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

MMTC Communications – Review

Vol. 13, No. 1, February 2022

IEEE COMMUNICATIONS SOCIETY

TABLE OF CONTENTS

Message from the Review Board Directors

Welcome to the February 2022 issue of the IEEE ComSoc MMTC Communications – Review.

This issue comprises three reviews that cover multiple facets of multimedia communication research including wireless distributed learning networks, point clouds, and video coding for machiens. These reviews are briefly introduced below.

The first paper, published in IEEE Journal on Selected Areas in Communications and edited by Dr. Cong Shen, studies asynchronous federated learning (FL) in a wireless distributed learning network. The authors propose a metric, called an effectivity score to represent the amount of learning from asynchronous FL. An Asynchronous Learning-aware transmission Scheduling problem is then formulated to maximize the effectivity score.

The second paper, edited by Dr. Mengbai Xiao, was published in IEEE Transactions on Multimedia. This paper proposes a point cloud clustering algorithm. It clusters with i) a constant number of points, ii) compact clusters, i.e., with low dispersion, iii) non-overlapping clusters, i.e., not intersecting each other, iv) the ability to scale with the number of points, and v) low complexity.

The third paper, edited by Dr. Debashis Sen, was published in IEEE Transactions on Multimedia. This paper proposes a portrait relighting method based on deep reinforcement learning (called PR-RL). The PR-RL model could conduct portrait relighting by sequentially predicting local light editing strokes, and use strokes to conduct dodge and burn operations on the image lightness, simulating image editing by artists using brush strokes.

The fourth paper, edited by Dr. Tiesong Zhao, was published in IEEE International Conference on Multimedia & Expo. The paper proposes leveraging the strength of predictive and generative models to support advanced compression techniques for both machine and human vision tasks simultaneously.

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
George Mason University, USA
Email: zyan4@gmu.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

Asynchronous Federated Learning in Wireless Distributed Learning Networks

A short review for "Adaptive Transmission Scheduling in Wireless Networks for Asynchronous Federated Learning"

Edited by Cong Shen

Hyun-Suk Lee and Jang-Won Lee, "Adaptive Transmission Scheduling in Wireless Networks for Asynchronous Federated Learning," IEEE Journal on Selected Areas in Communications, Vol. 39, No. 12, pp. 3673-3687, Dec. 2021.

Nowadays, a massive amount of data is generated from devices, such as mobile phones and wearable devices, which can be used for a wide range of machine learning (ML) applications from healthcare to autonomous driving. As the computational and storage capabilities of such distributed devices keep growing, federated learning (FL) has been widely studied as a potentially viable solution for distributed learning [1], [2]. FL learns a central model by using the locally trained models of distributed devices under the coordination of a central server even without sharing the local training data at the devices with the central server.

Recently, asynchronous FL has been widely studied to utilize distributed resources more effectively [3-6]. To this end, in asynchronous FL, each device continually trains its local model by using the arriving local data regardless of its transmission scheduling. Then, the device transmits its local model to the central server if it is scheduled; otherwise, it stores its local model for later use. This prevents the waste of local computation resources and too much pileup of the local data at devices. However, at the same time, it causes a time lag between the stored local models and the current central model. Such stored locally trained models with the time lag may cause an adverse effect to the convergence of the central model [5].

Several works on asynchronous FL have addressed such harmful effects [3-6]. However, they mainly focused on addressing the incurred stragglers and did not take into account more fundamental issues on the occurrence of the time lag due to network circumstances such as time-varying channels and scarce radio resources. Hence, for effectively utilizing asynchronous FL in a wireless distributed learning network (WDLN), it is necessary to study an asynchronous FL procedure that carefully addresses those key challenges occurring when implementing asynchronous FL in wireless networks.

To address such key challenges, this paper proposes an asynchronous FL procedure. In the procedure, transmission scheduling is determined over multiple rounds while carefully considering the characteristics of asynchronous FL, time-varying channels, and stochastic data arrivals of the edge devices. Also, the convergence of the asynchronous FL procedure is analyzed.

The significant idea of this paper for transmission scheduling is a metric called an effectivity score. It represents the amount of learning from asynchronous FL considering both amounts of local data used for learning and properties of asynchronous FL such as the harmful effects on learning due to the time lag. Hence, to achieve effective learning in the asynchronous FL procedure, an asynchronous learning-aware transmission scheduling (ALS) problem is formulated which maximizes the effectivity score while considering the system uncertainties (i.e., the time-varying channels and stochastic data arrivals).

