Display Adaptive 3D Content Remapping

Belen Masia[a,b], Gordon Wetzstein[b], Carlos Aliaga[a], Ramesh Raskar[b], Diego Gutierrez[a]

[a] *Universidad de Zaragoza*
[b] *MIT Media Lab*

Abstract

Glasses-free automultiscopic displays are on the verge of becoming a standard technology in consumer products. These displays are capable of producing the illusion of 3D content without the need of any additional eyewear. However, due to limitations in angular resolution, they can only show a limited depth of field, which translates into blurred-out areas whenever an object extrudes beyond a certain depth. Moreover, the blurring is device-specific, due to the different constraints of each display. We introduce a novel display-adaptive light field retargeting method, to provide high-quality, blur-free viewing experiences of the same content on a variety of display types, ranging from hand-held devices to movie theaters. We pose the problem as an optimization, which aims at modifying the original light field so that the displayed content appears sharp while preserving the original perception of depth. In particular, we run the optimization on the central view and use warping to synthesize the rest of the light field. We validate our method using existing objective metrics for both image quality (blur) and perceived depth. The proposed framework can also be applied to retargeting disparities in stereoscopic image displays, supporting both dichotomous and non-dichotomous comfort zones.

Keywords: stereo, displays, automultiscopic, content retargeting.

1. Introduction

Within the last years, stereoscopic and automultiscopic displays have started to enter the consumer market from all angles. These displays can show three-dimensional objects that appear to be floating in front of or behind the physical screen, even without the use of additional eyewear. Capable of electronically switching between a full-resolution 2D and a lower-resolution 3D mode, parallax barrier technology [1] is dominant for hand-held and tablet-sized devices, while medium-sized displays most often employ arrays of microlenses [2]. Although most cinema screens today are stereoscopic and rely on additional eyewear, large-scale automultiscopic projection systems are an emerging technology [3]. Each technology has its own particular characteristics, including field of view, depth of field, contrast, resolution, and screen size. Counterintuitively, produced content is usually targeted toward a single display configuration, making labor-intense, manual post-processing of the recorded or rendered data necessary.

Display-adaptive content retargeting is common practice for attributes such as image size, dynamic range (tone mapping), color gamut, and spatial resolution [4]. In order to counteract the accommodation-convergence mismatch of stereoscopic displays, stereoscopic disparity retargeting methods have recently been explored [5, 6, 7, 8, 9]. These techniques are successful in modifying the disparities of a stereo image pair so that visual discomfort of the observer is mitigated while preserving the three-dimensional appearance of the scene as much as possible. Inspired by these techniques, we tackle the problem of 3D content retargeting for glasses-free light field (i.e. automultiscopic) displays. These displays exhibit a device-specific

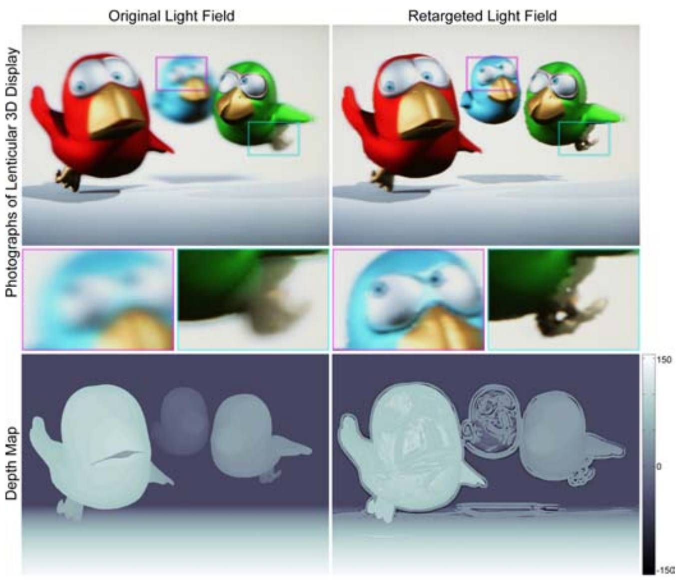

Figure 1: Our 3D content retargeting for a glasses-free lenticular display. Due to the limited depth of field of all light field displays, some objects in a 3D scene will appear blurred. Our remapping approach selectively fits the 3D content into the depth budget of the display, while preserving the perceived depth of the original scene. Top: actual photographs of the original and retargeted scenes, as seen on a Toshiba GL1 lenticular display. Notice the improvement in the blue bird or the legs of the green bird in the retargeted version. Middle: close-ups. Bottom: original and retargeted depths yielded by our method.

depth of field (DOF) that is governed by their limited angular resolution [10, 11]. Due to the fact that most light field displays only provide a low angular resolution, that is the number of viewing zones, the supported DOF is so shallow that virtual

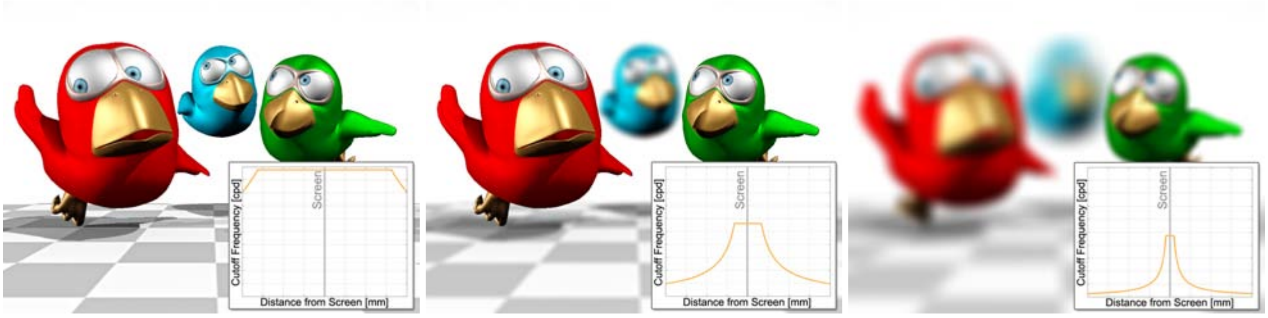

Figure 2: Simulated views of the *three-birds* scene for three different displays. From left to right: Holografika HoloVizio C80 movie screen, desktop and cell phone displays. The last two displays fail to reproduce it properly, due to their intrinsic depth-of-field limitations. The insets plot the depth vs. cut-off frequency charts for each display.

3D objects extruding from the physical display enclosure appear blurred out (see Figs. 1, left, and 2 for a real photograph and a simulation showing the effect, respectively). We propose here a framework that remaps the disparities in a 3D scene to fit the DOF constraints of a target display by means of an optimization scheme that leverages perceptual models of the human visual system. Our optimization approach runs on the central view of an input light field and uses warping to synthesize the rest of the views.

Contributions. Our nonlinear optimization framework for 3D content retargeting specifically provides the following contributions:

- We propose a solution to handle the intrinsic trade-off between the spatial frequency that can be shown and the perceived depth of a given scene. This is a fundamental limitation of automultiscopic displays (see Section 3).
- We combine exact formulations of display-specific depth of field limitations with models of human perception, to find an optimized solution. In particular, we consider the frequency-dependent sensitivity to contrast of the human visual system, and the sensitivity to binocular disparity. Based on this combination, a first objective term minimizes the perceived luminance and contrast difference between the original and the displayed scene, effectively minimizing DOF blur, while a second term strives to preserve the perceived depth.
- We validate our results with existing state-of-the-art, objective metrics for both image quality and perceived depth.
- We show how our framework can be easily extended to the particular case of *stereoscopic* disparity, thus demonstrating its versatility.
- For this extension, we account for a non-dichotomous zone of viewing comfort which constitutes a more accurate model of discomfort associated with the viewing experience.

As a result of our algorithm, the depth of a given 3D scene is modified to fit the DOF constraints imposed by the target display, while preserving the perceived 3D appearance and the desired 2D image fidelity (Figure 1, right).

Limitations. We do not aim at providing an accurate model of the behavior of the human visual system; investigating all the complex interactions between its individual components remains an open problem as well, largely studied by both psychologists and physiologists. Instead, we rely on existing computational models of human perception and apply them to the specific application of 3D content retargeting. For this purpose, we currently consider sensitivities to luminance contrast and depth, but only approximate the complex interaction between these cues using a heuristic linear blending, which works well in our particular setting. Using the contrast sensitivity function in our context (Section 4) is a convenient but conservative choice. Finally, depth perception from motion parallax exhibits strong similarities in terms of sensitivity with that of binocular disparity, suggesting a close relationship between both [12]; but existing studies on sensitivity to motion parallax are not as exhaustive as those on binocular disparity, and therefore a reliable model cannot be derived yet. Moreover, some studies have shown that, while both cues are effective, stereopsis is more relevant by an order of magnitude [13]. In any case, our approach is general enough so that as studies on these and other cues advance and new, more sophisticated models of human perception become available, they could be incorporated to our framework.

2. Related Work

Glasses-free 3D displays were invented more than a century ago, but even today, the two dominating technologies are parallax barriers [1] and integral imaging [2]. Nowadays, the palette of existing 3D display technologies, however, is much larger and includes holograms, volumetric displays, multilayer displays and directional backlighting among many others. State of the art reviews of conventional stereoscopic and automultiscopic displays [14] and computational displays [15] can be found in the literature. With the widespread use of stereoscopic image capture and displays, optimal acquisition parameters and capture systems [16, 17, 18, 19, 20], editing tools [21, 22], and spatial resolution retargeting algorithms for light fields [23] have recently emerged. In this paper, we deal with the problem of depth remapping of light field information to the specific constraints of each display.

