Joint View Expansion and Filtering for Automultiscopic 3D Displays

Piotr Didyk Pitchaya Sitthi-Amorn William Freeman Frédo Durand Wojciech Matusik

MIT CSAIL

Figure 1: *The method presented in this paper takes a stream of stereo images as an input and synthesizes additional views required for an automultiscopic display. The output views are also filtered to remove inter-view aliasing. ("Big Buck Bunny" © by Blender Foundation)*

Abstract

Multi-view autostereoscopic displays provide an immersive, glasses-free 3D viewing experience, but they require correctly filtered content from multiple viewpoints. This, however, cannot be easily obtained with current stereoscopic production pipelines. We provide a practical solution that takes a stereoscopic video as an input and converts it to multi-view and filtered video streams that can be used to drive multi-view autostereoscopic displays. The method combines a phase-based video magnification and an interperspective antialiasing into a single filtering process. The whole algorithm is simple and can be efficiently implemented on current GPUs to yield a near real-time performance. Furthermore, the ability to retarget disparity is naturally supported. Our method is robust and works well for challenging video scenes with defocus blur, motion blur, transparent materials, and specularities. We show that our results are superior when compared to the state-of-the-art depth-based rendering methods. Finally, we showcase the method in the context of a real-time 3D videoconferencing system that requires only two cameras.

CR Categories: I.3.3 [Computer Graphics]: Picture/Image generation—display algorithms,viewing algorithms;

Keywords: automultiscopic displays, view synthesis, interperspective antialiasing

Links: DL PDF WEB VIDEO

1 Introduction

Stereoscopic 3D content is becoming more popular as it reaches an increasing number of home users. While most of current TV sets are 3D-enabled, and there is plenty of 3D movies and sports programming available, the adoption of stereoscopic 3D is hampered by the use of 3D glasses required to view the content. Multi-view autostereoscopic (or automultiscopic) displays offer a superior visual experience since they provide both binocular and motion parallax without the use of special glasses. A viewer is not restricted to be in a particular position and many viewers can watch the display at the same time. Furthermore, automultiscopic displays can be manufactured inexpensively, for example, by adding a parallax barrier or a lenticular screen to a standard display.

However, there are three major problems that need to be addressed in order for a multi-view autostereoscopic TV to become a reality. First, current 3D content production pipelines provide only two views, while multi-view stereoscopic displays require images from many viewpoints. Capturing TV-quality scenes with dense camera rigs is impractical because of the size and cost of professional quality cameras. A solution to use view-interpolation to generate these additional views requires an accurate depth and inpainting of missing scene regions. There has been a steady progress in stereo depth reconstruction algorithms, but the quality is not yet good enough for TV broadcast and movies. Handling scenes that include defocus blur, motion blur, transparent materials, and specularities is especially difficult. Second, multi-view autostereoscopic displays require special filtering to remove interperspective aliasing – all image content that is not supported by a given display [Zwicker et al. 2006]. Without performing this step severe ghosting and flickering can be seen. However, in order to properly antialias a multi-view video, a dense light field is necessary. Finally, to assure viewing comfort, image disparities usually have to be modified according to the display type, size, and viewer preference. This disparity retargeting step also requires rerendering the scene with the adjusted disparities.

We propose a method that addresses all these three limitations. Our method takes a stereoscopic stream as an input and produces a correctly filtered multi-view video for a given automultiscopic display as shown in Figure 1. The solution does not require any changes to the current stereoscopic production and content delivery pipelines.

All additional processing can be done by the client (e.g., at home). The proposed method is simple and it can be implemented in hardware. Our current implementation on GPU in CUDA achieves a near real-time performance. The key to our solution is a steerable pyramid decomposition and filtering that has been recently successfully used for motion magnification in video sequences [Wadhwa et al. 2013]. We show how similar concepts can be used for view interpolation and how the antialiasing filter and disparity remapping can be incorporated with almost no additional cost. We demonstrate the results on a variety of different scenes including defocus blur, motion blur, and complex appearance. We compare our results to both the ground truth and depth-based rendering. Finally, we demonstrate our method on a real-time 3D video conferencing system that requires only two video cameras and provides multi-view autostereoscopic experience.

To summarize, our contributions are an efficient algorithm for joint view expansion, filtering and disparity remapping for multi-view autosteresocopic displays as well as an evaluation of the method on many different scenes along with a comparison to both the ground truth and the state-off-the-art depth-based rendering techniques.

2 Previous Work

An automultiscopic display reproduces multiple views corresponding to different viewing angles. This allows for glasses-free 3D and more immersive experience. In order to achieve this, all the views need to be provided to the display. A standard technique to acquire multiple images from different locations is to use a camera array. Such systems usually consist of calibrated and synchronized sensors, which record the scene from different locations. The number of cameras can range from a dozen [Matusik and Pfister 2004] to over a hundred [Wilburn et al. 2001]. However, such setups are usually impractical [Farre et al. 2011] and too expensive for commercial use. Instead, it is possible to use image-based techniques to generate missing views. Most of these techniques need to recover depth information first, and then a view synthesis method is used for computing additional views [Smolic et al. 2008]. Although there is a number of techniques that try to recover depth information from stereo views [Brown et al. 2003], this is an ill-posed problem. Most of existing methods is prone to artifacts and temporal inconsistency. The quality of estimated depth maps can be improved in a post-processing step [Richardt et al. 2012]. This, however, is usually a time consuming process. Instead of recovering dense map correspondence, it is possible to recover only sparse depth maps and use a warping technique to compute new views [Farre et al. 2011]. Such methods can produce good result but at expense of computational time which prevents real-time solutions.

