OVERVIEW OF MULTIVIEW VIDEO CODING AND ANTI-ALIASING FOR 3D DISPLAYS

Anthony Vetro[†], *Sehoon Yea*[†], *Matthias Zwicker*[‡], *Wojciech Matusik*[†], *Hanspeter Pfister*[†]

[†] Mitsubishi Electric Research Labs
Cambridge, USA

[‡] Department of Computer Science and Engineering
University of California, San Diego, USA

ABSTRACT

This paper addresses signal processing issues related to coded representation, reconstruction and rendering of multiview video for 3D displays. We provide an overview of standardization efforts for multiview video that are aimed at reducing data rates required to represent the multiview video in compressed form. We then present an anti-aliasing filtering technique that effectively eliminates ghosting artifacts when rendering multiview video on 3D displays. Since high-frequency components of the signal are removed, substantial reductions in the compressed data rate could also be realized. Finally, we discuss the importance of scalability in the context of multiview video coding and suggest a combined anti-aliasing and scalable decoding scheme to minimize decoding resources for a given 3D display.

***Index Terms*—** Multiview video, 3D display, anti-aliasing

1. INTRODUCTION

Multiview video will be used to drive next-generation 3D displays. These next-generation displays enable viewing of high-resolution stereoscopic images from arbitrary positions and without the need for glasses. The multiview video provides the 3D display with view-dependent pixels that render a different color to the observer based on viewing angle. For 3D systems based on multiview video to become feasible, there are still some obstacles to overcome as outlined below.

Compression is certainly needed since the increase in the amount of data is quite substantial. To reduce the data transmission and storage requirements, new compression techniques that exploit not only the temporal correlation in the video, but also the inter-view correlation between the videos are being extensively studied in the context of H.264/AVC [1]. In this paper, we provide a brief update on ongoing standardization of multiview video coding (MVC), including a review of techniques that have been adopted by the standard as well as some promising techniques being explored. We show that substantial gains compared to independent coding of each view are achieved by the new MVC standard.

While data reduction is important, it is not likely to be the major impediment in 3D systems. The viewing experience is more significant. For instance, it has been found that directly rendering multiview video on a 3D display may result in disturbing artifacts that cause an unpleasant viewing experience. One source of this effect is aliasing. This has prompted researchers to investigate anti-aliasing filters for multiview 3D displays. In this paper, we review a particular method for anti-aliasing that is based on a resampling framework to match the input signal with the display characteristics. We present experimental results that compare compression performance of multiview videos under different filtering conditions and a range of bit-rates. Subjective results demonstrate the value of the proposed anti-aliasing filter in terms of 3D display quality. As an added benefit, substantial reductions in bit-rate could be achieved when the anti-aliasing filter is used as a pre-filter prior to coding since the high frequency content that contributes to aliasing is suppressed.

Pre-filtering of the multiview video is applicable for systems in which the 3D display is known and the signal bandwidth could be matched to the capabilities of the display prior to compression. This type of scenario might exist for gaming systems or cinema applications. However, applications such as consumer broadcast or video conferencing, a more diverse set of display devices with varying capabilities could be present. Therefore, it is not possible to pre-filter the multiview video signal. In this paper, we also consider the relation between scalable multiview coding and anti-aliasing prior to display. Specifically, we highlight the importance of view and spatial scalability to minimize decoding resources and provide the anti-aliasing filter with an input that more closely matches the display characteristics.

The rest of this paper is organized as follows. The next section presents an overview of multiview video coding with emphasis on standardization activities. Section 3 addresses anti-aliasing techniques and provides related experimental results. Section 4 introduces a scalable multiview coding scheme that aims to integrate scalable decoding and anti-alias filtering to minimize decoding resources. Concluding remarks are given in Section 5.

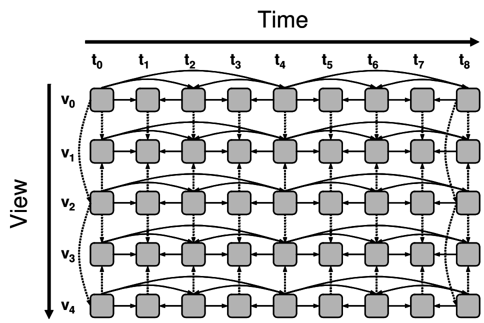

Fig. 1. Sample prediction structure for multiview video coding that exploits both temporal and spatial redundancy.

