Towards a 3D Video Format for Auto-Stereoscopic Displays

Anthony Vetro, Sehoon Yea, Aljoscha Smolic

Abstract

There has been increased momentum recently in the production of 3D content for cinema applications; for the most part, this has been limited to stereo content. There are also a variety of display technologies on the market that support 3DTV, each offering a different viewing experience and having different input requirements. More specifically, stereoscopic displays support stereo content and require glasses, while auto-stereoscopic displays avoid the need for glasses by rending view-dependent stereo pairs for a multitude of viewing angles. To realize high quality auto-stereoscopic displays, multiple views of the video must either be provided as input to the display, or these views must be created locally at the display. The former approach has difficulties in that the production environment is typically limited to stereo, and transmission bandwidth for a large number of views is not likely to be available. This paper discusses an emerging 3D data format that enables the latter approach to be realized. A new framework for efficiently representing a 3D scene and enabling the reconstruction of an arbitrarily large number of views prior to rendering is introduced. Several design challenges are also highlighted through experimental results.

Towards a 3D video format for auto-stereoscopic displays

Anthony Vetro[a], Sehoon Yea[a], Aljoscha Smolic[b]

[a] Mitsubishi Electric Research Laboratories, 201 Broadway, Cambridge, MA 02139, USA
[b] Fraunhofer Institute for Telecommunications - Heinrich-Hertz-Institut, Berlin, Germany

ABSTRACT

There has been increased momentum recently in the production of 3D content for cinema applications; for the most part, this has been limited to stereo content. There are also a variety of display technologies on the market that support 3DTV, each offering a different viewing experience and having different input requirements. More specifically, stereoscopic displays support stereo content and require glasses, while auto-stereoscopic displays avoid the need for glasses by rendering view-dependent stereo pairs for a multitude of viewing angles. To realize high quality auto-stereoscopic displays, multiple views of the video must either be provided as input to the display, or these views must be created locally at the display. The former approach has difficulties in that the production environment is typically limited to stereo, and transmission bandwidth for a large number of views is not likely to be available. This paper discusses an emerging 3D data format that enables the latter approach to be realized. A new framework for efficiently representing a 3D scene and enabling the reconstruction of an arbitrarily large number of views prior to rendering is introduced. Several design challenges are also highlighted through experimental results.

Keywords: 3D video, auto-stereoscopic display, multiview video coding, depth maps, H.264/AVC, view synthesis, depth estimation, rendering, stereo, MPEG.

1. INTRODUCTION

3D depth perception of observed visual scenery can be provided by 3D display systems which ensure that the user sees a specific different view with each eye [1]. Such a stereo pair of views must correspond to the human eye positions. Then the brain can compute the 3D depth perception. History of 3D displays dates back almost as long as classical 2D cinematography. In the past, users had to wear specific glasses (anaglyph, polarization, or shutter) to ensure separation of left and right view which were displayed simultaneously. Together with limited visual quality this is regarded as main obstacle for wide success of 3D video systems in home user environments.

In a 3D cinema environment, wearing glasses is more acceptable for users, since it is for a limited time only and has more the character of an event. Moreover, modern 3D cinema theaters provide an excellent visual quality. Therefore, 3D cinema has become very popular recently. Awareness of and interest in 3D video is rapidly increasing, among users who wish to experience the extended visual sensation, as well as among content producers, equipment providers, and distributors, who discover new business opportunities. At the same time technology is maturing from capture to display. The market of 3D cinema is expected to continue growing rapidly over the next years.

Naturally, 3D video is also an increasingly interesting technology for home user living room applications. However, the user habits and requirements in the home are very different from cinema applications, where everyone is sitting in a comfortable chair for a given time and paying full attention to the presented content without moving or interacting much. 3D video content will arrive to the home by 3DTV broadcast, 3D-DVD/Blu-ray, Internet, etc. Standard formats and efficient compression are crucial for the success of 3D video at home. A generic, flexible and efficient 3D video format that would serve a wide range of different 3D video systems is highly desirable in this context.

