Journal of Humanitarian and Applied Sciences مجلة العلوم الإنسانية والتطبيقية
(رقم الإيداع المحلي - 90-2020)
كلية الآداب والعلوم قصر خيار - جامعة المرقب
JHAS
Volume 9 - Issue 18 المجلد 9 – العدد 18

The State-of-the-art Advancements and Challenges of 3D Video Formats

Khalid Mohamed Alajel
Associate Professor
Image Processing/Wireless Communication
Department of Computer Engineering,
Faculty of Engineering,
Elmergib University
kmalajel@elmergib.edu.ly

Mustafa Almahdi Algaet
Associate Professor
Image Processing/Medical Processing
Department of Networks
Faculty of Information Technology,
Elmergib University
malgaet@elmergib.edu.ly

Ali Ahmad Milad
Associate Professor
Image Processing/Network Simulation
Department of Networks
Faculty of Information Technology
Elmergib University
alimilad@elmergib.edu.ly

Received: 15/11/2025 | Accepted:17/12/2025

Abstract

3D video communication has become one of the most prominent ways of sharing and exchanging rich visual information, driven by the growing demand for immersive media in applications such as virtual reality, augmented reality, telepresence, and advanced broadcasting systems. To achieve this objective, the source coding should compress the original video sequence as much as possible, and the compressed video data should be robust and resilient to channel errors. In this respect, it is necessary to provide both efficient video formats and advanced compression standards capable of balancing bitrate reduction with high visual fidelity while including mechanisms that improve error resilience, granting reliable transmission over bandwidth-limited and error-prone communication channels. The aim of this paper is to provide a comprehensive survey of up-to-date formats, recent developments, and challenges concerning the efficient representation, compression, and delivery of immersive content in 3D formats.

Keywords :3D video, video coding, 3D video formats, video transmission.

أحدث التطورات والتحديات في تنسيقات الفيديو ثلاثية الأبعاد

خالد محمد العجيلي
أستاذ مشارك
معالجة الصور/إتصالات لاسلكية
قسم هندسة الحاسوب
كلية الهندسة، جامعة المرقب
kmalajel@elmergib.edu.ly

مصطفى المهدي القط
أستاذ مشارك
معالجة الصور/معالجة طبية
قسم الشبكات، كلية تقنية المعلومات
جامعة المرقب
malgaet@elmergib.edu.ly

علي أحمد ميلاد
أستاذ مشارك
معالجة الصور/محاكاة الشبكات
قسم الشبكات، كلية تقنية المعلومات
جامعة المرقب
alimilad@elmergib.edu.ly

الملخص

أصبح الإتصال المرئي ثلاثي الأبعاد من أبرز وسائل مشاركة وتبادل المعلومات البصرية الغنية، مدفوعةً بالطلب المتزايد على تطبيقات مثل الواقع الافتراضي، والواقع المعزز، والتواجد عن بُعد، وأنظمة البث المتقدّمة. ولتحقيق هذا الهدف، يجب

أن يضغط ترميز المصدر تسلسل الفيديو الأصلي قدر الإمكان، وأن تكون البيانات المضغوطة متينةً ومقاومةً لأخطاء القناة. لذا، يلزم وجود صيغ فيديو فعّالة ومعايير ضغط متقدّمة لتحقيق التوازن بين خفض معدّل البت ودقة الصورة العالية، مع دمج آليات تُعزز مقاومة الأخطاء وتدعم نقلًا موثوقًا به عبر قنوات إتصال محدودة النطاق وعرضة للأخطاء. تُقدّم هذه الورقة مسحًا متطورًا لتنسيقات الفيديو ثلاثية الأبعاد، وتُلخص المعايير الحالية والتطورات الأخيرة والتحدّيات المرتبطة بالتمثيل الفّعال والضغط وتقديم المحتوى الغامر.

الكلمات المفتاحية: فيديو ثلاثي الأبعاد، ترميز الفيديو، صيغ الفيديو ثلاثية الأبعاد، نقل الفيديو.

I. Introduction

3D video communication is an essential aspect of creating a visually enhanced experience for users and also can be used in applications where high-quality immersive experiences are needed. There have been many different formats, coding standards, and transmission techniques developed to provide secure reliable transmission of 3D video communication across bandwidth-constrained systems with a low amount of errors. All of these advances have allowed for the effective delivery of 3D video communication across channels that have little to no bandwidth or are prone to errors (Munteanu & Timmermann, 2021), (Quach et al., 2019). The rise in the popularity of 3D video formats over the past several years has been fuelled by the growing need for rich visual media experiences. Applications that have benefited from this growth include virtual reality, augmented reality, telepresence, and high-definition broadcasting technologies (Mendiburu, 2012), (Vetro et al., 2011). In addition to provide greater visual depth and realism than 2D images, 3D videos have spurred extensive research and development of better content compression methods, more reliable transmission systems, and improved rendering technologies (ISO/IEC, 2022), (Nightingale et al., 2018), (Schwarz et al., 2014).

This paper provides a thorough evaluation of 3D video formats through the synthesis of present literature with in-depth reviews of the techniques used to create 3D content. Each of the reviewed formats has both advantages and disadvantages and they show how these techniques are being incorporated into creating the newest robust and effective video communication formats (Mendiburu, 2012), (Vetro et al., 2011). The latest formats of 3D videos (stereoscopic, multi-view, depth-enhanced, and volumetric) provide superior viewing and depth perception but create additional difficulties in terms of large amounts of data, low compression efficiencies (i.e. large amounts to transmit), and reduced error tolerance, (ISO/IEC, 2021). Coding and compressing techniques to enhance bandwidth/digital capacity while maintaining acceptable visual quality are critical. By synthesizing all these perspectives, an overall consolidation of the latest trends and challenges facing the development of 3D video communications, as well as potential future directions for research and optimization of formats and transmission options, are attained (Nightingale et al., 2018), (Hosseini & Swaminathan, 2017). Furthermore, the paper discusses the challenges associated with secure transmission, compression efficiency, and error resilience in modern 3D video communication systems, offering insights into future research directions in both format development and transmission analysis (Munteanu & Timmermann, 2021) , (Quach et al., 2019), (Boyce et al., 2021).

Most of the techniques presented in previous studies were developed to enhance video coding efficiency, reduce bandwidth usage, and improve their ability to withstand channel transmission errors, all focusing on 2D video. 3D video communication builds upon this existing knowledge base while including new aspects and difficulties associated with the transmission of 3D video. Things like encoding multiple viewpoints, maintaining depth information, and maintaining stereoscopic or volumetric consistency have introduced many more issues into the space of 3D video communication (ISO/IEC, 2021). To address all these needs, new High-End 3D video formats have been created, and new coding standards for these formats have been developed (e.g., Stereoscopic, Multiview, Depth Size, and Volumetric) (Quach, et al., 2019), (Munteanu, et al., 2021).

A single consolidated and up-to-date resource for 3D video formats and associated techniques does not currently exist. Few reviews have even been published regarding 3D video formats; these are typically limited to discussing 'immersion' (3D content delivery) without providing any quantitative analysis of the 3D formats themselves which is unfortunate considering the increasing demand for 3D content. This current review aims to provide a more comprehensive and current overview of 3D video formats as well as explore the various technical methods associated with each type of format based on the wealth of published research that currently exists relating to these two areas of 3D video. What does make this review unique the scope in which it covers, also including some new neural representation (Mildenhall et al., 2020) or depth enhanced representation based formats, light field video (Ng et al., 2005) and hybrid encoding schemes (Schwarz et al., 2014). In contrast to previous reviews, the analysis in this document systematically compares each type of 3D video format for their compression efficiency, rendering requirements, interoperability, scalability and suitability for current immersive applications thereby providing a more comprehensive analysis for researchers looking to apply 3D video technology to create content.

1.1. Motivation

In this section, several review papers addressing this different 3D video formats are discussed for clearly illustration of the novelty of our work. To the best of the authors' knowledge, a comprehensive survey addressing the recent developments of 3D video formats, highlighting a major gap in the literature, does not currently exist. Thus, this current review is critical to providing a current and thorough examination of the current 3D video technologies along with their features and challenges related to compression, transmission and security (Munteanu & Timmermann, 2021), (Nightingale et al., 2018).

Established video formats like MPEG-2, H.264/AVC, H.265/HEVC, VP9, and others have been extensively studied in research on 2D video (Sullivan et al., 2012), (Wiegand et al., 2003), (Han, 2020). These formats have been under intensive investigation in terms of compression efficiency, rate–distortion performance, scalability, and adaptability across diverse delivery platforms. Consequent results have made the evolution of 2D video coding quite well-documented, with recognized benchmarks and widely adopted standards (Boyce et al., 2015). Quite the opposite is the case for research on 3D video formats, which to date is not adequately consolidated; stereoscopic, multiview, immersive, and neural-representation domains host quite scattered studies. This reflects a strong need for a holistic and comprehensive review of modern 3D video formats (Munteanu & Timmermann, 2021), (Mildenhall et al., 2020).

