Display Pre-filtering for Multi-view Video Compression

Matthias Zwicker
University of California
San Diego

Sehoon Yea
Mitsubishi Electric
Research Laboratories

Anthony Vetro
Mitsubishi Electric
Research Laboratories

Clifton Forlines
Mitsubishi Electric
Research Laboratories

Wojciech Matusik
Mitsubishi Electric
Research Laboratories

Hanspeter Pfister
Mitsubishi Electric
Research Laboratories

ABSTRACT

Multi-view 3D displays are preferable to other stereoscopic display technologies because they provide autostereoscopic viewing from any viewpoint without special glasses. However, they require a large number of pixels to achieve high image quality. Therefore, data compression is a major issue for this approach. In this paper, we present a framework for efficient compression of multi-view video streams for multi-view 3D displays. Our goal is to optimize image quality without increasing the required data bandwidth. We achieve this by taking into account a precise notion of the multi-dimensional display bandwidth. The display bandwidth implies that scene elements that appear at a given distance from the display become increasingly blurry as the distance grows. Our main contribution is to enhance conventional multi-view compression pipelines with an additional pre-filtering step that bandlimits the multi-view signal to the display bandwidth. This imposes a shallow depth of field on the input images, thereby removing high frequency content. We show that this pre-filtering step leads to increased image quality compared to state-of-the-art multi-view coding at equal bitrate. We present results of an extensive user study that corroborate the benefits of our approach. Our work suggests that display pre-filtering will be a fundamental component in signal processing for 3D displays, and that any multi-view compression scheme will benefit from our pre-filtering technique.

Categories and Subject Descriptors

I.4.2 [**Computing Methodologies**]: Image Processing and Computer Vision—*Compression (Coding)*

General Terms

Human Factors

Keywords

3D Displays, Multi-View Compression, Antialiasing

1. INTRODUCTION

Multi-view 3D displays offer viewing of high-resolution stereoscopic images from arbitrary positions without glasses. These displays consist of view-dependent pixels that reveal a different color to the observer based on the viewing angle. View-dependent pixels can be implemented using conventional high-resolution displays and parallax-barriers, lenticular sheets, or holographic screens. Although the basic optical principles of multi-view auto-stereoscopy have been known for over a century [17], only recently displays with increased resolution, or systems based on multiple projectors, have made this approach practical. Today, commercial availability ranges from multi-view desktop monitors [16] to large-scale displays based on multi-projector systems [1, 7].

Multi-view 3D displays feature several advantages over competing autostereoscopic display technologies, such as stereoprojection systems using shuttered or polarized glasses. Most importantly, automultiscopic displays do not require users to wear any special glasses, which leads to a more natural and unrestricted viewing experience. They also do not require head tracking to provide motion parallax; instead, they provide accurate perspective views from arbitrary points inside a viewing frustum *simultaneously.* They are truly multi-user capable, since none of the display parameters needs to be adjusted to a individual user. For these reasons, we believe that multi-view 3D displays will become the device of choice for a large number of applications such as scientific visualization or remote collaboration. They have the potential to replace conventional 2D displays in the mass markets of digital entertainment [13].

However, the amount of data that needs to be processed, rendered, and transmitted to such displays is an order of magnitude larger than for systems based on stereo-image pairs. Therefore, data compression is of paramount importance for such systems. In this paper, we address this problem by introducing a framework for efficient compression of multi-view video streams that complements current techniques. Our approach reduces the required data rate to a minimum by taking into account the multi-dimensional display bandwidth.

The limited bandwidth of multi-view 3D displays corresponds to a *shallow depth of field.* This means that only those scene elements that are within a certain distance from the display plane can be shown sharply. Scene elements that appear at larger distances become increasingly blurry. However, for current displays the depth of field is only a few centimeters [20]. Therefore, many interesting scenes exceed the display bandwidth. Rendering these scenes leads to *inter-perspective aliasing* [15].

As our main contribution, we propose to improve multi-view compression by adding a pre-filtering step that bandlimits the input signal to the display bandwidth. Pre-filtering has two desirable effects: First, it removes high frequencies that would appear as inter-perspective aliasing, and second, it reduces the signal bandwidth. Although this approach is conceptually straightforward, it has not been pursued before.

We evaluate our approach using an extensive user study that corroborates the benefits of the pre-filtering step. We show that, at equal signal bandwidth, our approach leads to higher perceived image quality compared to state-of-the-art multi-view coding without pre-filtering. Our work suggests that any compression scheme for multi-view 3D displays will benefit from our pre-filtering technique. Therefore, we believe it will be an integral part of any multi-view compression pipeline.

2. PREVIOUS WORK

We distinguish three approaches to characterize display bandwidth. The first one, proposed by St. Hilaire [6], builds on wave optics. The main advantage of this approach is that it includes the effects of the diffraction limit. However, St. Hilaire's analysis shows that the diffraction limit does not play a significant role for typical multiview displays with not more than a few dozen views. A second approach, as described by Halle [5], is based on simple geometric considerations. Both the wave optics and the geometric technique share the disadvantage that they lead to a depth dependent formulation of display bandwidth. This means that they specify the display bandwidth as the bandwidth of two-dimensional images that appear at a given distance from the display plane.

