

## Resampling, Antialiasing, and Compression in Multiview 3-D Displays

Matthias Zwicker, Anthony Vetro, Sehoon Yea, Wojciech Matusik, Hanspeter Pfister, Fredo Durand

TR2007-084 April 2008

### Abstract

Multiview three-dimensional (3-D) displays offer viewing of high-resolution stereoscopic images from arbitrary positions without glasses. Such displays consist of view-dependent pixels that reveal a different color according to the viewing angle. Therefore, the left and right eye of an observer see slightly different images on the screen. This leads to the perception of 3-D depth and parallax effects when the observer moves. Although the basic optical principles of multiview auto-stereoscopy have been known for over a century [21], only recently displays with increased resolution, or systems based on multiple projectors, have made this approach practical. Today, commercial availability ranges from multiview desktop monitors [19] to large-scale displays based on multiprojector systems [1], [11].

*IEEE Signal Processing Magazine, November 2007*

This work may not be copied or reproduced in whole or in part for any commercial purpose. Permission to copy in whole or in part without payment of fee is granted for nonprofit educational and research purposes provided that all such whole or partial copies include the following: a notice that such copying is by permission of Mitsubishi Electric Research Laboratories, Inc.; an acknowledgment of the authors and individual contributions to the work; and all applicable portions of the copyright notice. Copying, reproduction, or republishing for any other purpose shall require a license with payment of fee to Mitsubishi Electric Research Laboratories, Inc. All rights reserved.



Matthias Zwicker, Anthony Vetro, Sehoon Yea,  
Wojciech Matusik, Hanspeter Pfister,  
and Frédo Durand



# Resampling, Antialiasing, and Compression in Multiview 3-D Displays

Addressing signal processing challenges

**M**ultiview three-dimensional (3-D) displays offer viewing of high-resolution stereoscopic images from arbitrary positions without glasses. Such displays consist of view-dependent pixels that reveal a different color according to the viewing angle. Therefore, the left and right eye of an observer see slightly different images on the screen. This leads to the perception of 3-D depth and parallax effects when the observer moves. Although the basic optical principles of multiview auto-stereoscopy have been known for over a century [21], only recently displays with increased resolution, or systems based on multiple projectors, have made this approach practical. Today, commercial availability ranges from multiview desktop monitors [19] to large-scale displays based on multiprojector systems [1], [11].

*Digital Object Identifier 10.1109/MSP.2007.905708*

View-dependent pixels can be implemented, for example, using conventional high-resolution displays and parallax barriers as shown in Figure 1. Other technologies include lenticular sheets and holographic screens [16]. Each view-dependent pixel can be thought of as emitting a small number of light rays in a set of discrete viewing directions, typically between eight and a few dozen. Often these directions are distributed in a horizontal plane, such that parallax effects are limited to horizontal motion of the observer. Figure 1 depicts a single scanline of a horizontal-parallax-only display.

Multiview 3-D displays feature a number of advantages over competing auto-stereoscopic display technologies, such as stereo-projection systems using shuttered or polarized glasses. Most important, multiview 3-D 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 multiuser capable, since none of the display parameters needs to be adjusted to an individual user. For these reasons, we believe that multiview 3-D displays will become the device of choice for a large number of applications such as scientific and medical visualization or remote collaboration. They have the potential to replace conventional two-dimensional (2-D) displays in the mass markets of digital entertainment.

In this article we survey recent advances that address signal processing challenges of multiview 3-D displays. In particular, effective signal processing algorithms are fundamental to display data at the highest quality without aliasing problems. They are also crucial to compress data effectively with minimal artifacts. At the core of these algorithms is an appropriate understanding of the sampling grid of multiview 3-D displays and their bandwidths.

## DISPLAY BANDWIDTH

We review three approaches to characterize display bandwidth. The first one, described in “Geometric Approach” builds on simple geometric considerations. Bandwidth, in this context, is defined as the smallest feature that can be reproduced by a display. This approach is useful to illustrate the fundamental properties of multiview 3-D displays, but it lacks a thorough theoretical foundation. The second approach, presented in “Fourier Optics,” models 3-D displays as optical imaging systems and analyzes them by using Fourier optics. The advantage of this approach is that it properly models diffraction effects. However, these effects are not significant for current displays with a small number of views. Finally, we describe how multiview 3-D displays can be modeled as higher-dimensional sampling grids in “Multidimensional Sampling.” Here, bandwidth is defined by the Shannon sampling theorem. This approach is preferable because it leads directly to practical algorithms for antialiasing, resampling, and compression.

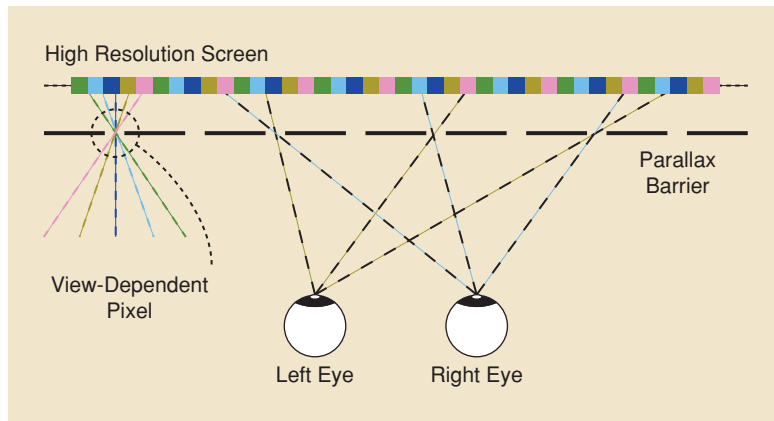
## GEOMETRIC APPROACH

A simple geometric approach for modeling multiview 3-D displays was proposed by Halle [8]. We explain this approach using Figure 2, which shows a typical geometry of the light rays emitted by a single scanline of a 3-D display. The discrete directions of all view-dependent pixels converge at a number of virtual perspective views at a given distance  $f$  from the display plane; i.e., this configuration multiplexes several perspective views on the display surface. For clarity of presentation we restrict ourselves to horizontal view-dependence, or so-called *horizontal-parallax-only displays*. It is straightforward, however, to extend the following analysis to include both horizontal and vertical view dependence.

