MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Next-Generation 3D Formats with Depth Map Support

Chen, Y.; Vetro, A.

TR2014-016 April 2014

Abstract

This article reviews the most recent extensions to the Advanced Video Coding (AVC) and High Efficiency Video Coding (HEVC) coding standards, which integrate depth video to support advanced multiview and 3D video functionalities. All the extensions provide single-view compatibility, while some extensions add depth support on top of conforming stereoscopic bitstreams. To achieve the highest gains in coding efficiency, depth information is utilized in coding the texture views. The compression formats described in this article can be used to support emerging auto-stereoscopic displays and free-viewpoint video functionalities.

IEEE Multimedia

Anthony Vetro
Mitsubishi Electric Research Labs

Next-Generation 3D Formats with Depth Map Support

Ying Chen
Qualcomm

Anthony Vetro
Mitsubishi Electric Research Labs

The majority of 3D video content today is represented as stereoscopic video, which matches the stereoscopic display systems that could be found in theaters and consumer electronics devices. In most cases, glasses are required to enable depth perception of the visual scene that is being rendered. Although the use of glasses is generally accepted in cinema environment, the same is not true for home viewing. To address this obstacle, a new generation of 3D displays referred to as auto-stereoscopic displays, which emit view-dependent pixels and do not require glasses for viewing, are being developed. These displays require many distinct viewpoints of the scene—many more than can be practically captured or transmitted. As a result, these displays often employ depth-based image rendering techniques.

Recognizing this future need, next-generation 3D formats standards that include support for depth maps are also being developed. The depth maps not only allow for high-quality image rendering, but also allow for more efficient coding of the associated multiview video. This article provides a brief overview of the 3D video coding extensions that have been developed or are under development.

Background

The committees responsible for the development of video coding standards are the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG). These committees have jointly developed the widely deployed Advanced Video Coding (AVC) standard[1] and the recently finalized High Efficiency Video Coding (HEVC) standard.[2,3] They are currently working on 3D extensions of these standards under the Joint Collaborative Team on 3D Video (JCT-3V), which was established in July 2012. The 3D video extensions support the improved coding of stereoscopic and multiview video and facilitate advanced 3D capabilities such as view rendering through the use of depth maps.

The depth information itself may be extracted from a stereo pair by solving for stereo correspondences or may be obtained directly through special range cameras; it may also be an inherent part of the content, for example, in 3D computer graphics generated imagery. Depth-image-based rendering techniques can then be used to render distinct viewpoints of the 3D scene. The objective of the standardization work has been to specify an efficient representation of multiview plus depth video, where the multiple viewpoints of a 3D scene are represented by texture views and corresponding depth views. In contrast to conventional video coding standards, the coding efficiency is evaluated based on the quality of not only the reconstructed views, but also the synthesized views. This helps ensure that the quality of the depth maps is maintained to enable high-quality rendering of views.

We should also mention that the main emphasis of the standardization committee's work to date has assumed linear camera arrangements with conventional stereoscopic capture, but an additional third view has also been considered to allow for wider baseline rendering. Alternative camera arrangements with a greater number of views are considered an interesting topic for further study.

In this article, we first describe simple extensions of the AVC and HEVC standards that enable inter-view prediction of texture and depth videos. Following this, we provide an overview of more advanced coding tools. Although the AVC and HEVC coding frameworks differ, a number of key principles have been realized in the advanced extensions of both standards. Figure 1 gives a map of the different standards that have been developed or are under development.

1070-986X/14/$31.00 © 2014 IEEE

Published by the IEEE Computer Society

Enabling Inter-view Prediction

The multiview extension of AVC, which is referred to as MVC, has been developed for the coding of two or more views. A key feature of this standard is that it maintains AVC compatibility, meaning that one view of the compressed multiview bitstream can be decoded by a legacy AVC decoder. Relative to independent coding of all views, compression gain is achieved by allowing the other views to be predicted using pictures from different views at the same time instances. In this process, known as *inter-view prediction*, a block-based disparity shift between the reference view and the current views is determined and is used to perform a disparity-compensated prediction. This is similar to the motion-compensated prediction used in conventional video coding, but it is based on pictures with different viewpoints rather than pictures at different time instances.