Generally speaking, content remapping is a standard approach to adapt spatial and temporal resolution, contrast, colors, and sizes of images to a display having limited capabilities in any of these dimensions [4]. For the particular case of disparity remapping, Lang et al. [6] define a set of non-linear disparity remapping operators, and propose a new stereoscopic warping technique for the generation of the remapped stereo pairs. A metric to assess the magnitude of perceived changes in binocular disparity is introduced by Didyk et al. [8], who also investigate the use of the Cornsweet illusion to enhance perceived depth [24]. Recently, the original disparity metric has been further refined including the effect of luminance-contrast [9]. Kim and colleagues [7] develop a a novel framework for flexible manipulation of binocular parallax, where a new stereo pair is created from two non-linear cuts of the EPI volume corresponding to multi-perspective images [25]. Inspired by Lang and colleagues [6], they explore linear and non-linear global remapping functions, and also non-linear disparity gradient compression. Here we focus on a remapping function that incorporates the specific depth of field limitations of the target display [26]. Section 8 provides direct comparisons with some of these approaches.

3. Display-specific Depth of Field Limitations

Automultiscopic displays are successful in creating convincing illusions of three-dimensional objects floating in front and behind physical display enclosures without the observer having to wear specialized glasses. Unfortunately, all such displays have a limited depth of field which, just as in wide-aperture photography, significantly blurs out-of-focus objects. The focal plane for 3D displays is directly on the physical device. Display-specific depth of field expressions have been derived for parallax barrier and lenslet-based systems [10], multilayer displays [11], and directional backlit displays [27]. In order to display an aliasing-free light field with any such device, four-dimensional spatio-angular pre-filters need to be applied before computing the display-specific patterns necessary to synthesize a light field, either by means of sampling or optimization. In practice, these filters model the depth-dependent blur of the individual displays and are described by a depth of field blur applied to the target light field. Intuitively, this approach fits the content into the DOF of the displays by blurring it as necessary. Figure 3 illustrates the supported depth of field of various automultiscopic displays for different display sizes.

Specifically, the depth of field of a display is modeled as the maximum spatial frequency f_ξ of a diffuse plane at a distance d_0 to the physical display enclosure. As shown by previous works [10, 11], the DOF of parallax barrier and lenslet-based displays is given by

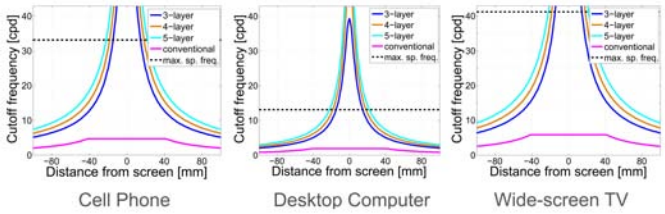

Figure 3: Depth of field for different display architectures and target displays. From left to right: cell phone (p = 0.09mm, v_D = 0.35m); desktop computer (p = 0.33mm, v_D = 0.5m); and widescreen TV (p = 0.53mm, v_D = 2.5m). For comparison purposes all depths of field are modeled for seven angular views.

$$\left|f_\xi\right| \leq \begin{cases} \frac{f_0}{N_a}, & for\, |d_0| + (h/2) \leq N_a h \\ (\frac{h}{(h/2)+|d_0|}) f_0, & otherwise \end{cases}, \qquad (1)$$

where N_a is the number of angular views, d_0 is the distance to the front plane of the display (i.e. the parallax barrier or lenslet array plane), h represents the thickness of the display, $f_0 = 1/(2p)$, and p is the size of the view-dependent subpixels of the back layer of the display, making the maximum resolution of the display at the front surface $f_\xi = f_0/N_a = 1/(2pN_a)$. For multilayered displays, the upper bound on the depth of field for a display of N layers was derived by Wetzstein et al. [11] to be

$$\left|f_\xi\right| \leq N f_0 \sqrt{\frac{(N+1)h^2}{(N+1)h^2 + 12(N-1)d_0^2}}. \qquad (2)$$

Note that in this case d_0 represents the distance to the middle of the display, and p the pixel size of the layers.

It can be seen how depth of field depends on display parameters such as pixel size p, number of viewing zones N_a, device thickness h, and number of layers N (for multilayer displays), and thus varies significantly for different displays. It also depends on the viewing distance v_D when expressed in cycles per degree. The above expressions can then be employed to predict an image displayed on a particular architecture, including loss of contrast and blur. Figure 2 shows three simulated views of the *three-birds* scene for three different displays: a Holografika HoloVizio C80 movie screen ($h = 100mm$, $p = 0.765mm$, $v_D = 6m$), a Toshiba automultiscopic monitor ($h = 20$, $p = 0.33$, $v_D = 1.5$) and a cell-phone-sized display ($h = 6$, $p = 0.09$, $v_D = 0.35$). The scene can be represented in the large movie screen without blurring artifacts (left); however, when displayed on a desktop display (middle), some areas appear blurred due to the depth-of-field limitations described above (see the blue bird). When seen on a cell-phone display (right), where the limitations are more severe, the whole scene appears badly blurred. In the following, we show how these predictions are used to optimize the perceived appearance of a presented scene in terms of image sharpness and contrast, where the particular parameters of the targeted display are an input to our method.

4. Optimization Framework

In order to mitigate display-specific DOF blur artifacts, we propose to scale the original scene into the provided depth budget while preserving the perceived 3D appearance as best as possible. As detailed in Section 3, this is not trivial, since there is an intrinsic trade-off between the two goals. We formulate this as a multi objective optimization problem, with our objective function made up of two terms. The first one minimizes the perceived luminance and contrast difference between the original and the displayed scene, for which display-specific expressions of the displayable frequencies are combined with a perceptual model of contrast sensitivity. The second term penalizes loss in perceived depth, for which we leverage disparity sensitivity metrics. Intuitively, the disparity term prevents the algorithm from yielding the obvious solution where the whole scene is flattened onto the display screen; this would guarantee perfect focus at the cost of losing any sensation of depth. The input to our algorithm is the depth map and the luminance image of the central view of the original light field, which we term d_{orig} and L_{orig}, respectively. The output is a retargeted depth map d, which is subsequently used to synthesize the retargeted light field.

Optimizing luminance and contrast: We model the display-specific frequency limitations by introducing spatially-varying, depth-dependent convolution kernels $k(d)$. They are defined as Gaussian kernels whose standard deviation σ is such that frequencies above the cut-off frequency at a certain depth $f_{\xi}(d)$ are reduced to less than 5% of its original magnitude. Although more accurate image formation models for defocus blur in scenes with occlusions can be found in the literature [28], their use is impractical in our optimization scenario, and we found the Gaussian spatially-varying kernels to give good results in practice. Kernels are normalized so as not to modify the total energy during convolution. As such, the kernel for a pixel i is:

$$k(d) = \frac{exp(-\frac{x_i^2+y_i^2}{2(\sigma(d))^2})}{\sum_j^K \left(exp(-\frac{x_j^2+y_j^2}{2(\sigma(d))^2})\right)} \quad (3)$$

where K is its number of pixels. The standard deviation σ is computed as:

$$\sigma(d) = \frac{\sqrt{-2log(0.05)}}{2\pi p f_{\xi}(d)} \quad (4)$$

with p being the pixel size in $mm/pixel$.

To take into account how frequency changes are perceived by a human observer, we rely on the fact that the visual system is more sensitive to near-threshold changes in contrast and less sensitive at high contrast levels [29]. We adopt a conservative approach and employ sensitivities at near-threshold levels as defined by the contrast sensitivity function (CSF). We follow the expression for contrast sensitivities ω_{CSF} proposed by Mantiuk et al. [30], which in turn builds on the model proposed by Barten [31]:

$$\omega_{CSF}(l, f_l) = p_4 s_A(l) \frac{MTF(f_l)}{\sqrt{(1+(p_1 f_l)^{p_2})(1-e^{-(f_l/7)^2})^{-p_3}}}, \quad (5)$$

where l is the adapting luminance in [cd/m^2], f_l represents the spatial frequency of the luminance signal in [cpd] and p_i are the fitted parameters provided in Mantiuk's paper[1]. MTF (modulation transfer function) and s_A represent the optical and the luminance-based components respectively, and are given by:

$$MTF(f_l) = \sum_{k=1..4} a_k e^{-b_k f_l} \quad (6)$$

$$s_A(l) = p_5 \left(\left(\frac{p_6}{l}\right)^{p_7} + 1\right)^{-p_8} \quad (7)$$

where a_k and b_k can again be found in the original paper. Figure 4 (left) shows contrast sensitivity functions for varying adaptation luminances, as described by Equations 5-7. In our context we deal with complex images, as opposed to a uniform field; we thus use the steerable pyramid [32] $\rho_S(\cdot)$ to decompose a luminance image into a multi-scale frequency representation. The steerable pyramid is chosen over other commonly used types of decomposition (e.g. Cortex Transform) since it is mostly free of ringing artifacts that can cause false masking signals [30].

Taking into account both the display-specific frequency limitations and the HVS response to contrast, we have the following final expression for the first term of our optimization:

$$\left\| \omega_{CSF} \left(\rho_S \left(L_{orig} \right) - \rho_S \left(\phi_b \left(L_{orig}, d \right) \right) \right) \right\|_2^2, \quad (8)$$

where ω_{CSF}, defined by Equation 5, are frequency-dependent weighting factors, and the operator $\phi_b(L, d) = k(d) * L$ models the display-specific, depth-dependent blur (see Section 3 and Figure 3). Note that we omit the dependency of ω_{CSF} on (l, f_l) for clarity. Figure 5 (*left*) shows representative weights ω_{CSF} for different spatial frequency luminance levels of the pyramid for a sample scene.