It is easier to analyze the *feature size* using geometric considerations, instead of deriving the display bandwidth. The feature size describes how the display renders an infinitesimal point located in 3-D space. Intuitively, any point will be rendered at least as large as the feature size. In Figure 2 we see that if the 3-D point happens to lie on the display plane, such as point  $A$ , it will appear at the same location on the display in all views. Therefore, the feature size corresponds to the size  $\Delta t$  of the view-dependent pixels, which is the finest detail that the display can render. However, if the point lies at a certain depth  $d$  from the display plane, as for example point  $B$  in Figure 2, it is projected to two different locations in adjacent views. In effect, the point is spread over the extent between these two locations, which is also known as the *disparity* of the point. Therefore, the feature size  $w(d)$  for a point at a given depth  $d$  is

$$w(d) = \max \left( \Delta t, \left\| \frac{d}{f+d} \right\| \Delta v \right), \quad (1)$$

where we derived the second term on the right-hand side using similar triangles. Note that the size of multiview pixels  $\Delta t$  is much smaller than  $\Delta v$  in practice. Equation (1) means that scenes rendered using this 3-D display exhibit a shallow *depth of field*, with the display surface corresponding to the *focal plane*. By setting  $w(d) = \Delta t$  and solving for  $d$ , we see that the feature

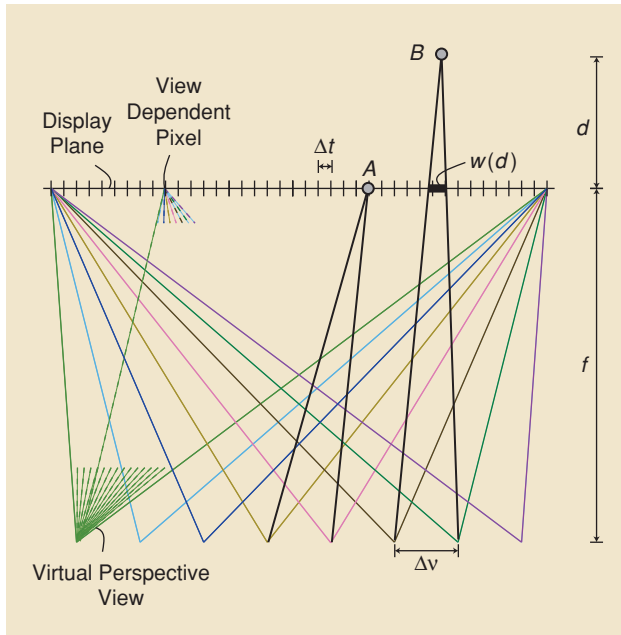


**[FIG1]** A multiview 3-D display implemented using a high-resolution screen and a parallax barrier. Each slit in the parallax barrier is a view-dependent pixel that reveals different colors based on the viewing direction.

size is  $\Delta t$  if  $\|d\| \leq f\Delta t/(\Delta v - \Delta t)$ . This means that scene elements are sharp if their distance to the focal plane is small enough. As the distance becomes larger, however, the feature size grows and scene elements become increasingly blurry.

### FOURIER OPTICS

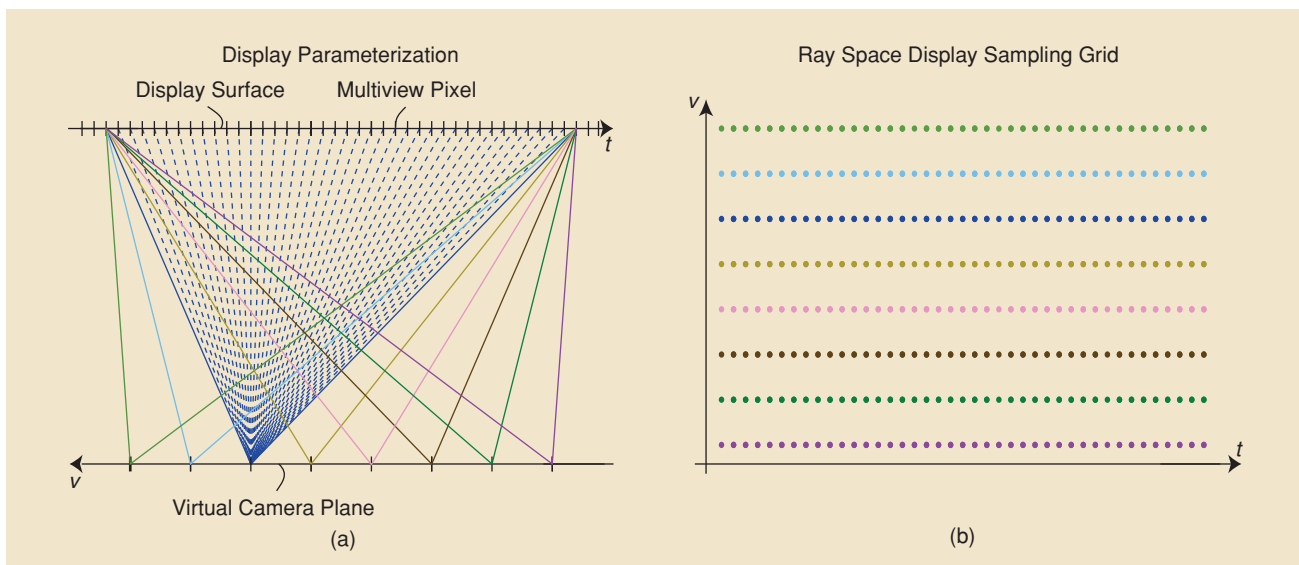
Rather than using geometric considerations, a more sophisticated approach is to take into account the wave nature of light and to model multiview 3-D displays as *optical imaging systems*.



**[FIG2]** Deriving the feature size based on geometric considerations. The spacing of view-dependent pixels is denoted by  $\Delta t$ . Virtual views converge at distance  $f$  from the display, and their spacing is  $\Delta v$ . Points at distance  $d$  from the display have disparity  $w(d)$ .

This analysis has its roots in the context of holographic stereograms [2], which multiplex several views using holographic materials and recording processes. Optical imaging systems are conveniently analyzed using Fourier optics [6]. Most important, Fourier optics provides ways to determine the *point spread function* and its Fourier transform, the *optical transfer function* (OTF), of an imaging system. Similar as the feature size discussed earlier, the point-spread function of a 3-D display describes how an infinitesimal point located in 3-D space is rendered. It can be interpreted as a low-pass filter that blurs the point to a finite size. The main advantage of the Fourier optics analysis, however, is that it properly models diffraction effects.