Recently, there has been a significant development in display designs [Holliman et al. 2011]. Commercial automultiscopic displays are usually based on either parallax barriers or lenticular sheets. Both, placed atop a high resolution panel, trade spatial resolution for angular resolution, and produce multiple images encoded as one image on the panel [Lipton and Feldman 2002; Schmidt and Grasnick 2002]. Multi-view projector systems have been also proposed [Matusik and Pfister 2004; Balogh 2006]. Even more recently, there has been many attempts of building a display which would reproduce the entire light field. One such example is a display with 256 views proposed by Takaki et al. [2010]. Also so-called compressive and multi-layer displays try to achive this goal by introducing more sophisticated hardware solutions [Akeley et al. 2004; Wetzstein et al. 2012]. This trend makes multi-view autostereoscopic display a promising solution for the future.

Automultiscopic screens usually produce a light field, which is a continuous 4D function representing radiance with respect to a position and a viewing direction [Levoy and Hanrahan 1996]. Due to discrete nature of an acquisition (i. e., limited number of views), the recorded light field is usually aliased. Chai et al. [2000] as well as Isaksen et al. [2000] presented a plenoptic sampling theory which analyses the spectrum of reconstructed light field. Based on this, there has been presented a number of techniques that allow for antialiasing of the recorded light field [Isaksen et al. 2000; Stewart et al. 2003]. In the context of automultiscopic display, the aliasing is not only due to undersampling of the light field but also because of the limited bandwidth of the display. Zwicker et al. [2006] took both sources of the aliasing into account and presented a combined antialiasing framework which filters input views coming from a camera array. However, the large number of views required for their technique makes the solution impractical in a standard scenario when only 3D stereo content (two views) is available.

A sequence of images required for automultiscopic display usually corresponds to a set of views captured from different locations. Such a sequence can be captured by a camera moving horizontally on a straight line. In this context, a problem of creating additional views is similar to a motion editing problem when the only motion in the scene comes from the camera movement. Recently, a number of techniques have been presented which seek to magnify invisible motions. For example, in the Lagrangian approach [Liu et al. 2005], the motion is explicitly estimated and then magnified. Later, an image based technique is used to compute frames that correspond to modified flow. Wu et al. [2012] proposed an Eulerian approach which eliminates the need of flow computation. Instead, it processes the video in space and time to amplify the temporal color changes. More recently, Wadhwa et al. [2013] proposed a phase-based technique. This method benefits from the observation that in many cases motion is encoded in a complex-valued steerable pyramid decomposition as coefficients variation. Compared to previous techniques, this method does not require motion computation and can handle much bigger displacements then the Eulerian approach. Our technique is inspired by these methods. Instead of estimating correspondence (depth) between two stereo views, we assume that it is encoded in the phase shift once the left and right views are decomposed into complex-valued steerable pyramids.

3 View Expansion

In this section, we describe our approach for view expansion. The goal of this method is to take as an input a standard 3D stereo video stream (i. e., left and right view), and create additional views that can be later used on an automultiscopic display. Our method is inspired by the phase-based motion magnification technique. Therefore, we first give a short overview of this method, and then explain how it can be adapted to create additional views for automultiscopic display.

3.1 Phase-based Motion Magnification

The phase-based motion magnification exploits the steerable pyramid decomposition [Simoncelli et al. 1992; Simoncelli and Freeman 1995], which decomposes images according to the spatial scale and orientation. Assuming that the signal is a sine wave, a small motion is encoded in the phase shift between frames. Therefore, the motion can be magnified by modifying the temporal changes of the phase.

In order to compute the steerable pyramid a series of filters $\Psi_{\omega,\Theta}$ is used. These filters correspond to one filter, which is scaled and rotated according to the scale ω and the orientation Θ. The steerable pyramid is then built by applying these filters to the discrete Fourier transform (DFT) $\tilde{\mathrm{I}}$ of each image I from the video sequence. This way each frame is decomposed into a number of frequency bands $\mathcal{S}_{\omega,\Theta}$ which have DFT $\tilde{\mathcal{S}}_{\omega,\Theta} = \tilde{\mathrm{I}}\Psi_{\omega,\Theta}$. A great advantage of such a

decomposition is that the response of each filter is localized, which enables processing of phases locally.

