International Journal of Engineering Science Invention
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org ||Volume 6 Issue 2|| February 2017 || PP. 26-36

3-D Video Formats and Coding- A review

Khalid Mohamed Alajel[1], Khairi Muftah Abusabee[1], Ali Tamtum[1]
[1](Electrical and Computer Engineering Department, Faculty of Engineering / Al-Mergib University, Libya)

Abstract: *The objective of a video communication system is to deliver the maximum of video data from the source to the destination through a communication channel using all of its available bandwidth. 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. However, while achieving a high coding efficiency, compression also makes the coded video bitstream vulnerable to transmission errors. Thus, the process of video data compression tends to work against the objectives of robustness and resilience to errors. Therefore, extra information that needs to be transmitted in 3-D video has brought new challenge and consumer applications will not gain more popularity unless the 3-D video coding problems are addressed.*

Keywords: *2-D video; 3-D video; 3-D Formats; 3-D video coding.*

I. Introduction

A digital video sequence consists of images, which are known as frames. Each frame consists of small picture elements, pixels that describe the color at that point in the frame. To describe fully the video sequence, a huge amount of data is required. Therefore, the video sequence is compressed to reduce the amount of data to make possible transmission over channels with limited bandwidth. "Fig. 1" describes a two-dimensional (2-D) video transmission system, where the encoder is used to compress the input sequence before the transmission over the channel. The reconstructed video sequence at the decoder side contains distortion introduced by the compression and the distortion in the channel respectively.

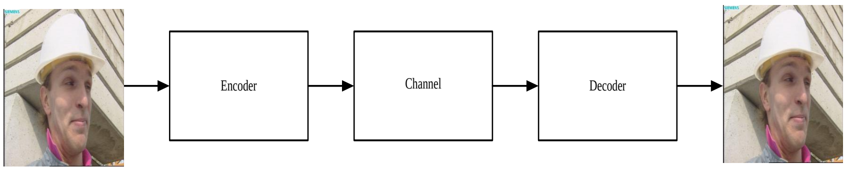


Fig. 1: 2-D video transmission system.

Video data creates a tremendous amount of data that needs to be transmitted or stored. The huge amount of data is a heavy burden for both transmission and decoding processes. Therefore, video data needs to be compacted into a smaller number of bits for practical storage and transmission. Source coding is the first important part in a communication system chain. The objective of this part is to remove the redundancy in the source as much as possible. Although there are many different categories of source coding techniques, depending on the source information itself, this section will focus on 2-D image and video coding techniques widely used in the recent international video standards.

There are currently many data compression techniques used for different purposes in video coding. One compression method employs statistical and subjective redundancy. Statistical redundancy can be efficiently compressed using lossless compression, so that the reconstructed data after compression are identical to the original video data. However, only a moderate amount of compression is achievable using lossless compression. In subjective redundancy, elements of video sequence can be removed without significantly affecting the visual quality. As a result, much higher compression is achievable at the expense of a loss of visual quality. The compression method employs both statistical and subjective redundancy which forms the basis of current video standards. Generally, most of all the recent video techniques used in today's video encoders are based on exploiting both temporal and spatial redundancy in the original video data (see "Fig. 2"). The following paragraph will describe the background of predictive video coding.

In spatial redundancy, there is a high correlation between successive frames of video. The process of removing redundancy within a frame is called intraframe coding. On the other hand, in temporal redundancy, there is a high correlation between pixels (samples) that are close to each other. The process of removing redundancy between frames is called interframe coding. Redundancy reduction is used to predict the value of pixels based on the values previously coded and code the prediction error. This method is called differential pulse code modulation (DCPM). Most of video coding standards such as MPEG-1 [4] MPEG-2 [5], MPEG-4 [6] by the moving picture experts group (MPEG) of the inter-national organization for standardization (ISO), and H.261 [7], H.263 [8], and H.264 [1] by the video coding expert group (VCEG) of international telecommunication union-telecommunication (ITU-T), employ a predictive coding system and variable-length code (VLC) techniques which are the root cause of error propagation.

Fig. 2: Spatial and temporal correlation of video sequence.

A new field in signal processing is the representation of three-dimensional (3-D) scenes. Interest in 3-D data representation for 3-D video communication has grown rapidly within the last few years. 3-D video may be captured in different ways such as stereoscopic dual-camera and multi-view settings. Since 3-D video formats consist of at least two video sequences and possibly additional depth data, many different coding techniques have been proposed [2, 3].

II. Three-Dimensional (3-D) Video

A three-dimensional (3-D) video system is able to offer to the user a sense of “being there” and thus provide a more impressive and realistic experience than two-dimensional (2-D) video. Recently, 3-D video has received increased attention due to the recent advances in capturing, coding and display technologies and it is anticipated that the 3-D video applications will increase rapidly in the near future. 3-D video system is able to offer to the user a depth perception of the observed scene. Such 3-D depth perception can be achieved by special 3-D display systems which allow 3-D visual data to be viewed by the user with each eye. There exist a variety of ways to represent 3-D content, such as conventional stereo video, multiview video and video-plus-depth [9]. As a consequence, there are a variety of compression and coding algorithms that are available for the different 3-Dvideo formats [10, 11]. In general, the additional dimension that the 3-D video provides results in tremendous amount of data that needs to be transmitted or stored. Consequently, there is a significant increase in the complexity of the whole 3-D video transmission system.

2-1 Human 3-D Visual System

Understanding how the human visual system (HVS) [12] works is crucial to understanding how 3-D imaging works. The HVS consists of two parts, the two eyes and the brain. Each eye has a retina that collects information and transfers it to a region of the brain called lateral geniculate body and then to the visual cortex through the optic nerve. The pictures produced at each of the retinas are one up-side-down and as the pieces of visual information are processed by the visual cortex, one single upright image is produced. As the two human eyes of an individual are separated by about 6-8 cm, the 3-D depth perception is realized by two slightly different images projected to the left and right eye retinas (binocular parallax) and then the brain fuses the two images to give the depth perception (See "Fig.3").

