Special Issue on Open Media Compression: Overview, Design Criteria, and Outlook on Emerging Standards

By CHRISTIAN TIMMERER, *Senior Member IEEE*
Guest Editor
MATHIAS WIEN, *Member IEEE*
Guest Editor
LU YU, *Senior Member IEEE*
Guest Editor
AMY REIBMAN, *Fellow IEEE*
Guest Editor

Universal access to and provisioning of multimedia content is now a reality. It is easy to generate, distribute, share, and consume any multimedia content, anywhere, anytime, or any device. Open media standards took a crucial role toward enabling all these use cases leading to a plethora of applications and services that have now become a commodity in our daily life. Interestingly, most of these services adopt a streaming paradigm, are typically deployed over the open, unmanaged Internet, and account for most of today's Internet traffic. Currently, the global video traffic is greater than 60% of all Internet traffic [1], and it is expected that this share will grow to more than 80% in the near future [2]. In addition, Nielsen's law of Internet bandwidth states that the users' bandwidth grows by 50% per year, which roughly fits data from 1983 to 2019 [3]. Thus, the users' bandwidth can be expected to reach approximately 1 Gb/s by 2022. At the same time, network applications will grow and utilize the bandwidth provided, just like programs and their data expand to fill the memory available in a computer system. Most of the available bandwidth today is consumed by video applications, and the amount of data is further increasing due to already established and emerging applications, e.g., ultrahigh definition, high dynamic range, or virtual, augmented, mixed realities, or immersive media applications in general.

This month's special issue reviews multimedia compression and system standards, with a focus on the technology developed in the context of these standards and addresses open research questions in the field.

In order to cope with these ever-demanding applications and services, efficient, open media compression and system standards are inevitable. Such standards provide interoperability that is a key factor, but it is a thin line as standards aim at specifying only the minimum required for interoperability to foster innovation through competition among vendors. Open media standards mostly define the *bitstream format* without specifying: 1) how to create (i.e., produce, encode, provision, and so on) such standardized formats or/and

Digital Object Identifier 10.1109/JPROC.2021.3098048

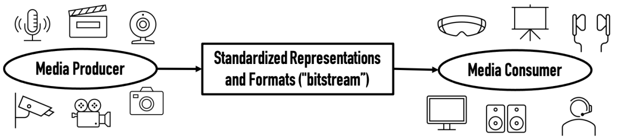


Fig. 1. ***Role of open media standards in the context of R&D efforts.***

2) how to make use of such standardized formats (i.e., decode, process, and so on). These aspects are deliberately outside the scope of standards and left open for competition (cf. Fig. 1). Thus, open media standards can be considered both *source* and *sink* of research and development (R&D) efforts, i.e., R&D can lead to new standards being created, or/and standards can also trigger new R&D efforts.

Open media standards are in the scope of this special issue, specifically, those specifying *data representations and formats* providing basic tools to be adopted directly by industry or referenced by other standards developing organizations (SDOs) (or parts thereof), i.e., Joint Photographic Experts Group (JPEG),[1] and Moving Picture Experts Group (MPEG),[2] [4]. JPEG—formally known as the International Organization for Standardization and the International Electrotechnical Commission (ISO/IEC) Joint Technical Committee 1 (JTC 1)/Subcommittee 29 (SC 29)/Working Group 1 (WG 1)—defines standards in the area of still picture coding, while MPEG—nowadays, a well-defined set of working groups (WGs) under ISO/IEC JTC 1/SC 29—defines standards in the area of moving picture coding and audio coding. Both SDOs develop also system standards for their respective media formats, e.g., file and delivery/streaming formats.

The remainder of this article is structured as follows. Section I reviews and discusses the term *open standards* in the context of this special issue. Section II describes design criteria addressing the computational complexity of open media standards. Section III provides an outlook that briefly highlights emerging standards. Finally, Section IV provides an overview of the articles within this special issue. Section V concludes this article.

I. OPEN STANDARDS

A. General

Standards are the premiere requisite to enable any type of communication or distribution of information among users, devices, and across borders. They represent the agreement of the participating industry on technical requirements and capabilities of any devices that are supposed to participate in or contribute to the service, which is to be established or maintained. A standard strives to provide a complete and unambiguous specification of all aspects of the standardized process. Thereby, it enables the industry to create interoperable devices and applications. At the same time, the level of accessibility and the availability of the standard can be a means of policy or market control. For example, standards may be agreed upon in an industry consortium at a national or international level in order to establish or protect an ecosystem among the consortium partners, and access from outside is restricted through privacy and licensing regulations. Even schemes or concepts developed by a single company may be established as *de facto* standards if the company manages to achieve broad adoption in the respective market. While such standards may be broadly established, they might typically not be classified as "open standards" since control over their use and contents is with a—potentially very—limited group of stakeholders. The aforementioned examples may rather be characterized as "closed standards," as control over the supported features of the standard and control over licensing conditions are in the hand of those stakeholders.

In order to be classified as an open standard, the group developing the standard should be open to interested participants who may bring in their requirements on features and capabilities supported by the standard and solutions enabling the standard to meet the defined goals. This is achieved, e.g., at the level of international standards setting organizations (SSOs) or SDOs, such as the ITU Telecommunication Standardization Sector (ITU-T) or ISO/IEC. These organizations (as well as others) have been established for the purpose of generating and managing standards that, thereby, are also referred to as *de jure* standards [5]. Interested industry and academia may join these organizations and contribute to standards development. Membership in such organizations may not come for free and may be subject to agreement membership conditions. However, access is not precluded by other means.

Another level of openness of a standard is access to the approved standards. For example, the recommendations of ITU-T can generally be downloaded from the websites of the organizations without any charge. ISO/IEC generally charges a fee for granting access to their standards. Independent of the form of accessibility the application and implementation of a standard may imply licensing costs, which is a topic out of the scope of the technical discussion but a topic of awareness among users and a potential key factor for the adoption of standards. Standardization organizations, such as ITU-T or ISO/IEC, do not interfere with the actual definition of licensing models. They strive to achieve the definition of standards that are available to industry under reasonable and nondiscriminatory terms (RANDs) and actually preclude publication of standards against which patents are claimed under non-RAND conditions.

---

[1] Accessed: April 16, 2021. [Online] Available: https://jpeg.org/

[2] Accessed: April 16, 2021. [Online] Available: http://mpeg.org/

With the foundation of the Alliance for Open Media (AOM) in 2015,[3] an industry consortium has been built which strives to generate open standards with the additional goal of providing these standards on a "royalty-free" basis. The first representative of such standards delivered by AOM is AV1, which is also discussed in this special issue. As a royalty-free standard, no licensing cost is supposed to be charged from implementers or users of this standard. While potential licensing fees for standards released by SDOs may be interpreted as compensation for the investment of contributors who were developing the standardized technology, the revenue for the investment of contributing to a new standard in this model is supposed to come in an indirect fashion.

In order to enforce a general definition of "Open Standards," the ITU-T provides a definition of the term as follows.

> "Open Standards" are standards made available to the general public and are developed (or approved) and maintained via a collaborative and consensus-driven process.
> "Open Standards" facilitate interoperability and data exchange among different products or services and are intended for widespread adoption.[4]

B. Example: Video Coding Standards

In the field of video coding, the application of a collaborative and consensus-driven process is realized on a cross-SDO basis between ITU-T and ISO/IEC. These SDOs have merged their standardization activities in the field of video coding to a very large extent for about the last 20 years by forming collaborative standardization groups, such as the Joint Video Team (JVT), the Joint Collaborative Team on Video Coding (JCT-VC), and the Joint Video Experts Team (JVET), which produced the video coding standards Advanced Video Coding (AVC, ITU-T H.264, and ISO/IEC 14496-10) [6], High Efficiency Video Coding (HEVC, ITU-T H.265, and ISO/IEC 23008-2) [7], and Versatile Video Coding (VVC, ITU-T H.266, and ISO/IEC 23090-3) [8], respectively. The specifications resulting from this collaborative work were published as so-called *twin text* by both ITU-T and ISO/IEC. Twin-text specifications can be published in a somewhat asynchronous fashion and share the common text base [9]. They may further include—to a limited extent—nonsynchronized additions. For these standardization projects, the input contributions and the reference software implementations have been made open access and can be reviewed by interested parties. Other recent ISO/IEC video coding standards comprise Essential Video Coding (EVC) [10], [11] and Low Complexity Enhancement Video Coding (LCEVC) [12]. These have been developed in MPEG as a parallel activity to JVET.

