Content-Aware Display Adaptation and Interactive Editing for Stereoscopic Images

Che-Han Chang, Chia-Kai Liang, and Yung-Yu Chuang

***Abstract*—We propose a content-aware stereoscopic image display adaptation method which simultaneously resizes a binocular image to the target resolution and adapts its depth to the comfort zone of the display while preserving the perceived shapes of prominent objects. This method does not require depth information or dense correspondences. Given the specification of the target display and a sparse set of correspondences, our method efficiently deforms the input stereoscopic images for display adaptation by solving a least-squares energy minimization problem. This can be used to adjust stereoscopic images to fit displays with different real estates, aspect ratios, and comfort zones. In addition, with slight modifications to the energy function, our method allows users to interactively adjust the sizes, locations, and depths of the selected objects, giving users aesthetic control for depth perception. User studies show that the method is effective at editing depth and reducing occurrences of diplopia and distortions.**

***Index Terms*—Content-aware image retargeting, stereoscopic image editing, depth adaptation.**

I. Introduction

The rapid deployment of stereoscopic equipment like displays and cameras will soon lead to a demand for users to be able to manipulate stereoscopic media similar to the way they manipulate 2D media. Stereoscopic media delivers not only an additional dimension and added enjoyment, but also additional challenges and constraints in creating a comfortable and enjoyable 3D experience. Because they do not address these constraints, naïve extensions of existing 2D media manipulation algorithms usually fail to deliver a comfortable 3D viewing experience. Thus, nontrivial adjustments are often required to accommodate new constraints and take advantage of new opportunities.

Most stereoscopic displays rely on the principle of *stereopsis*; human eyes are horizontally separated and the separation causes an interocular difference in the images projected onto the left and right retinas. When each eye is presented with the proper image, humans perceive depth by fusing the left and the right images. The fusibility of stereoscopic images depends not only on properly calibrated displays but also depends heavily on perfect matches between the left and right images. Mismatches in image pairs, or *binocular asymmetries*, can lead to serious viewing discomfort. In severe cases, the user experiences *diplopia* (double vision) and 3D scene perception is totally disrupted or highly inaccurate. Even if the user is able to perceive a consistent 3D view, the effort required to resolve conflicts caused by binocular imperfections can lead to serious fatigue, eyestrain, and headache, and may reduce the sense of realism [1], [2]. Such mismatch is often caused by asymmetrical optical geometry or photometric characteristics [3]. For example, the viewer may experience viewing discomfort if the left and the right images are misaligned horizontally.

Due to the diversity among display resolutions and aspect ratios, similar to 2D media, binocular images[1] require adaptation to be displayed properly on different devices. In addition to adapting to the device resolution and aspect ratio (retargeting along the x and y directions on the screen plane), for stereoscopic displays, we often must adapt images to its comfort zone (that is, depth adaptation along the z direction perpendicular to the display). In addition to adapting to different displays, depth adaptation is often required for binocular images with excessive depth ranges. However, existing solutions are not content-aware and can lead to noticeable object shape distortion. In this paper we present a *content-aware stereoscopic image display adaptation and editing* method, in which we perform image retargeting and depth adaptation simultaneously without dense correspondences. Our method is inspired by recent warping-based image manipulation methods which have been shown successful in many different problems, such as image/video retargeting [4], [5], video stabilization [6], artistic perspective manipulation [7] and wide-angle projection [8], just to name a few. The key idea is that as long as the warped results of salient regions are close to the physically-correct constraints or user intention, the distortions introduced in the non-salient regions are unnoticeable to the user.

Our method first detects a sparse set of robust correspondence points and then optimizes the warping fields of the image pair according to the target display parameters, correspondence constraints, and other constraints that prevent the results from distortions. Our method can achieve various retargeting scenarios, including changing the display size, aspect ratio, allowable depth range, and viewing configuration. It can also achieve effects not supported in traditional depth adaptation methods, such as changes to the scene depth that do not affect its scale. In addition, by modeling the user interaction as constraints, our system can be extended to an interactive stereoscopic image editing system. The user can specify the transformation of the disparity/depth values, and our system accordingly warps the input to generate a new stereoscopic image. The user can also select a single object and specify its position, depth, or even explicit 3D location. Our system automatically identifies the depths of other regions

Che-Han Chang, Chia-Kai Liang and Yung-Yu Chuang are with National Taiwan University, Taipei, Taiwan. (e-mail: frank@cmlab.csie.ntu.edu.tw, liangck@gmail.com, cyy@csie.ntu.edu.tw).

[1]We focus on stereoscopic images with two views and call them stereoscopic images and binocular images interchangeably.

and warps the input to match user's intention. The resultant system is the first content-aware system to simultaneously allow retargeting, depth adaptation, and interactive editing of stereoscopic images.

II. Related work

Stereoscopy and 3D Cinematography. Stereopsis was first described by Charles Wheatstone in 1838 [9]. Along with the development of 3D cinema, tremendous effort has been put into better understanding the biological and physiological foundation of stereopsis, such as the study of the stereo asymmetry effect [2] and visual discomfort [3]. There are editing tools for stereoscopic cinema [10], but most of them directly manipulate the disparity maps without high-level parameters such as eye positions. Recently, Koppal *et al.* proposed a viewer-centric editor for stereoscopic cinema which allows manipulation using stereo parameters such as interocular distance, field of view, and location [11]. However, their method requires accurate, dense depth maps, which are usually difficult to obtain for general footage.

Media Retargeting. The problem of 2D image and video retargeting, that is, adapting the images or videos for displays with different sizes and aspect ratios, has received considerable attention these years. While traditional scaling and cropping methods can easily cause significant distortions or information loss, modern content-aware approaches take into account the saliency distribution of the image and attempt to keep the salient features uncontaminated. These approaches can roughly be categorized as discrete approaches or continuous approaches [12].

The seam carving method [13] is a well-known discrete approach that uses dynamic programming to find the optimal seam to be removed in an image according to its saliency map. A seam is a path of pixels from top to bottom or side to side. Rubinstein *et al.*improved the original method and extended it for video resizing [14]. However, because of their discrete nature, those approaches do not preserve structured objects well, and lead to disturbing artifacts.

For continuous approaches, several warping-based methods have been proposed [4], [15], [16]. These methods treat retargeting as a mesh deformation/warping problem; prominent regions are constrained so that their shapes are preserved as much as possible while less salient areas are allowed to be distorted more. The optimal warping field is usually obtained by minimizing certain energy functions. A direct application of these 2D content-aware retargeting algorithms to binocular images could, however, lead to visual discomfort because the binocular disparity cues in the input are not properly preserved. Moreover, stereoscopic content introduces an additional retargeting dimension along the depth axis.

Depth Adaptation. For retargeting along the depth axis, or controlling depth perception in the 3D content, researchers in the stereoscopic display community have proposed a variety of techniques, such as false eye separation, alpha-false eye separation, image scaling, image shifting, view scaling, etc. [1] Unfortunately, none of these methods is content-aware, and hence they may cause large distortions on the image plane. Because most methods use global image transformations, they have limited control over depths or disparities (see the next section for more details). For example, Kim suggested using a uniform adaptation that scales the image uniformly [17]. However, this can lead to distortion of the object shape if the horizontal and vertical scaling factors are different. Moreover, the perceived depth range varies with the scaling factor.

Recently, Lang *et al.* [18] developed a nonlinear disparity mapping system for perceived depth change of stereoscopic videos. As does our method, their method uses image warping with disparity constraints to manipulate input images; also, both methods take into account depth adaptation applications. However, in terms of the problems to be solved, their method can handle videos, but is limited to altering the perceived depth. While our method deals only with images, it simultaneously supports image retargeting, depth adaptation, and interactive depth editing for stereoscopic images. More specifically, Lang *et al.* focus on the central aspects of disparity in stereoscopy and the resulting requirements for stereoscopic content production and display, and develop several disparity mapping operations for post-processing. However, they do not take into account the distortions when the display aspect ratio is changed; their depth adaptation method presumably fails in this situation. Moreover, they do not attempt to explicitly map between disparity values and 3D coordinates in their formulation. This could lead to the miniaturization or gigantism effects that will be discussed in Section III. In contrast, our method directly modifies the perceived depths and shapes through a more natural explicit mapping.

