3D Research (2010) 01: 43-47
10.1007/3DRes.01(2010)6

3DR REVIEW

# An Overview of 3D Video Representation and Coding

**Lianlian Jiang • Jiangqian He • Nan Zhang • Tiejun Huang**

Received: 26 November 2009 / Revised: 27 January 2010 /Accepted: 20 February 2010


**Abstract** As the next generation of television, three-dimensional television (3DTV) is enkindling various new applications. Since the representation and coding of three-dimensional video has direct effect on successive transmission, synthesis and display, it becomes crucially important and receives a wide publicity. This paper gives an overview of existing 3D video representation, relevantly state-of-the-art coding standards and various coding approaches consisting of conventional stereo video, multi-view video, video plus depth, multi-view video plus depth and layered depth video.

**Keywords** Three-dimensional television; 3D video representation; Video coding; Multi-view video; Depth map

## 1. Introduction

Three-dimensional television has entertained us with its advance in display, just like a feast for the eyes. Different kinds of display technologies have succeeded in presenting the vivid and lifelike scenes. With the development of display technologies and the requirement of audience to enjoy the natural 3D world, how to strive to create realistic 3D impressions of natural 3D scenes seems to be more urgent. The major challenge is that natural 3D scenes representing results in tremendous amount of data. In other words, the key to success in 3DTV is efficient compression since there are too much data needed to be transmitted under limited bandwidth.

Various 3D display technologies use various kinds of data representation formats, adopting different coding methods. But their objective remains the same, which is to encode 3D video by removing temporal and spatial redundancy. This paper focuses on coding algorithms of various representation formats and gives an overview of the state-of-the-art coding technology for 3D video.

Earlier in 1990s, MPEG-2 proposed multi-view profile. This is included in ITU-T Rec. H.262/ISO/IEC 13818-2[1,2]. Besides, the ISO/IEC JTC1 Moving Pictures Experts Group (MPEG) has organized a group named 3DAV (3D Audio-Visual) which is devoted to researching the technology of 3D audio and video in 2001[3,4]. In the following years, there appear different types of international standards or representation formats. One of the practical and well-developed approaches is using multiple views to render a three-dimensional scene. The direct solution is to independently encode all the separated videos using a current video codec such as H.264/MPEG-4 AVC[5]. Due to containing a large amount of inter-view statistical dependencies, it needs to be implemented exploiting combined temporal/inter-view prediction, referred to multi-view video coding (MVC)[6,7].

Apart from MVC, conventional color video and an associated depth map is another popular format for 3DTV, which is also well-known as video plus depth[8]. Furthermore, Multi-view video plus depth (MVD) extends the mentioned video plus depth approach. In addition to high quality rendering, MVD is also able to increase viewpoint flexibility to cover viewpoints that do not lie on the camera baselines[9].

At the Lausanne MPEG meeting a proposal brought forward that hidden (background) texture and hidden depth information should be added to amend MPEG-C Part 3 (ISO/IEC 23002-3), which is also known as Layered Depth Video (LDV)[10]. As the high-quality auto-stereoscopic displays will enter the consumer market in the next few years, the application and requirement of a new 3D video (3DV) is in process, extending the Free-view TV (FTV) from MVC to 3DV to fulfill the new requirement[11].

The following section gives an overview of 3D video representation approaches and coding technologies which have been proposed in the standards and several novelty ideas. Section 2 will focus on the stereo video coding and multi-view coding. Coding of the video and auxiliary depth information will be devoted in section 3. Finally, section 4 summarizes and concludes the paper.

Lianlian Jiang[1] • Jianqiang He[1] • Nan Zhang[2] (✉) • Tiejun Huang[1]
[1] National Engineering Laboratory for Video Technology, School of EE & CS, Peking University, Beijing, China
[2] School of Biomedical Engineering, Capital Medical University, Beijing, China
e-mail: nzhang@jdl.ac.cn

## 2. Coding of stereo video and multi-view video

Some 3DTV display system can use a group of video sequences captured by multiple cameras simultaneously for the same scene to show natural and vivid effect. Usually, the system will choose 8 or 16 views to represent 3D scene, while conventional stereo video adopts only two cameras.