(Vetro et al., 2011) reviewed the algorithmic design adopted to extend H.264/MPEG-4 AVC toward Multiview Video Coding (MVC). They presented the essential approach of MVC, focusing on how interview prediction and view scalability can be enabled within the H.264/MPEG-4 AVC framework. In their literature review, stated that the MVC standard indeed improves the compression efficiency of stereo and multiview video due to support for both inter-view prediction and temporal inter-picture prediction. Unlike this prior work, our survey is not limited to stereo and multiview video formats; it rather extends to more recent and advanced state-of-the-art 3D video representations.

Additionally, (Merkle et al., 2010) reviewed the available 3D video formats for both video-only and depth enhanced 3D representations. An overview of existing and upcoming 3D video coding standards is also given. Despite, efficient standardized coding algorithms for video-only formats are available, the authors emphasize on the MPEG 3D video coding standardization, aiming at 2-3 view depth enhanced formats and support of advanced stereoscopic processing as well as future auto-stereoscopic displays. In our survey, we do not limit ourselves to video-only and depth enhanced 3D representations, but cover all technologies that make use of use of 3D scene capture, representation, and rendering, including modern neural and learning-based 3D video formats. (Alajel et al., 2017) presented early detailed review on existing techniques for 3D video formats and coding. The authors have surveyed state-of-the-art 3-D video formats and coding. Various types of 3-D video representation techniques were reviewed and the major 3-D video coding techniques and standards in the literature were discussed. In their early literature review, they came to the conclusion that, with 3D video coding standards that could be adopted or extended from 2D to 3D formats, which are integral in resolving these issues. They conclude these techniques are very promising for 3-D video transmission.

The review of (Kakkar & Ragothaman, 2024) introduces a thorough overview of the current state of research concerning volumetric video. They strive to provide a comprehensive overview of this fast-changing topic and outline some areas of possible future research, so as to more fully develop the vision of volumetric video. Not explicitly covering stereoscopic 3D in their work, they discuss the general benefits and various applications of volumetric representations. In contrast to their broad focus, our survey explicitly focuses on aspects related to immersive compression efficiency, rendering requirements, interoperability, scalability, and suitability of different formats for modern immersive applications. (Shafi et al., 2020) the authors' work surveys the technology and resources available for streaming 360-degree video. They present a wide variety of capture and display paradigms, with some examples being from the viewpoint of the equipment capturing the video to the media which is used to display the video. Additionally, the authors outline the many different ways of representing 360-degree video using different projection methods, compression schemes, and streaming methods based on either visual characteristics or spherical features of the video. While certainly summarizing these areas, the authors also identify the most important elements of 360-degree video, including technical hurdles and issues associated with using it in real world applications. We do not restrict ourselves to specific projects but give a more comprehensive overview of research in the field of research across the wider field of 3D video formats.

(Shi et al., 2025), this paper begins by outlining the process of video streaming, reviewing metrics relevant to its evaluation, and considering some key issues that intelligent solutions must address. It then discusses the workflow of intelligent enhancement in video

streaming, analyzes a few representative models for content enhancement, and highlights their distinctive characteristics. In this way, the authors lead the discussion from basic concepts to advanced knowledge of intelligent techniques that will result in better quality and more efficient video streaming. This survey (Wang et al., 2025) reviews the state-of-the-art in extended reality (XR) streaming, focusing on multiple paradigms. First, the authors define XR, introduce several XR headsets, and their multimodal interaction methods to provide a basic understanding. They also discuss aspects affecting the quality of experience in XR systems. Second, they examine the factors determining XR Quality of Experience (QoE) to ensure that systems meet user expectations for compelling, immersive experiences. (Zhao et al., 2024), reviews methodologies on action recognition, which are organized in a systematic manner regarding model architecture and input modality. This ranges from traditional techniques to RGB-based neural networks, skeleton-based models, and advanced pose estimation methods in order to extract skeletal data, hence providing structured and holistic insight into the field.

Clear evidence from the studies listed above suggests that most of the existing survey publications either provide very broad, high-level summaries of 3D video formats or target mainly transmission-related aspects. In contrast, this paper provides an extensive survey based on categorizing and analyzing the various techniques adopted in the digital representation of 3D videos. Consolidation of the scattered knowledge into a single comprehensive overview is essential for providing the community of researchers, academicians, industry professionals, and end users with an understanding of the current status and guidelines toward further development. The objectives of this paper are stated as follows:

- ✓ It has reviewed and analyzed the current frame-compatible stereo formats, underlining their principles, bandwidth efficiency, and practical deployment scenarios.
- ✓ To give an overview of current full-resolution stereo formats, with a special emphasis on their respective compression strategies, visual quality, and compatibility with established video coding standards.
- ✓ Review current virtual reality and immersive 3D formats, including omnidirectional and head-mounted display–oriented representations.
- ✓ This paper reviews state-of-the-art volumetric and light-field 3D formats with respect to their data structures, capture complexity, and rendering pipelines.
- ✓ To provide a comparative summary on the use of formats for compression efficiency, rendering requirements, interoperability, scalability, and suitability for modern immersive applications.
- ✓ To identify current trends, open challenges, and emerging research directions that will shape the future development of 3D video representation and delivery.

1.2. Organization of paper

The paper bring-sixth a detailed and thorough review of the state-of- the-art methods for 3D video, with a specific emphasis on their compression efficiency, rendering requirements, interoperability, scalability, suitability for modern immersive applications, and current state-of-the-art research focuses.

The paper's organization is as follows: The Frame-Compatible Stereo Formats techniques are mentioned in Section II. Section III discusses the Full-Resolution Stereo Formats. In addition, VR and Immersive 3D Formats techniques are described in Section IV.

Moreover, the Volumetric and Light-Field 3D Formats are mentioned in Section V. In Section VI, the current trends and research directions are given. Finally, conclusions of this comprehensive review are drawn in Section VII.

II. Frame-Compatible Stereo Formats

Stereo formats that are frame-compatible are an accepted means of providing stereoscopic 3D video as a separate format via traditional means of 2D video encoding and transmission. The left view and right view of an image are spatially multiplexed into one video frame allowing for compatibility with the major compression schemes such as H.264/AVC and H.265/HEVC and legacy television and streaming networks. The typical configurations for packing the video frames are Side by Side (SBS), in which the image frame is a horizontal squeeze of the left and right views, the other configuration is Top and Bottom (TaB), in which the left and right views are vertically stacked; there are many packing options for specialized applications; among them are interleaved and checkerboard formats (Vetro, 2010).

Since this type of packing design slot does not require dual-stream encoding or dedicated codecs, it has an inherent limitation of spatial resolution for each view, so that at least 50% of the original resolution has been lost. Thus, this loss of spatial resolution negatively impacts the ability of viewers to perceive depth and extreme fine detail. Due to the simplistic design of frame-compatible formats, they have an idealized application for transmission via real-time restricted bandwidth applications and 3D television and 3D images through various internet and broadcast systems. The continuing support of frame-compatible formats is also evident in that they serve to balance the real-world requirements of delivering and transmitting stereoscopic video content efficiently (Van Duc et al., 2021), (Pejman et al., 2024).

The packaging of 3D stereo video streams is done through layout configurations like side-by-side (SBS) and top-and-bottom (TaB), where views are horizontally or vertically subsampled to fit the standard dimensions of a video frame. Other designs, like line-interleaving and checkerboard patterns to provide a proper balance between maintaining resolution and compatibility with the decoder, are used in some niche applications. Frame-compatible formats remove the need for dedicated stereo codecs or two stream transmission capabilities. In frame-compatible video format the viewer's depth and sharpness perception is significantly less than in full-resolution stereoscopic imaging due to the fact that the two channel views are each rendered at only half the pixel size of a single video stream, but the added benefits of having a frame-compatible video format to easily distribute via existing infrastructure and to support multiple devices has made it the most widely used means of distributing Real-time 3D broadcast, online video and consumer display systems.

The two stereo views can be combined into a single coded frame via physical packing using any of the established layout patterns, including Side by Side (SbS), Top and Bottom (TaB), Line Interleaved, or Checkerboard (shown in figure 1). In all instances, the original view or some number of dimensions have been downsampled so that both corresponding images are able to fit in the frame's original resolution, resulting in a view at approximately 50% spatial resolution. By packing in this way, it is able to be processed normally with 2D video encoders, as well as utilized with standard streaming or broadcasting infrastructure without adding additional bandwidth burden. At the point of the display of this frame, an

appropriate 3D-capable device can remove this packing, thereby re-creating the left and right views to allow for stereoscopic viewing, and it also provides a backward compatibility with previous systems.

