2023 IEEE International Conference on Metaverse Computing, Networking and Applications (MetaCom) | 979-8-3503-3333-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/MetaCom57706.2023.00069

Visual Data Compression for Metaverse: Technology, Standard, and Challenges

Peilin Chen
City University of Hong Kong
Kowloon, Hong Kong SAR, China

Bolin Chen
City University of Hong Kong
Kowloon, Hong Kong SAR, China

Meng Wang
City University of Hong Kong
Kowloon, Hong Kong SAR, China

Shiqi Wang
City University of Hong Kong
Kowloon, Hong Kong SAR, China

Zhu Li
University of Missouri-KC
Kansas City, MO 64111, USA

***Abstract*—There has been an increasing consensus that the emerging metaverse world relies on high efficiency visual data compression techniques for effective data storage, ultra-low delay interaction and ubiquitous communication. This paper provides a comprehensive survey on the recent advances of the visual data compression technology and standard, which could play essential roles in building the metaverse. In particular, we will review the visual data compression methodologies and standards, and present our vision on how the interactive coding could more efficiently support the metaverse. Finally, we discuss the distinguished requirements and fundamental challenges on the data compression for metaverse, and envision the future technology trend.**

***Index Terms*—Metaverse, visual data compression, interactive communication**

I. Introduction

The term "Metaverse" was first coined in Neal Stephenson's science fiction *snow crash* [1], which utilizes the prefix "meta" (meaning more comprehensive) to depict the word "universe". It illustrates a virtual realm that is achieved with the duality of the physical world by providing immersive inter-operations among users through individual avatars and specifically designed hardwares. In recent years, the digitalization in omnidirectional aspects of the real world has been regarded as the future of the Internet, aiming to provide seamless access and realistic communication beyond physical limitations in various fields, such as medicine [2], education [3], industry [4], and entertainment [5]. To achieve such an omnidirectional digitalization with promising user experience, the metaverse relies heavily on fundamental visual data compression techniques to support versatile communication and effective storage. More specifically, typical scenarios in communications include the interactions between humans and the virtual world, or between different virtual objects from one or multiple virtual worlds. Human beings can even exchange and interact through the virtual world, augmenting more immersive experience with reality. Moreover, the data compression technologies can also enable the recording of the virtual world history, such that any moment in the metaverse can be recovered in the future.

To empower the visual-centered applications in the metaverse and cultivate its ecosystem, three-dimensional (3-D) visual representations, such as point cloud [6], mesh [7], 360 degree virtual reality (VR) video [8], and Neural Radiance Fields [9], have been recognized to be more promising beyond the traditional 2D image in vividly rendering the virtual world. Considering the application scope of metaverse is far beyond conventional visual communication, there is no doubt that the application-specific visual data compression methods are highly desired. More specifically, there is an increasing consensus that the real-time interaction among human avatars imposes stringent challenges on the ultra-low delay compression, and the storage requirement typically demands ultra high efficiency compression technologies. As such, new challenges arising from the interaction and synchronization between the metaverse and the physical world, or within the metaverse, will certainly impact the new development of visual data compression methods.

This paper provides the review on visual data compression for metaverse, with the goals of surveying a wide spectrum of existing technologies and standards. It has been exciting to observe that the theoretical development and novel techniques on visual data coding, could support a wide variety of applications in metaverse. With the quick evolution of the compression methodologies in the era of Artificial Intelligence (AI), we also discuss how the AI powered compression technologies can be made best use of in these applications. Finally, we envision both the promise and challenges, when adopting these state-of-the-art compression schemes in the coming metaverse.

II. Visual data compression technology

Developing algorithms that can effectively compress the visual signals is now an intense focus of research. Considering the application scenarios of the metaverse, we introduce a wide variety of inventive compression schemes that operate under different application domains, to clarify how visual data can be effectively represented to pragmatically build the metaverse communication and storage systems.