To give an intuition how the phase-based motion magnification works, let us consider first a 1D case, i. e., 1D intensity profile f translating over time with a constant velocity. If the displacement is given by a function $\delta(t)$, the image changes over time according to $f(x+\delta(t))$. This function can be expressed in the Fourier domain as a sum of complex sinusoids:

$$f(x+\delta(t)) = \sum_{\omega=-\infty}^{\infty} A_\omega e^{i\omega(x+\delta(t))}, \tag{1}$$

where ω is a single frequency and A is amplitude of the sinusoid. From this, a band corresponding to the frequency ω is given by:

$$\mathcal{S}_\omega(x,t) = A_\omega e^{i\omega(x+\delta(t))}. \tag{2}$$

The $\omega(x+\delta(t))$ is the phase of the sinusoid, and it contains the motion information which can be directly amplified. However, changing individual phases does not lead to meaningful motion editing, as the motion is rather encoded in the relative changes of the phase over time. To amplify motion, first, the phase is filtered in the temporal direction to isolate desired phase changes $\mathrm{B}_\omega(x,t)$. Next, it is multiplied by a magnification factor α, and the original phase in band $\mathcal{S}_{\omega,\Theta}$ is increased by the amplified signal $\mathrm{B}_\omega(x,t)$. If we assume that the filtering applied to the phase removes only DC component, the new modified sub-band with amplified motion is:

$$\hat{\mathcal{S}}_\omega(x,y) = \hat{\mathcal{S}}_\omega(x,y)e^{i\alpha \mathrm{B}_\omega(x,t)} = A_\omega e^{i\omega(x+(1+\alpha)\delta(t))}. \tag{3}$$

The method generalizes to the 2D case, where the steerable pyramid decomposition uses filters with a finite spatial support. This enables detecting and amplifying local motions. For more details please refer to the original paper [Wadhwa et al. 2013].

3.2 Our Approach

In order to expand 3D stereo content to a multiview video stream (Figure 2), we make the following observation. Similarly to motion magnification, where the motion information is mostly encoded in the phase change, the parallax between two neighboring views is encoded in the phase difference. As we deal with two frames only (left and right), instead of analyzing the phase changes in the temporal domain, we need to account for the phase differences in corresponding bands between two input views. Because there is no notion of time, we denote the phase shift as δ instead of $\delta(t)$ in the rest of the paper.

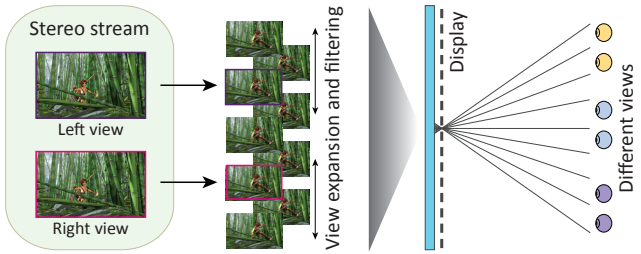

Figure 2: *Our method takes a 3D stereo stream as an input, and performs a view expansion together with an antialiasing filtering to obtain a correct input for an automultiscopic display. ("Sintel" © by Blender Foundation)*

In order to create the additional views we take the two stereo frames L and R, and perform the steerable pyramid decomposition on both. Then, we compute the phase difference for each complex coefficient. After modifying the phase differences according to the α value and collapsing the pyramids, two nearby views are created. The result is a stereo disparity expansion without a need of dense depth map reconstruction, which can be prone to significant artifacts.

The process can be defined as follows:

$$(L', R') = M(L, R, \alpha),$$

where M is the view generation process, and L' and R' are the nearby views according to the magnification factor α. The magnification factors are computed based on virtual camera positions that the images correspond to. If we assume that the input images coincide with locations $-x_0$ and x_0, for the left and the right view, respectively, the magnification factor for an arbitrary location x needs to be set to $\alpha = (|x| - x_0)/(2x_0)$. Because a new image is always reconstructed from the input view which is closest to the new location, the same α value is used for location x and $-x$. The process of choosing correct magnification factors is shown in Figure 3. For all our results, we use view expansion only in outward direction illustrated as the first example.

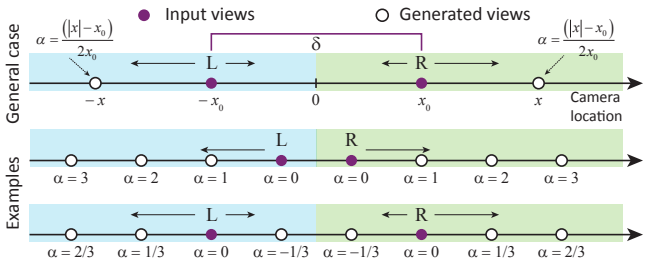

Figure 3: *The magnification factor α needs to be adjusted according to the position of the virtual camera for which the view is generated. Our method can synthesize new views in outward direction but also interpolate inbetween views. New views are reconstructed from the input image corresponding to the closest location, i. e., the left image is used to reconstruct images corresponding to locations in blue regions whereas the right one is used for green regions.*

4 Antialiasing for Automultiscopic Display

In this section, we show how our method for new views generation can be extended so that it produces images without interperspective aliasing. This requires the views to be filtered according to the local depth. The process is similar to adding a depth-of-field effect. A naïve and costly way to filter a single view is to generate a number of neighboring views and average them using weights corresponding to the distance from the original view. In contrast, a key advantage of our approach is that we can perform the filtering directly on the steerable pyramid decomposition. We derive a closed form solution that can be performed at almost no additional cost.

