Eurographics Symposium on Rendering (2006)
Tomas Akenine-Möller and Wolfgang Heidrich (Editors)

Antialiasing for Automultiscopic 3D Displays

M. Zwicker[1] W. Matusik[2] F. Durand[3] H. Pfister[2]

[1]Department of Computer Science and Engineering, University of California, San Diego
[2]Mitsubishi Electric Research Laboratories
[3]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology

Abstract

Automultiscopic displays show stereoscopic images that can be viewed from any viewpoint without special glasses. They hold great promise for the future of television and digital entertainment. However, the image quality on these 3D displays is currently not sufficient to appeal to the mass market. In this paper, we extend the frequency analysis of light fields to address some of the major issues in 3D cinematography for automultiscopic displays. First, we derive the bandwidth of 3D displays using ray-space analysis, and we introduce a method to quantify display depth of field. We show that this approach provides solid foundations to analyze and distinguish various aspects of aliasing. We then present an anti-aliasing technique for automultiscopic displays by combining a reconstruction and a display prefilter. Next, we show how to reparameterize multi-view inputs to optimally match the depth of field of a display to improve the image quality. Finally, we present guidelines for 3D content acquisition, such as optimal multi-view camera configuration and placement.

1. Introduction

For more than a century, the display of three-dimensional images has inspired the imagination and ingenuity of engineers and inventors. *Automultiscopic 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 (see Figure 1), lenticular sheets, or integral lens sheets. Although the optical principles of multiview auto-stereoscopy have been known for over a century [Oko76], it is only recently that displays with increased resolution have made them practical. Laptops with automultiscopic displays are shipping [Sha05], and high-quality automultiscopic desktop monitors cost about $3,000 [Opt05]. As a result, 3D television is getting renewed attention with Grundig's announcement that they will acquire, transmit, and display 3D content during the 2006 soccer world cup [Yah05].

However, automultiscopic displays today have several major drawbacks. Most importantly, they are plagued by disturbing visual artifacts that are most salient for moving observers, but also make long viewing from a static position an uncomfortable experience. Secondly, the acquisition of artifact-free 3D content is challenging. Photographers, videographers, and professionals in the broadcast and movie industry are unfamiliar with the complex setup required to record 3D content. There are currently no guidelines for multi-camera parameters, placement, and post-production processing. In particular, the image data captured by the acquisition system do not map to display pixels in a one-to-one fashion in most practical cases. This requires *resampling* of the content without introducing visual artifacts, which has never been studied for automultiscopic displays.

In this paper, we present a signal processing framework to deal with these problems in a principled way. We introduce a ray-space analysis of automultiscopic displays that allows us to characterize display bandwidth. In particular, our approach concisely explains the notion of display depth of field, and we show how to combat interperspective aliasing [Hal94,MT05] using linear filtering in ray space. We also extend our approach to resampling multi-view input that was acquired at a limited resolution without introducing aliasing artifacts. Next we introduce a method to cope with the shallow depth of field of practical 3D displays. It allows adapting the depth range of the acquired content to the depth of

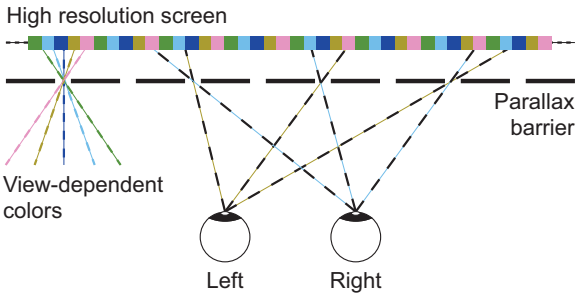

Figure 1: *An automultiscopic parallax barrier display with five view-dependent subpixels per multiview pixel. We distinguish automultiscopic displays with multiple view-dependent subpixels from autostereoscopic displays that project exactly two views, one for each eye.*

field of the display during post-production. Finally, we provide practical guidelines and formulas for 3D content acquisition. We explain how to compute the optimal configuration and placement of the multi-view camera given an estimate of the scene parameters.

Although the frequency analysis of ray space signals is a well known tool in computer graphics [CCST00], it has not been applied to study aliasing on 3D automultiscopic displays before. Isaksen et al. [IMG00] made the connection between light fields and automultiscopic displays, but they did not address aliasing artifacts due to limited display bandwidth. On the other hand, researchers in the 3D display community have been aware of this work, but they concluded that it could not be extended to address interperspective aliasing [MT05]. In this paper, we bridge the gap between previous work in the computer graphics and 3D display literature. Our approach provides novel algorithms for antialiasing and resampling, and precise guidelines for content acquisition.

2. Previous Work

Content for automultiscopic displays can be acquired using single cameras and fiber optics [JO02], stereo cameras [FKdB*02], or dense arrays of synchronized cameras [MP04, WSLH02]). Commercial automultiscopic displays [LF02, SG02] are based on parallax barriers or lenticular sheets placed on top of high resolution screens. Researchers have also used multi-projector systems [MP04, JJC*01]. While many acquisition and display systems provide only horizontal parallax, horizontal and vertical parallax can be obtained using fly's eye (or integral) lens sheets [JO02]. Even though we restrict our discussion to acquisition and display with horizontal parallax and few views per pixel, the principles presented in this paper apply to all these cases. However, we do not consider head-tracked stereoscopic display such as proposed by Perlin et al. [PPK00] and Sandin et al. [SMG*05]. The acquisition of stereo image pairs for such displays has been discussed in depth by Holliman et al. [Hol04, JLHE01]. However, stereo image pairs do not suffer from the aliasing artifacts that we are focusing on in this paper.

Most previous work has approached anti-aliasing for 3D displays using wave optics [MT05, HOY00, Hil94]. Halle [Hal94] provides both a geometrical and a wave optics perspective for anti-aliasing. All the previously proposed algorithms have difficulties to handle occlusion and specular surfaces. Furthermore, they require knowledge of per pixel scene depth for appropriate filtering. In the absence of depth information, the algorithms resort to a conservative worst case approach and filter based on the maximum depth in the scene [MT05]. In practice, this limits scenes that can be displayed with reasonable quality to ones with very shallow depth. The most critical conceptual difference with our work is that previous work analyzed 3D display aliasing by studying the reproduction of a single 3D point, while we cast it as a multidimensional sampling problem using the full signal.