Technically, we have made the following contributions. First, we provide a mapping between the perceived depths and the disparity values, which allows for more natural depth manipulation. Second, we show that, coupled with image warping, it is often sufficient to use sparse but accurate feature correspondences to alter depths or sizes. Third, we provide a mechanism to preserve relative disparities by automatically finding a similarity transformation of disparities.

III. Background

In this section, we first describe the basic model for stereo vision, in particular the relationship between perceived depth and image disparity. We then formulate the problem of stereoscopic image retargeting and discuss why linear scaling and other content-aware image retargeting methods do not solve this problem.

A. Stereo Vision

Fig. 1 shows a typical viewing configuration for stereoscopic displays. Here L and R denote the left and right eye of the viewer, respectively, e is the interocular distance between two eyes (this averages about 6.5 cm for adults), and D is the viewing distance to the screen. Without loss of generality, we assume that the eyes are aligned on the x-axis of the world coordinate and the origin is their midpoint. Note that in this paper we only deal with the viewing of existing stereoscopic images. Our purpose is to change the apparent depth (the perceived depth) to the viewer and not the real depth, which

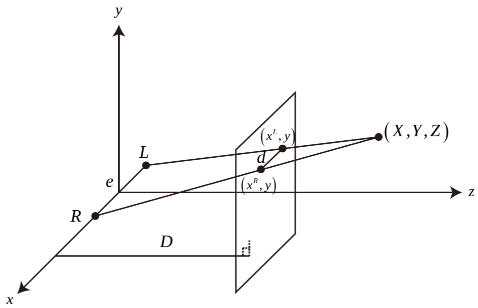


Fig. 1. The viewing configuration for stereoscopic displays.

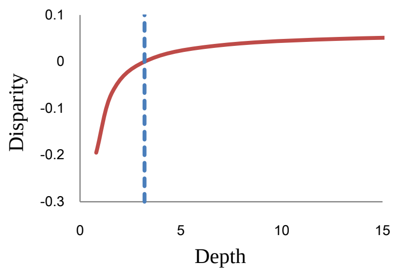


Fig. 2. The mapping between the disparity and the depth value. Here we use $e = 0.065$ m and $D = 3.2$ m in Eq. 6 (dotted line).

depends on the camera. For this, only the interocular distance matters and not the baseline distance, that is, the distance between two cameras during image capture.

A stereoscopic display delivers two different images to two eyes, and the viewer's brain fuses these images to achieve 3D perception. Therefore, to have a perception of point P at $[X_p, Y_p, Z_p]^T$ in 3D space, its projection is $p^L = [x_p^L, y_p]^T$ on the left image and $p^R = [x_p^R, y_p]^T$ on the right image, where

$$x_p^L = \left(X_p + \frac{e}{2}\right)\frac{D}{Z_p} - \frac{e}{2}, \tag{1}$$

$$x_p^R = \left(X_p - \frac{e}{2}\right)\frac{D}{Z_p} + \frac{e}{2}, \text{ and} \tag{2}$$

$$y_p = Y_p \frac{D}{Z_p}. \tag{3}$$

The horizontal shift of pixel p between the left and right eyes, $d_p = x_p^R - x_p^L$, is usually denoted as *disparity* and is related to its depth Z_p by

$$d_p = x_p^R - x_p^L = e\left(1 - \frac{D}{Z_p}\right). \tag{4}$$

Similarly, given the two corresponding points $p^L = [x_p^L, y_p]^T$ and $p^R = [x_p^R, y_p]^T$ on the left and the right images, the viewer perceives a 3D point P at $[X_p, Y_p, Z_p]^T$:

$$[X_p, Y_p, Z_p]^T = \frac{e}{e - d_p}\left[\frac{x_p^L + x_p^R}{2}, y_p, D\right]^T. \tag{5}$$

In particular, the perceived depth Z_p of the point is related to the disparity d_p as

$$Z_p = \frac{eD}{e - d_p}. \tag{6}$$

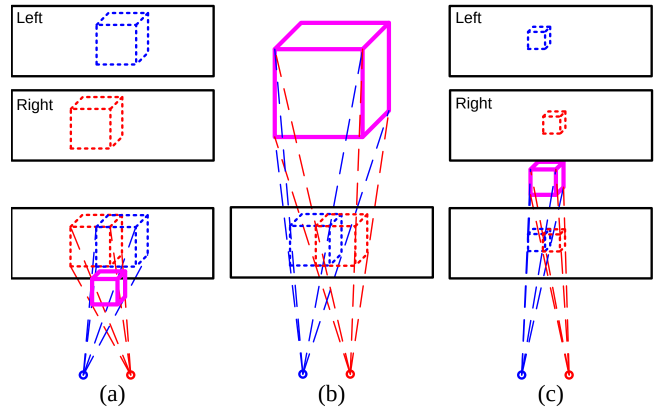


Fig. 3. Image shifting versus correct depth shift. (a) Given the original left/right images, viewers perceive the virtual magenta cube. (b) If we shift the left and right images, the cube moves backward and scales up at the same time. (c) To change the depth of the cube without scaling, its projections on the images must be scaled as well.

From Eq. (6) we can see that the perceived depth is related nonlinearly to disparity, as shown in Fig. 2. An object appears in front of the screen when its disparity is negative and vice versa. For a scene with objects at depths ranging from 0.8 m to infinity, the disparity values range from -0.2 m to 0.065 m. Note that the above formulas are measured in the physical domain, not the pixel domain. Therefore, to use these formulas, disparities measured in the pixel domain must be converted to the physical domain by dividing by the pixel density (pixels per inch) of the target display.

B. Depth Adaptation

We can also see that when the image is transformed, the disparities change subtly. When the image is stretched linearly along the x-axis, the disparities increase linearly. However, the disparities are unaffected by y-axis stretch. Therefore, when the real estate of the display increases, the depth range of the displayed image increases accordingly. In the worst case, the object can be pushed beyond infinity (i.e. , $d_p > e$ in Eq. (6)), leading to an incorrect and irritating 3D effect. Similarly, when the aspect ratio changes, the disparities change accordingly. These phenomena seriously hinder the distributions of stereoscopic content across different medium: a striking 3D effect in the cinema may look flat and boring on a 3D mobile phone, and a 3D effect that looks good on a mobile phone may lead to diplopia in the cinema.

Another crucial parameter of a stereoscopic display is its *comfort depth range*, or *comfort zone*. When viewing a stereoscopic display, our eyes fixate on the virtual 3D object, providing the convergence cue for 3D perception. We must focus on the screen for sharp images [19]; this lack of an accommodation cue (change of focus) informs the brain that the display is flat. This conflict between accommodation and convergence cues causes visual discomfort, especially for excessive disparity values. Thus the comfort zone is that range of depths where the conflict can be tolerated. Depth outside that zone can cause diplopia or blur. Because of optics properties, viewing distances, and other factors, different displays

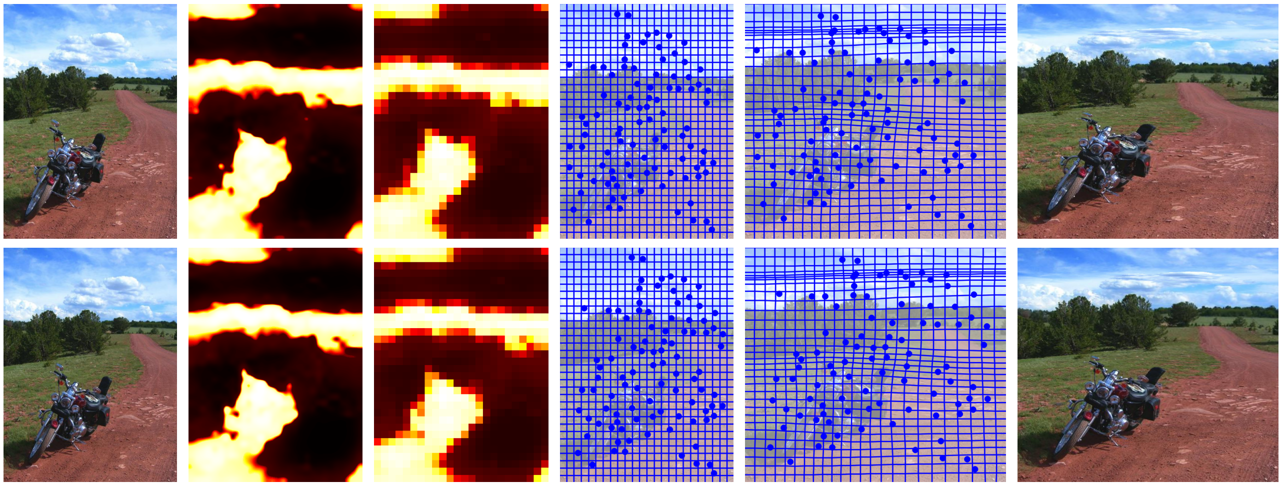


Fig. 4. Algorithm overview. The top row shows the left view and second row shows the right view. (a) The input binocular image pair, (b) their saliency maps, (c) quad importance maps, (d) original images with grid meshes and feature points, (e) retargeted image pair with deformed grid meshes and relocated feature points, and (f) retargeted image pair.

have different comfort zones. Even for the same viewing configuration, the comfort zone can vary among individuals.

