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

MMTC Communications – Review

Vol. 11, No. 3, June 2020

IEEE COMMUNICATIONS SOCIETY

TABLE OF CONTENTS

Message from the Review Board Directors 2

Intelligent Video Streaming between UAVs and Ground Clients
A short review for "Sensor-Augmented Neural Adaptive Bitrate Video Streaming on UAVs" (Edited by Ye Liu) 3

Cache-Assisted Dynamic Video Delivery for Mobile Users
A short review for "Markov Decision Policies for Dynamic Video Delivery in Wireless Caching Networks" (Edited by Cong Shen) 5

Recognizing Similar Species with Few Images
A short review for "Piecewise classifier mappings: learning fine-grained learners for novel categories with few examples" (Edited by Jun Zhou) 7

Message from the Review Board Directors

Welcome to the June 2020 issue of the IEEE ComSoc MMTC Communications – Review.

This issue comprises three reviews that cover multiple facets of multimedia communication research including video streaming, wireless caching and piecewise classifier mappings. These reviews are briefly introduced below.

The first paper is published in IEEE Transactions on Multimedia and edited by Dr. Ye Liu. It presents a sensor-augmented system for adaptive bitrate video streaming between UAVs and ground clients.

The second paper is published in IEEE Transactions on Wireless Communications and edited by Dr. Cong Shen. It presents a Markov decision policy for dynamic video delivery in wireless caching networks.

The thrid paper is published in IEEE Transactions on Image Processing and edited by Dr. Jun Zhou. It presents piecewise classifier mappings with learning fine-grained learners for novel categories with few examples.

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

IEEE ComSoc MMTC Communications – Review Directors

Qing Yang
University of North Texas, USA
Email: qing.yang@unt.edu

Roger Zimmermann
National University of Singapore, Singapore
Email: rogerz@comp.nus.edu.sg

Wei Wang
San Diego State University, USA
Email: wwang@mail.sdsu.edu

Zhou Su
Shanghai University, China
Email: zhousu@ieee.org

Intelligent Video Streaming between UAVs and Ground Clients

A short review for "Sensor-Augmented Neural Adaptive Bitrate Video Streaming on UAVs"

Edited by Ye Liu

X. Xiao, W. Wang, T. Chen, Y. Cao, T. Jiang, and Q. Zhang, " Sensor-Augmented Neural Adaptive Bitrate Video Streaming on UAVs," IEEE Transactions on Multimedia, vol. 22, vo. 6, June 2020.

Unmanned aerial vehicles (UAVs) [1] are playing important roles in different kinds of application fields. For example, UAVs can be used for aerial monitoring in smart agriculture [2]. In industrial area, the detection of gas leaks, oil spills and other mission critical applications is available by drones [3]. In these applications, it is critical to record high-definition videos of superior quality and seamlessly share them with ground clients [4]. However, most of the commercial UAVs currently adopt fixed-bitrate approaches to report real-time video stream, which often lead to poor service quality. On the other hand, although a number of adaptive bitrate algorithms have been proposed, but they all focused on scenarios in ground situations and do not fit well for air-to-ground wireless multimedia delivery. To address this issue, this paper presents a sensor-augmented system for adaptive bitrate video streaming between UAVs and ground clients.

A comprehensive measurement study was first conducted to analyze the impact of UAV's motions on throughput from the perspectives of link distance, velocity, and acceleration in both controlled flight states and uncontrolled flight conditions. In the tests, a DJI Matrice 100 UAV is used as transmitter to deliver the video stream to a laptop through IEEE 802.11n wireless protocol. The three sites of measurement are playground, plaza, and pool on a campus, respectively. Under controlled flight situations, the obtained observations are threefold, which are listed as follows: (1) There has been a sharp fall in the throughput when the link distance increase; (2) The throughput diminishes quickly when the velocity range increases, no matter whether the link distance is short or long; (3) The low value of acceleration data is dominated by the UAV vibrations, while the high value can indicate the changes in flight states.