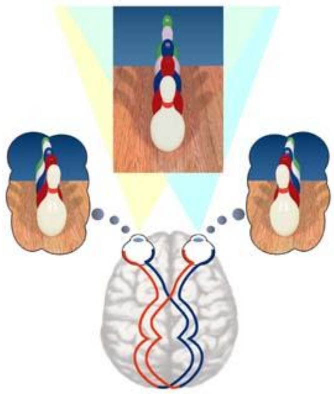


Fig.3: Human 3-D visual system.

Although the binocular parallax is the most dominant cue for depth perception, natural scenes contain a wide variety of visual cues known as monocular depth cues to determine depth. Monocular depth cues do not require the observer to have two eyes to perceive depth. Instead the HVS still uses several monocular depth cues such as motion parallax, relative size, and occlusion. The two eyes (binocular) are still the most important and widely used depth cues which provide enough information for the HVS. The binocular disparity is available because of the slight differences between the left and right eye points of view [13].

III. 3-D Video Formats And Coding

The contemporary interest in 3-D technology is now widespread and is manifested in different applications, including the 3-D cinema [16], 3-D video [17], and mobile phones [18]. Depending on the application, various choices of 3-D video formats are available. According to Merkle *et al*. [15], 3-D video formats can be presented in the following formats: conventional stereo video (CSV), video-plus- depth (V+D), multiview video (MVV), multiview video-plus-depth (MVD), and layered depth video (LDV). In this section, these formats are going to be briefly described along with their associated coding methods. The ballet 3-D video sequence [19, 20] will be used to illustrate these formats.

Conventional stereo video (CSV) is considered the least complex 3-D video format and it is a special case of multiview (2 views only). In CSV, the 3-D video consists of two videos (views) representing the left and right views of the same scene with slight difference in the angle of view corresponding to the distance of separation of the human eyes. Each view forms a normal 2-D video and the human brain can fuse these two different frames to generate the sensation of depth in the scene being viewed. "Fig.4" illustrates the CSV formats.


Left view


Right view

Fig. 4: CSV formats.

Since both cameras capture essentially the same scene, a straight-forward approach is to apply the existing 2-D video coding schemes. Using the 2-D video coding approach, the two separate views can be independently encoded, transmitted, and decoded with a 2-D video codec like H.264/AVC. This method is known as simulcast coding. However, since the two views have similar content, and therefore are highly redundant, coding efficiency can be increased by combined temporal/interview redundancy. This coding method is called multiview coding (MVC) [21, 22]. To achieve this goal, a corresponding standard has been defined in H.262/MPEG-2 multiview profile [23] as illustrated in "Fig. 5". The left view is encoded independently using MPEG-2 codec and for the right view, interview prediction is allowed in addition to temporal prediction. However, the gain in compression efficiency provided in the two views stereo video coding is limited compared to individual coding of each view. Some other coding methods are using view interpolation to compensate from camera geometry [24].

In CSV, the amount of data is twice that of 2-D video. Another alternative method for coding CSV data is called mixed resolution stereoscopic (MRS) coding [25]. In this method, the resolution of CSV data is downsampled to one fourth of its original resolution. Thus, a lower bit rate is achieved at equal quality. This makes the approach attractive for mobile devices [25]. MRS coding is illustrated in "Fig. 6" [26].

3-1 Video-plus-depth (V+D) format

One of the most popular formats for representing 3-D video is video-plus-depth (V+D), which consists of a conventional 2-D video with an associated per-pixel depth map represented with luma component only. For video and depth information, a stereo pair can be synthesized at the decoder. With this technique left and right views are generated at the display side by a method known as DIBR [14, 27]. The depth map represents the per-pixel distance from the camera and it is between Z_{near} = 255 and the maximum Z_{far} = 0, indicating the distance of the corresponding 3-D point from the camera, where the near objects appear brighter and the far objects appear darker. The V+D format is illustrated in "Fig.7".

Efficient coding of video-plus-depth format is necessary for mobile video services, due to its bandwidth and processing power limitations, for realizing 3-D video. For coding V+D format, both MPEG-2 and H.264/AVC can be used. If MPEG-2 is used, MPEG-C part 3 defines a video-plus-depth representation which allows encoding video and depth data as conventional 2-D video [28]. The video and depth sequences are encoded independently, where one view is transmitted simultaneously with the depth signal. The other view is synthesized by DIBR techniques at the receiver side. In this case, the transmission of a depth map increases the required bandwidth of 2-D video stream by about 20% [29].

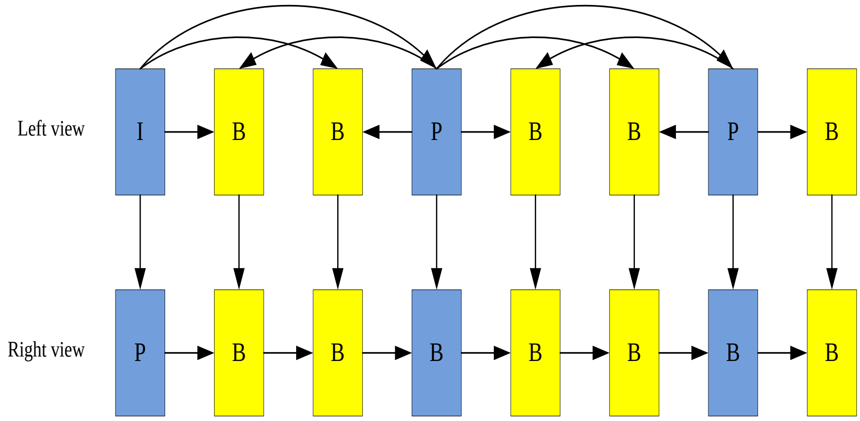
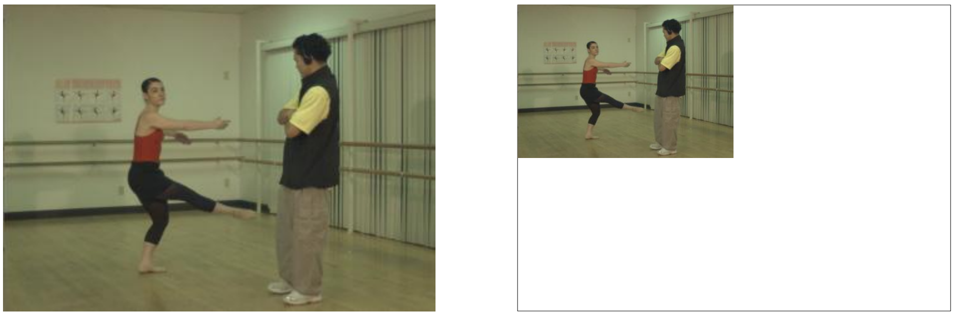


