EFFICIENT IMAGE RESAMPLING FOR MULTIVIEW DISPLAYS

Michael Schaffner[1,2]*, Pierre Greisen*[1,2]*, Simon Heinzle*[2]*, Aljoscha Smolic*[2]

[1]ETH Zurich, [2]Disney Research Zurich

ABSTRACT

We present an evaluation of different resampling strategies for autostereoscopic multiview displays. In particular, we compare the computational efficiency, memory requirements, and image quality of different resampling algorithms with focus on real-time architectures. Our assessment shows large differences in computational complexity for similar image quality, and aims at providing a recommendation for selecting a suitable resampling strategy.

Index Terms— Resampling, autostereoscopic, multiview

1. INTRODUCTION

Stereoscopic TV displays have dominated the consumer market in the past years. Unfortunately, current technology relies on special glasses to perceive the 3D illusion and thus reduces the overall attraction of stereoscopic 3D (S3D). As alternative, multiview autostereoscopic 3D (MA3D) displays are able to provide a glasses-free 3D impression, and the first high-quality TV sets are emerging into the consumer market.

One key challenge of MA3D is to provide multiple views of the same scene to the viewer. Beside the technical difficulty of acquiring multiview (MV) content, the number of views is display-specific and can currently range from 5 to more than 20 views. Thus, producing MV is not feasible and S3D-to-MV conversion is a more attractive option. The main technology employed today is depth-image based rendering (DIBR) [1]; the more recent image domain warping (IDW) approach [2] is also gaining increased interest.

Any S3D-to-MV algorithm involves a non-linear image transformation step. Although each output view can be treated as a separate resampling problem, MV displays require adapted anti-aliasing methods as described by Konrad et. al. [3–7]. Based on these findings, our work highlights the specific effects of MV resampling and provides a comparison of several resampling techniques in terms of visual artifacts and hardware costs (number of operations and memory requirements). The evaluation is based on the IDW framework [2], but the results can also be applied to DIBR methods. While a large body of previous work is concerned with single image resampling [8–10], their findings cannot be directly extended to MV sampling.

2. BACKGROUND

In this work, we analyze the following model for MV display resampling. High definition S3D video is converted to n output views, using a non-linear forward warp function that maps each input view to a transformed output view. The warped output views are then resampled to form an *interleaved* MV image. In the following, we summarize these steps in more detail, based on [3–10].

Single View Image Resampling. An image warp is defined as continuous mapping $\mathbf{x} = \mathbf{m}(\mathbf{u})$ that maps any coordinate $\mathbf{u} \in \mathbb{R}^2$ from the input domain to a location $\mathbf{x} \in \mathbb{R}^2$ in the output domain. In order to transform a discrete input to a discrete output, four steps are generally employed. First, a continuous signal $f_{\mathrm{in}}^{\mathrm{C}}(\mathbf{u})$ is reconstructed by applying an interpolation filter h_{i} to the discrete input. Second, the continuous signal is warped to $f_{\mathrm{warp}}^{\mathrm{C}}(\mathbf{x})$ using $\mathbf{m}(\mathbf{u})$. In a third step, an anti-aliasing filter h_{aa} is applied to achieve an aliasing-free output signal $f_{\mathrm{aa}}^{\mathrm{C}}(\mathbf{x})$. Finally, this signal is sampled to form the discrete output image. See Fig. 1 for an illustration.

Sampling for Multiview Displays. Image sampling is usually performed using orthogonal sampling grids. In contrast, MA3D displays often employ slanted lenticular lenses to achieve view-multiplexing [3], and thus result in non-orthogonal or irregular sampling grids (Fig. 1). Unfortunately, classical sampling theory cannot be applied to define the anti-aliasing filter h_{aa} in such cases. As shown by Konrad et. al. [4–7], an ideal filter needs to limit the frequency content to a *primitive cell* of the *inverse sampling lattice*. In general, this results in non-separable anti-aliasing filters with an intricate shape. For example, the lattice in Fig. 1 leads to a filter with a hexagonally shaped frequency response. However, for natural images, this ideal filter can be approximated by using more efficient separable filters [6].