4.1 Filtering Equation

For the simplicity, let us assume that instead of the previously defined function M we have two: M_R and M_L. These return only one of the views, i. e., R' or L' respectively. The process of antialising is analogous for both views, and we will describe the case of R' view.

In order to be filtered, R' needs to be averaged with its neighboring views according to the weights given by a low pass filter along the viewpoint dimension. Let us assume that the filter is given as a function $\mathcal{F}$. The antaliased view $\hat{R}'$ corresponding to fixed α value,

Figure 4: *Comparison of different content creation approaches for automultiscopic display. When each frame is rendered, but antialiasing is not applied a significant ghosting is visible for objects located further from the screen plane (green insets). These artifacts can be removed when the content is properly filtered but this requires rendering hundreds of views. Image-based techniques combined with filtering can produce good results for many cases, but also can introduce significant artifacts when depth estimation fails (red insets). Our method produces results similar to rendering with filtering, but at cost similar to real-time image-based techniques. ("Big Buck Bunny" © by Blender Foundation)*

can be computed as follows:

$$\hat{R}' = \int \mathcal{F}(\beta - \alpha) M_R(L, R, \beta) d\beta.$$

In order to perform the filtering directly on the pyramid decomposition, the above integration can be approximated before the reconstruction of the pyramid for each sub-band of R' separately. Let us consider one band $\hat{\mathcal{S}}_\omega(x, y, \alpha)$ of the decomposition of R'. The corresponding filtered sub-band can be computed as:

$$\tilde{\mathcal{S}}_\omega(x, y, \alpha) = \int \mathcal{F}(\beta - \alpha) \cdot \hat{\mathcal{S}}_\omega(x, y)\, d\beta,$$

which can be further transformed:

$$\begin{aligned}\tilde{\mathcal{S}}_\omega(x, y, \alpha) &= \int \mathcal{F}(\beta - \alpha) \cdot A_\omega e^{i\omega(x + (1+\beta)\delta)}\, d\beta \\ &= A_\omega e^{i\omega(x+\delta(t))} \int \mathcal{F}(\beta - \alpha) \cdot e^{i\omega\beta\delta}\, d\beta \\ &= \mathcal{S}_\omega(x, y) \int \mathcal{F}(\beta - \alpha) \cdot e^{i\omega\beta\delta}\, d\beta.\end{aligned}$$

The final filtered sub-band consists of two components. The first one, $\mathcal{S}_\omega(x, y)$, is a sub-band of original view R. The second component is the integral component, which depends only on δ. This is very convenient because in most cases it has a closed form solution, or it can be precomputed and stored as a lookup table parametrized by δ.

In our implementation we chose $\mathcal{F}$ to be a Gaussian filter:

$$\mathcal{F}_\sigma(\beta) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{\beta^2}{2\sigma^2}},$$

which results in each sub-band of view R' being:

$$\tilde{\mathcal{S}}_\omega(x, y, \alpha) = \frac{\sigma}{2} \cdot e^{i\alpha\delta - \sigma^2\delta^2/2} \cdot \mathcal{S}_\omega(x, y).$$

The above equation assumes a good estimation of the phase shift δ. In practice, the phase-based approach [Wadhwa et al. 2013] may underestimate it, which leads to insufficient filtering. This happens when the assumption that the correspondence between two views is encoded in the phase difference fails. In that case, we propose to correct the phase shift in each sub-band separately, based on the phase shift in the corresponding sub-band for the lower frequency. To this end, before applying the factor responsible for the filtering, we process the entire pyramid starting from the lowest frequency level. Whenever the phase shift on the level below is greater then $\pi/2$, the phase shift at the current level may be underestimated. In such a case, we correct the phase shift by setting its value to twice the phase shift on the lower level. This provides a correct phase shift estimation under the assumption that the correspondence between the input views behaves locally as a translation. The correct phase shift estimation is not crucial for the motion magnification nor for the nearby view synthesis. However, it is important for the correct filtering.

5 Results

In this section, we provide an extensive set of results to evaluate our method. First, we include implementation details and standard running times. Second, we provide a detailed comparison with a state-oft-the-art depth image-based rendering technique (DIBR). We also present our real-time 3D video conferencing system to showcase the robustness and efficiency. Then, we show an application of our method to depth remapping. Finally, the limitations are discussed.

Implementation details We have implemented our method on a GPU using CUDA API, and processed all sequences using NVIDIA GTX Titan graphics card on an Intel Xeon machine. We have used eight orientations in the steerable pyramid, which gives us a good trade-off between quality and performance. The time required for building a pyramid and reconstructing one additional view is independent of the image content, and it is 15 ms and 12 ms for building and reconstructing respectively, assuming a content with 816×512 resolution. This enables reconstruction of eight views for a standard automultiscopic display at 8.3 FPS rate. The memory requirement for our method is relatively low. Each pyramid requires 137 MB of memory. Hence, to process an input stereo sequence, 3×137 MB of memory is required – 2×137 MB for two input views and 137 MB for the synthesized view.