The authors propose the following three ALS algorithms that determine the transmission scheduling by solving the ALS problem; First, an ALS algorithm with the perfect statistical information about the system uncertainties (ALSA-PI) optimally and efficiently solves the problem using the state information reported from the edge devices in the asynchronous FL procedure and the statistical information. Second, a Bayesian ALS algorithm (BALSA) solves the problem using the state information without requiring any a priori information. Instead, it learns the system uncertainties based on a Bayesian approach. It is proven that BALSA is optimal in terms of the long-term average effectivity score by its regret bound analysis.

Third, a Bayesian ALS algorithm for a partially observable WDLN (BALSA-PO) solves the problem only using partial state information (i.e., channel conditions). It addresses a more restrictive WDLN in practice, where each edge device is allowed to report only its current channel condition to the AP.

Experimental results show that the transmission scheduling strategy based on the effectivity score, which is adopted in the proposed algorithms, is effective for asynchronous FL. In specific, ALSA-PI and BALSAs (i.e., BALSA and BALSA-PO) achieve performance close to an ideal benchmark that has no radio resource constraints and does not fail transmission. Also, they outperform other state-of-the-art scheduling algorithms in terms of training loss, test accuracy, learning speed, and robustness of learning. Those results clearly show the proposed BALSAs effectively schedule the transmissions even without any a priori information by learning the system uncertainties.

In summary, this paper studies asynchronous FL in a WDLN. To represent the amount of learning by asynchronous FL, a novel metric called an effectivity score is proposed. Then, the transmission scheduling problem for asynchronous FL is formulated to maximize the effectivity score while carefully considering system uncertainties. The proposed transmission scheduling algorithms can achieve effective learning by solving the problem even without any a priori information on the uncertainties. Experimental results demonstrate that the adaptive scheduling strategy in the proposed algorithms is effective to asynchronous FL.

References:

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2017.
[2] W. Y. B. Lim, N. C. Luong, D. T. Hoang, Y. Jiao, Y.-C. Liang, Q. Yang,D. Niyato, and C. Miao, "Federated learning in mobile edge networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, no. 3, pp. 2031–2063, 2020.
[3] S. Zheng, Q. Meng, T. Wang, W. Chen, N. Yu, Z.-M. Ma, and T.-Y. Liu, "Asynchronous stochastic gradient descent with delay compensation," in *Proc. International Conference on Machine Learning (ICML),* 2017.
[4] 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.
[5] Y. Chen, Y. Ning, M. Slawski, and H. Rangwala, "Asynchronous online federated learning for edge devices with non-IID data," in *Proc. 2020 IEEE Int. Conf. on Big Data (Big Data)*, 2020.
[6] C. Xie, S. Koyejo, and I. Gupta, "Asynchronous federated optimization," *arXiv preprint* arXiv:1903.03934, 2020.

Cong Shen received his B.S. and M.S. degrees, in 2002 and 2004 respectively, from the Department of Electronic Engineering, Tsinghua University, China. He obtained the Ph.D. degree from the Electrical Engineering Department, UCLA, in 2009. From 2009 to 2014, He worked for Qualcomm Research in San Diego, CA. He is currently an Assistant Professor in the Electrical and Computer Engineering Department at University of Virginia. He is a Senior Member of IEEE and serves as editor for the IEEE Transactions on Green Communications and Networking and IEEE Wireless Communications Letters.

A Practical Method towards Compact Clustering of Point Clouds

A short review for "Constant Size Point Cloud Clustering: A Compact, Non-Overlapping Solution"

Edited by Mengbai Xiao

A. Guarda, N. Rodrigues, and F. Pereira, "Constant Size Point Cloud Clustering: A Compact, Non-Overlapping Solution," in IEEE Trans. Multimedia, vol. 23, pp. 77-91, 2021.

With the fast development of virtual reality (VR) and augmented reality (AR) techniques, point clouds have emerged as a popular representation for both static and dynamic 3D objects. A point cloud is composed of points that are associated with the <x, y, z> coordinates and other optional attributes like the RGB color, reflectance, normal, etc. Compared to the meshes, another popular 3D representation, point clouds are more lightweight and have attracted increasing attention in the immersive and interactive visual environments in practice.