Fig. 5: Combined temporal and interview prediction for stereo coding.



Left view　　Right view

Fig. 6: Right view downsampling for MRS.

Color video　　Depth data

Fig. 7: V+D format.

If the H.264/AVC is used, the H.264 codec is applied to both sequences simultaneously but independently, where the video is the primary coded picture and the depth is the auxiliary coded picture. In this case, the required bandwidth increases by only 8% as mentioned by [1, 29]. The following coding standards are applicable to the video-plus-depth format, namely MPEG-C 3, H.264/AVC, H.264/MVC.

3-2 MPEG-C PART 3

The video-plus-depth format has been standardized within MPEG by a joint effort of Philips and Fraunhofer Heinrich Hertz Institute (HHI). The new standard has been finalized at the MPEG meeting in Marrakech, Morocco (January 2007). According to (ISO/IEC 23002-3), MPEG-C part 3 was presented for standardization of the video-plus-depth coding which allows encoding the depth maps as conventional 2-D video. Due to the very nature of the depth data, higher coding efficiency of depth data could be achieved than the video data which results in small extra needed bandwidth for transmitting the depth data. Thus, the total bandwidth required for video-plus-depth is reduced compared to that of stereo video.

MPEG-C part 3 is combined with H.264/AVC for coding video-plus-depth as illustrated in "Fig. 8". H.264/AVC is used to encode the video and depth data sequences independently. The two coded bitstreams are interleaved in the multiplexer frame-by-frame resulting in one stream for transmission. The demultiplexer separates the transmitted stream back into two bitstreams which are then decoded independently using the H.264/AVC decoder after transmission over wireless channels. This technique has been adopted in [30].

3-3 H.264/AVC

For coding video-plus-depth format using H.264/AVC, the auxiliary picture syntax specifies that extra monochrome pictures must be sent with the video stream. The monochrome picture must contain the exact number of macroblocks as the primary picture. Thus, auxiliary coded pictures should have the same syntactic and semantic restrictions. The overview diagram in "Fig. 9" illustrates the coding procedure of H.264/AVC for color-plus-depth format. The depth and video sequences are interlaced line by-line into one sequence, where the top field contains the video data and the bottom field the depth data. H.264/AVC coder is applied to both sequences simultaneously but independently where the video is the primary coded picture and the depth the auxiliary coded picture, resulting in one coded bit-stream. After transmission, this stream is decoded resulting in the distorted video and depth sequences. However, with this approach the backward compatibility is not supported.

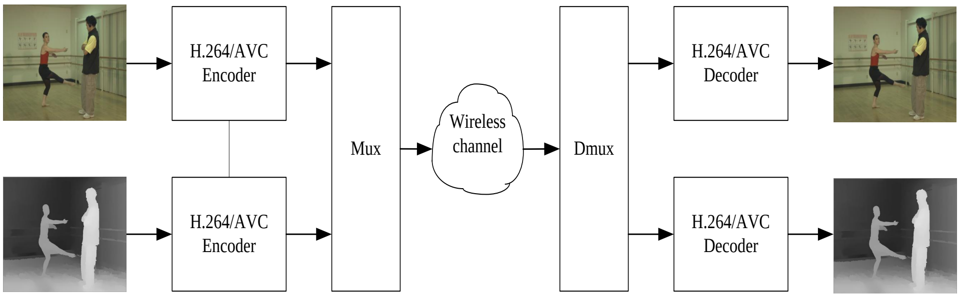


Fig. 8: Block diagram of MPEG-C part 3 coding for video-plus-depth representation.

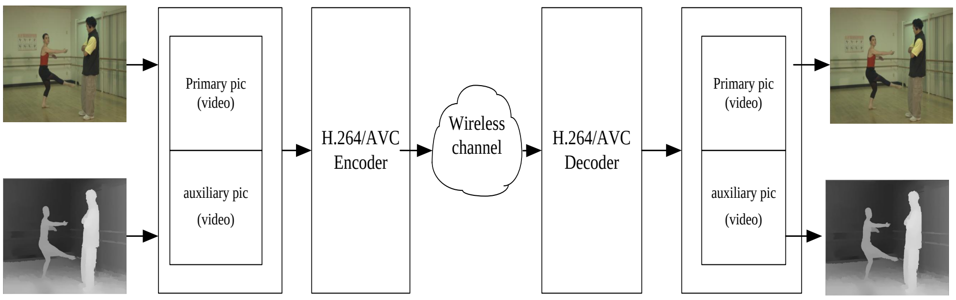


Fig. 9: H.264/AVC coding for video-plus-depth representation.

3-4 H.264/MVC

In multiview video coding, the picture can have temporal and interview prediction, respectively. "Fig. 10" shows the MVC coding process for video-plus-depth data. Interview predictive coding is applied through the H.264/AVC encoder for both sequences. Since the H.264/MVC combines temporal and interview prediction, thus, the input sequences must be with identical resolution. The advantage of this method is the backward compatibility.