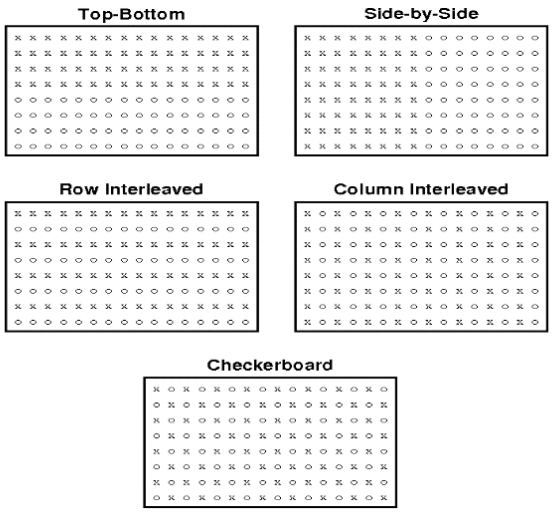

Fig. 1: Common frame-compatible formats where 'x' represents the samples from one view and 'o' represents the samples from the other view (Vetro, 2010).

Though frame-compatible formats are very useful and effective to use, they have multiple limitations due to being frame compatible by design. One limitation that affects the quality of the two 3D views displayed together as a single image is the reduction in the spatial resolution of each view, which creates a noticeable loss in quality when viewed on a large display or high-definition (HD) screen, particularly for portraying detail or depth cues that require a lot of attention. Additionally, the spatial proximity of the two views causes video codecs, which typically work by exploiting redundancies between views, to misinterpret the video data and treat views as if they are non-homogeneous signals, resulting in a reduction in the efficiency of compression and perceived quality. Thus, the types of stereo formats that can be used with current compression techniques for producing superior quality stereoscopic images will likely constrain both the quality of the stereoscopic representations and the amount of compression that can be performed on them. However, due to the simple structure of the stereo frame-compatible format, its low computational workload and capability of being integrated into existing infrastructures that deliver video, frame compatible formats are still considered an optimal solution for those who have limited bandwidth and/or will use them for producing 3D videos that are backward-compatible with 2D video systems.

Frame-compatible formats are most advantageous as they provide a way to utilize existing consumer equipment and video infrastructure to distribute stereoscopic 3D services without the need for new dedicated hardware or custom codecs, as they can be encoded and decoded the same way as standard video. Standard encoders can be used to compress these videos, and they can be transmitted via established broadcast and streaming channels, and decoded by legacy receivers without any changes. However, a significant disadvantage of these frame-compatible formats is that monoscopic devices do not interpret the stereo information contained within the stereo video. Instead, they may display the packed frame in its raw format (e.g., side-by-side) rather than extracting and reconstructing the intended

stereo pair. This illustrates the trade-off between universal compatibility and optimal viewing on non-3D capable devices.

2-1 Side-by-Side (SBS)

In Side by Side (SbS) frame compatibility formats, the two views left-eye and right-eye are contained together in the same frame, generally arranged horizontally (figure 2). However, there is no universal standard for how these views are ordered, so both configurations should be checked: "Cross-Eyed" (right view is on the left; left view is on the right) and "Parallel Pair" (left view is on the left; right view is on the right). If the image looks visually correct but produces discomfort when being viewed, it most likely means the erroneous view order has been selected. Generally speaking, images with the file extensions .JPS (JPEG stereo) and PNS (PNG stereo) are meant for versions of a cross-eyed view, while movies and other stereo images can use either configuration, depending on their source. Accurate ordering of the views is required to achieve both accurate depth perception and a comfortable stereoscopic viewing experience (Van Duc et al., 2024), (Vetro et al., 2011).

Fig. 2: The side-by-side 3D format (JVC Professional Video Visual Systems, 2025).

2-2 Top-and-Bottom (Over-Under)

In the Top-and-Bottom (TaB) layout, both left eye view and right eye view images are stored together in a single image or video; as such, they are stacked vertically (figure 3). As with Side-by-Side layouts, placing views in the proper order is important; therefore, if you see an image presented properly and everything appears normal but feels uncomfortable, your views are most likely assigned incorrectly. Because of how much frame space is taken up as well as how easy it is to use with equipment set up to process regular videos, the layout has become widely adopted by producers of 3D content created using stereoscopic techniques (Vetro et al., 2011), (Tripathi et al., 2011).

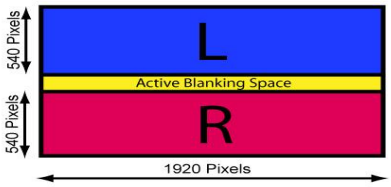

Fig. 3: The Top and Bottom or Over-Under 3D format (JVC Professional Video Visual Systems, 2025).

2-3 Interleaved Row/Column

When the left eye and right eye are represented in an interleaved row/column configuration, there are alternating horizontal or vertical lines representing left and right images. They are interlaced on a line by line basis, creating an alternating effect. Viewing such images on standard 2D displays or without the use of a dedicated stereoscopic viewer will frequently yield an image that appears to have a great deal of noise or confusion, due to having both images mixed together. This format is most commonly used by systems that

utilize optical separation for stereoscopic viewing, e.g., certain types of three-dimensional monitors or projectors, and allows for depth perception without losing the single frame representation. (Ju et al., 2025). To demonstrate the advantages of using an interleaver, consider when the 16-QAM modulation is used with an input sequence length of 8000 symbol inputs; interleaver matrix can be represented as a 4 × 2000 matrix. The gain from using an interleaving function is seen clearly by the increase in coding performance shown by figure 4 (Xiong et al., 2021).

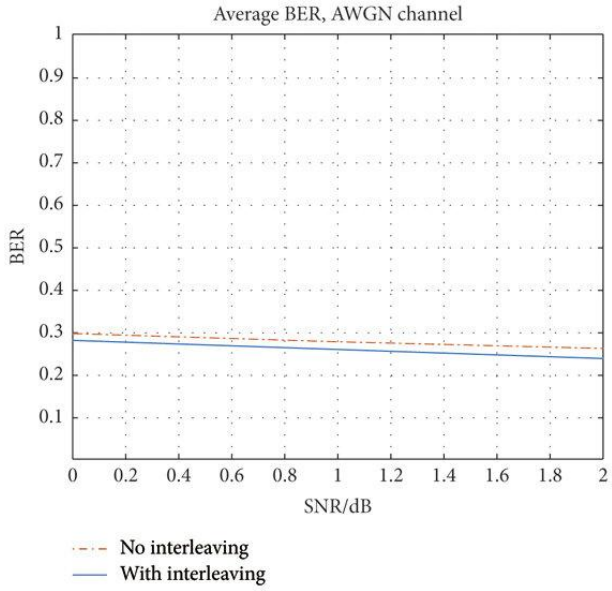

Fig. 4: Comparison of BER before and after interleaving (Xiong et al., 2021).

2-4 Checkerboard

A way to encode left and right eye images in a checkerboard pattern (checkerboard) is to alternate pixels, this provides an optical balance of maintaining the resolution of an image and providing enough vertical separation between two images so that the user can view them in 3D (figure 5). The disadvantages of the checkerboard format are similar to the interleaved format; the resulting left and right image will appear severely distorted and noisy when displayed using a traditional 2-D display technology. (Chiang et al., 2012) proposes a stereo packing scheme using checkerboard subsampling in order to combine the left and right views into one frame for efficient encoding and transmission under conventional video coding standards (for example, H.264/AVC).

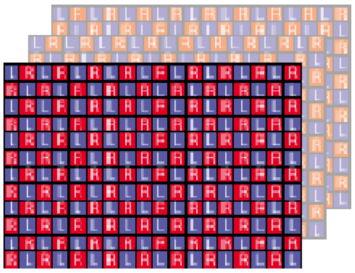

Fig. 5: The checkerboard 3D format (JVC Professional Video Visual Systems, 2025).

All of these frame-compatible formats effectively reduce the resolution of each view approximately half of the original frame size, thus they can fit both views comfortably within the confines of their original size frame without creating an increase in the amount of bandwidth to transmit/store these images. The fact that they work with current video encoders, all transmission channels, and all existing 3D compatible legacy devices make these formats easy to implement and provide less degradation of perception, i.e. lower sharpness and fidelity of depth perception; regardless of that, there is a large market for using these formats because of their simplicity, minimal amount of required storage space for implementation, and ease in using them for live broadcasts, streaming, and in the creation and use of 3D consumer digital displays.

III. Full-Resolution Stereo Formats

Full-resolution stereo formats refer to stereoscopic 3D video representations that have a full-resolution image for each eye. By comparison to frame-compatible formats (e.g., side-by-side half, top-and-bottom), where both views have been reduced in spatial resolution to fit into one frame, full-resolution stereo formats retain all the fidelity of the left and right views. This provides higher-quality images, stronger perception of depth, and greater comfort with viewing images, particularly on high-end displays and in professional uses.

3-1 Dual-Stream Full Resolution

The Stereoscopic 3D video format (Dual Stream Full Resolution), represents an essentially independent pair of full resolution video files that were simultaneously recorded and processed from both the left and right eyes. Both files retain their respective spatial resolutions (e.g., 1080p, 4K, or higher) figure 6, which maximizes image fidelity when viewing them in stereo.