Moreover, some counter-intuitive findings were observed under uncontrolled flight states. For example, it is not a linear relationship between the throughput and link distance, and should consider it as a probability distribution. In terms of the impact of acceleration, the throughput also fluctuates dramatically due to sudden change in velocity. Based on these observations, the conclusion is that the relationship between throughput and sensor data cannot be described in an analytical expression way. Therefore, it needs to find an intelligent solution, that can extract the relationship by the system itself, and does not rely on preconfigured analytical expressions.

Neural networks [5] are good candidates to perform such tasks, which motives the authors to design a Sensor-Augmented Adaptive BitRate algorithm (SA-ABR) for video streaming. It leverages deep reinforcement learning, long short-term memory [6], and inherent sensor data.

A training methodology is first designed to faithfully model the dynamics of video streaming in client applications, in which the new video states are directly calculated based on current information. Second, the sensor data is preprocessed to eliminate noises and disturbances. Both distance data and velocity data are quantized, as well as acceleration data. The results are as follows: (1) the link distance over 50 meter is encoded as "1" , while it is marked as "0" when the distance is less 50 meter; (2) The three levels of velocity are divided into below 8 m/s, 8-12 m/s, and over 12 m/s, which are encoded as "0", "1", and "2", respectively; (3) For acceleration, the "0" indicates it is below 18 m/s^2, and the "1" means the acceleration is over 18 m/s^2. Third, DRL Training process is performed. Its network architecture consists of UAV and video state, long short-term memory, actor network, and critic network. The design of deep reinforcement learning policy, training algorithm, data sampling, and reward strategy are carefully considered.

Finally, the performance of the proposed approach was evaluated through a series of field test, including the comparison between the proposed SA-ABR and existing state-of-the-art ABR algorithms, the analysis of benefits from LSTM network, and the advantage of feeding UAV sensor data into neural networks. The results shown its feasibility and outperformance for video streaming from the air to ground.

In summary, the work introduces a novel sensor-augmented neural adaptive bitrate solution for real-time high-definition video streaming between UAVs and ground clients.

References:

[1] C. Zhan, H. Hu, Z. Wang, R. Fan and D. Niyato, "Unmanned Aircraft System Aided Adaptive Video Streaming: A Joint Optimization Approach," in IEEE Transactions on Multimedia, vol. 22, no. 3, pp. 795-807, March 2020.

[2] Y. Liu, X. Ma, L. Shu, G. P. Hancke and A. M. Abu-Mahfouz, "From Industry 4.0 to Agriculture 4.0: Current Status, Enabling Technologies, and Research Challenges," in IEEE Transactions on Industrial Informatics, 2020.

[3] Z. Zhou, C. Zhang, C. Xu, F. Xiong, Y. Zhang and T. Umer, "Energy-Efficient Industrial Internet of UAVs for Power Line Inspection in Smart Grid," in IEEE Transactions on Industrial Informatics, vol. 14, no. 6, pp. 2705-2714, June 2018.

[4] J. Yu, A. Vandanapu, C. Qu, S. Wang and P. Calyam, "Energy-aware Dynamic Computation Offloading for Video Analytics in Multi-UAV Systems," 2020 International Conference on Computing, Networking and Communications (ICNC), Big Island, HI, USA, 2020, pp. 641-647.

[5] M. Chen, U. Challita, W. Saad, C. Yin and M. Debbah, "Artificial Neural Networks-Based Machine Learning for Wireless Networks: A Tutorial," in IEEE Communications Surveys & Tutorials, vol. 21, no. 4, pp. 3039-3071, Fourthquarter 2019.

[6] S. Agethen and W. H. Hsu, "Deep Multi-Kernel Convolutional LSTM Networks and an Attention-Based Mechanism for Videos," in IEEE Transactions on Multimedia, vol. 22, no. 3, pp. 819-829, March 2020.