The third approach [20] is based on a ray space representation of multiview 3D displays. Essentially, it casts the analysis of display bandwidth as a multidimensional sampling problem in three- or four-dimensional ray space. This approach is related to the concept of light fields [12], which has been studied extensively in the computer graphics community. The frequency analysis of light fields, also known as *plenoptic sampling theory*, has been studied by Chai et al. [2] and Isaksen et al. [8]. An analysis of the display bandwidth using plenoptic sampling theory [2] reveals important properties, such as the shallow depth of field of practical displays.

Moller et al. [15] describe a method to prevent interperspective aliasing that is based on St Hilaire's [6] display bandwidth analysis. Unfortunately, this approach requires the knowledge of per pixel scene depth. In addition, it leads to a spatially varying 2D filter. Zwicker et al. [20] derive 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. It does not require the knowledge of scene depth and it is implemented as a linear convolution rather than relying on spatially varying filtering. Therefore, we base our pre-filtering technique on this approach.

Konrad et al. [11] 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 interperspective aliasing.

Multiview 3D displays require, at least, an order of magnitude more samples than conventional 2D displays to achieve comparable image quality because of the higher dimensionality of the input signals. Therefore, data compression plays a crucial role in making these displays practical. Compression of multi-view video data is a highly active area of research, and standardization efforts for multi-view video compression are well under way in the MPEG-4 community. Various extensions of the H.264/MPEG-4 AVC video compression standard to the multi-view setting have been proposed recently [14].

However, to the best of our knowledge, none of the previous multi-view video or light field compression techniques take the three- or four-dimensional bandwidth of multi-view 3D displays into account. This means that parts of the frequency content of the encoded signal will appear as interperspective aliasing when rendered on a 3D display. This can reduce image quality and lead to inefficient compression. Our multi-view compression scheme includes a low-pass filtering stage to ensure that the encoded signal does not exceed the bandwidth of a target 3D display. This approach has two advantages over previous techniques. First, it avoids interperspective aliasing artifacts, and second, our approach increases compression efficiency.

3. PRELIMINARIES

In this section we briefly review the concept of display bandwidth as introduced by Zwicker et al. [20], which is based on plenoptic sampling theory [2]. The main idea of our approach is to ensure that we do not include any frequency content beyond the display bandwidth in the compression pipeline.

Multi-view 3D displays seek to reproduce the full *light field* [12, 4] of an input scene. Similar to light fields, we parameterize the light rays emitted by the display by their intersection with two parallel planes. The intersection coordinates of each ray correspond to a point in *ray space*, and the set of all rays forms a higher-dimensional grid in ray space.

We illustrate the parameterization and the sampling grid of a typical display in Figure 1. For simplicity, we choose a display that provides only horizontal view-dependency, which is a common restriction of practical displays [16, 7]. This allows us to focus on an individual horizontal scanline and reduce the visualization to two dimensions, denoted by v and t. We depict the geometry of the display rays at the top, and the corresponding sampling grid in ray space at the bottom of Figure 1. We place the t coordinate axis at the display plane, such that all rays belonging to one view-dependent pixel have the same t coordinate. The v axis corresponds to a virtual camera plane, where the rays of individual display views converge. Note that the sampling grid can also be interpreted as an epipolar plane image (EPI). The parameterization in Figure 1 matches the construction of the displays manufactured by Newsight [16]; however, for other devices different choices for the parameter planes may be more appropriate.

In Figure 2, we show an example of a multi-view signal sampled on a display grid corresponding to Figure 1. Note that the foreground character has the smallest disparity, i.e., each point is projected to almost the same pixel in all views.

Therefore, it will appear to lie on the display plane. The background, however, is shifted in each view and will appear at a distance behind the display plane.

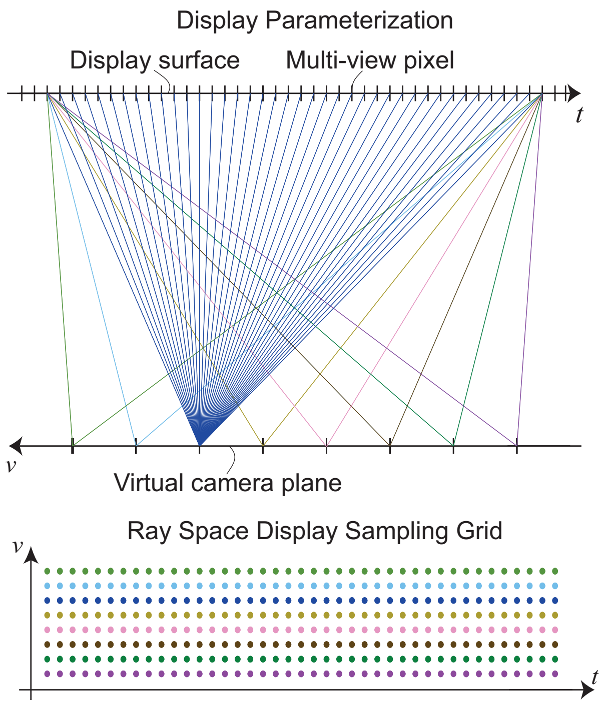

Figure 1: We illustrate the parameterization of a multi-view display at the top, and the corresponding ray-space sampling grid at the bottom.