Halle [8] gives an intuitive account on how the wave properties of light explain the limited bandwidth of holographic stereograms. In addition, St. Hilaire [10] develops a more detailed derivation of their OTFs. He models stereograms as imaging systems with a generalized exit pupil [6] that corresponds in size to the spacing  $\Delta v$  between the virtual perspective views (Figure 2). In addition, he includes a focusing error that models the fact that light waves originate at a distance  $f$  from the viewer instead of the true distance  $f + d$ . St. Hilaire's results are applicable not only to holographic stereograms but also to other techniques to multiplex perspective views, such as parallax barriers and lenticular arrays [18]. His analysis shows that the OTF (and the point spread function) is affected by diffraction effects only if the spacing between the centers of projection of the perspective views ( $\Delta v$  in Figure 2) is smaller than about 1 mm at a typical viewing distance of about 60 cm. However, this implies that the display would multiplex hundreds of views. At most a few dozen views and much larger spacings are common in current displays. In this case, diffraction becomes negligible and the extent of the point spread function is determined by the same factor as in (1), confirming the geometric intuition.



**[FIG3]** Display parameterization in (a) ray space and the corresponding (b) ray space sampling grid. Each ray is parameterized by its intersection coordinates with the  $t$  and  $v$  axes, and the set of rays forms a rectangular sampling grid in ray space.

## MULTIDIMENSIONAL SAMPLING

The third approach is to analyze a *ray space* representation of multiview 3-D displays. This approach is related to the concept of light fields [7], [15], [22], which has been investigated extensively in the computer graphics and computer vision communities. The frequency analysis of light fields, also known as *plenoptic sampling theory*, has been studied by Chai et al. [3] and Isaksen et al. [12]. Here, we review the light field analysis of 3-D displays by Zwicker et al. [26].

Multiview 3-D displays seek to reproduce the array of light rays at a discrete number of locations and directions in the viewing zone. Each ray emitted by the display is parameterized by its intersection with two parallel planes, similar to the light field parameterization [7], [15]. In Figure 3(a), a  $t$ -plane coincides with the display plane, and a  $v$ -plane contains the centers of the virtual views. As discussed in “Geometric Approach,” the figure depicts a single scanline of a horizontal-parallax-only display. The  $t, v$ -intersection coordinates of all display rays form a 2-D rectangular sampling grid in ray space illustrated on the right in Figure 3. The combination of all scanlines can be interpreted as a 3-D sampling grid; displays with horizontal and vertical parallax would lead to four-dimensional (4-D) grids.

According to the Shannon sampling theorem, the geometry of the sampling grid immediately characterizes the display bandwidth as a rectangular box shown in Figure 4. The boundaries of the box correspond to the Nyquist limit of the grid. The box describes the bandwidth in a *joint domain* of frequencies  $\phi$  depending on the viewing direction (vertical axis, corresponding to  $v$ ) and spatial frequencies  $\theta$  (horizontal axis, corresponding to  $t$ ). Using insights from plenoptic sampling [3], we can derive a depth-dependent spatial bandwidth from the joint bandwidth. For example, the spectrum of a scene at infinity is a line with slope  $\partial\phi/\partial\theta = 1$ , and the corresponding spatial bandwidth is  $2\pi/\Delta v$ . The depth  $d = f\Delta t/(\Delta v - \Delta t)$  leads to a spectrum with slope  $\partial\phi/\partial\theta = \Delta t/\Delta v$ , which is the steepest slope that preserves the maximum spatial bandwidth  $2\pi/\Delta t$ . Reassuringly, this matches the results derived from (1). Ng [20] describes a similar observation in the context of light field cameras.

The advantage of the multidimensional sampling approach is that it describes the display bandwidth in a joint spatial and directional domain. This is in contrast to the geometric and the Fourier optics techniques, which derive a 2-D spatial bandwidth that depends on the distance of scene points to the display plane. Another difference is that the geometric and the Fourier optics approaches analyze the ability of the display to render infinitesimal points, while the multidimensional sampling approach is based on higher-dimensional sampling grids and the Shannon sampling theorem. In terms of signal processing, the former study post-aliasing and the latter prealiasing [5]. The multidimensional sampling approach leads to simple signal processing algorithms that can be expressed as linear filters in ray space. In addition, it can be used to model not only 3-D displays but also sampled multiview input data. This allows a unified treatment of multiview data acquisition, processing, and rendering as we will discuss in “Resampling.”

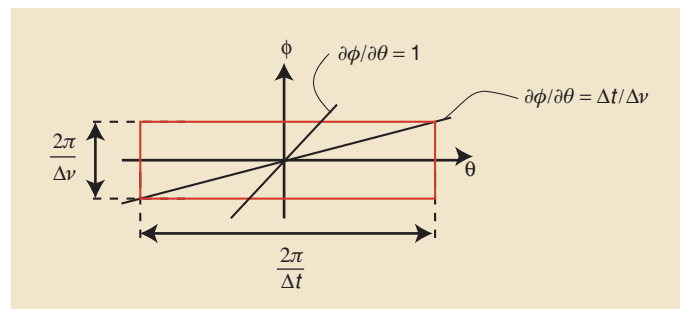
## ANTIALIASING

If we attempt to display scenes containing features that are smaller than the point spread function, or equivalently features that contain frequencies above the Nyquist limit, *interperspective aliasing* artifacts will appear. Halle [8] and Moller and Travis [18] derive several antialiasing algorithms from the depth dependent point spread function in (1). One approach is to filter the image plane projection of each point based on its depth. Alternatively, the depth planes of 3-D scenes can be filtered before projection. The drawback of both techniques, however, is that the depth of each point in the scene needs to be known. In addition, depth-dependent filtering leads to a spatially varying filter kernel.

In contrast, bandlimiting a multiview signal in its ray space representation does not require the knowledge of scene depth. One simply convolves the ray-space representation of the input signal with a 2-D *display prefilter* corresponding to the bandwidth in Figure 3 as proposed by Zwicker et al. [26]. This approach has also been described informally by Halle [8] and Moller and Travis [18]. Intuitively, it averages the perspective views that lie between the centers of projection sampled by the display.

