

## Trends of International Standardization of Three-dimensional Video Coding

*Hideaki Kimata*<sup>†</sup>

### Abstract

This article introduces the trends of international standardization of three-dimensional video, which should lead to new services that differ from conventional television.

### 1. Introduction

Video compression technologies have made possible various exciting new services since their international standardization started. While the main target has been high compression efficiency, a lot of attention has recently been paid to three-dimensional (3D) video. International standardization of 3D video has started and it is expected to lead to novel video services. This article introduces the trends of 3D video format and compression standardization, which have been studied in the Motion Picture Experts Group (MPEG), and mentions the background, applications, and features of the H.264 Annex H (MVC) standard, which has been adopted in the fixed-media Blu-ray 3D format (MVC: Multiview Video Coding).

### 2. 3D video and multiview video

We live in a 3D space. Therefore, if we could transmit the 3D space itself, that would be a straightforward way to communicate correctly what one sees as the scenery and appearance of objects to other people. However, it is currently impossible to capture the 3D space itself. Instead, a two-dimensional (2D) image projected from the 3D space is captured by a camera and transmitted. Using several camera images captured by multiple cameras enables a large amount of detailed spatial information to be communicated. A

collection of multiple videos capturing the same scene from different viewpoints is called multiview video. 3D video is a more general medium for representing 3D space and it may include 3D geometry information of the scene as well as multiview video. By transmitting 3D video, we can communicate a more realistic impression of the scenery and appearance of objects than conventional TV can.

### 3. 3D video and H.264 standardization

3D video representation and compression formats have been standardized to achieve 3D video services. 3D video standardization activities started about ten years ago. This was the time when TV services began to change from analog broadcasting to digital broadcasting, and also it was when personal video delivery services such as video on demand started. Against this background, high-compression-efficiency video coding H.264/MPEG-4 AVC (advanced video coding) standardization by the joint team of ITU-T and ISO/IEC JTC1 was started in 2001 (ITU-T: International Telecommunication Union, Telecommunication Standardization Sector, ISO: International Organization for Standardization, IEC: International Electrotechnical Commission, JTC1: Joint Technical Committee 1). On the other hand, expansion of services using new video formats was expected at the same time (**Fig. 1**). Since it was predicted that high-definition television (HDTV) displays would be popularized within a few years, such new video formats needed to handle different representations from

<sup>†</sup> NTT Cyber Space Laboratories  
Yokosuka-shi, 239-0847 Japan

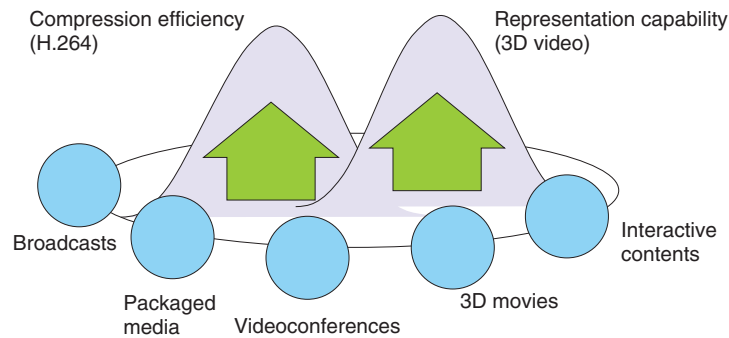


Fig. 1. Expansion of services resulting from video coding standards.

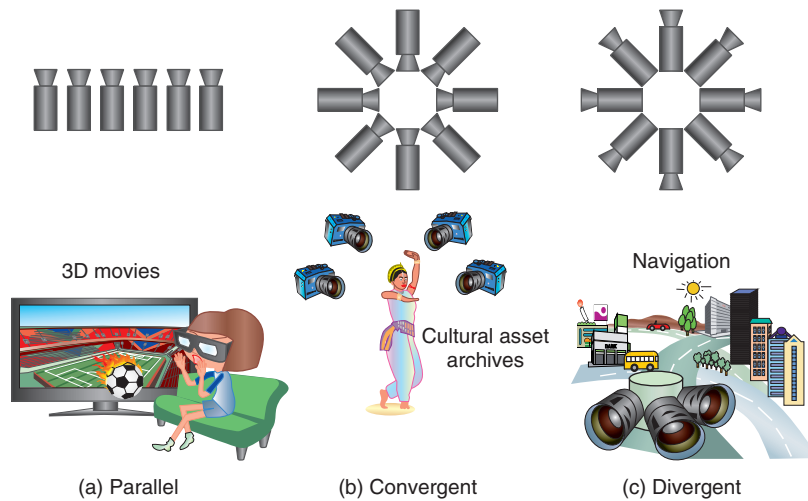


Fig. 2. Camera arrangements and applications.

ordinary TV and Internet video delivery services. Moreover, a new video format that would treat 3D space was expected. The standardization of video representation and compression formats for 3D space was started in 2001 by MPEG, which is a Working Group under ISO/IEC JTC1. The MPEG 3D space activity is called 3D video standardization.