It has to be noted that technological developments in the field of media compression are strongly influenced and shaped by the output of standardization groups, such as the joint teams mentioned above. In the field of video compression, many authors rely on the test models of HEVC or VVC as the state of the art to compare with, and the encoder configurations defined by the common testing conditions (CTCs) of the standardization groups, such as JCT-VC or JVET, mark the common reference point for many research activities. In fact, it seems that, in many cases, authors to some extent confuse the reference software and the encoder configuration settings provided with the CTCs with the standards themselves. It should be noted that both are maintained for the purpose of standards development and maintenance. In real applications, encoder implementations may follow very different optimization criteria and also may make use of very different encoder configuration settings than those provided with the CTC.

II. SPECIAL TOPIC: COMPUTATIONAL COMPLEXITY

In the original setup of this special issue, the guest editors intended to include an article on the important and overarching topic of computational and implementation complexity in both, software and hardware. It turned out, however, that it was not possible to find experts from the relevant industry to write a public article on this topic.

In this section, we highlight some complexity aspects with a focus on the design of video coding standards, mostly using the most recent VVC standard as the example [8]. VVC has been chosen as all input and output documents of the group in charge, JVET, are publicly available.[5] The description here stays at a conceptual level. For details on VVC, the reader is referred to the VVC standard [8] or/and the VVC overview article provided in this issue.

The complexity of video encoding and decoding algorithms has a direct impact on the usability of the coding scheme. Specifically, for the decoder, since consumption of video is ubiquitous on TV sets, computers, and—power critical—mobile devices, a hardware-friendly design leverages the use of the scheme. This becomes even more evident by the increasing resolutions that decoders have to cope with. Given that modern mobile devices often have displays with ultrahigh-definition (UHD) resolutions of 3840×2160 and more, the capability of real-time UHD decoding on battery-driven devices becomes essential. At the same time, encoder implementations are required that support real-time encoding of such resolutions, e.g., in order to support usage of the video coding scheme in a live-broadcasting environment.

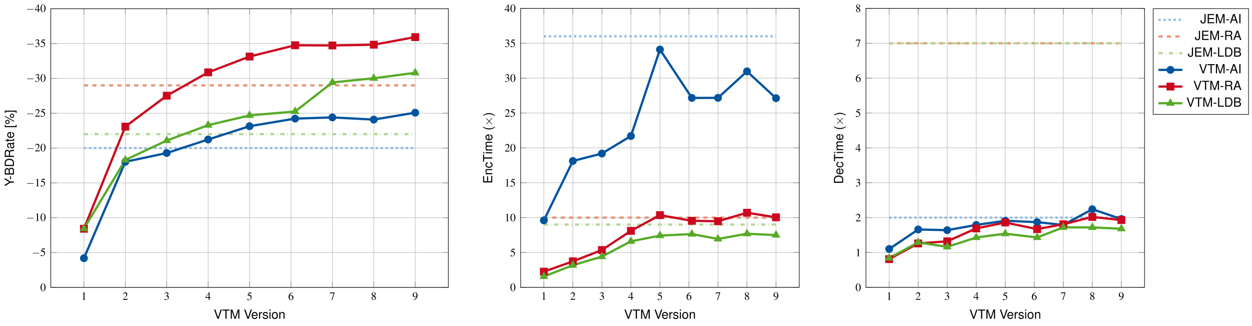
In the following, aspects of concern are presented and approaches of

---

[3] Accessed: April 16, 2021. [Online] Available: https://aomedia.org/about/

[4] Accessed: April 16, 2021. [Online] Available: https://www.itu.int/en/ITU-T/ipr/Pages/open.aspx

[5] Accessed: April 16, 2021. [Online] Available: https://jvet-experts.org/



(a) (b) (c)

Fig. 2. *Evolution of the performance of the VTM over its versions for the all intra (AI), random access (RA), and low delay (LD) configurations [13]. The dashed lines represent the performance of the Joint Evaluation Model JEM used for study before the development of the VTM. (a) Luma Bjontegaard Delta rate savings [14] for the VTM, (b) VTM encoding time, and (c) VTM decoding time. VTM data from JVET AHG3 report [15], and data for JEM from [16].*

dealing with the topic in the development of the VVC standard are addressed.

A. Monitoring of Computational Complexity

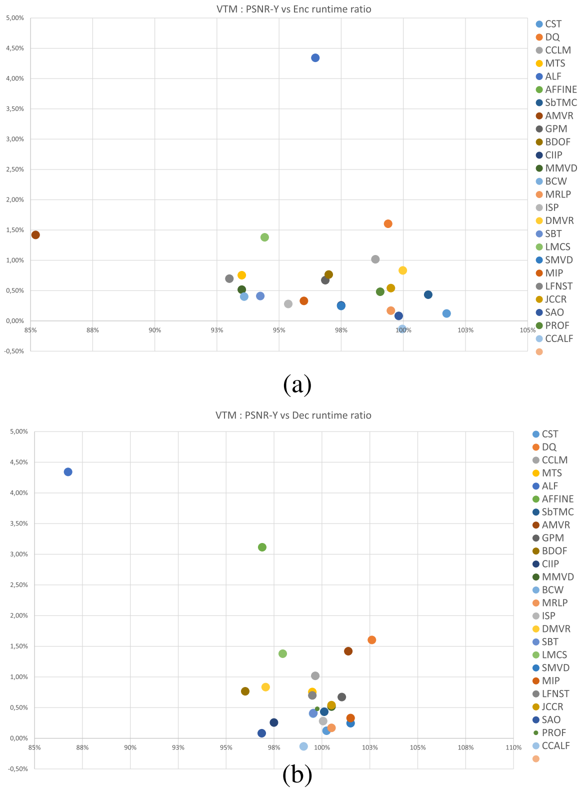
Throughout the development process of VVC, the rate-distortion performance of the VVC test model (VTM) and the computational complexity of the tools were monitored. A very high number of tools were proposed for VVC, some of which originally had a significant complexity impact. The tolerable computational load of each tool was, therefore, carefully balanced against the resulting rate-distortion benefit. Since this aspect was considered to be of utmost importance, a number of *ad hoc* groups (AHGs) dedicated to the topics of compression efficiency evolution and complexity consideration were active throughout the development process of VVC. They are briefly described as follows.

1) *AHG 3: Test model software development [15]:* This AHG monitored the rate-distortion performance according to the CTCs for 1) AI; 2) RA; and 3) LD configurations, and also reported the encoding and decoding times as a general measure for the computational complexity. The evolution of these data over the versions of the VTM software is shown in Fig. 2.
2) *AHG 10: Encoding algorithm optimization [17]:* This AHG worked on encoder performance optimization with respect to objective and subjective quality, and potentially improved settings for coding tools available in the specification. The mandates of this group further included the study of adaptive coding structures and multipass encoding. While its scope was mostly on quality aspects, the work conducted in this context influenced the final VTM encoder design and also impacted the encoder configuration used in the CTCs [13].
3) *AHG 13: Tool reporting procedure and testing [18]:* The task of AHG 13 was to study the impact of switching ON or OFF the tools implemented in the VTM software. These “tool ON/OFF” tests had a crucial impact on understanding the benefit of the numerous tools adopted into the VVC design. Fig. 3 exemplary shows the impact of a set of tools on the luma Bjontegaard Delta rate savings over encoding and decoding time. For example, it can be seen from the figure that the activation of the adaptive loop filter (ALF) has the most significant impact in terms of rate savings while, at the same time, also having the greatest impact on the decoding time.
4) *AHG 16: Implementation studies [19]:* This AHG focused on the analysis of proposed coding tools to identify implementation issues relating to: 1) decoder pipelines; 2) decoder throughput; and 3) other aspects of implementation difficulty. It further solicited hardware analysis of complex tools and provide feedback on potential solutions to address identified issues, thereby giving guidance throughout the standardization process whenever concerns on implementation aspects were raised.