The ray-space sampling grid imposes a strict limit on the signal bandwidth that can be represented by the display, known as the Nyquist limit. Using basic ideas from plenoptic sampling theory [2] it is straightforward to show that a scene does not exceed the Nyquist limit if the maximum disparity in its EPI, parameterized as in Figure 1, is less than one pixel. On the other hand, band-limiting the light-field of an arbitrary scene corresponds to limiting its depth of field [12]. For current displays the depth of field spans only a few centimeters [20], which means that many interesting scenes exceed the display bandwidth. If band-limitation is omitted, disturbing interperspective aliasing artifacts [15], such as ghosting artifacts, will appear on the display. Band-limiting a light field to the display-bandwidth can be implemented using simple linear filtering in ray space [20]. We apply this approach in our pre-filtering step described in Section 4.1.

4. COMPRESSION PIPELINE

In this section we describe our compression pipeline for multi-view 3D displays, which consists of two main steps shown in Figure 3. In the first step, described in Section 4.1, we perform a display pre-filtering operation. This step removes frequency content from the input signal that is beyond the Nyquist limit of the display. Because these frequencies would appear as aliasing on the multi-view display, the pre-filtering step does not reduce image quality. However, it increases the compression efficiency by zeroing out parts of the spectrum of the input signal. In the second step of our pipeline, we run the pre-filtered signal through a state-of-the-art multi-view compression algorithm, which we summarize in Section 4.2.

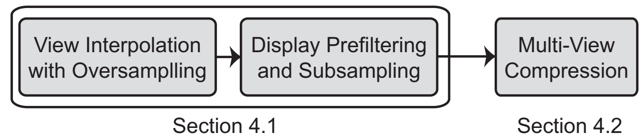

Figure 3: Compression pipeline.

4.1 Display Pre-filtering

The objective of this step is to obtain input data for the subsequent compression stage. We ensure two criteria: First, the data needs to be sampled on the display grid as shown in Figure 1. This can be achieved using view interpolation. Second, it needs to obey the Nyquist limit of the display. This is achieved by appropriate pre-filtering.

View Interpolation. Multi-view data acquired by camera arrays does in general not correspond to the sampling grid as shown in Figure 1. Therefore, it needs to be reparameterized or even resampled first, depending on the configuration of the camera array. There is a vast number of view interpolation techniques [19, 3] that can be used to achieve this, but a discussion of these is beyond the scope of this paper. We obtained test data using different approaches. We used the data from Zitnick et al. [19], which is available publicly. This data includes depth maps, such that view interpolation can be achieved using straightforward reprojection. We also rendered animated multi-view sequences from sytnhetic 3D geometry. Finally, we composited video sequences with opacity maps to generate different depth layers.

Pre-filtering. With the multi-view data available in the display parameterization as shown in Figure 1, our goal is now to band-limit this signal to the bandwidth of the display sampling grid. This can be achieved by convolving the EPI of each horizontal scanline with the display prefilter. However, the multi-view input signal, sampled at the display resolution, will contain aliasing for most scenes as discussed by Chai et al. [2]. Unfortunately, pre-filtering aliased signals leads to ghosting artifacts. Therefore, we first oversample the signal to remove aliasing from the input, then band-limit and subsample to the display resolution.

Aliasing in the input signal is illustrated in the frequency domain in Figure 4 on the left. Here, the frequencies ϕ and θ correspond to spatial coordinates v and t. The spectra exhibit the bow-tie shapes typical for light-field data [2]. The vertical spacing of the replicas of the continuous spectrum is inversely proportional to the sampling density along the v axis. Practical displays, however, provide only a small number of samples along the v-axis, typically not more than a few dozen. Therefore, non-central replicas overlap with the display bandwidth (Figure 4a). Applying the display prefilter directly to this data will lead to aliasing artifacts (Figure 4b).

To avoid this situation, we oversample the signal along the v-direction such that it is free of aliasing within the display bandwidth. This means we interpolate more views at a smaller spacing in v than the display actually provides. We prevent aliasing if none of the bow tie spectra except

Figure 2: Example of a multi-view signal sampled on a display grid corresponding to Figure 1. Each image represents a slice with $v = const$, the color coding matches Figure 1.