A. Hybrid Video Coding

Video compression is typically characterized by the throughput of the channel and the distortion of reconstructed

979-8-3503-3333-6/23/$31.00 ©2023 IEEE
DOI 10.1109/MetaCom57706.2023.00069

videos. The hybrid coding framework, which has been widely adopted in dominated video coding standards, play the key role in improving the video coding performance. The state-of-the-art video coding standards including VVC [10] and AVS3 [11] which are dedicated to 4K/8K video compression, could deliver the ever best coding performance at the expense of high computational complexity. On top of VVC, the Enhanced Compression Model (ECM) was also developed towards the next generation video coding standard. These sophisticated designed video compression algorithms aim to remove redundancies, including spatial, temporal, statistical, and perceptual redundancies in video sequences. In particular, the predictive video coding framework adopts advanced intra and inter prediction techniques [12], such that spatial and temporal redundancies are removed, and only residuals are coded. Subsequently, the residuals are transformed, quantized, and entropy-coded [13], and encoded residuals with the side information are finally conveyed in the compact bitstream. Various rate-distortion optimization technologies [14] have been designed to ensure optimal coding performance, and deep learning techniques have also been incorporated into the encoder/decoder, to improve the performance of the hybrid coding framework [15], [16]. One dominate type of visual data in Metaverse could be the synthetic data rendered by computers. Recently, screen content coding has attracted numerous attentions and the associated coding technologies such as intra block copy prediction, palette prediction, and transform skip have been investigated and adopted in the extension of HEVC [17] and VVC standards [18]. The optimization of screen content encoder has also received attentions, by replacing the distortion measure or optimizing the bit allocation [19], [20]. Although the elaborately designed hybrid coding framework has achieved remarkable progress in traditional 2D video compression, how to properly incorporate them into metaverse communication to support a vast assortment of applications is still non-trivial.

B. High Dimensional Data Compression

High-dimensional visual data representations that depict the world from multiple views are critical in providing pleasant immersive experiences. In recent years, diverse high-dimensional visual data types, such as multiview video [21], point cloud [22], light field [23], and 360-degree virtual reality (VR) video [8], have been developed beyond the traditional 2D frames to vividly render the virtual environment. However, the bottleneck of utilizing such high-dimensional visual data lies in how to compactly represent them in an effective way.

One straightforward way for compressing these data is to reshape them into 2D-frame-like video sequences, such that we could resort to the mature video coding schemes. Given the decoded video sequence, the inverse operation for projecting back to high-dimension data is subsequently performed at the decoder. For example, the light field data could be decomposed according to view directions and subsequently re-arranged into an inter-correlated video sequence which can be further encoded by VVC. The Video-based Point Cloud Compression (V-PCC) scheme [24], [25] suggests projecting the 3-D point cloud into different 2-D maps, and 2-D video codecs can be utilized to faithfully compress them. The 2-D to 3-D remapping is conducted at the V-PCC decoder [26]. Such a geometry conversion is also an essential part during the compression of 360-degree video in the VVC standard. Another line in compressing the high-dimensional data is exploiting the geometry characteristics. Typically, the Geometry-based Point Cloud Compression (G-PCC) scheme [27] uses the pruned octree form to efficiently approximate the original data [28]. Recently, inspired by the fantastic progress of deep learning, the end-to-end high-dimensional visual data coding framework has also been widely studied with favorable performance achieved [29]–[31].

C. Model-based Coding

Model-based coding (MBC) leverages the analysis-synthesis techniques which exploit the intrinsic statistics and structural information, showing the very promising possibility for low bitrate visual communication. More specifically, hand-crafted models have been employed to economically describe structural representations at the encoder side, whilst the compact feature representation can further facilitate the reconstruction of target videos via the synthesis models at the decoder side. The earliest work of MBC could be dated back to 1950's when Schreiber *et al.* [32] designed a complete TV bandwidth reduction system, which can compactly project the video content into the edge information. The decoded edge can be further used to synthesize the original video signal. From then on, the MBC technique has been developed from the perspective of different granularity levels, including pixel-based coding, block-based coding, mesh-based coding, region-based coding, object-based coding, knowledge-based coding and perceptual coding. Early MBC techniques mainly adopted traditional pulse code modulation model [33], block based translation model [34], [35] and triangle mesh model [36], [37] to actualize pixel/block/mesh based video coding. Besides, region-based coding [38], [39] also played an important role in the following evolution of MBC. Moreover, object-based coding [40]–[42] belonging to a further prolongation of region-based coding could describe an identified object in a more precise way, achieving better rate-distortion performance. Regarding knowledge-based coding [43]–[45], the specific object, especially for talking face or moving body, enjoyed the strong priors via parameterized models. Perceptual coding [46]–[48], which is proposed based on the characteristics of human visual system, attempted to achieve significant compression gains by incorporating perception models into the codecs. However, the video reconstruction quality of MBC techniques are govern by these conventional analysis and synthesis models, thus hindering their practical applications.