Resolution Ratio. Due to the view-multiplexing of MA3D displays, the effectively used resolution of each input view is often considerably smaller than the overall resolution. More formally, a *resolution ratio* between the input images and the interleaved output image can be defined as the quotient of the density of their sampling lattices [4]. This ratio is denoted as $\rho(\Lambda_1, \Lambda_2) = d_2/d_1$ in the following, where d_1 and d_2 are the corresponding densities of the lattices Λ_1 and Λ_2. The ratio allows to describe a change in resolution for non-orthogonal sampling grids (e.g. downsampling for $\rho(\Lambda_1, \Lambda_2) < 1$), and plays an important role in our evaluations.

3. PRACTICAL MULTIVIEW RESAMPLING

Having introduced the general steps of MV resampling, we now address the problems that arise with computationally efficient filter kernels. We then show different strategies for MV rendering and highlight their potential pitfalls.

3.1. Resampling Artifacts

Generally, there is no closed-form analytical expression for the rendering equation $f_{\mathrm{view}}^{\mathrm{D}}(\mathbf{x})$ that is useable in practice. Thus, it has to be approximated with non-ideal filters and intermediate rasterization steps, which in turn gives rise to blur and aliasing. Blur is

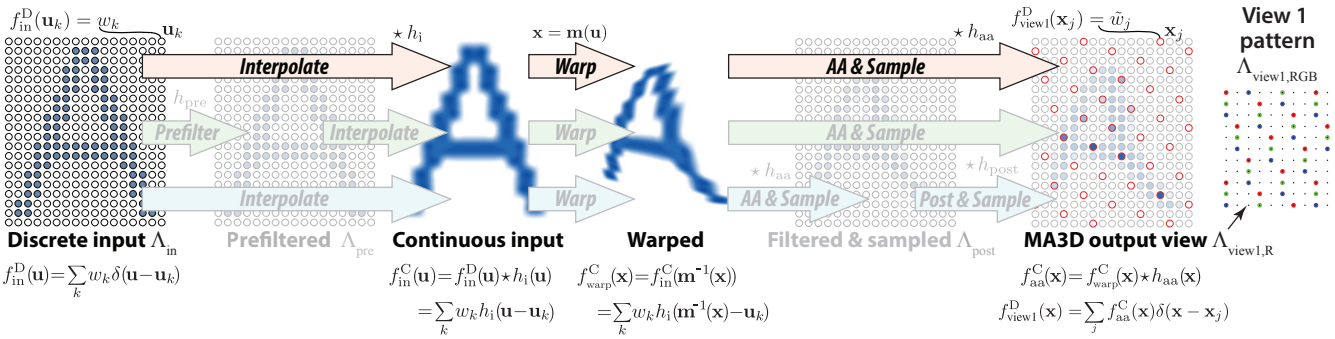

Fig. 1: *MV resampling: the opaque red arrows (top) reflect the steps of the analytical resampling problem as presented in Sec. 2. The faded green arrows (middle) and blue arrows (bottom) indicate the steps of practical implementations using* pre-filtering *and* post-filtering*, as presented in Sec. 3.2. Note that the MA3D output view is sampled on a non-orthogonal sampling grid defined by Λ_{view1}, illustrated for the Alioscopy HD 42" display [11] on the right.*

due to attenuation in the filter pass band and can be reduced by using higher order filters. Aliasing is due to insufficient attenuation in the stop band which causes components of the periodic spectrum of the input image to alias back into the pass band upon re-sampling. We will focus on aliasing artifacts beause aliasing is specifically noticeable in multiview resampling applications due to non-orthogonal sampling grids. There are two potential sources for aliasing: the non-orthogonal display sampling and the warp transformation.