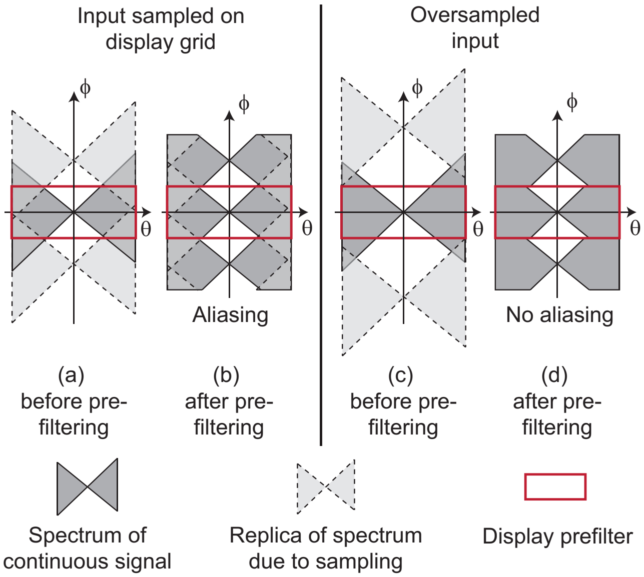

Figure 4: Display pre-filtering without oversampling leads to aliasing artifacts, as shown on the left. Oversampling the input avoids these problems, as shown on the right. Note that the display prefilter is a unit square. The visualization is stretched horizontally to emphasize the difference in resolution between spatial sampling (the resolution of the multiview display) and angular sampling (the number of views).

the central overlaps with the display prefilter. Let us assume that the multi-view signal, sampled at the display resolution, has a maximum disparity of d pixels. Note that the slopes of the bow-ties correspond to the maximum disparity d in the EPIs [2]. Therefore, their vertical spacing needs to be at least $(d+1)/2$ to remove overlap with the prefilter. This implies an oversampling factor of $(d+1)/2$: For a display with k views, we need to interpolate at least $k \times (d+1)/2$ views. Prefiltering with oversampling is shown in Figure 4c on the right. The output signal is free of aliasing problems as illustrated in Figure 4d.

We band-limit the oversampled signal by convolving it with the display pre-filter. We implement this step as a convolution with a Gaussian filter in the spatial domain. Of course, other filter kernels could be used alternatively. After pre-filtering we subsample at the original display resolution.

We compare results of pre-filtering with and without oversampling in Figure 5 using the data set shown in Figure 2. If we pre-filter these eight views directly, aliasing problems appear as disturbing ghosting artifacts as shown in Figure 5 at the top. Oversampling before prefiltering avoids these problems as shown in Figure 5 at the bottom. The maximum disparity in this data set is 7 pixels, therefore we interpolated $8 \times (7+1)/2 = 32$ views before pre-filtering.

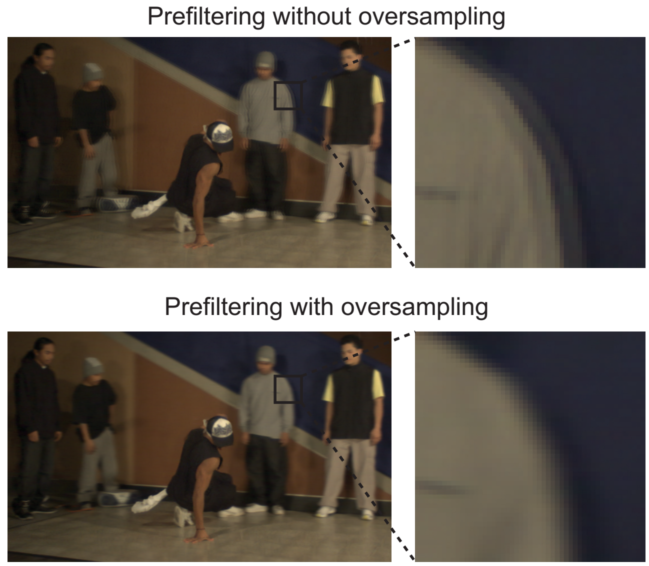

Figure 5: Comparison of pre-filtering with and without oversampling. Pre-filtering without oversampling leads to aliasing problems appearing as ghosting artifacts, shown at the top. Oversampling avoids these problems as shown at the bottom.

Results. We show results of our display pre-filtering procedure in Figure 6. At the top, we illustrate the effect of pre-filtering on input views for the display. These images are zoomed-in areas of the blue-framed view in Figure 2. Without pre-filtering, the image shows sharp edges even in the background area cropped out here. Pre-filtering, on the right, blurs the background, i.e., it reduces the depth of field of the image. At the bottom we show simulated views of the multiview display as a single eye would see it. Without pre-filtering, aliasing artifacts become apparent as ghosting. On the other hand, the simulated view with pre-filtering does not show any artifacts. Our compression scheme banks on the fact that the pre-filtered input views as shown at the top right are easier to compress than without pre-filtering, while the simulated views with and without pre-filtering are of very similar quality.

4.2 Multi-view Compression

One solution for compressing multiview videos is to encode each view independently using a state-of-the-art video codec such as H.264/AVC [10]. The main advantage of this approach is that current standards and existing hardware could be used. To achieve further gains in coding efficiency, extensions to the H.264/AVC standard are now being developed to exploit not only the redundancy in pictures over time, but also the redundancy between pictures in different camera views.