B. Picture Partitioning

The partitioning of pictures plays a key role in the layout of video encoders and decoders. The units and blocks, which the picture is partitioned into, set the basic buffer and storage constraints since the whole design works on the level of rectangular blocks. The scheme includes different grouping mechanisms that enable parallel processing on both the encoder and decoder sides.

Relevant VVC partitioning concepts include the coding tree unit (CTU) of size 128×128, which is further partitioned into rectangular coding units (CUs) by using a multitype tree (MTT). The CU size ranges from 128×128 down to 4×4 samples, including nonsquare sizes. The CTUs are organized in slices, which includes a set of CTUs in raster-scan order. Each slice can be independently decoded. The area covered by the slice is controlled by horizontal and vertical tile boundaries that allow for the formation of rectangular regions. Tiles share the feature



(a)

(b)

Fig. 3. ***Illustration of VTM "tool ON/OFF" tests: impact of deactivating various tools of the VTM in terms of Luma Bjontegaard Delta rate savings over (a) encoding time and (b) decoding time. Plots cited from [18].***

of slices to be independently decodable, thereby providing a means for a flexible partitioning of the encoded picture into a set of rectangular regions. Finally, subpictures are a concept newly introduced in VVC, which allows for more flexible and independent handling of partitioned regions. In contrast to tiles and slices, subpicture boundaries are treated like picture boundaries, i.e., no information from outside the subpicture area is used for motion compensated prediction from reference pictures. This is specifically useful for the processing of ultrahigh-resolution video, which needs to be encoded, e.g., in 360° video applications.

C. Buffer Concepts

Two major buffers are implemented in a typical video decoder: 1) the coded picture buffer (CPB) that contains the compressed bitstream at the input side and 2) the decoded picture buffer (DPB) that holds the reconstructed pictures until the display and for use as reference pictures for motion-compensated interprediction. Besides the reconstructed pictures, motion vectors of the reference pictures are stored for the purpose of temporal prediction of motion information. In order to reduce memory requirements, the motion information is subsampled to an 8×8 grid, thereby omitting motion information from smaller block sizes.

The decoding process itself relies on a line-buffer concept: for the decoding of a CTU, only information from the neighboring CTUs on the top and left (each of them if available) may be employed for prediction inside the current CTU. Thereby, only one CTU line needs to be stored in the decoder for future reference during the decoding process. The required information is even more reduced as it is restricted to reconstructed samples and motion information of the CUs at the adjacent boundary of the current CTU.

Compared to its predecessor HEVC, the maximum block size has been extended from 64×64 to 128×128, which induces an increased burden on the decoding process. Specifically, due to the very large size of isolated blocks, the processing time of parallel processes may become very unbalanced depending on the neighboring block sizes. In order to mitigate this effect, the concept of virtual pipeline data units (VPDUs) is introduced. The VPDUs introduce an additional virtual partitioning into a maximum of 64×64, which affects the maximum block size for intraprediction and transform blocks. Blocks of larger size are inherently split to a 64×64 grid without further signaling need. This additional partitioning is intended to increase the throughput for pipeline processing in hardware decoders.

D. Integer Arithmetic

Since the release of AVC [6], video coding standards of ITU-T and ISO/IEC follow the principle of only applying integer arithmetic in the decoding process. Furthermore, divisions by numbers other than 2^N are completely replaced by approximation through integer multiplication and bit shifts. Thereby, computationally more complex division operations are completely avoided. The concept further enables a bit-exact specification, as any rounding operations are explicitly expressed as part of the computation. As a result, any compliant decoder will generate exactly the same reconstructed video.

E. Decoder-Side Search

Decoder-side search algorithms have been considered too complex to be realized in previous video coding standards until HEVC [7]. VVC does include decoder-side search operations. Search-based tools derive information at the decoder side based on the available reconstructed pictures. The adoption of such tools was enabled by optimized algorithms that include a limited number of search steps in the decoder process while providing rate-distortion benefits.

VVC includes methods refining the motion information on a 4 × 4 subblock level based on optical flow and decoder-side motion vector refinement (DMVR). Prediction refinement with the optical flow (PROF) for blocks with affine motion and bidirectional optical flow (BDOF) for bipredicted blocks refine the motion information on a 4 × 4 subblock level by minimizing the difference between the two prediction blocks in the two reference pictures. The scheme relies on a simplified optical flow equation. DMVR refinement applies an actual search over a 2 × 2 integer sample range with subsample refinement. This scheme is applied on complete blocks for CUs smaller than 16 × 16 samples and 16 × 16 subblocks otherwise.

F. Loop Filters

Loop filters operate on the reconstructed video after block prediction and reconstruction of the quantized residual signal. VVC includes three major loop filters with different purposes: 1) deblocking filter; 2) sample-adaptive offset (SAO); and 3) ALF.

The block-based nature of the coding scheme implies the occurrence of visible and potentially very annoying block boundaries in the reconstructed picture, especially for low bitrates. The deblocking filter is designed to mitigate these artifacts. While, in AVC, the filter implied a significant complexity burden and a bottleneck in terms of throughput [20], the design of the deblocking filter in HEVC and the refined version in VVC allow for parallelized and block-based operations.

The sample adaptive offset (SAO) filter operates on a sample basis using either a 3 × 3 neighborhood for modifying the sample value or just operating on the intensity value of the sample itself. It has already been included in HEVC and is specifically helpful to reduce ringing or banding artifact [21]. Due to its very local structure, it can be efficiently implemented and parallelized.

The ALF is the most complex loop filter. The filter provides signal enhancement by adaptively attenuating or enhancing relevant frequency components in the signal. In VVC, the filter operates on a 4 × 4 block basis and applies a diamond-shaped filter kernel of 5 × 5 or 7 × 7 samples with 25 sets of filter coefficients available and selected based on local activity and directionality. As depicted in Fig. 3, the activation of ALF has the most significant impact on the rate-distortion performance of the coding scheme while, at the same time, having the highest impact on the decoder runtime for the VTM.

G. Concluding Remarks

The main purpose of this section is to highlight complexity aspects of video coding standards using VVC as an example. In the following, we highlight articles in this area. Viitanen *et al.* [22] provide a complexity analysis of the HEVC decoder, whereas Pakdaman *et al.* [23] provide such an analysis for VVC encoding and decoding. Recently, Bossen *et al.* [24] provide a complexity analysis of VVC and its VTM reference software. However, these articles focus on video coding standards defined in the context of ISO/IEC and ITU-T. For other video coding standards (e.g., AVC, HEVC, VVC, EVC, LCEVC, VP9, and AV1), mainly performance comparisons are available, such as [25] targeting bitrate savings and encoding/decoding time. In particular, LCEVC is designed to provide a lightweight enhancement to provide high-resolution videos based on any legacy codec providing a low-resolution version of the same video. There are two sublayers in the enhancement layer and low complexity coding tools, including default and customizable upscaling filters, picture-level-switchable colocated-only temporal prediction, and 2 × 2 or 4 × 4 fixed point transforms. The verification test [26] demonstrates the ability of LCEVC to outperform state-of-the-art codecs AVC, HEVC, EVC, and VVC as half-resolution base layer codecs over the corresponding full-resolution single layer anchors and unguided up-sampling ones. Further performance evaluations are highlighted in Section IV.

