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

MMTC Communications – Review

Vol. 13, No. 2, April 2022

IEEE COMMUNICATIONS SOCIETY

TABLE OF CONTENTS

Message from the Review Board Directors

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

This issue comprises four reviews that cover multiple facets of multimedia communication research including subjective evaluations of 3D content and multisensory 360-degree videos, novel generative adversarial networks for multimedia applications, and deep learning for wireless network optimization. These reviews are briefly introduced below.

The first paper, published in ACM Transactions on Applied Perception and edited by Dr. Carsten Griwodz, evaluates three classical methods for subjective quality assessment of 3D visual content, which is increasingly used in VR and AR applications nowadays. The results show that the double stimulus impairment scale (DSIS) method outperforms the other two in various scenarios.

The second paper, edited by Dr. Yong Luo, was published in IEEE Transactions on Multimedia. This paper proposes TWGAN, a generative adversarial network model that consists of one generator and a twin discriminator and uses the combination of saturating and non-saturating loss functions. Experiments show that TWGAN can effectively address the training instability problem.

The third paper, edited by Dr. Shengjie Xu, was published in IEEE Wireless Communications. This paper reviews recent works that applies deep learning in wireless network optimization.

The fourth paper, edited by Dr. Roberto G. de A. Azevedo, is to be published in IEEE Trans. on Visualization and Computer Graphics. This paper presents a study on how the combination of 360-degree video degradation and multisensory content affects user's quality of experience.

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
Rutgers University, USA
Email: yao.liu@rutgers.edu

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

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

Subjective QoE Assessment for Visual 3D Content

A short review for
"Comparison of Subjective Methods for Quality Assessment of 3D Graphics in Virtual Reality"

Edited by Carsten Griwodz

Y. Nehmé, J-P. Farrugia, F. Dupont, P. Le Callet, and G. Lavoué, "Comparison of Subjective Methods for Quality Assessment of 3D Graphics in Virtual Reality", in. ACM Trans. Appl. Percept. 18(1), 2021

In both our private and professional lives, 3D content is increasingly present, and can increase our understanding of reality, virtuality and every intermediate form **[1]**. The 3D content that we see and interact with comes in a vast multitude of forms, ranging from simple labels attached to real-world imagery in AR **[2]**, to fully rendered virtual environments in VR that are may be encoded as textured meshes **[3]** or pointclouds **[4]**.

Frequently, the content for such virtual experiences needs to be delivered over the network, and due to the vast amount of information that makes up virtual objects and virtual scenes, relevant data must be selected from a large overall dataset, and it must be compressed using lossy codecs for timely transfer. Like the uses of virtual content, the means of compression can vary widely, and efficient compression strategies are a hot research topic. An understanding of suitable compression methods for a particular scenario is required to select the best combination of compression and transfer method to achieve a high degree of user satisfaction. It is therefore essential that we learn about methods that are suitable for the assessment of users' quality-of experience (QoE) in this quickly developing visual medium.

The paper by Nehmé et al. that is reviewed in this letter takes a detailed look at classical methods for subjective quality assessment that have proven their worth in video compression and transfer. Subjective quality assessment provides us with an understanding of real human users' perception of content's visual quality. Truly understanding how users perceive the unavoidable temporal and spatial inaccuracies due to lossy compression and latency introduced by network transfer is the basis for decision-making at all levels of new virtual reality systems, whether the decisions are made by hand-coded or data-driven algorithms.

Of course, not all possible contents can be investigated in a single study, and Nehmé et al. focus in this paper on 3D objects that are observed by a human user through a VR headset. The authors aim at a strictly controlled scenario and put the 3D object into a neutral room with grey walls and use lighting that appears ambient and does not generate shadows. Although the participants in the user study observe the object through an HTC Vive headset in this study, their physical motions do not generate any motion of the camera in the virtual world. This virtual scenario is a translation of the well-known standardized setups for the assessment of video quality into the virtual world **[5]**. Furthermore, where videos in a video quality assessment study would obviously present content that changes over time, Nehmé et al. create a fixed motion path of the virtual camera around the object, while the object itself is static.

For a multimedia systems researcher, these restrictions may appear unfortunate because desirable multimedia systems allow the user at least 3 and often 6 degrees of freedom of motion in the environment, and environments in VR applications become increasingly dynamic as well. To understand these restrictions, it is important to look at the authors' motivation.