Performing efficient compression relies on having good predictors. While the correlation between temporally neigh-

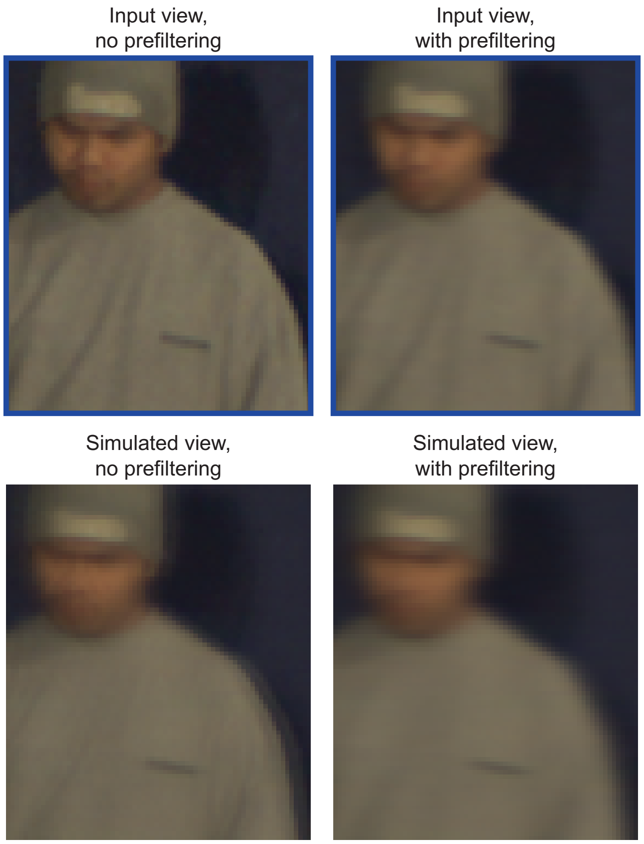

Figure 6: Comparison of simulated display views with and without pre-filtering. At the top, we show an input view with and without prefiltering. At the bottom, we show simulated views of the display as seen from a single eye. Ghosting artifacts appear if we omit pre-filtering, as shown on the left. The simulated view with pre-filtering, on the right, has a shallow depth of field, but t does not show any artifacts.

boring pictures is often very strong, including spatially neighboring pictures offers some advantages. For example, spatially neighboring pictures are useful predictors in uncovered regions of the scene, during fast object motion, or when objects appear in one view that are already present in neighborings views at the same time instant.

It has been shown that coding multiview video with inter-view prediction does give significantly better results compared to independent coding of each view [14]. Improvements of more than 2 dB have been reported for the same bit-rate, 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 more comprehensive review of recent developments in multi-view coding can be found in [9]. Figure 7 shows a sample rate-distortion comparison of the quality obtained by simulcast coding, i.e., independent coding of each view, with the quality of the latest inter-view coding scheme that employs prediction from spatial neighbors. Due to this improved coding efficiency, all compression experiments that follow utilize inter-view prediction using an algorithm based on Merkle et al.'s approach [14].

Since our pre-filtering approach suppresses high frequency

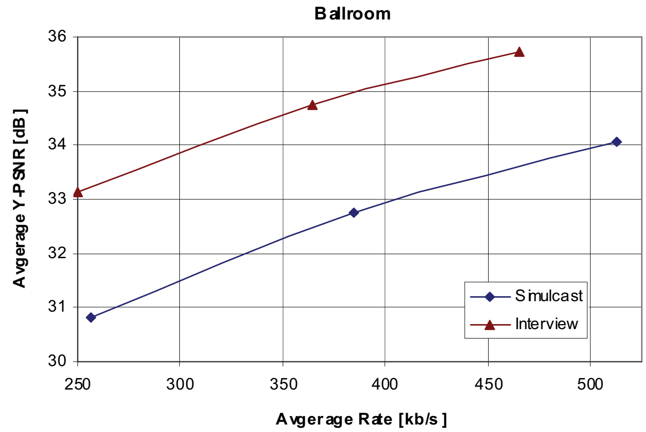

Figure 7: Sample RD curves comparing performance of simulcast to multiview video coding that employs inter-view prediction.

of the input signal to avoid anti-aliasing, the multiview signal becomes even easier to compress. To demonstrate the reduction in data rate that is possible, we plot the rate-distortion curves comparing the quality of the compression of multi-view videos with and without pre-filtering at different bit-rates in Figure 8. We performed the measurements using the *ballroom* data set from the MPEG-4 standardization community [18]. 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.

We compare the result of compression of pre-filtered views and original views in Figure 9. The images are from the *Waterfall* test sequence, which was also included in our user study (Section 5). We show results of compression without pre-filtering at the top, and with pre-filtering at the bottom. We reduced the bitrate of both sequences to 110 kbps per second. The images in Figure 9 are simulated display views similar to Figure 6. The foreground character shows stronger blocking artifacts in the version without pre-filtering, at the top, than with pre-filtering, at the bottom. In addition, pre-filtering removes ghosting artifacts, which appear without pre-filtering in the background in the top image.

5. USER STUDY

We conducted a preferential study designed to shed light on the effect of our pre-filtering approach on user preference when viewing compressed 3D videos. Twelve subjects participated in our study, six males and six females between the ages of 23 and 45 years old. These individuals were recruited from an on-line community bulletin board and from the administrative and speech departments at our organization. Participants from outside our organization were paid $10 compensation for their time.