However, a point cloud has to carry a massive number of points for accurately and clearly representing visual objects, which leads to the demands of efficient compression solutions towards the point clouds. Currently, two codecs for point clouds are being developed in MPEG that support both static and dynamic point clouds, namely Geometry-based Point Cloud Compression (G-PCC) and Video-based Point Cloud Compression (V-PCC). In the coding process, the input point cloud should be segmented into small clusters for more efficiently discovering both the inner-frame and the inter-frame redundancy, which is similar to defining the blocks in JPEG and the macroblocks in H.264/AVC. Moreover, the point clouds have no connectivity information so that it is challenging in semantically understanding the content. Thanks to the fast development of deep learning (DL) techniques, the neural network (NN) based schemes are considered the most effective solutions to this problem. However, in order to make the point cloud as the legal input of the NN, segmenting the point cloud into compact, non-overlapping clusters of a constant size is necessary.

In this paper, Guarda et al. [1] propose a novel point cloud clustering algorithm that is able to generate compact and non-overlapping point clusters of a constant size. Though previous studies [2], [4] have proposed the algorithms achieving the same goal, their solutions can hardly scale to large datasets, e.g., the popular MPEG point cloud dataset [3] where commonly hundreds of thousands of points form a point cloud. The proposed constant size, compact, non-overlapping clustering (C2NO) algorithm is thus expected to have low computation complexity so that it could be applied to the datasets with massive points.

The proposed algorithm is realized into four stages: *initialization*, *iterative clustering*, *clusters refinement*, and *clusters padding* (optional). In the initialization phase, preliminary clustering that considers different local densities of the point cloud is performed. Specifically, the space is regularly divided into equal volumes and each volume is further partitioned adaptively according to its point density. With this method, roughly equal number of points are attributed to each cluster centroid. After the initialization stage, the algorithm iteratively adjusts the point clusters until the point numbers of the clusters are equal. For each adjusting iteration, the algorithm always starts from the cluster with the most/least points and then pushes/pulls points to/from the adjacent clusters. Such a process repeats until all clusters have been checked. With this so-called adjacency restriction method, the point clusters will still overlap with each other, and a cluster refinement stage is required to construct point clusters that are non-overlapping. To achieve this, for each cluster, the algorithm calculates the distance of the points from the centroid and the distances from the centroids of the neighboring clusters. If one point is found its distance from the current cluster centroid is farther than that from one of the neighboring clusters, this implies the point should belong to that neighboring cluster. Once a similar point is found in that neighboring cluster, i.e., the point is closer to the current cluster centroid than the neighboring cluster, swapping is carried out so

that both clusters are more compact. After the cluster refinement stage, overlapping is almost eliminated. In the last step, an optional cluster padding operation is applied to fill the incomplete clusters, making all point clusters have the exactly same point number.

In the evaluation, C2NO is compared to BSC-clustering [4] and SSk-Means clustering [5], which both generate point clusters of a constant size. In the evaluation, two datasets are used: the MPEG dataset (~1 million of points per point cloud) and the ICME dataset (~30 thousands of points per point cloud). To highlight the strength of the proposed algorithm, the following metrics are measured: 1) the cluster dispersion that measures the compactness of the resulting point clusters and 2) the computation complexity that measures the efficiency of the methods used in the comparison.

In terms of the cluster dispersion, the smallest largest and average distance between points and centroids are measured. C2NO clearly outperforms SSk-Means, which is a heuristic-based solution. This indicates that C2NO generates more compact point clusters. As for BSC-clustering, the clusters generated are slightly more compact. This is because BSC-clustering is one of the optimization-based clustering methods. To evaluate the computation complexity, the authors measure the running time of different algorithms on the same machine. While SSk-Means clustering is implemented by Python, BSC-clustering and C2NO are implemented in MATLAB. The experimental results show that the running time of C2NO is almost two orders lower than the other two methods on the ICME dataset. For the MPEG dataset having much larger point clouds, BSC-clustering and SSk-Means clustering even fail converging in a reasonable time.

In summary, the authors of this paper propose an effective and efficient algorithm for partitioning a point cloud into non-overlapping clusters of a constant size. Compared to the state-of-the-art solutions, the proposed algorithm is faster and could scale for the datasets with massive points. I believe this technique would substantially support the future development of compression algorithms towards the point cloud data type as well as the neural networks.