Ye Liu received the M.S. and Ph.D. degrees from Southeast University, Nanjing, China in 2013 and 2018, respectively. He was a visiting scholar at Montana State University, Bozeman, USA from October 2014 to October 2015. He was a visiting PhD student from February 2017 to January 2018 in Networked Embedded Systems Group at RISE SICS (Swedish Institute of Computer Science). He is currently a researcher at Nanjing Agricultural University. His current research interests include wireless sensor networks, energy harvesting systems, and smart agriculture. He has published papers in prestigious journals such as IEEE Communications Magazine, IEEE Transactions on Industrial Informatics, IEEE Network, IEEE Internet of Things Journal, ACM Transactions on Embedded Computing System, and so on. He was awarded the 1st place of the EWSN Dependability Competition in 2019.

.

Cache-Assisted Dynamic Video Delivery for Mobile Users

A short review for "Markov Decision Policies for Dynamic Video Delivery in Wireless Caching Networks"

Edited by Cong Shen

Minseok Choi, Albert No, Mingyue Ji and Joongheon Kim, "Markov Decision Policies for Dynamic Video Delivery in Wireless Caching Networks," IEEE Transactions on Wireless Communications, vol. 18, no. 12, pp. 5705-5718, Dec. 2019.

Within few years, it has been expected that tens of exabytes of global data traffic be handled on daily basis, and multimedia services, in particular, on-demand video streaming will account for about 70% of them [1]. In such services, a relatively small number of popular contents is requested at ultra high rates and playback delay is one of the most important measurement criteria of goodness [2]. To deal with these challenges, wireless caching technologies have been studied for video streaming services by storing popular videos in caching helpers located nearby users during off-peak time [3]-[7]. Therefore, it is obvious that storing and delivery of video files are of major research interests in wireless caching networks.

This paper proposes a dynamic video delivery policy in the wireless caching network depending on stochastic network states. The proposed policy makes three different but necessary decisions for the streaming user: 1) caching node for video delivery, 2) video quality and 3) the quantity of video chunks to receive. Here, caching node association and decisions of video quality and amounts of receiving chunks are conducted in different timescales. Since wireless link activation for video delivery is time-consuming, it is reasonable that caching node association is performed slower than other decisions. The proposed content delivery method reflects user mobility. Therefore, the user enables to associate with the caching node based on estimation of the short-term decisions on quality and receiving chunk amounts in future.

The core idea of this paper comes from the tradeoff between video quality and playback delay that the user experiences while receiving the desired video chunks. If the user wants to receive the high-quality file, the caching node cannot deliver many chunks at once. Since the playback delay is generated when the next chunk is not arrived in the user queue, if the user is assumed to receive chunks in order, the playback delay is generated when the queue is empty. Therefore, the user has to adaptively control the queue length, i.e., the number of receiving chunks for avoiding the playback latency. Additionally, in the wireless caching network model where different caching nodes can store the identical videos but different qualities, this tradeoff between video quality and playback delay is dependent also on node association.

The proposed video delivery scheme maximizes the average streaming quality while averting playback latency. The authors adopt the long-term average video quality in each received chunk as a performance metric, because different videos can have different number of chunks. The long-term time-averaged playback delay occurrence rate can be limited by pursuing the strong stability of the converted queue model which is defined as a large constant minus the user queue length. The optimization framework of video delivery policy in two different timescales is constructed based on frame-based Lyapunov optimization theory and Markov decision process (MDP). For every caching node candidate, the user can perform MDP to determine they are suitable for video delivery. After that, the user can find the most appropriate node for association according to the frame-based Lyapunov optimization theory. Also, the decisions on video quality and the quantity of video chunks to receive in shorter timescales are obtained by the MDP.

For the proposed scheme, the authors also present the distance-based interference management. The two different safety distances for streaming users and their associated caching nodes respectively to keep the interference levels below the predetermined threshold. A new streaming user who wants to exploit the wireless caching network should be generated outside the safety radii of all caching nodes associated with the existing users. In addition, the new user has to find the caching node to receive the desired content outside the safety radii of all existing users. Based on this concept, the proposed video delivery scheme can be performed with multiple activated delivery links.