For these reasons, *depth adaptation* is required to ensure a vivid and enjoyable 3D experience. Given a stereoscopic image pair captured for a specific viewing configuration, the depth adaptation process attempts to adjust the content such that the 3D perception delivered in another viewing configuration is identical or similar to the original one. The method most commonly used in commercial stereo displays is the image shifting method. By horizontally shifting one of the images, we can increase/decrease the disparities and thus the depths. However, because the mapping between the disparity and 3D coordinate is nonlinear, this simple method causes undesirable miniaturization or gigantism effects, as illustrated in Fig. 3. Thus, when image shifting is used to adjust the binocular image, the perceived scene scale changes accordingly as an unwanted side effect. Other methods that rely on global image transformations have the same drawback [1].

Theoretically, for perfect depth adaptation, one should first reconstruct the scene from the input images, transform the scene to fit the display comfort zone, and finally re-project the scene to obtain the new stereoscopic images. One such example is shown in Fig. 3 (c), where we see that the projected images of the object are properly transformed. However, this approach requires dense scene geometry, which is typically noisy or even unavailable. Moreover, in the scene transformation and re-projection process, the system must recover scene content occluded in the original input, which itself is a challenging and unsolved research problem. One solution is to sample more data during acquisition by using multi-rigging techniques or camera arrays [20], which allow for better scene reconstruction. If the footage is computer-generated, it is possible to re-render it for each display [21]. However, these approaches are expensive for amateurs. Another solution is to edit the stereoscopic content by manual authoring, which can be very time-consuming. As described in the next section, our warping-based approach can avoid these difficulties and still generate appealing results.

IV. The Proposed Retargeting Method

We developed our content-aware binocular image display adaptation algorithm based on warping-based image manipulation methods [4] with two stereoscopic constraints: *vertical alignment* and *horizontal disparity consistency*. The former requires that corresponding points in the two views remain horizontally aligned on the same scanline after processing, and the latter requires that horizontal disparities are manipulated in a consistent way. As explained in Section III, these constraints are important for enjoyable 3D viewing experiences.

Fig. 4 illustrates the overview of our method. First, a saliency detection algorithm is applied to measure per-pixel importance of the image pair (Fig. 4(b)). Then, we represent each image as a grid mesh and measure the per-quad importance by averaging and normalizing per-pixel saliency (Fig. 4(c)). Next, we use feature extraction and matching to obtain sparse matching pairs between the left and right images (Fig. 4(d)). Given the retargeting parameters, we obtain the warping functions on the mesh vertices by optimizing an energy function (Fig. 4(e)). Finally, we interpolate the full warping fields and the final output using bilinear interpolation (Fig. 4(f)). We detail each step below.

Given the binocular image pair $\{\mathbf{I}^L, \mathbf{I}^R\}$, we estimate the saliency maps $\{\Phi^L, \Phi^R\}$ using a graph-based visual saliency algorithm [22]. To build the stereoscopic constraints, we must extract the correspondences between $\mathbf{I}^L$ and $\mathbf{I}^R$. Although we could use two-frame stereo correspondence algorithms to estimate dense correspondences [23], unfortunately, state-of-the-art stereo methods are computationally expensive and still far from perfect. Thus, we only extract sparse but reliable features from the image pair. In our system, we first detect SIFT features [24] from both images. For each feature point in $\mathbf{I}^L$, we find its best match in $\mathbf{I}^R$ and then verify all matches using the fundamental matrix estimated using RANSAC [25]. To ensure a better spatial distribution of features, we use non-maximum suppression [26] to remove cluttered features. As we will see, warped images driven by sparse correspondences are enough to generate satisfactory 3D effects.

A. Energy Minimization

In this subsection, we describe the energy minimization approach we use to obtain the warping fields. We denote the set of n matched features as $\mathbf{F} = \{(\mathbf{f}_i^L, \mathbf{f}_i^R)|i = 1..n\}$. Similar to other warping-based methods, we use a uniform grid mesh to guide the image deformation. Let $\{\mathbf{V}^L, \mathbf{E}^L, \mathbf{Q}^L\}$ and $\{\mathbf{V}^R, \mathbf{E}^R, \mathbf{Q}^R\}$ denote the grid meshes for both images, where $\mathbf{V}$, $\mathbf{E}$ and $\mathbf{Q}$ represent the vertex sets, edge sets, and quad face sets, respectively. Our content-aware display adaptation algorithm attempts to find two sets of deformed vertex positions $\tilde{\mathbf{V}}^L = \{\tilde{\mathbf{v}}_i^L\}$ and $\tilde{\mathbf{V}}^R = \{\tilde{\mathbf{v}}_i^R\}$ for the left and right images such that the energy function $\Psi(\tilde{\mathbf{V}}^L, \tilde{\mathbf{V}}^R)$ is minimized. The energy function consists of four parts: distortion energy Ψ_d, line bending energy Ψ_b, alignment energy Ψ_a, and disparity consistency energy Ψ_c.

Distortion energy. This term prevents important quads from being non-uniformly scaled and is defined similarly to that used by Wang *et al.* [4]. For each quad q with four edges $\mathbf{E}(q)$, the distortion energy for the quad is defined as

$$\Psi_q(q) = \sum_{(i,j)\in\mathbf{E}(q)} \| (\tilde{\mathbf{v}}_i - \tilde{\mathbf{v}}_j) - s_q (\mathbf{v}_i - \mathbf{v}_j) \|^2, \quad (7)$$

where s_q is the scale factor defined by $\tilde{\mathbf{v}}_i$ and $\mathbf{v}_i$. Please refer to [4] for more details. The total distortion energy is the weighted sum of the distortions of all quads in both views, defined as

$$\Psi_d = \sum_{q\in\mathbf{Q}^L} \varpi(q)\Psi_q(q) + \sum_{q\in\mathbf{Q}^R} \varpi(q)\Psi_q(q), \quad (8)$$

where $\varpi(q)$ is the quad importance of q (Fig. 4 (c)). We initialize $\varpi(q)$ as the average of the saliency values of all pixels in q and then normalize it to $[\varepsilon, 1]$, where ε is a small constant (we set $\varepsilon = 0.05$).