In 3D video standardization, first, video representation formats were classified according to the arrangement of cameras constructing the 3D video. If the relative differences in positions and directions among cameras are simple, camera arrangements are classified into three types: parallel, convergent, and divergent arrangements (**Fig. 2**). Applications of 3D video are also classified according to the camera arrangement.

The parallel arrangement provides enhanced depth

impression, as typified by the current boom in 3D movies. The convergent arrangement is suitable for free viewing around a tangible or intangible cultural asset. The divergent arrangement is suitable for free viewing of city scenery from the street, as typified by the current boom in web maps and navigation services.

At about the same time, the depth camera, which captures depth (distance) information in 2D arrays, appeared. A depth camera can be used alongside a normal video camera to simultaneously acquire depth and video (color) information (**Fig. 3**). Alternatively, depth information can be estimated by computer vision technologies from multiple cameras images. And from the depth and video (color) information, a virtual camera image representing the image that would have been taken by a camera having a slightly

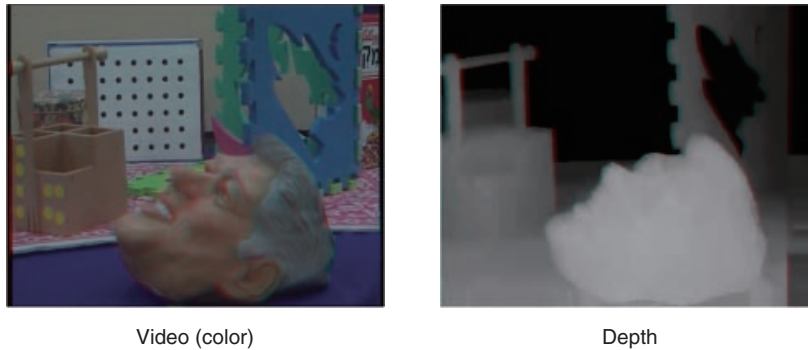


Fig. 3. Video (color) and depth information.

Table 1. Characteristics of MVC profiles.

Profile	Multiview High Profile	Stereo High Profile
Maximum views	1024	2
Interlace scan	Not supported	Supported

different position or angle from the original cameras, can be synthesized by computer graphics (CG) technologies. MPEG also discussed a 3D video data format supporting image synthesis from depth information. For instance, depth information can be treated as a gray-scale image because it is obtained in a 2D array. The assumed application was 3D TV.

Because applications were classified according to camera arrangement and camera type, experiments exploring the efficiency of data formats were conducted in case studies with end users for 3D video standardization discussions. From the deep discussion based on the evaluation results, it was concluded that the fundamental data format for 3D video was multiview video, geometry information about the videos, and depth information according to the conditions. Standardization of Multiview Video Coding (MVC) as a compression format of multiview video started in 2005.

When MVC standardization was about to start, the standardization of the state-of-the-art high-efficiency video compression format H.264 had just finished and MVC was expected to be standardized over H.264. Sure enough, MVC standardization started as an extension of H.264 and it was finished in 2009 and specified as H.264 Annex H. In H.264, the operational guidelines are specified as profiles, and MVC specifies two profiles. The main characteristics of MVC profiles are given in **Table 1**. Historically, the

Multiview High Profile was specified first and the Stereo High Profile was specified a couple of years later.

#### 4. Characteristics of MVC standard

Here, the main characteristics of the MVC standard are introduced. The main video format is multiview video, and several functional extensions were required to support the assumed 3D video applications. The main features discussed in MVC standardization are listed in **Table 2**. In MVC, virtual camera image synthesis is considered as the final process before display. The series of processes from capture to display assumed in MVC is depicted in **Fig. 4**. On the subject of video streaming, the case of a feedback streaming channel from a user was also discussed. When no feedback streaming channel is available, it is necessary to transmit all video streams; however, when it is available, it is possible to transmit only the requested part of the compressed stream. Accordingly, at the acquisition and encoding side, all information should be compressed and then at the display side, a user will see part of the total. The idea of transmitting only part of the stream is an extension of conventional scalable video coding. In scalable video coding, the stream is composed of multiple layers: the fundamental layer called the base layer should always be transmitted and other layers are optionally transmitted in addition to the base layer. On the other hand, in MVC, because an arbitrary part of stream should be extractable and transmittable, a base layer is not defined as a specific layer. Instead, the novel unit of a *view* is introduced. The function of extracting an arbitrary part of the stream is called view scalability in MVC. Note that the use of view was

Table 2. Main features of MVC standard.