Preserving perceived depth: This term penalizes the perceived difference in depth between target and retargeted scene using disparity sensitivity metrics. As noted by different researchers, the effect of binocular disparity in the perception of depth works in a manner similar to the effect of contrast in the perception of luminance [8, 33, 34]. In particular, our ability to detect and discriminate depth from binocular disparity depends on the frequency and amplitude of the disparity signal. Human sensitivity to binocular disparity is given by the following equation [8] (see also Figure 4, right):

$$\omega_{BD}(a, f) = (0.4223 + 0.007576a + 0.5593log_{10}(f) + 0.03742alog_{10}(f) + 0.0005623a^2 + 0.7114log_{10}^2(f))^{-1} \quad (9)$$

[1] sourceforge.net/apps/mediawiki/hdrvdp/

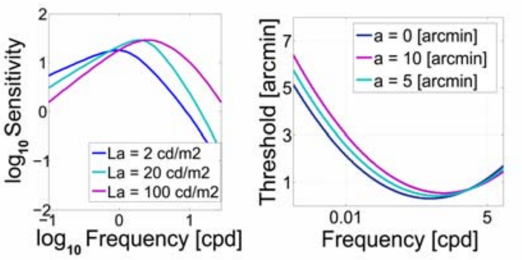

Figure 4: Thresholds and sensitivity values from which the weights for our optimization are drawn. Left: Contrast sensitivity functions. Right: Binocular disparity discrimination thresholds (thresholds are the inverse of sensitivities).

where frequency f is expressed in [cpd], a is the amplitude in [arcmin], and ω_{BD} is the sensitivity in [arcmin^{-1}]. In a similar way to ω_{CSF} in Equation 8, the weights ω_{BD} account for our sensitivity to disparity amplitude and frequency. Given this dependency on frequency, the need for a multi-scale decomposition of image disparities arises again, for which we use a Laplacian pyramid $\rho_L(\cdot)$ for efficiency reasons, following the proposal by Didyk et al. [8]. Figure 5 (*right*), shows representative weights ω_{BD}.

The error in perceived depth incorporating these sensitivities is then modeled with the following term:

$$\left\|\omega_{BD}\left(\rho_L\left(\phi_v\left(d_{orig}\right)\right)-\rho_L\left(\phi_v\left(d\right)\right)\right)\right\|_2^2. \tag{10}$$

Given the viewing distance v_D and interaxial distance e, the operator $\phi_v(\cdot)$ converts depth into vergence as follows:

$$\phi_v(d)=acos\left(\frac{\mathbf{v_L}\cdot\mathbf{v_R}}{\|\mathbf{v_L}\|\,\|\mathbf{v_R}\|}\right), \tag{11}$$

where vectors $\mathbf{v_L}$ and $\mathbf{v_R}$ are illustrated in Figure 6. The Laplacian decomposition transforms this vergence into frequency-dependent disparity levels.

Objective function: Our final objective function is a combination of Equations 8 and 10:

$$\begin{aligned}\arg\min_d &\left(\mu_{DOF}\left\|\omega_{CSF}\left(\rho_S\left(L_{orig}\right)-\rho_S\left(\phi_b\left(L_{orig},d\right)\right)\right)\right\|_2^2\right.\\ &\left.+\mu_D\left\|\omega_{BD}\left(\rho_L\left(\phi_v\left(d_{orig}\right)\right)-\rho_L\left(\phi_v\left(d\right)\right)\right)\right\|_2^2\right).\end{aligned} \tag{12}$$

For multilayer displays, we empirically set the values of $\mu_{DOF}=10$ and $\mu_D=0.003$, while for conventional displays $\mu_D=0.0003$ due to the different depth of field expressions.

5. Implementation Details

We employ a large-scale trust region method [35] to solve Equation 12. This requires finding the expressions for the analytic gradients of the objective function used to compute the Jacobian, which can be found in Annex A. The objective term in Equation 8 models a single view of the light field, i.e. the central view, in a display-specific field of view (FOV). Within a moderate FOV, as provided by commercially-available displays, this is a reasonable approximation; we obtain the rest of the light field by warping. In the following, we describe this and other additional implementation details.

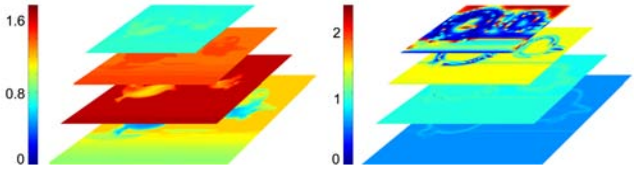

Figure 5: Left: Weights ω_{CSF} (contrast sensitivity values) for different luminance spatial frequency levels for a sample scene (*birds*). Right: Weights ω_{BD} (inverse of discrimination threshold values) for different disparity spatial frequency levels for the same scene.

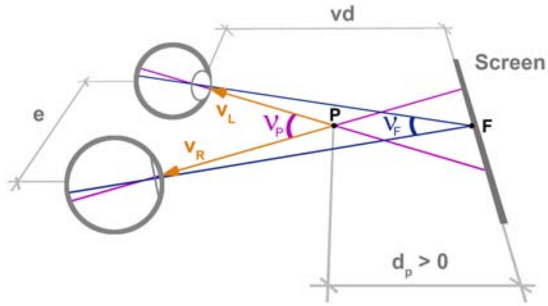

Figure 6: Computing vergence values. Vergence ν_P of a point P depends on its position, the viewing distance v_D and the interaxial e. The corresponding disparity for P is $(\nu_P-\nu_F)$. vd refers to the viewing distance and d_P is the depth of point P.

Sensitivity weights and target values: The weights used in the different terms, ω_{CSF} and ω_{BD} are pre-computed based on the values of the original depth and luminance, d_{orig} and L_{orig}. The transformation from d_{orig} to vergence, its pyramid decomposition and the decomposition of L_{orig} are also pre-computed.

Contrast sensitivity function: As reported by Mantiuk et al. [30], no suitable data exists to separate L- and M-cone sensitivity. Following their approach, we rely on the *achromatic* CSF using only luminance values.

Depth-of-field simulation: The depth-dependent image blur of automultiscopic displays is modeled as a spatially-varying convolution in each iteration of the optimization procedure. Due to limited computational resources, we approximate this expensive operation as a blend between multiple shift-invariant convolutions corresponding to a quantized depth map, making the process much more efficient. For all scenes shown in this paper, we use $n_c=20$ quantized depth clusters.

Warping: View warping is orthogonal to the proposed retargeting approach; we implement here the method described by Didyk et al. [36], although other methods could be employed instead ([7, 37, 38]). To reduce warping artifacts due to large depth gradients at the limits of the field of view for each light field, we median-filter the depth and constrain depth values around the edges.

6. Retargeting for Stereoscopic Displays

One of the advantages of our framework is its versatility, which allows to adapt it for display-specific disparity remapping of stereo pairs. We simply drop the depth of field term from Equation 12, and incorporate a new term that models the comfort zone. This is an area around the screen within which the 3D content does not create fatigue or discomfort in the viewer in stereoscopic displays, and is usually considered as a dichotomous subset of the fusional area. Although any comfort-zone model could be directly plugged into our framework, we incorporate the more accurate, non-dichotomous model suggested by Shibata et al. [39]. This model provides a more accurate description of its underlying psychological and physiological effects. Additionally, this zone of comfort depends on the viewing distance v_D, resulting on different expressions for different displays, as shown in Figure 7. Please refer to Annex B for details on how to incorporate the simpler, but less precise, dichotomous model.

Our objective function thus becomes:

$$\left\|\omega_{BD}\left(\rho_L\left(\phi_\upsilon\left(D_{orig}\right)\right)-\rho_L\left(\phi_\upsilon\left(d\right)\right)\right)\right\|_2^2+\mu_{CZ}\left\|\varphi\left(d\right)\right\|_2^2, \quad (13)$$

where $\varphi(\cdot)$ is a function mapping depth values to visual discomfort:

$$\varphi(d)=\begin{cases}1-\frac{s_{far}}{v_D-d}-T_{far} & \text{for} \quad d<0\\ 1-\frac{s_{near}}{v_D-d}-T_{near} & \text{for} \quad d\geq 0\end{cases} \quad (14)$$

where v_D is the distance from the viewer to the central plane of the screen and s_{far}, s_{near}, T_{far}, and T_{near} are values obtained in a user study carried out with 24 subjects.

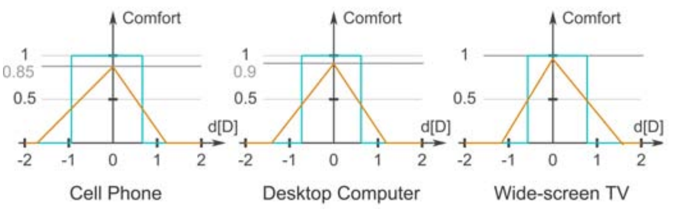

Figure 7: Dichotomous (blue) and non-dichotomous (orange) zones of comfort for different devices. From left to right: cell phone ($v_D = 0.35m$), desktop computer ($v_D = 0.5m$) and wide-screen TV ($v_D = 2.5m$).

7. Results

We have implemented the proposed algorithm for different types of automultiscopic displays including a commercial Toshiba GL1 lenticular-based display providing horizontal-only parallax with nine discrete viewing zones, and custom multilayer displays. The Toshiba panel has a native resolution of 3840×2400 pixels with a specially engineered subpixel structure that results in a resolution of 1280×800 pixels for each of the nine views. Note that even a highly-engineered device such as this suffers from a narrow depth of field due to the limited angular sampling. We consider a viewing distance of 1.5 m for the Toshiba display and 0.5 m for the multilayer prototypes.

Figures 1 and 8 show results of our algorithm for the Toshiba display. The target scenes have been originally rendered as light fields with a resolution of 9×9, with a field of view of 10°. Since the Toshiba display only supports horizontal parallax, we only use the nine horizontal views for these examples. Note how depth is compressed to fit the display's constraints in those areas with visible loss of contrast due to blur (blue bird or far away pins, for instance), while enhancing details to preserve the *perceived* depth; areas with no visible blur are left untouched (eyes of the green bird, for instance). This results into sharper retargeted scenes that can be shown within the limitations of the display. The remapping for the teaser image took two hours for a resolution of 1024×768, using our unoptimized Matlab code.