Line bending energy. In addition to non-uniform scaling, we want to minimize the bending of the grid edges. That is, we want the angle between the original edge e and deformed edge $\tilde{e}$ to be as small as possible. Wang *et al.* [4] define a non-linear line bending energy which is complex to minimize. Here we propose a new linear line bending energy. Consider edge e which has the two vertices $\mathbf{v}_i$ and $\mathbf{v}_j$ and its deformed version $\tilde{e} = (\tilde{\mathbf{v}}_i, \tilde{\mathbf{v}}_j)$. Define vectors $\mathbf{e} = \mathbf{v}_i - \mathbf{v}_j$ and $\tilde{\mathbf{e}} = \tilde{\mathbf{v}}_i - \tilde{\mathbf{v}}_j$; we use the following term to approximate the angle between $\mathbf{e}$ and $\tilde{\mathbf{e}}$:

$$\Delta(\tilde{\mathbf{e}}) = \|s_e\mathbf{e} - \tilde{\mathbf{e}}\|^2, \quad (9)$$

where s_e is a scale parameter we wish to optimize. Taking the partial derivative of Δ with respect to s_e, we obtain the optimal s_e^* as

$$s_e^* = (\mathbf{e}^T\mathbf{e})^{-1}\mathbf{e}^T\tilde{\mathbf{e}}. \quad (10)$$

Substituting s_e^* back into Eq. (9) yields a function of $\tilde{\mathbf{e}}$:

$$\begin{aligned}\Delta(\tilde{\mathbf{e}}) &= \|s_e\mathbf{e} - \tilde{\mathbf{e}}\|^2 \\ &= \|\mathbf{e}(\mathbf{e}^T\mathbf{e})^{-1}\mathbf{e}^T\tilde{\mathbf{e}} - \tilde{\mathbf{e}}\|^2 \\ &= \|\mathbf{C}\tilde{\mathbf{e}}\|^2, \end{aligned} \quad (11)$$

where $\mathbf{C} = \mathbf{e}(\mathbf{e}^T\mathbf{e})^{-1}\mathbf{e}^T - \mathbf{I}$ and $\mathbf{I}$ is the identity matrix.

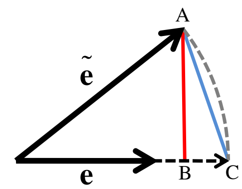


Fig. 5. The approximated line bending energy. Wang *et al.* used the line AC to approximate the arc AC. This however leads to a nonlinear term. We instead use line AB as the approximation. The results are similar to theirs but without nonlinear terms.

Eq. (11) can be further rewritten as a function of $\tilde{\mathbf{v}}_i$ and $\tilde{\mathbf{v}}_j$:

$$\Delta(\tilde{\mathbf{v}}_i, \tilde{\mathbf{v}}_j) = \left\| \mathbf{C} \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{v}}_i \\ \tilde{\mathbf{v}}_j \end{bmatrix} \right\|^2. \quad (12)$$

Finally, the total line bending energy can be defined as

$$\Psi_b = \sum_{(i,j)\in\mathbf{E}^L} \Delta(\tilde{\mathbf{v}}_i^L, \tilde{\mathbf{v}}_j^L) + \sum_{(i,j)\in\mathbf{E}^R} \Delta(\tilde{\mathbf{v}}_i^R, \tilde{\mathbf{v}}_j^R). \quad (13)$$

We use the example shown in Fig. 5 to compare different line bending measurements. In this example, the ideal line bending measurement should be the angle between $\mathbf{e}$ and $\tilde{\mathbf{e}}$, or equivalently the length of the arc AC. Wang *et al.* use the line AC to approximate the arc, which is still a non-linear function of $\tilde{\mathbf{e}}$. In our formulation (Eq. (9)), the parameter s_e allows us move freely on the line along $\mathbf{e}$. Thus, Eq. (9) essentially finds the minimal distance from A to the line along $\mathbf{e}$, giving us the line AB. That is, we use the line AB to approximate the arc AC, and in so doing eliminate the nonlinear terms in the energy function. Although it is not as close as the line AC, it is good enough for our application. Only when the arc AC is large does the approximation deviate significantly, when the warping result is implausible and useless anyway.

Fig. 6 shows that the proposed line bending energy yields results similar to Wang *et al.*'s line bending energy. However, it is more efficient. Because Wang *et al.*'s line bending energy is nonlinear, it requires an iterative nonlinear optimization. The number of iterations is typically around 10. The proposed linear line bending energy requires only one iteration and is therefore more efficient.

Alignment energy. This term is used to ensure vertical alignment of features after deformation to avoid binocular asymmetries. The energy Ψ_a is defined as

$$\Psi_a = \frac{1}{n}\sum_{i=1}^{n} \left(\tilde{\mathbf{f}}_i^L[y] - \tilde{\mathbf{f}}_i^R[y]\right)^2, \quad (14)$$

where we use the notation $\mathbf{v}[y]$ to represent the y component of the vector $\mathbf{v}$, and similarly $\mathbf{v}[x]$ for the x component.

Note that the relocated feature $\tilde{\mathbf{f}}$ can be expressed as a linear combination of the vertices after deformation $\tilde{\mathbf{v}}_i$ using barycentric coordinates. Assume that, before deformation, the feature $\mathbf{f}$ is related to the vertices $\mathbf{v}_i$ of the quad where it is located in as $\mathbf{f} = \sum_{i=1}^4 \beta_i\mathbf{v}_i$, where β_i are the barycentric coordinates. The relocated feature $\tilde{\mathbf{f}}$ can then be written as a linear combination of deformed vertices, $\tilde{\mathbf{f}} = \sum_{i=1}^4 \beta_i\tilde{\mathbf{v}}_i$, using the same barycentric coordinates. Therefore, Eq. (14) can be written as a function of the warped vertices $\tilde{\mathbf{v}}_i$.

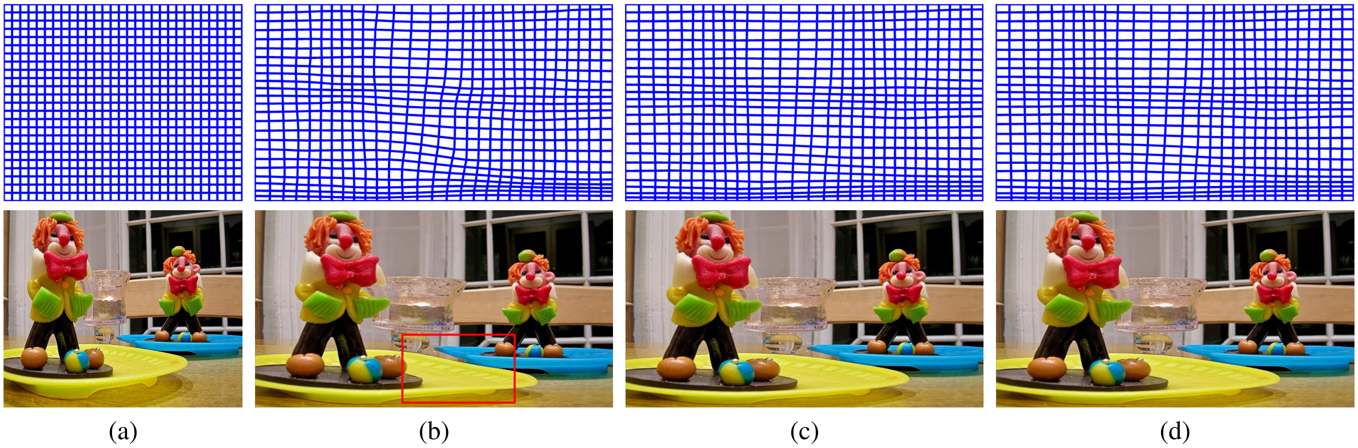


Fig. 6. Resizing results with different line bending energies. (a) Original single image. (b) Without line bending energy. (c) Wang *et al.*'s line bending energy. (d) The proposed line bending energy. The proposed line bending energy gives similar results to (c), but it is linear and more efficient.

Disparity consistency energy. This term is used to ensure that the disparities of features are manipulated in a consistent way to avoid distortion of the perceived depths. We propose two different disparity consistency energies, each of which is useful for different applications. The first energy is an attempt to keep the perceived depths identical to those before deformation. This is useful for the situation when the image size changes while the viewing configuration is the same. In such cases, we would like to maintain the same disparity so that the perceived depth is the same after resizing. For this option, the disparity consistency energy Ψ_c is defined as

$$\Psi_c = \frac{1}{n}\sum_{i=1}^{n}\left(d_i - \tilde{d}_i\right)^2, \tag{15}$$

where $d_i = \mathbf{f}_i^R[x] - \mathbf{f}_i^L[x]$ and $\tilde{d}_i = \tilde{\mathbf{f}}_i^R[x] - \tilde{\mathbf{f}}_i^L[x]$ are the disparity values in the pixel domain before and after deformation respectively.