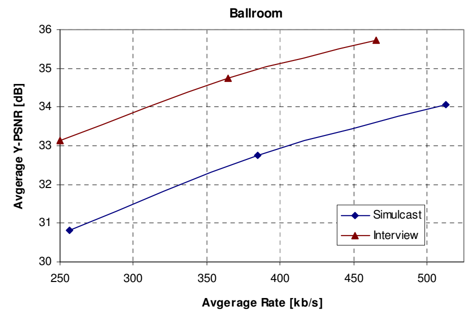

Fig. 2. Sample RD curves comparing performance of simulcast to multiview video coding that employs inter=view prediction.

2. MULTIVIEW VIDEO CODING STANDARD

A straight-forward solution for coding multiview video is to encode each view independently using a state-of-the-art video codec such as H.264/AVC [1]. The advantage of this approach is that it could be achieved with current standards and existing hardware. However, it does not exploit the redundancy across views, thereby not realizing the most efficient representation in compressed form.

The basic idea employed in all related works on efficient multiview video coding is to exploit both spatial and temporal redundancy for compression. Since all cameras capture the same scene from different viewpoints, inter-view redundancy is present. A sample prediction structure is shown in Fig. 1. Pictures are not only predicted from temporal neighbors, but also from spatial neighbors in adjacent views.

It has been shown that coding multiview video in this way does give significantly better results compared to the simple H.264/AVC simulcast solution [2]. 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. Fig. 2 shows sample RD curves comparing the performance of simulcast coding (without the use of hierarchical B-pictures) with the performance of the JMVM 2.0 software [3] that employs hierarchical predictions in both spatial and view dimensions. Of course, the use of hierarchical B-pictures in the simulcast solution will also provide some gains, but they are not shown in this plot.

There are many variation on the prediction structure considering both temporal and spatial dependencies. The structure not only affects coding performance, but has notable impact on delay, memory requirements and random access. It has been confirmed that the majority of gains are obtained using inter-view prediction at anchor positions, i.e., set of pictures that have no temporal prediction, only spatial prediction such as t_0 and t_8 in Fig. 1. Rate penalties of approximately 5-15% could be expected if the spatial predictions at non-anchor positions are removed [4]. The upside is that delay and required memory would also be reduced.

Enabling spatio-temporal prediction in the H.264/AVC standard is realized with no change to the macroblock level coding tools used for coding single-view video. In fact, the only major changes are the means by which reference pictures are managed and the high-level syntax carried in the bitstream [5]. With regards to managing the multiview reference pictures, it is important to define processes that construct appropriate reference picture lists for each picture to be decoded and that reference pictures be discarded when no longer needed for prediction. Additions to the high-level syntax of the bitstream are also necessary to specify coding dependencies among the different views.

Beyond these extensions, there are a variety of other macroblock level coding tools that are also being explored under core experiments within the standards committee. One such tool is illumination compensation [6], which has shown to be very useful when the illumination or color characteristics vary in different views. The proposed method employs predictive coding for the DC component of inter prediction residues, where the predictor for illumination change is formed from neighboring blocks since illumination differences tend to be spatially correlated. Coding gains up to 0.6 dB have been reported in comparison to the existing weighted prediction tool of H.264/AVC.

Another novel macroblock level coding tool that is currently being explored is view synthesis prediction, where a picture in a given view is predicted from synthesized references generated from neighboring views. One approach is to encode depth for each block, which is then used at the decoder to generate the view synthesis data used for prediction [7]. A second approach estimates pixel-level disparities at both the encoder and decoder and encodes only disparity correction values [8].

Finally, extensions to the skip and direct coding modes that infer side information from spatial references rather than temporal references are also under consideration. Such coding modes are expected to provide gains at lower bit-rates where a block could be reconstructed from spatial references with a minimal amount of overhead.

3. ANTI-ALIASING FOR 3D DISPLAYS

This section first reviews fundamental principles on rendering multiview image data on 3D displays and discusses related work on anti-aliasing. We then present experimental results that validate the subjective improvements achieved with anti-aliasing and the bandwidth reduction that is possible.

3.1. Resampling Framework

Multiview image data can be interpreted as sampled higher dimensional signals. The resulting sampling grids are similar to the grids of multiview displays, but in general there is no one-to-one correspondence between camera rays and display rays. To render a sampled light field on an automultiscopic display, the samples of the input light field need to be mapped to the samples, or pixels, of the display. This involves a resampling operation, which consists of three conceptual steps: continuous reconstruction of the input signal, band-limiting the continuous signal to the bandwidth of the display, and sampling of the band-limited signal on the display pixel grid.