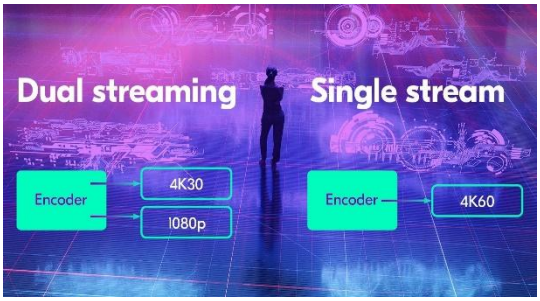

Fig. 6: Dual streaming (Ofek, 2024).

(Liu et al., 2023) Presents the process of no-reference quality evaluation of stereoscopic images on a dual-stream network, using two parallel streams to process stereo information. This paper is on a no-reference stereoscopic image quality assessment that makes use of a dual-stream network; hence, relevant as it uses two parallel streams for the processing of stereo information. (Cao et al., 2011) shows that it is possible to achieve a hybrid camera system for recording video with both high spatial resolution and high spectral resolution by integrating an RGB camera with high spatial resolution and a multispectral grayscale camera with high spectral resolution using an efficient propagation algorithm for the result. It is evident from the experiments that the system is able to provide useful high-resolution multispectral video that facilitates various computer vision tasks like dynamic

white balance adjustment, object tracking, among other tasks that RGB cameras cannot provide.

3-2 Multiview Video Coding (MVC) – H.264 Extension

Multiview Video Coding (MVC) builds upon the H.264/AVC standard by providing a more effective method for encoding multiple camera views, including those used for stereoscopic 3D and other multiview video applications. MVC provides a technique called inter-view prediction that allows encoders to utilize the redundancy between different camera views, allowing for a much lower bitrate when encoding more than one view at the same time than if they were encoded independently as illustrated in figure 7 (Merkle et al., 2007), (Mendiburu, 2012). MVC was first established in ISO/IEC 14496-10 (part of MPEG 4 AVC) and ITU-T H.264 and was the technology behind Blu-ray 3D and similar high definition 3D technology (ISO/IEC, 2014), (Mendiburu, 2012).

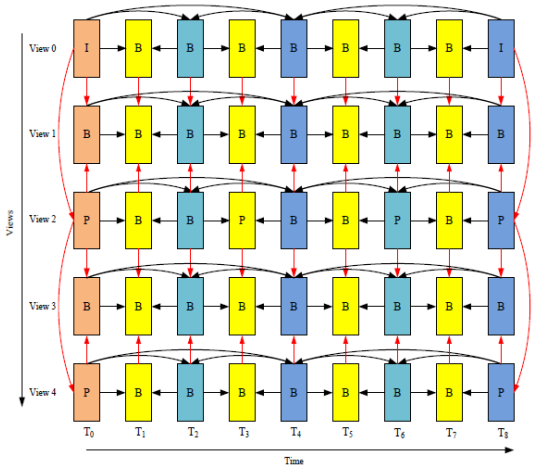

Fig. 7: Multiview coding structure with temporal/interview prediction (Mendiburu, 2012).

(Liu et al., 2022) mainly presented a two-stream interactive network for no-reference stereoscopic image quality assessment, jointly exploiting local and global information. By parallelly designing local and global streams for feature extraction and enabling interactions between them, the model captures fine-grained distortion details and the overall perceptual structure of stereoscopic image pairs.

3-3 MV-HEVC (H.265 Multiview)

Multiview HEVC (MV-HEVC) is an extension of the standard H.265/HEVC that supports efficient coding of a multiscopic video environment in stereoscopic 3D and multi-camera settings. MV-HEVC builds on top of HEVC by including inter-view predictions which allow for dependent views to use previously decoded frames from other views. The result is a much lower bit rate than if each view were independently coded while preserving full spatial resolution across all views (Chen et al., 2017). These additional inter-view reference pictures are emphasized in the multiview prediction structure shown in figure 8, with each view representing a different layer. The left view does not depend on other layers, so it is fully HEVC compatible, whereas the right view depends on the left view.

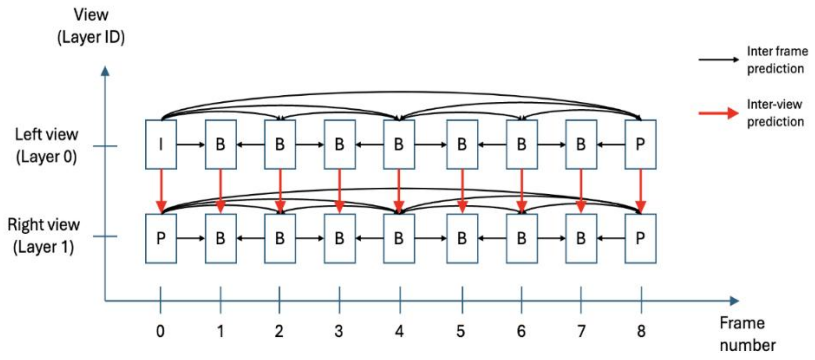

Fig. 8: Typical multiview prediction structure for stereo video.

Another important aspect of the design of MV-HEVC is backward compatibility; the first encoded view is referred to as the base view and is coded in such a way that any decoder that complies with HEVC can decode it. The second and subsequent views are coded as a dependent layer and share the same predictive coding structure as the base view. Compared to the MVC (Multi-view Coding or H.264 version), MV-HEVC improves on MVC by adding many of the improvements in HEVC such as improved motion compensation, larger coding units, and better in-loop filtering. This results in a more efficient way of compressing video and better scalability to business applications (Cao et al., 2025).

(Li et al., 2024) demonstrate that message embedding based on motion vector these differences (MVD) or MVP index disrupts the local optimality of motion vector prediction, and this disruption can be used as a robust steganalysis feature. They define the optimal rate of MVP as a one-dimensional feature that perfectly distinguishes between cover and stego videos under most tested conditions. Further, their method requires no machine learning training, exhibits low computational complexity, and performs efficiently in practical scenarios.

3-4 VVC (H.266) Multiview Extensions

The next generation of multiview (stereoscopic) video compression is the Versatile Video Coding (VVC/H.266) Multiview Extensions. They provide a robust multiview video compression framework that expands on VVC's current level of efficiency to include NEW multi-camera, immersive, and free-viewpoint video applications. The VVC multiview architecture is based on the same layered video coding structure as MVC (H.264) and MV-HEVC (H.265). It is comprised of a full resolution base layer view and several dependent layer views. The dependent layers utilize both temporal predictions and inter-views references to help reduce the bitstream size compared to the full video codec (VVC) (Mingyuan et al., 2026).

The VVC Multiview Extensions offer greater compression efficiency than MV-HEVC due to the addition of tools such as affine motion compensation, decoder-side motion refinements, triangular prediction units, advanced block partitioning, and improved In-Loop filtering provided in H.266. With these additional capabilities, VVC Multiview has shown very significant improvements in the compression of high-resolution stereoscopic video,

with substantial benefits when used with large numbers of cameras and dynamic, free-viewpoint applications (Bull et al., 2021).

(Chlubna et al., 2026) proposes a focus-aware compression framework specialized for 3D displays without glasses. It leverages the fact that multiple views seen simultaneously create out-of-focus areas due to visual blending. The authors introduce new objective visual quality metrics and an automatic method to detect optimal focusing distance from input 3D views. Using this focus information, out-of-focus regions can be compressed more aggressively or have their high-frequency content reduced (via depth-of-field effects), improving compression efficiency with minimal perceived quality loss**.**

Table 1: Comparison of MVC (H.264), MV-HEVC (H.265), and VVC Multiview (H.266).

Feature / Aspect	MVC (H.264 Multiview)	MV-HEVC (H.265 Multiview)	VVC Multiview (H.266 Multiview)
Standard	H.264/AVC Extension	H.265/HEVC Extension	H.266/VVC Extension
Base Layer Compatibility	Base view is a valid H.264 stream	Base view is a valid HEVC stream	Base view is a valid VVC stream
Dependent Views	Inter-view predicted using MVC tools	Inter-view predicted using HEVC motion and transform tools	Inter-view predicted using advanced VVC tools (affine MC, decoder-side refinement, geometric partitioning)
Coding Efficiency	Lowest among the three	30–40% better than MVC	30–50% better than MV-HEVC (depending on configuration)
Prediction Tools	Temporal + basic inter-view prediction	Improved temporal prediction, efficient motion compensation, SAO filtering	Affine motion, intra-block copy, MIP, triangular prediction units, improved in-loop filters
Scalability (Views)	Limited; suitable mainly for stereo	Improved multi-view support	Designed for large-scale multiview camera arrays
Strengths	Backward compatibility; simple architecture	Better compression; efficient for stereoscopic and multiview	Best efficiency; scalable; supports advanced immersive and free-viewpoint use cases
Limitations	Lower efficiency; aging standard	Higher complexity, limited industry use	Very high complexity; hardware support still emerging

IV. VR and Immersive 3D Formats

Specialized video formats are used by virtual reality (VR) and immersive 3D Systems to display a wide field of view (FOV) with accurate depth perception and the ability for users to interact in 6 degrees of freedom (6DoF). VR has unique virtual video formats that allow for head-tracked rendering. They also support high spatial resolution, ultra-low latency and an efficient method for mapping spherical and volumetric scene content. The current immersive format landscape ranges from monoscopic 360 video to Stereoscopic 360 VR to advanced light-field representations and volumetric representations enabling free viewpoint navigation (Shafi et al., 2026), (Kim et al., 2020), (Vadakital et al., 2022).