Automultiscopic displays emit light fields [LH96], which represent radiance as a function of position and direction in regions of space free of occluders. A frequency analysis of light fields, which is also known as *plenoptic sampling theory*, was first introduced by Chai et al. [CCST00] and Isaksen et al. [IMG00]. They analyze the spectrum of a scene as a function of object depth and show that most light fields captured using camera arrays are aliased. They propose reconstruction filters that remove aliasing and preserve as much as possible of the original spectrum. Isaksen et al. [IMG00] also demonstrate that reparameterization can be used to display light fields on automultiscopic displays. More recently, Stewart et al. [SYGM03] proposed to enhance Chai's reconstruction filter with a wide aperture filter. This leads to images with a larger depth of field than Isaksen's approach without sacrificing the sharpness on the focal plane. Our approach builds on Isaksen et al.'s work [IMG00] on re-parameterizing light fields for automultiscopic display. However, they focus on combatting aliasing during light field reconstruction and they do not take display bandwidth into consideration. In contrast, we present a resampling framework that addresses both reconstruction aliasing and aliasing due to limited display bandwidth.

3. Signal Processing for 3D Displays

In this section, we introduce a multidimensional signal processing framework to study 3D displays. Our analysis is based on a ray space representation of automultiscopic 3D displays, which we explain in Section 3.1. In Section 3.2 we show how this allows us to characterize the display bandwidth, and in Section 3.3 we describe the connection between display bandwidth and depth of field. Our approach marks a significant departure from previous work in the automultiscopic display community that analyzed the reproduction of a single 3D point [Hil94, Hal94, MT05] rather than a signal representing a full scene.

3.1. Ray Space Representation

Multiview autostereoscopic displays seek to reproduce the light array for every location and direction in the viewing zone. We parameterize light rays by their intersection with two planes [LH96, GGSC]. For a parallax-barrier display, we use the parallax barrier plane as the t coordinate, and the high resolution screen as the v coordinate (Figure 2). We follow Chai et al.'s [CCST00] conventions that the t and v axes have opposite orientations and the v coordinate of a ray is relative to its t coordinate. All rays intersecting the t plane at one location correspond to one *multi-view pixel*, and each intersection with the v plane is a *view-dependent subpixel*. We call the number of multi-view pixels the *spatial resolution* and the number of view-dependent subpixels per multi-view pixel the *angular resolution*. The display rays form a higher-

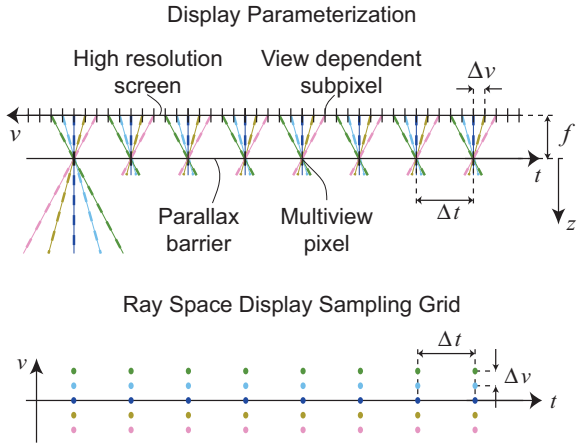

Figure 2: *Parameterization of a scanline of a multiview display and corresponding 2D sampling grid. Because the v coordinate of a ray is relative to its t coordinate, all vertical rays (shown at the top of the figure) lie on the t axis with v=0 (shown at the bottom).*

dimensional grid in ray space[†] (Figure 2, bottom). Each ray in the top of Figure 2 corresponds to one sample point at the bottom of the figure. Current digital automultiscopic displays provide only horizontal parallax, i.e., they sample only in the horizontal direction on the v plane. Hence, we can treat each scanline on the t plane independently, which leads to a two-dimensional ray space. We use the term *display view* to denote a slice of ray space with $v = const$ (note that these views are parallel projections of the scene). Without loss of generality, we assume the distance between the planes f has unit length (see Figure 2), and we omit factors f in our equations. However, this unit factor needs to be considered to ensure the consistency of physical units.

3.2. Display Bandwidth

The sampling grid in Figure 2 imposes a strict limit on the bandwidth that can be represented by the display, known as the *Nyquist limit* (Figure 3 left). Let us denote angular and spatial frequencies by ϕ and θ, and sample spacing by Δv and Δt. Then the *display bandwidth* is given by

$$H(\phi,\theta) = \begin{cases} 1 & \text{for} \quad |\phi| \le \pi/\Delta v \quad \text{and} \quad |\theta| \le \pi/\Delta t, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

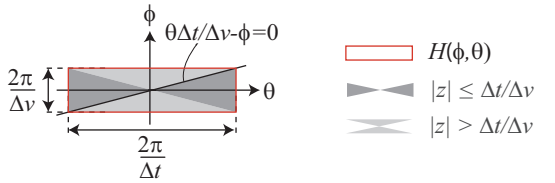

Figure 3: *Ray-space bandwidth of automultiscopic displays. Note that the relative scaling of the ϕ and θ axes is arbitrary. We chose the scaling to indicate the relative resolution of the two axes, which is usually two orders of magnitude larger in the spatial direction (θ axis) than in the angular direction (ϕ axis). The depth of field of the display is given by the diagonals of its rectangular bandwidth.*

3.3. Display Depth of Field

Chai et al. [CCST00] and Isaksen et al. [IMG00] have shown that the spectrum of a light field, or ray space signal, of a scene with constant depth is given by a line $\phi/z + \theta = 0$, where z is the distance from the t-plane. From Figure 3 we see that for scenes at depths $|z| \le \Delta t/\Delta v$, their spectral lines intersect the rectangular display bandwidth on its left and right vertical boundary. This means these scenes can be shown at the highest spatial resolution $\theta = \pi/\Delta t$ of the display. However, for scenes with $|z| > \Delta t/\Delta v$, their spectra intersect the display bandwidth on the horizontal boundary. As a consequence, their spatial frequency is reduced to $\theta = \pi/(\Delta v z)$. This is below the spatial resolution of the display, hence these scenes will appear blurry.