### 2.1 Conventional Stereo Video Coding

Conventional stereo video coding is the most simple and traditional representation of 3D video. Corresponding to the distance of human eyes, two cameras simultaneously take capture of the same scene to acquire stereo video from slight different viewpoints. The two similar images benefit compression by predicting one image from the other. That is to say, after encoding one sequence independently, the matched pair can be predicted with reference to previous video exploiting the correlations between adjacent cameras.

Multi-view profile (MVP) is specified in ITU-T Rec. H.262/ISO/IEC 13818-2 around 15 years ago. MVP's video coding scheme is two-layer (base layer and enhancement layer) as illustrated in Fig. 1[2]. The base layer video is coded as MPEG-2 main profile (MP) bit stream. With the purpose of better efficiency, the enhancement layer video is coded exploiting correlation from both temporal and spatial dependency.

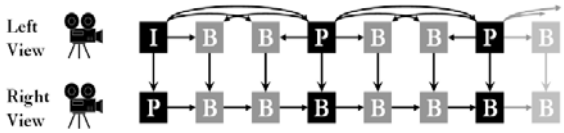


Fig.1 The prediction for conventional stereo video coding

In stereoscopic video applications the base layer is assigned to the left eye view and the enhancement layer to the right eye view, while the base layer of the main profile bit stream can also be displayed in two dimensional monitor[12].

Since the method guarantees the backward compatibility, stereo video is easier to be wildly deployed in short time. However, the enhancement of compression efficiency is relatively limited, since the temporal prediction is already well-developed. For the more, the 3D information provided by stereo video is limited especially when capturing in close distance, more views are investigated widely in recent years.

### 2.2 Multi-view Video Coding

Multi-view video is acquired by simultaneously taking capture of the same scene from different viewpoints. As the first step of free viewpoint video, the user can enjoy the scene interactively from multiple orientations.

Independently coding of separated videos, called simulcast, is the most straightforward method using a state-of-the-art codec such as H.264/AVC[5]. However, inter-view statistical redundancies can be removed to decrease the amount of data. These redundancies can be classified into two types, inter-view similarity between adjacent camera views and temporal similarity between temporally successive images of each video. Motion compensation techniques that are well-developed for single-view video compression can be used to temporal prediction. Likewise, disparity compensation techniques can be utilized to reduce inter-view redundancies[13]. As shown in Fig. 2, algorithms that are based on hierarchical B-pictures[14] as supported by H.264/AVC syntax in temporal and inter-view dimension has proved best performance[15], which has been selected by JVT as the joint multi-view view model (JMVM). Inter-view prediction starts from P picture and then uses the hierarchical B pictures, in order to decrease the redundancy from adjacent views.

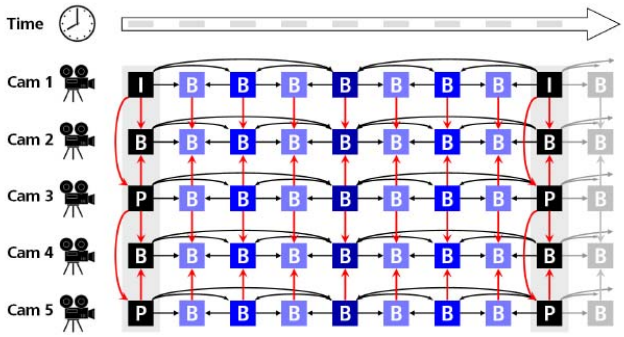


Fig. 2 The structure scheme of multi-view video coding with combined temporal/interview prediction

MVC specified by ISO/IEC 14496-10 | ITU-T Rec. H.264 supports the direct coding of the synchronous information from multiple views using a single stream and exploits inter-camera redundancy to reduce the bit rate.