Their goal is to assess the suitability of methods that were developed with video quality assessment for conducting QoE studies. Their goal is not to understand the user experience in a particular VR application. To focus on this goal, it is important to maximize repeatability and avoid a strong engagement of the participants with the content.

The paper is rather meant to several questions about the assessment methods:

- consistency of results,
- accuracy of scores, and
- confidence of scores.

The assessment methods that are considered in the paper are the Absolute Category Rating with Hidden Reference (ACR-HR), the Double Stimulus Impairment Scale (DSIS), and the Subjective Assessment Methodology for Video Quality (SAMVIQ), all of which are standardized by the International Telecommunication Union **[5]**.

Nehmé et al. present the properties of these three methods, where ACR-HR is a single stimulus method, DSIS a double stimulus method, and SAMVIQ a multi-stimuli method with random access. Due to the major differences between the three methods, the authors compare the performance of ACR-HR and DSIS first before investigating SAMVIQ and finally comparing all of them.

ACR-HR and DSIS are compared first. Obviously, there is not baseline for true QoE values of the 3D models used for the study. Lacking this, the authors conduct the study with two groups of participants and assess the consistency of their responses. Although the general trend of quality assessment between both approaches and with both groups of participants is the same, Nehmé et al. observe that the consistency of DSIS results across groups is higher, probably due to the built-in normalization of a double stimulus method. The method is also shown to be more accurate in the sense of reducing the randomness of results. Concluding that DSIS is preferrable over ACR-HR due to the reference views, the authors do then compare it to the random-access method SAMVIQ. In principle, both methods yield the same results. However, by studying the influence of the number of participants on confidence intervals, the authors can conclude that the SAMVIQ method achieves consistent QoE assessment with fewer participants. When two groups of 12 participants are combined, however, DSIS achieves the same confidence as SAMVIQ. Due to the extensive time that participants spend on rating using the SAMVIQ method, the authors conclude that the DSIS method is the better method for real-world testing although it requires more participants.

Consequently, Nehmé et al. provide with their work a clear answer to the question about the best performing and most suitable QoE assessment method for VR content that many multimedia researchers ask. They determine that a simple double stimulus method like DSIS can achieve consistent, accurate results faster than a complex method like SAMVIQ that provides participants with the option to explore. This establishes important baseline knowledge for more complex VR scenarios that we will be facing in a lot of upcoming multimedia research.

References:

[1] Paul Milgram, Haruo Takemura, Akira Utsumi, Fumio Kishino, "Augmented reality: a class of displays on the reality-virtuality continuum," Proc. SPIE 2351, Telemanipulator and Telepresence Technologies, Dec 1995.

[2] Y. Siriwardhana, P. Porambage, M. Liyanage and M. Ylianttila, "A Survey on Mobile Augmented Reality With 5G Mobile Edge Computing: Architectures, Applications, and Technical Aspects," in IEEE Communications Surveys & Tutorials, 23(2), 2021.

[3] A. Maglo, G. Lavoué, F. Dupont, and C. Hudelot, "3D Mesh Compression: Survey, Comparisons, and Emerging Trends". ACM Comput. Surv. 47(3), 2015.

[4] P. E. J. Kivi, M. J. Mäkitalo, J. Žádník, J. Ikkala, V. K. M. Vadakital and P. O. Jääskeläinen, "Real-Time Rendering of Point Clouds With Photorealistic Effects: A Survey," in *IEEE Access* 10, 2022

[5] ITU-R BT.500-13. "Methodology for the subjective assessment of the quality of television pictures BT series broadcasting service". In Proceedings of the International Telecommunication Union, 2012.

Dr. **Carsten Griwodz** is professor at the University of Oslo and leader of the Section for Distributed Infrastructure and Security. He is also an associate researcher at Simula Research Laboratory. His research interest is the performance of multimedia systems. He is concerned with streaming media, which includes all kinds of media that are transported over the Internet with a temporal demands, including stored and live video as well as games and immersive systems. To achieve this, he wants to advance operating system and protocol support, parallel processing, and the understanding of the human experience.

Towards Stable Training for Generative Adversarial Networks

A short review for "TWGAN: Twin Discriminator Generative Adversarial Networks"

Edited by Yong Luo

Z. Zhang, M. Li, H. Xie, J. Yu, T. Liu, and C. W. Chen, "TWGAN: Twin Discriminator Generative Adversarial Networks," in IEEE Trans. Multimedia, Vol. 24, pp. 677-688, 2022.