Currently, there is a great variety of different 3D display systems designed for the home user applications, starting from classical 2-view stereo systems with glasses. More sophisticated candidates for 3D vision in living rooms are multiview auto-stereoscopic displays, which do not require glasses [1]. They emit more than one view at a time but the technology ensures that users only see a stereo pair from a specific viewpoint. Today's 3D displays are capable of showing 9 or more different images at the same time, of which only a stereo pair is visible from a specific viewpoint. This supports multi-user 3D vision without glasses in a living room environment. Motion parallax viewing can be supported if consecutive views are stereo pairs and arranged properly.
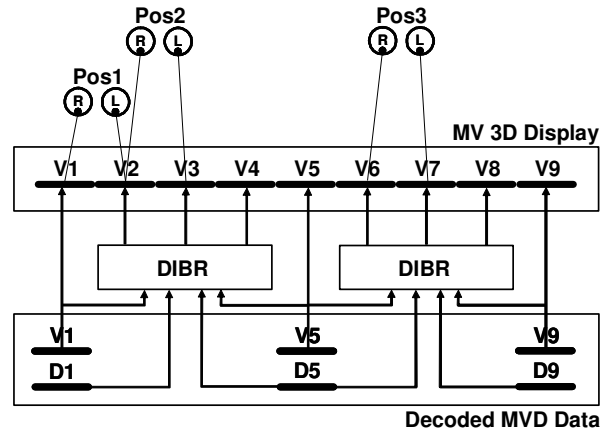
Fig. 1. Advanced 3D video concept based on a multiview plus depth format; Pos: viewpoint, R: right eye, L: left eye, V: view/image, D: depth.

As illustrated in Fig. 1, a user at position 1 sees views 1 and 2 with right and left eye, respectively. Another user at position 3 sees views 6 and 7, hence multi-user 3D viewing is supported. Assume a user moves from position 1 to position 2. Now views 2 and 3 are visible with the right and left eye, respectively. If V1 and V2 is a stereo pair with proper human eye distance baseline, then V2 and V3 are also a pair, and so on; a user moving in front of such a 3D display system will perceive a 3D impression with disocclusions and occlusions of objects in the scenery depending on their depth. However, this motion parallax impression will not be seamless and the number of different positions is restricted to N-1.

To enable the functionality described above results in a tremendous increase of data rate, i.e., N times the bit rate for compressed transmission compared to 2D video, if all views are treated independently. Multiview video coding (MVC) including inter-view prediction typically reduces the overall bit rate by 20% [2], which appears to be still far too high for most application scenarios.

A more efficient approach for stereo video (i.e., 2 views only) is to use a video plus depth format [3]. MPEG recently released a corresponding standard known as MPEG-C Part 3, which will be briefly further in this paper. With this format, a receiver can reproduce the stereo video by depth image-based rendering (DIBR) of the second video. It has been shown that the depth data can be compressed very efficiently, and that the resulting bit rate is much smaller than the bit rate for the corresponding stereo video, while providing the same visual quality. This concept works well if the virtual view to be rendered is close to the available view. Extrapolation artifacts increase with the distance of the virtual view. as a result, this format is not suitable for 3D video systems with a large number of views.

The solution for advanced 3D video systems presented in this paper is the extension and combination to multiview video plus depth. In the example in Fig. 1, only a subset of M = 3 views is transmitted to the receiver. For these views, sample-accurate depth maps have to be generated at the sender side and conveyed together with the video signal. All other views to be displayed are generated by DIBR at the receiver. Such a format is generic and flexible since it supports any type of 3D display from simple 2 view stereo to an arbitrary large number of output views. It is efficient since only a few views with depth are actually coded. This approach is currently being investigated in MPEG as the future 3D video format. The design of such a 3D video system includes a number of difficult and partially unresolved issues that still require further research [4]. This includes multiview capture, depth estimation/generation, parameterization of the system (such as the number of input views), efficient compression of the depth data, transmission and rendering.

The rest of this paper is organized as follows. In the next section, a brief overview of existing 3D video formats is provided. In section 3, the emerging 3D video framework being considered by MPEG is introduced. Section 4 provides some experimental results and highlights several issues that must be considered in the evaluation. Concluding remarks are given in section 5.

2. EXISTING 3D VIDEO FORMATS

In this section, a brief overview of existing 3D video formats, including both stereo and multiview formats, is provided. The merits of each format will be discussed along with the drawbacks and limitations.

2.1 Simulcast

The most obvious and straightforward means to represent stereo or multi-view video is simulcast, where each view is encoded independent of the other. This solution has low complexity since dependencies between views are not exploited, thereby keeping computation and processing delay to a minimum. It is also a backward compatible solution since one of the views could be decoded for legacy 2D displays. With simulcast, each view is assumed to be encoded with full spatial resolution, and the main drawback is that coding efficiency is not maximized since redundancy between views is not considered. However, prior studies on asymmetrical coding of stereo, whereby one of the views is encoded with less quality, suggest that substantial savings in bitrate for the second view could be achieved. In this way, one of the views is more coarsely quantized than the other [5] or coded with a reduced spatial resolution [6], yielding an imperceptible impact on the stereo quality. Further study is needed to understand how this phenomenon extends to multiview video.