We have also fabricated a prototype multilayer display (Figure 9). This display is composed of five inkjet-printed transparency patterns spaced by clear acrylic sheets. The size of each layer is 60×45 mm, while each spacer has a thickness of 1/8". The transparencies are conventional films for office use and the printer is an Epson Stylus Photo 2200. This multilayer display supports 7×7 views within a field of view of 7° for both horizontal and vertical parallax. The patterns are generated with the computed tomography solver provided by Wetzstein et al. [11]. Notice the significant sharpening of the blue bird and, to a lesser extent, of the red bird. It should be noted that these are lab prototypes: scattering, inter-reflections between the acrylic sheets, and imperfect color reproduction with the desktop inkjet printer influence the overall quality of the physical results. In Figure 10, we show sharper, simulated results for the *dice* scene for a similar multilayer display.

We show additional results using more complex data sets, with varying degrees of depth and texture, and different object shapes and surface material properties. In particular, we use the Heidelberg light field archive[2], which includes ground-truth depth information. The scenes are optimized for a three-layer multilayer display, similar to the one shown in Figure 9. They have been optimized for a viewing distance of 0.5 m and have resolutions ranging from 768×768 to 1024×720. The weights used in the optimization are again $\mu_{DOF} = 10$ and $\mu_D = 0.003$. Figure 11 shows the results for the *papillon*, *buddha2* and *statue* data sets. Our algorithm recovers most of the high frequency content of the original scenes, lost by the physical limitations of the display. The anaglyph representations allow to compare the perceived depth of the original and the retargeted scenes

[2] http://hci.iwr.uni-heidelberg.de/HCI/Research/LightField/lf_archive.php

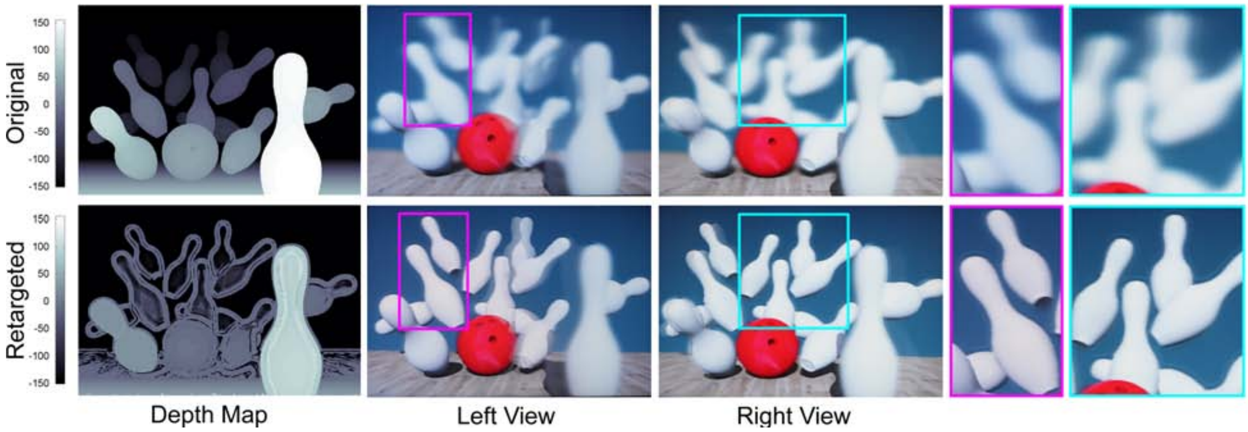

Figure 8: Additional results for commercial lenticular display (actual photographs). Top row: depth map, perspective from left, and perspective from right for original scene. Bottom row: depth map and similar perspectives for the retargeted scene. The slight double-view of some of the pins in the left view is due to interview cross-talk in the Toshiba display.

(please refer to the supplementary material for larger versions to ensure proper visualization). Figure 12 shows additional views of the *buddha2* and *statue* light fields.

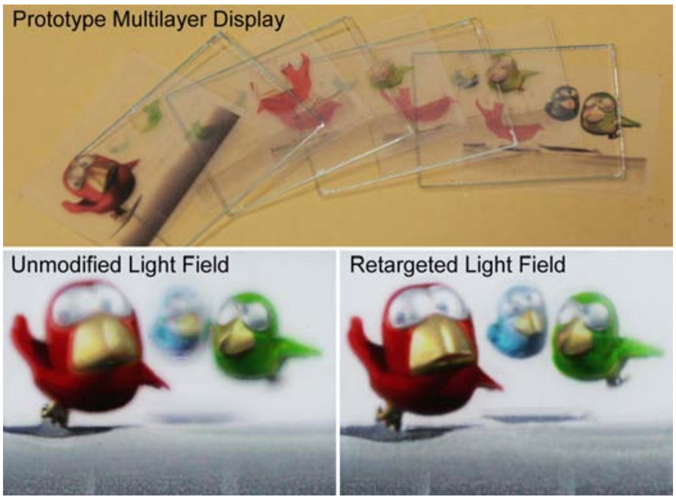

Figure 9: 3D content retargeting for multilayer light field displays (actual photographs). Even five attenuating layers (top) can only provide a limited depth of field for a displayed scene (bottom left). Our retargeting algorithm maps the multiview content into the provided depth budget (bottom right).

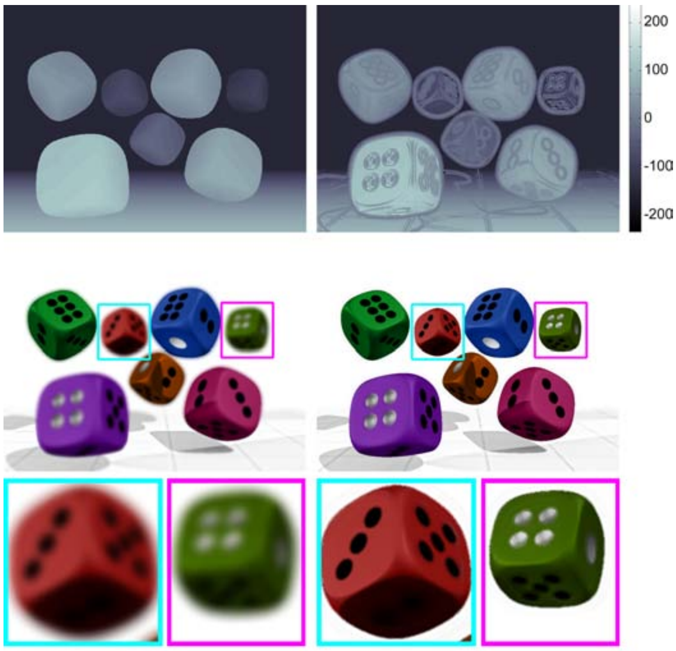

Figure 10: Results of simulations for a multilayer display (five layers). Top row: initial and retargeted depth. Middle row: initial and retargeted luminance. Bottom row: close-ups.

As shown in this section, our algorithm works well within a wide range of displays and data sets of different complexities. However, in areas of very high frequency content, the warping step may accumulate errors which end up being visible in the extreme views of the light fields. Figure 13 shows this: the *horses* data set contains a background made up of a texture containing printed text. Although the details are successfully recovered by our algorithm, the warping step cannot deal with the extremely high frequency of the text, and the words appear broken and illegible.

Finally, Figure 14 shows the result of applying our adapted model to the particular case of stereo retargeting, as described in Section 6.

8. Comparison to Other Methods

Our method is the first to specifically deal with the particular limitations of automultiscopic displays (depth vs. blur trade-off), and thus it is difficult to directly compare with others. However, we can make use of two recently published *objective* computational metrics, to measure distortions both in the observed 2D image fidelity, and in the perception of depth. This also provides an objective background to compare against existing approaches for stereoscopic disparity retargeting, for which