Generative Adversarial Networks (GANs) [1] is one kind of generative model, in which a game is played between two competing neural networks. A generator (G) which produces synthetic data with some given noise and a discriminator (D) which distinguishes whether the data is from the output of the generator or real data.

GAN can produce very visually appealing samples and has become more and more popular in multimedia applications. Although GANs have achieved great success, they are often difficult to train and suffer from the training instability problem [2]. This is because in the Standard GAN (SGAN) [1], D is a classifier, and its output is the probability that whether the input data is real. When D is optimal, the loss function of SGAN is approximately equal to the Jensen Shannon divergence (JSD) [2] between the distribution of generator (P_g) and the distribution of discriminator (P_{data}). While in a typical case, P_g and P_{data} do not overlap or the overlapping parts can be ignored. This phenomenon will cause the JSD item cannot be optimized further and lead to issue of the gradient vanishing, and thus training instability. Additionally, the loss function of SGAN is also called saturating loss function [3].

Recent attempts have been made to address the training instability problem by improving the training of GAN. Arjovsky et al. [2] first proposed a method named $-\log D$ alternative. They change the loss function of the generator from $1-\log D$ to $-\log D$ to address training instability. Although this method is effective, unfortunately, it will incur another issue that the optimization of this kind of loss function will cause the gradient of generator volatile when D is optimal. This kind of loss function is named as non-saturating loss [3] to distinguish it from the SGAN's loss function.

Then, Arjovsky et.al [4] propose a distance to take place of the JS divergence when we optimize the generator under optimal discriminator. They name this distance as Wasserstein distance, and they propose a novel GAN called Wasserstein GAN (WGAN) [4]. Although this method is effective enough, the performance of WGAN is still not satisfied. Lately, Simon Jenni and Paolo Favaro [5] proposed a method that added the same noise both in G and D to address the training instability problem. This method will make the gradient of the generator not vanish when the optimal discriminator is obtained. Although this method is useful, however, it may change the distribution of real data (P_{data}) a little when we optimize the GAN, this will decrease the performance of the GAN.

Concentrating on the training instability problem, in this paper, we propose a novel GAN named TWGAN, which consists of one generator and a twin discriminator. The twin discriminator consists of two discriminators and the saturating and non-saturating loss functions are applied to the two discriminators respectively to balance the distribution optimization between the theory and ideal case.

The conception of TWGAN and the other two-discriminator GAN models (such as D2GAN [6] and STDGAN [7]) are fundamentally different. In other two-discriminator GAN model, two discriminators give opposite probability to real data and synthetic data. The first discriminator D_1 rewards high scores for the samples from the data distribution, while the second one D_2 favors the samples from the generator conversely. While in TWGAN, two discriminators have identical architecture and both of them favor the samples from the real data. Besides, the other two-discriminator GAN model aims to address the

mode collapse problem, the function of D_2 is to avoid producing similar samples from generator. While TWGAN aims to address the training instability problem, the function of D_2 with non-saturating loss is to balance the distribution optimization between the theory and ideal case.

The authors also introduce whitening and coloring transform, Spectral Normalization and ResBlock in TWGAN to obtain the higher inception score [8] and lower FID [9]. Besides, the authors analyze the distribution optimization process of SGAN and TWGAN and show that the distribution optimization process of SGAN is different from that of the ideal case, while the distribution optimization process of TWGAN is similar to that of the ideal case. Furthermore, TWGAN can address the gradient volatile issue in $-\log D$ alternative method [2]. The using of the twin discriminator can balance the $-2JS(p_{data} \parallel p_G)$ item which the non-saturating loss function produced.

The lower variances obtained from multiple times run compared with other GANs in the extensive experiments on diverse datasets show that the benefits of the twin discriminator and the combination of saturating and non-saturating loss functions. The higher inception score and lower FID, as well as the faster convergence speed obtained in the experiments also shows the superiority of TWGAN.

In summary, this paper proposes a novel GAN, namely TWGAN, which consists of a generator and two discriminators. The saturating and non-saturating loss functions are applied to the two discriminators respectively to balance the distribution optimization between the theory and actual training. A drawback of TWGAN is that the number of network parameters and the training time cost will be increased, and thus a future work may be to employ some efficient learning methods to reduce the number of parameters.

References:

[1] I. Goodfellow et al., "Generative adversarial nets," in Proc. Adv. Neural Inf. Process. Syst., 2014, pp. 2672-2680.

[2] M. Arjovsky and L. Bottou, "Towards principled methods for training generative adversarial networks," in Proc. Int. Conf. Learn. Representations, 2017.