2.2 Stereo Interleaving

There is a class of formats for stereo content that we collectively refer to as stereo interleaving. This category includes both time multiplexed and spatial multiplexed formats as shown in Fig. 2. In the time multiplexed format, the left and right views would be interleaved as alternating frames or fields. With spatial multiplexing, the left and right views would appear in either a side-by-side or over/under format. As is often the case with spatial multiplexing, the respective views are "squeezed" in the horizontal dimension or vertical dimension to fit within the size of an original frame. The drawback of representing the stereo signal in this way is that spatial resolution would be lost.

To distinguish the left and right views, some additional out-of-band signaling is necessary. For instance, the H.264/AVC standard specifies a Stereo SEI message that identifies the left view and right view; it also has the capability of indicating whether the encoding of a particular view is self-contained, i.e., frame or field corresponding to the left view are only predicted from other frames or fields in the left view. Inter-view prediction for stereo is possible when the self-contained flag is disabled. Similar type of signaling would be needed for the spatially multiplexed content.

The major drawback of all stereo interleaving approaches is that it is not possible for legacy 2D devices to extract, decode and display a 2D version of the 3D program. These legacy devices are not designed to understand the special signaling for stereo content and will only be able to decode and display a stream with alternating frames or in the side-by-side format. While this might not be so problematic for packaged media since special 3D disc players could be upgraded and new discs could store both 2D and 3D formats, it is certainly a major issue for broadcast services where the transmission channel is limited and devices cannot be upgraded.

(a)　　　　　　　　　(b)　　　　　　　　　(c)
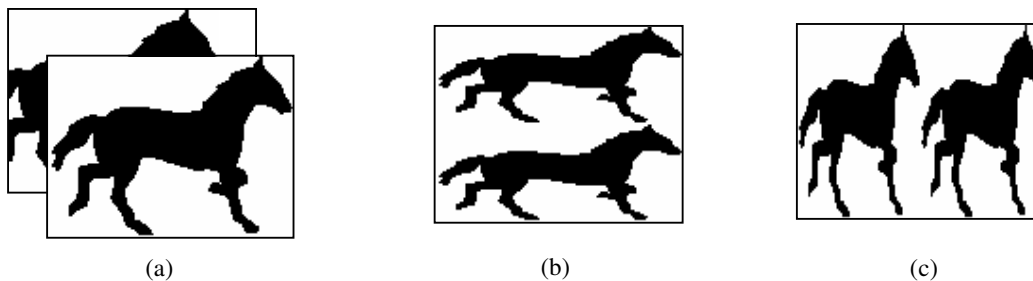
Fig. 2. Various stereo interleaving formats (a) time multiplexed frames, (b) spatial multiplexed as over/under, (c) spatial multiplexed as side-by-side. Images from http://www.stereo3d.com.

2.3 2D + Depth

Another well-known representation format is the 2D plus depth format. The inclusion of depth enables a display-independent solution for 3D that supports generation of an increased number of views as need by any stereoscopic display. A key advantage is that the main 2D video provides backward compatibility with legacy devices. Also, this representation is agnostic of coding format, i.e., the approach works with both MPEG-2 and H.264/AVC. The main drawback is that the format is only capable of rendering a limited depth range and has problems with occlusions. ISO/IEC 23002-3 (also referred to as MPEG-C Part 3) specifies the representation of auxiliary video and supplemental information. In particular, it enables signaling for depth map streams to support 3D video applications.

2.4 Multiview Video Coding

To improve coding efficiency of multiview video, the redundancy over time and across views could be exploited. In this way, pictures are not only predicted from temporal neighbors, but also from spatial neighbors in adjacent views. The Multiview Video Coding (MVC) standard has been specified as an amendment of H.264/AVC. It has been shown that MVC gives significantly better results compared to simple AVC-based simulcast [2]. Specifically, improvements of more than 2 dB have been reported for the same bitrate, and subjective testing has indicated that the same quality could be achieved with approximately half the bit-rate for a number of test sequences.

A key aspect of the MVC design is that it does not require any changes to lower-level syntax, so it is quite compatible with single-layer AVC hardware. Also, it is mandatory for the compressed multiview stream to include a base layer stream that could be easily extracted and used for backward compatibility with legacy 2D displays; this base layer stream is identified by the NAL unit type in H.264/AVC. Furthermore, inter-view prediction is enabled through flexible reference picture management, where decoded pictures from other views are made available in the decoded reference picture buffer. It is important to emphasize that the core decoding modules do not need to be aware of whether reference picture is a time reference or multiview reference from another view. In terms of syntax, the standard only requires small changes to high-level syntax, e.g., view dependency needs to be known for decoding.