D. Learning-based Visual Coding

The deep neural networks, including convolutional neural network (CNN) [49], variational autoencoder (VAE) [50] and generative adversarial network (GAN) [51], have greatly

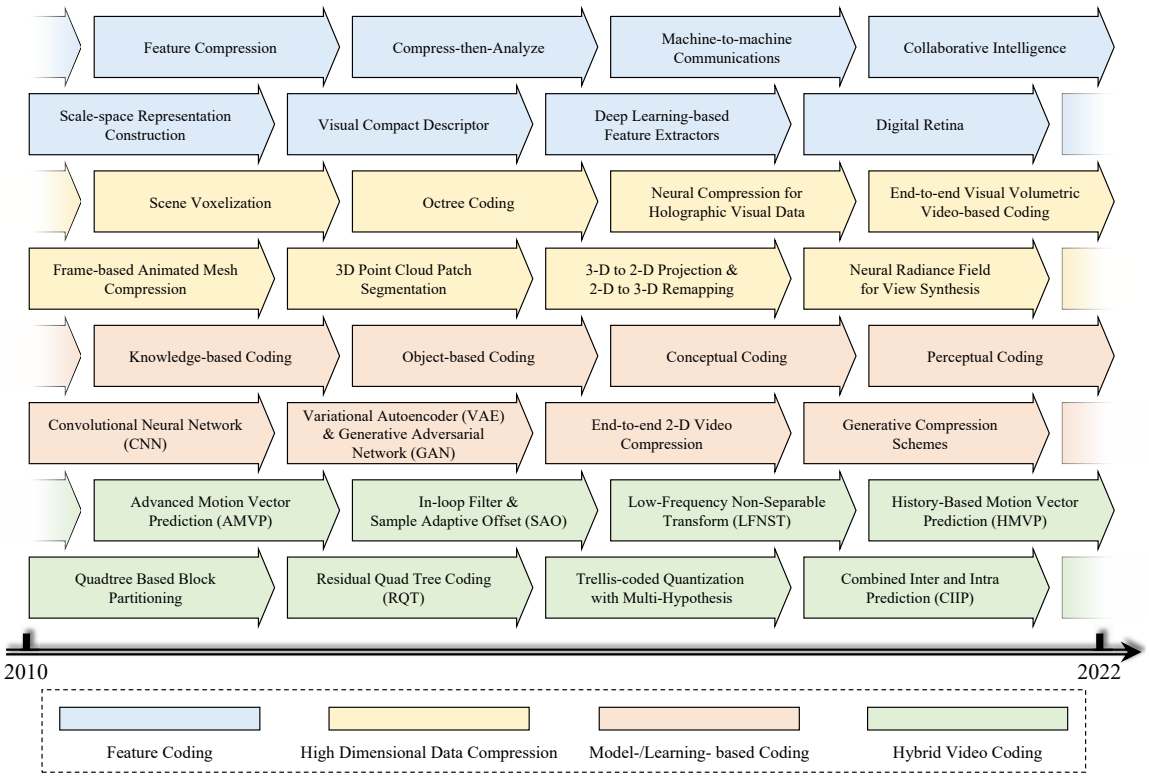

Fig. 1. The roadmap of visual data compression technologies over the past decade.

advanced the paradigm of visual compression, such that the visual data in metaverse can be promisingly characterized into a latent code, the abstract visual concepts (*e.g.,* structure) or semantic representations (*e.g.,* landmark or keypoint). Along this vein, the high-quality visual communication under ultra-low bandwidth can be actualized in the metaverse. Accordingly, the learning-based visual coding techniques can be categorized into end-to-end compression, conceptual coding and generative compression. End-to-end compression [52]–[54] aims to utilize the deep learning models to jointly optimize entire encoder/decoder by considering the rate-distortion trade-off, delivering the promising perceptual rate-distortion performance. Regarding conceptual coding [55], [56], a novel analysis-friendly compression paradigm is able to encode images into highly-abstract and fully-interpretable concepts, and high-quality images are reconstructed in a deep synthesis manner to support various vision tasks. Recently, the latest generative compression schemes [57], [58] have also achieved the competitive rate-distortion performance. Unlike conceptual coding algorithms, generative compression schemes directly transform the images into very compact representations that can be used to reconstruct images with generative models. However, these learning-based visual coding techniques are still faced with the unprecedented challenges, including how to quantify the visual quality of the compressed visual data from a technically sound way in the specific application scenarios, and how these quality measures can be used to optimize the entire coding framework.