In practice, the continuous reconstruction step is usually approximated by upsampling the input data to a sufficiently high resolution. State-of-the-art view interpolation techniques could be used to solve this problem. In the absence of any filtering, aliasing may occur in the sampling step along several dimensions. First, aliasing may appear within each view due to the discrete 2D pixel grid of each view. Second, inter-perspective aliasing occurs as visually disturbing ghosting problems due to the discrete number of views.

Konrad et al. [9] address aliasing due to the discrete 2D pixel grid of each view. Because these grids are usually not rectangular, they derive custom filters using an optimization process to provide optimal image quality. However, their analysis does not take into account inter-perspective aliasing. Moller et al. [10] describe a method to prevent inter-perspective aliasing that is based on a display bandwidth analysis. Unfortunately, this approach requires the knowledge of per pixel scene depth and leads to a spatially varying 2D filter.

Zwicker et al. [11] describe a unifying approach that derives a low-pass filter directly from the ray-space sampling grid of the multiview 3D display. This approach prevents aliasing within each view as well as inter-perspective aliasing. In addition, it does not require the knowledge of scene depth and it is implemented as a linear convolution rather than relying on spatially varying filters. In its present form, this scheme is realized by first oversampling the multiview video signal so that it is free of aliasing within the display bandwidth. This means that more views are generated through view interpolation at a smaller spacing than the display actually provides. After pre-filtering, the multiview signal is then subsampled at the original display resolution. The performance of this filtering technique is examined more closely in the next section.

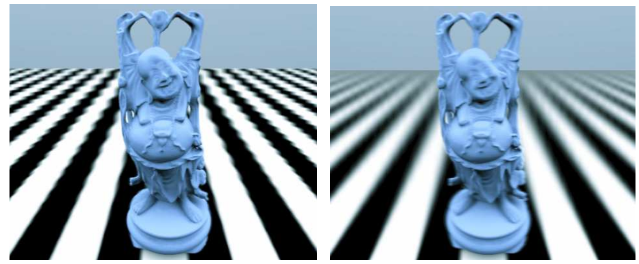

Fig. 3. Subjective comparison of a simulated views [11]: without pre-filtering (left), with pre-filtering (right).

3.2. Evaluation

We first show the subjective improvements of the anti-aliasing prefilter [11]. The synthetic Buddha sequence is used for this purpose. To simulate the display effects, sample frames of the multiview video are used to simulate the perspective views of an automultiscopic display. The results with an without display prefiltering are shown in Fig. 3. It is evident that without prefiltering, the displayed image will incur aliasing problems that appear as staircase effects in the sample frame shown. In natural video sequences that we have tested, the aliasing problems appear as ghosting artifacts on edges in the scene. Applying the display pre-filter avoids these problems. Although the anti-aliased images are blurred compared to the original, subjective viewing on a 3D display reveals a more comfortable viewing experience.

Since high frequency of the input signal is suppressed to avoid anti-aliasing, the multiview signal becomes easier to compress. To demonstrate the reduction in data rate that is possible, we plot the RD curves comparing two natural video sequences with and without display pre-filtering in Fig. 4. These plots show that the rate could be reduced by approximately half in the medium to higher rate ranges. It is important to note that this should not be viewed as a gain in coding efficiency since the references used for each curve are indeed different. The purpose of these plots are just to demonstrate the degree of rate savings that are achieved when the multiview signal has been pre-filtered with the primary purpose of removing anti-aliasing artifacts.

4. SCALABLE DECODING WITH ANTI-ALIASING

One disadvantage of the resampling and pre-filtering scheme presented in the previous section is that it requires knowledge of the target display prior to compression. Ideally, the coding scheme should be designed to accommodate various decoding and display capabilities, therefore scalability in the context of multiview video coding should also be considered. This section discusses the benefits of a scalable multiview coding scheme to facilitate efficient decoding with anti-aliasing prior to display. Two dimensions of scalability are considered: the