MVC was designed mainly to support auto-stereoscopic displays that require a large number of views. However, large camera arrays are not common in the current acquisition and production environments. Furthermore, although MVC is more efficient than simulcast, the rate of MVC encoded video is still proportional to the number of views. Of course, this varies with factors such as scene complexity, resolution and camera arrangement, but when considering a high number of views, the achievable rate reduction might not be significant enough to overcome constraints on channel bandwidth.

It is important to note that in the near-term MVC is still useful for delivery of stereo contents. While some rate reduction could certainly be achieved relative to simulcast, the more important factor might be the backward compatibility that it provides to existing 2D services. In contrast to some of the stereo interleaving solutions, such as side-by-side, the full resolution could also be maintained. MVC also has benefits relative to the 2D + Depth format for stereo in terms of rendering quality for generic scenes.

3. EMERGING 3D VIDEO FRAMEWORK

The target for the emerging 3D Video (3DV) framework is shown in Fig. 3. Due to limitations in the production environment, the data format is assumed to be based on a limited camera input; stereo content is most likely, but up to three views might also be considered. In order to support a wide range of auto-stereoscopic displays, it should be possible for an arbitrarily large number of views to be generated from this data format. Finally, and very importantly, the rate required for transmitting this representation of a 3D scene should be fixed. This means that for a given resolution, the introduction of 3D services should not incur substantial rate overhead. Additionally, there should not be an increase in the rate simply because the display requires a higher number of views. This decoupling of the transmission rate and the number of output views, while achieving a high quality output, is the main novelty of this framework.
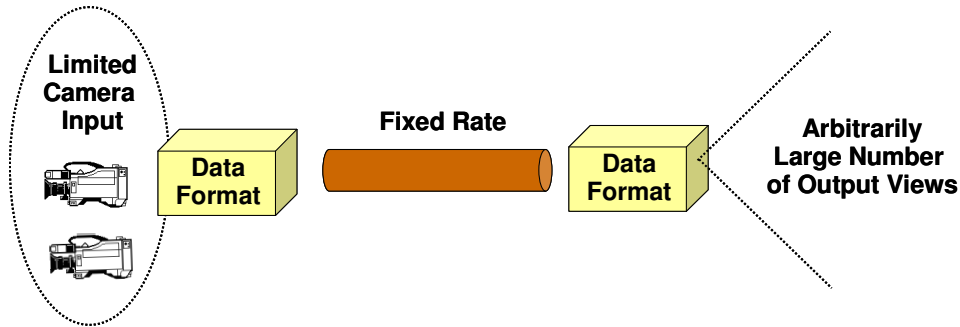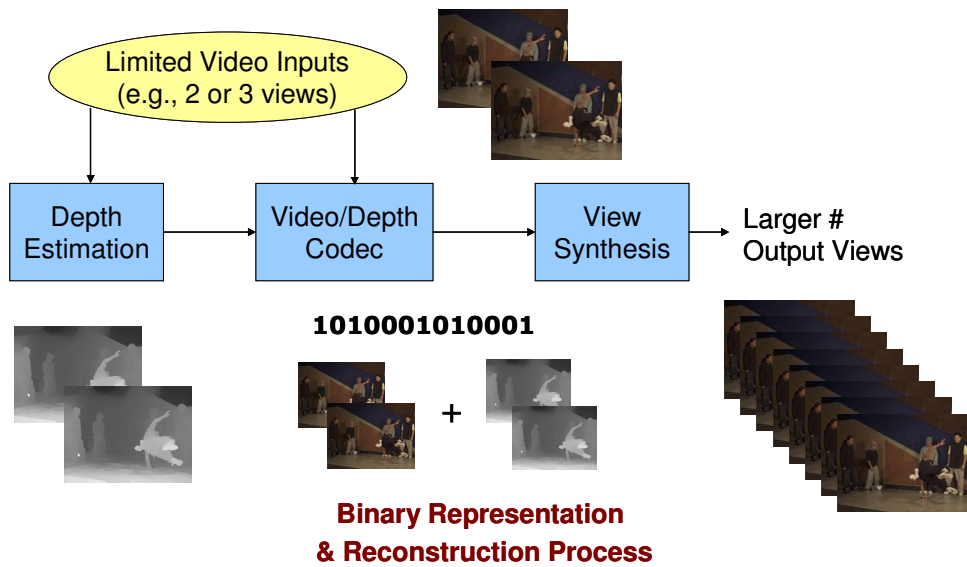
Fig. 3. Target of 3D video framework including limited camera input, fixed rate transmission and capability to render an arbitrarily large number of output views.