E. Feature Coding

The end-edge-cloud computing paradigm of the metaverse pushes the intelligent computing capabilities from the cloud to the network edges and end users, such that the decentralized, collaborative and intelligent communication paradigm inspires new visual coding schemes. The virtual entities of the metaverse leverage the machine intelligence to understand the visual information. Therefore, feature compression, which naturally accommodates the end-edge-cloud computing, has received great attentions. The remarkable approaches, referred to as Analyze-then-Compress (ATC) [59] and digital retina [60], lay the foundation for compact feature representation. The standards of Compact Descriptors for Visual Search (CDVS) [61] and Compact Descriptors for Video Analysis (CDVA) [62] have also been finalized to facilitate visual retrieval and analytics. With the unprecedented success of deep learning-based feature extractors, the compact representation of deep learning features has also been widely studied. The local and global features can both be compressed based on the philosophy of hybrid video coding [63]. The lossy compression of intermediate features was proposed for intelligent sensing [64] and collaborative intelligence [65].

In particular, the bit allocation based on Pareto optimization has been proposed for multi-stream feature communication systems [66]. Moreover, a lightweight compression algorithm for the intermediate feature is designed for splitting neural networks [67]. Targeting at video compression toward machine vision, the exploration of video coding for machines (VCM) [68] was also with a series of works focusing on feature coding [69], [70].

III. Challenges & Research Directions

Remarkable progress has been made in the field of visual data compression in the past decades, evidenced by these advanced compression algorithms which achieve very promising compression performance on natural images/videos/high dimensional data. Nevertheless, this does not necessarily mean that the visual data compression research for metaverse has reached a sufficient level of maturity, especially when considering the arising new application scenarios and challenges from metaverse.

First, metaverse could be greatly benefited from the traditional cloud gaming infrastructure, where the visual data could be rendered from the cloud and transmitted to the client side for decoding [71]. Herein, the compression performance becomes the bottleneck, as it is always desirable to convey the visual data with high quality and low bitrate consumption, to accommodate the dynamically varying bandwidth. The research on joint rendering and compression in the cloud-edge-end infrastructure enables the collaboraitve rendering at cloud, edge and thin client sides. In particular, the rendering tasks can be coherently scheduled, to jointly optimize the rate-distortion-delay-complexity. This direction is still in preliminary stages, and there is a large space to explore in the future.

Second, the interactive coding that is featured with ultra low delay and high compression performance is vital for metaverse. To this end, it is highly expected that the visual data could be analyzed in a scientifically sound way, and represented in a style that is friendly for interactive coding [72]. The alternative but promising solutions are conceptual coding and semantic coding, which could be naturally supportive for the interactivity with promising performance due to the representation capability with the deep neural networks. Though existing research results have shown the promise, these methods often suffer from generalization problem.

Third, in most of the compression algorithms and standards discussed, the inputs are natural visual data. Besides these, the compression algorithms can be extended to an even broader range of applications. The compression of AI generated content (AIGC), which are becoming increasingly important, has not received sufficient attentions. Other types of signals, such as spiking signals, could play prominent roles in facilitating the brain-computer interface. However, the spike signal compression is still in the infancy age. The new challenges arising from the metaverse could also impact the new development of the compression algorithms.

IV. Conclusions and Future remarks

In this paper, we have conducted a thorough review of visual data compression techniques for use in the metaverse. Rather than covering all existing algorithms and standards, we have focused on those that have the most potential to be applied in the metaverse and highlighted the main challenges that should be addressed in the future. It is clear that the development of new applications and requirements in the metaverse will drive advances in compression technology, which will also shape the future of the metaverse.

Acknowledgments

This work was supported in part by the Shenzhen Science and Technology Program (Project No. JCYJ20220530140816037), in part by the National Natural Science Foundation of China under 62022002, and in part by the Hong Kong GRF-RGC General Research Fund under Grant 11203220 (CityU 9042957). This work was also supported in part by NSF-2148382.