References:

[1] A. Guarda, N. Rodrigues, and F. Pereira, “Constant Size Point Cloud Clustering: A Compact, Non-Overlapping Solution,” in IEEE Transactions on Multimedia, vol. 23, pp. 77-91, 2021.

[2] M. Malinen and P. Franti, “Balanced K-Means for Clustering,” in Proc. Joint IAPR Int. Workshop Structural, Syntactic, Statistical Pattern Recognit., Joensuu, Finland, Aug. 2014, pp. 32-41.

[3] S. Schwarz et al., Common Test Conditions for Point Cloud Compression. Gwangju, Korea: Doc. MPEG N17345, Jan. 2018.

[4] W. Tang et al., “Optimizing MSE for Clustering with Balanced Size Constraints,” Symmetry, vol. 11, p. 338, Mar. 2019.

[5] E. Schubert et al., “A Framework for Clustering Uncertain Data,” Proc. VLDB Endowment, vol. 8, no. 12, pp. 1976-1979, Aug. 2015.

Mengbai Xiao, Ph.D., is a Professor in the School of Computer Science and Technology at Shandong University, China. He received the Ph.D. degree in Computer Science from George Mason University in 2018, and the M.S. degree in Software Engineering from University of Science and Technology of China in 2011. He was a postdoctoral researcher at the HPCS Lab, the Ohio State University. His research interests include multimedia systems, parallel and distributed systems. He has published papers in prestigious conferences such as ACM Multimedia, ACM ICS, IEEE ICDE, IEEE ICDCS, IEEE INFOCOM.

Deep Reinforcement Learning based Brush Stroke Simulation for Image Relighting

A short review for "PR-RL: Portrait Relighting via Deep Reinforcement Learning"

Edited by Debashis Sen

X. Zhang, Y. Song, Z. Li and J. Jiang, " PR-RL: Portrait Relighting via Deep Reinforcement Learning," in IEEE Transactions on Multimedia (Early Access), 2021.

Image editing using brush strokes is a procedure often used by artist to retouch an image for various modification or enhancement purposes [1]. One such activity is to change the lighting effects to alter the lighting condition, which is often performed following a sequential coarse-to-fine strategy. Relighting portrait photos are of particular interest to casual photographers, who can immensely benefit from an automatic and accurate relighting tool that is based on the standard guidelines followed by the artists.

Portrait photos are dominated by faces that have complex and intricate geometric details [2], but at the same time the geometry is predominantly regular and symmetric. Therefore, convolutional neural network (CNN) based end-to-end deep learning approaches are directly applicable, which can be trained on a dataset to perform effective relighting [3], [4]. However, such approaches may introduce locally-confined artifacts depending on the similarity of the image at hand to the dataset images. The CNN-based end-to-end encoder-decoder architectures used are often trained considering overall image and lighting error measures [4] sometimes along with an adversarial loss, which are distant from the guidelines followed by human artists. Hence, a deep learning approach seeking optimality in line with artistic procedures that avoids local errors is highly desirable.

The paper being reviewed here provides one such solution where deep reinforcement learning is applied to simulate image relighting by artists using brush strokes. The presented approach is claimed to be locally effective, scale-invariant and interpretable. The coarse-to-fine procedure followed by artists is modeled as an action of an automatic agent, which is trained through deep reinforcement learning (DRL). The training involves predicting actions and providing feedback reward against an action. The method is a sequential local light editing process, where stroke selection and dodge & burn operations are performed in a coarse-to-find manner in the image being relighted.

The image to be worked upon and a spherical harmonic lighting [5] (SHL) vector are fed as inputs to the approach, where a provision has been given to use a reference image instead of the lighting vector. A light editing stroke is parameterized as its position, shape and lightness and a continuous and high dimensional action space is considered. Deep deterministic policy gradient is employed to model the agent, which generates the action in the continuous space which leads to the rendering of a 'soft' stroke. A deep CNN based encoder-decoder architecture (but not end-to-end) is employed to implement the actor - stroke rendered module. Interpretable image editing by the stroke is performed by simulating the dodge & burn retouching techniques in photography, where the exposure of the relevant area is altered to increase or decrease the light as required. This allows the light in the altered image to be a seamless variation of that in the input image, which simulates a real lighting condition change quite effectively. The process is repeated based on a predefined parameter referred to as the action bundle.