Comparison to depth-based techniques We are not aware of any real-time method that directly computes properly filtered content for automultiscopic 3D displays based on a stereoscopic video stream. However, in order to evaluate our method we compared our technique to a combination of depth-based rendering and antialiasing. Our hypothetical competitive method takes a stereoscopic video stream as an input, and reconstructs a depth map for each image pair. Then, it applies a real-time warping technique for additional views synthesis. To obtain one antialiased view the method averages 30

neighboring views according to Gaussian weights similar to these that are used in Section 4. For estimating depth we used a recent technique proposed by Hosni et al. [2013]. The view sythesis is similar to the approach presented by Didyk et al. [2010]. Combination of these two techniques provides a good trade-off between quality and performance.

We compare this depth-based rendering with our method on three different examples. Two of them are computer generated animations (Figures 4 and 5). The third example is a photograph taken using a 3D camera (LG Olympus P725 camera) (Figure 6). This example is challenging for both techniques as the captured scene contains both reflections and transparent objects. In addition, for the sequence from Figure 4 we computed a dense light field (100 of views). This allowed us to use the antaliasing technique proposed by Zwicker et al. [2006]. We refer to this as to the ground truth method.

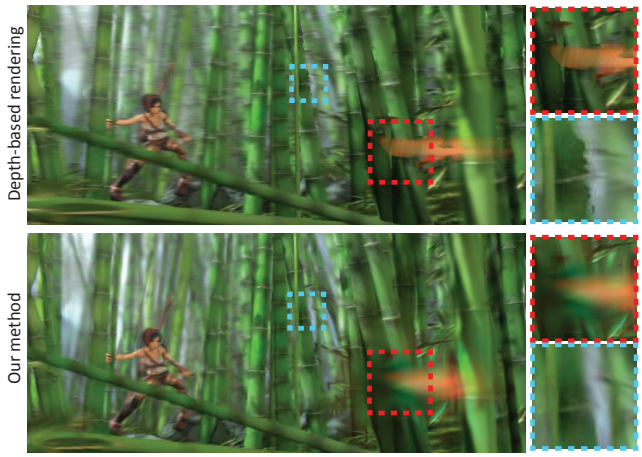

Figure 5: *The comparison between our method and depth-based rendering for one of the synthesized views. Please note the artifacts due to the poor depth estimation for depth-based rendering. Blue inset shows how incorrect depth estimation results in jaggy depth discontinuities. In red insets, depth estimation fails in reconstructing depth of the out-of-focus butterfly. In both cases our method produces more correct results. ("Sintel" © by Blender Foundation)*

In all cases our method produces more graceful degradation of the image quality comparing to the DIBR method. It is important to note, that artifacts produced by the depth-based technique are mostly due to poor depth estimation and not due to incorrect view-synthesis. Depth estimation is an ill-posed problem, and such methods cannot handle regions with non-obvious per-pixel depth values (e. g., transparencies, reflections, motion blur, defocus blur, and thin structures that have partial coverage) as shown in Figures 5 and 6. Real-time depth estimation methods also have problems with temporal coherence. In contrast, our method does not produce visible and disturbing artifacts even though the coherence is not explicitly enforced. This can be observed in the accompanying video. Additionally, Figure 7 visualizes errors of our and the depth-based techniques when compared to the ground truth using the SSIM metric [Wang et al. 2004]. The error produced by the latter is localized mostly around depth discontinuities. Our method provides error which is distributed more uniformly across the image, and therefore less disturbing. It is important to mention that the error of our technique is significantly influenced by the different types of blur introduced by the compared methods. While the ground-truth and the depth-based techniques filter images only in the horizontal direction, our method results in a more uniform blur. This can be observed in Figure 4 (green inset).

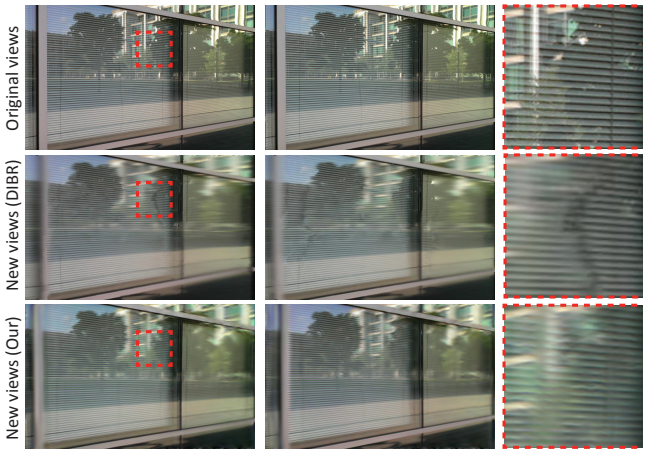

Figure 6: *Transparent and highly reflective objects are very challenging for any depth estimation and view synthesis methods. The figure shows the input images (top) and views that were generated using depth image-based technique (middle) and our method (bottom). The technique that relies on the depth estimation fails to reconstruct highly reflective and transparent objects.*