This behavior is dual to photographic depth of field effects and the range of exact refocusing in light field photography [Ng05] . The range $|z| \le \Delta t/\Delta v$ is the range that can be reproduced by a 3D display at maximum spatial resolution and we call it the *depth of field of the display*. Similar to light field photography, the depth of field is proportional to $1/\Delta v$, or the Nyquist limit in the angular dimension.

[†] We are aware that most physical displays do not correspond to a quadrilateral sampling grid as shown in Figure 2. A similar analysis can be developed for these grids. Non-quadrilateral grids have been investigated by Konrad et al. [KA05], but they only considered sampling in individual views ($v = const$ slices) instead of (v,t) ray space.

Since current displays have a very limited angular bandwidth, they exhibit a *shallow* depth of field. A commercial $23''$ parallax barrier display [Opt05] has a pitch of $0.25mm$ for the view-dependent subpixels and a distance of $4mm$ between the high-resolution screen and the parallax barrier. This corresponds to 8 view-dependent subpixels per multiview pixel that subtend an angle, or a *field of view*, of about 25 degrees. To parameterize the display rays, we choose the parallax barrier to coincide with the t plane and place the v plane at unit distance. Therefore, we get $\Delta t = 2mm$ and $\Delta v = 0.0625mm$, which yields a depth of field of a mere $\pm 32mm$. This means that any scene element that appears at a distance larger than $32mm$ from the display surface will be blurry. Although this seems like a very small range, it is sufficient to create a convincing illusion of depth perception for viewing distances up to a few meters.

To characterize scenes with respect to a given display, it is useful to specify scene depth relative to the depth of field of the display. Interestingly, the ratio of scene depth over depth of field $d(z) = z\Delta v/\Delta t$ corresponds to the *disparity* between views on the display. By this definition, scenes with maximum disparity $d < 1$ lie within the depth of field of the display. A given disparity $d > 1$ means that the spatial bandwidth is reduced by a factor of $1/d$. This is equivalent to Halle's geometric sampling criterion [Hal94].

4. Resampling for 3D Displays

With the bandwidth analysis from the previous section it is straightforward to antialias *continuous* input signals. We would simply multiply the spectrum of the input signal with a low-pass filter that has the same bandwidth as the display. In practice, however, light fields are represented as *sampled* signals, which are usually acquired using camera arrays. To show a sampled light field on an automultiscopic display, the samples of the input light field need to be mapped to the samples, i.e., pixels, of the display. Unfortunately, the sampling patterns of typical light field acquisition devices, such as a camera rigs, and automultiscopic displays do not lead to a one-to-one correspondence of rays. Hence, showing a light field on an automultiscopic display involves a *resampling* operation.

In this section, we introduce a resampling framework that avoids aliasing artifacts due to both sampling steps involved in a light-field acquisition and display pipeline, i.e., the sampling that occurs during scene acquisition, and sampling that is performed when mapping camera samples to display pixels. Our approach is based on the resampling methodology proposed by Heckbert [Hec89] in the context of texture mapping. In Section 4.1, we first describe how to reparameterize the input light field and represent it in the same coordinate system as the display. This allows us to derive a resampling filter, which combines reconstruction and prefiltering, as described in Section 4.2.

4.1. Reparameterization

Before deriving a combined resampling filter, we need to establish a common parameterization for the input light field and the 3D display (Figure 4, left). We restrict the discussion to the most common case where the light field parameterizations are parallel to the display. Denote input coordinates of the camera and the focal plane by t_{in} and v_{in}, the distance from the t_{in} axis by z_{in}, and the sampling distances by Δt_{in} and Δv_{in}. The t_{in} axis is also called the *camera baseline*. Similarly, we use display coordinates t_d, v_d, z_d, Δt_d, and Δv_d. Without loss of generality, we assume a unit distance between the t- and v-planes for both the display and the input light field.

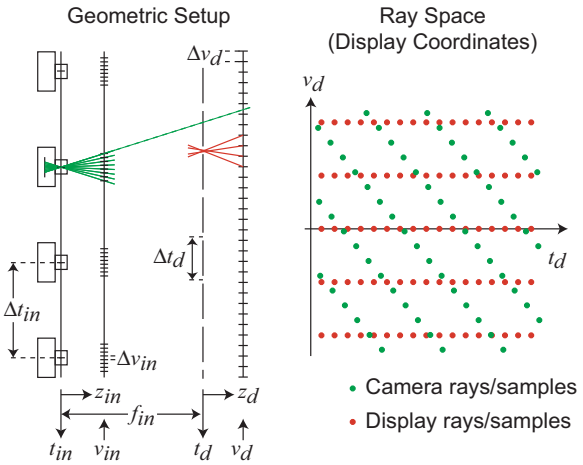

Figure 4: *Mapping the input rays to display coordinates: Geometric setup (left) and sampling grids in ray space (right). The acquisition parameterization is translated by a distance f_{in} compared to the display, which corresponds to a shear in ray space. Each ray on the left corresponds to one sample point on the right.*

The relation between input and display coordinates is given by a single parameter f_{in}, which is the distance between the camera plane t_{in} and the zero-disparity plane t_d of the display. This translation corresponds to a shear in ray space

$$\begin{bmatrix} v_{in} \\ t_{in} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ f_{in} & 1 \end{bmatrix} \begin{bmatrix} v_d \\ t_d \end{bmatrix} = \mathbf{M} \begin{bmatrix} v_d \\ t_d \end{bmatrix}. \quad (2)$$

Automultiscopic displays based on parallax barriers usually have a high spatial resolution (several hundred multiview-pixels per scanline) and low angular resolution (around ten view-dependent sub-pixels), while acquired light fields have a low spatial resolution (a few dozen cameras) and high angular resolution (several hundred pixels per scanline). This leads to two sampling grids that are highly anisotropic and that do not have the same orientations (Figure 4, right). Of

course it is also possible to construct automultiscopic displays with a one-to-one correspondence between camera and display rays, for example using projectors [MP04]. However, this does not solve the aliasing problem that we are addressing.