The novel design of the reward to carry out the DRL through a critic deep network, which takes the original and the relighted images and lighting vectors (states) as inputs, is primarily based on two observations. They are:

- More emphasis is given by a human artist on bright and shadow areas of faces in portrait images that are highly susceptible to variation in light.
- A human artist always carries out a coarse-to-fine editing for photo relighting, where retouching is done in a sequential manner.

The overall reward function used is a combination of four specific rewards, which are called the PatchGAN reward, Content L2 reward, Shading reward and Stroke reward. While the PatchGAN reward forces the generated image to be realistic, the Content L2 reward makes the relighted image look similar to the target (given or generated using the given SHL vector). The Shading reward makes relighting look natural and realistic emphasizing on the right face shadow and highlight. The Stroke reward is used to restrict the stroke parameters based on its number in the sequence of such operations required as per the action bundle, forcing the approach to work in a coarse-to-fine manner.

Mathematical models of the brush stroke, the dodge & burn operations and the reward function are provided and clearly explained in the paper. Both the actor (stroke generation) and the critic networks in the approach use the ResNet-18 [6] architecture followed by a fully-connected layer predicting actions and Q values, respectively. Extensive experimental results are provided in the paper justifying that the approach reported performs well on quantitative and qualitative evaluations in comparison to the state-of-the-art.

References:

[1] Antonio Criminisi, Toby Sharp, Carsten Rother, and Patrick P'erez. 2010. Geodesic image and video editing. ACM Trans. Graph. 29, 5, Article 134 (October 2010), 15 pages.

[2] S. Sengupta, A. Kanazawa, C. D. Castillo and D. W. Jacobs, "SfSNet: Learning Shape, Reflectance and Illuminance of Faces 'in the Wild'," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 6296-6305.

[3] H. Zhou, S. Hadap, K. Sunkavalli and D. Jacobs, "Deep Single-Image Portrait Relighting," 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 7193-7201

[4] Tiancheng Sun, Jonathan T. Barron, Yun-Ta Tsai, Zexiang Xu, Xueming Yu, Graham Fyffe, Christoph Rhemann, Jay Busch, Paul Debevec, and Ravi Ramamoorthi. 2019. Single image portrait relighting. ACM Trans. Graph. 38, 4, Article 79 (August 2019), 12 pages.

[5] R. Basri and D. W. Jacobs, "Lambertian reflectance and linear subspaces," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 2, pp. 218-233, Feb. 2003

[6] Z. Huang, S. Zhou and W. Heng, "Learning to Paint With Model-Based Deep Reinforcement Learning," *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 8708-8717.

Debashis Sen is an Assistant Professor in the Department of Electronics and Electrical Communication Engineering and a faculty member 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.

Joint Feature and Video Compression in Scalable Video Coding

A short review for "An Emerging Coding Paradigm VCM: A Scalable Coding Approach Beyond Feature and Signal"

Edited by Tiesong Zhao

S. Xia, K. Liang, W. Yang, L.-Y. Duan and J. Liu. "An Emerging Coding Paradigm VCM: A Scalable Coding Approach Beyond Feature and Signal," IEEE International Conference on Multimedia & Expo (ICME), 2020.

The past decades have witnessed a booming of lossy video codecs, such as H.26x [1], VC-x [2] and AVSx [3] series, which adopted hybrid coding structures with Rate-Distortion Optimization (RDO) to improve their coding performances. Until now, the most advanced lossy encoder is capable of processing 8K videos with high efficiency. In their applications, the ultimate receiver of videos was usually considered as a human user. Therefore, these lossy codecs have been designed with perception-based quality metrics, such as PSNR and SSIM.

Recently, to process the IoT's huge amount of video data, the Video Coding for Machine (VCM) [4] has been exploited. In IoT with massive front-end cameras, large-scale video analytics require a high-performance cloud server, and if necessary, enormous front-end nodes with edge computing. The machine-based processing essentially calls for new machine-oriented video codecs that are more communication-efficient than reigning lossy video codecs. To this end, the MPEG CDVS/CDVA [5] were proposed to finish the video analytics with compressed features instead of high-bitrate videos.