5.1 Method and Procedure

Subjects were first shown an example video that shipped

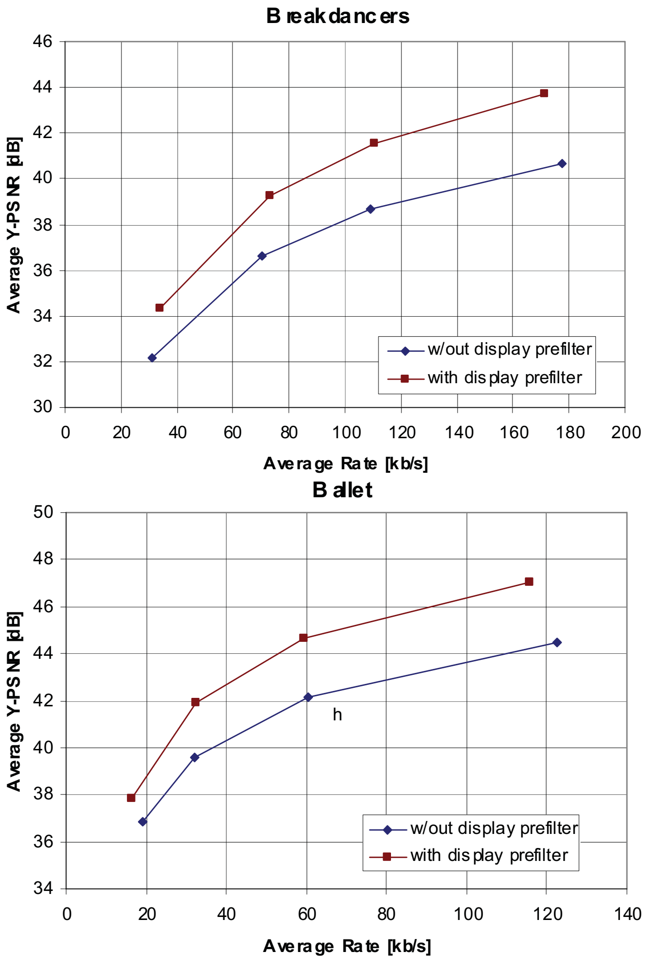

Figure 8: Comparison of RD curves for breakdancer and ballet sequences with and without pre-filtering.

with our display in order to demonstrate the capabilities of the device. We used a 23" display by Newsight [16] that provides eight views with a resolution of 640 × 384 pixels each. For most of our participants, this was their first experience viewing a 3D display. All of our participants were able to perceive depth in the image, and all of them had normal or corrected normal vision.

The experiment proceeded in the following manner. Participants were shown a series of video pairs with a short segment of blank grey video inserted between them. Each video in the pair contained the same content, compressed with and without pre-filtering as described in Section 4. Participants were allowed to view the pair of videos as many times as they wanted to in order to answer the question, "Which video do you prefer overall?" Five different video clips were used, ranging in length between six and ten seconds. These included a variety of different content - a video of a ballerina, a video of several break-dancers, a synthetic scene of a model dragon, a man standing in front of a waterfall, and a man standing in front of a pedestrian walkway. All video sequences have eight views with a resolution of 640 × 384 pixels. Each video pair was compressed at three different bitrates for a total of 15 pairs. We manually adjusted the quality parameter of the compression algorithm to achieve similar bitrates with and without pre-filtering. We report

Figure 9: Comparison of compressed frames of a video sequence with and without pre-filtering. The images show simulated views of a multi-view 3D display. The version without pre-filtering at the top shows stronger blocking artifacts than the version with pre-filtering at the bottom. In addition, pre-filtering avoids ghosting artifacts, which appear without pre-filtering in the background in the top image.

the bitrates for the test sequences in Table 1. Although they do not match exactly, we verified empirically that the remaining differences are too small to influence the perceptual study.

Scene	Low	Medium	High
Walkway	52.8/51.2	82.6/83.2	138.4/136.4
Breakdancers	37.8/40.6	46.3/48.9	70.2/73.0
Waterfall	58.0/61.2	91.3/95.1	132.1/128.7
Dragon	59.4/54.4	121.9/124.9	179.4/181.8
Ballet	31.9/32.4	60.3/59.4	122.7/115.7

Table 1: Low, medium, and high bitrates in kbps for the five test sequences with/without pre-filtering. We manually adjusted the parameters of the compression algorithm to obtain similar bitrates with and without pre-filtering.

An experimenter recorded the participant's preference after each pair, as well as their comments during the trials and at the end of the experiment when they were asked to explain what characteristics of the videos were helping them in their decisions. The order of presentation of the two compression techniques was random within each trial, and the order of presentation of each of the bitrates and clips was balanced among participants using a Latin square. In summary, our design was:

- 2 compression techniques (with/without pre-filtering),
- 5 video clips (ballerina, break-dancers, dragon, waterfall, and walkway),
- 3 bitrates (low, medium, and high),
- 12 participants,
- resulting in 180 trials in total.