Simultaneous output of multiple views	Application to 3D displays
Predictive coding between views	High compression efficiency
Support for view scalability and extraction of arbitrary view	Support for various types of 3D display and improved transmission efficiency
Inexpensive implementation of extension from video codec	Low cost because only predictive coding is used between views

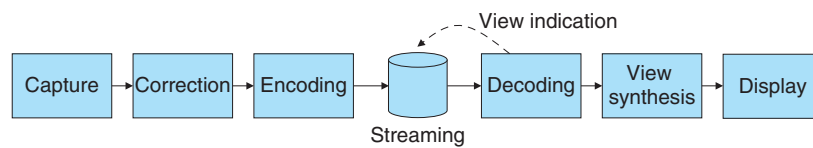


Fig. 4. Flow of processes assumed in the MVC standard.

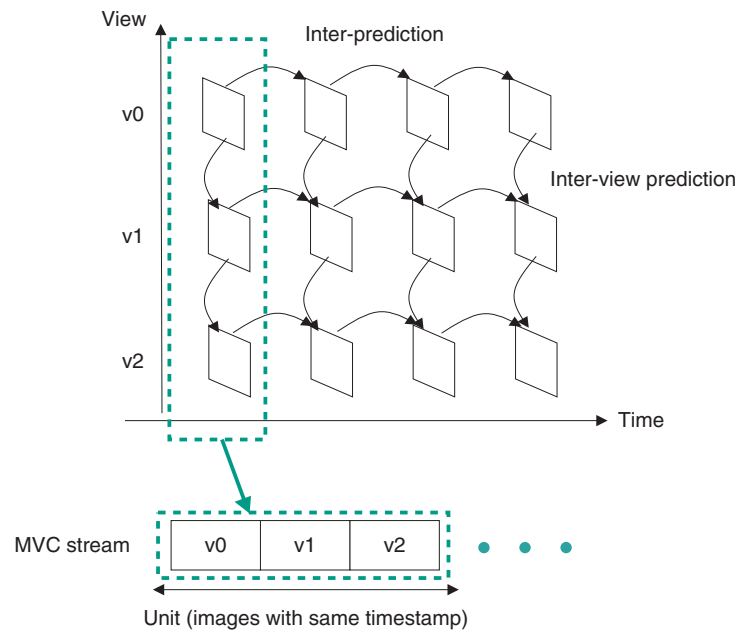


Fig. 5. Structure of MVC stream.

intended to distinguish camera images, but in the MVC standard, it is not limited to camera images. For instance, a view can be assigned to a region of an image and it can be used to extract part of a huge image. MVC supports many data formats besides multiview video as a result of the introduction of this versatile view concept.

In MVC, multiple views are synchronized and

compressed together. Images with the same timestamp for views are compressed sequentially according to the predetermined ordering of views. At the decoding side, the decoder just decodes part of the stream and obtains the images with the same timestamp as the target views (Figs. 5 and 6). Thus, synchronization is achieved at the codec level and extraction of the target view stream is easily achieved. And

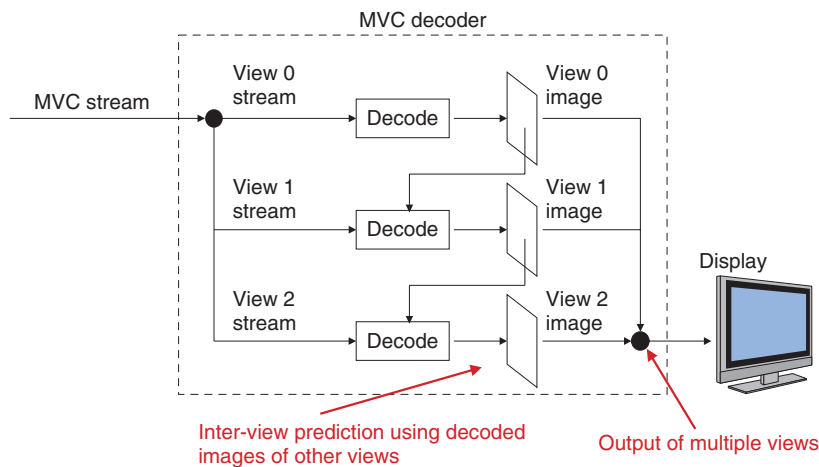


Fig. 6. Example of MVC decoder component blocks.

to improve compression efficiency, predictive coding between views can be applied when images having the same timestamp are compressed. This feature is efficient for compressing 3D video and achieves a significant reduction in the data rate because the similarity among views is high in 3D video.