To analyze aliasing artifacts due to non-orthogonal sampling, we assume a simplified setup with unit mapping $\mathbf{m}(\mathbf{u}) = \mathbf{u}$ and similar sampling densities $\rho(\Lambda_{in}, \Lambda_{view}) \approx 1$. In this case, both filters h_i and h_{aa} in Fig. 1 (red arrows) can be combined into one resampling filter $h_{i\_aa} = h_i \star h_{aa}$. When a low-order separable filter such as the bilinear kernel is used, aliasing can appear due to the mismatch to the bandwidth shape. While this effect is usually negligible for orthogonal sampling lattices, it is much more visible for non-orthogonal lattices. As the spectral components of the input are also replicated non-orthogonally, these replicas can bleed into very different regions of the hexagonally-shaped display bandwidth, shown in Fig. 2 (a-b). This results in aliasing patterns with different directionality, which can be seen in the reconstructed[1] view Fig. 2 (c).

Aliasing due to the warp manifests itself in similar but more localized artifacts. These artifacts can occur in regions that are strongly compressed. In S3D-to-MV conversion, this type of aliasing is typically less visible, as image compressions are only performed in less salient regions such as backgrounds or uniform regions [2]. Furthermore, in S3D applications, transformations are mostly restricted to horizontal translations since the transformations originate from *horizontal disparities* [1].

3.2. Removing Aliasing Artifacts

There are two diverse ways to mitigate aliasing due to the non-orthogonal sampling: first, the spectral copies can be attenuated better by using a higher order filter h_{i_aa}, as shown in Fig. 2 (c-e); secondly, the spectral copies can be spaced further apart by using oversampled input images, as shown in Fig. 2 (f-h). If a unit mapping is employed, both techniques could be integrated into the resampling pipeline easily in one single step. However, when using arbitrary transformation mappings $\mathbf{m}(\mathbf{u})$, combining h_i and h_{aa} becomes computationally extremely challenging, as most filter kernels cannot be transformed analytically.

Thus one solution is to perform the interpolation with a low order interpolation kernel h_i and apply a high quality *post-filter* h_{post} to the transformed image (the technique is illustrated in Fig. 1 with blue arrows). This post-filter is then independent of the transformation and can be implemented efficiently. Aliasing artifacts due to the warp are barely visible in this setting, as no directionality change happens when the transformed image is sampled on the orthogonal grid Λ_{post} (Fig. 3a shows a compressed image region that was rendered with a post-filtering method and no anti-aliasing filter h_{aa}). Anti-aliasing is not strictly required, but it could be performed by adaptively super-sampling the output image on a higher resolution along the horizontal direction, for example.

Using the fact that for stereo applications the warp is approximately a global translation, i.e., only has small local variations, we propose a second strategy that moves the high-quality filter to the input image domain. That is, we use oversampling in the input domain, by applying a high quality *pre-filter* h_{pre} to up-sample the input image (illustrated in Fig. 1 with green arrows). Deviations from a global translation in the warp now causes aliasing artifacts that exhibit a directionality change since the transformed image is sampled on the non-orthogonal grid Λ_{view}. Thus aliasing due to the warp is in this case more visible than with post-filtering (Fig. 3b). The effect can be mitigated using adaptive supersampling or by using adaptive splatting [12] as shown in Fig. 3 (c-d).

4. EVALUATION

In this section, we compare the resampling quality and complexity of efficient and practical resampling strategies for MV rendering. First, we will describe the evaluation setup. Then, we discuss the evaluated resampling strategies. Finally we will discuss the benefits and drawbacks of the evaluated algorithms. Fig. 4 summarizes the results of our evaluation.

PSNR Measurements. In order to assess the resampling quality of different resampling strategies, we measure the mean *peak sig-*

[1]The display views are reconstructed on a rectangular 1080p lattice with a very high order (51x51), hexagonal FIR filter in order to simulate the MA3D-screen-to-eye transfer function. The filter was designed with the fwind2 method from Matlab's image processing toolbox (using a Hann window).