Owing to taking advantage of the temporal and interview correlations, some experiments conducted in the context of MPEG standardization[6] have shown that dedicated MVC exceeds simulcast significantly with the gain of peak signal-to-noise ratio (PSNR) from 0.5 dB to 3 dB, so at the request of industry, ISO/MPEG and ITU/VCEG decided to develop a specific MVC standard as the extension of H.264/AVC and finished the task in July 2008[15,16].

During the standardization of MVC, JMVM is set as a project focusing on the potential use. Consequently, illumination compensation[17] and motion skip have been adopted into JMVM instead of the current MVC standard. Illumination compensation incorporates illumination changes with the motion compensation process. Motion skip mode[18]or inter-view direct mode[19] refers to motion information from the corresponding block in neighboring view. Other tools may contribute further gain and deserve taking into consideration. Adaptive reference filtering[20] could compensate mismatches between views. View synthesis prediction[21] can generate virtual views for prediction using neighboring views and estimated depth.

All these tools would require changes to slice or macroblock level in syntax.

Scalability and adaptation of multi-view video technology is already accepted to support a variety of applications, including free-view video and three-dimensional video. However, the huge amount of data will

impose heavy burden on storage and transmission. Though hierarchical B pictures offer efficient compression, the algorithm has increased complexity of the decoder and the encoder, which limits applications especially for mobile devices. Besides, MVC is based on fixed camera inputs and cannot vary the baseline distance to adjust the depth perception unless it is equipped with extra data for virtual view synthesis.

## 3. Coding of the video and auxiliary depth information

Since depth information is crucial to advanced stereoscopic display and auto-stereoscopic N-view displays, multi-view video plus depth is towards the standardization of free viewpoint video, as the extension of video plus depth. Subsequently, as an alternative to MVD, Layered depth video can represent 3D video more efficiently.

### 3.1 Coding of video plus depth

The European Information Society Technologies (IST) launched "Advanced Three-Dimensional Television System Technologies" (ATTEST) project to design a novel, backwards-compatible and flexible broadcast 3DTV system[22], which is based on a more flexible distribution of an image-based layered 3D data representation format consisting of monoscopic color video and associated per-pixel depth information. The format is illustrated in Fig. 3. The depth can provide the 3D cues and show viewer how far the objects are with the format of 8-bit gray values. The gray level 0 specifies the furthest and the 255 represents the closest. With the help of so-called depth-image-based rendering (DIBR) technique, the receiver side can obtain high flexibility to render the virtual views of the 3D scene by 3D image warping algorithm[23]. According to the principle of human vision system, as long as the two eyes can have the specific view with disparity, the brain can have the 3D effect after synthesis.

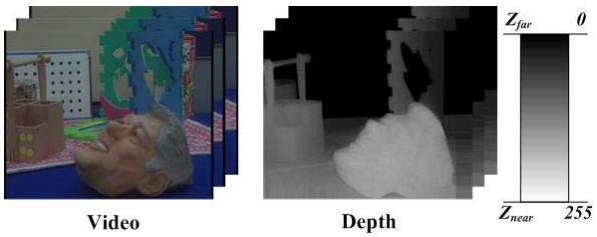


Fig. 3 Video plus depth

MPEG-C Part 3 (ISO/IEC 23002-3) is the coding standard of single video plus per sample depth[24]. As the auxiliary video, the depth is compressed as conventional luminance signals and can be generated by estimation from original left or right view at the sender side.

Moreover, Auxiliary Picture Syntax in the H.264/MPEG-4 AVC specifies the main video stream could be sent along with the extra depth video. An auxiliary coded picture supplements the primary coded picture with the same syntactic and semantic restrictions. The H.264/MPEG-4 AVC codec encodes video as primary coded picture and depth as the auxiliary coded picture simultaneously but independently.

Besides the state-of-the-art video codec, more depth map coding algorithms are proposed, including sharing the motion information of the corresponding texture video[25] or rate-distortion optimization of the quadtree decomposition[26] and so on[27].