4.2. Combined Resampling Filter

Our resampling technique for 3D display antialiasing proceeds in three steps as illustrated graphically in Figure 5, where we represent signals and filters in the frequency domain. First, a continuous signal is reconstructed from the input data given in its original parameterization, which we denote by angular and spatial frequencies ϕ_{in} and θ_{in}. Care has to be taken to avoid aliasing problems in this step and to make optimal use of the input signal. We apply advanced reconstruction filters from the light field rendering literature [SYGM03, CCST00] to address these issues. These techniques strive to extract a maximum area of the central replica from the sampled spectrum while discarding areas that overlap with neighboring replicas. Next, we reparameterize the signal to *display coordinates*, denoted by ϕ_d and θ_d, using the mapping derived in Section 4.1. The signal is then prefiltered to match the Nyquist limit of the display pixel grid as described in Section 3.1 and sampled on the display pixel grid. Prefiltering guarantees that replicas of the sampled signal in display coordinates do not overlap.

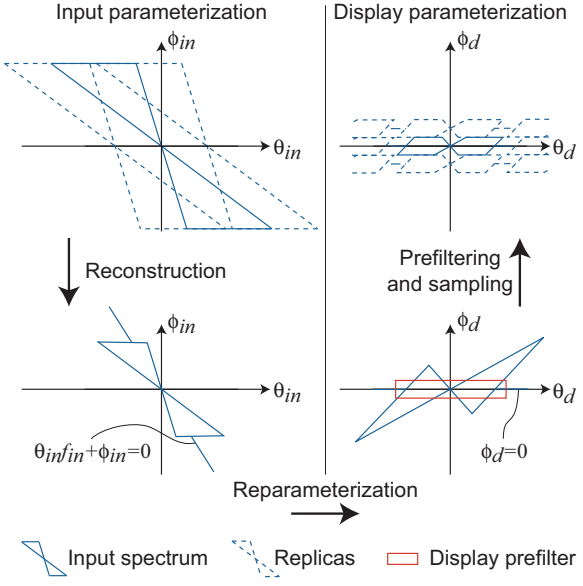

Figure 5: *The light field spectrum in input coordinates (left) and display coordinates (right).*

We now derive a unified resampling filter by combining the three steps described above. We operate in the spatial domain, which is more useful for practical implementation. We proceed as follows:

1. Given samples $\xi_{i,j}$ of an input light field, we recover a continuous light field l_{in}:

$$l_{in}(v_{in},t_{in}) = \sum_{i,j} \xi_{i,j}\, r\left(\begin{bmatrix} v_{in} - i\Delta v_{in} \\ t_{in} - j\Delta t_{in} \end{bmatrix}\right), \qquad (3)$$

   where r denotes the light field reconstruction kernel. Figure 5 indicates that we are using a Stewart [SYGM03] filter.
2. Using Equation 2, we express the reconstructed light field in display coordinates as

$$l_d(v_d,t_d) = l_{in}\left(\mathbf{M}\begin{bmatrix} v_d \\ t_d \end{bmatrix}\right). \qquad (4)$$

3. We convolve the reconstructed light field, represented in display coordinates, with the display prefilter h, which yields the band-limited signal

$$\tilde{l}_d(v_d,t_d) = (l_d \otimes h)(v_d,t_d). \qquad (5)$$

   Sampling this signal on the display grid will not produce any aliasing artifacts.

By combining the above three steps, we express the band-limited signal as a weighted sum of input samples

$$\tilde{l}_d(v_d,t_d) = \sum_{i,j} \xi_{i,j}\, \rho\left(\begin{bmatrix} v_d \\ t_d \end{bmatrix} - \mathbf{M}^{-1}\begin{bmatrix} i\Delta v_{in} \\ j\Delta t_{in} \end{bmatrix}\right). \qquad (6)$$

The weighting kernel ρ is the so-called *resampling filter*. It is defined as the convolution of the reconstruction kernel, expressed in display coordinates, and the prefilter

$$\rho(v_d,t_d) = (r(\mathbf{M}[\cdot]) \otimes h)(v_d,t_d). \qquad (7)$$

We implemented light field resampling filters using Gaussians [Hec89] to produce all results shown in this paper.

Since both the reconstruction filter and the prefilter are highly anisotropic, as shown in Figure 5, we need to carefully align them to preserve as much signal bandwidth as possible. Note that Equation 2 implies $[\phi_{in},\theta_{in}] = [\phi_d,\theta_d]\,\mathbf{M}^{-1}$. Therefore, the input spectrum is sheared down along the vertical axis. We also observe that the line $\theta_{in}f_{in} + \phi_{in} = 0$, corresponding to depth $z_{in} = f_{in}$, is mapped to the zero-disparity plane of the display. Hence, the depth of field of the display, expressed in input coordinates, lies at distances $f_{in} \pm \Delta t/\Delta v$ from the cameras. This means that the distance f_{in} between the camera plane and the display plane should be chosen such that for objects of interest $z_{in} - f_{in} = z_d < \Delta_t/\Delta v$.

Figure 6 illustrates the effectiveness of the resampling framework using simulated perspective views of an automultiscopic display. The views have a horizontal resolution of 472 multiview pixels, and the scene has a maximum display disparity of 56 pixels. We used 100 input views to allow for high quality reconstruction using a Stewart [SYGM03] reconstruction filter. In this situation, the reconstructed signal has a higher bandwidth than the display. If we omit the display prefilter *pre-aliasing* becomes apparent as rippling artifacts in Figure 6a. With the display prefilter, the simulated

view in Figure 6a exhibits a shallow depth of field as predicted by our bandwidth analysis and aliasing artifacts are avoided.