Simulation results show that the proposed video delivery balances the tradeoff between video quality and playback latency in various environments, e.g., node distributions, caching distributions, Lyapunov parameter, and interference-to-noise ratio. The authors intentionally compare the proposed one with two extreme cases: 1) the nearest caching node association which is optimal for delay minimization, and 2) association with the caching node having the highest-quality file. The proposed scheme generates the small latency in the same scale of the first extreme case, while achieving the very high quality measure almost the same as that of the second extreme case. Also, one can see that the existing delivery scheme is not appropriate for delivering the video contents consisting of multiple chunks in order. Further, the proposed technique can adjust the tradeoff between performances of video quality and playback delay by controlling the system parameter.

In summary, this paper studies the dynamic delivery policy of video files of various quality levels in the wireless caching network. When the user is moving, the streaming user makes decisions on caching node to receive the desired file, video quality, and the number of receiving chunks. The different timescales are considered for the caching node association and decisions on quality and the number of receiving chunks. Extensive simulation results show that the proposed delivery can balance as well as adjust the tradeoff between video quality and playback latency in various environments.

References:

[1] *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2017–2022 White Paper*. Accessed: Sep. 3, 2019. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/serviceprovider/visual-networking-index-vni/white-paper-c11-738429.htm.

[2] X. Cheng, J. Liu, and C. Dale, “Understanding the characteristics of Internet short video sharing: A YouTube-based measurement study,” *IEEE Trans. Multimedia*, vol. 15, no. 5, pp. 1184–1194, Aug. 2013.

[3] N. Golrezaei, K. Shanmugam, A. G. Dimakis, A. F. Molisch, and G. Caire, “FemtoCaching: Wireless video content delivery through distributed caching helpers,” in *Proc. IEEE INFOCOM*, Orlando, FL, USA, Mar. 2012, pp. 1107–1115.

[4] C. Yang, Y. Yao, Z. Chen, and B. Xia, “Analysis on cache-enabled wireless heterogeneous networks,” *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 131–145, Jan. 2016.

[5] L. Zhang, M. Xiao, G. Wu and S. Li, "Efficient Scheduling and Power Allocation for D2D-Assisted Wireless Caching Networks," in IEEE Transactions on Communications, vol. 64, no. 6, pp. 2438-2452, June 2016.

[6] W. Jiang, G. Feng, and S. Qin, “Optimal cooperative content caching and delivery policy for heterogeneous cellular networks,” *IEEE Trans. Mobile Comput.*, vol. 16, no. 5, pp. 1382–1393, May 2017.

[7] M. Choi, J. Kim, and J. Moon., “Wireless video caching and dynamic streaming under differentiated quality requirements,” *IEEE J. Sel. Areas Commun.*, vol. 36, no. 6, pp. 1245–1257, Jun. 2018.

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 Wireless Communications and IEEE Wireless Communications Letters.

Recognizing Similar Species with Few Images

A short review for "Piecewise classifier mappings: learning fine-grained learners for novel categories with few examples"

Edited by Jun Zhou

Xiu-Shen Wei, Peng Wang, Lingqiao Liu, Chunhua Shen and Jianxin Wu, "Piecewise classifier mappings: learning fine-grained learners for novel categories with few examples", IEEE Transactions on Image Processing, 28 (12), 6116-6125, 2019

Fine-grained image recognition aims at distinguishing categories that are very similar to teach other. It is a highly useful technique for species recognition in which many species have only subtle differences and sometimes can only be distinguished by experts. As in other computer vision tasks, deep learning approaches have been adopted to tackle the fine-grained recognition problem, and have achieved amazing performance [1]. However, these approaches require categories be known in advance and provided in the training data, so they cannot be easily extended to novel categories. Furthermore, the success of deep learning is usually built on top of a large amount of training data which in reality is difficult or expensive to obtain in particular for many species recognition applications. It is therefore interesting to explore solutions that learn from only few examples per category. Few-shot learning is such a learning paradigm [2], which can learn new concepts from several or even one example. This paper proposes to integrate few-shot learning with fine-grained image recognition which is a challenging topic.