4-1 Stereoscopic VR180

Designed specifically for providing high quality stereo 3D immersive visuals within a specified 180 for Main Viewing Directions, (MVD), Stereoscopic VR180 is a unique video

format that facilitates the 3D stereoscopic experience. Unlike traditional VR formats where the viewer can rotate around themselves, this allows for capturing images from a central point towards the main focal area (180). Since this stereo VR format provides an unparalleled level of visual detail within the area in front of the viewer, it allows for greater image resolution and a greater depth perception; thereby requiring significantly less bandwidth and computational processing than an equivalent full 360x360 format (Lavrushkin et al., 2021).

(Sassatelli et al., 2020) propose new ways to manage limited network resources as they relate to 360 deg. video. Rather than using traditional methods of adapting the amount of compression or other measures based on the amount of available resources, the authors provide two other forms of interaction-based interference, specifically, Virtual Walls (VWs) and Slow Downs (SDs). These alternative approaches allow for a decrease in the rate at which content is delivered by reducing the amount of data required for each user while still allowing users' visual experiences to remain intact.

4-2 Stereoscopic 360° VR

360° stereoscopic VR is an immersive video format providing a full-spherical surround that creates an authentic two-eye depth perception experience to any direction or area viewed by a viewer while simultaneously viewing stereoscopic 3D footage. Stereoscopic 360° VR differs from VR180's focus on forward viewing, allowing users to access all eight directions (up/down) as well as having access to all of the space around the viewer for both 360° 6DoF (rotation and some limited translation) interactivity when utilized with either depth view synthesis or depth view synthesis combined with scene content creation and/or depth view synthesis combined with view-synthesis creation capabilities.

A stereoscopic 360° scene representation that supports head-motion parallax to enhance immersion in VR has proposed in (Luo et al., 2018). (Artois et al., 2023) proposes a system that augments the conventional 360° video with depth information to achieve an immersive VR experience. Using the depth information to recreate the image as a 3D representation, it provides the user with the capability for motion parallax and real depth perception by head movement, whereas the other approaches for 360° video support viewing only with rotation (3DoF) in monoscopic or stereoscopic settings. It tackles the issue of rendering hidden areas by using inpainting techniques for a seamless experience with the capability of adding more virtual objects. Experimental results indicate that the newly proposed system is highly effective in providing an immersive VR experience with real depth perception. (Pirker et al., 2021) offers a literature review of using 360° Virtual Reality (VR) Videos and Full Interactive Virtual Reality sessions in the educational setting, as well as their potential for benefits and drawbacks. It is shown that 360° Virtual Reality Videos can improve learning by adding to learner motivation, engagement, presence, perception, and empathies over other video forms despite the lack of direct evidence that it provides benefits of enhanced learning through technological means such as better educational results or learning retention.

4-3 Cubemap 3D

A cubemap 3D stereoscopic and monoscopic image format lays the spherical view of the complete scene into six sides of a cube that enables efficient representation and rendering of the 360° content as illustrated in figure 9. Cubemaps eliminate much of the geometric

distortion associated with spherical projection and thus provide better spatial uniformity than using spherical projection for the virtual reality (VR), augmented reality (AR), and immersive video applications (Budagavi et al., 2016).

(Chieh et al., 2021) proposes a region-level bit allocation scheme tailored for rate control in 360-degree video coding using cubemap projection. The approach first detects high HEVC coding cost regions on each face of the cubemap using machine-learning based features, namely texture, motion magnitude, motion density, temporal coherence. Then, a surface-fitting based bit allocation function will be applied in assigning bits between high-cost and nonhigh-cost regions. Experimental results show that this method improves bitrate accuracy and BD-WS-PSNR compared with the original R-model rate control in HEVC.

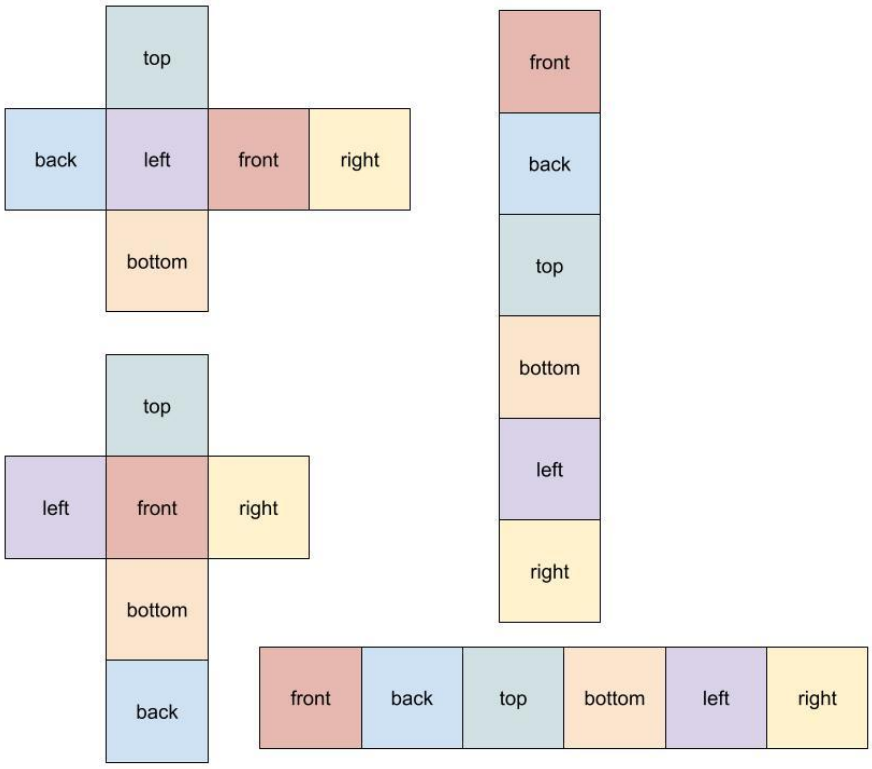

Fig. 9: cubemap layout (Trek View, 2025).

4-4 Apple Immersive Video

Apple Immersive Video is Apple's proprietary immersive media format, designed primarily for the Apple Vision Pro headset. It delivers stereoscopic 3D content with a wide field of view (≈180°), high resolution (up to 8K), high frame rates, and spatial audio, allowing viewers to feel fully "inside" the scene. The format supports multiview or dual-layer encoding (using MV-HEVC principles), where each eye receives a high-fidelity video layer, and metadata enables accurate depth rendering and head-tracked spatial audio (HEVC stereo video, Apple Inc., 2023), (Chen et al., 2017).

Table 2: Comparison of Immersive Video Formats

Feature / Aspect	VR180	Stereoscopic 360° VR	Light-Field / Volumetric	Apple Immersive Video
Field of View	~180° forward-facing	360° spherical	Full 360° or arbitrary viewpoints	~180° forward-facing
Degrees of Freedom (DoF)	3DoF / limited 3DoF+	3DoF / 3DoF+	6DoF (full free-viewpoint)	Primarily 3DoF+ (forward-facing)
Resolution	High per-eye resolution	Very high, doubles data per eye	High per view, depends on sampling density	Up to 8K per eye
Encoding	Frame-compatible SBS/TAB, H.264/H.265	SBS/TAB, equirectangular/cubemap, H.265/VVC	Point-cloud compression (PCC), MVD, VVC	MV-HEVC based multiview layers (proprietary workflow)
Rendering / Display	VR headset	VR headset / mobile VR	VR/AR, holographic displays	Apple Vision Pro (exclusive)
Bandwidth / Data Efficiency	Moderate	High	Very high	High (optimized for forward-facing 180°)
Applications	Immersive video, storytelling, training	Full VR environments, events, documentaries	Telepresence, 6DoF VR/AR, holograms	Cinematic storytelling, immersive films, concerts, travel
Advantages	Efficient, high-quality forward-facing depth	Full environment immersion	Realistic free-viewpoint navigation	Premium cinematic quality, high-fidelity stereoscopy, spatial audio
Limitations	Limited to forward view	Bandwidth-heavy, computational load	Very high complexity, specialized hardware	Proprietary, Apple ecosystem, high production requirements

V. Volumetric and Light-Field 3D Formats

3D Video created in either volumetric or light-field format is an advanced representation of 3D video where not only does it contain the geometry (X/Y/Z Coordinates), Depth (Distance from Camera) and Appearance (Color, Texture or Material) of every individual object within the Scene, but also allows for a 6 Degrees-of-Freedom (6DoF) Interactive Experience. In contrast with Stereoscopic or Multiview Videos, which allow you to view Objects from just one fixed point of view, Volumetric or Light-field Videos allow Users to move around freely within the Virtual Scene to experience realistic motion parallax, Occlusion, or View-dependent effects (Kerbl et al., 2023), (ISO/IEC, 2025).