We had two experimental hypotheses:

- As a group, participants would prefer video clips compressed using the pre-filtering technique over clips compressed without pre-filtering.
- The preference for videos rendered using the pre-filtering technique over those without pre-filtering would be inversely correlated with the bitrate of the encoded video.

5.2 Results and Discussion

As predicted by hypothesis one, our participants preferred the pre-filtering technique in a majority of the experimental trials (60.7% vs. 39.3% of trials for anti-aliased and bilinear respectively), with nine of our twelve participants preferring the pre-filtered technique overall. While this difference is not statistically significant, the lack of significance is likely due to the strong preference for one technique or the other on the part of three of our participants and the resulting large standard deviations for the mean preference scores. Figure 10 shows the mean preference scores for both compression techniques.

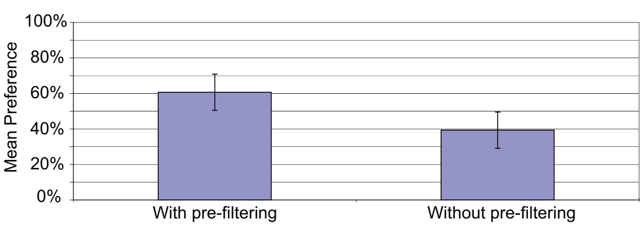

Figure 10: The percent of trials in which participants preferred each of the rendering techniques. Bars indicate standard error.

In accord with hypothesis two, there appears to be an interaction between compression technique and bitrate, as shown in Figure 11. As seen in the figure, the lower bitrates resulted in a higher preference for the pre-filtered technique, while the highest bitrate was the only bitrate in which our participants preferred the technique without pre-filtering overall.

This interaction may be due to the participants' multiple viewing of the videos during the experiment. Each scene used in our study had an object or person that was the focus of the scene. During the first viewing of the scene, our participants tended to focus on this main object; however, after viewing the videos several times, they began to inspect the background, the foreground, or other secondary objects in the scene. When the bitrate was high, and compression

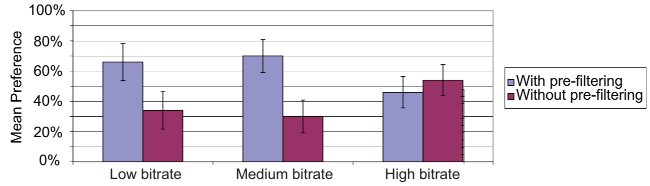

Figure 11: The percent of trials in which participants preferred each rendering technique for each of the three bitrates.

artifacts were few, the compression without pre-filtering produces clearer images in these regions of the scene farthest from the focal point. While this clarity results in object ghosting, which is a quality that the majority of our participants identified as distracting, participants were able to make out more detail in their subsequent viewings of the video. Given this observation and comments from our participants to this effect, we hypothesize that the preference for pre-filtered rendering would grow for 3D video viewed only a single time.

There also appears to be an interaction between the video clip shown and the rendering technique preferred by our participants, indicating that content is an important consideration when choosing compression techniques. Figure 12 shows the mean preference for both compression techniques for each of the five video clips used in the study. The scene with the ballerina is the only scene for which our participants preferred the compression without pre-filtering. This scene included not only a dancing ballerina, which appears in focus on the display, but also a dance partner that is closer to the viewer and slightly out of focus. Several participants mentioned that they could not see as much detail of the partner when viewing the pre-filtered version of the video.

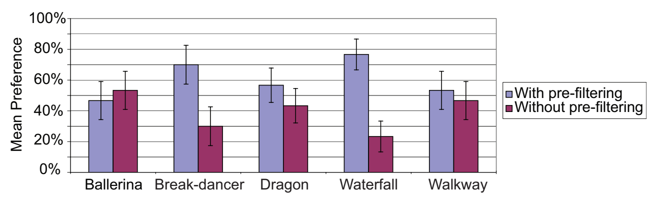

Figure 12: The percent of trials in which participants preferred each rendering technique for each of the five different video clips.

As shown in Figure 12, the waterfall clip resulted in the largest preference difference between rendering techniques. This clip contains a man's face in the foreground, and a complex moving background. Compression without pre-filtering wastes many bits on the complex background, such that the face in the foreground exhibits significantly more compression artifacts compared to the pre-filtering version. Because humans are very sensitive to the qualities of faces, these artifacts may have driven up our participants' preference for the pre-filtered version of this video. Without pre-filtering, the motion of the waterfall in the background also interacts with ghosting artifacts to produce visual noise. Several of our participants commented that the background of this scene was especially "chaotic" and "disorienting" when presented without pre-filtering. Interestingly, the same face is present in the foreground of the walkway video clip; however, this clip resulted in a lower difference between preference scores. The more static background of this sequence apparently leaves more bits available to encode the actor's face than does the complex, high frequency details of the waterfall. This indicates that complex scenes may benefit more from our pre-filtering technique.