References

[1] N. Stephenson, *Snow crash: A novel*. Spectra, 2003.
[2] R. Chengoden et al., "Metaverse for healthcare: A survey on potential applications, challenges and future directions," *arXiv preprint arXiv:2209.04160*, 2022.
[3] B. Kye et al, "Educational applications of metaverse: possibilities and limitations," *J. Educ. evaluation health Prof.*, vol. 18, 2021.
[4] Y. Hu and H. Chen, "The trend of industrial design from the perspective of metaverse," in *HCII*. Springer, 2022, pp. 397–406.
[5] X. Niu and W. Feng, "Immersive entertainment environments-from theme parks to metaverse," in *HCII*. Springer, 2022, pp. 392–403.
[6] S. Schwarz et al., "Emerging mpeg standards for point cloud compression," *IEEE J. Emerg. Sel. Top. Circuits Syst.*, vol. 9, no. 1, pp. 133–148, 2018.
[7] A. Maglo, G. Lavoué, F. Dupont, and C. Hudelot, "3d mesh compression: Survey, comparisons, and emerging trends," *ACM Computing Surveys (CSUR)*, vol. 47, no. 3, pp. 1–41, 2015.
[8] Y. Zhou, L. Tian, C. Zhu, X. Jin, and Y. Sun, "Video coding optimization for virtual reality 360-degree source," *IEEE J. Sel. Top. Signal Process.*, vol. 14, no. 1, pp. 118–129, 2019.
[9] B. Mildenhall et al., "Nerf: Representing scenes as neural radiance fields for view synthesis," *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.
[10] B. Bross et al, "Overview of the versatile video coding (VVC) standard and its applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 10, pp. 3736–3764, 2021.
[11] J. Zhang et al, "Recent development of AVS video coding standard: AVS3," in *2019 IEEE PCS*, pp. 1–5.
[12] M. Lei et al., "Look-ahead prediction based coding unit size pruning for VVC intra coding," in *2019 IEEE ICIP*, pp. 4120–4124.
[13] Y. Fan, Y. Zeng, H. Sun, J. Katto, and X. Zeng, "A pipelined 2d transform architecture supporting mixed block sizes for the VVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, pp. 1–1, 2019.
[14] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "SSIM-motivated rate-distortion optimization for video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 4, pp. 516–529, April 2012.
[15] Y. Zhang, T. Shen, X. Ji, Y. Zhang, R. Xiong, and Q. Dai, "Residual highway convolutional neural networks for in-loop filtering in HEVC," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 3827–3841, Aug 2018.
[16] N. Yan, D. Liu, H. Li, B. Li, L. Li, and F. Wu, "Convolutional neural network-based fractional-pixel motion compensation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 3, pp. 840–853, March 2019.
[17] D. Flynn et al, "Overview of the range extensions for the HEVC standard: Tools, profiles, and performance," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 4–19, 2016.

[18] T. Nguyen et al, "Overview of the screen content support in VVC: Applications, coding tools, and performance," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 10, pp. 3801–3817, 2021.

[19] S. Wang et al, "Utility-driven adaptive preprocessing for screen content video compression," *IEEE Trans. Multimed.*, vol. 19, no. 3, pp. 660–667, 2016.

[20] S. Wang, K. Gu, K. Zeng, Z. Wang, and W. Lin, "Objective quality assessment and perceptual compression of screen content images," *IEEE computer graphics and applications*, vol. 38, no. 1, pp. 47–58, 2016.

[21] T. Hussain et al, "A comprehensive survey of multi-view video summarization," *Pattern Recognition*, vol. 109, p. 107567, 2021.

[22] M. Berger et al, "A survey of surface reconstruction from point clouds," in *Computer Graphics Forum*, vol. 36, no. 1, 2017, pp. 301–329.

[23] L. Li and Z. Li, "Light field and plenoptic point cloud compression," in *Handbook of Dynamic Data Driven Applications Systems*. Springer, 2022, pp. 557–583.

[24] L. Li et al, "Efficient projected frame padding for video-based point cloud compression," *IEEE Trans. Multimed.*, vol. 23, pp. 2806–2819, 2020.

[25] L. Li, Z. Li, V. Zakharchenko, J. Chen, and H. Li, "Advanced 3d motion prediction for video-based dynamic point cloud compression," *IEEE Trans. Image Process.*, vol. 29, pp. 289–302, 2019.