The MVC approach is simply defined by two features: extending the high-level syntax so that the appropriate signaling of view identifiers and their references is supported and defining a process by which decoded pictures of other views can be used to predict a current picture in another view. Further details on MVC are available elsewhere.[4,5]

A similar multiview extension of HEVC, referred to as MV-HEVC, will be finalized in mid-2014. This specification follows the same design principles of the prior MVC extension in the AVC framework. As with the MVC design, this scheme enables inter-view prediction so that pictures from other views at the same time instance can be used to predict a picture in the current view, and it provides compatibility with single-view coding of HEVC.

Adding Depth Support

Building on standards for multiview coding, depth support has been added to enable depth-image-based rendering for the generation of additional viewpoints. For a number of use cases, it was considered desirable to maintain compatibility with the compressed stereoscopic bitstream.

In the case of AVC, it was decided to specify an independent second stream for the representation of depth as well as high-level syntax signaling of the necessary information to express the interpretation of the depth data and its association with the video data.[6] This approach does not involve macroblock-level changes to the AVC or MVC syntax, semantics, and decoding processes. The corresponding 3D video codec is referred to as MVC+D.

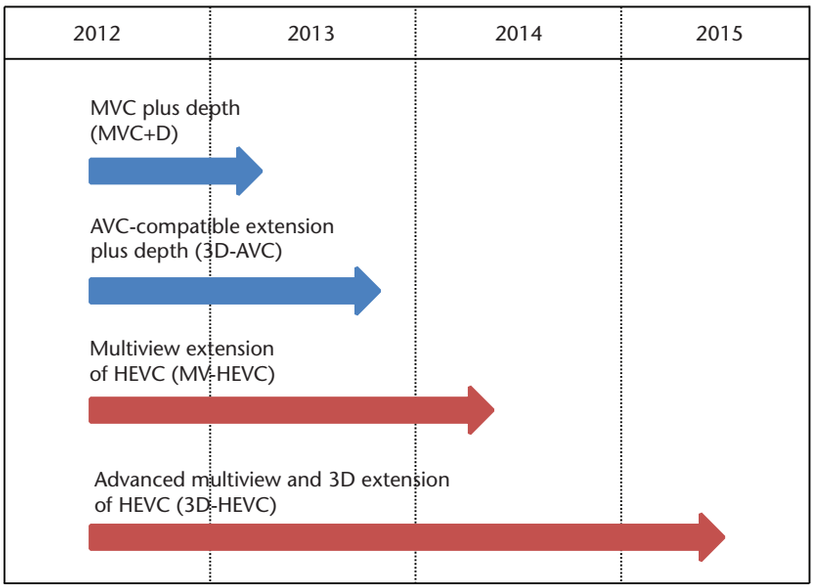

Figure 1. Various extensions to the Advanced Video Coding (AVC) and High Efficiency Video Coding (HEVC) standard. These extension have been developed or are under development.

In the MV-HEVC framework, support for depth maps will be realized with auxiliary picture syntax. The auxiliary picture decoding process would the same for video or multiview video, but it may not necessarily have normative decoding requirements as part of a profile. This would enable applications to optionally decode depth maps. If the industry desires an additional level of interoperability, a profile that requires the capability to decode depth would be added at a later stage. This approach differs from that taken in the MVC+D extension, which adds support for depth using specially designated network abstraction layer (NAL) units for depth and defines a dedicated profile for the decoding of multiview plus depth video.[7]

Advanced Coding Tools for Texture

To achieve higher coding efficiency, researchers have studied and evaluated advanced coding tools that better exploit inter-view redundancy. In contrast to the standards discussed in the previous section, block-level changes to the syntax and decoding process are considered to maximize the possible coding gain.