5-1 MPEG-I MIV (MPEG Immersive Video)

The MPEG Immersive Video (MIV) is a standardised framework created by MPEG to encode, transmit and render interactive video contents; it includes stereoscopic, multiview and volumetric videos. The MIV specification is part of the MPEG-I (Immersive) series of specifications and provides scalable, interoperable solutions to the future of immersive media applications, including VR, AR, free-viewpoint video and telepresence (Kerbl et al., 2023).

(Vadakital et al., 2022), presents an overview of the MPEG Immersive Video (MIV) standard developed within the MPEG-I framework, aiming to enable efficient representation

and compression of immersive video content supporting six degrees of freedom (6DoF). The core concept of MIV is based on multi-view video plus depth (MVD) representations combined with geometry and occupancy information. Instead of transmitting full volumetric data, the standard encodes a selected set of camera views and associated depth maps, which are then used at the decoder to synthesize intermediate virtual views.

5-2 MPEG OMAF (Omnidirectional Media Format)

OMAF by MPEG – the Moving Picture Experts Group is a framework developed by ISO for distributing 360-degree videos and performing audio-visual (AV) Media Delivery. It has created an interoperable format to allow for both supply and playback of Omnidirectional media (which are media produced in such a way as to cover all directions or all around). The purpose of creating OMAF was to allow for the efficient decoding, rendering and usages of OMAF formatted media through devices that use HMDs (Head-Mounted Displays) and or OMAF-compliant VR players to allow for the best experience of watching 360-degree video and experience live and on-demand VR Video experiences with such media (Vadakital et al., 2022).

5-3 Point Cloud Compression (V-PCC / G-PCC)

Point Cloud Compression (PCC), introduced in MPEG's (ISO/IEC 23090-9) is a standard that encodes 3D point clouds for the efficient storage and transmission of this data. A point cloud is made up of individual points that represent the geometrical representation (XYZ) of an environment or object, as well as additional attribute information (colour, reflectance, etc). The application of PCC is vital to the success of volumetric video, holographic displays, 6DoF Virtual/Augmented Reality (VR/AR), and Immersive Telepresence, where large amounts of raw point cloud data exist and therefore cannot be stored or transmitted without becoming impractical. (Zhang et al., 2024), article on the "Current Development of MPEG Geometry-based Point Cloud Compression (G-PCC) Edition 2" provides an overview of the history of developing and finalizing G-PCC, in conjunction with recent developments on the Edition 2 standardization process. The focus of the article is to provide a review of how the Edition 2 has improved the efficiencies of both static and dynamic point cloud compression as well as the new features added to allow for additional types of attributes (color, reflectance, etc.) that can now be compressed.

VI. Key challenges and future research directions for 3D video formats

Technical and practical challenges pertaining to 3-D video formats include high data rates, computational complexity, and interoperability between different devices and platforms. Effective compression, streaming, and rendering techniques are in strong need for emerging immersive formats such as volumetric video, multiview video, and neural representations that will guarantee premium quality with limited bandwidth consumption or latency. Future research is foreseen to be concentrated on hybrid representations, learning-based compression and rendering methods, perceptually optimized quality metrics, and adaptive streaming strategies responsive to viewer focus or device capabilities.

1 - High-level trends

- **Shift from per-view video to volumetric / neural representations.** Traditional multi-view/left-right stereoscopic formats are being complemented or replaced in many research and industry efforts by volumetric scene representations (point clouds, meshes + textures, multi-plane/light-field, and neural fields such as NeRFs). These allow 6DoF experiences and better free-viewpoint rendering (Vadakital et al., 2022).
- **Standards and industry consolidation around MPEG families (MIV, V-PCC, G-PCC).** MPEG's Immersive Video (MIV) and point-cloud compression standards provide practical, interoperable formats that are now mature enough for experiments and early deployments (ISO/IEC, 2023).
- **Neural representations + hybrid pipelines.** Neural Radiance Fields (NeRF) and derivatives (and newer techniques like 3D Gaussian Splatting) have rapidly advanced novel-view synthesis and compact scene encoding; but practical systems often combine neural and classical elements. (Kerbl et al., 2023), Shows how to move neural radiance representations toward real-time rendering using 3D anisotropic Gaussians (significant practical speedups vs classic NeRF). Useful if you care about real-time/interactive pipelines. (Mildenhall et al. 2020) classic paper that started modern neural-field view synthesis. Read for the core idea (ML model that maps 3D location + view direction → density + radiance), evaluation methodology, and the baseline for almost all later neural scene work.
- **From offline capture to real-time and streaming.** Research focus is moving toward lower-latency capture, on-the-fly compression, and streaming (adaptive bitrates for viewpoint changes) to enable live volumetric/6DoF experiences. Standards and experiments are explicitly addressing streaming constraints (Vadakital et al., 2022).

2. Major technical advances (state-of-the-art) (2025)

- **MPEG Immersive Video (MIV):** a practical standard that wraps pre/post-processing around conventional codecs to support limited 6DoF immersive playback enabling interoperability and industry uptake. Useful baseline for experiments and deployments. The MPEG specification for MIV (6DoF-limited immersive playback). Read this to understand practical interoperability, the expected bitstream model, and how industry wraps 3D/2D components for deployable systems (Mildenhall et al. 2020).
- **Point-cloud compression (V-PCC / G-PCC):** mature methods for coding dense and sparse point clouds respectively; they make point-cloud streaming feasible over networks and are the de-facto standards in many demos and trials. V-PCC maps dense point clouds to 2D patches and leverages existing video codecs — a pragmatic approach used for dense, camera-like captures. Good for streaming-focused implementations and for comparing projection-based compression tradeoffs (ISO/IEC, 2025)
- **Neural fields and Gaussian splatting:** NeRFs gave huge gains in quality for novel-view synthesis; Gaussian Splatting and related work drastically speed up rendering and make neural methods more practical for near-real-time visualization. Reviews show rapid progress in dynamic NeRFs (handling motion, temporal coherence) (Bao et al., 2024).

- **Learned compression / ML post-processing:** applying neural networks to compress, denoise, and enhance compressed point clouds/light-fields — improving visual quality at lower bitrates (e.g., learning color correction after V-PCC) (Gao et al., 2022).

3 - Core technical challenges (open problems)

- **Rate-distortion vs. interactivity tradeoff.** High-quality volumetric representations are large. Compressing them while keeping the ability to rapidly change viewpoint (low latency) is hard. Streaming systems must trade per-view quality for bandwidth and time-to-first-render (ISO/IEC, 2023).
- **Real-time capture and representation conversion.** Converting multi-camera/fisheye captures into compact point clouds / neural fields fast enough for live use remains challenging (alignment, temporal consistency, and reconstruction speed) (Lin, 2024). **Temporal stability and dynamic scenes.** Neural methods (NeRF) originally targeted static scenes; making them robust for dynamic, deforming, or specular scenes (people, cloth, reflections) while maintaining efficiency is active research (Lin, 2024).
- **Perceptual metrics and QA.** Objective metrics for perceived quality in 6DoF/viewpoint-varying environments are immature — we need perceptual metrics that account for viewpoint changes, motion, and occlusions (ISO/IEC, 2023).
- **Interoperability and toolchain complexity.** Multiple formats (point clouds, mesh+texture, multi-plane images, neural fields, MIV wrappers) mean complex toolchains; moving between these reliably and efficiently is nontrivial (MPEG Experts, 2025).

VII- Research directions and open problems

1. **Real-time NeRF/3DGS pipelines for live capture.**
   - Goal: reduce end-to-end latency from capture to render for dynamic scenes (people, small groups). Evaluate tradeoffs in fidelity vs latency.
   - Why: moves neural methods from offline to live. Use dynamic-NeRF literature as starting point (Lin, 2024).
2. **Learned compression for hybrid representations.**
   - Goal: design codecs that combine V-PCC/G-PCC with neural residuals — a base classical codec plus learned enhancement layer that is compact and streamable.
   - Why: practical path to better rate-distortion with existing standards (Lin, 2024).
3. **Perceptual 6DoF quality metrics and benchmark.**
   - Goal: build a benchmark dataset with multi-view captures, user studies, and a metric that correlates with subjective quality across viewpoints and motion.
4. Why: enables fair comparison across codecs and representations; current metrics (PSNR, SSIM) are insufficient (ISO/IEC, 2023).
5. **Efficient temporal compression for dynamic point clouds.**
   - Goal: exploit temporal redundancy across frames in V-PCC/G-PCC pipelines or via learned motion-compensation for point clouds.
   - Why: big bitrate savings for streaming moving scenes.

