and PU-GAN) to generate the dense point cloud for the current frame. Comprehensive experiments demonstrate the effectiveness of the proposed method for upsampling point cloud sequences.

Zhao *et al.* [A11] propose a learning-based mesh normal denoising scheme, called NormalNet, which employs deep neural networks to find the correlation between the volumetric representation and denoised face normal. The proposed NormalNet follows the iterative framework of filtering-based mesh denoising and introduces normalization into mesh denoising. Some other main contributions of this work include a classification-based training-faces selection strategy for balancing the training set and a mismatched-faces rejection strategy for removing the mismatched faces between noisy mesh and ground truth. Compared to some state-of-the-art works, NormalNet can effectively remove noise while preserving the original features and avoiding pseudo-features.

V. Point Cloud High-Level Vision

Liu and Xu [A12] present a two-stream baseline method referred to as GeometryMotion-Net for 3D action recognition. In the proposed mechanism, the authors first represent each point cloud video as a limited number of randomly sampled frames with each frame consisting of a sparse set of points and then use a new two-stream (i.e., the geometry stream and the motion stream) framework for effective 3D action recognition. Specifically, for each point in the current frame, a set of 3D offset features relative to the neighboring points in the reference frame are first produced and then local neighborhood information of this point in the offset feature space is exploited. Based on the newly generated virtual overall geometry point cloud and multiple virtual forward/backward motion point clouds, any existing point cloud analysis methods (e.g., PointNet) can be used for extracting discriminant geometry and bidirectional motion features in the geometry and motion streams, respectively, which are further aggregated to make the proposed two-stream network trainable in an end-to-end fashion. Comprehensive experiments on both large-scale and small-scale datasets demonstrated the effectiveness and efficiency of the proposed method for 3D action recognition.

Deng *et al.* [A13] deem point clouds as the hollow-3D data and present a new object detection architecture hallucinated hollow-3D R-CNN (H23D R-CNN). First, multi-view features are obtained by sequentially projecting point clouds into the perspective and bird's-eye views. Second, the 3D representation is hallucinated by a novel bilaterally guided multi-view fusion block. Finally, 3D objects can be detected via a box refinement module with a novel Hierarchical Voxel RoI Pooling operation. In this way, the complementary information can be effectively utilized in the proposed framework, and finally, the extensive experiments show the superiority of the proposed method.

Zhao *et al.* [A14] propose a simple but effective 3D object detection method called Transformer3D-Det (T3D), in which they additionally introduce a transformer-based vote refinement module to refine the voting results of VoteNet and can thus significantly improve the 3D object detection performance. Specifically, their T3D framework consists of three modules: a vote generation module, a vote refinement module, and a bounding box generation module. Given an input point cloud, they first utilize the vote generation module to generate multiple coarse vote clusters. Then, the clustered coarse votes will be refined by using a transformer-based vote refinement module to produce more accurate and meaningful votes. Finally, the bounding box generation module takes the refined vote clusters as the input and generates the final detection results for the input point cloud. To suppress the effect of the inaccurate votes, they also propose a new non-vote loss function to train the T3D. As a result, the T3D framework can achieve better 3D object detection performance. Comprehensive experiments demonstrate the effectiveness of the T3D framework for 3D object detection.

VI. Conclusion

The accepted articles in this Special Issue provide an overview of the state-of-the-art as well as new results in the field of point cloud compression and processing. The breadth of the topics reported in this issue demonstrates the interest of the community in this active research area. It is our hope that this special issue will encourage further researches in this area. As part of the new digital initiative for IEEE Transactions on Circuits and Systems for Video (TCSVT), accepted articles can also share a short 10-min presentation of the work in the TCSVT YouTube channel (https://www. youtube. com/channel/UC46cVfjkp7XLdDXsgGuG_Lg) to help disseminating the main ideas. You are encouraged to check that out.

Zhu Li, *Lead Guest Editor*
Department of Computer Science and Electrical Engineering
University of Missouri–Kansas City
Kansas City, MO 64111 USA
e-mail: lizhu@ustc.edu.cn

Shan Liu, *Guest Editor*
Tencent Media Laboratory
Palo Alto, CA 94306 USA
e-mail: shanl@tencent.com

Frederic Dufaux, *Guest Editor*
Université Paris-Saclay, CNRS,
CentraleSupélec, L2S
91190 Gif-sur-Yvette, France
e-mail: frederic.dufaux@l2s.centralesupelec.fr

Li Li, *Guest Editor*
Department of Electronic Engineering and Information Science
University of Science and Technology of China
Hefei 230026, China
e-mail: lil1@ustc.edu.cn