III. OUTLOOK: EMERGING MEDIA STANDARDS

This section provides an overview about emerging JPEG and MPEG media standards for which roadmaps are publicly available at their respective web sites.[6,7] Finally, we briefly highlight our own view about future video coding aspects.

A. JPEG

JPEG recently started a number of relevant activities that are worth to be reported here. In particular, Foessel *et al.* [27] provide a status and progress report. In terms of emerging media standards not covered in this special issue, *JPEG Pleno* [28]–[30] is probably the most advanced standard that aims to address novel image modalities, i.e., 1) light field; 2) point cloud; and 3) holographic imaging. Ebrahimi *et al.* [28] describe the rationale behind the vision for the JPEG Pleno standard that is based on the plenoptic function [31]. Astola *et al.* [29] provide a comprehensive overview of JPEG Pleno, including use cases, requirements, and parts 1–4 of JPEG Pleno (ISO/IEC 21794). Part 1 defines the JPL file format, part 2 is about light field coding, and parts 3 and 4 define conformance and reference software, respectively. Finally, it outlines future work items of JPEG Pleno, i.e., point cloud and holography. For example, holographic content compression challenges and standardization efforts are described in [30], while streaming issues are addressed in [32].

Another emerging JPEG standard is related to learning-based image coding and referred to as *JPEG AI*.[8] The goal of JPEG AI is to define an *end-to-end learning-based coding* approach

---

[6] Accessed: June 18, 2021. [Online] Available: https://jpeg.org/

[7] Accessed: June 18, 2021. [Online] Available: http://www.mpeg.org/

[8] Accessed: April 16, 2021. [Online] Available: https://jpeg.org/jpegai/

that is currently subject to research in multiple grand challenges colocated with scientific conferences, such as IEEE MMSP'20 learning-based image coding challenge[9] [33] and CVPR's Workshop and Challenge on Learned Image Compression (CLIC)[10] [34]. These challenges are typically defined around a common dataset with the aim to improve coding efficiency with respect to a predefined anchor. An overview of recent advances in end-to-end learned image and video compression can be found in [35].

JPEG DNA[11] refers to another JPEG exploration activity, which aims to define an efficient image coding format by using deoxyribonucleic acid (DNA) that is represented by a quaternary representation basis rather than a binary unit like in traditional image coding. Thus, artificial DNA molecules are created representing image data resulting in DNA-based media storage. Dimopoulou *et al.* [36] propose a first closed-loop solution for generating a constrained fixed-length quaternary code, which is extended in [37] to a variable-length encoding solution based on the JPEG standard. Li *et al.* [38] describe a novel image-based DNA system (IMG-DNA) for the efficient storage of images in DNA with improved robustness resulting in higher fault-tolerance compared to the state of the art.

Finally, *JPEG Fake Media*[12] aims to address various issues related to media manipulation utilizing deep learning (DL)-based methods that produce realistic media content almost indistinguishable from authentic content. Such a kind of content is generally referred to as "deepfakes" or simply "fake media" resulting in various issues concerning security, privacy, integrity, and so on. In this context, Temmermans *et al.* [39] describe the current state and scope of media blockchain technologies for JPEG Privacy and Security that is now continued as part of the JPEG Fake Media efforts. JPEG Fake Media basically targets image descriptors in two categories: 1) modification descriptors and 2) secure signaling of authenticity information.

B. MPEG

MPEG's roadmap[13] in terms of open media standards is roughly clustered into *media coding*, *systems and tools*, and *beyond media* and further detailed below.

1) Media Coding: It can be further divided into: 1) audio; 2) visual; and 3) haptics data. Audio-related aspects are covered in this special issue by Quackenbush and Herre, and video/visual-related aspects are extensively discussed. Deep neural network-based video coding (DNNVC; sometimes simply referred to as AI-based video coding) is at the time of writing this article an exploratory activity and targets end-to-end learned video compression [35]. A survey of machine learning-based video coding optimizations is provided by Zhang *et al.* [40]. Liu *et al.* [41] provide a review and case study about DL-based video coding. Another aspect is related to dense light field coding for which a survey is provided by Conti *et al.* [42] that cover both JPEG and MPEG. Another emerging topic in the video coding space is known as video coding for machines (VCM) targeting nonhuman users in the context of machine vision. Duan *et al.* [43] provide a comprehensive introduction to this topic, while Fischer *et al.* [44] specifically address feature-based rate-distortion optimization in the context of VCM. Finally, MPEG has been recently started working on the coded representation of haptics that enable efficient representation and compression of time-dependent haptic signals and are suitable for the coding of timed haptic tracks that can be synchronized with audio and/or video media. Holland *et al.* [45] provide an overview of the IEEE 1918.1 "Tactile Internet (TI)" standards workings group and its standards, including haptic codecs for the TI.

2) Systems and Tools: It covers a broad range of topics with the aim to enable efficient delivery of all kinds of media to future consumption devices stimulating human senses, mostly related to sight and hearing but also others such as touch and potentially also smell and taste [46]. The OMAF standard is already covered in this special issue, which is also extensively covered in the scientific literature [47]–[49] addressing dynamic, and adaptive 360° video streaming use cases [50]–[52]. At the time of writing this article, the second edition of the OMAF standard has been ratified in December 2020, which provides support for more than one viewport and the dynamic bitstream generation according to the viewport among others. The processing of media components within the network is addressed in MPEG network-based media processing (NBMP) that is further described in [60] and [61], including its use in the context of OMAF [55]. Another exploration activity includes the delivery of [six degrees of freedom (6DoF)] media for emerging (immersive) display technologies, such as introduced in [56]–[59], which certainly also impacts the Quality of Experience (QoE) [60].

3) Beyond Media: It covers all kinds of coded representation formats that are not referred to as traditional media content (i.e., audio or/and video/visual data) and currently comprises two work items: 1) neural network compression for multimedia and 2) genome compression. Regarding the former, i.e., neural network compression, Sattler *et al.* [61] describe trends and advances in deep neural network communication as a new field of research at the intersection of machine learning and communications. In particular,

---

[9] Accessed: April 16, 2021. [Online] Available: https://jpegai.github.io/

[10] Accessed: April 16, 2021. [Online] Available: http://compression.cc/

[11] Accessed: April 16, 2021. [Online] Available: https://jpeg.org/jpegdna/

[12] Accessed: April 16, 2021. [Online] Available: https://jpeg.org/jpegfakemedia/

[13] Accessed: April 16, 2021. [Online] Available: https://www.mpegstandards.org/

a survey of methods is provided, which communicates neural networks toward clients for the purpose of: 1) on-device inference; 2) federated learning; 3) peer-to-peer learning; and 4) distributed learning within data centers. In the latter, genome compression, Voges *et al.* describe the first open ISO/IEC standard for the compression and exchange of genomic sequencing data referred to as MPEG-G.

C. Future of Media Standards

As outlined above, an immediate next step in the field of visual media standards will be related to: 1) the optimization of existing codecs (i.e., with new coding tools, including neural network-based coding tools) and 2) so-called end-to-end deep neural network-based video codecs. Section IV provides an overview of the articles covered in this special issue, and specifically, Ding *et al.* address this aspect of video coding.

It is our belief that future steps should aim to better understand the human visual system (HVS) and usage for perceptual coding tools. Some of these aspects are already covered in today's video coding standards/tools (e.g., scene cuts and Group of Picture structure with hierarchical quantization) but not so much emphasized like in audio coding.

Finally, multimedia content currently dominates traffic on today's networks that are expected to further increase in the future. Future media standards will certainly provide specifications enabling interoperable, immersive media applications, and services, but this will coexist with traditional, 2-D rectangular video for a while. Other topics are related to video coding for (autonomous) machines [43], TI (including haptics) [62]–[64], or/and enhancing traditional video with sensory effects [46], [65] enabling truly immersive experiences [60]. In summary, the future of media standards provides a lot of open aspects for R&D.