[3] A. Jolicoeur-Martineau, "The relativistic discriminator: A key element missing from standard GAN," in Proc. Int. Conf. Learn. Representations, 2018.

[4] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," in Proc. Int. Conf. Learn. Representation, 2017.

[5] S. Jenni and P. Favaro, "On stabilizing generative adversarial training with noise," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2019, pp. 12 145-12 153.

[6] T. Nguyen, T. Le, H. Vu, and D. Phung, "Dual discriminator generative adversarial nets," in Proc. Adv. Neural Inf. Process. Syst., 2017, pp. 2667-2677.

[7] Z. Zhang and J. Yu, "STDGAN: Resblock based generative adversarial nets using spectral normalization and two different discriminators," in Proc. 27th ACM Int. Conf. Multimedia., 2019, pp. 674-682.

[8] T. Salimans, "Improved techniques for training GANs," in Proc. Adv. Neural Inf. Process. Syst., 2016, pp. 2234-2242.

[9] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local nash equilibrium," in Proc. Adv. Neural Inf. Process. Syst., 2017, pp. 6626-6637.

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

His research interests are primarily on machine learning and data mining with applications to multimedia information understanding and analysis. He has authored or co-authored over 50 papers in top journals and prestigious conferences including IEEE T-PAMI, IEEE T-NNLS, IEEE T-IP, IEEE T-KDE, IEEE T-MM, WWW, IJCAI, AAAI, CIKM, ICDM and ICME. He received the IEEE Globecom 2016 Best Paper Award.

Deep Learning for Future Wireless Networking

A short review for "Deep Learning for Wireless Networking: The Next Frontier"

Edited by Shengjie Xu

Y. Cheng, B. Yin and S. Zhang, "Deep Learning for Wireless Networking: The Next Frontier," in IEEE Wireless Communications, vol. 28, no. 6, pp. 176-183, December 2021.

This paper reviews recent efforts in leveraging deep learning (DL) for addressing wireless network optimization problems, presenting a fundamental understanding of where and how the supremacy of DL based approaches comes versus the conventional modeling-based approaches. The authors also present some basic research challenges and some promising research directions for fully exploiting the potential of DL in wireless network. The authors also illustrated the effectiveness of DL with an innovative case study of integrating DL with multi-hop wireless network flow optimization.

The authors first present the tremendous benefits in DL. Due to its recent success in various domains [1], DL has been identified as a disruptive enabler for automatic and autonomous network management. Incorporating DL intelligence into wireless networks not only has the potential to replace the manual interventions involved in the current engineering-intensive network management tasks, but also give rise to novel network optimization approaches that deliver superior system performance in real-time. Given the strong capability in big data analytics, DL techniques can be leveraged to distill insightful knowledge (e.g., the intricate correlations between the network configurations and the achievable performances) from the abundant data over modern wireless networks to enable innovative control and optimization methods.

In particular, deep reinforcement learning (DRL) techniques, which have demonstrated impressive results in areas such as robotics and video games, provide promising opportunities for developing online control policies in complex and large-scale networking scenarios. Moreover, the emerging DL hardware accelerators significantly speed up the DL-related operations; the network controllers equipped with modern DL technologies are obtaining the capability to promptly adjust and optimize resource allocation efficiently in response to rapidly changing networking conditions.

The authors then examine some legacy strategies that fail to fully exploit the computation experience of solving all the historical problem instances. Wireless network optimization tasks usually involve the management or allocation of network resources (e.g., radios, channels and transmit power), intending to deliver good network performance. Most existing approaches to such tasks follow the paradigm of mathematical programming, in which a wireless network utility maximization (WNUM) problem is formulated and solved. When designing algorithms for WNUM problems, computational overhead and optimality are the two most important concerns. However, wireless networks are inherently dynamic and require adaptive control. Regarding a specific network optimization task, it is not uncommon that the network controller has to solve many WNUM problems with the same or similar structure to guarantee stable network performance [2].

After recognizing these limitations in wireless network optimization, the authors present the supremacy of DL in wireless network optimization. DL technologies can advance the state of the art of wireless network optimization from four aspects: establishing practical and informative formulations of the optimization tasks; alleviating the computational overhead of generating approximated solutions; exploring approaches that are superior to the existing ones; and discovering new latent knowledge that facilitates efficient and effective optimization.