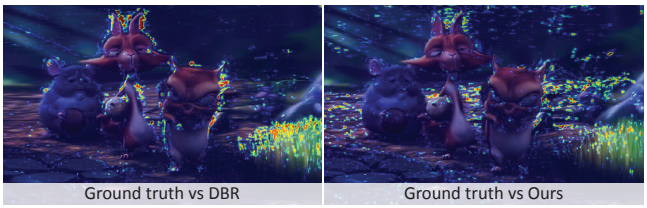

Figure 7: *A colormap visualizing errors between depth-based rendering and ground truth (left) as well as our method and ground truth (right) for the example from Figure 4. The difference is computed using the SSIM metric. The error of the depth-based technique is localized mostly around depth discontinuities. In contrast, the error introduced by our method is distributed more uniformly across the entire image. ("Big Buck Bunny" © by Blender Foundation)*

We attribute the better results produced by our method to the overcomplete representation that we use in our method. While depth-based approaches estimate only one depth value per pixel, which can lead to artifacts in complex cases where no such single value exists, our technique captures the correspondence between views using phase differences for multiple spatial frequencies and orientations separately. As a result the local depth is not represented as one value but many, which can lead to better performance where the depth is not well-defined.

Standard 3D Stereo Content To demonstrate the robustness of our method, we have tested it on many sequences downloaded from the internet. These sequences often have severe compression artifacts, vertical misalignment, and visible color differences between cameras. We have used our approach to expand a stereoscopic video stream to a multiview stream and to display it on an 8-view automultiscopic screen. The method works very well with most of these sequences. We show two of them in the accompanying video.

3D Video Conferencing System Based on our fast view expansion technique, we have built a light-weight, real-time 3D video conferencing system. It consists of eight cameras mounted on a linear ring as well as an automultiscopic display. The system operates

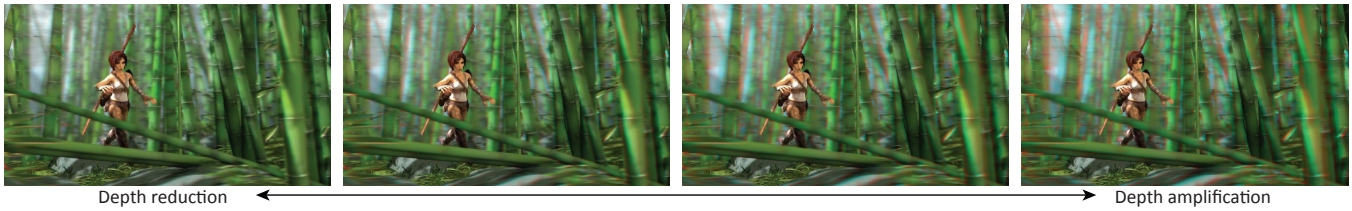

Figure 8: *Our method can also support disparity manipulations. We show stereo images in anaglyph version (red channel for the left eye and cyan for the right one) for the same scene with different depth ranges. ("Sintel" © by Blender Foundation)*

in two modes: it either uses all cameras to acquire 8 views, or it uses only 2 of them and computes the additional six views using our method. In both cases, the 8 views are streamed in real-time to the screen, providing an interactive feedback for the users. See the supplementary video for the comparison between views captured using all cameras and those generated using our technique. Note that the views rendered by our method are filtered to avoid aliasing and this does not add any additional cost to the processing. In contrast, original views captured by eight cameras contain aliasing. This could be removed using the method presented by Zwicker et al. [2006] with the aid of depth image-based rendering. However, it would be prohibitively expensive for a real-time system. We showcase our video conferencing system in the accompanying video.

Disparity Manipulations Our method can also be used for remapping disparities in stereoscopic images and videos. Such modifications are often desired and necessary in order to adjust disparity range in the scene to a given comfort range [Lambooij et al. 2009], viewer preferences or for an artistic purpose [Lang et al. 2010]. For example, NVIDIA 3D Vision allows users to change depth range using a simple knob. Also, methods that target directly automultiscopic displays exist [Didyk et al. 2012]. Using our method, disparity range in the image can be changed by adjusting α value in our view expansion (Section 3). The result is a global scaling of disparities. The example of such manipulations is presented in Figure 8.

Limitations The phase-based approach is limited to processing video that exhibits small displacements [Wadhwa et al. 2013]. For larger displacements the locality assumption of the motion does not hold. Therefore, for larger displacements, only lower spatial frequencies can be correctly reconstructed. In the context of view synthesis for multi-view autostereoscopic displays, this limitation is largely alleviated due to the need of the interperspective antialiasing. While the view synthesis cannot correctly reconstruct high frequencies for scene elements with large disparity, these frequencies need to be removed anyway since they usually lie outside of the display bandwidth and lead to aliasing artifacts. For extreme cases, where either magnification factors or the interaxial between input images are large, some artifacts can remain visible. Therefore, our technique is not a universal substitute for large camera arrays. However, in such cases the method can reduce the number of required cameras significantly. Figure 9 visualizes a case where α values are drastically increased.