Given the above requirements, it is quite natural to expect that the format will be comprised of stereo video with a depth map representation of the scene, possibly multiple depth maps corresponding to different views. Such a format would almost certainly incur some rate increase relative to a conventional 2D or the 2D + Depth format discussed in the previous section. However, rendering quality is expected to increase due to the availability of more than a single 2D view since occlusions could be handled better. It would be a target to enable rendering quality similar to that simulcast or MVC in which all output views are directly encoded. The main advantage of 3DV in this case would be a substantial reduction in bit-rate compared to these formats. The trade-off between 3D rendering capability and bit rate for the different formats discussed is illustrated in Fig. 4.
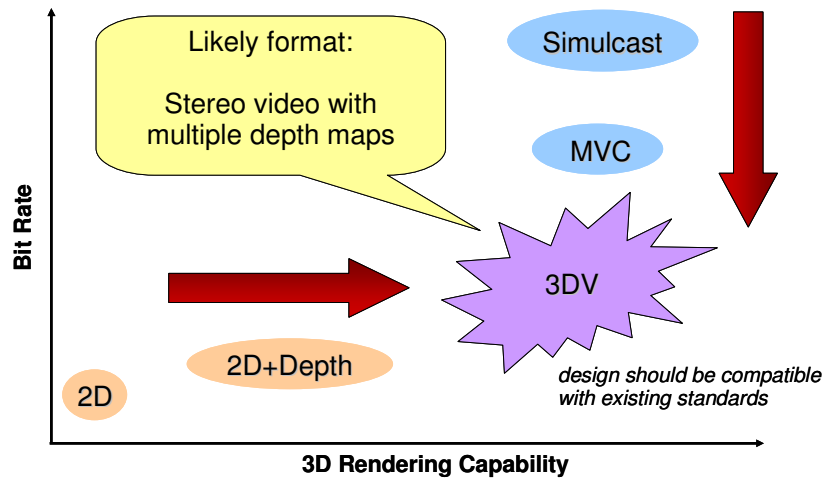
Fig. 4. Illustration of 3D rendering capability versus bit rate for different formats, where 3D Video aims to improve rendering capability of 2D+Depth format while reducing bit rate requirements relative to simulcast and MVC.

Assuming limited video inputs, the main processing steps to realize a 3DV framework is provided in Fig. 5. From the video inputs, depth estimation would be performed. The multiview video and estimated depth data would then be encoded, transmitted and reconstructed. At the receiver, a view synthesis module would be used to generate the required number of output views for the target display. The steps outlined in this framework present a number of challenges. In the following, some issues related to depth estimation and accuracy, as well as the view generation process, are highlighted.

Fig. 5. Processing steps in 3D video framework including depth estimation, video and depth codec and view synthesis (intermediate view generation). In this framework, a larger number of output views are generated from limited video inputs.

Estimating depth from stereo image pairs is a classic computer vision problem that has been well studied in the literature and continues to be an active area of research, see e.g., [7][8]. As evident from these references, there are a variety of algorithms that provide different estimation results. Of course, it is very important that the estimated depth values accurately represent the scene so that high-quality intermediate views could be generated at the receiver. Furthermore, in the context of a compression system, the sampling precision of the estimated depth will have a notable impact on the required transmission rate as well as the view generation process; this aspect of the design requires thorough consideration. From a pure rendering perspective, it is always advantageous to have finer sampling precision, e.g., pixel-level depth values without much quantization noise. However, such a fine sampling of the depth information will typically lead to a large rate overhead, so the depth estimation and codec must consider ways to reduce this information, e.g., through quantization or sub-sampling. Some of the trade-offs involved in reducing the sampling precision of the depth will be further discussed in the next section.

A popular technique for performing view generation is to perform a 3D warping of the available reference images using depth information of the scene. Conceptually, this is achieved by projecting points from the reference images into the 3D space according to the depth, then re-projecting these values to a new image plane corresponding to the view being generated. Clearly, this view generation process is highly dependent on the sampling precision of the depth. Additionally, there will typically be artifacts resulting from the 3D warping process due to occluded areas in the reference views. The availability of two reference images often helps alleviate this problem to some extent since areas not visible by one view are sometimes visible by the other view, but still special post-processing such as blending or hole-filling is typically needed to recover these undefined areas of the generated view. This problem will be illustrated and discussed further in the next section.