We demonstrate antialiasing using ray-space filtering in Figure 7. In (a) we show a simulated view of a multiview 3-D display without band limiting the input signal. Interperspective aliasing appears as ghosting artifacts and low-frequency rippling patterns. The maximum size of the point spread function in this scene is shown as a black bar at the top of the images. We band-limit the scene to the display bandwidth using ray-space filtering on the right. The result exhibits the typical shallow depth of field imposed by the display bandwidth.

Our analysis so far assumed multiview displays with rectangular sampling grids, such that each scan line can be treated independently. However, most physical displays have irregular 2-D grids on the image plane (corresponding to the  $t$ -axis). Konrad et al. [14] describe the design of optimal filters for these grids, but they do not take into account sampling in the directional dimension (the  $v$ -axis). Therefore, their filters do not avoid inter-perspective aliasing. The design of optimal



**[FIG4]** The display bandwidth in the joint domain of frequencies  $\phi$  depending on the viewing direction (vertical axis) and spatial frequencies  $\theta$  (horizontal axis). A scene at infinity, i.e.,  $d = \infty$ , has a spectrum along the line with slope  $\partial\phi/\partial\theta = 1$ . This leads to a spatial bandwidth  $2\pi/\Delta v$ . A scene at a depth  $d = f\Delta t/(\Delta v - \Delta t)$  has a spectrum with slope  $\partial\phi/\partial\theta = \Delta t/\Delta v$ , which is the largest slope that preserves the maximum spatial frequency  $2\pi/\Delta t$ .



filters spanning the spatial and the directional dimensions is an interesting problem for future work.

## RESAMPLING

The analysis in the preceding section focused exclusively on the sampling characteristics of the display. In practice, however, input signals are represented in a sampled form as well, for example as multiview video streams. Here we focus on image acquisition systems where cameras are placed on a horizontal line with their optical axes perpendicular to that line as shown in Figure 8(a). This type of data is known as light fields, and it is conveniently analyzed using the ray-space analysis described in “Multidimensional Sampling.”

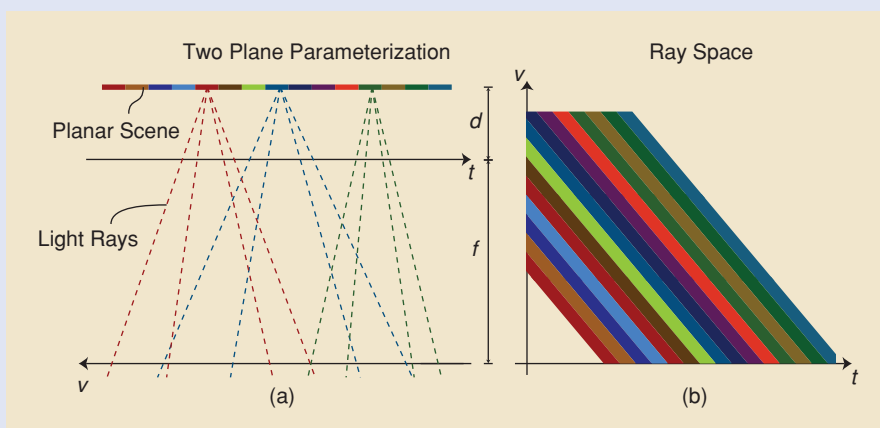
Each camera pixel corresponds to a ray, or a sample, that can be parameterized using its intersection with two planes  $t_{in}, v_{in}$  similar as in Figure 3 for the display samples. We assume that the camera plane is parallel to the display, but it is not necessarily at the same distance as the plane containing the virtual views. In addition, the number of cameras may be different from the number of virtual views. This means that there is no one-to-one correspondence between input and display samples as shown in Figure 8(b). Therefore, a full display pipeline involves a *resampling operation* that consists of the following steps: continuous reconstruction of the input signal, reparameterization to represent the signal in display coordinates, pre-filtering according to the display characteristics, and sampling

### FREQUENCY ANALYSIS OF LIGHT FIELDS

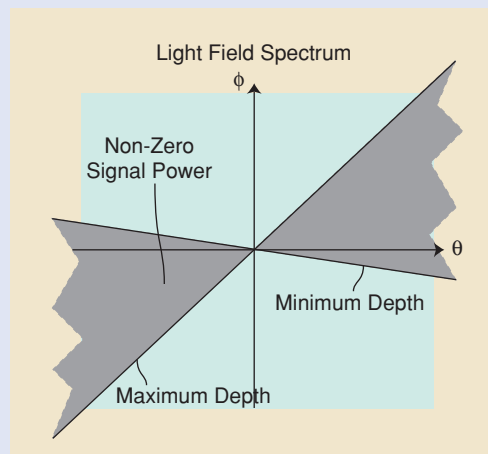
Light fields are a representation of the set of all light rays that are reflected by a 3-D scene. Each ray has a color and a set of coordinates that describes its trajectory. Light fields define each ray’s coordinates by its intersection with two parallel planes. Therefore, light fields constitute 4-D functions, or signals. Any arbitrary perspective view of the scene corresponds to a certain 2-D slice through the 4-D function. In practice, discrete light fields are acquired, or sampled, using camera arrays. Each camera pixel corresponds to a light ray, which is identified using the two-plane parameterization. Light fields can be visualized by extracting 2-D slices that are shown on regular displays. As an alternative, they can be reproduced using multiview 3-D displays as discussed in this article. Because these operations involve resampling the discrete input signals, it is useful to understand the frequency contents of these signals.

Here we summarize the main results of a frequency analysis of light fields due to Chai et al. [3] and Isaksen et al. [12]. For simplicity we illustrate the ideas in two dimensions. We first consider the special case of a planar scene that lies parallel to the parameterization planes as shown in Figure 5(a). We also assume that the scene consists of a perfectly diffuse material, which means that the reflected light at each point on the surface does not vary with direction. In Figure 5(b) we plot the color of each ray at its corresponding location in *ray space*, which is a 2-D plane parameterized by ray coordinates. We observe that each point in the scene corresponds to a straight line of constant color in ray space. The slope of the line is determined by the distance of the point from the  $t$  parameter plane as  $\partial v/\partial t(d) = -(d+f)/d$ . This means that the slope is infinity for points that coincide with the  $t$  plane, i.e.,  $d = 0$ , and it is  $-1$  for points at infinity.