In cases in which viewing configurations change, the disparity values should be scaled and shifted accordingly. Thus, for the second option, we try to maintain the relative depths of the feature points in the input images by finding a monotonic increasing mapping of depths. In this way, the depth order of the objects is preserved but their absolute depths are flexible. A trivial choice is to find a proper 1D similarity transform of *depths* to maintain the relative depths. However, this makes the energy term nonlinear to deformed features. We choose instead to find a proper 1D similarity transformation of *disparities* to maintain the relative depths:

$$\Psi_c = \frac{1}{n}\sum_{i=1}^{n}\left((s_d d_i + t_d) - \tilde{d}_i\right)^2, \tag{16}$$

where s_d represents the global scaling factor of disparity and t_d represents the shift. Using the same approach used to obtain the optimal s_e^* in Eq. (9), we eliminate s_d and t_d from Eq. (16) and turn it into a function of deformed features, each of them a linear combination of deformed vertices. After defining matrix $\mathbf{E}$ and vector $\tilde{\mathbf{E}}$ as

$$\mathbf{E} = \begin{bmatrix} d_1 & 1 \\ d_2 & 1 \\ \vdots & \vdots \\ d_n & 1 \end{bmatrix} \text{ and } \tilde{\mathbf{E}} = \begin{bmatrix} \tilde{d}_1 \\ \tilde{d}_2 \\ \vdots \\ \tilde{d}_n \end{bmatrix},$$

Eq. (16) is re-written as

$$\Psi_c = \frac{1}{n}\left\|\mathbf{E}\begin{bmatrix} s_d \\ t_d \end{bmatrix} - \tilde{\mathbf{E}}\right\|^2, \tag{17}$$

and the optimal scale s_d^* and shift t_d^* are

$$\begin{bmatrix} s_d^* \\ t_d^* \end{bmatrix} = (\mathbf{E}^T\mathbf{E})^{-1}\mathbf{E}^T\tilde{\mathbf{E}}. \tag{18}$$

By substituting s_d^* and t_d^* back into Eq. (17), the energy Ψ_c can be written as a function of deformed features:

$$\Psi_c = \frac{1}{n}\|\mathbf{B}\mathbf{A}\tilde{\mathbf{f}}\|^2, \tag{19}$$

where $\mathbf{B} = \mathbf{E}(\mathbf{E}^T\mathbf{E})^{-1}\mathbf{E}^T - \mathbf{I}$, $\mathbf{A} = [-\mathbf{I}|\mathbf{I}]$ and

$$\tilde{\mathbf{f}} = \begin{bmatrix} \tilde{\mathbf{f}}_1^L[x] \\ \vdots \\ \tilde{\mathbf{f}}_n^L[x] \\ \tilde{\mathbf{f}}_1^R[x] \\ \vdots \\ \tilde{\mathbf{f}}_n^R[x] \end{bmatrix}.$$

Again, $\tilde{\mathbf{f}}$ can be rewritten in terms of the deformed vertices. Note that the optimal scale and shift can be integrated into the energy function and determined automatically by optimization.

The final energy Ψ is the sum of the four defined energy terms:

$$\Psi = \Psi_d + \lambda_b\Psi_b + \lambda_a\Psi_a + \lambda_c\Psi_c. \tag{20}$$

In our experiments, we set $\lambda_b = 1$, $\lambda_a = 10$, and $\lambda_c = 500$. Our energy function is an interplay between 2D shape conservation and depth preservation. Note that these terms are all functions of the deformed grid vertices $\tilde{\mathbf{v}}_i^L$ and $\tilde{\mathbf{v}}_i^R$. Minimizing Ψ corresponds to solving a least-squares problem and leads to a linear system involving only $\tilde{\mathbf{v}}_i^L$ and $\tilde{\mathbf{v}}_i^R$. By

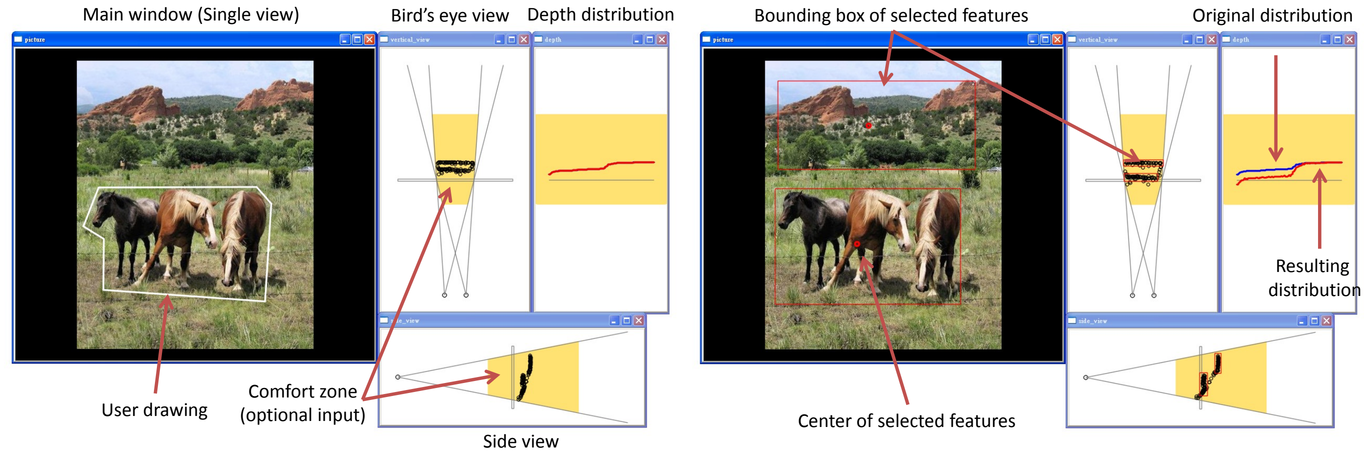


Fig. 7. Screenshots of the user interface for stereoscopic image editing. The left part shows the image, the side view, and bird's eye view of feature points and depth distributions. The right part shows those of the edited results.

finding the sets of deformed vertices $\tilde{\mathbf{V}}^L$ and $\tilde{\mathbf{V}}^R$ which minimize Ψ and satisfy the boundary conditions [4], we warp both images to the target resolution while maintaining the 3D shapes of important objects.

V. Interactive Stereoscopic Image Editing

Our framework can be extended to interactively edit depths of the whole scene or even a region in images. Note that it is more natural to edit depths than disparities. Thus, we use the formulas laid out in Section III to convert between depths and disparities (Eq. (1–3) and Eq. (5)). In this section we describe first our user interface design and then how to incorporate user edits into the energy defined in the previous section.

A. User Interface Design

We developed a graphical user interface (GUI) for interactive and direct manipulation of the stereoscopic images. Screenshots of our GUI are shown in Fig. 7. The main window shows the editing image, and can freely switch between the original input and the edited result. Depending on the display capability, it can also switch to the left view, the right view, the anaglyph image, and the binocular image, which allows the user to view the 3D effect during editing.

The user simply drags the image boundary to adjust the size and aspect ratio, and the system displays the retargeting image interactively. We provide several different visualization methods for depth adaptation or adjustment. Our GUI shows the 3D spatial distribution of the feature points from the side and from the top. It also displays the comfort zone, and the sorted depth distribution of feature points. The comfort zone is an optional input which is found either in the specification of the target display or is determined empirically.

We provide many options for editing depths. As the first option, the user can either specify a similarity transformation or directly draw the desired target depth distribution in the depth distribution view. Our system automatically calculates the resulting disparity value for each feature point. For the second option, the user can select an area by drawing a bounding polygon and edit its 3D position and scaling factor. In all these editing operations, our system can generate the warped result and update the disparity distribution and feature locations immediately. For example, in Fig. 7, we select the horses, and move them closer to the screen while keeping the background mountains fixed. The user can see this editing is performed correctly, as indicated in the feature and depth distributions.

B. Modified Energy

As mentioned, for the first option, the user first specifies the desired depth transformations for all features. This can be done by specifying 1) a 1D similarity transformation of depths, or 2) the target depth distribution. The system then converts the depths to the corresponding disparities and incorporates them into the disparity consistency energy Ψ_c.