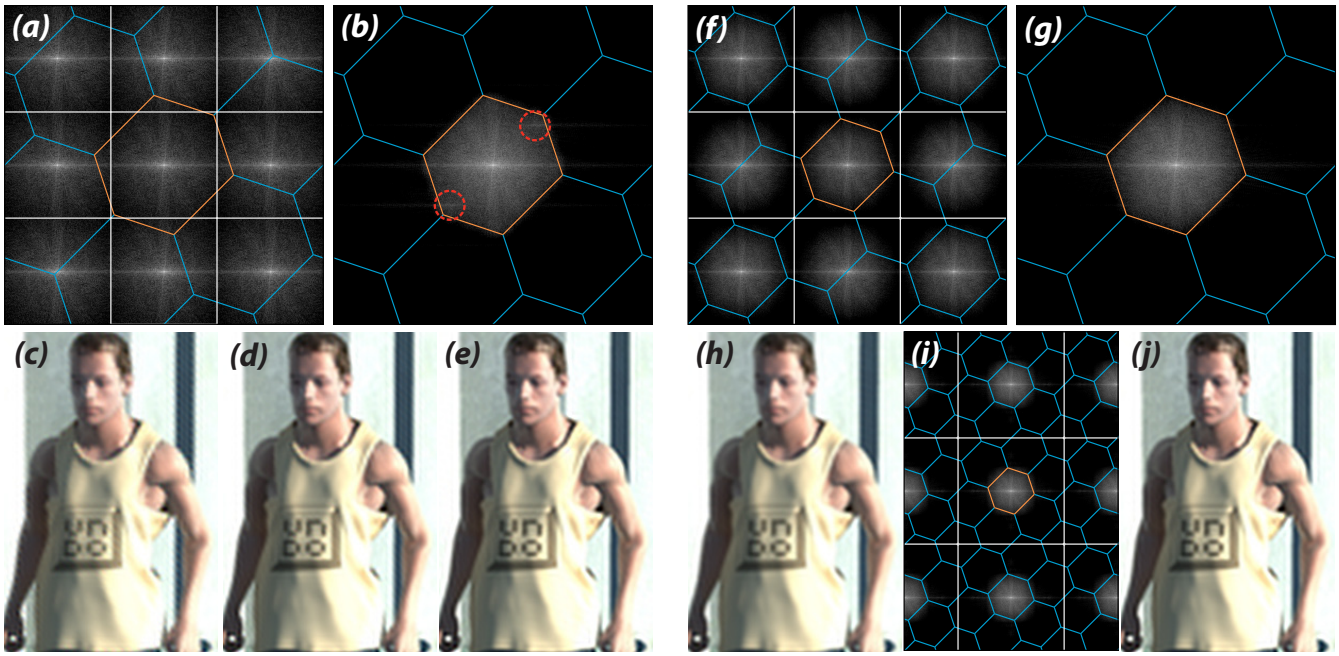

Fig. 2: *Aliasing artifacts due to non-orthogonal view sampling: a) shows the periodic input spectrum, where the white boxes illustrate the input bandwidth. The bandwidth of one view of the Alioscopy HD 42" display [11] is illustrated with the hexagonal cells (the primitive cell is drawn in orange). c)-d) show reconstructed display views for different resampling filters of increasing complexity: c) bilinear, d) bicubic, e) 6x6 polyphase Lanczos (use your document viewer too zoom in in order to better see the artifacts). b) shows the resulting spectrum of c). The frequency components in the red circles cause aliasing with directionality change. Alternatively, oversampling can be used instead of higher order filters to reduce aliasing. f)-h) show the effect for* $1.5\times$ *and i)-j) for* $3\times$ *oversampling, using a separable 6x6 polyphase Lanczos kernel to upsample the input image.*