From Figure 5 it is obvious that the Fourier transform of the ray space signal has a power spectrum that is zero everywhere except along a line *perpendicular* to the lines in ray space. Chai et al. [3] further conclude that the spectrum of an arbitrary scene is limited by the spectral lines corresponding to its minimum and maximum distance from the parameterization planes. This leads to the typical bow-tie shaped spectra as shown in Figure 6.



[FIG5] (a) Light field parameterization of a planar scene with constant depth  $d$ . (b) The ray space representation consists of parallel lines whose slope is determined by the depth of the scene.



[FIG6] The spectrum of a light field is limited by the spectral lines corresponding to the minimum and maximum depth in the scene.

on the display grid. Both the reconstruction and the sampling step are prone to aliasing artifacts, which can be avoided by the procedure outlined below.

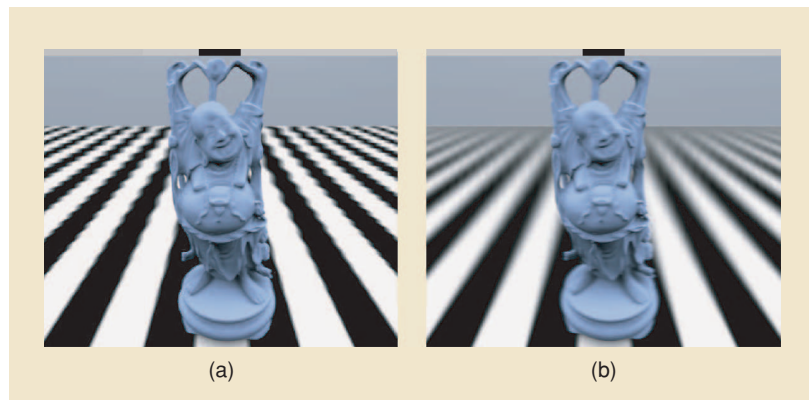
### RECONSTRUCTION FILTERS

Reconstructing light fields that are sampled using camera arrays as in Figure 3 has been studied extensively in the computer graphics literature. Chai et al. [3] realized that, in practice, most light fields contain aliasing because of the limited number of acquisition cameras. They propose to use reconstruction filters that are sheared in the directional dimension such that the bandwidth of the filter is as large as possible without including any aliasing. However, this approach still leads to the loss of spatial frequencies and blurry results in many practical situations. Stewart et al. [23] proposed an improved reconstruction filter that extends Chai's technique with a *wide aperture filter*. Wide aperture filters are very narrow in the directional domain but include all spatial frequencies. This filter preserves spatial frequencies at a specific depth at the cost of a small amount of aliasing. Reconstruction filters are usually applied in the spatial domain, but frequency domain computations may be advantageous in some cases [20].

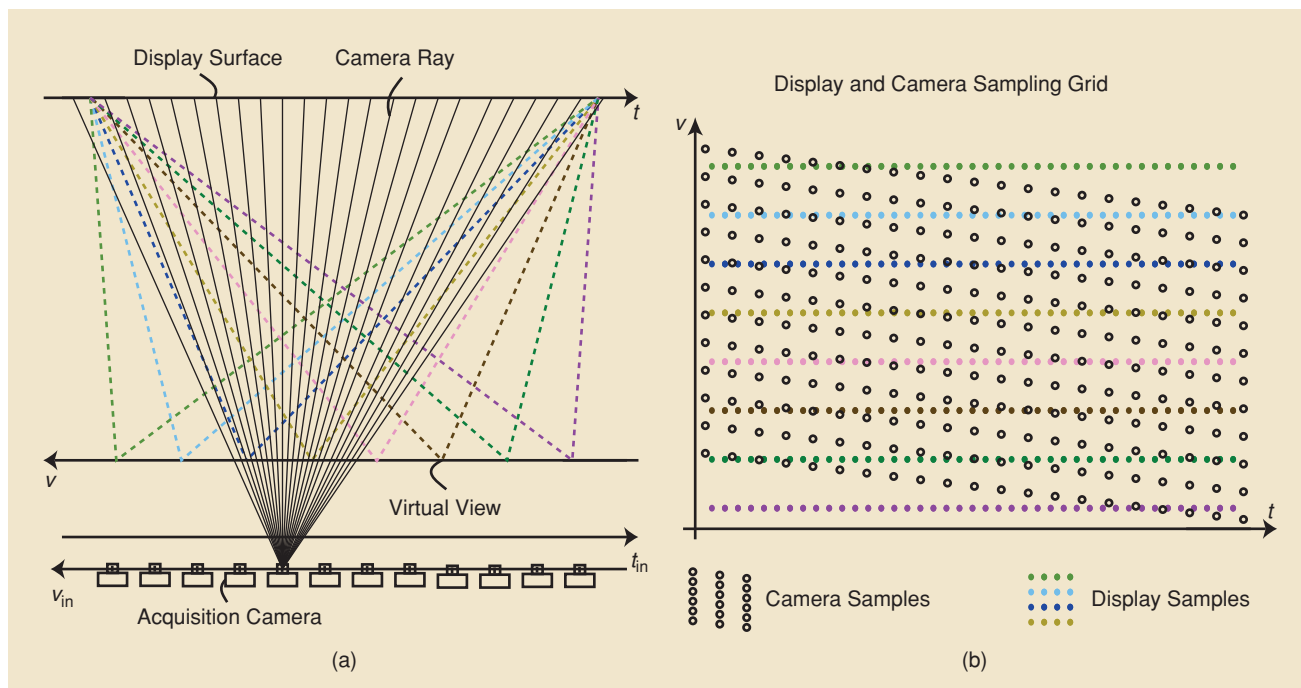
### COMBINED RESAMPLING FILTERS

After reconstruction, a display prefilter is applied to avoid aliasing due to sampling on the display grid. This requires the reparameteriza-

tion of the input signal in display coordinates. Note that Figure 8 implies a linear relationship, a vertical shear in this case, between display coordinates  $t, v$  and camera coordinates  $t_{in}, v_{in}$ . Therefore, it is straightforward to map the reconstruction filter from input to display coordinates [12], [26]. Zwicker et al. [16] construct a combined *resampling filter* by multiplying the reconstruction filter, after it is mapped to display coordinates, with the display prefilter. Figure 9 illustrates this procedure schematically. They approximate the resampling filter as a continuous kernel in the spatial domain using a combination of Gaussians, similar to image resampling algorithms [9]. In this example the Stewart reconstruction filter has been deliberately aligned with the display plane to preserve maximum signal bandwidth.