4. SIMULATION RESULTS

In this section, experimental results that consider rate overhead for depth information in the 3DV framework are discussed. Also, the effect of depth quality on the intermediate view generation results is analyzed. Additional results that also evaluate such effects may be found in [9].
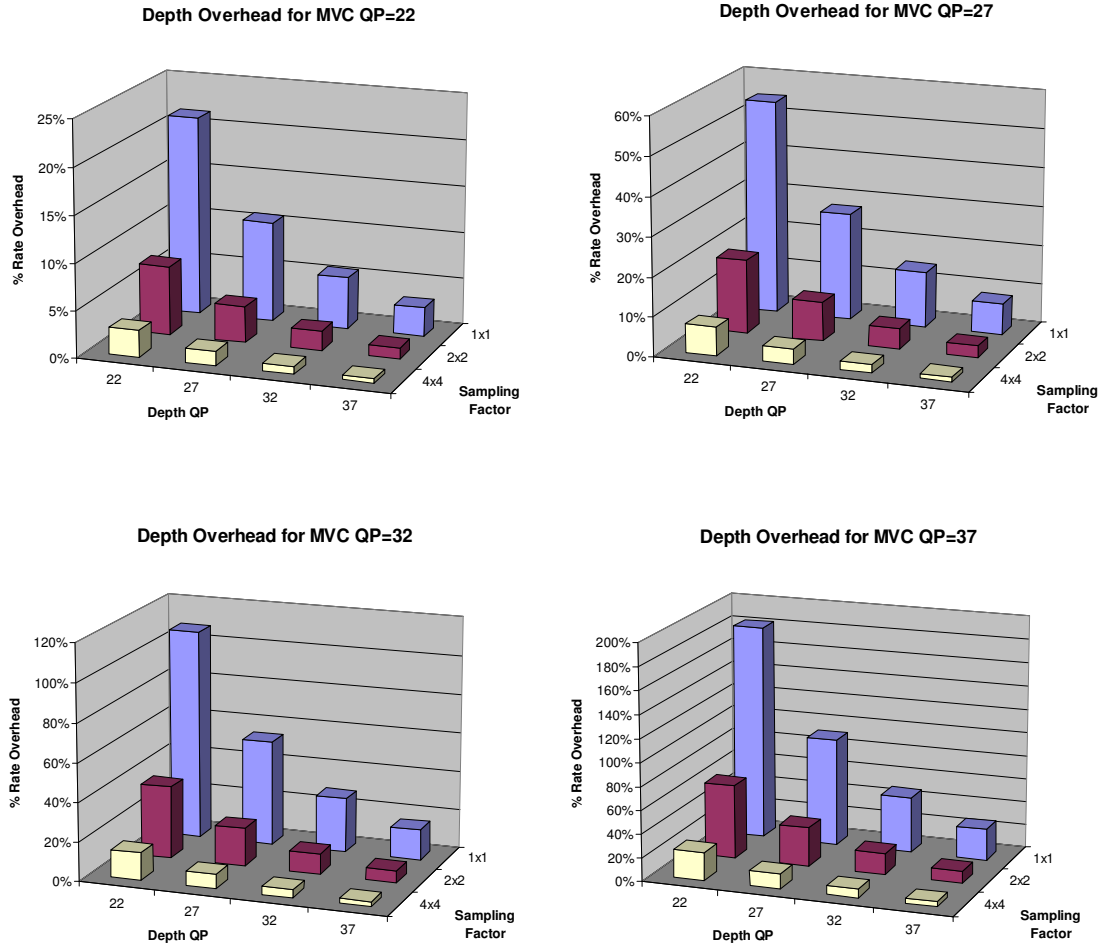
Fig. 6. Rate overhead for depth subject to various QP and sub-sampling factors, where percentage of depth overhead is calculated relative to the rate for coding a stereo pair of Breakdancer sequence with MVC at various QP values.

4.1 Depth overhead

As discussed in the previous subsection, depth information is useful for generating intermediate views from a reference set of views at the receiver. The results provided in this section illustrate potential rate overheads incurred by adding this additional information to the bitstream, and demonstrate the impact that quantization and sub-sampling have on the rate.