Ge Li, *Guest Editor*
School of Electronic and Computer Engineering
Peking University
Shenzhen 518055, China
e-mail: geli@pku.edu.cn

C.-C. JAY KUO, *Guest Editor*
Ming Hsieh Department of Electrical and
Computer Engineering
University of South California
Los Angeles, CA 90007 USA
e-mail: cckuo@sipi.usc.edu

APPENDIX: RELATED ARTICLES

[A1] X. Zhang and W. Gao, "Adaptive geometry partition for point cloud compression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 12, pp. 4561–4574, Dec. 2021.
[A2] W. Zhu, Y. Xu, D. Ding, Z. Ma, and M. Nilsson, "Lossy point cloud geometry compression via region-wise processing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 12, pp. 4575–4589, Dec. 2021.
[A3] B. Zhao, W. Lin, and C. Lv, "Fine-grained patch segmentation and rasterization for 3-D point cloud attribute compression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 12, pp. 4590–4602, Dec. 2021.
[A4] F. Song, Y. Shao, W. Gao, H. Wang, and T. Li, "Layer-wise geometry aggregation framework for lossless LiDAR point cloud compression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 12, pp. 4603–4616, Dec. 2021.
[A5] D. T. Nguyen, M. Quach, G. Valenzise, and P. Duhamel, "Lossless coding of point cloud geometry using a deep generative model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 12, pp. 4617–4629, Dec. 2021.
[A6] X. Wu, Y. Zhang, C. Fan, J. Hou, and S. Kwong, "Subjective quality database and objective study of compressed point clouds with 6DoF head-mounted display," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 12, pp. 4630–4644, Dec. 2021.
[A7] Q. Liu *et al.*, "PQA-Net: Deep no reference point cloud quality assessment via multi-view projection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 12, pp. 4645–4660, Dec. 2021.
[A8] D. Ding, C. Qiu, F. Liu, and Z. Pan, "Point cloud upsampling via perturbation learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 12, pp. 4661–4672, Dec. 2021.
[A9] P. Zhang, X. Wang, L. Ma, S. Wang, S. Kwong, and J. Jiang, "Progressive point cloud upsampling via differentiable rendering," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 12, pp. 4673–4685, Dec. 2021.
[A10] K. Wang, L. Sheng, S. Gu, and D. Xu, "Sequential point cloud upsampling by exploiting multi-scale temporal dependency," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 12, pp. 4686–4696, Dec. 2021.
[A11] W. Zhao, X. Liu, Y. Zhao, X. Fan, and D. Zhao, "NormalNet: Learning-based mesh normal denoising via local partition normalization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 12, pp. 4697–4710, Dec. 2021.
[A12] J. Liu and D. Xu, "GeometryMotion-Net: A strong two-stream baseline for 3D action recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 12, pp. 4711–4721, Dec. 2021.
[A13] J. Deng, W. Zhou, Y. Zhang, and H. Li, "From multi-view to hollow-3D: Hallucinated hollow-3D R-CNN for 3D object detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 12, pp. 4722–4734, Dec. 2021.
[A14] L. Zhao, J. Guo, D. Xu, and L. Sheng, "Transformer3D-Det: Improving 3D object detection by vote refinement," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 12, pp. 4735–4746, Dec. 2021.

Zhu Li (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering from Northwestern University, Evanston, in 2004. He was the AFOSR SFFP Summer Visiting Faculty at the UAV Research Center, U.S. Air Force Academy (USAFA), in 2016, 2017, 2018, and 2020. He was a Senior Staff Researcher/a Senior Manager with Samsung Research America's Multimedia Standards Research Laboratory, Richardson, TX, USA, from 2012 to 2015; a Senior Staff Researcher/a Media Analytics Lead with FutureWei (Huawei) Technology's Media Laboratory, Bridgewater, NJ, USA, from 2010 to 2012; an Assistant Professor with the Department of Computing, The Hong Kong Polytechnic University, from 2008 to 2010; and a Principal Staff Research Engineer with the Multimedia Research Laboratory (MRL), Motorola Labs, from 2000 to 2008. He is currently an Associate Professor with the Department of Computer Science and Electrical Engineering (CSEE), University of Missouri-Kansas City, and the Director with the NSF I/UCRC Center for Big Learning (CBL), UMKC. His research interests include point cloud and light field compression, graph signal processing and deep learning in the next-gen visual compression, and image processing and understanding. He has 47 issued or pending patents, more than 160 publications in book chapters, journals, and conferences in these areas. He serves as the Steering Committee Member for IEEE ICME (2015–2018). He is an Elected Member of the IEEE Multimedia Signal Processing (MMSP), IEEE Image, Video, and Multidimensional Signal Processing (IVMSP), and IEEE Visual Signal Processing and Communication (VSPC) Tech Committees. He received the Best Paper Award at IEEE Int'l Conf on Multimedia and Expo (ICME), Toronto, in 2006; and the Best Paper Award (DoCoMo Labs Innovative Paper) at IEEE Int'l Conf on Image Processing (ICIP), San Antonio, in 2007. He is the Program Co-Chair of IEEE Int'l Conf on Multimedia and Expo (ICME) 2019 and co-chaired the IEEE Visual Communication and Image Processing (VCIP) 2017. He is the Associate Editor-in-Chief of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY and an Associated Editor of IEEE TRANSACTIONS ON IMAGE PROCESSING (2020), IEEE TRANSACTIONS ON MULTIMEDIA (2015–2018), and IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (2016–2019).