[FIG7] This figure shows simulated views of a multiview 3-D display. The black bars at the top of the images indicate the maximum size of the point-spread function of the display. (a) An example of interspersive aliasing and (b) the antialiased image using filtering in ray space.



[FIG8] (a) Multiview video acquisition using a camera array. Camera and display rays can be parameterized using the same  $t$  and  $v$  coordinates. (b) However, there is no one-to-one correspondence of camera and display samples in general.



Zwicker et al. [26] also discuss how to derive minimum acquisition sampling rates, i.e., the minimum number of cameras and their resolution, that are required to achieve high-quality results on a target display. Intuitively, the sampling rate is sufficient for a given display when no reconstruction aliasing appears within the bandwidth of the display. Increasing the acquisition sampling rate beyond this criterion does not increase display quality.

### VIEW INTERPOLATION

If the acquisition cameras are set up in a more general configuration than shown in Figure 8, it is not possible to derive a unified resampling filter as discussed above. In this case, it is preferable to use more powerful view interpolation algorithms [4], [25] to generate the display views directly. To avoid aliasing, however, one should generate virtual views at a smaller spacing than  $\Delta v$ . This increases the vertical spacing of the spectral replicas (see Figure 9) such that they do not overlap with the display bandwidth. These views can be bandlimited by applying the display

prefilter directly and subsampled to the number of views provided by the display. The drawback of this approach is that high-quality view interpolation algorithms usually do not allow real-time processing and are not robust enough for complex scenes. For example, it is difficult to extract depth maps [25] in real time without special hardware.

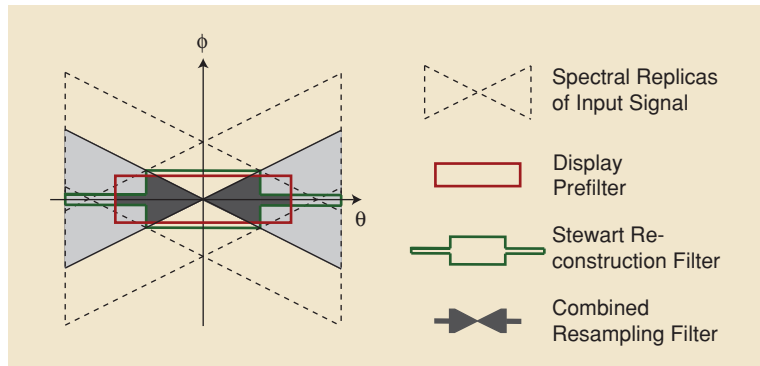
### DISPLAY PREFILTERING FOR MULTIVIEW COMPRESSION

Multiview 3-D displays require large amounts of data to achieve high image quality because they use video streams with multiple views. Therefore, data compression is of paramount importance for such systems. Here we review an application of the signal processing framework discussed in the previous sections to improve multiview video compression [17]. A straightforward approach is to add a *display prefiltering* step to conventional systems that is applied before compression [27]. The prefiltering step low-pass filters the multiview input signal as discussed in “Antialiasing.” Note that this approach is applicable only for systems in which the 3-D 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 scientific visualization, gaming systems, and 3-D television or cinema applications.

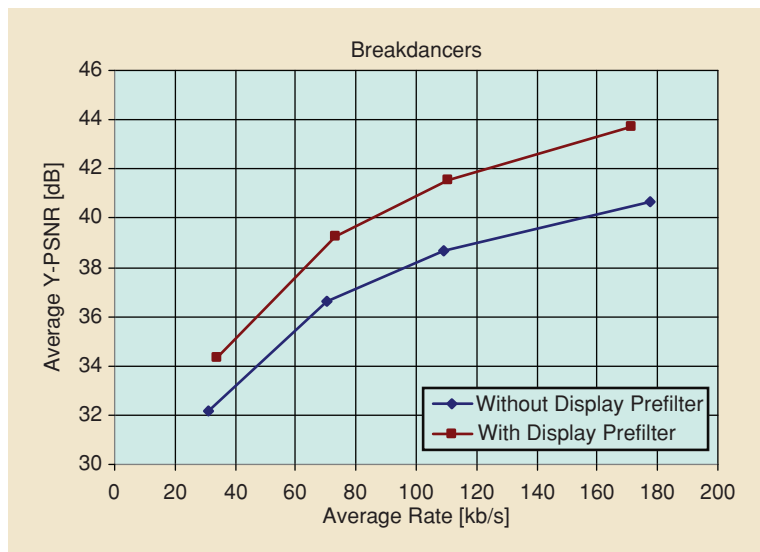
The key objective of the low-pass filter is that high-frequency content that is beyond the Nyquist limit of the display be removed from the input signal. Since these frequencies would appear as aliasing on the multiview display, the filtering step does not reduce image quality. However, it would have a positive effect on the compression efficiency by suppressing the energy in select parts of the input spectrum.

To demonstrate the effect of the display prefilter on compression, we plot rate-distortion curves comparing the quality of the compression of multiview videos with and without prefiltering at different bit rates in Figure 10. Simulations are conducted using the *breakdancer* data set [24]. These results 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 is just to demonstrate the degree of rate savings that are achieved when the multiview signal has been prefiltered with the primary purpose of removing antialiasing artifacts.