Video plus depth can fulfill four requirements including interoperability, display technology independency, backwards compatibility and compression efficiency. Particularly concerning the efficient compression, the European ATTEST project has given the results that depth occupies only 10%–20% of the bit rate to show a good quality comparing with the color video[23]. Unfortunately, it cannot support free viewpoint navigation. What's worse, the artifacts and occlusion problem is the fatal problem since the invisibly occluded section will become visible in virtual view from another viewpoint.

To reduce artifacts, a new three-dimensional video representation method proposes to combine two-dimensional texture, depth map with shape masks[28] as shown in Fig. 4. It can not only effectively erase the cloud noise caused by the latter synthesizing of the uncertain boundaries, but also conduct the self-adaptive adjustment of the depth to the practical needs.


Fig. 4 Two-dimensional texture+Depth map+Shape masks

### 3.2 Coding of Multi-view video plus depth

Multi-view video plus depth (MVD) is the extension of video plus depth where spectators will no longer be limited in fixed viewing zone and viewing angle. MVD is composed of multiple color videos with associated multiple depth data as shown in Fig. 5. Compared with MVC, MVD doesn't need to transmit so many sequences of the same scene for the intermediate views can be generated by the DIBR technique at the receiver side. Namely, fewer views and associated depth information are needed to be simultaneously transmitted. Consequently, the bit rates will be efficiently reduced.

Since depth data is smooth and monochromatic, it can be easily compressed when using standard video coding algorithms and exploiting the inter-view redundancy[29]. An efficient optimal solution is proposed by joint estimation and coding of the depth map using dynamic programming along the tree of wavelet coefficients[30]. Furthermore, jointly coding of depth and multi-view is another option by

using motion information of corresponding texture video[25] or depth view synthesis prediction[31]. The video/depth compression performances can be improved by performing joint video/depth rate allocation[32-34].

Fig. 5 Multi-view video plus depth

Although the multi-view video can support a variety of 3D displays applications, the large amount of data is a serious problem, especially when the number of cameras increases.

### 3.3 Coding of Layered depth video

Layered depth video (LDV), as a novel representation, is an efficient representing and rendering methods for 3D objects with complex geometries[35]. It is the alternative to MVD and can be acquired by MVD after warping *n* depth images into a common camera view[36]. It uses one color video with associated depth map and a background layer with associated depth map. The background layer includes image content which is covered by foreground objects in the main layer[37]. This is illustrated in Fig. 6.

Fig. 6 Layered depth video

LDV is the sequence of layered depth image (LDI), so the efficient compression of image is the key technology. LDI is consisting of an array of layered depth pixels ordered from closest to furthest from a single sight. Every layered depth pixel contains multiple depths at per pixel location. The function of multiple depth information is to avoid occlusion problem since occluded section will be visible as the viewpoint moves away from the center of LDI camera. Other than ordinary image, each LDI pixel contains 63 bit information including R,G and B components, alpha channel, depth showing the distance of the pixel to the camera and the index into a splat table, so the compression approach need to be updated. Moreover, there are three obstacles that are multiple layers, the sparse back layer and multiple property values of each pixel to be solved by the new compression algorithm[38].

Usually, the method is involved with several parts. After recording the number of layers (NOL) by pixel, the individual component of LDI data is preprocessed[39] and then separately encoded with current standards. In respect of the irregular and sparse shape of back layer image, there are two kinds of algorithms. One is aggregation of data in each layer into horizontal direction[38]. The other is called layer filling, which is to fill the empty pixel locations of all layer images using pixels in the first layer or background layer[35]. Besides, optimal allocation of LDI bit stream among different components has to be taken into consideration.

LDV is more efficient than MVD with lower coding complexity. In virtue of the imperfection of converting the multi-view video to layered depth video, it will bring about the artifacts when displaying.