The proposed framework adopts a meta-learning paradigm. Given few labelled training samples or exemplars, meta-learning maps these samples to their category classifier. The training process uses an auxiliary dataset to learn the mapping function. Then the learned model is evaluated on a testing set with novel categories in which one or several exemplars are available for few-shot learning. During the training process, a set of meta-training sets are generated from the auxiliary dataset. These sets contain randomly chosen categories and associated images. Similar to the testing stage, the meta-training set consists of limited exemplars for the few-shot learning setting, and a query set. The exemplars are fed into the mapping function to produce classifiers. Then the classifiers are used to classify the query set. The mapping function can be learned By minimizing the classification loss.

The above-mentioned process is implemented using a deep neural network with two modules for representation learning and classifier mapping respectively. The representation learning module uses a bilinear convolutional neural network (CNN) [3]. The bilinear CNN has two parallel feature extractors. Each feature extractor produces a feature vector at each spatial location. Then at each location, the outer product of two parallel generated features is calculated and vectorized. The feature vectors are pooled to get the image representation. An interesting observation here is that the feature vector can be converted into a set of sub-vectors that focus on certain part of the image.

The second module of the network, classifier mapping, then takes the mean image representation as input and map them into corresponding category classifiers. Here a piece-wise mapping approach is adopted which produces sub-classifiers from sub-vectors, and then combine the sub-classifiers into the final classifier. This strategy reduces the model parameters by two orders of magnitude from the traditional global mapping strategy.

The proposed method was evaluated on three fine-grained image datasets, including CUB Birds [4], Stanford Dogs [5] and Stanford Cars [6]. Each dataset contains 100+ categories and 10,000+ images. Each dataset was randomly split into an auxiliary training set and a testing set. Once the model had been learned from the training set, one or five images were chosen from each testing set for one-shot or five-shot learning, with the rest images for testing. The results show that high accuracy improvement was achieved on all three datasets, when compared against six alternative methods. In particular, the proposed piece-wise mapping strategy demonstrate significant benefit in both accuracy and model complexity over the global mapping strategy.

In summary, this paper showcase the integration of two challenging computer vision tasks, fine-grained image classification and few-shot learning. The proposed solution is based on a piecewise mappings function in the classifier mapping module, which produces optimized decision boundary. This idea can be very useful in designing future deep learning network structure. This work can also be adopted by real-world specie recognition systems given its effectiveness and efficiency.

References:

[1] S. Huang, Z. Xu, D. Tao, and Y. Zhang, “Part-stacked CNN for fine-grained visual categorization,” in CVPR, pp. 1173–1182, 2016.

[2] L. Bertinetto, J. F. Henriques, J. Valmadre, P. Torr, and A. Vedaldi, “Learning feed-forward one-shot learners,” in NIPS, pp. 523–531, 2016.

[3] T. Lin, A. RoyChowdhury, and S. Maji, “Bilinear convolutional neural networks for fine-grained visual recognition,” IEEE TPAMI, vol. 40, no. 6, pp. 1309–1322, 2017.

[4] C. Wah, S. Branson, P. Welinder, P. Perona, S. Belongie. “The Caltech-UCSD Birds-200-2011 Dataset.” Computation & Neural Systems Technical Report, CNS-TR-2011-001.

[5] A. Khosla, N. Jayadevaprakash, B. Yao, and L. Fei-Fei, “Novel dataset for fine-grained image categorization,” in CVPR Workshop on Fine-Grained Visual Categorization, pp. 806–813, 2011.

[6] J. Krause, M. Stark, J. Deng, and L. Fei-Fei, “3D object representations for fine-grained categorization,” in ICCV Workshop on 3D Representation and Recognition, pp. 554–561, 2013.

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

He is now an associate professor in the School of Information and Communication Technology in Griffith University. Prior to this appointment, he had been a research fellow in the Australian National University, and a researcher at NICTA, Australia. His research interests are in spectral imaging, pattern recognition, computer vision, and their applications to and environmental informatics and remote sensing.

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.