This paper proposes to bridge the gap between signal-level (*i.e.* videos) and task-level (*i.e.* features) representations and connect them in a scalable encoder. A base layer consisting of features extracted from sparse motion patterns is utilized for machine-based recognition and analytics. An enhancement layer consisting of key frames of videos can be further combined with base layer to reconstruct the video sequences with motion. As a result, the base layer serves machine-based processing while the complete scalable encoder (*i.e.,* base + enhancement layers) applies to human vision.

The bridge to connect base and enhancement layers is the sparse motion pattern. This work is designed for action recognition in large-scale networks, where the motion pattern is a compact feature that is critical for skeleton-based action recognition. Inspired by this, the authors extract a sparse representation of motion pattern by jointly considering the compactness and effectiveness of features. At the base layer, the sparse motion pattern is compressed as a feature stream; while at the enhancement layer, the motion information can be further combined with key frames to reconstruct all non-key frames of videos with a generative model. As a result, the complete scalable encoder provides both features and videos with different combinations of layers.

The proposed scalable encoder is examined via comprehensive test on PKU-MMD dataset [6], a large-scale dataset for human action recognition. Compared with H.265/HEVC [7], the base layer improves the action recognition accuracy by 9.4% with a bitrate reduction of 67.9%, which demonstrates the effectiveness of machine-oriented video coding. In addition, the complete scalable encoder improves the SSIM by 0.0063 with a bitrate reduction of 2.7%, as compared with H.265/HEVC, which demonstrates the efficiency of the complete scalable encoder. In conclusion, the proposed scalable encoder improves both machine-oriented and human-oriented applications with reductions in bitrate consumption.

Nowadays, the theoretical and experimental improvements of reigning lossy codecs have encountered a bottleneck. The deep-learning-based compression, which treats the videos as stacked features instead of high-dimensional signals, have attracted attentions of video coding community. This work provides a successful

paradigm to address concerns on both machine-based processing and human-oriented vision. Although the improvement of this work is not very significant, especially considering the newest milestone of lossy video codecs, Versatile Video Coding (VVC), it has provided an effective model to address both concerns in one encoder and can serve as a guidance to develop new VCM codecs.

References:

[1]. B. Bross, Y.-K. Wang, Y. Ye, et al, "Overview of the Versatile Video Coding (VVC) Standard and Its Applications," IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 10, pp. 3736-3764, 2021.

[2]. "ST 2042-1:2017 - SMPTE Standard - VC-2 video Compression," Doc. ST 2042-1, pp. 1–127, 2017.

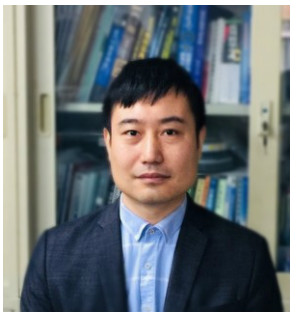
[3]. J. Zhang, C. Jia, M. Lei, et al, "Recent Development of AVS Video Coding Standard: AVS3," in Picture Coding Symposium (PCS), pp. 1–5, 2019.

[4]. L.-Y. Duan, J. Liu, W. Yang, et al, "Video Coding for Machines: A Paradigm of Collaborative Compression and Intelligent Analytics," IEEE Transactions on Image Processing, vol. 29, pp. 8680-8695, 2020.

[5]. L.-Y. Duan, V. Chandrasekhar, et al, "Compact Descriptors for Video Analysis: The Emerging MPEG Standard," IEEE Multimedia, vol. 26, no. 2, pp. 46-54, 2019.

[6]. C. Liu, Y. Hu, Y. Li, et al, "PKU-MMD: A large scale benchmark for skeleton-based human action understanding," in ACM International Conference on Multimedia (ACM MM), pp. 1-8, 2017.

[7]. G. Sullivan, J. Ohm, et al, "Overview of the High Efficiency Video Coding (HEVC) Standard," IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1649-1668, 2012.



Tiesong Zhao, Ph.D, is a Minjiang Distinguished Professor in Fuzhou University, Fujian, China. He received the B. S. and PhD degree from the University of Science and Technology of China and City University of Hong Kong, in 2006 and 2011, respectively.

His research interests include multimedia signal processing, coding, quality assessment and transmission. Due to his contributions in video coding and transmission, he received the Fujian Science and Technology Award for Young Scholars in 2017. He has also been serving as an Associate Editor of IET Electronics Letters since 2019.

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.