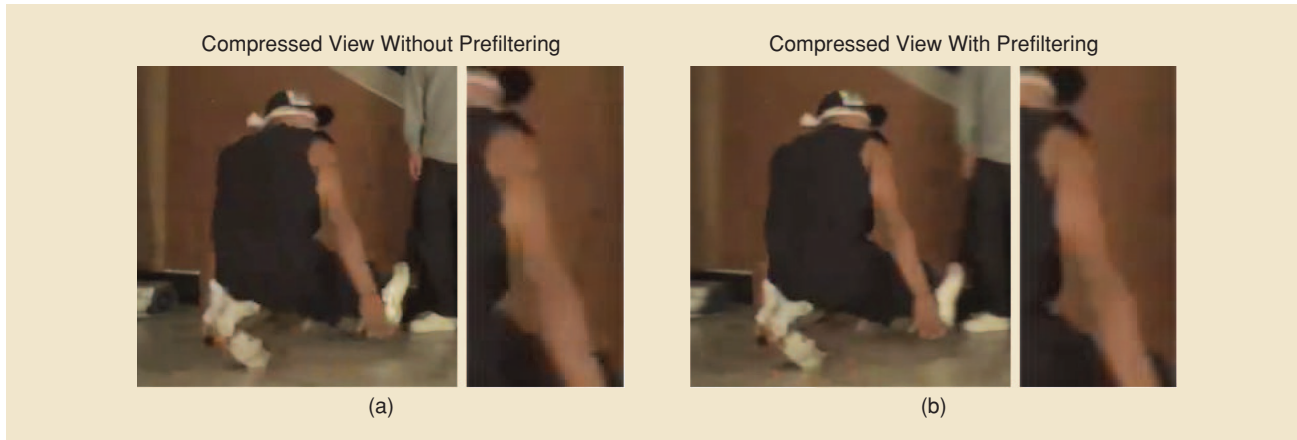
To evaluate the impact of the display prefilter on the reconstructed quality of the multiview video, we examine a sample frame from the *breakdancer* sequence. The multiview video is coded at an average bit-rate of 110 kb/s per view. Figure 11 shows the compressed frame (a) without prefiltering and (b) with prefiltering on the right. In (a), stronger blocking artifacts can be found in the foreground character compared to the case with prefiltering.



**[FIG9]** Construction of a resampling filter. The resampling filter is a product of a Stewart reconstruction filter and the display prefilter. The reconstruction filter eliminates aliasing from the sampled input signal, and the display prefilter band-limits the reconstructed signal to the display bandwidth.



**[FIG10]** Comparison of rate-distortion curves for the *breakdancer* sequence with and without prefiltering.



**[FIG11]** Comparison of compressed frames of a video sequence (a) without and (b) with prefiltering. Note that blocking artifacts are more pronounced when prefiltering is omitted.

This result suggests that the display prefiltering has benefits at lower bit rates where blocking artifacts begin to appear.

### CONCLUDING REMARKS

Multiview 3-D displays provide unrestricted stereoscopic viewing and parallax effects without special glasses or head tracking. They can be built by using high-resolution displays or multiprojector systems and parallax barriers or lenticular screens. Today, a variety of products are available commercially. Because of their advantages over other stereoscopic display technologies, multiview 3-D displays hold great promise for applications such as scientific visualization, remote collaboration, and digital entertainment.

This article surveyed different approaches to develop signal processing algorithms for these displays. Simple geometric considerations lead to the observation that multiview 3-D displays are subject to a limited depth of field. The geometric intuition is supported by a Fourier optics analysis. This analysis shows that as the number of views grows, diffraction effects will eventually limit the imaging capabilities of these displays. However, current displays operate far from the diffraction limit. A third approach models the displays as multidimensional sampling patterns, similar to light fields. This approach provides a unified analysis of multiview data acquisition using camera arrays and rendering on 3-D displays. It leads to simple algorithms for filtering, antialiasing, and resampling. For example, a display prefiltering stage can be integrated with conventional multiview video compression pipelines to increase compression efficiency.

Understanding the sampling properties of multiview 3-D displays leads to many useful insights. However, it may not be sufficient to develop algorithms that achieve the best possible results from a perceptual perspective. Designing perceptually optimal filters for processing, compressing and rendering multiview data on 3-D displays is an interesting and relatively unexplored avenue for future research. In particular, these filters should take into account the often nonrectangular geometry of the high-dimensional sampling grids.

### MULTIVIEW VIDEO CODING

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 [13]. 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 within the Joint Video Team (JVT), which consists of experts from ISO/IEC's Moving Pictures Experts Group (MPEG) and ITU-T's Video Coding Experts Group (VCEG). The aim is 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 neighboring 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 neighboring views at the same time instant. It has been shown that coding multiview video with interview prediction does give significantly better results compared to independent coding of each view [17]. 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.

Scalability is an important factor in the design of video codecs in general, and for multiview video coding in particular. Due to the large amount of data and potentially diverse receivers and displays, there exists a strong need to easily access specific views at a given time or output a desired spatial resolution. In contrast to nonscalable video coders, a scalable video coder produces a layered bit stream so that different fidelity, spatial resolutions, temporal resolutions, or views, in the case of multiview video, can be generated by simply truncating the scalable bit stream or accessing a select set of encoding units. With any layered scheme, dependencies between layers exist and will impact the portions of the bit-stream that are required to reconstruct the desired signal.

## AUTHORS

**Matthias Zwicker** (mzwicker@cs.ucsd.edu) is an assistant professor in Computer Science and Engineering at the University of California in San Diego. He earned his Ph.D. in computer science from the Federal Institute of Technology (ETH) in Zurich, Switzerland, in 2003. In 2003-2005, he was a post-doctoral associate at the the Massachusetts Institute of Technology. His research interests are in computer graphics, focusing on point-based methods, signal processing for image synthesis, 3-D geometry processing, and data-driven methods.

**Anthony Vetro** (avetro@merl.com) is with Mitsubishi Electric Research Labs, Cambridge, MA, where he is a Group Manager responsible for multimedia technology. His research interests include information coding, media streaming and multi-dimensional signal processing. He has published a number of papers in these areas and received several awards for his work on transcoding. He has also been an active member of MPEG and JVT standardization committees. Dr. Vetro is Chair-elect of the Technical Committee on Multimedia Signal Processing and an Associate Editor of IEEE Signal Processing Magazine. He is a Senior Member of the IEEE.

**Sehoon Yea** (yea@merl.com) received the B.S. degree in control and instrumentation engineering from Seoul National University, Korea, the M.S. degree in electrical engineering from Nagoya University, Japan, and the Ph.D. degree in electrical engineering from RPI (Rensselaer Polytechnic Institute), Troy, NY. From 1996 to 2001, he was a Research Engineer at IAE (Institute for Advanced Engineering) in Korea working on control systems design. In the summer of 2004, he was an Intern with Sarnoff Corporation, Princeton, NJ. He joined MERL in January 2006 as a member of Technical Staff. His research interests include digital image and video compression, enhancement and communication.

**Wojciech Matusik** (wmatusik@adobe.com) is a senior research scientist at Adobe Systems. He received a B.S. in EECS from the University of California at Berkeley in 1997, M.S. in EECS from MIT in 2001, and Ph.D. in 2003. In 2004, he was named one of the world's top 100 young innovators by MIT's Technology Review Magazine. His primary research lies in computer graphics, data-driven modelling, computational photography, and new display technologies.