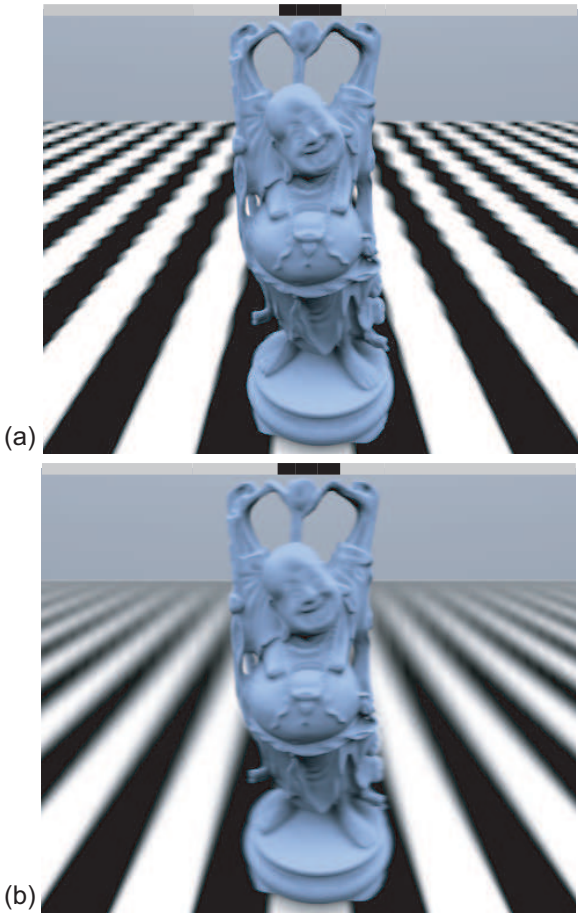

Figure 6: *Resampling results: (a) Simulated view using only the reconstruction filter. This leads to* pre-aliasing. *(b) The combined resampling filter eliminates pre-aliasing. The black bar at the top of the images indicates the maximum display disparity* (see also color plate).

5. Baseline and Depth of Field

The relation between the 3D display and the input light field as described in the last section implies that the display acts as a *virtual window* to a *uniformly scaled scene*. The display reproduces the light field of the scene at a different, usually smaller, scale. However, often it is neither desirable nor practically possible to achieve this. It is not unusual that the depth range of the scene by far exceeds the depth of field of the display, which is very shallow as shown in Section 3.1. This means that large parts of the scene are outside the display bandwidth, which may lead to overly blurred views. In addition, for scenes where the object of interest is far from the cameras, like in outdoor settings, the above assumption means that a very large camera baseline is required. It would also mean that the pair of stereoscopic views seen by an observer of the display would correspond to cameras that are physically far apart, much further than the two eyes of an observer in the real scene.

We show how these problems can be solved by changing the size of the camera baseline. As shown in Section 5.1, this can be expressed as an additional linear transformation of the input light field that reduces the displayed depth of the scene. We describe in Section 5.2 how this additional degree of freedom allows a user to specify a desired depth range in the input scene that needs to be in focus. We deduce the required baseline scaling that maps this depth range to the display depth of field.

5.1. Baseline Scaling

Modifying the camera baseline during acquisition corresponds to the transformation of the displayed configuration shown in Figure 7. An observer at a given position sees the perspective view that is acquired by a camera closer to the center of the baseline. That is, we remap each acquired camera ray such that its intersection with the baseline (i.e., the t_{in} plane) is scaled by a factor $s > 1$, while its intersection with the zero-disparity plane of the display (i.e., the t_d-plane) is preserved.

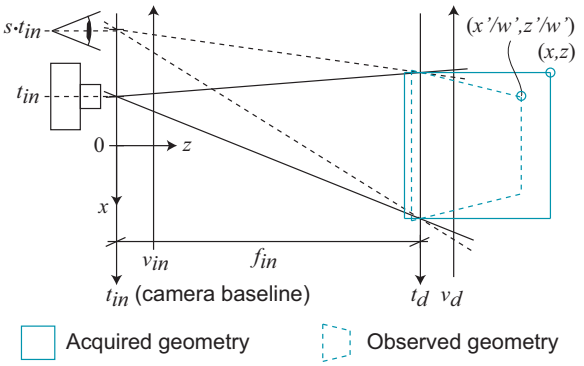

Figure 7: *To change the size of the baseline, we scale rays along t_{in} while preserving their intersection with t_d.*

It is easy to see that this mapping corresponds to a linear transformation of input ray space, and that any linear transformation of ray space corresponds to a projective transformation of the scene geometry. For the transformation shown in Figure 7, the projective transformation is

$$\begin{bmatrix} x' \\ z' \\ w' \end{bmatrix} = \begin{bmatrix} sf_{in} & 0 & 0 \\ 0 & sf_{in} & 0 \\ 0 & s-1 & f_{in} \end{bmatrix} \begin{bmatrix} x \\ z \\ 1 \end{bmatrix}, \qquad (8)$$

i.e., a point (x,z) in the scene is mapped to $(x'/w', z'/w')$.

The projective transformation of scene geometry is also illustrated in Figure 7. This scene transformation is closely related to depth reduction techniques from the stereoscopic displays literature [WHR99, JLHE01], which are used to aid stereo-view fusion. As noted by Wartell et al. [WHR99] and Jones et al. [JLHE01], we observe that the transformation moves points at infinity, i.e., $z = \infty$, to a finite depth $z'/w' = (f_{in}s/(s-1+f_{in})$. In addition, as s approaches infinity, z'/w' approaches f_{in}. This means that scene depth is compressed towards the zero-disparity plane of the display.

We generalize the transformation from display to input coordinates by including the mapping shown in Figure 7, which leads to

$$\begin{bmatrix} v_{in} \\ t_{in} \end{bmatrix} = \begin{bmatrix} \frac{1}{f_{in}} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & s-1 \\ 0 & s \end{bmatrix}^{-1} \begin{bmatrix} f_{in} & 0 \\ f_{in} & 1 \end{bmatrix} \begin{bmatrix} v_d \\ t_d \end{bmatrix} = \mathbf{M}(f_{in}, s) \begin{bmatrix} v_d \\ t_d \end{bmatrix}. \quad (9)$$

We call this mapping $\mathbf{M}(f_{in}, s)$ to emphasize that it is determined by the free parameters f_{in} and s.

5.2. Controlling Scene Depth of Field

In a practical scenario, a user wants to ensure that a given depth range in the scene is mapped into the depth of field of the display and appears sharp. Recall that the bandwidth of scene elements within a limited depth range is bounded by two spectral lines [CCST00]. In addition, the depth of field of the display is given by the diagonals of its rectangular bandwidth as shown in Section 3.1. Using the two free parameters in Equation 9, s for scaling the baseline and f_{in} for positioning the zero-disparity plane of the display with respect to the scene, we determine a mapping that aligns these two pairs of lines, which achieves the desired effect.