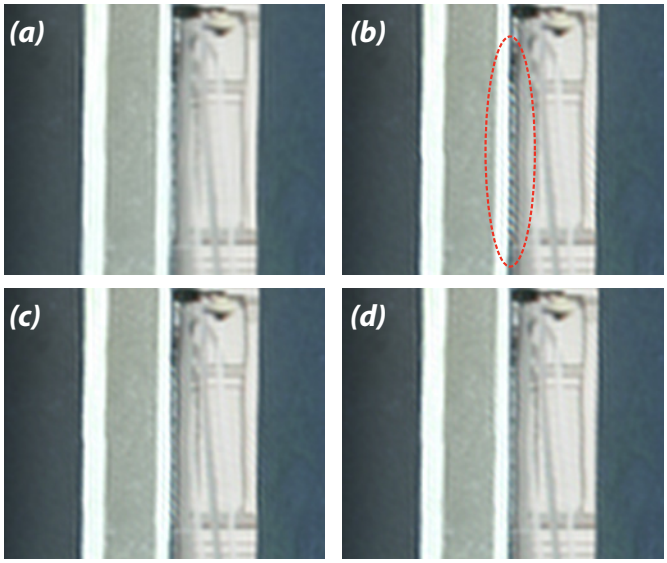

Fig. 3: *a) and b) show the effect of post- and pre-filtering with a bilinear kernel* h_i *and no filter* h_{aa}*. Aliasing is barely visible in a). c) shows similar results, with additional adaptive* $2\times$ *supersampling in horizontal direction using a simple box filter* h_{aa}*. c) shows pre-filtering with adaptive EWA splatting [12], where Gaussian kernels are used for* h_i *and* h_{aa}*.*

nal to noise ratio (PSNR) of one reconstructed view with respect to a high quality reference over 31 frames of the *Undo Dancer* test sequence [13]. The output views are reconstructed as described in Sec. 3.1. The reference image is rendered from 1080p input footage with a high quality two-step splatting algorithm and a nonseparable 51×51 post-filter h_{post}. The algorithm employs 6×6 Lanczos kernels for h_i and h_{aa} and an intermediate rasterization step with variable supersampling to transform the input images.

Complexity Measurements. To compare the computational complexity of the different approaches, the number of operations and memory costs are estimated analytically. We assume 10% of all rendered pixels require anti-aliasing for the adaptive anti-aliasing methods. Memory costs assume that the warp results in a vertical pixel translation of maximally two pixels (on Λ_{in_r1}).

Experimental Setup. Our evaluation is performed using the parameters of the Alioscopy HD 42" display [11], which supports 8 views interleaved into a super-lattice with resolution 1920×1080 RGB pixels. As the size of the input views has an impact on computational complexity and image quality, we evaluate each resampling configuration on three different input resolutions. A prefix $r1$ denotes $\rho\left(\Lambda_{in\_r1}, \Lambda_{view}\right) = 1/8$ (input resolution 1920×1080), prefix $r2$ denotes $\rho\left(\Lambda_{in\_r2}, \Lambda_{view}\right) = 1/2$ (input resolution 960×540), and prefix $r3$ denotes $\rho\left(\Lambda_{in\_r3}, \Lambda_{view}\right) = 9/8$ (input resolution 640×360).