Shan Liu (Senior Member, IEEE) received the B.Eng. degree in electronic engineering from Tsinghua University and the M.S. and Ph.D. degrees in electrical engineering from the University of Southern California. She is currently a Distinguished Scientist at Tencent and the General Manager at Tencent Media Laboratory. She was formerly the Director with the Media Technology Division, MediaTek, USA. She was also formerly with MERL and Sony. She holds more than 300 granted U.S. patents. She has been an active contributor to international standards for more than a decade and has numerous technical proposals adopted into various standards, such as VVC, HEVC, OMAF, DASH, MMT, and PCC. She received the Best AE Award from IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY in 2019 and 2020. She has been the Vice Chair of IEEE Data Compression Standards Committee since 2019. She was an Editor of H.265/HEVC SCC and H.266/VVC standards. She was named the APSIPA Distinguished Industry Leader in 2018.

Frederic Dufaux (Fellow, IEEE) received the M.Sc. degree in physics and the Ph.D. degree in electrical engineering from the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, in 1990 and 1994, respectively. He is currently the CNRS Research Director with the CNRS, CentraleSupélec, Laboratoire des Signaux et Systèmes (L2S, UMR 8506), Université Paris-Saclay, where he is the Head of the Telecom and Networking Hub. He is the author or coauthor of three books and more than 200 research publications. He served as the Chair for the IEEE SPS Multimedia Signal Processing (MMSP) Technical Committee in 2018 and 2019. He was the Vice General Chair of ICIP 2014, the General Chair for MMSP 2018, and the Technical Program Co-Chair for ICIP 2019 and ICIP 2021. He is the Chair of the Steering Committee of ICME in 2022 and 2023. He is also a Founding Member and the Chair of the EURASIP Technical Area Committee on Visual Information Processing.

Li Li (Member, IEEE) received the B.S. and Ph.D. degrees in electronic engineering from the University of Science and Technology of China (USTC), Hefei, Anhui, China, in 2011 and 2016, respectively. He was a Visiting Assistant Professor with the University of Missouri–Kansas City from 2016 to 2020. He is currently an Associate Professor with the University of Science and Technology of China. His research interests include image/video coding and processing. He has more than 70 publications in book chapters, journals, and conferences in these areas. He received the Top 10% Paper Award at the 2016 IEEE Visual Communications and Image Processing (VCIP) Conference and the 2019 IEEE International Conference on Image Processing (ICIP) Conference. He was the winner of the 2016 IEEE International Conference on Multimedia and Expo Grand Challenge on Light Field Image Compression.

Ge Li received the Ph.D. degree in electrical and computer engineering from Auburn University in 1999. He did his post-doctoral research work with the University of California at Davis in 2002. He is currently a Professor at the School of Electronic and Computer Engineering, Peking University Shenzhen Graduate School, China. He has more than 20 years of experience in research and product design in the industry. He has published more than 70 technical papers and 70 patents (issued or pending). His research interests include image and video process and analysis, 3D video coding, and quality assessment.

C.-C. Jay Kuo (Fellow, IEEE) received the B.S. degree from the National Taiwan University, Taipei, Taiwan, in 1980, and the M.S. and Ph.D. degrees from the Massachusetts Institute of Technology, Cambridge, in 1985 and 1987, respectively, all in electrical engineering. He is currently the holder of William M. Hogue Professorship, a Distinguished Professor of electrical and computer engineering and computer science, and the Director of the USC Multimedia Communication Laboratory (MCL), University of Southern California. He is a fellow of the National Academy of Inventors (NAI), the American Association for the Advancement of Science (AAAS), and the International Society for Optical Engineers (SPIE).