6. CONCLUSIONS

We have presented a framework for multi-view video compression for 3D displays that takes into account the multi-dimensional display bandwidth. We proposed a pre-filtering step that band-limits the input to the display bandwidth before compression. Since the display bandwidth imposes a shallow depth of field, this removes high frequencies from the input signal, making it easier to compress. In addition, if pre-filtering is omitted, high frequencies that appear out of focus on the display lead to ghosting artifacts. The pre-filtering step also avoids these *inter-perspective aliasing* artifacts. Therefore, pre-filtering is beneficial for compression for two reasons: it reduces compression artifacts, and it avoids aliasing.

We have evaluated our technique with a preferential user study. We prepared pairs of multi-view video sequences of the same scene, compressed at the same bitrate, with and without our pre-filtering approach. We included three different bitrates and five scenes. We asked our subjects to indicate their "overall preference" for each pair of sequences. Our study finds that pre-filtering is an important parameter to optimize the visual quality of compressed 3D videos. Pre-filtering improves the quality most at lower bitrates, and in scenes with complex and moving backgrounds. However, if given the chance to scrutinize sequences at higher bitrates by repeated viewing, subjects tend to prefer compression without pre-filtering. This seems to indicate that they prefer sharper images with ghosting artifacts to images with a shallower depth of field.

The perceptual evaluation of compression techniques will play an important role in making multi-view 3D displays practical. We have introduced the notion of display pre-filtering in the compression pipeline, and shown its significance to optimize perceptual quality. However, much more work needs to be done in this area. Testing should be performed on sequences much longer than ours to take into account eye strain. The amount of pre-filtering could be adjusted to better explore the trade-off between ghosting and bluriness. In terms of compression, scalable techniques need to be developed that can produce the right amount of depth of field for different displays with different numbers of views. Finally, algorithms should be developed that allocate more bandwidth to objects that are in focus and less bandwidth to scene elements that are out of focus.

Acknowledgments

We are grateful to Microsoft Research for granting access to the *breakdancers* data set and to Aseem Agarwala for the waterfall scene.

7. REFERENCES

[1] T. Agocs, T. Balogh, T. Forgács, F. Bettio, E. Gobbetti, and G. Zanetti. A large scale interactive holographic display. In *Proc. IEEE VR 2006 Workshop on Emerging Display Technologies*, 2006.

[2] J. X. Chai, S. C. Chan, H. Y. Shum, and X. Tong. Plenoptic sampling. In *Computer Graphics*, SIGGRAPH 2000 Proceedings, pages 307–318, Los Angeles, CA, July 2000.

[3] A. Fitzgibbon, Y. Wexler, and A. Zisserman. Image-based rendering using image-based priors. *International Journal of Computer Vision*, 63(2):141–151, 2005.

[4] S. Gortler, R. Grzeszczuk, R. Szeliski, and M. Cohen. The lumigraph. In *Computer Graphics*.

[5] M. Halle. Holographic stereograms as discrete imaging systems. In *Practical Holography VIII*, volume 2176 of *SPIE Proceedings*, pages 73–84, 1994.

[6] P. S. Hilaire. Modulation transfer function and optimum sampling of holographics stereograms. *Applied Optics*, 33(5), February 1994.

[7] Holografika. Holovizio displays. http://www.holografika.com/, 2006. Cited September, 2006.

[8] A. Isaksen, L. McMillan, and S. J. Gortler. Dynamically reparameterized light fields. In *Computer Graphics*, SIGGRAPH 2000 Proc., July 2000.

[9] ISO/IEC JTC1/SC29/WG11. Survey of algorithms used for multi-view video coding (MVC), MPEG document MPEG2005/N6909, January 2005.

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

[11] J. Konrad and P. Agniel. Subsampling models and anti-alias filters for 3-d automultiscopic displays. *IEEE Trans. Image Proc.*, 2005.

[12] M. Levoy and P. Hanrahan. Light field rendering. In *Computer Graphics*, SIGGRAPH 96 Proceedings, pages 31–42, New Orleans, LS, Aug. 1996.

[13] L. Meesters, W. IJsselsteijn, and P. Seuntiens. A survey of perceptual evaluations and requirements of three-dimensional TV. *IEEE Transactions on Circuits and Systems for Video Technology*, (3):381–391, 2004.

[14] 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.

[15] C. N. Moller and A. Travis. Correcting interperspective aliasing in autostereoscopic displays. *IEEE Transactions on Visualization and Computer Graphics*, 11(2):228–236, March/April 2005.

[16] Newsight. http://www.newsight.com/, 2006. Cited January, 2006.

[17] T. Okoshi. *Three-Dimensional Imaging Techniques.* Academic Press, 1976.

[18] Y. Su, A. Vetro, and A. Smolic. Common test conditions for multiview video coding. Doc. JVT-U211, available online at http://ftp3.itu.ch/av-arch/jvt-site/, 2006.

[19] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High-quality video view interpolation using a layered representation. *ACM Trans. Graph.*, 23(3):600–608, 2004.

[20] M. Zwicker, W. Matusik, F. Durand, and H. Pfister. Antialiasing for automultiscopic 3d displays. In *Eurographics Symposium on Rendering*, 2006.