By exploiting the depth data features, however, higher coding efficiency can be achieved. For instance, the existing correlation between the 2-D video sequence and its corresponding depth map sequence can be exploited to improve the compression ratio as proposed by [31, 32]. Alternative approaches based on so-called Platelets were also proposed [33]. The V+D concept is highly interesting due to the backward compatibility and the use of the available video codec. This format is alternative to CSV for mobile 3-D services and is being investigated by Fraunhover Institute for telecommunications. However, the advantages of V+D format come at the cost of increased encoder/decoder complexity [34].

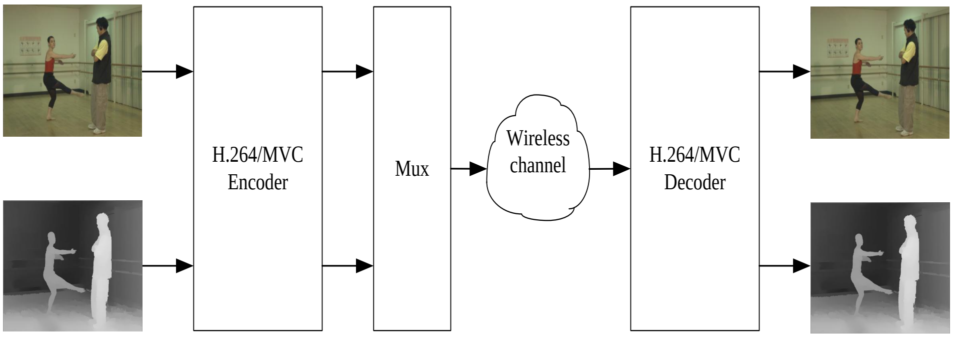


Fig. 10: H.264/MVC coding for video-plus-depth format.

3-5 MULTIVIEW VIDEO (MVV) FORMAT

One drawback of stereo video is that it only provides 3-D from one direction, whereas the HVS has the ability to see different parts of objects if the head is moved. Multiview can provide all the necessary depth cues and is considered one of the most promising techniques for 3-D video. For more than two views of the CSV this is easily extended to multiview video (MVV) [10, 35]. Transmission of a huge amount of data is the major challenge with multiview video applications, which require a high coding efficiency scheme. In MVV, *N* cameras are arranged to capture the same scene from different view-points. Therefore, they all share common scene contents.

The straight-forward method to encode multiview video is a Simulcast coding where, each view is coded independently. Simulcast coding can be done with any video codec including H.264/AVC where, the temporal and spatial correlation within one view is exploited. However, multiview video contains a large amount of interview statical dependencies which can be exploited for combined temporal/interview prediction. The multiple correlations makes multiview video coding have a different structure from single view, where the images are predicted temporally from neighbouring images within the same view and also from corresponding images in adjacent views, as illustrated in "Fig. 11". Significant gain can be achieved by combining temporal/interview prediction as proposed by [36, 37].

In July 2008, H.264/MVC standard [38] was specified as an extension to H.264/AVC. H.264/MVC uses the intra prediction for each view to reduce interview dependency. At the same time, it applies interview prediction from neighbouring views to every 2nd view using previously encoded frames from adjacent views, as depicted in "Fig. 11". Several researchers addressing that interview/temporal prediction technique efficiency exploited statical redundancy in multiview video data [39, 41]. Among them, algorithms that are based on hierarchical prediction (B) are proposed. This structure outperforms the simulcast coding by 20% of coding efficiency as reported in [42, 43]. According to Merkle *et al*. [43], H.264/AVC and hierarchical B-frames have been shown to achieve the highest coding efficiency. As H.264/MVC combines temporal and interview prediction, the identical resolution of the input video sequence is required.

Although this approach enhances the coding efficiency of multiview video, its drawback is increased complexity. To address this issue of complexity, one solution is to allow interview prediction only at key frames, which slightly reduces the coding efficiency compared to the one using key frames for all frames. However, as shown by Merkle *et al*. in [43], in the case of sparsely positioned cameras, interview prediction may not have any impact on coding efficiency while the complexity of the encoder is reduced substantially. For more details of MVC, the reader is referred to [35, 44].

3-6 MULTIVIEW PLUS DEPTH (MVD) FORMAT

Transmitting all views requires a high bit rate where, the number of views increases the bit rate linearly. Therefore, MVC is inefficient if the number of views to be transmitted is large. At the same time, V+D format provides a very limited free viewpoint video (FVV) functionality. The solution to the problem of high bit rate when transmitting all views, and the limited FVV, is multiview plus depth (MVD) format. MVD format contains multiple views and associated depth information for each view as illustrated in "Fig.12".

Multiview plus depth format is an extension of V+D and is included by MPEG in recent proposals [45, 46]. In MVD, depth has to be estimated for the *N* views and then *N* color with *N* depth videos have to be encoded and transmitted. MVD video sequence can be coded using methods for multiview video coding where, the depth image is estimated for each view of the multiview videos. In coding depth map sequences, Fehn *et al*.

[14] showed that the depth data contains only 10-20% of the data in color sequence. Many algorithms have been proposed for coding MVD such as [3, 47]. The coding of MVD has been improved by using Platelet-based depth coding as shown in [33].