IV. OVERVIEW OF THIS SPECIAL ISSUE

The goal of this special issue is to provide an overview of open media standards focusing on the technology developed in the context of those. This special issue covers the following aspects of open media standards: 1) video coding; 2) image coding; 3) audio coding; 4) point cloud coding (PCC); 5) systems aspects (i.e., omnidirectional media); 5) compression and exchange of genomic sequencing data. In this section, we provide a brief overview of the articles within this special issue.

In the context of *traditional video coding*, this special issue features two articles covering video coding formats from both ISO/IEC/ITU-T (i.e., HEVC and VVC) and AOMedia (i.e., AV1). Han *et al.* provide a "A technical overview of AV1" that comprises a deep insight into the AV1 video compression format, specifically a technical overview of the AV1 codec design, including: 1) high-level syntax; 2) reference frame system (reference frames, alternate reference frame, and frame scaling); 3) superblock and tiling structures; 4) coding block operations (partitioning, intra/interframe prediction, dynamic motion vector referencing scheme, transform coding, and quantization); 5) entropy coding system (probability model, arithmetic coding, and level map transform coefficient coding system); 6) postprocessing filters (deblocking, constrained directional enhancement, loop restoration, frame super-resolution, and film grain synthesis); and 7) profile and level definition. Finally, a performance evaluation is provided, which compares AV1 to its predecessor VP9. Bross *et al.* provide an overview titled "Developments in international video coding standardization after AVC, with an overview of versatile video coding (VVC)." In particular, this article reviews video coding standards after AVC, i.e., HEVC and VVC, each of them targeting a 50% bit-rate reduction compared to its respective predecessor. In the beginning, this article provides an overview of video coding standards in general with a block diagram of a hybrid video encoder, including the modeling of the decoder within the encoder. Furthermore, it briefly highlights HEVC focusing on its first version and extensions, i.e., range extensions, scalable HEVC, multiview, 3-D, and screen content coding. VVC is the most recent video coding standard in this series of standards at the time of writing of this article, which is described next, including: 1) an overview of the standardization and development process; 2) a description of coding tools (block partitioning, inter/intrapicture prediction, transform, quantization, entropy coding, in-loop filtering, screen content coding, and 360° video coding); and 3) systems and transport interfaces (RA support, adaptation parameter set, picture header, reference picture management, high-level picture partitioning, picture resolution changes with interpicture prediction, scalability support, and profile/tier/level aspects). Finally, this article provides both objective and subjective video coding efficiency analysis of VVC with respect to its predecessor and also an optimized, open-source VVC encoder. It is noted that these two articles deliberately do not report performance evaluations among the codecs under discussion since this should ideally be based on comparable encoder implementations that are not trivially achieved. The interested reader is referred to various articles documenting such kinds of evaluations, e.g., [66], [67], [68], or [69], but we note to consider them with caution.

Recently, (deep) neural networks have been applied in the context of video coding resulting in approaches that could be collectively referred to as deep neural network-based video coding (DNNVC). In this special issue, Ding *et al.* provide a report titled "Advances in video compression system using deep neural network: A review and case studies" that targets video coding improvements for both individual blocks of the hybrid video encoder and jointly across multiple blocks, including end-to-end approaches. In particular, this article clusters DNNVC in three modules: 1) preprocessing (based on saliency and analysis/synthesis); 2) coding (modularized and end-to-end neural video coding); and 3) postprocessing (in-loop and postfiltering). Finally, it provides case studies for each of the three modules, i.e., switchable texture-based video preprocessing that leverages DNN-based semantic understanding for subsequent coding improvement, a framework for end-to-end neural video coding, and in-loop/postfiltering using stacked DNN-based neural filters for quality enhancement of reconstructed frames.

Immersive video is becoming more and more important. Thus, this special issue contributes the article of Boyce *et al.* titled "MPEG immersive video coding standard" targeting immersive volumetric content captured by multiple cameras with 6DoF. This article provides an overview of the MPEG Immersive Video (MIV) codec by providing: 1) a description of the given source material; 2) details about the codec structure, atlases, and patches; and 3) view parameters. Furthermore, it describes the Test Model for Immersive Video (TMIV) reference software encoder, decoder, and renderer. The alignment with the V3C/V-PCC specification and experimental results concludes this article.

PCC is addressed in the article by Cao *et al.* titled "Compression of sparse and dense dynamic point clouds—Methods and standards." This article provides a comprehensive survey of state-of-the-art methods clustered into compression: 1) for static point cloud objects; 2) for dynamic point cloud objects; and 3) using DL methods. The core of the article describes MPEGs recent standards for Video-based Point Cloud Compression (V-PCC) and Geometry-based Point Cloud Compression (G-PCC). For both compression formats, V-PCC and G-PCC, coding tools are described in detail, and evaluation results are presented as well.

As the representative article for the field of *image coding*, Descampe *et al.* present an overview on the state of the art in image coding for mezzanine applications in their article titled "JPEG XS—A new standard for visually lossless low-latency lightweight image coding." The goal of this standard is to provide a means for very-high-quality image storage and transmission, e.g., for display or sensor links. The paper provides a broad introduction, including use cases, key features, and comparison with other image codecs before arriving at its core, i.e., the technical overview comprising the following aspects: 1) intercomponent and intracomponent decorrelation; 2) quantization; 3) entropy coding; 4) packetization; 5) rate computation; 6) smoothing buffer and decoding steps; and 7) system integration and encoding architecture. Finally, this article concludes with a latency analysis, an overview of profiles/levels/formats, performance evaluation (i.e., objective, subjective, and complexity analysis), and status of the standardization process and upcoming extensions.

In addition to visual media, audio provides yet another important modality in the context of this special issue, not only but specifically since the standardization of MPEG-1 Audio Layer III also known as MP3 [70]. Recently, immersiveness has become increasingly important for audio, and thus, Quackenbush and Herre provide an overview titled "MPEG standards for compressed representation of immersive audio." This article focuses on two standards, namely, MPEG-H Audio (universal immersive audio coding) and MPEG-I Immersive Audio (compressed representation for virtual and augmented reality). The former is described in detail with respect to coding tools and performance, whereas the latter is—at the time of writing of this article—still under development, and the article focuses on the differences from other standards and the requirements and development process of an MPEG-I 6DoF immersive audio standard.

While video, image, and audio codings are fundamental to an efficient representation of audio–visual content, *systems aspects*, such as storage and delivery (including streaming) formats, are required to enable applications and services with high QoE. Hannuksela and Wang provide "An overview of Omnidirectional MediA Format (OMAF)." which is arguably the first virtual reality (VR) system standard that includes support for 360° video (among others). This article introduces the end-to-end OMAF architecture from content authoring to the player and highlights representation formats of omnidirectional video and images, i.e., the coordinate system used in OMAF, projection formats, regionwise packaging, mesh omnidirectional video, fisheye omnidirectional video and images, and supplemental metadata for omnidirectional video and images. Furthermore, various aspects of 360° video streaming are described, and finally, video/images profiles and toolset brands are highlighted.

Last but not least, open media standards are not limited to audio-visual information and associated data but may also include other types of data. In this context, Voges *et al.* provide "An introduction to MPEG-G: The first open ISO/IEC standard for the compression and exchange of genomic sequencing data." The amount of data generated by genomic sequencing machines calls for efficient representation formats. This article provides an overview of the MPEG-G standard comprising five parts: 1) transport and storage of genomic information; 2) coding of genomic information; 3) metadata and application programming interfaces; 4) reference software; and 5) conformance. The core of the specification deals with the coding of genomic information comprising preprocessing, descriptor generation, transformation and entropy coding, and the decoding process. Finally, results about the compression performance are described, including a discussion thereof.

We hope that this special issue provides readers a better understanding of open media standards in the areas of visual and audio coding, and systems aspects.