Moreover, one feature for implementation is that the compression process is not very different from that in a conventional video compression codec. Since the only difference is predictive coding between views, an MVC decoder can easily be constructed from a conventional video decoder when images of views are decoded sequentially.

## 5. Standardization of related technologies

Several related technologies have been standardized. For the delivery of video contents, the system layer standard MPEG-2 TS (transport stream) has been extended to support MVC. This extension has led to not only the transmission of 3D video but also the adoption of MVC as the fixed-media standard Blu-ray 3D format.

For 3D video viewing without the need for special glasses, the delivery of supplementary information about the view synthesis process should be considered because the positions and directions of displayed views are usually dependent on the capabilities of the display system. Two types of supplementary information have been standardized. One is the camera parameter of views. This is a description of the camera's position and direction. The camera parameter is necessary in order to assume the position of signals of

views in the 3D space, which are exploited to synthesize a virtual camera image by the CG method of 3D warping. The camera parameter description is specified in the MVC standard as supplementary information about the compressed stream. The other supplementary information is the relationship between 3D scene depth and disparity. This relationship is applied when the depth information that is compressed in the stream is converted into disparity information for 3D display. The supplementary information about this relationship has been standardized as MPEG-C part 3, which is independent of compression standards.

For stereoscopic video, the conventional compression format is currently used in TV broadcasting instead of the brand-new compression format MVC. In order to transmit two views, stereoscopic video is transformed into a single side-by-side (SBS) formatted video, and mapping information about the SBS transformation is necessary for the display side to perform the inverse transformation. This mapping information has been standardized for H.264, and it is under consideration for MPEG-2 Video.

## 6. 3D video applications

Let us return to the topic of 3D video applications. First, for the parallel camera arrangement, 3D TV is assumed. Stereoscopic video is becoming popular as a result of the recent boom in 3D movies, and the MVC standard has been adopted as the Blu-ray 3D format to deliver high-quality stereoscopic video to homes. In the Blu-ray 3D format, the view scalability feature of MVC is exploited efficiently. For 3D

display, both views are extracted and displayed; for ordinary display, either one view or the other is extracted and displayed.

On the other hand, the convergent and divergent camera arrangements are expected to lead to more exciting services in the near future. For instance, the convergent arrangement may be used for free view-point video covering the whole of a sports stadium, and the divergent arrangement may be used for panoramic video that gives us a strong feeling of actually being at the site of an event. My coworkers and I have studied video processing technologies for a panoramic video service.

## 7. Conclusions and future activities

This article introduced the trends of 3D video standardization and the novel video services that may result from it. In particular, features and applications of H.264 Annex H (MVC), which has been adopted in the fixed-media Blu-ray 3D format, were explained.

In MPEG, a new activity concerning the next standard after MVC is now under way. To support glasses-free 3D video more efficiently, a new representation and compression format of video and depth are

being studied. This activity is called FTV (free view-point TV) in MPEG. An efficient standard should be specified in the very near future, in time for the current 3D movie boom.

On the other hand, video compression standardization of High Efficiency Video Coding (HEVC), which should be the successor to H.264, is also under way. Like the relation between 3D video and H.264, another 3D video standardization linked to HEVC can be expected in the future.

## References

- [1] A. Smolic, H. Kimata, and A. Vetro, "Development of MPEG Standards for 3D and Free Viewpoint Video," Proc. of SPIE Optics East, Three-Dimensional TV, Video, and Display IV, Boston, MA, USA, Oct. 23–26, 2005.
- [2] H. Kimata, "3D Video Formats and Compression for Content Distributions," Proc. of the 17th International Display Workshops, Fukuoka, Japan, Dec. 2010.
- [3] H. Kimata, M. Isogai, H. Noto, M. Inoue, K. Fukazawa, and N. Matsuura, "Interactive Panorama Video Distribution System," Proc. of ITU Telecom World 2011 Symposium, Oct. 2011 (to be published).
- [4] H. Kimata, M. Kitahara, K. Kamikura, and Y. Yashima, "Free-view-point Video Communication Using Multi-view Video Coding," NTT Technical Review, Vol. 2, No. 8, pp. 21–26, Aug. 2004. <https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr200408021.pdf>



**Hideaki Kimata**

Senior Research Engineer, Visual Media Communications Project, NTT Cyber Space Laboratories.

He received the B.E. and M.E. degrees in applied physics and the Ph.D. degree in electrical engineering from Nagoya University, Aichi, in 1993, 1995, and 2006, respectively. Since joining NTT in 1995, he has been engaged in R&D of high-efficiency video coding and error-resilient video coding algorithms and high-reality visual communication systems. He acted as an editor of the ITU-T and ISO/IEC international standard H.264 Annex H (MVC). He is a member of the Institute of Electronics, Information and Communication Engineers.