In the AVC family of standards, the 3D-AVC extension has recently been finalized, which

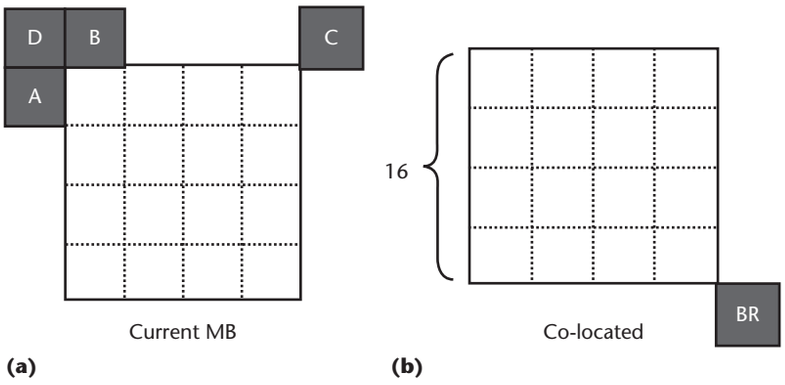

Figure 2. Neighboring blocks used as part of the NBDV process in 3D-AVC. (a) Spatial neighbors, (b) temporal neighbor.

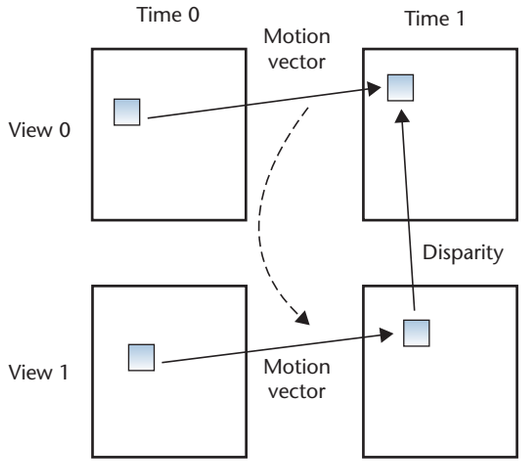

Figure 3. Motion prediction between views. The motion vector of view 1 is inferred from the motion vector of view 0 from corresponding blocks at time 1 based on the disparity between those blocks as derived by the NBDV process.

supports new block-level coding tools for texture views.[8] Depth views in 3D-AVC are coded similar to MVC+D, and no block-level changes for depth coding have been introduced. In 3D-AVC, the texture information of a single view is coded in a manner that is compatible with AVC.

In the HEVC family, the 3D-HEVC extension is in the process of being finalized.[9,10] As with 3D-AVC, block-level coding tools for texture views have been introduced into the draft specification of 3D-HEVC. In addition, novel techniques for improved coding of depth have been accepted into the draft specifications.

The following sections provide an overview of the most notable coding tools in 3D-AVC and 3D-HEVC. We focus on the main functionality and highlight differences in the designs between each standard.

Neighboring Block-Based Disparity Vector Derivation (NBDV)

As the name implies, this technique derives a disparity vector for a current block using an available disparity motion vector of spatial and temporal neighboring blocks. The derivation principle is the same in both 3D-AVC and 3D-HEVC, but the location of neighboring blocks differs slightly. In 3D-AVC, the bottom-right (BR) block of the colocated block is used as the temporal neighbor (see Figure 2), whereas 3D-HEVC uses the center block of the colocated block. Once a disparity vector in a neighboring block is identified, the NBDV process terminates and the derived disparity vector is set equal to the identified disparity vector.

The main benefit of this technique is that disparity vectors to be used for inter-view prediction can be directly derived without additional bits and independent of an associated depth picture. Disparity information can also be derived from the decoded depth picture when camera parameters are available.

Inter-view Motion Prediction

The motion information between views exhibit a high degree of correlation, and inferring it from one view to another view leads to notable gains in coding efficiency because good predictions generally reduce the bit rate required to send such information. To achieve this, the disparity, such as that derived by the NBDV process, is used to establish a correspondence between the blocks in each view. Figure 3 illustrates motion prediction between views.

The concept of inter-view motion prediction is supported in both the 3D-AVC and 3D-HEVC, but the designs differ. In 3D-AVC, inter-view motion prediction is realized with a new prediction mode, whereas in 3D-HEVC, it is realized by leveraging the syntax and decoding processes of the merge and advance motion vector prediction (AMVP) modes that were newly introduced by the HEVC standard. Additional information about this technique in 3D-HEVC is available elsewhere.[10]

View Synthesis Prediction (VSP)

This method uses the depth information to warp texture data from a reference view to the current view in order to generate a predictor for the current view. Although depth is often available with pixel-level precision, a block-based