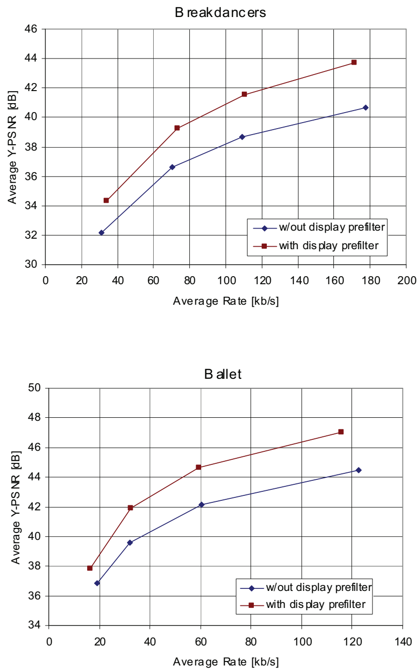

Fig. 4. Comparison of RD curves for breakdancer and ballet sequences when prefiltering is applied with and without display prefiltering.

number of views and the spatial resolution of each view.

View scalability is achieved by coding the multiview video with hierarchical dependency in the view dimension. Consider the prediction structure with 5 views shown in Fig. 1. If the display only requires that 3 of these views be decoded, then there are two choices. We may either discard the two views with bi-directional dependency, i.e., v_1 and v_3, or the views that come later in decoding order, i.e., v_3 and v_4. The first option would increase the relative disparity between views and hence the amount of upsampling that is required. Therefore, to minimize the upsampling rate, the second option is a better choice. This sampling of the view dimension generally implies selection of views that are spaced closer together, which should be factored into the design of prediction structures.

The spatial resolution of each view affects the spectrum of the input signal. In [11], it was stated that the minimum sampling rate (view and spatial resolution) could be derived by finding the tightest packing of replicas of the input spectrum such that none of the non-central replicas overlap with the display prefilter. If the number of views to be decoded are determined based on the criteria discussed above and acquisition parameters such as camera aperture are fixed, then the only remaining degree of freedom is the spatial resolution. Therefore, the receiver does not need to support the decoding of spatial resolution layers beyond its display capability and support for spatial scalability could alleviate the need to fully decode a high resolution video and sample it to the appropriate display resolution.

5. CONCLUDING REMARKS

We provided an overview of current standardization activities on multiview video coding. We also addressed the topic of anti-aliasing for 3D display of multiview video including its impact on compression in the context of a pre-filtering scheme. Finally, the potential benefits of a scalable multiview coding design has been discussed to minimize resources required for decoding and anti-aliasing prior to display.

6. REFERENCES

[1] ITU-T Rec. & ISO/IEC 14496-10 AVC, "Advanced video coding for generic audiovisual services," 2005.

[2] P. Merkle, K. Müller, A. Smolic, and T. Wiegand, "Efficient compression of multiview video exploiting inter-view dependencies based on H.264/AVC," in *Proc. IEEE Int'l Conf. Multimedia & Expo*, 2006.

[3] P. Pandit and A. Vetro, "JMVM 2.0 software," in *Doc. JVT-U208*, Hangzhou, China, 2006.

[4] M. Droese and C. Clemens, "Results of CE1-D on multiview video coding," in *ISO/IEC JTC1/SC29/WG11 Doc. m13247*, Montreux, Switzerland, 2006.

[5] E. Martinian, A. Behrens, J. Xin, A. Vetro, and H. Sun, "Extensions of H.264/AVC for multiview video compression," in *Proc. IEEE Int'l Conf. Image Proc.*, 2006.

[6] Y.L. Lee et al., "CE11: Illumination compensation," in *Doc. JVT-U052*, Hangzhou, China, 2006.

[7] E. Martinian, A. Behrens, J. Xin, and A. Vetro, "View synthesis for multiview video compression," in *Proc. Picture Coding Symposium*, 2006.

[8] M. Kitahara et al., "Multi-view video coding using view interpolation and reference picture selection," in *Proc. IEEE Int'l Conf. Multimedia & Expo*, 2006, pp. 97–100.

[9] J. Konrad and P. Agniel, "Subsampling models and anti-alias filters for 3-D automultiscopic displays," *IEEE Trans. Image Proc.*, January 2006.

[10] C. N. Moller and A.R.L. Travis, "Correcting interperspective aliasing in autostereoscopic displays," *IEEE Transactions on Visualization and Computer Graphics*, vol. 11, no. 2, pp. 228–236, March/April 2005.

[11] M. Zwicker, W. Matusik, F. Durand, and H. Pfister, "Antialiasing for automultiscopic 3D displays," in *Eurographics Symposium on Rendering*, 2006.