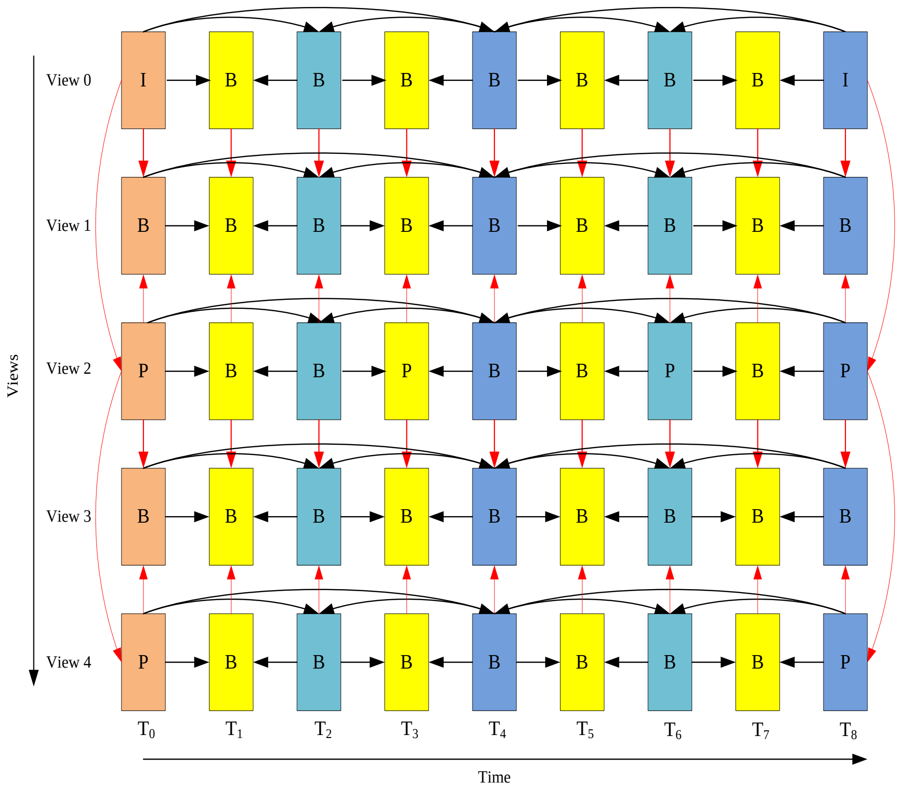


Fig. 11: Multiview coding structure with temporal/interview prediction.

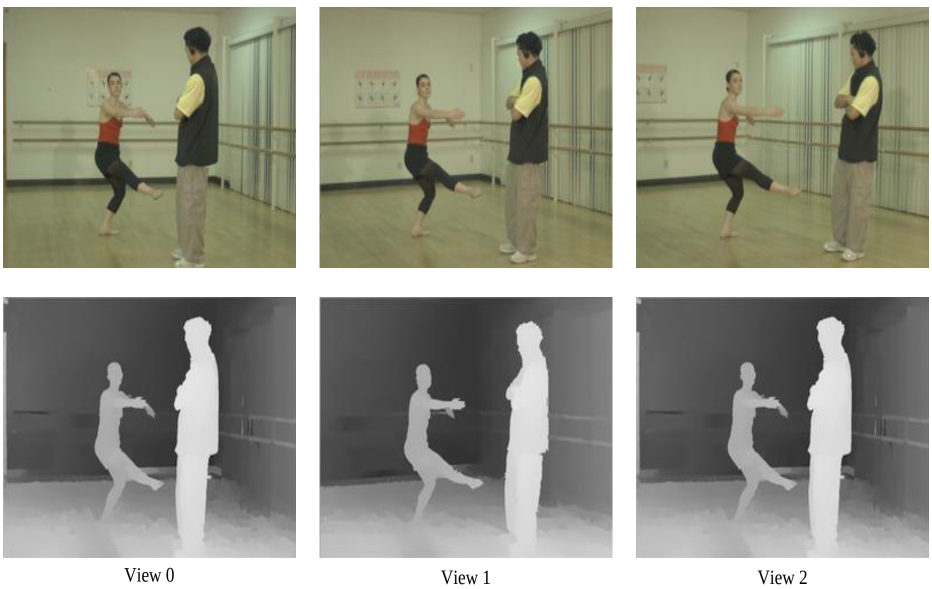


Fig. 12: Multiview video-plus-depth.

3-7 LAYERED DEPTH VIDEO (LDV) FORMAT

Although MVD can reduce the required bandwidth to transmit the color and depth data for all views, the overall required bandwidth is still very large. To further reduce the bit rate, LDV is an effective technique. Layered depth video (LDV) [48, 49] is a derivative and alternative to MVD where only one full view with additional residual data is transmitted. One representation of LDV again uses color video with associated depth map (V+D) representation and an additional component called the background layer with its associated depth map, as illustrated in "Fig. 13".

Another type of LDV consists of a main layer that contains one full or central view and one or more residual layers of color and depth data to represent the side views. One major problem with LDV is disocclusions, where blank spots appear as the distance between the central view and side views increases. Hence, the extra information enables a correct rendering of disoccluded objects. For more details on LDV, the reader is refereed to [49, 50].

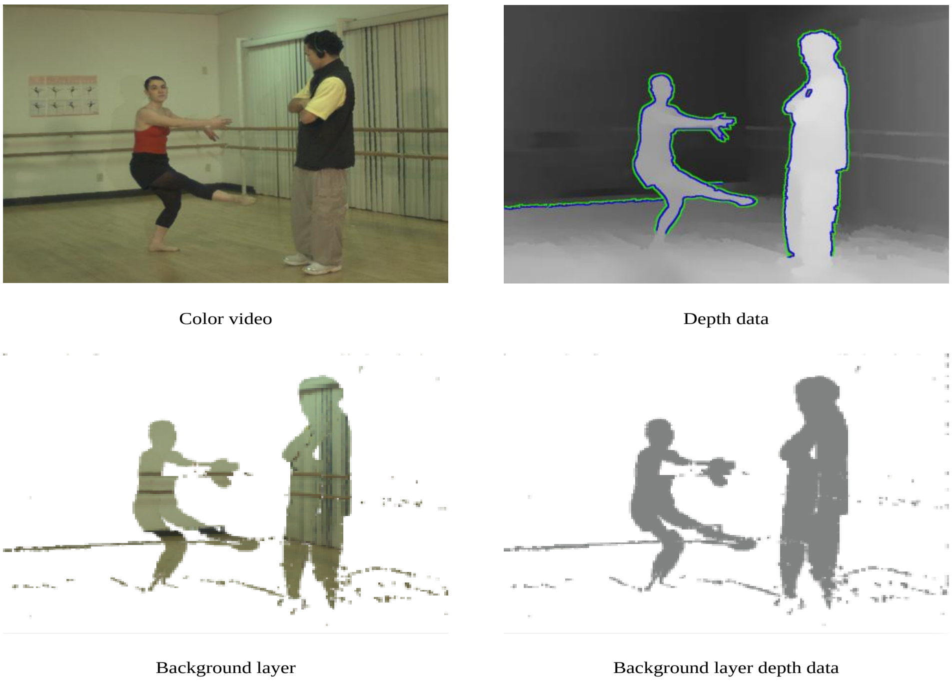


Fig. 13: Layered depth video.