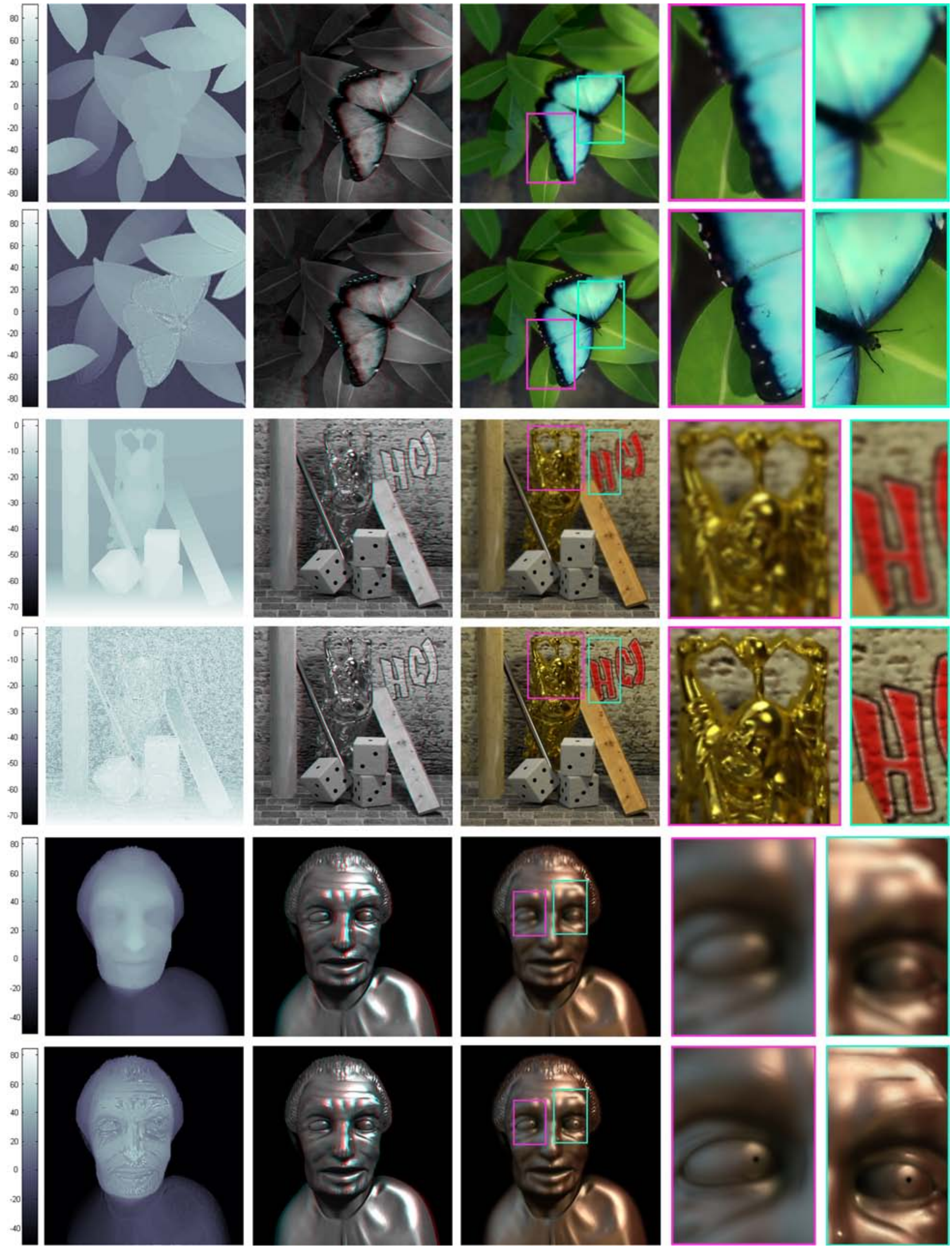

Figure 11: Results for the *papillon* (top), *buddha2* (middle) and *statue* (bottom) data sets from the Heidelberg light field archive. For each data set, the top row shows the original scene, while the bottom row shows our retargeted result. From left to right: depth map, anaglyph representation, central view image, and selected zoomed-in regions. Notice how our method recovers most of the high frequency details of the scenes, while preserving the sensation of depth (larger versions of the anaglyphs appear in the supplementary material). Note: please wear anaglyph glasses with cyan filter on left and red filter on right eye; for an optimal viewing experience please resize the anaglyph to about 10 cm wide in screen space and view it at a distance of 0.5 m.

alternative methods do exist.

Figure 12: Additional non-central views of the retargeted *buddha2* and *statue* light fields, with corresponding close-ups.

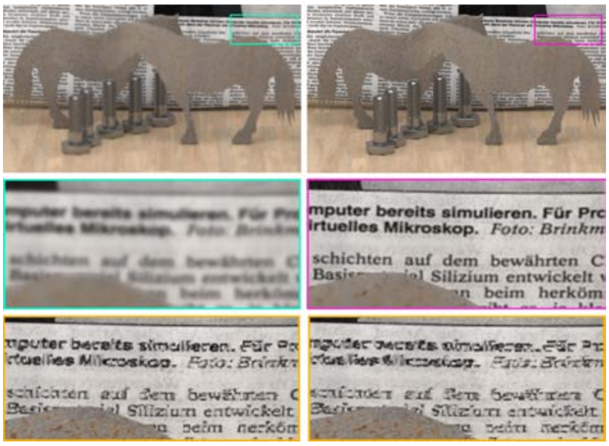

Figure 13: Results for the *horses* data set from the Heidelberg light field archive. Very high frequencies that have been initially cut off by the display (green box) are successfully recovered by our algorithm (pink). However, subsequent warping can introduce visible artifacts in those cases, which progressively increase as we depart from the central view of the light field. This progression is shown in the bottom row (yellow boxes).

Metrics: We need to measure *both* observed 2D image quality *and* resulting degradations in perceived depth. For image quality, numerous metrics exist. We rely on the HDR-VDP 2 calibration reports provided by Mantiuk and colleagues [30] in their website[3] , where the authors compare quality predictions from six different metrics and two image databases: LIVE [40] and TID2008 [41]. According to the prediction errors, reported as Spearman's correlation coefficient, multi-scale SSIM (MS-SSIM, [42]) performs best across both databases for the blurred image distortions observed in our application. The mapping function we use, $log(1-\text{MS-SSIM})$, yields the highest correlation for Gaussian blur distortions.

Fewer metrics exist to evaluate distortions in depth. We use the metric recently proposed by Didyk and colleagues to estimate the magnitude of the perceived disparity change between two stereo images [8]. The metric outputs a heat map of the differences between the original and the retargeted disparity maps in Just Noticeable Difference (JND) units.

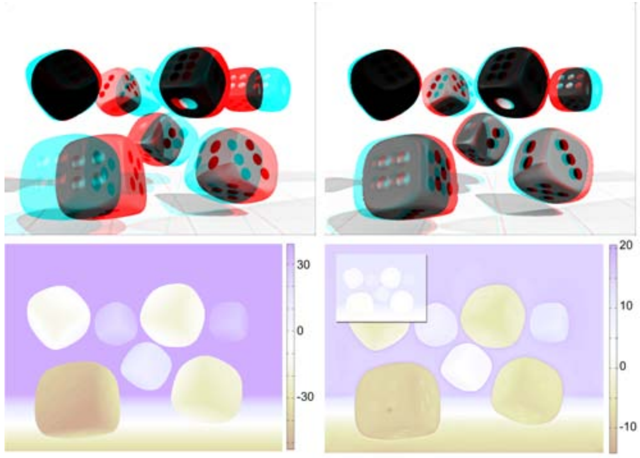

Figure 14: Retargeting for stereo content. *Left column:* Anaglyph and corresponding pixel disparity map of the original scene. For a common (around 0.5m) viewing distance on a desktop display, left and right images cannot be fused. *Right column:* Anaglyph and corresponding pixel disparity map of the retargeted scene. Images can now be fused without discomfort, and perception of depth is still present despite the aggressive depth compression. Note that the scales of the disparity maps are different for visualization purposes; the small inset shows the retargeted disparity map for the same scale as the original. Note: please wear anaglyph glasses with cyan filter on left and red filter on right eye; for an optimal viewing experience please resize the anaglyph to about 10 cm wide in screen space and view it at a distance of 0.5 m.

Alternative Methods: There is a large space of linear and non-linear global remapping operators, as well as of local approaches. Also, these operators can be made more sophisticated, for instance by incorporating information from saliency maps, or adding the temporal domain [6]. To provide some context to the results of the objective metrics, we compare our method with a representative subset of alternatives, including global operators, local operators, and a recent operator based on a perceptual model for disparity. In particular, we compare against six other results using different approaches for stereo retargeting: a linear scaling of pixel disparity (*linear*), a linear scaling followed by the addition of bounded Cornsweet profiles at depth discontinuities (*Cornsweet* [24])[4], a logarithmic remapping (*log*, see e.g. [6]), and the recently proposed remapping of disparity in a perceptually linear space (*perc. linear* [8]). For the last two, we present two results using different parameters. This selection of methods covers a wide range from very simple to more sophisticated.

The linear scaling is straightforward to implement. For the bounded Cornsweet profiles method, where profiles are carefully controlled so that they do not exceed the given disparity bounds and create disturbing artifacts, we choose $n = 5$ levels as suggested by the authors. For the logarithmic remapping, we

[3] http://hdrvdp.sourceforge.net/reports/2.1/quality_live/ http://hdrvdp.sourceforge.net/reports/2.1/quality_tid2008/

[4] In our tests, this consistently yielded better results than a naive application of unbounded Cornsweet profiles, as originally reported by Didyk and colleagues [24]

use the following expression, inspired by Lang et al. [6]:

$$\delta_o = K \cdot log(1 + s \cdot \delta_i), \tag{15}$$

where δ_i and δ_o are the input and output pixel disparities, s is a parameter that controls the scaling and K is chosen so that the output pixel disparities fit inside the allowed range. We include results for $s = 0.5$ and $s = 5$. Finally, for the perceptually linear method, disparity values are mapped via transducers into a perceptually linear space, and then linearly scaled by a factor k. The choice of k implies a trade-off between the improvement in contrast enhancement and how faithful to the original disparities we want to remain. We choose $k = 0.75$ and $k = 0.95$ as good representative values for both options respectively.

Comparisons: Some of the methods we compare against (*linear*, *Cornsweet* and *log*) require to explicitly define a minimum spatial cut-off frequency, which will in turn fix a certain target depth range. We run comparisons on different data sets and for a varied range of cut-off frequencies: For the *birds* scene, where the viewing distance is $v_D = 1.5$ m, we test two cut-off frequencies: $f_{cpmm} = 0.12$ cycles per mm ($f_{cpd} = 3.14$ cycles per degree), and $f_{cpmm} = 0.19$ ($f_{cpd} = 5.03$), the latter of which corresponds to remapping to the depth range which offers the maximum spatial resolution of the display (see DOF plots in Figure 16b). For the *statue*, *papillon* and *buddha2* scenes, optimized for a multilayer display with $v_D = 0.5$ m, we set the frequencies to $f_{cpmm} = 0.4$, 0.5 and 1.1, respectively (corresponding $f_{cpd} = 3.49$, 4.36 and 9.60). The frequencies are chosen so that they yield a fair compromise between image quality and perceived depth, given the trade-off between these magnitudes; they vary across scenes due to the different spatial frequencies of the image content in the different data sets.

Figure 15 shows a comparison to the results obtained with the other methods both in terms of image quality and of perceived depth for three different scenes from the Heidelberg data set (*papillon*, *buddha2*, and *statue*). Heat maps depict the error in perceived depth (in JNDs) given by Didyk et al.'s metric. Visual inspection shows that our method consistently leads to less error in perceived depth (white areas mean error below the 1 JND threshold). Close-ups correspond to zoomed-in regions from the resulting images obtained with each of the methods, where the amount of DOF blur can be observed (please refer to the supplementary material for the complete images). Our method systematically yields sharper images, even if it also preserves depth perception better. Only in one case, in the *statue* scene, perceptually linear remapping yields sharper results, but at the cost of a significantly higher error in depth perception, as the corresponding heat maps show.