We compute the mapping by simply equating the two corresponding pairs of spectral lines, i.e., the pair that bounds the user specified depth range mapped to display coordinates, and the pair that defines the depth of field of the display. Let us denote the minimum and maximum scene depth that the user desires to be in focus on the display by z_{front} and z_{back}. The solution for the parameters s and f_{in} is

$$f_{in} = \frac{2z_{max}z_{min} + \frac{\Delta t}{\Delta v}(z_{max} - z_{min})}{(z_{min} + z_{max})}, \quad (10)$$

$$s = \frac{\frac{\Delta v}{\Delta t}(z_{min} + z_{max})^2/(1 - \frac{\Delta v}{\Delta t}z_{max})}{2(z_{min} - \frac{\Delta v}{\Delta t}z_{max}z_{min} - z_{max} + \frac{\Delta v}{\Delta t}z_{min}^2)}. \quad (11)$$

We visualize the effect of our depth reduction technique in Figure 8. We show an input spectrum (left) that is mapped to display coordinates without depth compression (middle). On the right, the spectrum is compressed such that a specified depth range lies in the depth of field of the display. In Figure 9 we compare simulated views corresponding to Figure 7. We simulate a display with an angular resolution of 8 views spanning a field of view of 25 degrees. The views

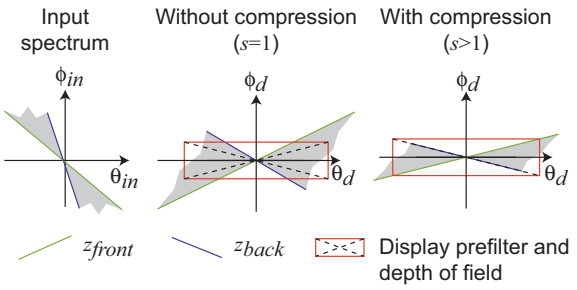

Figure 8: *Visualization of depth compression: input spectrum (left), spectrum mapped to display coordinates without depth compression (middle), spectrum mapped to display coordinates such that a given scene depth range is in focus (right).*

have a spatial resolution of 375 pixels. Without depth compression, the example scene has a maximum disparity of 206 pixels. Since most of the scene lies outside the depth of field of the display, we obtain an extremely blurry image as seen on the left in the figure. Applying our depth compression method to get the whole locomotive into focus, we compute a factor $s = 6.7$ and reduce the maximum disparity to 10 pixels. A resulting simulated view is shown on the right in Figure 9.

Figure 9: *Simulated display views without depth compression (left) and with depth compression (right). We map the locomotive into the depth of field of the display, which reduces the maximum display disparity from* 206 *to* 10 *pixels.*

6. Optimizing Acquisition

The spectra and aliasing of light fields shown on 3D displays depend on a number of *acquisition* and *display* parameters, such as the number of cameras, their spacing, their aperture, the scene depth range, and display resolution. For the practical use of 3D displays it is important that 3D cinematographers do not need to rely on trial and error to determine the

acquisition parameters. In this section, we describe how to derive minimum acquisition sampling rates 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.

We present precise rules to find acquisition parameters for a common scenario: a user wishes to capture a given scene such that certain objects can be kept in focus on the target display. We summarize the input scene and display parameters for two practical examples in Table 1. The scene depth range specifies the minimum and maximum distance from objects in the scene to the camera array. The scene depth of field is the depth range in the scene that should be kept in focus on the display. The display will show the scene as seen through a virtual window of a given width. The placement of this virtual window corresponds to the location of the display plane in the scene, i.e., the f_{in} parameter, which we compute as described below. In addition, the user specifies the spatial and angular resolution, and the field of view of the display. Finally, he chooses the camera aperture. The camera aperture is important because increasing it can reduce reconstruction aliasing [LH96].

Parameter	(a)	(b)
Scene depth range (z_{min}, z_{max})	$(4.5m, 8m)$	
Scene depth of field (z_{front}, z_{back})	$(5m, 5.5m)$	
Virtual window width	$4m$	
Display resolution (spat., ang.)	$(640, 8)$	
Display field of view	$30deg.$	
Camera aperture a	$3.5cm$	$13cm$

Table 1: *Summary of input parameters for determining minimum sampling requirements.*

Given the input parameters in Table 1, we first use Equation 11 to compute the focal distance f_{in} and the baseline scaling s, which determine the mapping from input to display coordinates. We then derive the minimum sampling rate, i.e., the minimum number and resolution of cameras, by finding the tightest packing of replicas of the input spectrum such that none of the non-central replicas overlap with the display prefilter. This is illustrated in Figure 10. Note that for a camera aperture a the replicas of the input spectrum are limited to parallel slabs of width $2\pi/a$ [LH96]. We choose the focal distance of the cameras to coincide with f_{in}, such that the slabs are oriented horizontally in display coordinates. The resulting acquisition parameters are summarized in Table 2. Observe that in the second example, shown in Figure 10b, we almost reduce the number of cameras to the angular resolution of the display. However, achieving this is often impractical because of the large camera apertures it requires.