## 4. Conclusions

Since 3D technology has stepped into a new stage, it will be widely applied to broadcasting, telemedicine, interactive video, cinema, gaming and other applications. Research on coding of stereo video, multi-view video with associated depth or disparity data is comparatively well-developed with available international standards. Inspired by requirements of a new 3DV initiated by MPEG, MVD will become the hot topic.

**Acknowledgement** This paper was supported in part by the NSFC of China (No.60803069), the “863” Program of China (No.2007AA01Z315), the “973” Program of China (No.2009CB320904).

## References

1. B. Haskell, A. Puri and A. Netrevali (1998) Digital video: an introduction to MPEG-2, *Journal of Electronic Imaging* 7: 265-266
2. Generic coding of moving pictures and associated audio information-Part 2: video, *ITU-T Rec. H.222.0|ISO/IEC 13818-1 (MPEG 2 Systems), ITU-T and ISO/IEC JTC1* (1994)
3. M. Jose (2003) MPEG 3DAV AhG activities report, *65th MPEG Meeting*, Trondheim, Norway
4. A. Smolic and D. McCutchen (2004) 3DAV exploration of video-based rendering technology in MPEG, *IEEE Transactions on Circuits and Systems for Video Technology* 14: 348-356
5. Advanced video coding for generic audio-visual services, ITU-T Recommendation H.264 & ISO/IEC 14496-10 AVC (2003)
6. P.Merkle, K. Müller, A. Smolic, and T. Wiegand (2006) Efficient compression of multi-view video exploiting inter-view dependencies based on H.264/MPEG4-AVC, *IEEE International Conference on Multimedia and Exposition*
7. K-J. Oh and Y-S. Ho (2006) Multi-view video coding based on the lattice-like pyramid GOP structure, Picture Coding Symposium
8. S. Shinya, K. Hideaki and Y. Ohtani (2009) Real-time free-viewpoint viewer from Multi-view video plus depth representation coded by H.264/AVC MVC extension, 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 1-6
9. Introduction to 3D video, ISO/IEC JTC1/SC29/WG11 coding of moving pictures and audio/N9784, Archamps, France (2008)