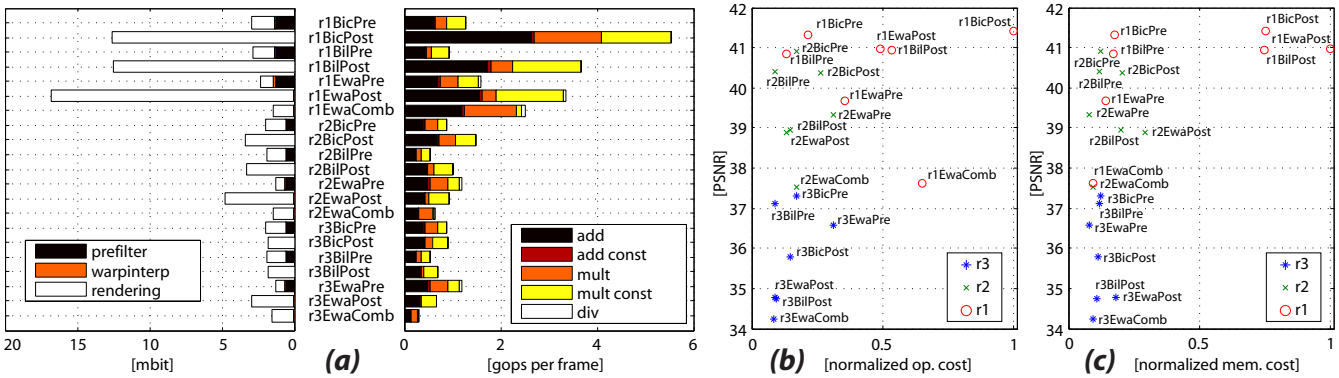

Fig. 4: *a) shows estimates of the computational cost of the evaluated methods. b) and c) show the operation and memory costs (normalized) versus image quality (mean PSNR value over 31 frames of our test sequence, evaluated on one view). The PSNR is calculated w.r.t. a high quality reference image (Sec. 4). The normalized operation cost is calculated by relative weighting of the operation types. The weights were estimated by synthesizing 20bit fixed point operators in UMC 65nm CMOS technology at 400MHz: add* $= 0.0571$*, add const* $= 0.0429$*, mult* $= 0.2857$*, mult const* $= 0.0929$*, div* $= 1.0$*.*

4.1. Resampling strategies

In our evaluation, we focus on practical and efficient resampling strategies. More specifically, we compare bilinear, bicubic and Gaussian interpolation kernels h_i and their pre- and post-filtering variants. All pre-filtering variants up-sample the input images onto a 1920×1080 grid Λ_{pre}. All post-filtering variants retain the same resolution $\Lambda_{in} = \Lambda_{post}$ before subsampling to Λ_{view}.

The bilinear and bicubic kernels h_i are implemented using backward mapping, and will thus require the inversion of $\mathbf{m}(\mathbf{u})$. The 4 variants will be denoted as *BilPre*, *BilPost*, *BicPre* and *BicPost*. Aiasing due the warping is mitigated using adaptive $2\times$ supersampling in horizontal direction using a simple box filter h_{aa}.

Gaussian kernels allow for efficient forward mapping and do not require the costly warp inversion of backward mappings. Furthermore, Gaussians are closed under affine transformations which allows to analytically combine interpolation h_i and anti-aliasing h_{aa} into one single filter (adaptive elliptical weighted average (EWA) splatting [12]). The respective pre- and post-filtering methods are denoted as *EwaPre* and *EwaPos*. As additional simplification, the post-filter h_{post} can be chosen as Gaussian and merged with h_i and h_{aa} into one single filter kernel, such that only one filtering step has to be performed. This combined EWA variant will be denoted as *EwaComb*. Note, that neither of the three variants requires super-sampling, as forward EWA rendering is implemented as an area sampler and thus implicitly includes anti-aliasing.

Warp Interpolation and Inversion. The warp $\mathbf{m}$ is usually not provided as analytical function but described by samples of a quadrilateral mesh [2]. Furthermore, this mesh is often not provided at pixel-resolution but approximated at lower resolution, usually at $10\times$ smaller widths and heights. This has implications for backward-mapping methods, as this quadrilateral warp mesh must be inverted for every output sample. Note that higher-resolution meshes require less intermediate values to be stored in memory, but also result in higher computational complexity. In our evaluation, this warp inversion is performed using barycentric coordinates for performance reasons, however, this may lead to visible artifacts for coarser mesh sizes. The complexity of backward mapping methods is thus reduced in applications that use the inverse mapping $\mathbf{m}^{-1}()$.

4.2. Discussion

Fig. 4 a) shows the estimated complexity of the evaluated algorithm variants and (b-c) relate those estimates to the image quality. As can be seen from the PSNR plots, rendering from higher resolution input footage leads to significantly better quality. However, all post-filtering variants suffer from increased operation and memory cost in this case. This becomes even more pronounced for a bigger number of views, as all intermediate results Λ_{post} need to be stored and downsampled. In contrast, the pre-filtering variants offer considerably lower complexity when rendering from higher input resolution, as only a small amount of super-samples is required in addition to the exact number of output samples.