To better assess the quality of the deblurring of the retargeted images, Figure 16a shows the MS-SSIM metric for the different methods averaged over the scenes tested, together with the associated standard error (we plot the absolute value of $log(1 - \text{MS-SSIM})$). We have added the result of the original image, without any retargeting method applied (N for *none* in the chart). Our method yields the best perceived image quality (highest MS-SSIM value), and as shown in Figure15, the lowest error in depth perception as well. This can be intuitively explained by the fact that our proposed multi-objective optimization (Eq. 12) explicitly optimizes *both* luminance and depth, whereas existing algorithms are either heuristic or take into account only one of the two aspects.

To further explore this image quality vs. depth perception trade-off, we have run the comparisons for the *birds* scene for two different cut-off spatial frequencies. Figure 16b shows comparisons of all tested algorithms for the *birds* scene retargeted for a lenslet-based display. For two of the methods, ours and the perceptually linear remapping (with $k = 0.75$ and $k = 0.95$), defining this minimum spatial frequency is not necessary. Error in depth for these is shown in the top row. For the other four methods (*linear*, *Cornsweet*, *log* $s = 0.5$, *log* $s = 5$), the cut-off frequency needs to be explicitly defined: we set it to two different values of $f_{cpmm} = 0.12$ and $f_{cpmm} = 0.19$, which correspond to an intermediate value and to remapping the content to the maximum spatial frequency of the display, respectively. The resulting error in depth is shown in the middle and bottom rows of Figure 16b. Error in perceived depth clearly increases as the cut-off frequency is increased. The bar graph at the top left of Figure 16b shows image quality results for $f_{cpmm} = 0.12$. Note that for $f_{cpmm} = 0.19$, the methods *linear*, *Cornsweet* and *log* yield perfectly sharp images (since we explicitly chose that frequency to remap to the maximum resolution of the display), but at the cost of large errors in perceived depth.

9. Conclusions and Future Work

Automultiscopic displays are an emerging technology with form factors ranging from hand-held devices to movie theater screens. Commercially successful implementations, however, face major technological challenges, including limited depth of field, resolution, and contrast. We argue that compelling multiview content will soon be widely available and tackle a crucial part of the multiview production pipeline: display-adaptive 3D content retargeting. Our computational depth retargeting algorithm extends the capabilities of existing glasses-free 3D displays, and deals with a part of the content production pipeline that will become commonplace in the future.

As shown in the paper, there is an inherent trade-off in automultiscopic displays between depth budget and displayed spatial frequencies (blur): depth *has to* be altered if spatial frequencies in luminance are to be recovered. This is not a limitation of our algorithm, but of the targeted hardware (Figure 3). Our algorithm aims at finding the best possible trade-off, so that the inevitable depth distortions introduced to improve image quality have a minimal perceptual impact. Therefore, the amount of blur (the cut-off frequency) in the retargeted scene depends on the actual visibility of the blur in a particular area, according to the CSF. Should the user need to further control the amount of defocus deblurring, it could be added to the optimization in the form of constraints over the depth values according to the

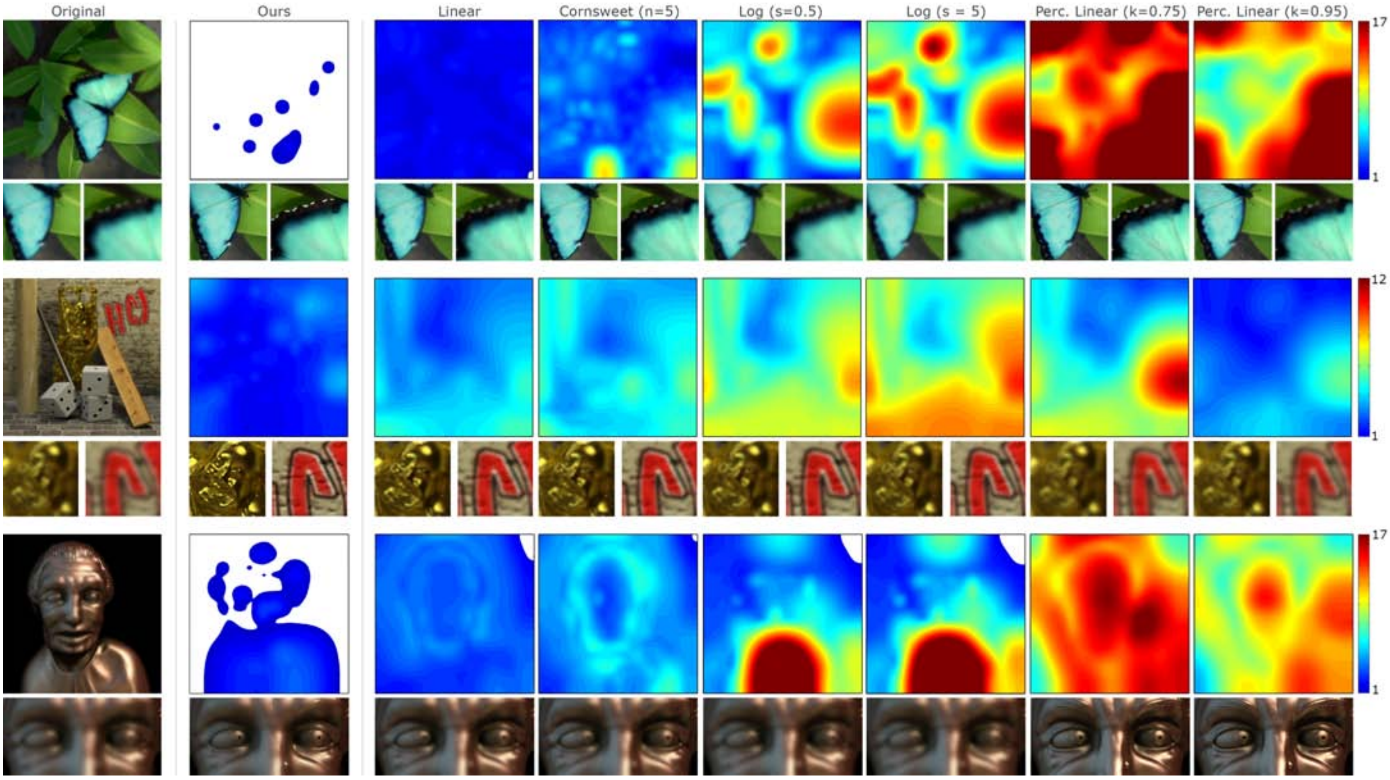

Figure 15: Comparison against other methods for three different scenes from the Heidelberg light field archive. From top to bottom: *papillon* (f_{cpmm} = 0.4, f_{cpd} = 3.49), *buddha2* (f_{cpmm} = 1.1, f_{cpd} = 9.60), and *statue* (f_{cpmm} = 0.5, f_{cpd} = 4.36). Errors in depth are shown as heat maps (lower is better) according to the metric by Didyk and colleagues [8]; white areas correspond to differences below one JND. Viewing distance is 0.5 m.

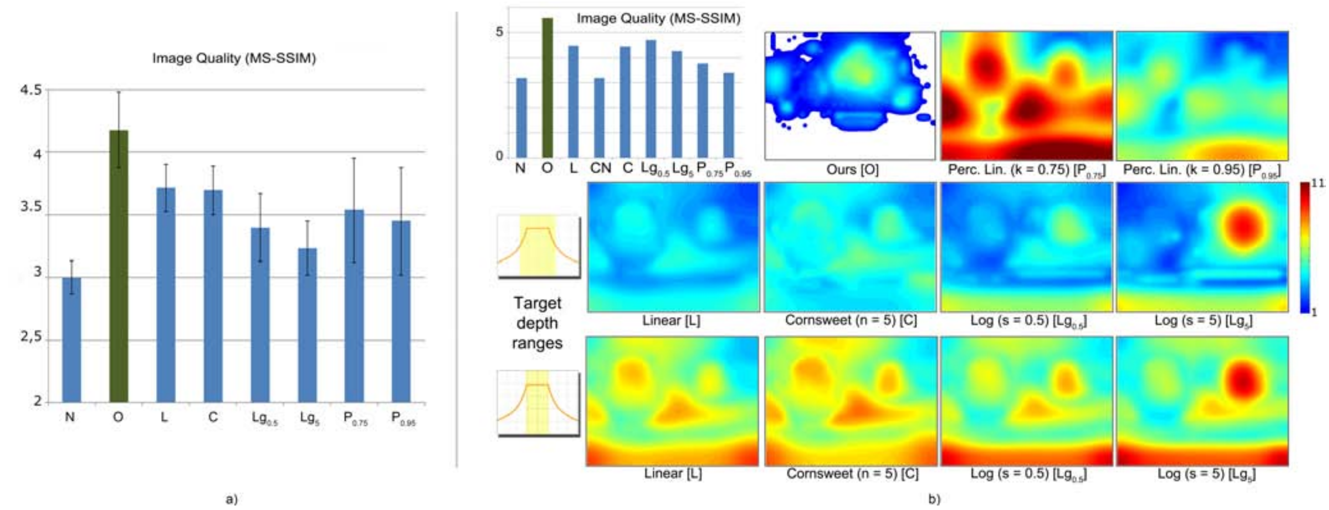

Figure 16: (a) Comparison of average luminance quality (lack of blur) according to the MS-SSIM metric for all the data sets used in this comparisons (higher is better). (b) Comparison against other methods for the *birds* scene, for two different cut-off frequencies. Top row, from left to right: resulting image quality as predicted by MS-SSIM for f_{cpmm} = 0.12, and error in depth for the two methods that do not require providing a target depth range. Middle row: error in depth for the three methods requiring a target depth range, for a cut-off frequency f_{cpmm} = 0.12 (f_{cpd} = 3.14). The smaller image represents the depth vs. cut-off frequency function of the display, with the target depth range highlighted in yellow. Bottom row: same as middle row for a cut-off frequency f_{cpmm} = 0.19 (f_{cpd} = 5.03), corresponding to the maximum spatial frequency allowed by the display (flat region of the DOF function). Errors in depth are shown as heat maps (lower is better) according to Didyk et al's metric [8]; white areas correspond to differences below one JND. Note the intrinsic trade-off between image quality and depth perception for the methods requiring a specific target depth range: when remapping to the maximum spatial frequency of the display, error in perceived depth significantly increases. Viewing distance is 1.5 m.

corresponding DOF function.