10. Proposal to amendment MPEG-C Part 3. organization international de normalization, ISO/IEC JTC1/SC29/WG11 Coding Of Moving Pictures and Audio/ M14700, Lausanne, Switzerland (2007)
11. Vision on 3D video organization international de normalization, ISO/IEC JTC1/SC29/WG11MPEG 2009/ N10357, Lausanne, Switzerland (2009)
12. X. Chen and A. Luthra (1997) MPEG-2 multi-view profile and its application in 3D TV, *Proceedings of SPIE*, 3021: 212-223
13. M. Flierl, A. Mavlankar, and B. Girod (2007) Motion and disparity compensated coding for multi-view video, *IEEE Transactions on Circuits and Systems for Video Technology*
14. H. Schwarz, D. Marpe, and T. Wiegand (2006) Analysis of hierarchical B-pictures and MCTF, *IEEE International Conference on Multimedia and Exposition*
15. A. Smolic, K. Mueller, N. Stefanoski, J. Ostermann, A. Gotchev, G.B. Akar, G. Triantafyllidis, and A. Koz (2007) Coding Algorithms for 3DTV-A Survey, *IEEE Transactions on Circuits and Systems for Video Technology*, 17:1606-1621
16. Joint Draft 8.0 on Multiview video coding, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG/JVT-AB204, 28th Meeting: Hannover (2008)
17. J. H. Kim, P. Lai, J. Lopez, A, Ortega, Y. Su, P. Yin, C. Gomila (2007) New coding tools for illumination and focus mismatch compensation in Multiview video coding, *IEEE Transactions on Circuits and Systems for Video Technology*, 17: 1519-1535
18. H. Yang, Y. Chang and J. Huo (2009) Fine-granular motion matching for inter-view motion skip mode in Multi-view video coding, *IEEE Transactions on Circuits and Systems for Video Technology*, 19: 887-892
19. X. Guo, Y. Lu, F. Wu, W. Gao (2006) Inter-view direct mode for Multiview video coding, *IEEE Transactions on Circuits and Systems for Video Technology*,16:1527-1532
20. P. Lai, A. Ortega, P. Pandit, Y. Peng and C. Gomila (2008) Adaptive reference filtering for bidirectional disparity compensation with focus mismatches, *International Conference on Image Processing* , 6: 2456-2459
21. Y. Sehoon, A.Vetro (2007) RD-optimized view synthesis prediction for multi-view video coding, *International Conference on Image Processing* , 1: 209 -212
22. C. Fehn, P. Kauff, M. de Beeck, M. Ernst, W. IJsselsteijn, M. Pollefeys, L. Gool, E.Ofek, and I. Sexton (2002) An evolutionary and optimized approach on 3D-TV, *Proceedings of International Broadcast Conference*, 357-365
23. C. Fehn, R.Delabarre. S. Pastoor (2006) Interactive 3DTV-concepts and key technologies, *Proceedings of the IEEE* , 94: 524-538
24. Information technology -- MPEG video technologies -- Part 3: Representation of auxiliary video and supplemental information ISO/IEC 23002-3 (2007)
25. H. Oh, Y-S. Ho (2006) H.264-based depth map sequence coding using motion information of corresponding texture video, *PSIVT*,898-907
26. Y. Morvan,D. Farin and P. H.N. de With (2007) Depth image compression based on an R-D optimized quadtree decomposition for the transmission of multivew images, International Conference on Image Processing, 105-108
27. B. Zhu, G. Y. Jiang, M. Yu, P. An, Zh. Y. Zhang (2009) Depth map compression for view synthesis in FTV, ISO/IEC JTC1/SC29/ WG11 MPEG2008/M 16021
28. N. Zhang, S. Ma and W. Gao (2009) Shape-based depth map coding, *IEEE The Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*
29. T. Wiegand, G. Sullivan, J. Reichel, H. Schwarz, and M. Wien (2006) Joint draft 8 of SVC amendment, ISO/IECJTC1/SC29/WG11 and ITU-T SG16 Q.69 (JVT-U201)
30. M. Maitre, Y. Shinagawa and M. N. Do (2007) Rate-distortion optimal depth maps in the wavelet domain for free-viewpoint rendering, *International Conference on Image Processing*, 125-128
31. S-T. Na, K-J. Oh, C. Lee and Y-S. Ho (2008) Multi-view depth video coding using depth view synthesis, International Symposium on Circuits and Systems, 1400-1403
32. Y. Morvan, D. Farin and P. H. N. de (2007) Joint depth/texture bit-allocation for multi-view video compression, Proceedings of Picture Coding Symposium
33. Y. Liu, Q. Huang, S. Ma, D. Zhao and W. Gao (2009) Joint video/depth rate allocation for 3D video coding based on view synthesis distortion model, *Signal Processing: Image Communication*, 24: 666-681
34. Y. Liu, S. Ma, Q. Huang, D. Zhao, W. Gao, N. Zhang (2009) Compression-induced rendering distortion analysis for texture/depth rate allocation in 3D video compression, *Proceedings of Data Compression Conference*, 352-361
35. X. Cheng, L. Sun and S. Yang (2007) A multi-view video coding approach using layered depth image, Multimedia Signal Processing, 143-146
36. J. Shade, S. Gortler, L. Hey, and R. Szeliskiz (1998) Layered depth images, *Proceedings of ACM SIGGRAPH*, 231-242
37. Project no. FP7-213349. D5.1-requirements and specifications for 3D video (2008)
38. J. Duan and J. Li (2003) Compression of the layered depth image, *IEEE Transactions on Image Processing*, 12:365-372
39. S.U. Yoon, S.Y. Kim, and Y.S. Ho (2004) Preprocessing of depth and color information for layered depth image coding, *Lecture Notes Computer Science*, 3333: 622-699