IV. 3-D Video Coding Standards

Coding and compression of 3-D video formats is the next block in the 3-D video processing chain. To realize an efficient transmission over bandwidth limited channels, 3-D video representation formats discussed in the previous section, have to be compressed efficiently. In the last few years, the ISO-MPEG and ITU-VCEG international organizations mainly focused on improving the coding efficiency of the H.264/AVC standard and on multiview video coding. The joint team between ISO-MPEG and ITU-VCEG set up a joint collaboration in 2010 to develop a video coding standard which aimed to improve the coding efficiency of H.264/AVC by up to 50%. The scope of this subsection is to describe related compression standards. In particular, H.264/AVC and H.264/MVC, are briefly reviewed.

4-1 H.264/AVC codec

Apart from the deblocking filter, most of H.264 standard functions (prediction, transform, and entropy coding) are presented in prior standards but the most important changes in H.264 appear in the details of each function. The input to the H.264 encoder is video frames in YUV format. H.264/AVC encoder will try to exploit redundancies to reduce the amount of bits necessary to represent it. Then, the decoder will identify the syntax of representation and decode the received bit stream to reconstruct the video at the receiver side. The H.264/AVC standard consists of two layers, known as the video coding layer (VCL) and the network abstraction layer (NAL). The reader is referred to the standard itself [51] and some overview papers that have discussed this matter [35, 52, 53].

4-2 MVC extension of H.264 standard

The large amount of data required to represent multiview video applications, which requires the development of highly efficient coding schemes, is the major challenge for multiview video transmission. MVC is based on the single-view video compression standard. For the general case of two or more views, the joint video team (JVT) of the ITU-T video coding is developing a multiview extension of the H.264/AVC standard, known as H.264/MVC extension. MVC provides a new technique to improve coding efficiency by exploiting temporal as well as interview statical dependencies between neighboring views. Consequently, MVC takes advantage of the redundancies among the inter-pictures of one view and the interview pictures of other views.

A straightforward approach for coding multiview video content is simulcast coding where each view is encoded and decoded separately. This can be done with any video codec including H.264/AVC. In simulcast coding, the prediction process is limited to the reference pictures in the temporal dimension. "Fig. 14" shows the simulcast coding structure with hierarchical bi-directional B pictures for temporal prediction with two views and a group of pictures (GOP) of length of 8. This scheme is simple, but is an inefficient way to compress multiview video sequences because it does not benefit from the existing correlation between the different views.

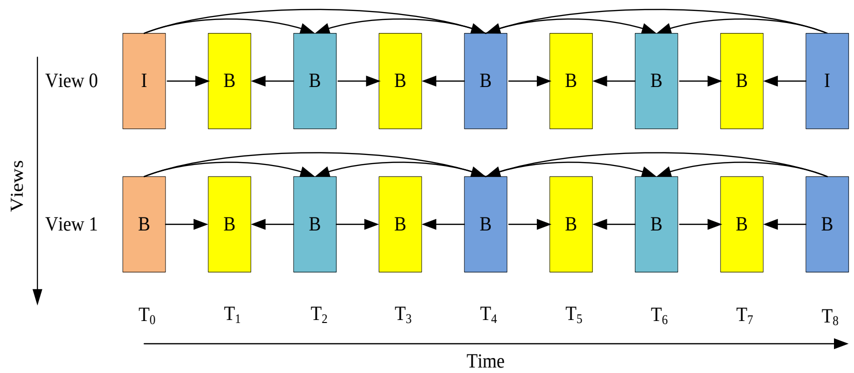


Fig. 14: Simulcast coding structure with B pictures for temporal prediction.

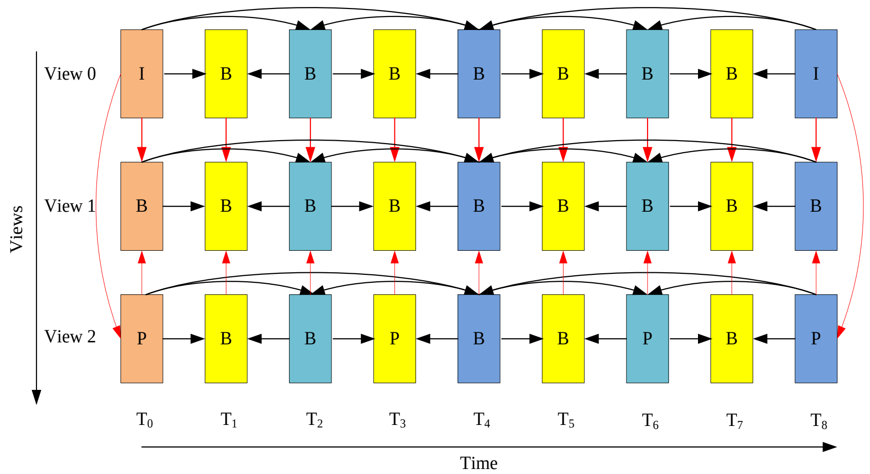


Fig.15: Typical MVC prediction structure.

The MVC coding was added as an extension to H.264/AVC in July 2008 and ultimately standardized in early 2010. The H.264/MVC standard uses hierarchical B pictures for each view and at the same time, applies inter-view prediction to every second view in order to exploit all statical dependencies. "Fig.15" illustrates how temporal prediction is combined with inter-view prediction. The first view is coded independently, as in simulcast coding, and for the remaining views, interview reference pictures are additionally used for prediction. As a consequence, MVC provides up to 40% bit rate reduction for multiview data in comparison to single view AVC coding. This is at the cost of random access delay. A more detailed description of the H.264/MVC is given in [35, 44, 54]. As discussed in this section, the 3-D video coding standards are mainly 3-D extensions of existing 2-D video coding standards modified to support 3-D application requirements.