- G-PCC targets sparse / LiDAR-style point clouds using native 3D structures (octrees, etc.). Essential reading when geometry sparsity, scalability, and low-overhead storage matter (Lin, 2024).

6. **Robust view synthesis for challenging materials (specular, translucent).**
   - Goal: integrate physics-aware rendering priors or learn specularity models into neural fields to correctly synthesize shiny/translucent objects.
7. **Interoperability and converter toolkits.**
   - Goal: create reliable, open-source toolchains to convert between camera feeds → MIV / V-PCC / NeRF / 3DGS with reference implementations and performance baselines.
   - Why: broad adoption needs simple tools and reproducible pipelines (Vadakital et al., 2022).

VIII. Conclusion

To summarise, the rapid advancement of 3D video telecommunication channels has made it necessary to implement an efficient, effective method for representing, compressing, and subsequently transmitting this type of video over medium suitable for use with modern immersive apps. The growing uses of VR, AR, telepresence technology, and advanced broadcast technologies have created an increased demand for the ability to deliver high-quality visuals, yet consume fewer bitstreams while continuing to be resilient to failures in the transmission medium(s). At the same time, while new emerging 3D video representations and compression standards will create additional complexity in the use of 3D video technology by introducing new types of compression, which they will require digital forensics techniques such as steganography to locate any hidden information in an extremely compressed multimedia stream. This survey provides an overview of the current state of the field, identifies areas of need, and provides directions for future research on how best to address the developing need for more efficient, secure, and reliable 3D video telecommunication technology.

References

1. Alajel, K. M., Abusabee, K. M., and Tamtum, A. (2017). 3D video formats and coding: A review. *International Journal of Engineering Science Invention, 6*(2), 26–36.
2. Apple Inc. (2023). *ISO base media file format and Apple HEVC stereo video*. Apple Developer Documentation.
3. Artois, J., Van Wallendael, G., and Lambert, P. (2023, January). 360DIV: 360° video plus depth for fully immersive VR experiences. In *Proceedings of the IEEE International Conference on Consumer Electronics (ICCE)* (pp. 1–2). Las Vegas, NV, United States.
4. Bao, J., Liu, Y., Li, Z., Zhu, S., and Yeung, S.-K. A. (2024, December). Color enhancement for V-PCC compressed point cloud via 2D attribute map optimization. In *Proceedings of the IEEE International Conference on Visual Communications and Image Processing (VCIP)* (pp. 1–5). Tokyo, Japan.
5. Boyce, J., *et al.* (2021). *Error resilience and robust transmission in immersive video streaming*. *IEEE Access.*
6. Boyce, J., Ye, Y., Li, J., and Seregin, V. (2015). Overview of SHVC and MV-HEVC. *IEEE Transactions on Circuits and Systems for Video Technology.*

7. Budagavi, M., Furton, J., Saxena, A., and Wang, J. (2016). 360 degrees video coding using cubemap projection. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)* (pp. 378–382). Phoenix, AZ, United States.
8. Bull, D. R., and Zhang, F. (2021). Video coding standards and formats. In *Intelligent image and video compression* (2nd ed., pp. 435–484). Academic Press.
9. Cao, M., Tian, L., and Li, C. (2026). A HEVC video steganalysis algorithm for transform unit partition modes. *Expert Systems with Applications, 297*(A).
10. Cao, X., Tong, X., Dai, Q., and Lin, S. (2011). High resolution multispectral video capture with a hybrid camera system. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 297–304). Colorado Springs, CO, USA.
11. Chen, J., Müller, K., Ohm, J.-R., Vetro, A., and Wang, Y. (2017). An overview of multiview high efficiency video coding. *IEEE Transactions on Circuits and Systems for Video Technology, 27*(10), 2124–2140.
12. Chen, J., Müller, K., Ohm, J.-R., Vetro, A., and Wang, Y. (2017). An overview of multiview high efficiency video coding. *IEEE Transactions on Circuits and Systems for Video Technology, 27*(10), 2124–2140.
13. Chiang, A.-T., Wang, H.-M., Yang, J.-F., and Wang, J.-F. (2012). A new stereo packing format based on checkerboard sub-sampling for efficient stereo video coding. In *2012 IEEE International Symposium on Circuits and Systems (ISCAS)* (pp. 385–388). Seoul, Korea (South).
14. Chlubna, T., Vlnas, M., Bařina, D., Milet, T., and Zemčík, P. (2026). Focus aware compression and image quality metric for 3D displays. *Signal Processing, 238*, 91–110.
15. Dziembowski, A., Lafruit, G., Thudor, F., Lee, G., and Alface, P. R. (2022). The MPEG immersive video standard—Current status and future outlook. *IEEE MultiMedia, 29*(2), 90–99.
16. Gao, K., Gao, Y., He, H., Lu, D., Xu, L., and Li, J. (2022). NeRF: Neural radiance field in 3D vision: A comprehensive review. *arXiv preprint arXiv:2210.00379*.
17. Han, O. (2020). Performance evaluation of VP9 and H.265 video codecs. *IEEE Access, 8*,
18. Hosseini, M., and Swaminathan, V. (2017). View-aware tile-based adaptations in 360 virtual reality video streaming. In *Proceedings of the 2017 IEEE Virtual Reality (VR)* (pp. 423–424).
19. ISO/IEC. (2014). *Information technology—Coding of audio-visual objects—Part 10: Advanced Video Coding* (ISO/IEC 14496-10).
20. ISO/IEC. (2021). *Geometry-based point cloud compression (G-PCC)* (ISO/IEC 23090-9).
21. ISO/IEC. (2021). *Video-based point cloud compression (V-PCC)* (ISO/IEC 23090-5).
22. ISO/IEC. (2022). *MPEG immersive video (MIV)* (ISO/IEC 23090-12).
23. ISO/IEC. (2022, February). *Information technology—Coded representation of immersive media—Part 9: Point cloud compression (PCC)* (ISO/IEC Standard 23090-9). International Organization for Standardization.
24. ISO/IEC. (2023). *Information technology—Coded representation of immersive media—Part 12: Immersive video* (ISO/IEC Standard 23090-12). International Organization for Standardization.
25. ISO/IEC. (2025, March). *Information technology—Coded representation of immersive media—Part 5: Visual volumetric video-based coding (V3C) and video-based point cloud compression (V-PCC)* (ISO/IEC Standard 23090-5). International Organization for Standardization.
26. ISO/IEC. (2025, March). *Information technology—Coded representation of immersive media—Part 5: Visual volumetric video-based coding (V3C) and video-based point cloud compression (V-PCC)* (ISO/IEC Standard 23090-5). International Organization for Standardization.

27. Ju, Y., Hu, J., Luo, Z., Deng, H., Zhao, H., Du, L., Wu, C., Hao, D., Wang, X., and Pan, T. (2025, July). CI-VID: A coherent interleaved text-video dataset.
28. JVC Professional Video Visual Systems. (n.d.). *Technical description for model MDL101911 – Feature 02*. JVC Pro. Retrieved December 14, 2025, from http://pro.jvc.com/prof/attributes/tech_desc.jsp?model_id=MDL101911andfeature_id=02
29. Kakkar, P., and Ragothaman, H. (2024). The evolution of volumetric video: A survey of smart transcoding and compression approaches. *International Journal of Computer Graphics and Animation, 14*(1–4), 1–11.
30. Kerbl, T., Kopanas, G., Leimkühler, T., and Drettakis, G. (2023). 3D Gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics, 42*(4), 1–14.
31. Kim, J., Kim, K., and Kim, W. S. (2022). Impact of immersive virtual reality content using 360 degree videos in undergraduate education. *IEEE Transactions on Learning Technologies, 15*(1), 137–149.
32. Lavrushkin, S., Molodetskikh, I., Kozhemyakov, K., and Vatolin, D. (2021). Stereoscopic quality assessment of 1,000 VR180 videos using 8 metrics. In Proceedings of the ISandT International Symposium on Electronic Imaging: Stereoscopic Displays and Applications XXXII (pp. 350-1–350-7).
33. Li, J., Zhang, M., Niu, K., Zhang, Y., and Yang, X. (2024). A HEVC video steganalysis method using the optimality of motion vector prediction. *Computers, Materials and Continua, 79*(2), 2085–2103.
34. Lin, J. (2024). Dynamic NeRF: A review. *arXiv preprint arXiv:2405.08609*.
35. Liu, B., Liu, X., Dai, A., Zeng, Z., Wang, D., Cui, Z., and Yang, J. (2023). *Dual stream diffusion net for text-to-video generation*.
36. Liu, Y., Huang, B., Yue, G., Wu, J., Wang, X., and Zheng, Z. (2022). Two-stream interactive network based on local and global information for no-reference stereoscopic image quality assessment. *Journal of Visual Communication and Image Representation, 87.*
37. Luo, B., Xu, F., Richardt, C., and Yong, J. H. (2018). Parallax360: Stereoscopic 360 scene representation for head motion parallax. In *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (pp. 343–351).
38. Mendiburu, L. (2012). *3D movie making: Stereoscopic digital cinema*. Focal Press.
39. Mendiburu, L. (2012). *3D movie making: Stereoscopic digital cinema*. Focal Press.
40. Merkle, P., Müller, K., and Wiegand, T. (2010, July). 3D video coding – An overview of present and upcoming standards. In *Proceedings of the IEEE Visual Communications and Image Processing (VCIP)*, Huangshan, China.
41. Merkle, P., Smolic, A., Müller, K., and Wiegand, T. (2007). Efficient prediction structures for multiview video coding. *IEEE Transactions on Circuits and Systems for Video Technology, 17*(11), 1461–1473
42. Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., and Ng, R. (2020). NeRF: Representing scenes as neural radiance fields for view synthesis. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 405–421). Glasgow, United Kingdom.
43. Mildenhall, B., Srinivasan, P., Tancik, M., Barron, J. T., Ramamoorthi, R., and Ng, R. (2020). NeRF: Representing scenes as neural radiance fields for view synthesis. In *Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IV* (pp. 405–421). Springer.
44. Mildenhall, B., Srinivasan, P., Tancik, M., Barron, J. T., Ramamoorthi, R., and Ng, R. (2020). NeRF: Representing scenes as neural radiance fields for view synthesis. In *Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IV* (pp. 405–421). Springer.