6 Conclusions

We presented a novel method which, for the first time, combines view synthesis and antialiasing for automultiscopic display. In contrast to prior solutions, our algorithm does not perform explicit depth estimation and alleviates this source of artifacts. Instead, we leverage the link between parallax and the local phase of Gabor-like wavelets, in practice complex-valued steerable pyramids. This allows us to exploit the translation-shift theorem and simply extrapolate the phase difference measured in the two input views. Importantly, the pyramid representation allows us to integrate antialiasing directly and avoid expensive numerical prefiltering. We derive a closed-form approximation to the prefiltering integral that results in a simple attenuation of coefficients based on the band and phase difference. The simplicity of our method is key because it enables an interactive implementation and makes it behave robustly even for difficult cases. It is also guaranteed to avoid artifacts at the focal plane because the measured phase difference is zero. For displays that do not reproduce only horizontal parallax but also vertical, our method can be extended to generate small light fields. In Figure 10, we created additional views in the horizontal as well as the vertical direction using four input images.

Acknowledgments

We would like to thank Neal Wadhwa and Michael Rubinstein for providing their code, Krzysztof Templin for proofreading, Moira Forberg and Javier Ramos for their help with the video, Eric Deren / Dzignlight Studios for providing the content for our video as well as all anonymous reviewers for their helpful comments and guidance. This work was partially supported by NSF IIS-1111415, NSF IIS-1116296 and Quanta Computer.

References

AKELEY, K., WATT, S. J., GIRSHICK, A. R., AND BANKS, M. S. 2004. A stereo display prototype with multiple focal distances. *ACM Trans. Graph. 23*, 3, 804–813.

BALOGH, T. 2006. The holovizio system. In *Electronic Imaging 2006*, 60550U–60550U.

BROWN, M. Z., BURSCHKA, D., AND HAGER, G. D. 2003. Advances in computational stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence 25*, 8, 993–1008.

CHAI, J.-X., TONG, X., CHAN, S.-C., AND SHUM, H.-Y. 2000. Plenoptic sampling. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, ACM Press/Addison-Wesley Publishing Co., 307–318.

DIDYK, P., RITSCHEL, T., EISEMANN, E., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2010. Adaptive image-based stereo view synthesis. In *Proc. VMV*.

DIDYK, P., RITSCHEL, T., EISEMANN, E., MYSZKOWSKI, K., SEIDEL, H.-P., AND MATUSIK, W. 2012. A luminance-contrast-aware disparity model and applications. *ACM Trans. Graph. (Proc. SIGGRAPH Asia) 31*, 6, 184:1–184:10.

FARRE, M., WANG, O., LANG, M., STEFANOSKI, N., HORNUNG, A., AND SMOLIC, A. 2011. Automatic content creation for multiview autostereoscopic displays using image domain warping. In *IEEE International Conference on Multimedia and Expo*.

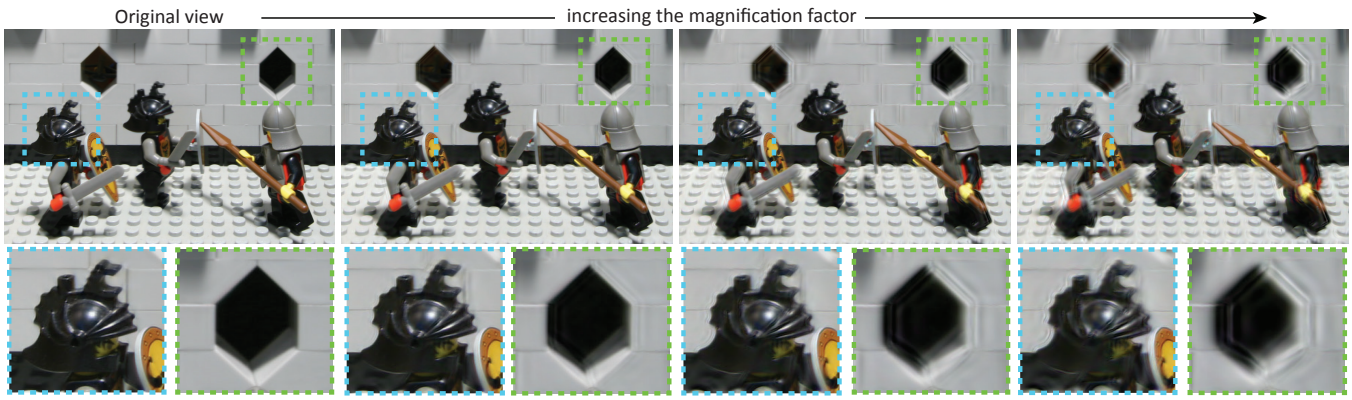

Figure 9: *The figure shows how very large magnification factors (increasing from left to right) affect the final quality of results. For the visualization purpose the inter-view antialiasing was reduced to make the artifacts more visible. The input images come from "The Stanford Light Field Archive" (*`http://lightfield.stanford.edu/lfs.html`*).*

HOLLIMAN, N. S., DODGSON, N. A., FAVALORA, G. E., AND POCKETT, L. 2011. Three-dimensional displays: a review and applications analysis. *IEEE Transactions on Broadcasting 57*, 2, 362–371.