In our experiments, we consider two views of the Breakdancer test sequence (view 2 and view 4), where each view has a video sequence and corresponding depth video associated with it [10]. These two views are encoded with the MVC reference software [11] using a set of fixed QP values (22, 27, 32, 37). For the depth, in addition to the full resolution depth videos for each of the two input views, we also prepare lower resolution depth videos that have been sub-sampled by a factor of 2 and 4 in each dimension. The different resolution depth videos are referred to by the sub-sampling factors, i.e., 1x1 for full resolution, 2x2 for the quarter resolution, and 4x4 for the sixteenth resolution. We then encode the depth videos corresponding to views 2 and 4 with the same MVC software and with the same set of QP values as used for the video sequences.

The percentages of rate overhead for the various depth video encodings that might be associated with the stereo video inputs are plotted in Fig. 6. The percentages are calculated by dividing the rate for the depth videos by the rate for the video inputs. As one would expect, the depth overhead is relatively low when the video inputs are coded with higher quality (e.g., QP=22), but this overhead could be quite significant when the rate for the video inputs are reduced (e.g., QP=32 or QP=37). Clearly, encoding the depth with a coarser quantizer or sub-sampling the depth videos helps to reduce the overhead. Assuming conventional block-based coding techniques, an encoder would need to determine the optimal sub-sampling factor and quantization scale for depth. For example, given a fixed level of quality for the input videos and an allowable percentage overhead, an encoder would determine (perhaps adaptively) whether it is better to sub-sample the depth images or quantize them more coarsely. This issue will be revisited to some extent in the next subsection when we consider view generation.

It is noted that the results presented in this section only consider sub-sampling and quantization strategies to reduce the overhead incurred by the depth information. However, there might also be other means to reduce the rate for depth. For instance, it might be possible to filter the depth video to make it both easier to encode, while still retaining sample accuracy from depth estimation process. More efficient coding of depth information could also reduce the overhead, e.g., platelet-based coding scheme proposed in [12],[13]. The depth may also be used to reduce the rate of the input videos by using view synthesis prediction techniques on the input video as described in [14].

4.2 Intermediate view generation quality

The quality of intermediate view generation with various encodings of the depth is considered in the following. The same encoded data as used in the previous subsection are utilized here, and objective results as well as sample frames are provided. To perform view synthesis in our experiments, the software from Nagoya University has been used [15]. Recall from the previous subsection that only views 2 and 4 are coded, i.e., the depth for view 3 is not transmitted to the decoder. Based on information from the two input views, view 3 is synthesized as an intermediate view.

For reference, the synthesized view using uncoded texture and depth data (from views 2 and 4) is provided in Fig. 7(a). It is shown here that while some artifacts exist along some edges in the scene, the synthesis quality is generally acceptable; this quality could likely be improved with some minor improvements to the software. Note that the sample frames in this figure have been cropped to better show quality differences.

In Fig 7(b), the synthesized view using QP=27 for texture and QP=22 for depth without sub-sampling is shown. According to the results from the previous subsection, this combination of these parameters corresponds to a 54% increase in rate to accommodate the depth with this representation and encoding. Even with this relatively high overhead, an increase in the artifacts around edges in the scene is observed.

Two sample results using a reduced depth overhead of approximately 8% are shown in Figs. 7(c) and 7(d); both results are also generated using QP=27 for texture. The results in Fig 7(c) use depth encoded with QP=37 and no sub-sampling, while the results in Fig. 7(d) use depth with QP=22 and 4x4 sub-sampling. As shown in both of these figures, there is a notable increase in artifacts, and the video is generally unacceptable. The coarse quantization of depth results in severe view generation artifacts around edges of the scene; these appear either as broken contours or spurious patches around the object borders. These artifacts are also visible with the sub-sampled depth. Furthermore, there is additional blurring (or smearing) of edges when sub-sampling is applied. This is most likely attributed to the simple averaging operation that is used to down-sample the depth maps. This effect is reduced with 2x2 sub-sampling, and it could be expected that better filters would also lessen this blurring.

Table 1 provides a summary of average PSNR results across all depth QP values and sub-sampling factors tested. Under our current simulation conditions, the quality is generally higher when there is no sub-sampling applied; however, as mentioned above, the synthesis results could be increased with improved filtering. For the case when no sub-sampling is applied, there is a 1dB difference in average PSNR between the synthesized results using depth QP=22 versus depth QP=37. Generally speaking, these numerical measures do not correspond with the subjective results shown in Fig. 7, therefore PSNR should only be used as a very rough indicator of the quality.

Table 1. Comparison of average PSNR in dB for intermediate view generation results of view 3 of breakdancer sequence for various sub-sampling factors and QP for depth with QP=27 for texture.