V. Conclusion

This paper has 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. Coding of 3-D video for limited bandwidth is an important problem that needs to be addressed. The paper concluded with 3-D video coding standards that could be adopted or extended from 2-D to 3-D formats, which are integral in resolving these issues. From the state-of-the-art literature, it is evident that these techniques are very promising for 3-D video transmission.

References

[1]. H.264 and ISO/IEC 14496-10 (MPEG-4 AVC), Advanced video coding for generic audiovisual services, 2007.
[2]. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity, *IEEE Trans. Broadcast*., 13(4), 2004, 600-612.
[3]. S.-U. Yoon and Y.-S. Ho, Multiple color and depth video coding using a hierarchical representation, *IEEE Trans. Circuits Syst. Video Technol*., 17(11), 2007, 1450-1460.
[4]. ISO/IEC JTC1/SC29/WG1N3797, Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/spart 2: Video, ISO/IEC 11172-2 (MPEG-1 Video), ISO/IEC JTC 1, 1993.
[5]. ITU-T and ISO/IEC JTC 1, Generic coding of moving pictures and associated audio information part 2: Video, ITU-T Rec. H.262 and ISO/IEC 13818-2(MPEG-2 Video), 1994.
[6]. ISO/IEC JTC 1, Coding of audio-visual objects part 2: Visual, ISO/IEC 14496-2 (MPEG-4 Part 2), 1999.
[7]. ITU-T Rec. H.261, ITU-T, Video codec for audiovisual services at p x 64 kbit/s, 1993.
[8]. TU-T Rec. H.263, ITU-T, Video coding for low bit rate communications, 2000.
[9]. V. Anthony, T. M. Alexis, M. Karsten, and C. Tao, 3D-TV content storage and transmission, *IEEE Trans. Broadcast*., 57(2), 2011, 384-394.
[10]. A. Smolic, K. Mueller, N. Stefanoski, J. Ostermann, A. Gotchev, G. B. Akar, G. Triantafyllidis, and A. Koz, Coding algorithms for 3DTV-a survey, IEEE *Trans. Circuits Syst. Video Technol*., 17(11), 2007, 1606-1621.
[11]. G.-M. Su, Y.-C. Lai, A. Kwasinski, and H. Wang, 3D video communications: challenges and opportunities, *International journal of communication systems*, (24), (10), 2011, 1261-1281.
[12]. B. Wandell, Fundations of Vision, Sinauer Associates. Sunderland, MA: Wiley, 1995.
[13]. B. Girod, Eye movements and coding of video sequences, *Proc. SPIE, Visual Communications and Image Processing*, VCIP'88, 1001, Cambridge, MA, USA, 1988, 398-405.
[14]. C. Fehn, Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV, *Proc. SPIE Conf. Stereoscopic Displays and Virtual Reality Systems XI*, San Jose, CA, USA, 2004, 93-104.
[15]. P. Merkle, K. Muller, and T. Wiegand, 3D video: acquisition, coding, and display, *IEEE Trans. Consum. Electron*., 56(2), 2010, 946-950.
[16]. E. A. Umble, Making it real: the future of stereoscopic 3D film technology, AMC Siggraph Computer Graphics, 40(1), 2008, 925-932.
[17]. Y. Morvan, D. Farin, and P. H. N. de. With, System architecture for free viewpoint video and 3D-TV, *IEEE Trans. Consum. Electron*., 54 (2), 2008, 925-932.
[18]. J. Flack, J. Harrold, and G. J. Woodgate, A prototype 3D mobile phone equipped with a next generation autostereoscopic display, *Proc. SPIE Stereoscopic Displaysand Virtual Reality Systems XIV*, San Jose, CA, USA, 2007, 502-523.
[19]. MSR 3D video sequence: Microsoft Research, “Interview and ballet sequences,” [Online]. Available: http://research.microsoft.com/en-us/um/ people/sbkang/3dvideodownload/, [Viewed: Feb. 2015].
[20]. C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, High-quality video view interpolation using a layered representation, *ACM Transactions on Graphics*, 23(3), 2004, 600-608.
[21]. M. Flierl and B. Girod, Multiview video compression, IEEE *Signal Process*. Mag., 24(6), 2007, 66-76.
[22]. M. Philipp, M. Karsten, and W. Tomas, Efficient compression of multiview video exploiting inter-view dependencies based on H.264/MPEG4-AVC, *Proc. ICME'06*, Toronto, Canada, 2006, 1717-1720.
[23]. B. Haskell, A. Puri, and A. Netrevali, Digital Video: An Introduction to MPEG-2. Chapman and Hall, 1997.
[24]. K. Yamamoto, M. Kitahara, H. Kimata, T. Yendo, T. Fujii, M. Tanimoto, S. Shimizu, K. Kamikura, and Y. Yashima, Multiview video coding using view interpolation and color correction, *IEEE Trans. Circuits Syst. Video Technol*., 17, (11), 2007, 1436-1449.
[25]. H. Brust, A. Smolic, K. Mueller, G. Tech, and T. Wiegand, Mixed resolution coding of stereoscopic video for mobile devices, *Proc. 3DTV- Conference 2009, The True Vision: Capture, Transmission and Display of 3D Video*, Potsdam, Germany, 2009, 1-4.
[26]. N. Atzpadin, P. Kau_, and O. Schreer, Stereo analysis by hybrid recursive matching for real-time immersive video conferencing, *IEEE Trans. Circuits Syst. Video Technol*., 14(3), 2004, 321-334.
[27]. C. Fehn, P. Kau_, M. O. D. Beeck, F. Ernst, W. I Jsselsteijn, M. Polle feys, L. V. Gool, E. Ofek, and I. Sexton, An evolutionary and optimised approach on 3D-TV, *Proc. International Broadcast Conference*, Amsterdam, The Netherlands, 2002, 357-365.
[28]. ISO/IEC JTC1/SC29/WG11, Text of ISO/IEC FDIS 23002-3 representation of auxiliary video and supplemental information, *Proc.* Doc. N8768, Marrakech, Morocco, 2007.
[29]. C. Fehn, A 3D-TV system based on video plus depth information, *Proc. The Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, California, USA, 2003, 1529-1533.
[30]. K. M. Alajel, BER Performance Analysis of a New Hybrid Relay Selection Protocol, *Proceedings of International Journal of Signal Processing Systems*, 4, (1), 2016, 13-16.
[31]. S. Grewntsch and E. Miiller, Sharing of motion vectors in 3D video coding, *Proc. IEEE ICIP'04*, Singapore, 2004, 3271-3274.
[32]. M. T. Pourazad, P. Nasiopoulos, and R. K. Ward, An H.264-based video encoding scheme for 3D TV, *Proc. 14th European Signal Processing Conference (EUSIPCO)*, Florence, Italy, 2004, 3271-3274.
[33]. P. Merkle, Y. Morvan, A. Smolic, D. Farin, K. Mller, P. H. N. de With, and T. Wiegand, The effect of depth compression on multiview rendering quality,” *Proc. 3DTV-Conference 2008, The True Vision: Capture, Transmission and Display of 3D Video*, Istanbul, Turkey, 2008, 245-248.
[34]. C. J. Chartres, R. S. Green, and G. W. Ford, Multiview video compression, *IEEE Signal Process*. Mag., 31(6), 2007, 66-76.
[35]. A. Vetro, T. Wiegand, and G. J. Sullivan, Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard, *Proceedings of the IEEE*, 99(4), 2011, 626-642.
[36]. P. Merkle, K. Muller, A. Smolic, and T. Wiegand, Statistical evaluation of spatiotemporal prediction for multiview video coding, *Proc. 2nd Work- shop On Immersive Communication And Broadcast Systems, ICOB 2005*, Berlin, Germany, 2005, 27-28.
[37]. A. Kaup and U. Fecker, Analysis of multireference block matching for multiview video coding, *Proc. 7th Workshop Digital Broadcasting*, Erlangen, Germany, 2006, 33-39.
[38]. ISO/IEC JTC1/SC29/WG11, Text of ISO/IEC 14496-10:200X/FDAM 1 multiview video coding, Doc. N9978, Germany, 2008.
[39]. X. Cheng, L. Sun, and S. Yang, A multiview video coding scheme using shared key frames for high interactive application, *Proc. Picture Coding Symposium, PCS'06*, Beijing, China, 2006.
[40]. A. Vetro, W. Matusik, H. P. P_ster, and J. Xin, Coding approaches for end-to-end 3-D TV systems, *Proc. Picture Coding Symposium, PCS'04*, San Francisco, CA, USA, 2004, 319-324.
[41]. D. Socek, D. Culibrk, H. Kalva, O. Marques, and B. Furht, Permutation- based low-complexity alternate coding in multiview H.264/AVC, *Proc. IEEE ICME'06*, Toronto, Canada, 2006, 2141-2144.