We have demonstrated significant improvements in sharpness and contrast of displayed images without compromising the perceived three-dimensional appearance of the scene, as our results and validation with objective metrics show. For the special case of disparity retargeting in stereoscopic image pairs, our method is the first to handle display-specific non-dichotomous zones of comfort: these model the underlying physical and physiological aspects of perception better than binary zones used in previous work. In the supplementary video, we also show an animated sequence for retargeted content. It is shown as an anaglyph, so it can be seen in 3D on a regular display. Although the frames of this video clip have been processed separately, our algorithm provides temporally stable retargeting results.

A complete model of depth perception remains an open problem. One of the main challenges is the large number of cues that our brain uses when processing visual information, along with their complex interactions [43, 44]. A possible avenue of future work would be to extend the proposed optimization framework by including perceptual terms modeling human sensitivity to accommodation, temporal changes in displayed images, sensitivity of depth perception due to motion parallax or the interplay between different perceptual cues. However, this is not trivial and will require significant advances in related fields. Another interesting avenue of future work would be to extend our optimization framework to deal with all the views in the light field, thus exploiting angular resolution.

We hope that our work will provide a foundation for the emerging multiview content production pipeline and inspire others to explore the close relationship between light field acquisition, processing, and display limitations in novel yet unforeseen ways. We believe bringing the human visual system into the design pipeline [45, 46] is a great avenue of future work to overcome current hardware limitations in all areas of the imaging pipeline, from capture to display.

References

[1] Ives FE. Parallax stereogram and process of making same. U.S. Patent 725,567; 1903.

[2] Lippmann G. Épreuves réversibles donnant la sensation du relief. Journal of Physics 1908;7:821–5.

[3] Holografika . HoloVizio C80 3D Cinema System. `http://www.holografika.com`; 2012.

[4] Banterle F, Artusi A, Aydin T, Didyk P, Eisemann E, Gutierrez D, et al. Multidimensional image retargeting. ACM SIGGRAPH Asia Course Notes; 2011.

[5] Kim M, Lee S, Choi C, Um GM, Hur N, Kim J. Depth Scaling of Multiview Images for Automultiscopic 3D Monitors. In: 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video. 2008,.

[6] Lang M, Hornung A, Wang O, Poulakos S, Smolic A, Gross M. Nonlinear disparity mapping for stereoscopic 3d. ACM Trans Graph (SIGGRAPH) 2010;29:75:1–75:10.

[7] Kim C, Hornung A, Heinzle S, Matusik W, Gross M. Multi-perspective stereoscopy from light fields. ACM Trans Graph 2011;30:190:1–190:10.

[8] Didyk P, Ritschel T, Eisemann E, Myszkowski K, Seidel HP. A perceptual model for disparity. ACM Trans Graph (SIGGRAPH) 2011;30:96:1–96:10.

[9] Didyk P, Ritschel T, Eisemann E, Myszkowski K, Seidel HP, Matusik W. A luminance-contrast-aware disparity model and applications. ACM Transactions on Graphics (Proceedings SIGGRAPH Asia 2012) 2012;31.

[10] Zwicker M, Matusik W, Durand F, Pfister H. Antialiasing for Automultiscopic 3D Displays. In: Proc. EGSR. 2006,.

[11] Wetzstein G, Lanman D, Heidrich W, Raskar R. Layered 3D: Tomographic Image Synthesis for Attenuation-based Light Field and High Dynamic Range Displays. ACM Trans Graph (SIGGRAPH) 2011;30:1–11.

[12] Rogers B, Graham M. Similarities between motion parallax and stereopsis in human depth perception. Vision Research 1982;22:261–70.

[13] Braun M, Leiner U, Ruschin D. Evaluating motion parallax and stereopsis as depth cues for autostereoscopic displays. In: Proc. SPIE 7863, Stereoscopic Displays and Applications XXII; vol. 7863. 2011,.

[14] Urey H, Chellappan KV, Erden E, Surman P. State of the Art in Stereoscopic and Autostereoscopic Displays. Proc IEEE 2011;99:540–55.

[15] Wetzstein G, Lanman D, Gutierrez D, Hirsch M. Computational Displays. ACM SIGGRAPH Course Notes; 2012.

[16] Lipton L. Foundations of the Stereoscopic Cinema: a study in depth. Van Nostrand Reinhold; 1982.

[17] Mendiburu B. 3D Movie Making: Stereoscopic Digital Cinema from Script to Screen. Focal Press; 2009.

[18] Jones G, Lee D, Holliman N, Ezra D. Controlling Perceived Depth in Stereoscopic Images. In: SPIE Stereoscopic Displays and Virtual Systems VIII; vol. 4297. 2001, p. 42–53.

[19] Oskam T, Hornung A, Bowles H, Mitchell K, Gross M. OSCAM - optimized stereoscopic camera control for interactive 3D. ACM Trans Graph (SIGGRAPH) 2011;30:189:1–:=:8.

[20] Heinzle S, Greisen P, Gallup D, Chen C, Saner D, Smolic A, et al. Computational stereo camera system with programmable control loop. ACM Trans Graph 2011;30:94:1–94:10.

[21] Koppal SJ, Zitnick CL, Cohen MF, Kang SB, Ressler B, Colburn A. A viewer-centric editor for 3d movies. IEEE Computer Graphics and Applications 2011;31:20–35.

[22] Ward B, Kang SB, Bennett EP. Depth director: A system for adding depth to movies. IEEE Computer Graphics and Applications 2011;31:36–48.

[23] Birklbauer C, Bimber O. Light-Field Retargeting. Computer Graphics Forum (Proc Eurographics) 2012;31:295–303.

[24] Didyk P, Ritschel T, Eisemann E, Myszkowski K, Seidel HP. Apparent stereo: The cornsweet illusion can enhance perceived depth. In: Proc. Electronic Imaging. 2012, p. 1–12.

[25] Seitz SM, Kim J. The space of all stereo images. Int J Comput Vision 2002;48:21–38.

[26] Masia B, Wetzstein G, Aliaga C, Raskar R, Gutierrez D. Perceptually-optimized content remapping for automultiscopic displays. In: ACM SIGGRAPH 2012 Posters. SIGGRAPH '12; 2012, p. 63:1–:=:.

[27] Wetzstein G, Lanman D, Hirsch M, Raskar R. Tensor Displays: Compressive Light Field Synthesis using Multilayer Displays with Directional Backlighting. ACM Trans Graph (SIGGRAPH) 2012;31:1–11.

[28] Hasinoff SW, Kutulakos KN. Multiple-aperture photography for high dynamic range and post-capture refocusing. Tech. Rep.; University of Toronto, Dept. of Computer Science; 2009.

[29] Mantiuk R, Daly S, Kerofsky L. Display adaptive tone mapping. ACM Trans Graph (SIGGRAPH) 2008;27:68:1–68:10.

[30] Mantiuk R, Kim KJ, Rempel AG, Heidrich W. HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. ACM Trans Graph (SIGGRAPH) 2011;30:40:1–40:13.

[31] Barten PGJ. Contrast Sensitivity of the Human Eye and its Effects on Image Quality. SPIE Press; 1999.

[32] Simoncelli EP, Freeman WT. The Steerable Pyramid: A Flexible Architecture for Multi-Scale Derivative Computation. In: Proc. ICIP. 1995, p. 444–7.

[33] Bradshaw MF, Rogers BJ. Sensitivity to horizontal and vertical corrugations defined by binocular disparity. Vision Research 1999;39:3049–56.

[34] Anstis SM, Howard IP. A Craik-O'Brien-Cornsweet Illusion for Visual Depth. Vision Research 1978;18:213–7.

[35] Coleman T, Li Y. A reflective newton method for minimizing a quadratic function subject to bounds on some of the variables. SIAM Journal on Optimization 1996;6:1040–58.

[36] Didyk P, Ritschel T, Eisemann E, Myszkowski K, Seidel HP. Adaptive Image-space Stereo View Synthesis. In: Proc. VMV. 2010, p. 299–306.

[37] Luo SJ, Shen IC, Chen BY, Cheng WH, Chuang YY. Perspective-aware warping for seamless stereoscopic image cloning. Transactions on Graphics (Proceedings of ACM SIGGRAPH Asia 2012) 2012;31:182:1–:=:8.

[38] Niu Y, Feng WC, Liu F. Enabling warping on stereoscopic images. ACM

Trans Graph 2012;31:183:1–:=:7.

[39] Shibata T, Kim J, Hoffman DM, Banks MS. The zone of comfort: Predicting visual discomfort with stereo displays. Journal of Vision 2011;11:1–29.

[40] Sheikh H, Sabir M, Bovik A. A statistical evaluation of recent full reference image quality assessment algorithms. IEEE Trans Image Processing 2006;15:3440–51.

[41] Ponomarenko N, Lukin V, Zelensky A, Egiazarain K, Carli M, Battisti F. Tid2008-a database for evaluation of full-reference visual quality assessment metrics. Advances of Modern Radioelectronics 2009;10:30–45.