To illustrate, assume that a similarity transformation is used. First, the disparities d_i of all features $\mathbf{F} = \{\mathbf{f}_i\}$ are converted from the pixel domain to the physical domain. Next, the depths Z_i are calculated from their disparities d_i using Eq. (6), $Z_i = \frac{eD}{e-d_i}$. The target depth $\hat{Z}_i$ is calculated as $\hat{Z}_i = s_z Z_i + t_z$ and then converted back as the target disparity $\hat{d}_i = e\left(1 - \frac{D}{\hat{Z}_i}\right)$ using Eq. (4). After converting $\hat{d}_i$ from the physical domain to the pixel domain, the disparity consistency energy is then modified using the target disparities as

$$\Psi_c = \frac{1}{n}\sum_{i=1}^{n}\left(\tilde{d}_i - \hat{d}_i\right)^2 . \quad (21)$$

If the target depth distribution is specified, the target depths $\hat{Z}_i$ are given by the user and the resulting procedure is similar.

Our method also allows users to change the size and position of the object. First, the user selects features on the object by drawing a closed region. The user can then input the 3D scaling factor (s_x, s_y, s_z) and translations (t_x, t_y, t_z) for this object. The set of selected features $\hat{F} = (\mathbf{f}_i^L, \mathbf{f}_i^R)$ is projected back to its 3D position (X_i, Y_i, Z_i) using Eq. (5). Scaling and translating the 3D position accordingly yields the target 3D position $\hat{X}_i, \hat{Y}_i, \hat{Z}_i$, which we then project onto both views to obtain $\hat{\mathbf{f}}_i^L$ and $\hat{\mathbf{f}}_i^R$ using Eqs (1)–(3). For the remaining features, either Eq. (15) or (16) is used as the constraint.

(a) (b) (c) (d) (e) (f)

Fig. 8. Resizing results with different aspect ratios. (a) The original stereoscopic image, resolution 472×425. (b) and (c): The results using linear scaling. (d) The saliency map of the left image. (e) and (f): Our results.

Assuming that Eq. (16) is used, the disparity consistency energy is then modified as

$$\Psi_c = \frac{1}{|\hat{\mathbf{F}}|} \sum_{i \in \hat{\mathbf{F}}} \left(\|\tilde{\mathbf{f}}_i^L - \hat{\mathbf{f}}_i^L\|^2 + \|\tilde{\mathbf{f}}_i^R - \hat{\mathbf{f}}_i^R\|^2 \right) + \frac{\lambda}{|\mathbf{F} \setminus \hat{\mathbf{F}}|} \sum_{i \in \mathbf{F} \setminus \hat{\mathbf{F}}} \left((s_d d_i + t_d) - \tilde{d}_i \right)^2. \quad (22)$$

λ is the weight between these two parts of energy (we set $\lambda = 0.1$). Finally, Ψ with the modified Ψ_c is minimized to deform the images to match our constraints in a content-aware manner.

VI. Experimental Results

We implemented our system on a PC with a 2.39GHz Pentium Duo CPU and 3.5GB RAM. Our system is very efficient since it only involves linear system solving. Furthermore, the matrix in the linear system is fixed and its factorization can be pre-computed. Thus, retargeting to different sizes only requires one back substitution to solve the linear system. For a 480×600 stereoscopic image with quad size 20×20, factorizing a matrix typically takes less than 2 seconds, and a back substitution takes 0.0086 second. Overall, our retargeting system takes 0.018 second for one retargeting operation, that is, achieving real-time editing with 55 FPS.

For our experiments, we collected binocular images from the stereoscopic image repository on Flickr. Limited by the medium, the results shown in this section are presented as red/cyan color anaglyph images; they are best viewed under their original resolutions. For user studies, a Samsung 2233 RZ 22" 3D monitor with shuttered glasses and an Nvidia GeForce 3D Vision Solution were used for the best viewing results.

A. Stereoscopic Image Resizing

In Fig. 8, we resize the original image pair to two different aspect ratios. The depths are preserved by using Eq. (16) as the disparity consistency energy. Compared with the traditional scaling method, our results do not distort the people in the image or alter their relative perceived depths.

In Fig. 9, we show the resizing results using different disparity consistency energies. While we could keep the absolute depths identical to the original ones using Eq. (15) as shown in Fig. 9(c), because this enforces an identity transformation of disparities, there would be more distortions in the non-salient regions, such as at the image boundary. On the other hand, if we only constrain the relative depths using Eq. (16), the additional degree of freedom would allow us to better preserve the more prominent objects, as shown in Fig. 9(d). However, because of the unconstrained depth range, the objects might end up outside of the comfort zone of the display. Therefore, although Eq. (16) generally leads to better content-aware results, Eq. (15) is a better choice for large resizing factors.

(a) (b) (c) (d)

Fig. 9. Resizing results using different disparity consistency energy functions. (a) The original 444×610 stereoscopic image. Results of (b) traditional scaling, (c) our method using Eq. (15), and (d) Eq. (16) as Ψ_c.

In Fig. 10, we compare our method with linear scaling and with the optimized scale-and-stretch (OSS) method [4]. With linear scaling, the image is stretched uniformly and features are kept aligned horizontally after resizing, thus maintaining the required binocular symmetry. However, it seriously distorts the object shapes and scales up all disparities. Some objects may go beyond the comfort zone and cause diplopia. With the OSS method, while the image is resized to the target resolution in a content-aware manner, the resulting unwanted vertical parallax causes viewing discomfort. In addition, the original absolute or relative horizontal disparity values are totally destroyed, which leads to incorrect and inconsistent 3D perception. Our method yields the best results by jointly performing image resizing and depth adaptation in a content-aware manner.

B. Depth Adaptation and Editing

In Fig. 11, by specifying different scale factors for feature depths, we retarget the binocular image to different aspect ratios and suppress or expand the depth range of the scene using the procedure in Section V. In this example, the depths of feature points are calculated from their disparities using Eq. (6); e and D are set to 6.5 cm and 35 cm, respectively. In Fig. 12, by specifying different 1D similarity transformations of depths, we can put the clouds in the image at different Z-positions relative to the screen.

In Fig. 13, the user edits the position and shape of the boat, and our system generates the result by using the modified disparity consistency energy (Eq. (22)). In Fig. 13, the boat is placed at different depths: (b) in front of the screen and (c) behind the screen. Because we specify a rigid transformation to 3D coordinates of features on the boat, our method does not change the object size as the image shifting method does, as illustrated in Fig. 3(b). Note that although the boat in Fig. 13(c) looks smaller than the boat in Fig. 13(b), both actually have the same 3D size. The boat in Fig. 13(b) looks bigger because it has been edited to be closer to the viewer and thus has a

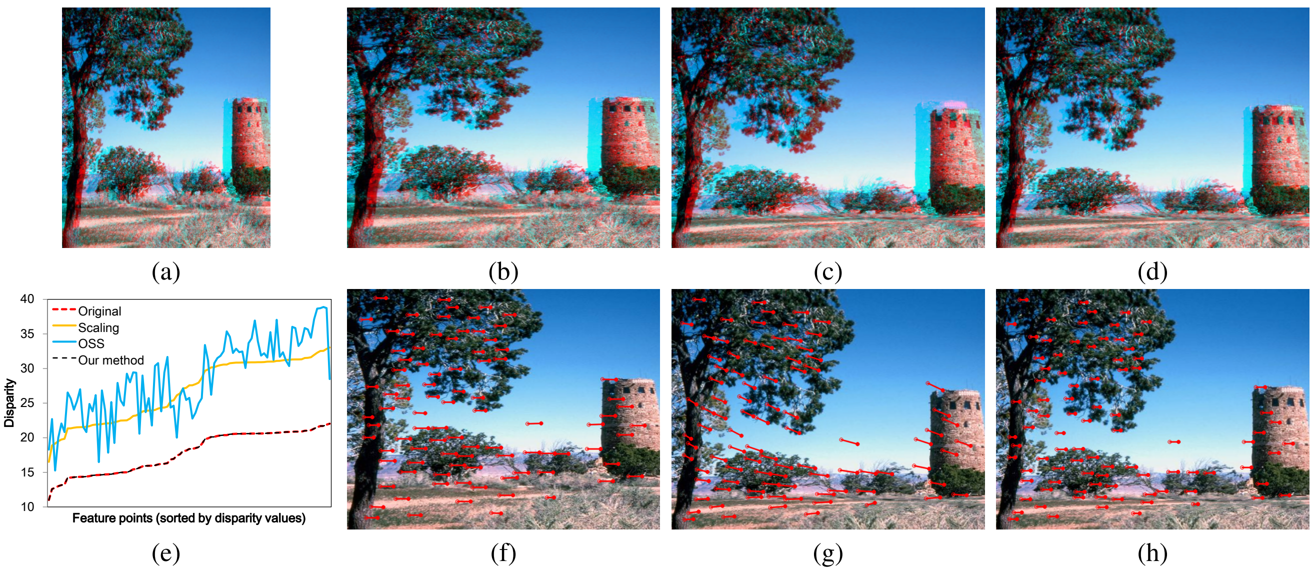


Fig. 10. Comparisons among different retargeting approaches. (a) The original 472×545 stereoscopic image pair. We expand its width to 1.5x using (b) linear scaling, (c) OSS, and (d) the proposed method. The disparities of the feature points are shown in (e). Note that our method preserves most disparity values (the distribution overlaps with the original one) without causing any noticeable distortion. The feature points and their disparities are shown in (f)–(h) for the three methods in (b)–(d). Note the expanded horizontal disparities in (f) and the vertical parallax in (g).