[42]. K. Muller, P. Merkle, H. Schwarz, T. Hinz, A. Smolic, and T. Wiegand, Multi-view video coding based on H.264/AVC using hierarchical B-frames, *Proc. Picture Coding Symposium, PCS'06*, Beijing, China, 2006.

[43]. A. S. P. Merkle, K. Muller, and T. Wiegand, Efficient prediction structures for multiview video coding, *IEEE Trans. Circuits Syst. Video Technol*., 17(11), 2007, 1461-147.

[44]. Y. Chen, Y.-K. Wang, K. Ugur, M. M. Hannuksela, J. Lainnema, and M. Gabbouj, The emerging MVC standard for 3D video services, *EURASIP J. Adv. Signal Process*., 2009(1), 1-13, 2009.

[45]. ISO/IEC JTC1/SC29/WG11, Overview of 3D video coding, *Proc*. Doc. N9784, Archamps, France, 2008.

[46]. P. Kau_, N. Atzpadin, C. Fehn, M. Muller, O. Schreer, A. Smolic, and R. Tanger, Depth map creation and image based rendering for advanced 3DTV services providing interoperability and scalability, *Signal Process. Image Commun*., 22(2), 2007, 217-234.

[47]. P. Merkle, A. Smolic, and T. Wiegand, Multi-view video plus depth representation and coding, *Proc*. *IEEE ICIP'07*, San Antonio, TX, USA, 2007, 201-204.

[48]. K. Muller, 3D visual content compression for communications, *IEEE E-Letter*, 4(7), 2009, 22-24.

[49]. K. Muller, A. Smolic, K. Dix, P. Merkle, P. Kau_, and T. Wiegand, View synthesis for advanced 3D video systems, *EURASIP J. Image Video Process*., 2008(7), 2008, 1-11.

[50]. K. Muller, A. Smolic, K. Dix, P. Kau_, and T. Wiegand, Reliability- based generation and view synthesis in layered depth video, *Proc. IEEE 10th Workshop on Multimedia Signal Processing*, Cairns, Queensland, Australia, 2008, 34-39.

[51]. ITU-T and ISO/IEC JTC 1, Advanced video coding for generic audiovisual services, ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG-4 AVC), 2010.

[52]. T. Wiegand, G. J. Sullivan, G. Bjntegaard, and A. Luthra, Overview of the H.264/AVC video coding standard, *IEEE Trans. Circuits Syst. Video Technol*., 13(7), 2003, 560-576.

[53]. G. J. Sullivan and T. Wiegand, Video compression - from concepts to the H.264/AVC standard, *Proceedings of the IEEE*, 93(1), 2005, 18-31.

[54]. K. Muller, P. Merkle, and T. Wiegand, 3-D video representation using depth maps, *Proceedings of the IEEE*, 99(4), 2011, 643-656.