DL technologies provide a universal approach to model the optimization tasks that are intractable to

formulate mathematically. The work in [3] employs DNNs to develop an end-to-end wireless communication system in which several key functions, including encoding, decoding, modulation, and demodulation, are performed in an integrated fashion. By leveraging the near-term predictions of some system parameters (e.g., traffic loads, content requests, and user trajectory), those approaches can serve the demands proactively to improve the performance, which also demonstrate good adaptability to the dynamic environments.

Recent research attempts at embedding DL models in the traditional algorithmic frameworks for combinatorial optimization also reveal promising results in advancing the state-of-the-art methodologies [4, 5], which shed light on a hybrid optimization framework for algorithm design. The former work focuses on a greedy heuristic framework in which the criterion for selecting the next step option is learned using DRL methods. The latter work investigates the local search framework, where the search direction is guided by the outputs of DNNs. Given the pervasiveness of heuristics in the wireless networking domain, the authors envision that the idea of leveraging DL to promote existing algorithms can be leveraged to address wireless network optimization tasks in a broader sense.

Despite the advantages of incorporating DL models in wireless network optimization, some fundamental research challenges have yet to be addressed to fully unleash the potential of DL technologies in simplifying network management and enhancing network performance. The authors vision some challenges and illustrate some of the research topics that deserve further considerations from four aspects: informative training data generation; effective training objective design; performance guarantees, and scalability.

The authors provide a survey with a holistic perspective, of the recent efforts in leveraging DL for wireless network optimization, probing insightfully where and how the supremacy of DL based approaches comes versus the conventional modeling-based approaches. The challenges of applying the state of the art from the machine learning community to general wireless network optimization problems are discussed and several promising research directions are pointed out.

References:

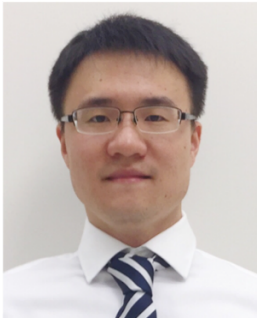
[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature*, vol. 521, no. 7553, 2015, pp. 436–44.

[2] L. Liu et al., "Deep Learning Meets Wireless Network Optimization: Identify Critical Links," *IEEE Trans. Network Science and Engineering*, vol. 7, no. 1, 2020, pp. 167–80.

[3] H. Ye et al., "Deep Learning-Based End-to-End Wireless Communication Systems with Conditional GANs as Unknown Channels," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, 2020, pp. 3133–43.

[4] E. Khalil et al., "Learning Combinatorial Optimization Algorithms over Graphs," *Advances in Neural Information Processing Systems*, 2017, pp. 6348–58.

[5] X. Chen and Y. Tian, "Learning to Perform Local Rewriting for Combinatorial Optimization," *Advances in Neural Information Processing Systems*, 2019, pp. 6278–89.



Shengjie Xu [SM'14-M'19] received a Ph.D. degree in Computer Engineering from University of Nebraska-Lincoln, and an M.S. degree in Telecommunications from University of Pittsburgh. Before that, he held a B.E. degree in Computer Science and Information Security. Presently, he is an Assistant Professor of Computer and Cyber Sciences in the Beacom College of Computer and Cyber Sciences at Dakota State University. His research interests include AI-driven cybersecurity, secure edge computing, and critical infrastructure protection. He serves as a Technical Editor for IEEE Wireless Communications and an Editor for International Journal of Sensor Networks. He was awarded the 2020 IET Journals Premium Award for Best Paper. He is a member of IEEE and ACM. He holds multiple professional certifications in cybersecurity and computer networking.

On the impact of additional Sensory Stimuli on the Quality of Experience of 360-degree videos

A short review for "Multisensory 360◦ Videos Under Varying Resolution Levels Enhance Presence"

Edited by Roberto G. de A. Azevedo

A. Covaci, E. B. Saleme, G. Mesfin, I-S. Comsa, R. Trestian, C. A. S. Santos, and G. Ghinea, "Multisensory 360◦ Videos Under Varying Resolution Levels Enhance Presence," IEEE Trans. on Visualization and Computer Graphics, 2022 (Early access) doi: 10.1109/TVCG.2022.3140875

Immersive media technologies have been gaining popularity in the last few years, and it is expected that they will drastically affect the way we consume multimedia content. Indeed, there has been a considerable amount of research in both industry and academy, resulting in new hardware (e.g., head-mounted displays, controllers, etc.) and software technologies that provide more and more immersive experiences. Technologies such as point clouds, light fields, and 360-degree videos are examples of visual media formats that can be used to provide a more immersive experience to end users.