[42] Zhou Wang EPS, Bovik AC. Multi-scale Structural Similarity for Image Quality Assessment. In: Proc. IEEE Asilomar Conference on Signals, Systems and Computers. 2003,.

[43] Cutting JE, Vishton PM. Perception of Space and Motion; chap. Perceiving Layout and Knowing Distances: The integration, relative potency, and contextual use of different information about depth. Academic Press; 1995,.

[44] Hillis JM, Watt SJ, Landy MS, Banks MS. Slant from texture and disparity cues: Optimal cue combination. Journal of Vision 2004;4:967–92.

[45] Didyk P, Eisemann E, Ritschel T, Myszkowski K, Seidel HP. Apparent display resolution enhancement for moving images. ACM Transactions on Graphics (Proceedings SIGGRAPH 2010, Los Angeles) 2010;29(4).

[46] Masia B, Presa L, Corrales A, Gutierrez D. Perceptually-optimized coded apertures for defocus deblurring. Computer Graphics Forum 2012;31(6).

Appendix A. Objective Function and Analytical Derivatives

In this section we go through the mathematical expressions of the two terms of the objective function in detail. We also include their derivatives, necessary for computing the analytical Jacobian used in the optimization process.

Appendix A.1. Term 1: Optimizing Luminance and Contrast

This term, as shown in Equation (8) of the main text, has the following form:

$$T_1 = \omega_{CSF}\left(\rho_S\left(L_{orig}\right) - \rho_S\left(\phi_b\left(L_{orig}, d\right)\right)\right) \tag{A.1}$$

Note that this expression yields a vector of length N_{pyr} (N_{pyr} being the number of pixels in the pyramid $\rho_S\left(L_{orig}\right)$ or $\rho_S\left(\phi_b\left(L_{orig}, d\right)\right)$), which is a vector of differences with respect to the target luminance L_{orig}, weighted by contrast sensitivity values. This vector of errors thus contains the residuals that `lsqnonlin` optimizes for the depth of field term. The weighting factor μ_{DOF} is left out of this derivation for the sake of simplicity, since it is just a product by a constant both in the objective function term and in its derivatives. This is valid also for the second term of the objective function.

Since the multi-scale decomposition is a linear operation, we can write:

$$T_1 = \omega_{CSF}\left(M_S \cdot L_{orig} - M_S \cdot \phi_b\left(L_{orig}, d\right)\right) \tag{A.2}$$

where M_S is a matrix of size $N_{pyr} \times N_{im}$, N_{im} being the number of pixels in the luminance image L_{orig}. Substituting the blurring function $\phi_b(\cdot,\cdot)$ by its actual expression

$$\frac{\partial T_{1,i}}{\partial d} = \omega_{CSF,i}\left(-M_{S,i} \cdot (L_{orig} * \frac{\partial k(d)}{\partial d})\right), \tag{A.3}$$

where $M_{S,i}$ is the $i-th$ row of M_S. The derivative of the kernels $k(d)$ is:

$$\frac{\partial k(d)}{\partial d} = \frac{\left(exp(-\frac{x_i^2+y_i^2}{2(\sigma(d))^2})\right)\left(\frac{(x_i^2+y_i^2)4\sigma(d)\frac{\partial\sigma(d)}{\partial d}}{(2(\sigma(d))^2)^2}\right)\sum_j^K\left[exp(-\frac{x_j^2+y_j^2}{2(\sigma(d))^2})\right]}{\left(\sum_j^K\left[exp(-\frac{x_j^2+y_j^2}{2(\sigma(d))^2})\right]\right)^2} - \tag{A.4}$$

$$\frac{\sum_j^K\left[\left(exp(-\frac{x_j^2+y_j^2}{2(\sigma(d))^2})\right)\left(\frac{(x_j^2+y_j^2)4\sigma(d)\frac{\partial\sigma(d)}{\partial d}}{(2(\sigma(d))^2)^2}\right)\right]\left(exp(-\frac{x_i^2+y_i^2}{2(\sigma(d))^2})\right)}{\left(\sum_j^K\left[exp(-\frac{x_j^2+y_j^2}{2(\sigma(d))^2})\right]\right)^2}.$$

The derivative of the standard deviation σ is straightforward, knowing $\partial(f_\xi(d))/\partial d$. As described in the main text, the expression for $f_\xi(d)$ depends on the type of automultiscopic display. For a conventional display [10]:

$$f_\xi(d) = \begin{cases} \frac{f_0}{N_a}, & for\,|d| + (h/2) \leq N_a h \\ (\frac{h}{(h/2)+|d|})f_0, & otherwise \end{cases}, \tag{A.5}$$

where N_a is the number of angular views, h represents the thickness of the display and $f_o = 1/(2p)$ is the spatial cut-off frequency of a mask layer with a pixel of size p. For multilayered displays, the upper bound on the depth of field for a display of N layers is [11]:

$$f_\xi(d) = N f_0 \sqrt{\frac{(N+1)h^2}{(N+1)h^2 + 12(N-1)d^2}}. \tag{A.6}$$

The derivatives are as follows:

$$\frac{\partial f_\xi(d)}{\partial d} = \begin{cases} 0, & for\,|d| + (h/2) \leq N_a h \\ (\frac{-hd/|d|}{((h/2)+|d|)^2})f_0, & otherwise \end{cases} \tag{A.7}$$

for a conventional display and

$$\frac{\partial f_\xi(d)}{\partial d} = N f_0 \frac{12\sqrt{N+1}(N-1)hd}{((N+1)h^2 + 12(N-1)d^2)^{3/2}}. \tag{A.8}$$

for a multilayered display.

Appendix A.2. Term 2: Preserving Perceived Depth

This term, introduced in Equation 10 of the main text, is modeled as follows:

$$T_2 = \omega_{BD}\left(\rho_L\left(\phi_\upsilon\left(D_{orig}\right)\right) - \rho_L\left(\phi_\upsilon(d)\right)\right) \tag{A.9}$$

Again, since the multi-scale decomposition is a linear operation, we write:

$$T_2 = \omega_{BD}\left(M_L \cdot \phi_\upsilon\left(D_{orig}\right) - M_L \cdot \phi_\upsilon(d)\right) \tag{A.10}$$

where M_L is a matrix of size $N_{dpyr} \times N_d$, N_d being the number of pixels in the depth map D_{orig}. Taking the derivative with respect to d yields the following expression for each element $T_{2,i}$ of the residuals vector for this term:

$$\frac{\partial T_{2,i}}{\partial d} = \omega_{BD,i}\left(-M_{L,i} \cdot \frac{\partial \phi_\upsilon(d)}{\partial d}\right), \tag{A.11}$$

where $M_{L,i}$ is the $i-th$ row of M_L. As explained in the main text, $\phi_v(d)$ converts depth d_P of a point P into vergence v_P. This, given the viewing distance v_D and the interaxial distance e, is done using function $\phi_v(\cdot)$:

$$\phi_v(d) = acos\left(\frac{\mathbf{v_L} \cdot \mathbf{v_R}}{\|\mathbf{v_L}\| \, \|\mathbf{v_R}\|}\right), \tag{A.12}$$

where vectors $\mathbf{v_L}$ and $\mathbf{v_R}$ have their origins in P and end in the eyes (please also see Figure 6 in the main text). Placing the coordinate origin in the center of the screen (z-axis normal to the screen, x-axis in the horizontal direction) we can rewrite the previous equation for a point $P = (x_i, y_i, d_i)$ as:

$$v_d = \phi_v(d) = acos\left(\frac{\kappa}{\sqrt{\eta}\sqrt{\zeta}}\right), \tag{A.13}$$

where:

$\kappa = (x_L - x_i)(x_R - x_i) + (v_D - d_i)^2,$

$$\eta = (x_L - x_i)^2 + (v_D - d_i)^2,$$

$$\zeta = (x_R - x_i)^2 + (v_D - d_i)^2.$$

Finally, differentiating Equation A.13 with respect to depth:

$$\frac{\partial \phi_v(d)}{\partial d} = -\left(1 - \left(\frac{\kappa}{\sqrt{\eta}\sqrt{\zeta}}\right)^2\right)^{-1/2} \cdot \left(\frac{-2(v_D - d_i)\sqrt{\eta}\sqrt{\zeta} - \kappa\Psi(d_i)}{\eta\zeta}\right)$$

where $\Psi(d_i)$ is as follows:

$$\Psi(d_i) = -d_i(v_D - d_i)\eta^{-1/2}\zeta^{1/2} - d_i(v_D - d_i)\zeta^{-1/2}\eta^{1/2}$$

Appendix B. A Dichotomous Zone of Comfort

As explained in the paper, Equation B.1 describes our objective function for the simplified case of stereo remapping:

$$\left\|\omega_{BD}\left(\rho_L\left(\phi_v\left(D_{orig}\right)\right) - \rho_L\left(\phi_v(d)\right)\right)\right\|_2^2 + \mu_{CZ}\|\varphi(d)\|_2^2, \tag{B.1}$$

where $\varphi(\cdot)$ is a function mapping depth values to visual discomfort. To incorporate a dichotomous model (such as those shown in cyan in Figure 7 for different devices and viewing distances v_D), instead of the non-dichotomous model described in the paper (shown in orange in the same figure), we can define a binary indicator function, such as

$$\varphi_{dc}(d) = \begin{cases} 0 & \text{for} \quad d_{comfort}^{min} \leq d \leq d_{comfort}^{max} \\ \infty & \text{otherwise} \end{cases} \tag{B.2}$$

For a practical, numerically-robust implementation, a smooth function that approximates Equation B.2 is preferable, ensuring C^1 continuity. Our choice for such a function is the Butterworth function which is commonly used as a low-pass filter in signal processing:

$$\varphi_{bf}(d) = 1 - \sqrt{\frac{1}{1 + (\gamma d)^{2s}}} \tag{B.3}$$

where γ controls the position of the cut-off locations and s the slope of such cut-offs.