[26] A. Akhtar et al, "Video-based point cloud compression artifact removal," *IEEE Trans. Multimed.*, 2021.

[27] D. Graziosi, O. Nakagami, S. Kuma, A. Zaghetto, T. Suzuki, and A. Tabatabai, "An overview of ongoing point cloud compression standardization activities: Video-based (V-PCC) and geometry-based (G-PCC)," *APSIPA Transactions on Signal and Information Processing*, vol. 9, 2020.

[28] K. Ainala et al, "An improved enhancement layer for octree based point cloud compression with plane projection approximation," in *Appl. digit. image process. XXXIX*, vol. 9971. SPIE, 2016, pp. 223–231.

[29] X. He, Q. Liu, and Y. Yang, "Mv-gnn: Multi-view graph neural network for compression artifacts reduction," *IEEE Trans. Image Process.*, vol. 29, pp. 6829–6840, 2020.

[30] J. Wang, D. Ding, Z. Li, and Z. Ma, "Multiscale point cloud geometry compression," in *2021 IEEE DCC*, pp. 73–82.

[31] A. Akhtar, Z. Li, and G. Van der Auwera, "Inter-frame compression for dynamic point cloud geometry coding," *arXiv preprint arXiv:2207.12554*, 2022.

[32] W. F. Schreiber, C. F. Knapp, and N. D. Kay, "Synthetic highs — an experimental tv bandwidth reduction system," *Journal of the SMPTE*, vol. 68, no. 8, pp. 525–537, 1959.

[33] R. L. Carbrey, "Video transmission over telephone cable pairs by pulse code modulation," *Proceedings of the IRE*, vol. 48, no. 9, pp. 1546–1561, 1960.

[34] I. Dinstein, K. Rose, and A. Heiman, "Variable block-size transform image coder," *IEEE Transactions on Communications*, vol. 38, no. 11, pp. 2073–2078, 1990.

[35] V. E. Seferidis and M. Ghanbari, "General approach to block-matching motion estimation," *Opt. Eng.*, vol. 32, no. 7, pp. 1464 – 1474, 1993.

[36] D. Kubasov and C. Guillemot, "Mesh-based motion-compensated interpolation for side information extraction in distributed video coding," in *2006 ICIP*, pp. 261–264.

[37] G. J. Sullivan and R. Baker, "Motion compensation for video compression using control grid interpolation," in *1991 IEEE ICASSP*, pp. 2713–2716.

[38] D. Graham, "Image transmission by two-dimensional contour coding," *Proceedings of the IEEE*, vol. 55, no. 3, pp. 336–346, 1967.

[39] M. Biggar, O. Morris, and A. Constantinides, "Segmented-image coding: performance comparison with the discrete cosine transform," *Radar and Signal Processing, IEE Proceedings F*, vol. 135, pp. 121 – 132, 05 1988.

[40] M. Hans Georg et al, "Object-oriented analysis-synthesis coding of moving images," *Signal Process.: Image Commun.*, vol. 1, no. 2, pp. 117–138, 1989.

[41] P. Salembier et al, "Segmentation-based video coding system allowing the manipulation of objects," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 1, pp. 60–74, 1997.

[42] A. Vetro, T. Haga, K. Sumi, and H. Sun, "Object-based coding for long-term archive of surveillance video," in *2003 ICME*, vol. 2, pp. II–417.

[43] R. Lopez and T. Huang, "Head pose computation for very low bit-rate video coding," in *CAIP*, 1995, pp. 440–447.

[44] Parke, "Parameterized models for facial animation," *IEEE Computer Graphics and Applications*, vol. 2, no. 9, pp. 61–68, 1982.

[45] K. Aizawa, H. Harashima, and T. Saito, "Model-based analysis synthesis image coding (mbasic) system for a person's face," *Signal Processing: Image Communication*, vol. 1, no. 2, pp. 139–152, 1989.

[46] N. Jayant, J. Johnston, and R. Safranek, "Signal compression based on models of human perception," *Proceedings of the IEEE*, vol. 81, no. 10, pp. 1385–1422, 1993.

[47] R. Safranek and J. Johnston, "A perceptually tuned sub-band image coder with image dependent quantization and post-quantization data compression," in *ICASSP*, 1989, pp. 1945–1948 vol.3.