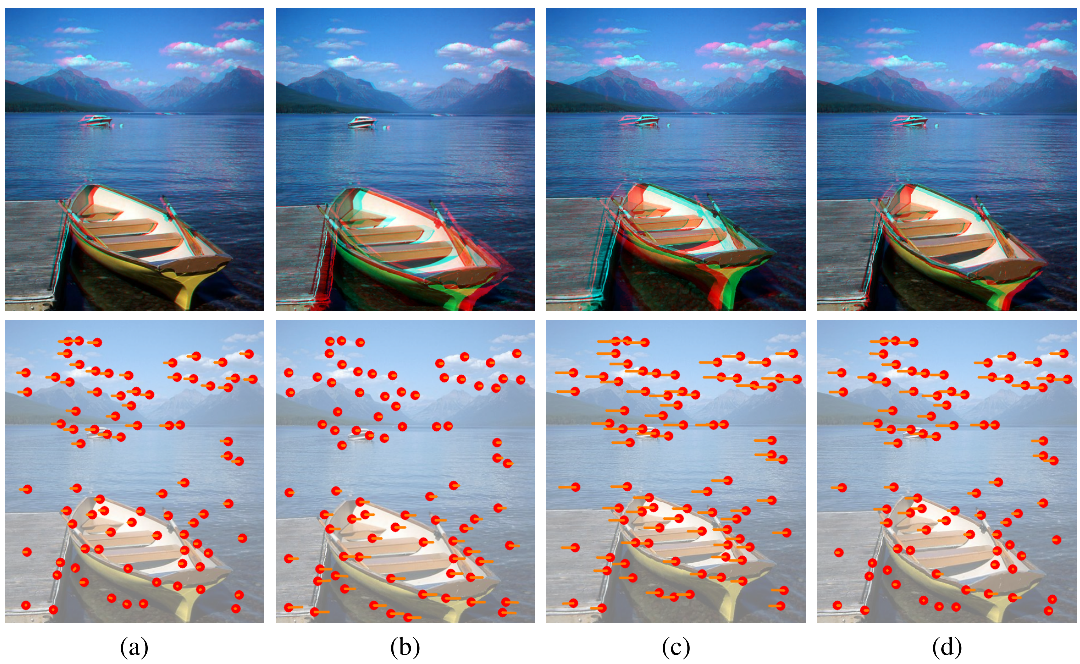


Fig. 13. Results of depth adaptation by user editing. The boat is placed at different depths. The first row presents the results as red/cyan images; and the second row displays the right-view images with features and disparities. (a) The original 476×555 stereoscopic image. All objects are behind the screen. (b) The boat is moved to the front of the screen (note that sign changes of disparities), and (c) behind the screen. (d) The depth range of the boat is tripled.

bigger 2D projection on the image plane. Note that for this effect, the projection of the boat on the image plane must change with its depth, leading to the gradual scale change in Fig 13(b,c). In Fig. 13(d), the boat is stretched along the Z-axis by specifying a new depth range while its X and Y coordinates are fixed: the boat is essentially prolonged in shape while keeping its position. Note that because we use Eq. (16) to preserve relative disparities in the image, the depths of background objects change accordingly. The use of Eq. (15) allows us to modify the location of the boat while fixing all other objects.

C. *User Study*

Since it is difficult to define an objective evaluation for stereoscopic retargeting, we performed two subjective user studies to evaluate the performance of our method. The target for the first user study was image resizing and that for the second was depth editing. 24 subjects (13 male, 11 female, average age 24.7) with normal stereoscopic vision were invited to participate in the user study. For both studies, each subject was presented with eight test cases.

Image resizing. Since there were no other content-aware binocular image resizing methods, we compared it to the naïve extension of Wang *et al.*'s method (OSS) by applying

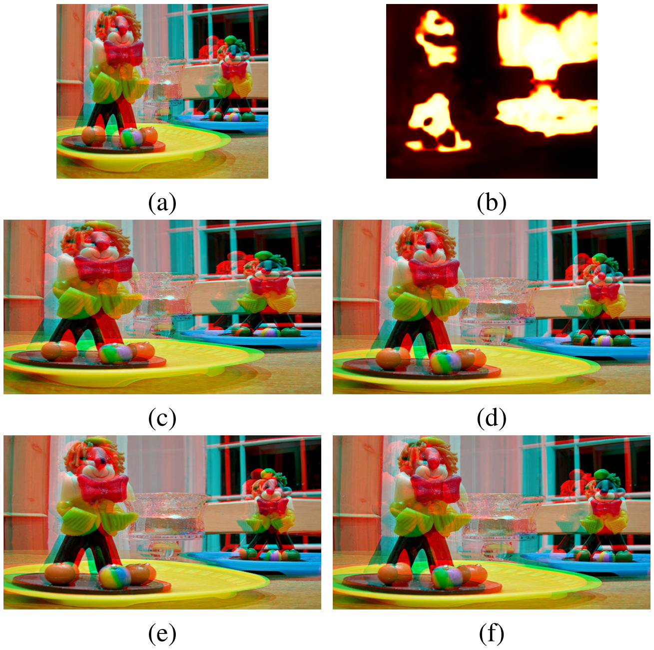


Fig. 11. Resizing results of different depth ranges by user editing. (a) The original 574×473 input. (b) The saliency map of the left image, and the results of (c) traditional scaling, (d) our method which maintains the absolute depths, (e) our method which suppresses the depth range to a half, and (f) our method which expands the depth range to be 1.5x. Note that less prominent areas are distorted more to complete the required depth range change in more prominent areas.

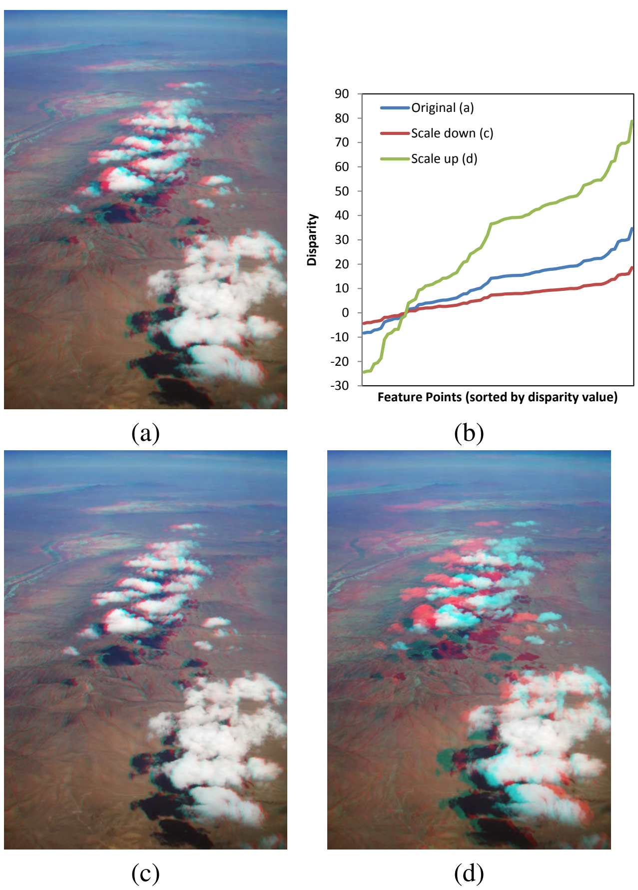


Fig. 12. Depth adaptation results. (a) The original 484×680 stereoscopic image. (b) The disparity distributions. (c) The depths are scaled down. (d) The depths are scaled up.