HOSNI, A., RHEMANN, C., BLEYER, M., ROTHER, C., AND GELAUTZ, M. 2013. Fast cost-volume filtering for visual correspondence and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence 35*, 2, 504 – 511.

ISAKSEN, A., MCMILLAN, L., AND GORTLER, S. J. 2000. Dynamically reparameterized light fields. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, ACM Press/Addison-Wesley Publishing Co., 297–306.

LAMBOOIJ, M., IJSSELSTEIJN, W., FORTUIN, M., AND HEYNDERICKX, I. 2009. Visual discomfort and visual fatigue of stereoscopic displays: a review. *J Imaging Science and Technology 53*, 030201–14.

LANG, M., HORNUNG, A., WANG, O., POULAKOS, S., SMOLIC, A., AND GROSS, M. 2010. Nonlinear disparity mapping for stereoscopic 3D. *ACM Trans. Graph. 29*, 4, 75:1–75:10.

LEVOY, M., AND HANRAHAN, P. 1996. Light field rendering. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, ACM, 31–42.

LIPTON, L., AND FELDMAN, M. H. 2002. New autostereoscopic display technology: the synthagram. In *Electronic Imaging 2002*, International Society for Optics and Photonics, 229–235.

LIU, C., TORRALBA, A., FREEMAN, W. T., DURAND, F., AND ADELSON, E. H. 2005. Motion magnification. *ACM Trans. Graph. 24*, 3, 519–526.

MATUSIK, W., AND PFISTER, H. 2004. 3d tv: a scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes. *ACM Trans. Graph. 23*, 3, 814–824.

RICHARDT, C., STOLL, C., DODGSON, N. A., SEIDEL, H.-P., AND THEOBALT, C. 2012. Coherent spatiotemporal filtering, upsampling and rendering of RGBZ videos. *Computer Graphics Forum (Proc. Eurographics) 31*, 2, 247–256.

SCHMIDT, A., AND GRASNICK, A. 2002. Multiviewpoint autostereoscopic dispays from 4d-vision gmbh. In *Electronic Imaging 2002*, 212–221.

SIMONCELLI, E. P., AND FREEMAN, W. T. 1995. The steerable pyramid: A flexible architecture for multi-scale derivative computation. In *IEEE International Conference on Image Processing*, vol. 3, 444–447.

SIMONCELLI, E. P., FREEMAN, W. T., ADELSON, E. H., AND HEEGER, D. J. 1992. Shiftable multiscale transforms. *IEEE Transactions on Information Theory 38*, 2, 587–607.

SMOLIC, A., MULLER, K., DIX, K., MERKLE, P., KAUFF, P., AND WIEGAND, T. 2008. Intermediate view interpolation based on multiview video plus depth for advanced 3d video systems. In *IEEE International Conference on Image Processing*, 2448–2451.

STEWART, J., YU, J., GORTLER, S. J., AND MCMILLAN, L. 2003. A new reconstruction filter for undersampled light fields. In *Proceedings of the 14th Eurographics workshop on Rendering*, Eurographics Association, 150–156.

TAKAKI, Y., AND NAGO, N. 2010. Multi-projection of lenticular displays to construct a 256-view super multi-view display. *Optics Express 18*, 9, 8824–8835.

WADHWA, N., RUBINSTEIN, M., GUTTAG, J., DURAND, F., AND FREEMAN, W. T. 2013. Phase-based video motion processing. *ACM Trans. Graph. (Proc. SIGGRAPH) 32*, 4, 80:1–80:10.

WANG, Z., BOVIK, A. C., SHEIKH, H. R., AND SIMONCELLI, E. P. 2004. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing 13*, 4, 600–612.

WETZSTEIN, G., LANMAN, D., HIRSCH, M., AND RASKAR, R. 2012. Tensor Displays: Compressive Light Field Synthesis using Multilayer Displays with Directional Backlighting. *ACM Trans. Graph. (Proc. SIGGRAPH) 31*, 4, 1–11.

WILBURN, B. S., SMULSKI, M., LEE, H.-H. K., AND HOROWITZ, M. A. 2001. Light field video camera. In *Electronic Imaging 2002*, International Society for Optics and Photonics, 29–36.

WU, H.-Y., RUBINSTEIN, M., SHIH, E., GUTTAG, J., DURAND, F., AND FREEMAN, W. T. 2012. Eulerian video magnification for revealing subtle changes in the world. *ACM Trans. Graph. (Proc. SIGGRAPH) 31*, 4, 65:1–65:8.

ZWICKER, M., MATUSIK, W., DURAND, F., AND PFISTER, H. 2006. Antialiasing for automultiscopic 3d displays. In *Proceedings of the 17th Eurographics conference on Rendering Techniques*, Eurographics Association, 73–82.

Figure 10: *The top image array corresponds to a small light field created from four images marked in green. The small insets bellow present magnified fragments of the reconstructed images. Please refer to the electronic version of the paper to see the images in a zoomed-in version. The input images come from "The Stanford Light Field Archive" (*`http://lightfield.stanford.edu/lfs.html`*).*