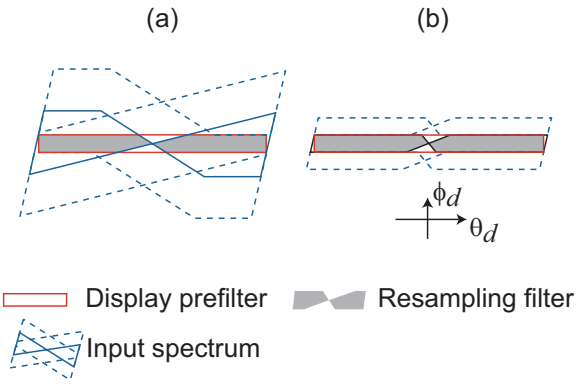

Figure 10: *Illustration of minimum sampling using display coordinates. (a) Minimum sampling of a given scene by packing replicas of the input signal as tightly as possible. (b) Increasing the camera aperture allows for tighter packing and reduces the number of required cameras.*

7. Results

We have implemented a light field resampling algorithm based on Equation 7 and using Gaussian filters [Hec89]. In Figure 11 we compare our approach to resampling without prefiltering and conservative per-view filtering as proposed by Moller and Travis [MT05]. The input data consists of a densely sampled 1D light field such that reconstruction aliasing is avoided. We simulate perspective views of an automultiscopic display with 8 views and a horizontal resolution of 566 pixels. The maximum disparity is 11 pixels, as indicated by the black bars at the top of the images. As shown in Figure 11a, *pre-aliasing* appears as ghosting artifacts if prefiltering is omitted. Conservative filtering removes ghosting, but leads to overly blurry results (Figure 11c). Note how our prefilter preserves spatial frequencies around the zero-disparity plane, which passes through the tusk of the elephant (Figure 11d).

In Figure 12c and Figure 12e, we only used the display prefilter instead of the combined resampling filter. Here the view has a spatial resolution of 302 pixels, and the maximum display disparity in the scene is 10 pixels. Our input

Parameter	(a)	(b)
Camera focal distance f_{in}	$5.24m$	$5.24m$
Camera spacing Δt_{in}	$5.5cm$	$13.1cm$
Camera baseline	$1.05m$	$1.05m$
Number of cameras	20	10
Camera field of view	51.4	51.4
Camera resolution	823	823

Table 2: *Parameters for optimized sampling of a scene as specified by Table 1.*

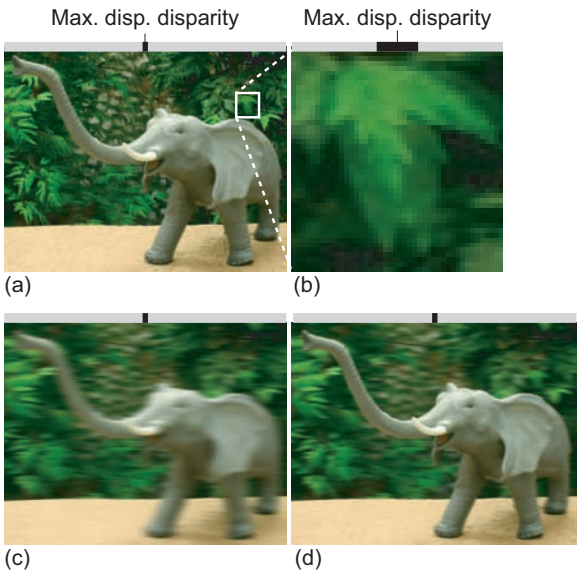

Figure 11: *Comparison of aliasing, conservative filtering of individual views, and linear filtering in ray space. (a) Simulated view with aliasing apparent as ghosting artifacts, shown in close-up (b). (c) Simulated view with conservative filtering, (d) with ray-space filtering. Note how our approach preserves spatial frequencies on the zero-disparity plane.*

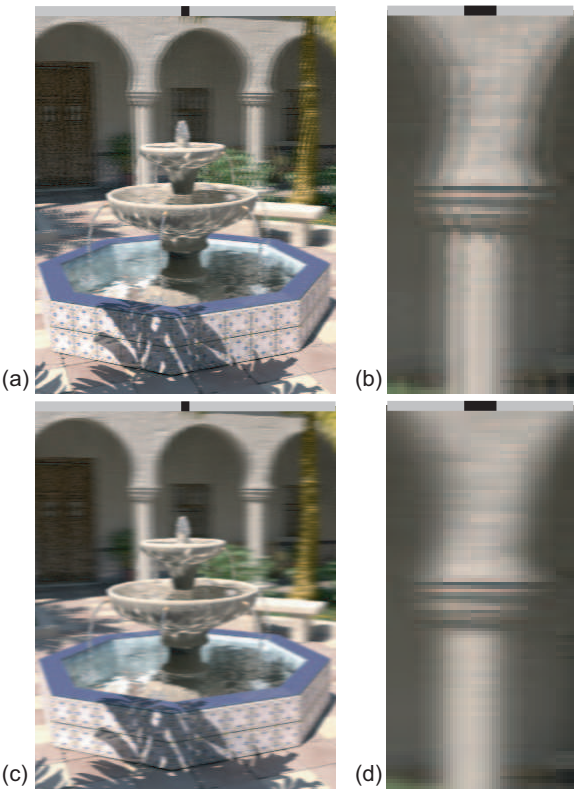

Figure 12: *(a) Simulated view showing* reconstruction aliasing*, (b) close-up of ghosting artifacts; (c) using the combined resampling filter, (d) close-up reveals no ghosting problems. The black bar at the top of the images indicates the maximum display disparity* (see also color plate).

data contained 16 camera views. This leads to *reconstruction aliasing*, since the bandwidth of the display overlaps with some of the non-central replicas of the input signal. Only the combination of reconstruction and prefilter can avoid aliasing in all scenarios.

8. Conclusions and Future Work

We have presented a framework for studying sampling and aliasing for 3D displays. The framework is based on a ray space analysis, which makes our problem amenable to signal processing methods. We derive the bandwidth of 3D displays, quantitatively explain their shallow depth of field behavior, and show that antialiasing is achieved by linear filtering in ray space. We then derive a resampling algorithm that allows us to render high quality scenes acquired at a limited resolution without aliasing on 3D displays. We also alleviate the shallow depth of field of current displays by allowing a user to specify a depth range in the scene that should be mapped to the depth of field of the display. Finally, we derive minimum sampling requirements for high quality display. We believe that these results will benefit better engineering of multi-view acquisition and 3D display devices.

We are currently studying the perceptual impact of our antialiasing scheme, and a preliminary user study led to inconclusive results. While a significant part of the test subjects seemed to prefer antialiased scenes, others liked the aliased versions better, which are sharper in general. We are conducting further experiments to find the perceptually optimal balance between ghosting due to aliasing and blurriness caused by the shallow display depth of field. Our preliminary study included only static scenes, and we will include the effect of moving scenes in future experiments. In addition, we believe that antialiasing has profound implications on light field compression and transmission algorithms. Current compression algorithms compress the original light field data. Prefiltering the data before compressing and sending it to the end user will lead to greater compression efficiency and high quality display.