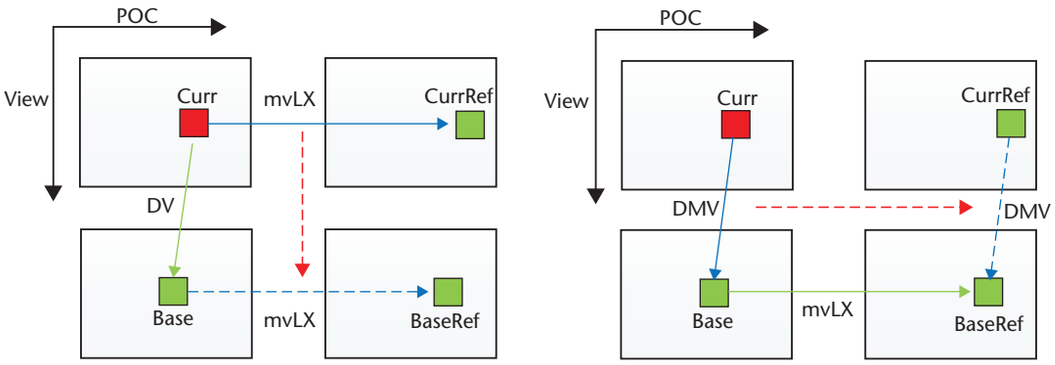

Figure 4. Advanced residual prediction (ARP) designs. (a) Temporal ARP, (b) inter-view ARP.

VSP scheme has been specified in both 3D-AVC and 3D-HEVC to align this type of prediction with existing modules for motion compensation. To perform VSP, the depth information of the current block is used to determine the corresponding pixels in the inter-view reference picture. Because texture is typically coded prior to depth, the depth of the current block can be estimated using the NBDV process we described earlier. In 3D-AVC, it is also possible to code depth prior to texture and hence obtain the depth information directly.

As with inter-view motion prediction, the same VSP concept is supported in both 3D-AVC and 3D-HEVC, but the designs differ significantly. VSP is supported in 3D-AVC with a high-level syntax flag that determines whether the reference picture to be used for prediction is an inter-view reference picture or a synthesized reference picture as well as a low-level syntax flag to indicate when skip/direct mode prediction is relative to a synthesized reference picture. In 3D-HEVC, the VSP design is realized by extensions of the merge mode, whereby the disparity and inter-view reference picture corresponding to the VSP operation is added to the merge candidate list.[10]

Illumination Compensation

Prediction may fail when cameras capturing the same scene are not calibrated in color transfer or by lighting effects. Therefore, illumination compensation was introduced to improve the coding efficiency for blocks predicted from inter-view reference pictures. This mode only applies to blocks that are predicted by an inter-view reference picture.

Inter-view Residual Prediction

Also known as advanced residual prediction (ARP) and only supported in 3D-HEVC, this prediction mode increases the accuracy of the residual predictor. In ARP, the motion vector is aligned for the current block and the reference block, so the similarity between the residual predictor and the residual signal of the current block is much higher, and the remaining energy after ARP is significantly reduced.

As Figure 4 illustrates, two types of ARP designs exist: temporal ARP and inter-view ARP. In temporal ARP, the residual predictor is calculated as a difference between the reference block (Base) and its reference block (BaseRef). With inter-view ARP, an inter-view residual is calculated from the temporal reference block in a different view (BaseRef) and its inter-view reference block, hypothetically generated by the disparity (DMV) that is signaled for the current block.

Advanced Coding Tools for Depth

To achieve higher compression efficiency, new coding tools have been adopted in 3D-HEVC for the coding of depth views. We introduce a few notable coding features here.

Depth Motion Prediction

Similar to motion prediction in texture coding, depth motion prediction is achieved by adding new candidates into the merge candidate list. The additional candidates include an inter-view merge candidate, a subblock motion parameter inheritance candidate, and a disparity-derived depth candidate.

Partition-Based Depth Intra Coding

To better represent the particular characteristics of depth, each depth block may be geometrically partitioned and more efficiently represented. In 3D-HEVC, these nonrectangular partitions are collectively referred to as depth modeling modes (DMMs). Two types of partitioning patterns are applied, including the wedglet pattern, which

segments the depth block with a straight line, and a contour pattern, which can support two irregular partitions.