Although traditionally most multimedia content has been developed to stimulate two human senses, i.e., sight and hearing, to provide a fully immersive experience, additional effects stimulating other human senses are also needed [1]. In such a context, Covaci et al. [2] bring a new study on how the overall user quality of experience (QoE) is impacted by such sensorial effects together with 360-degree video degradation. The topic of the paper is timely and similar studies are much needed towards a better understanding of end-user experiences and optimizing computing resources to provide immersive services with the best QoE possible.

Covaci et al.'s study focuses on understanding how the additional sensorial effects affect the sense of presence on 360-degree videos and how such effects can be used to mask video quality drops, e.g., in adaptive bitrate (ABR) scenarios. In particular, the authors focus on answering the following research questions: RQ.1) "what is the impact of sensorial effects on the sense of presence in 360-degree videos?"; RQ.2) "what is the impact of video degradation on the sense of presence in 360-degree video with sensorial effects?"; and RQ.3) "what is the impact of video motion levels on the sense of presence in 360-degree videos with sensorial effects?"

The stimuli for the participants were composed by 3 different 360-degree videos augmented with sensorial effects (Coffe Shop, Lavender field, and Rollercoaster). The setup included a Smartphone-based HMD, a scent device, and a wind-emitter device. Three independent variables ("Encoding Quality", "Motion Dynamism", and "Sensory Effects") and one dependent variable ("Sense of Presence") were part of the study. Sense of presence is measured using the SUS (Slatter-Usoh-Steed) questionnaire [3].

Similar to previous studies [4, 5], and answering RQ.1, Covaci et al.'s paper shows that the additional effects induce a higher level of presence when compared to a similar 360-degree visual quality level without the sensorial effects. Indeed, the three presence themes on SUS (Being there, Dominant reality, and Images or places) have higher values.

Even more interesting, related to RQ.2, the study also shows that there are cases in which increasing the video bitrate level does not necessarily lead to an improvement in the perceived sense of presence. This can be partially explained by the "masking effects" that additional sensorial content can induce in the users. Since sensory effects are basically metadata that can be sent cheaply, a better understanding of how such additional effects impact the overall user QoE might be used to optimize network resources in new scenarios in the future.

Finally, related to RQ.3, the paper discusses the impact of the visual content itself on the sense of

presence. Although the visual content characteristics (e.g., content dynamism) have an impact on how the users feel immersed, the authors argue that in a multi-sensorial environment, such variable cannot be considered independently. The initial results on such a direction are very interesting. However, since the study only includes 3 different contents, more research should be done in such a direction.

Overall, Covaci et al.'s paper discusses a timely topic and highlights the need for considering the user experience of multimedia content in a broader sense for resource allocation strategies. In particular, they also point out the need for adaptive bandwith-allocation protocols to consider user-centric measures (such as the sense of presence) in practice. Such user-centric approach is even more necessary when considering future immersive scenarios, in which, for instance, it is possible to (cheaply) provide additional sensory effects that can mask quality fluctuations on the main (network-hungry) visual media.

References:

[1] G. Ghinea, C. Timmerer, W. Lin, & S. R. Gulliver, Mulsemedia: State of the art, perspectives, and challenges. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 11(1s), 1-23.

[2] A. Covaci, E. B. Saleme, G. Mesfin, I-S. Comsa, R. Trestian, C. A. S. Santos, and G. Ghinea. "Multisensory 360◦ Videos Under Varying Resolution Levels Enhance Presence," in IEEE Trans. on Visualization and Computer Graphics, 2022 (early access), doi: 10.1109/TVCG.2022.3140875.

[3] M. Usoh, Martin, E. Catena, S. Arman, and M. Slater. "Using presence questionnaires in reality." Presence 9, no. 5 (2000): 497-503.

[4] Guedes, A. L., Azevedo, R. G. A., Frossard, P., Colcher, S., and Barbosa, S. D. J. "Subjective evaluation of 360-degree sensory experiences." In 2019 IEEE 21st International Workshop on Multimedia Signal Processing (MMSP) (pp. 1-6).

[5] A. Covaci, R. Trestian, E. B. Saleme, I.-S. Comsa, G. Assres, C. A. S. Santos, and G. Ghinea. "360 Mulsemedia: a way to improve subjective QoE in 360 videos." In Proceedings of the 27th ACM International Conference on Multimedia, pp. 2378-2386. 2019.

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

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.