Acknowledgments

Frédo Durand acknowledges a Microsoft Research New Faculty Fellowship.

References

[CCST00] CHAI J. X., CHAN S. C., SHUM H. Y., TONG X.: Plenoptic sampling. In *Computer Graphics* (Los Angeles, CA, July 2000), SIGGRAPH 2000 Proceedings, pp. 307–318.

[FKdB*02] FEHN C., KAUFF P., DE BEECK M. O., ERNST F., IJSSELSTEIJN W., POLLEFEYS M., GOOL L. V., OFEK E., SEXTON I.: An evolutionary and optimised approach on 3D-TV. In *Proceedings of International Broadcast Conference* (Amsterdam, Netherlands, Sept. 2002), pp. 357–365.

[GGSC] GORTLER S., GRZESZCZUK R., SZELISKI R., COHEN M.: The lumigraph. In *Computer Graphics*.

[Hal94] HALLE M.: Holographic stereograms as discrete imaging systems. In *Practical Holography VIII* (1994), vol. 2176 of *SPIE Proceedings*, pp. 73–84.

[Hec89] HECKBERT P.: *Fundamentals of Texture Mapping and Image Warping*. Ucb/csd 89/516, U.C. Berkeley, June 1989.

[Hil94] HILAIRE P. S.: Modulation transfer function and optimum sampling of holographics stereograms. *Applied Optics 33*, 5 (February 1994).

[Hol04] HOLLIMAN N. S.: Mapping perceived depth to regions of interest in stereoscopic images. In *Proc. SPIE Vol. 5291, Stereoscopic Displays and Virtual Reality Systems XI* (2004).

[HOY00] HOSHINO H., OKANO F., YUYAMA I.: A study on resolution and aliasing for multi-viewpoint image acquisition. *IEEE Transactions on Circuits and Systems for Video Technology 10*, 2 (2000), 366–375.

[IMG00] ISAKSEN A., MCMILLAN L., GORTLER S. J.: Dynamically reparameterized light fields. In *Computer Graphics* (Los Angeles, CA, July 2000), SIGGRAPH 2000 Proceedings.

[JJC*01] JEON H.-I., JUNG N.-H., CHOI J.-S., JUNG Y., HUH Y., KIM J.-S.: Super multiview 3d display system using reflective vibrating scanner array. In *Stereoscopic Displays and Virtual Reality Systems VIII* (june 2001), Proceedings of SPIE, pp. 175–186.

[JLHE01] JONES G. R., LEE D., HOLLIMAN N. S., EZRA D.: Controlling perceived depth in stereoscopic images. In *Proc. SPIE Vol. 4297, Stereoscopic Displays and Virtual Reality Systems VIII* (June 2001), pp. 42–53.

[JO02] JAVIDI B., OKANO F. (Eds.): *Three-Dimensional Television, Video, and Display Technologies*. Springer-Verlag, 2002.

[KA05] KONRAD J., AGNIEL P.: Subsampling models and anti-alias filters for 3-d automultiscopic displays. *IEEE Trans. Image Proc.* (2005).

[LF02] LIPTON L., FELDMAN M.: A new stereoscopic display technology: The synthagram. In *Proc. SPIE Stereoscopic Displays and Virtual Reality Systems* (Jan. 2002), vol. 4660, pp. 229–235.

[LH96] LEVOY M., HANRAHAN P.: Light field rendering. In *Computer Graphics* (New Orleans, LS, Aug. 1996), SIGGRAPH 96 Proceedings, pp. 31–42.

[MP04] MATUSIK W., PFISTER H.: 3d tv: A scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes. *ACM Transaction on Graphics 23*, 3 (Aug. 2004), 811–821.

[MT05] MOLLER C. N., TRAVIS A.: Correcting interperspective aliasing in autostereoscopic displays. *IEEE Transactions on Visualization and Computer Graphics 11*, 2 (March/April 2005), 228–236.

[Ng05] NG R.: Fourier slice photography. *ACM Trans. Graph. 24*, 3 (2005), 735–744.

[Oko76] OKOSHI T.: *Three-Dimensional Imaging Techniques*. Academic Press, 1976.

[Opt05] OPTICALITY:. http://www.opticalitycorporation.com/, 2005. Cited November, 2005.

[PPK00] PERLIN K., PAXIA S., KOLLIN J.: An autostereoscopic display. In *SIGGRAPH 2000 Conference Proceedings* (New Orleans, LS, July 2000), vol. 33, pp. 319–326.

[SG02] SCHMIDT A., GRASNICK A.: Multi-viewpoint autostereoscopic displays from 4d-vision. In *SPIE Stereoscopic Displays and Virtual Reality Systems* (Jan. 2002), vol. 4660, pp. 212–221.

[Sha05] SHARP:. http://www.sharp3d.com/, 2005. Cited November, 2005.

[SMG*05] SANDIN D., MARGOLIS T., GE J., GIRADO J., PETERKA T., DEFANTI T.: The varrier^{TM} autostereoscopic virtual reality display. *ACM Transactions on Graphics 24*, 3 (2005), 894–903.

[SYGM03] STEWART J., YU J., GORTLER S., MCMILLAN L.: A new reconstruction filter for undersampled light fields. In *Eurographics Symposium on Rendering* (Leuven, Belgium, 2003), ACM International Conference Proceeding Series, pp. 150–156.

[WHR99] WARTELL Z., HODGES L. F., RIBARSKY W.: Balancing fusion, image depth and distortion in stereoscopic head-tracked displays. In *SIGGRAPH '99: Proceedings of the 26th annual conference on Computer graphics and interactive techniques* (1999), pp. 351–358.

[WSLH02] WILBURN B., SMULSKI M., LEE H. K., HOROWITZ M.: The light field video camera. In *Media Processors 2002* (Jan. 2002), vol. 4674 of *SPIE*, pp. 29–36.

[Yah05] YAHOO: A world first at the ifa: Live 3d tv without glasses. http://biz.yahoo.com/prnews/050901/ukth016.html?.v=13, 2005. Cited November, 2005.