Segment-Wise DC Coding (SDC)

This coding mode enables the transform and quantization process to be skipped so that depth prediction residuals are directly coded. It also supports a depth look-up table (DLT) to convert the depth values to a reduced dynamic range. SDC can be applied to both intra and inter prediction, including the new DMM modes. When the SDC mode is applied, only one DC predictor is derived for each partition, and based on that, only one DC difference value is coded as the residual for the whole partition.

Concluding Remarks

This article reviewed the most recent extensions to the AVC and HEVC coding standards, which integrate depth video to support advanced multiview and 3D video functionalities. All the extensions provide single-view compatibility, while some extensions add depth support on top of conforming stereoscopic bitstreams, such as MVC+D and MV-HEVC. To achieve the highest gains in coding efficiency, the 3D-AVC and 3D-HEVC standards utilize depth/disparity information in coding the additional texture views, but this requires changes to the block-level syntax and decoding process relative to single-view video decoding.

In terms of coding efficiency performance, experiments have shown that notable bit-rate reductions on the order of 30 to 40 percent relative to simulcast can be achieved by enabling inter-view prediction. It has also been reported that an additional 20 percent gain in coding performance could be achieved by 3D-AVC and 3D-HEVC with the additional block-level coding tools.

Although it has enjoyed great success and acceptance in theaters, 3D video has yet to realize its full potential for home entertainment and consumer electronics. The availability of 3D content will continue to increase and high-quality autostereoscopic displays that do not require glasses for 3D viewing will eventually be introduced into the market. It is also expected that 3D functionality will also become more prevalent on mobile devices in the coming years. The compression formats described here that include depth support could be used to support the multiview and 3D rendering capabilities that would be required by these emerging 3D applications and services. **MM**

References

1. *Rec. ITU-T H.264 and ISO/IEC 14496-10, Advanced Video Coding for Generic Audiovisual Services*, MPEG-4 AVC, 2014.
2. *Rec. ITU-T H.265 and ISO/IEC 23008-2, High Efficiency Video Coding*, Jan. 2013.
3. G.J. Sullivan et al., "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 22, no. 12, 2012, pp. 1649–1668.
4. Y. Chen et al., "The Emerging MVC Standard for 3D Video Services," *EURASIP J. Advances in Signal Processing*, vol. 2009, no. 1, 2009, article no. 8.
5. A. Vetro, T. Wiegand, and G. J. Sullivan, "Overview of the Stereo and Multiview Video Coding Extensions of the H.264/AVC Standard," *Proc. IEEE*, vol. 99, no. 4, 2011, pp. 626–642.
6. Y. Chen et al., "Overview of the MVC+D 3D video coding standard," *J. Visual Communication and Image Representation*, Apr. 2013; doi: http://dx.doi.org/10.1016/j.jvcir.2013.03.013.
7. G. Tech et al., "MV-HEVC Draft Text 5," Joint Collaborative Team on 3D Video Coding Extensions (JCT-3V) document JCT3V-E1004, 5th Meeting, Vienna, AT, 27 July-2 Aug. 2013.
8. M.M. Hannuksela et al., "3D-AVC Draft Text 8," Joint Collaborative Team on 3D Video Coding Extensions (JCT-3V) document JCT3V-F1002, 6th Meeting, Geneva, CH, 25 Oct. 1 Nov. 2013.
9. G. Tech et al., "3D-HEVC Draft Text 3," Joint Collaborative Team on 3D Video Coding Extensions (JCT-3V) document JCT3V-G1001, 7th Meeting: San Jose, USA, 11–17 Jan. 2014.
10. G.J. Sullivan et al., "Standardized Extensions of High Efficiency Video Coding (HEVC)," *IEEE J. Selected Topics in Signal Processing*, vol. 7, no. 6, 2013, pp. 1001–1016.

Ying Chen is a senior engineer at Qualcomm. His research interests include video coding, multimedia processing and communication. Chen has a PhD in electrical engineering from Tampere University of Technology. Contact him at cheny@qti.qualcomm.com.

Anthony Vetro is the multimedia group manager at Mitsubishi Electric Research Labs. His research interests include various aspects of multimedia and signal processing. Vetro has a PhD in electrical engineering from Polytechnic University. Contact him at avetro@merl.com.