[48] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "Perceptual video coding based on ssim-inspired divisive normalization," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1418–1429, 2012.

[49] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE CVPR*, 2016, pp. 770–778.

[50] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *Proceedings of the 2nd ICLR*, 2014, p. 14.

[51] I. Goodfellow et al, "Generative adversarial nets," vol. 27, 2014.

[52] D. Minnen, J. Ballé, and G. D. Toderici, "Joint autoregressive and hierarchical priors for learned image compression," in *NeurIPS*, vol. 31, 2018.

[53] G. Lu, W. Ouyang, D. Xu, X. Zhang, C. Cai, and Z. Gao, "DVC: An end-to-end deep video compression framework," in *Proceedings of IEEE/CVF CVPR*, 2019.

[54] R. Yang, F. Mentzer, L. Van Gool, and R. Timofte, "Learning for video compression with recurrent auto-encoder and recurrent probability model," *IEEE J. Sel. Top. Signal Process*, vol. 15, no. 2, pp. 388–401, 2020.

[55] J. Chang et al, "Conceptual compression via deep structure and texture synthesis," *IEEE Trans. Image Process.*, vol. 31, pp. 2809–2823, 2022.

[56] ——, "Thousand to one: Semantic prior modeling for conceptual coding," in *2021 IEEE ICME*, pp. 1–6.

[57] B. Chen et al, "Beyond keypoint coding: Temporal evolution inference with compact feature representation for talking face video compression," in *2022 IEEE DCC*, pp. 13–22.

[58] G. Konuko et al, "H-DAC: Hybrid coding with deep animation models for ultra-low bitrate video conferencing," in *IEEE ICIP*, 2022.

[59] A. Redondi et al, "Compress-then-analyze versus analyze-then-compress: What is best in visual sensor networks?" *IEEE Transactions on Mobile Computing*, vol. 15, no. 12, pp. 3000–3013, 2016.

[60] W. Gao et al, "Digital retina: A way to make the city brain more efficient by visual coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 11, pp. 4147–4161, 2021.

[61] L.-Y. Duan et al, "Overview of the MPEG-CDVS standard," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 179–194, 2015.

[62] ——, "Compact Descriptors for Video Analysis: The emerging MPEG standard," *IEEE MultiMedia*, vol. 26, no. 2, pp. 44–54, 2019.

[63] L. Ding, Y. Tian, H. Fan, C. Chen, and T. Huang, "Joint coding of local and global deep features in videos for visual search," *IEEE Transactions on Image Processing*, vol. 29, pp. 3734–3749, 2020.

[64] Z. Chen, K. Fan, S. Wang, L. Duan, W. Lin, and A. C. Kot, "Toward intelligent sensing: Intermediate deep feature compression," *IEEE Trans. Image Process.*, vol. 29, pp. 2230–2243, 2019.

[65] I. V. Bajić, W. Lin, and Y. Tian, "Collaborative intelligence: Challenges and opportunities," in *2021 IEEE ICASSP*, pp. 8493–8497.

[66] S. R. Alvar et al, "Pareto-optimal bit allocation for collaborative intelligence," *IEEE Trans. Image Process.*, vol. 30, pp. 3348–3361, 2021.

[67] R. A. Cohen, H. Choi, and I. V. Bajić, "Lightweight compression of intermediate neural network features for collaborative intelligence," *IEEE Open Journal of Circuits and Systems*, vol. 2, pp. 350–362, 2021.

[68] L. Duan, J. Liu, W. Yang, T. Huang, and W. Gao, "Video coding for machines: A paradigm of collaborative compression and intelligent analytics," *IEEE Trans. Image Process.*, vol. 29, pp. 8680–8695, 2020.

[69] N. Le et al, "Image coding for machines: an end-to-end learned approach," in *2021 IEEE ICASSP*, pp. 1590–1594.

[70] S. Wang et al, "Towards analysis-friendly face representation with scalable feature and texture compression," *IEEE Trans. Multimed.*, vol. 24, pp. 3169–3181, 2021.

[71] S.-P. Chuah and N.-M. Cheung, "Bandwidth efficient mobile cloud gaming with layered coding and scalable phong lighting," in *2014 IEEE ICIP*, pp. 6006–6010.

[72] B. Chen et al, "Interactive face video coding: A generative compression framework," *arXiv preprint arXiv:2302.09919*, 2023.