45. MPEG Experts. (n.d.). *MPEG Experts Area — Documents, datasets, conformance, software and reference resources*. Retrieved December 15, 2025.
46. Munteanu, C. F., and Timmermann, D. (2021). Volumetric video representations: A survey. *IEEE Transactions on Multimedia*. Advance online publication.
47. Munteanu, C. F., and Timmermann, D. (2021). *Volumetric video representations: A survey*. *IEEE Transactions on Multimedia*.
48. Munteanu, C. F., and Timmermann, D. (2021). *Volumetric video representations: A survey*. *IEEE Transactions on Multimedia*.
49. Munteanu, C. F., and Timmermann, D. (2021). *Volumetric video representations: A survey*. *IEEE Transactions on Multimedia*.
50. Munteanu, C. F., and Timmermann, D. (2021). *Volumetric video representations: A survey*. *IEEE Transactions on Multimedia*.
51. Ng, R., Levoy, M., Brédif, M., Duval, J., Horowitz, M., and Hanrahan, P. (2005). *Light field photography with a hand-held plenoptic camera* (Tech. Rep.). Stanford University.
52. Nien, Y.-C., and Tang, C.-W. (2021). Region-level bit allocation for rate control of 360-degree videos using cubemap projection. *Journal of Visual Communication and Image Representation, 79*, 103206.
53. Nightingale, J., Emmanouilidou, D., Quddus, A., and Ghanbari, M. (2018). A survey of error resilience techniques for video streaming. *ACM Computing Surveys, 51*(5), 1–38.
54. Nightingale, J., Emmanouilidou, D., Quddus, A., and Ghanbari, M. (2018). A survey of error resilience techniques for video streaming. *ACM Computing Surveys, 51*(5), 1–38.
55. Nightingale, J., Emmanouilidou, D., Quddus, A., and Ghanbari, M. (2018). A survey of error resilience techniques for video streaming. *ACM Computing Surveys, 51*(5), 1–38.
56. Ofek, E. (2024, May 21). *Understanding dual streaming in the AVoIP realm*. KramerAV. Retrieved December 15, 2025, from https://www.kramerav.com/content-hub/blog/understanding-dual-streaming-in-the-avoip-realm/
57. Pejman, H., Coulombe, S., Vázquez, C., Jamali, M., and Vakili, A. (2024, June). A novel region-dependent packing method for stereoscopic 360° videos using horizontal downsampling of equirectangular projection. In *Proceedings of the Picture Coding Symposium (PCS)* (pp. 1–5). Taichung, Taiwan.
58. Pirker, J., and Dengel, A. (2021). The potential of 360 degree virtual reality videos and real VR for education: A literature review. *IEEE Computer Graphics and Applications*.
59. Quach, M., Valenzise, G., and Dufaux, F. (2019). *Learning convolutional transforms for lossy point cloud geometry compression*. In *Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP)* (pp. 4320–4324). IEEE.
60. Quach, M., Valenzise, G., and Dufaux, F. (2019). Learning convolutional transforms for lossy point cloud geometry compression. In *2019 IEEE International Conference on Image Processing (ICIP)* (pp. 4320–4324).
61. Quach, M., Valenzise, G., and Dufaux, F. (2019). Learning convolutional transforms for lossy point cloud geometry compression. In *2019 IEEE International Conference on Image Processing (ICIP)* (pp. 4320–4324).
62. Sassatelli, L., Winckler, M., Fisichella, T., Dezarnaud, A., Lemaire, J., Aparicio-Pardo, R., and Trevisan, D. (2020). New interactive strategies for virtual reality streaming in degraded context of use. *Computers and Graphics, 86*, 27–41.
63. Schwarz, S., Sjöström, M., and Olsson, R. (2014). A combined pre-processing and scalable 3D video coding scheme. *IEEE Transactions on Multimedia, 16*(3), 665–676.
64. Schwarz, S., Sjöström, M., and Olsson, R. (2014). A combined pre-processing and scalable 3D video coding scheme. *IEEE Transactions on Multimedia, 16*(3), 665–676.
65. Shafi, R., Shuai, W., and Younus, M. U. (2020). 360 degree video streaming: A survey of the state of the art. *Symmetry, 12*(9).

66. Shafi, R., Shuai, W., and Younus, M. U. (2020). 360 degree video streaming: A survey of the state of the art. *Symmetry, 12*(9), 1491.
67. Shi, W., Li, Q., Yu, Q., Wang, F., Shen, G., Jiang, Y., Xu, Y., Ma, L., and Muntean, G. M. (2025). A survey on intelligent solutions for increased video delivery quality in cloud–edge–end networks. *IEEE Communications Surveys and Tutorials, 27*(2).
68. Sullivan, G. J., Ohm, J.-R., Han, W.-J., and Wiegand, T. (2012). Overview of the high efficiency video coding (HEVC) standard. *IEEE Transactions on Circuits and Systems for Video Technology, 22*(12), 1649–1668.
69. Trek View. *Projection types in 360 photography*. Trek View Blog. Retrieved December 15, 2025.
70. Tripathi, S., Piccinelli, E. M., and Aliprandi, D. (2010). H.264/AVC stereo video compression benchmarking. In *11th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 10)* (pp. 1–4). Desenzano del Garda, Italy.
71. Vadakital, V. K. M., Dziembowski, A., Lafruit, G., Thudor, F., Lee, G., and Alface, P. R. (2022). The MPEG immersive video standard—Current status and future outlook. IEEE MultiMedia.
72. Van Duc, P., Tin, P. T., Le, A. V., Nhan, N. H. K., and Elara, M. R. (2021). Inter-frame based interpolation for top–bottom packed frame of 3D video. *Symmetry, 13*(4), Article 702.
73. Vetro, A. (2010). *Frame compatible formats for 3D video distribution* (Technical Report No. TR2010-099). Mitsubishi Electric Research Laboratories (MERL).
74. Vetro, A., Wiegand, T., and Sullivan, G. J. (2011). Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard. *Proceedings of the IEEE, 99*(4), 626–642.
75. Vetro, A., Wiegand, T., and Sullivan, G. J. (2011). Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard. *Proceedings of the IEEE, 99*(4), 626–642.
76. Vetro, A., Wiegand, T., and Sullivan, G. J. (2011). Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard. *Proceedings of the IEEE, 99*(4), 626–642.
77. Wang, H., Dong, H., and El Saddik, A. (2025). Immersive multimedia communication: State of the art on eXtended reality streaming. *ACM Transactions on Multimedia Computing, Communications, and Applications, 21*(7), 1–33.
78. Wiegand, T., Sullivan, G. J., Bjøntegaard, G., and Luthra, A. (2003). Overview of the H.264/AVC video coding standard. *IEEE Transactions on Circuits and Systems for Video Technology, 13*(7), 560–576.
79. Xiong, X., Dai, Y., Hu, Z., Huo, K., Bai, Y., Li, H., and Liu, D. (2021). Hardware sharing for channel interleavers in 5G NR standard. *Security and Communication Networks, 2021*, 1–13.
80. Zhang, W., Yang, F., Xu, Y., and Preda, M. (2024). Standardization status of MPEG geometry-based point cloud compression (G-PCC) Edition 2. In *Proceedings of the 2024 Picture Coding Symposium (PCS)* (pp. 1–5). Taichung, Taiwan.
81. Zhao, L., Lin, Z., Sun, R., and Wang, A. (2024). A review of state of the art methodologies and applications in action recognition. *Electronics, 13*(23).