**Hanspeter Pfister** (pfister@seas.harvard.edu) is Professor of the Practice in the School of Engineering and Applied Sciences (SEAS) and Director of Visual Computing at the Initiative in Innovative Computing (IIC) at Harvard University. Before joining Harvard he worked for 11 years at Mitsubishi Electric Research Laboratories (MERL). His research is at the intersection of visualization, computer graphics, and computer vision. He has a Ph.D. in computer science from the State University of New York at Stony Brook. He is chair of the IEEE Visualization and Graphics Technical Committee (VGTC).

**Frédo Durand** (fredo@csail.mit.edu) is an associate professor in Electrical Engineering and Computer Science at the Massachusetts Institute of Technology, and a member of the Computer Science and Artificial Intelligence Laboratory (CSAIL).

He received his Ph.D. from Grenoble University, France, in 1999, supervised by Claude Puech and George Drettakis. From 1999 till 2002, he was a post-doc in the MIT Computer Graphics Group with Julie Dorsey. He works both on synthetic image generation and computational photography, where new algorithms afford powerful image enhancement and the design of imaging system that can record richer information about a scene.

## REFERENCES

- [1] T. Agocs, T. Balogh, T. Forgcs, F. Bettio, E. Gobbetti, and G. Zanetti, "A large scale interactive holographic display," in *Proc. IEEE VR 2006 Workshop on Emerging Display Technologies*, 2006, pp. 311–312.
- [2] S.A. Benton, "Survey of holographic stereograms in processing and display of three-dimensional data," *Proc. the SPIE* 367, 1983, pp. 254–261.
- [3] J.-X. Chai, X. Tong, S.-C. Chan, and H.-Y. Shum, "Plenoptic sampling," in *Proc. ACM SIGGRAPH 2000*, July 2000, pp. 307–318.
- [4] A. Fitzgibbon, Y. Wexler, and A. Zisserman, "Image-based rendering using image-based priors," *Int. J. Comput. Vision*, vol. 63, no. 2, pp. 141–151, 2005.
- [5] A. Glassner, *Principles of Digital Image Synthesis*. San Mateo, CA: Morgan Kaufmann, 1995.
- [6] J.W. Goodman, *Introduction to Fourier Optics*, 2nd ed. New York: McGraw-Hill, 1996.
- [7] S.J. Gortler, R. Grzeszczuk, R. Szeliski, and M.F. Cohen, "The lumigraph," in *Proc. SIGGRAPH 96*, Aug. 1996, pp. 43–54.
- [8] M. Halle, "Holographic stereograms as discrete imaging systems," in *Practical Holography VIII*, vol. 2176 of *SPIE Proceedings*, 1994, pp. 73–84.
- [9] P. Heckbert, "Fundamentals of Texture Mapping and Image Warping," M.S. thesis, Ucb/csd 89/516, U.C. Berkeley, June 1989.
- [10] P. St Hilaire, "Modulation transfer function and optimum sampling of holographic stereograms," *Applied Optics*, vol. 33, no. 5, pp. 768–774, Feb. 1994.
- [11] Holografika. Holovizio displays (2006 Sept.). [Online]. Available: <http://www.holografika.com/>
- [12] A. Isaksen, L. McMillan, and S.J. Gortler, "Dynamically reparameterized light fields," in *Proc. ACM SIGGRAPH 2000, Computer Graphics Proceedings*, July 2000, pp. 297–306.
- [13] Advanced video coding for generic audiovisual services, ITU-T Rec. & ISO/IEC 14496-10 AVC, 2005. [Online]. Available: <http://www.itu.int/rec/T-REC-H.264>
- [14] J. Konrad and P. Agniel, "Subsampling models and anti-alias filters for 3-d autostereoscopic displays," *IEEE Trans. Image Proc.*, vol. 15, no. 1, pp. 128–140, 2005.
- [15] M. Levoy and P.M. Hanrahan, "Light field rendering," in *Proc. SIGGRAPH 96*, Aug. 1996, pp. 31–42.
- [16] M. Lucente and T.A. Galyean, "Rendering interactive holographic images," in *Proc. SIGGRAPH 95*, Aug. 1995, pp. 387–394.
- [17] 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. Conf. Multimedia & Expo*, 2006, pp. 1717–1720.
- [18] C.N. Moller and A.R.L. Travis, "Correcting interperspective aliasing in autostereoscopic displays," *IEEE Trans. Visual. Comput. Graphics*, vol. 11, no. 2, pp. 228–236, Mar./Apr. 2005.
- [19] Newsight (2006 Jan.) [Online]. Available: <http://www.newsight.com/>
- [20] R. Ng, "Fourier slice photography," *ACM Tran. Graphics*, vol. 24, no. 3, pp. 735–744, Aug. 2005.
- [21] T. Okoshi, *Three-Dimensional Imaging Techniques*. New York: Academic, 1976.
- [22] H.-Y. Shum, S.B. Kang, and S.-C. Chan, "Survey of image-based representations and compression techniques," *IEEE Trans. Circuits Syst. Video Theory*, vol. 13, no. 11, pp. 1020–1037, 2003.
- [23] J. Stewart, J. Yu, S.J. Gortler, and L. McMillan, "A new reconstruction filter for undersampled light fields," in *Proc. 14th Eurographics Workshop on Rendering*, June 2003, pp. 150–156.
- [24] Y. Su, A. Vetro, and A. Smolic, "Common test conditions for multiview video coding," (2006) Doc. JVT-U211, [Online]. Available: <http://ftp3.itu.ch/av-arch/jvt-site/>
- [25] L. Zitnick, S.B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," *ACM Trans. Graphics*, vol. 23, no. 3, pp. 600–608, Aug. 2004.
- [26] M. Zwicker, W. Matusik, F. Durand, and H. Pfister, "Antialiasing for automultiscopic 3d displays," in *Rendering Techniques 2006: 17th Eurographics Workshop on Rendering*, June 2006, pp. 73–82.
- [27] M. Zwicker, A. Vetro, S. Yea, C. Forlines, W. Matusik, and H. Pfister, "Display pre-filtering for multi-view video compression," in *Proc. the 15th Annual ACM Int. Conf. on Multimedia MULTIMEDIA '07*, to be published. 