Depth QP	Sub-sampling Factor		
	1x1	2x2	4x4
22	31.54	29.25	28.36
27	31.32	29.21	28.31
32	30.93	29.25	28.12
37	30.47	28.98	28.12

(a)

(b)

(c)

(d)

Fig. 7. Comparison of view interpolation quality with QP=27 for texture and various depth encodings. (a) synthesized frame with uncompressed texture and depth, (b) synthesized frame with QP=22 for depth and no sub-sampling, (c) synthesized frame with QP=37 for depth and no sub-sampling, (d) synthesized frame with QP=22 for depth and 4x4 sub-sampling.

5. CONCLUDING REMARKS

This paper discusses an emerging 3D video format that aims to support auto-stereoscopic displays that require a high number of views for rendering. The format is expected to include multiview video and depth, and aims to overcome limitations of existing data formats in this space. In particular, the inclusion of depth information enables a decoupling between the number of views that need to be transmitted and the number of views required by the display. Limiting the number of views is also more practical for production environments.

Simulation results are provided to highlight the rate overhead incurred by depth information, as well as the quality of intermediate views that are generated using various encodings of the depth. While the quality of intermediate views was acceptable in the absence of any compression, the quality degraded substantially when the depth was coded with a coarse quantization or when sub-sampling was utilized. These results underscore a clear need for improved representation and coding of depth information.

REFERENCES

1. J. Konrad and M. Halle, "3-D Displays and Signal Processing – An Answer to 3-D Ills?," *IEEE Signal Processing Magazine*, Vol. 24, No. 6, Nov. 2007.
2. P. Merkle, A. Smolic, K. Mueller, and T. Wiegand, "Efficient Prediction Structures for Multiview Video Coding," *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 17, No. 11, Nov. 2007
3. C. Fehn, P. Kauff, M. Op de Beeck, F. Ernst, W. Ijsselsteijn, M. Pollefeys, L. Vangool, E. Ofek, and I. Sexton, "An Evolutionary and Optimised Approach on 3D-TV", *Int'l. Broadcast Convention (IBC)*, Amsterdam, Netherlands, Sept. 2002.
4. P. Kauff, N. Atzpadin, C. Fehn, M. Müller, O. Schreer, A. Smolic, and R. Tanger, "Depth Map Creation and Image Based Rendering for Advanced 3DTV Services Providing Interoperability and Scalability," *Signal Processing: Image Communication. Special Issue on 3DTV*, Feb. 2007.
5. L. Stelmach and W. J. Tam, " Stereoscopic Image Coding: Effect of Disparate Image-Quality in Left- and Right-Eye Views," *Signal Processing: Image Communication* , Vol. 14, pp. 111-117, 1998
6. L. Stelmach, W.J. Tam; D. Meegan, and A. Vincent, "Stereo image quality: effects of mixed spatio-temporal resolution," *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 10, No. 2, pp. 188-193, March 2000.
7. D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, 47(1/2/3):7-42, April-June 2002.
8. Middlebury Stereo Vision Page, http://vision.middlebury.edu/stereo/
9. P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Multi-view Video Plus Depth Representation and Coding", *Proc. IEEE International Conference on Image Processing*, San Antonio, TX, USA, Sept. 2007.
10. Microsoft 3D video test sequences (available online: http://research.microsoft.com/ivm/3DVideoDownload/)
11. P. Pandit, A. Vetro, and Y. Chen, "WD 1 Reference software for MVC," JVT-AA212, Geneva, Switzerland, April 2008 (available online: http://ftp3.itu.ch/av-arch/jvt-site/2008_04_Geneva/JVT-AA212.zip)
12. Y. Morvan, D. Farin and P. H.N. de With, "Depth-image compression based on an R-D optimized quadtree decomposition for the transmission of multiview images," *Proc. IEEE International Conference on Image Processing*, San Antonio, TX, USA, Sept. 2007.
13. P. Merkle, Y. Morvan, A. Smolic, D. Farin, K. Müller, P.H.N. de With, and T. Wiegand, "The effect of depth compression on multiview rendering quality," *Proc. 3DTV-CON*, Istanbul, Turkey, May 2008.
14. S. Yea and A. Vetro, "View synthesis prediction for rate overhead reduction in FTV," *Proc. 3DTV-CON*, Istanbul, Turkey, May 2008.
15. M. Tanimoto, T. Fujii, K. Suzuki, N. Fukushima, Y. Mori, "Reference Softwares for Depth Estimation and View Synthesis," ISO/IEC JTC1/SC29/WG11 Doc. m15377, Archamps, France, April 2008.