V. CONCLUSION

The aim of this article is to provide an overview of the articles covered in this special issue on open media standards. Media standards are important with respect to interoperability among vendors or providers of applications and services that can be both source and sink of scientific R&Ds. We also discussed the term *open standards* from a general point of view and specific to most recent video coding standards. This article also includes a special topic covering the aspect of *computational complexity* using the VVC standard as an example. Finally, we provide an outlook of emerging media standards (i.e., JPEG and MPEG) based on information publicly available (i.e., roadmap) at the time of writing this article. We do hope that this special issue provides a tutorial-style overview of standards and technologies in the multimedia domain. Furthermore, we do hope that it will stimulate further R&D efforts in this exciting area, and we invite everyone to get involved in and excited about media standards. ■

Acknowledgment

The authors would like to thank authors for their efforts in preparing the manuscripts for this special session and all reviewers for their valuable feedback. They would also like to thank the many contributors to the various standardization activities covered in this special issue and their efforts in realizing these open media standards. They are deeply grateful to Prof. Gianluca Setti, the current Editor-in-Chief, for his support and guidance. They would also like to thank Vaishali Damle and Jo Sun from PROCEEDINGS OF THE IEEE for their incredible effort and support in preparing this special issue. The financial support of the Austrian Federal Ministry for Digital and Economic Affairs, the National Foundation for Research, Technology and Development, and the Christian Doppler Research Association, is gratefully acknowledged. Christian Doppler Laboratory ATHENA: https://athena.itec.aau.at/.

REFERENCES

[1] Sandvine. (2020). *The Mobile Internet Phenomena Report*. [Online]. Available: https://www.sandvine.com/phenomena

[2] "Cisco visual networking index: Forecast and trends, 2017–2022," Cisco, San Jose, CA, USA, White Paper, 2019. [Online]. Available: http://bit.ly/2wmdZJb

[3] J. Nielsen, "Nielsen's law of internet bandwidth (updated 2019)," Nielsen Norman Group, Fremont, CA, USA, White Paper, 2019. [Online]. Available: https://www.nngroup.com/articles/law-of-bandwidth/

[4] C. Timmerer, "Immersive media delivery: Overview of ongoing standardization activities," *IEEE Commun. Standards Mag.*, vol. 1, no. 4, pp. 71–74, Dec. 2017.

[5] *Standards and Patents*, document WIPO, SCP-13-2, Standing Committee on the Law of Patents, 2009. [Online]. Available: https://www.wipo.int/edocs/mdocs/scp/en/scp 13/scp132.pdf

[6] *Advanced Video Coding*, document ITU-T Recommendation H.264 and ISO/IEC 14496-10 (AVC), ITU-T and ISO/IEC JTC1, Tech. Rep., May 2003. [Online]. Available: http://www.itu.int/rec/T-REC-H.264

[7] *High Efficiency Video Coding*, document ITU-T Recommendation H.265 and ISO/IEC 23008-2 (HEVC), ITU-T and ISO/IEC JTC1, Tech. Rep., Apr. 2013. [Online]. Available: http://www.itu.int/rec/T-REC-H.265

[8] *Versatile Video Coding*, document ITU-T Recommendation H.266 and ISO/IEC 23090-3 (VVC), ITU-T and ISO/IEC JTC1, Tech. Rep., Aug. 2020. [Online]. Available: https://www.itu.int/rec/T-REC-H.266

[9] *Guide for ITU-T and ISO/IEC JTC 1 Cooperation*, document ITU-T and ISO/IEC JTC 1 Rec. A.23 Annex A, Oct. 2010. [Online]. Available: http://www.itu.int/rec/T-REC-A.23/en

[10] K. Choi *et al.*, "MPEG-5: Essential video coding standard," *Proc. SPIE*, vol. 11137, pp. 250–258, Sep. 2019, doi: 10.1117/12.2530429.

[11] K. Choi, J. Chen, D. Rusanovskyy, K.-P. Choi, and E. S. Jang, "An overview of the MPEG-5 essential video coding standard [standards in a nutshell]," *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 160–167, May 2020.

[12] G. Meardi *et al.*, "MPEG-5—Part 2: Low complexity enhancement video coding (LCEVC): Overview and performance evaluation," *Proc. SPIE*, vol. 11510, pp. 238–257, Aug. 2020, doi: 10.1117/12.2569246.

[13] F. Bossen, J. Boyce, K. Suehring, X. Li, and V. Seregin, *JVET Common Test Conditions and Software Reference Configurations for SDR Video*, Joint Video Experts Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, document JVET-N1010, Mar. 2019.

[14] K. Andersson *et al.*, *Working Practices Using Objective Metrics for Evaluation of Video Coding Efficiency Experiments (Draft 4)*, Joint Video Experts Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, docunment JVET-T2016, Jan. 2021.

[15] F. Bossen, X. Li, and K. Suehring, *JVET AHG Report: Test Model Software Development (AHG3)*, Joint Video Experts Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, document JVET-T0003, Oct. 2020.

[16] M. Karczewicz and E. Alshina, *JVET AHG Report: Tool Evaluation (AHG1)*, Joint Video Exploration Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, document JVET-H0001, Oct. 2017.

[17] A. Duenas *et al.*, *JVET AHG Report: Encoding Algorithm Optimization (AHG10)*, document JVET-T0010, Oct. 2020. [Online]. Available: https://jvet-experts.org/doc_end_user/documents/20_Teleconference/wg11/%JVET-T0010-v2.zip

[18] W.-J. Chien *et al.*, *JVET AHG Report: Tool Reporting Procedure (AHG13)*, Joint Video Experts Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, document JVET-T0013, Oct. 2020.

[19] M. Zhou *et al.*, *JVET AHG Report: Implementation Studies (AHG16)*, Joint Video Experts Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, document JVET-S0016, Jun. 2020.

[20] A. Norkin *et al.*, "HEVC deblocking filter," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1746–1754, Dec. 2012.

[21] C.-M. Fu *et al.*, "Sample adaptive offset in the HEVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1755–1764, Dec. 2012.

[22] M. Viitanen, J. Vanne, T. D. Hamalainen, M. Gabbouj, and J. Lainema, "Complexity analysis of next-generation HEVC decoder," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2012, pp. 882–885.

[23] F. Pakdaman, M. A. Adelimanesh, M. Gabbouj, and M. R. Hashemi, "Complexity analysis of next-generation VVC encoding and decoding," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2020, pp. 3134–3138.

[24] F. Bossen, K. Suhring, A. Wieckowski, and S. Liu, "VVC complexity and software implementation analysis," *IEEE Trans. Circuits Syst. Video Technol.*, early access, Apr. 9, 2021, doi: 10.1109/TCSVT.2021.3072204.

[25] G. Minopoulos, V. A. Memos, K. E. Psannis, and Y. Ishibashi, "Comparison of video codecs performance for real-time transmission," in *Proc. 2nd Int. Conf. Comput. Commun. Internet (ICCCI)*, Jun. 2020, pp. 110–114.

[26] *Verification Test Report on the Compression Performance of Low Complexity Enhancement Video Coding*, ISO/IEC JTC 1/SC 29/WG 4, document N0076, Apr. 2021.

[27] S. Foessel *et al.*, "JPEG status and progress report," *SMPTE Motion Imag. J.*, vol. 129, no. 8, pp. 83–91, Sep. 2020.

[28] T. Ebrahimi, S. Foessel, F. Pereira, and P. Schelkens, "JPEG pleno: Toward an efficient representation of visual reality," *IEEE Multimedia*, vol. 23, no. 4, pp. 14–20, Oct./Dec. 2016.

[29] P. Astola *et al.*, "JPEG Pleno: Standardizing a coding framework and tools for plenoptic imaging modalities," *ICT Discoveries*, vol. 3, no. 1, pp. 1–15, 2020. [Online]. Available: http://handle.itu.int/11.1002/pub/8153d79a-en

[30] R. K. Muhamad *et al.*, "JPEG Pleno holography: Scope and technology validation procedures," *Appl. Opt.*, vol. 60, no. 3, pp. 641–651, Jan. 2021. [Online]. Available: http://ao.osa.org/abstract.cfm?URI=ao-60-3-641

[31] M. Landy and J. A. Movshon, *The Plenoptic Function and the Elements of Early Vision*, 1991, pp. 3–20.