The combined filtering approach *EwaComb* has low complexity and requires small memory sizes, especially for $r3$ footage. However it suffers from the low filter order and exhibits well visible, global aliasing artifacts due to the display sampling. This results in low PSNR values, which could only be mitigated by broadening the filter width at the cost of increased blurring. Thus the PSNR saturates around 37.6dB even for high resolution footage. While the other two forward mapping algorithms *EWAPre*, *EWAPost* show the promise of lower complexity due to the combined filter, the combination comes at the cost of lower PSNR ratios than their backward counterparts.

5. CONCLUSION

Our evaluations show that among the evaluated algorithm configurations the *pre-filtering* variants offer the best trade off between image quality and computational complexity – regardless of the input resolution. They also show good scalability with respect to the number of views, since only the required display pixels are calculated and the high quality filtering is performed on the input images. However, the highest visual quality is achieved by *post-filtering* variants using high resolution input images. Interestingly, the quality-complexity trade-offs for the EWA filtering are only convincing for smaller input resolutions. The highly optimized forward mapping algorithm of Berretty et al. [14] shows promising results and might be a better option to EWA filtering.

6. REFERENCES

[1] A. Smolic, P. Kauff, S. Knorr, A. Hornung, M. Kunter, M. Mueller, and M. Lang, "Three-dimensional video postproduction and processing," *Proceedings of the IEEE*, vol. 99, no. 4, pp. 607–625, 2011.

[2] M. Farre, O. Wang, M. Lang, N. Stefanoski, A. Hornung, and A. Smolic, "Automatic content creation for multiview autostereoscopic displays using image domain warping," in *2nd IEEE International Workshop on Hot Topics in 3D (Hot3D 2011)*. IEEE.

[3] J. Konrad and M. Halle, "3-D displays and signal processing," *Signal Processing Magazine, IEEE*, vol. 24, no. 6, pp. 97 –111, 2007.

[4] J. Konrad and P. Agniel, "Artifact reduction in lenticular multiscopic 3d displays by means of anti-alias filtering," *Stereoscopic Displays and Virtual Reality Systems*, vol. 5006, no. 1, pp. 336–347, 2003.

[5] ——, "Non-orthogonal subsampling and anti-alias filtering for multiscopic 3d displays," *Stereoscopic Displays and Virtual Reality Systems XI*, vol. 5291, no. 1, pp. 105–116, 2004.

[6] ——, "Subsampling models and anti-alias filters for 3-d automultiscopic displays," *IEEE Transactions on Image Processing*, vol. 15, no. 1, pp. 128 –140, January 2006.

[7] J. Konrad and A. Jain, "Crosstalk in automultiscopic 3-d displays: blessing in disguise?" *Stereoscopic Displays and Virtual Reality Systems XIV*, vol. 6490, no. 1, p. 649012, 2007.

[8] P. Heckbert, "Fundamentals of texture mapping and image warping," Masters Thesis, Univ. of California, Berkeley, Dept. of Electrical Eng. and Computer Science, 1989.

[9] G. Wolberg, *Digital image warping*. IEEE Computer Society press, 1990, vol. 3.

[10] N. A. Dodgson, "Technical report: Image resampling," 15 JJ Thomson Avenue, Cambridge CB3 0FD, United Kingdom, Tech. Rep., August 1992, this technical report is based on a dissertation submitted by the author for the degree of Doctor of Philosophy to the University of Cambridge, Wolfson College.

[11] Alioscopy, "Alioscopy 3d displays," November 2012, http://www.alioscopy.com/en/3Ddisplays.php.

[12] P. Greisen, M. Schaffner, S. Heinzle, M. Runo, A. Smolic, A. Burg, H. Kaeslin, and M. Gross, "Analysis and VLSI implementation of EWA rendering for real-time HD video applications," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 11, pp. 1577 –1589, nov. 2012.

[13] "Call for proposals on 3d video coding technology," ISO/IEC JTC1/SC29/WG11, Geneva, Switzerland, Tech. Rep. N12036, March 2011, approved.

[14] R.-P. M. Berretty, F. J. Peters, and G. T. G. Volleberg, "Real-time rendering for multiview autostereoscopic displays," pp. 60 550N–60 550N–12, 2006. [Online]. Available: http://dx.doi.org/10.1117/12.642125