[32] H. Amirpour, C. Timmerer, and M. Ghanbari, "Towards view-aware adaptive streaming of holographic content," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, London,

U.K., Jul. 2020, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/9106055

[33] A. Gotchev, D. Tian, A. Mesaros, G. Cheung, M. Sjöström, and G. Buracchi, in *Proc. IEEE 22nd Int. Workshop Multimedia Signal Process. (MMSP)*, 2020.

[34] T. Boult *et al.*, in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020. [Online]. Available: https://ieeexplore.ieee.org/xpl/conhome/9142308/proceeding, doi: 10.1109/CVPR42600.2020.

[35] W.-H. Peng and H.-M. Hang, "Recent advances in end-to-end learned image and video compression," in *Proc. IEEE Int. Conf. Vis. Commun. Image Process. (VCIP)*, Dec. 2020, pp. 1–2.

[36] M. Dimopoulou, E. G. S. Antonio, and M. Antonini, "A JPEG-based image coding solution for data storage on DNA," Tech. Rep. 2103.09616, 2021. [Online]. Available: https://arxiv.org/abs/2103.09616

[37] M. Dimopoulou and M. Antonini, "Image storage in DNA using vector quantization," in *Proc. 28th Eur. Signal Process. Conf. (EUSIPCO)*, Jan. 2021, pp. 516–520.

[38] B. Li, L. Ou, and D. Du, "Image-based approximate DNA storage system," Tech. Rep. 2103.02847, 2021. [Online]. Available: https://arxiv.org/abs/2103.02847

[39] F. Temmermans, D. Bhowmik, F. Pereira, T. Ebrahimi, and P. Schelkens, "Exploration of media block chain technologies for JPEG privacy and security," in *Proc. SPIE*, vol. 11353, Apr. 2020, Art. no. 113530U, doi: 10.1117/12.2556684.

[40] Y. Zhang, S. Kwong, and S. Wang, "Machine learning based video coding optimizations: A survey," *Inf. Sci.*, vol. 506, pp. 395–423, Jan. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025519307145

[41] D. Liu, Y. Li, J. Lin, H. Li, and F. Wu, "Deep learning-based video coding: A review and a case study," *ACM Comput. Surveys*, vol. 53, no. 1, pp. 1–35, May 2020, doi: 10.1145/3368405.

[42] C. Conti, L. D. Soares, and P. Nunes, "Dense light field coding: A survey," *IEEE Access*, vol. 8, pp. 49244–49284, 2020.

[43] 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.

[44] K. Fischer, F. Brand, C. Herglotz, and A. Kaup, "Video coding for machines with feature-based rate-distortion optimization," in *Proc. IEEE 22nd Int. Workshop Multimedia Signal Process. (MMSP)*, Sep. 2020, pp. 1–6.

[45] O. Holland *et al.*, "The IEEE 1918.1 'tactile Internet' standards working group and its standardss," *Proc. IEEE*, vol. 107, no. 2, pp. 256–279, Jan. 2019.

[46] C. Timmerer, M. Waltl, B. Rainer, and H. Hellwagner, "Assessing the quality of sensory experience for multimedia presentations," *Signal Process., Image Commun.*, vol. 27, no. 8, pp. 909–916, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0923596512000252

[47] R. Skupin, Y. Sanchez, D. Podborski, C. Hellge, and T. Schierl, "Viewport-dependent 360 degree video streaming based on the emerging omnidirectional media format (OMAF) standard," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, p. 4592.

[48] A. Zare, A. Aminlou, and M. M. Hannuksela, "6K effective resolution with 4K HEVC decoding capability for OMAF-compliant 360° video streaming," in *Proc. 23rd Packet Video Workshop*, Jun. 2018, pp. 72–77, doi: 10.1145/3210424.3210425.

[49] S. Ahsan, A. Hourunranta, I. D. D. Curcio, and E. Aksu, "FriSBE: Adaptive bit rate streaming of immersive tiled video," in *Proc. 25th ACM Workshop Packet Video*, Jun. 2020, pp. 28–34, doi: 10.1145/3386292.3397121.

[50] M. Graf, C. Timmerer, and C. Mueller, "Towards bandwidth efficient adaptive streaming of omnidirectional video over HTTP: Design, implementation, and evaluation," in *Proc. 8th ACM Multimedia Syst. Conf.*, Jun. 2017, pp. 261–271, doi: 10.1145/3083187.3084016.

[51] C.-L. Fan, W.-C. Lo, Y.-T. Pai, and C.-H. Hsu, "A survey on 360° video streaming: Acquisition, transmission, and display," *ACM Comput. Surv.*, vol. 52, no. 4, pp. 1–36, Sep. 2019, doi: 10.1145/3329119.

[52] A. Yaqoob, T. Bi, and G.-M. Muntean, "A survey on adaptive 360° video streaming: Solutions, challenges and opportunities," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 4, pp. 2801–2838, 4th Quart., 2020.

[53] C. Timmerer, "MPEG column: 125th MPEG meeting in Marrakesh, Morocco," *ACM SIGMultimedia Records*, vol. 11, no. 1, p. 1, Mar. 2019, doi: 10.1145/3458462.3458467.

[54] X. Ducloux and T. Fautier, "New MPEG codecs and immersive applications standards for 2020," *SMPTE Motion Imag. J.*, vol. 129, no. 8, pp. 103–107, Sep. 2020.

[55] Y. You, A. Hourunranta, and E. B. Aksu, "OMAF4Cloud: Standards-enabled 360° video creation as a service," *SMPTE Motion Imag. J.*, vol. 129, no. 9, pp. 18–23, Oct. 2020.

[56] D. Smalley, T.-C. Poon, H. Gao, J. Kvavle, and K. Qaderi, "Volumetric displays: Turning 3-D inside-out," *Opt. Photon. News*, vol. 29, no. 6, pp. 26–33, Jun. 2018. [Online]. Available: http://www.osa-opn.org/abstract.cfm?URI=opn-29-6-26

[57] T. Zhou, R. Tucker, J. Flynn, G. Fyffe, and N. Snavely, "Stereo magnification: Learning view synthesis using multiplane images," *ACM Trans. Graph.*, vol. 37, no. 4, pp. 1–12, Aug. 2018, doi: 10.1145/3197517.3201323.

[58] M. Broxton *et al.*, "Immersive light field video with a layered mesh representation," *ACM Trans. Graph.*, vol. 39, no. 4, p. 86, Jul. 2020, doi: 10.1145/3386569.3392485.

[59] J. V. D. Hooft *et al.*, "From capturing to rendering: Volumetric media delivery with six degrees of freedom," *IEEE Commun. Mag.*, vol. 58, no. 10, pp. 49–55, Oct. 2020.

[60] A. Perkis *et al.*, "QUALINET white paper on definitions of immersive media experience (IMEx)," Tech. Rep., 2020.

[61] F. Sattler, T. Wiegand, and W. Samek, "Trends and advancements in deep neural network communication," Tech. Rep., 2020.

[62] M. Maier, M. Chowdhury, B. P. Rimal, and D. P. Van, "The Tactile Internet: Vision, recent progress, and open challenges," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 138–145, May 2016.

[63] E. Steinbach *et al.*, "Haptic codecs for the tactile Internet," *Proc. IEEE*, vol. 107, no. 2, pp. 447–470, Feb. 2019.

[64] K. Antonakoglou, X. Xu, E. Steinbach, T. Mahmoodi, and M. Dohler, "Toward haptic communications over the 5G tactile internet," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3034–3059, 4th Quart., 2018.

[65] M. Waltl, C. Timmerer, and H. Hellwagner, "Improving the quality of multimedia experience through sensory effects," in *Proc. 2nd Int. Workshop Qual. Multimedia Exper. (QoMEX)*, Jun. 2010, pp. 124–129.

[66] D. Grois *et al.*, "Performance comparison of emerging EVC and VVC video coding standards with HEVC and AV1," *SMPTE Motion Imag. J.*, vol. 130, no. 4, pp. 1–12, May 2021.

[67] A. Zabrovskiy, C. Feldmann, and C. Timmerer, "Multi-codec DASH dataset," in *Proc. 9th ACM Multimedia Syst. Conf.*, Jun. 2018, pp. 438–443, doi: 10.1145/3204949.3208140.

[68] T. Laude, Y. G. Adhisantoso, J. Voges, M. Munderloh, and J. Ostermann, "A comprehensive video codec comparison," *APSIPA Trans. Signal Inf. Process.*, vol. 8, p. e30, Nov. 2019.

[69] J. D. Cock, A. Mavlankar, A. Moorthy, and A. Aaron, "A large-scale video codec comparison of x264, x265 and libvpx for practical VOD applications," *Proc. SPIE*, vol. 9971, Sep. 2016, Art. no. 997116, doi: 10.1117/12.2238495.

[70] K. Brandenburg and G. Stoll, "ISO/MPEG-1 audio: A generic standard for coding of high-quality digital audio," *J. Audio Eng. Soc.*, vol. 42, no. 10, pp. 780–792, Oct. 1994. [Online]. Available: https://www.aes.org/e-lib/browse.cfm?elib=6925

ABOUT THE GUEST EDITORS

Christian Timmerer (Senior Member, IEEE) received the M.Sc. (Dipl.Ing.) and Ph.D. (Dr.techn.) degrees (for research on the adaptation of scalable multimedia content in streaming and constraint environments) from Alpen-Adria-Universität (AAU), Klagenfurt, Austria, in January 2003 and June 2006, respectively.

He is currently an Associate Professor with the Institute of Information Technology (ITEC) and the Director of the Christian Doppler Laboratory ATHENA, AAU. In 2013, he cofounded Bitmovin, Klagenfurt, to provide professional services around MPEG-DASH where he holds the position of the Chief Innovation Officer (CIO)—Head of Research and Standardization. His research interests include immersive multimedia communication, streaming, adaptation, quality of experience, and sensory experience.

Dr. Timmerer is a member of the Association for Computing Machinery (ACM), the IEEE Computer Society, the IEEE Communications Society, and ACM Special Interest Group on Multimedia (SIGMM). He was the General Chair of Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS) 2008, International Conference on Quality of Multimedia Experience (QoMEX) 2013, ACM Multimedia Systems Conference (MMSys) 2016, and ACM Packet Video (PV) Workshop 2018 and has participated in several EC-funded projects, notably DANAE, ENTHRONE, P2P-Next, ALICANTE, SocialSensor, COST IC1003 QUALINET, and ICoSOLE. He also participated in ISO/MPEG work for several years, notably in the area of MPEG-21, MPEG-M, MPEG-V, and MPEG-DASH, where he also served as a Standard Editor. He was a Guest Editor of three special issues of IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (JSAC). He has served as an Associate Editor for IEEE TRANSACTIONS ON MULTIMEDIA. Further information is available at http://blog.timmerer.com.

Mathias Wien (Member, IEEE) received the Diploma and Dr.Ing. degrees from Rheinisch-Westfälische Technische Hochschule Aachen (RWTH Aachen University), Aachen, Germany, in 1997 and 2004, respectively. In 2018, he achieved the status of the Habilitation, which makes him an independent scientist in the field of visual media communication.

He was with the Institut für Nachrichtentechnik, RWTH Aachen University (Head: Prof. Jens-Rainer Ohm), as a Researcher from 1997 to 2006 and a Senior Researcher and the Head of Administration from 2006 to 2018. Since July 2018, he has been with the Lehrstuhl für Bildverarbeitung, RWTH Aachen University (Head: Prof. Dorit Merhof), as a Senior Researcher, a Leader of the Visual Media Communication Group, and the Head of Administration. He also serves as the Conventor of ISO/IEC JTC1 AG5 “Visual Quality Assessment.” He has been an Active Contributor to VVC, HEVC, and H.264/AVC. He has participated and contributed to ITU-T VCEG, ISO/IEC MPEG, the Joint Video Team (JVT), the Joint Collaborative Team on Video Coding (JCT-VC), and the Joint Video Experts Team (JVET) of VCEG and ISO/IEC MPEG. He has served as a Coeditor of the Scalability Amendment to H.264/AVC (SVC). In the aforementioned standardization bodies, he has co-chaired and coordinated several *ad hoc* groups and tool and core experiments. He has published more than 60 scientific articles and conference papers in the area of video coding and has coauthored several patents in this area. He has further authored and coauthored more than 200 standardization documents. He has published the Springer textbook *High Efficiency Video Coding: Coding Tools and Specification* that fully covers Version 1 of HEVC. His research interests include image and video processing, immersive, space–frequency adaptive and scalable video compression, and robust video transmission.

Dr. Wien is a member of the IEEE Signal Processing Society and the IEEE Circuits and Systems Society. He is also a member of IEEE Circuits and Systems Society Technical Committee Visual Signal Processing and Communications (CASS TC VSPC). He has co-organized and co-chaired special sessions at IEEE International Conference on Visual Communications and Image Processing (VCIP) and Picture Coding Symposium (PCS). He has co-organized and co-chaired the Grand Challenge on Video Compression Technology at IEEE International Conference on Image Processing (ICIP) 2017. He also serves as an Associate Editor for IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY and *Signal Processing: Image Communication*.

Lu Yu (Senior Member, IEEE) received the B.Eng. degree (honors) in radio engineering and the Ph.D. degree in communication and electronic systems from Zhejiang University, Hangzhou, China, in 1991 and 1996, respectively.

After that, she joined Zhejiang University as a Faculty Member. She is currently a Professor with the College of Information Science and Electronic Engineering and the Director of the Institute of Information and Communication Networks, Zhejiang University. She contributed more than 100 adopted technical proposals to video coding standards, including those defined by AVS, IEEE, ISO/IEC MPEG, and ITU-T SG16. Her research area includes visual perception and perceptual quality assessment, visual signal representation and coding, multimedia processing, and related architecture design, in which she published more than 160 articles and invented more than 80 granted patents.

Prof. Yu acted as the Video Sub-Group Co-Chair and the Chair of the Audio and Video Coding Standardization Working Group (AVS) for 16 years. She has been appointed as the Video Sub-Group Chair of MPEG in January 2018. She is also the Convenor of ISO/IEC JTC 1/SC 29/WG 4, MPEG Video Coding.

Amy Reibman (Fellow, IEEE) is currently a Professor of electrical and computer engineering with Purdue University, West Lafayette, IN, USA. She pursued 23 years of industrial research at AT&T Labs—Research, Atlanta, GA, USA, where she was a Distinguished Member of Technical Staff and a Lead Inventive Scientist. She has published over 40 journal articles and 120 conference papers. She holds more than 60 U.S. patents. Her research interests are in video analytics, video stabilization, and image and video quality assessment. She has also done pioneering work on video transmission over packet networks. Her current research focuses on video analytics for real-world applications where inputs may have substantial-quality impairments.

Dr. Reibman was elected as an IEEE Fellow in 2005 for her contributions to video transport over networks. She was a member of the IEEE Awards Committee from 2010 to 2012. In 1998, she won the IEEE Communications Society Leonard G. Abraham Prize Paper Award. She was the Chair of the IEEE Fellow Committee from 2016 to 2017 and a Judge on the IEEE Fellow Committee from 2012 to 2015 and the IEEE Fellow Strategic Planning Committee from 2013 to 2019. She was the Technical Co-Chair of the IEEE International Conference on Image Processing in 2002. She was a Distinguished Lecturer for the IEEE Signal Processing Society from 2008 to 2009.