Fig. 14. The 8 test cases (left images) used in the image resizing user study.

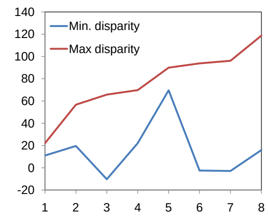


Fig. 15. The disparity ranges of 8 test cases in the image resizing user study.

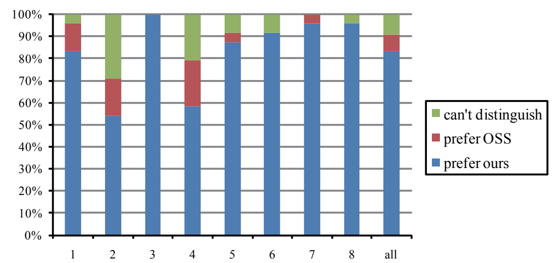


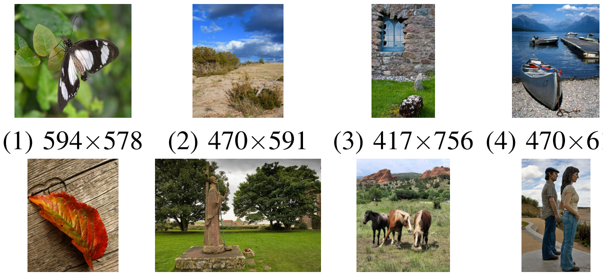
Fig. 16. The voting result for each image in the image resizing user study.

it independently on the left and right images. Subjects were asked to choose between the results of our method and OSS, taking into account viewing comfort and depth perception. In this user study, our method used in all cases the first kind of disparity consistency energy. All eight images were expanded on the x-axis to 150% of the original width. Fig. 14 shows the original images and their dimensions, and Fig. 15 shows their disparity ranges.

Applying OSS independently likely causes binocular asymmetries, including unwanted vertical parallax, inconsistent horizontal disparities, and excessive disparity, all of which can lead to uncomfortable visual phenomena such as blur and diplopia. Our method minimizes binocular asymmetries by taking into account stereoscopic constraints.

Subjects found it difficult to identify a superior method for 18 of the total 192 comparisons. Of the remaining 174 comparisons (90.6%) in which subject preferences were clear-cut, 160 votes (91.9%) favored the results of our method over those of OSS. Fig. 16 shows the distribution of votes for each test case. In general, the proposed method significantly outperformed OSS for cases with large depth ranges such as cases #3, #6, #7, and #8.

Depth editing. Fig. 17 shows the eight test cases used in the user study, and Fig. 18 shows their disparity ranges. In this study, the input depths were adapted using the procedure described in Section V. The depths of all features were



(1) 594×578 (2) 470×591 (3) 417×756 (4) 470×612

(5) 470×589 (6) 540×369 (7) 452×550 (8) 456×754

Fig. 17. The 8 test cases (left images) used in the depth editing user study.

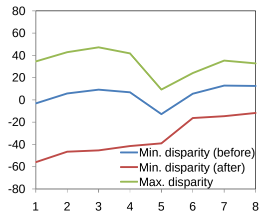


Fig. 18. The disparity ranges of 8 test cases in the depth editing user study.

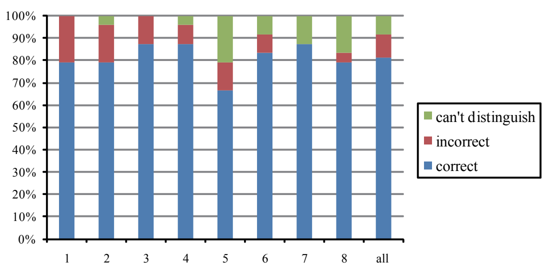


Fig. 19. The voting result for each image in the depth editing user study.

calculated from their disparities using Eq. (6) where e and D were set to 6.5 cm and 35 cm, respectively. Then, we applied the following 1D similarity transformation for depths Z_i

$$\hat{Z}_i = 2Z_i - Z_{max}, \tag{23}$$

where Z_{max} was the maximum depth of all features. This linear mapping kept the maximum depth the same, but doubled the depth range; this manipulation of the depth range effectively drew the foreground area closer to the user.

Subjects were asked to answer the question "*Which image's foreground area looks closer to you?*" Of the 192 comparisons, 16 were indecisive. Of the remaining 176 comparisons, 156 votes (88.6%) correctly recognized the images with closer foreground areas. Fig. 19 shows the distribution of votes for each test case. Note that depth range clearly influences whether the user can correctly recognize the depth editing. For example, case #5 in Fig. 17 has the narrowest disparity range and hence the worst recognition result. This is because the depths did not change as much as in other examples. In addition, other contextual factors such as perspective and occlusion cues affect human depth perception. This also explains why some were misrecognized. Nevertheless, in this evaluation more than 80% of the depth editing results were considered successful.

D. Discussions and Limitations

Our method has a few limitations. First, it shares the limitations that apply to all warping-based 2D retargeting methods. Thus, it may fail to preserve shapes of prominent structures if they are not well aligned with the mesh. Second, the performance of our algorithm is limited by the accuracy of the saliency map and the number of correct stereo correspondences. However, the continuous advance of saliency detection, feature matching, and stereo matching algorithms over time will no doubt alleviate this problem. Third, even with dense, per-pixel stereo correspondences, the warping-based approach inevitably introduces distortions around depth discontinuities. However, we found that when the degree of disparity change is moderate, and when no structuring elements appear around the discontinuity, these artifacts are hardly noticeable. Fourth, our method is not well-suited to scenes that contain objects with transparency or thin structures (such as smoke, glass, or hair) because feature extraction and matching in such cases is not reliable.

(a) (b) (c)

Fig. 20. An unsuccessful case. (a) The original stereoscopic image. (b) The result after increasing the depth range. (c) The resultant left view with the deformed grid mesh and features.

Fig. 20 shows an example where our method does not perform well. While the lamp pole is a thin structure and no features are associated with it, the lamp itself has a few features and the flowers on the bottom have many. In this example, we attempt to scale up the depth range to bring the flowers closer. Thus, the lamp remains where it is and the flowers move closer to increase the range. Because there are no features associated with the pole, it is not considered to be at the same depth as the lamp. Thus, with no constraints on the pole, the smooth warping field ensures that its top stays with the deeper lamp and its bottom goes with the shallower flowers, thereby curving the pole. This problem is related to feature extraction for thin structures and to the depth discontinuity mentioned above.

VII. Conclusion

In this paper, we have presented an efficient warping-based method for stereoscopic image retargeting. We have formulated it as an energy minimization problem based on sparse stereo correspondences and their disparity constraints for obtaining the optimal warping fields for the images, and have extended it to interactive stereoscopic image editing. Without explicit scene reconstruction and occlusion filling, our method still generates appealing results. We conducted two user studies to show that our method is effective for many different retargeting scenarios.

We have only used feature matches as constraints to maintain symmetry and consistency between views. Although this is sufficient in most cases, our system would be more robust and convenient if a dense depth map were available. Such depth maps also facilitate better ROI detection that takes depth into account [27]. People are often interested in the regions that pop out from the screen, and typically pay less attention to background objects. Also, as autostereoscopic displays often need more than two views, usually calling for eight to sixteen views, extending our method to handle multi-view images or light fields would be an interesting research venue. Similar to 2D image retargeting, it is natural to extend our method to stereoscopic video display adaptation, where additional depth cues such as motion parallax must be taken into account.

Acknowledgment

We would like to thank Chun-Wei Liu and Ming-Hsu Chang for their help with user study, and the following Flickr members for their images: 47609087@N00, balliolman, neilp2, rawshooter72, robbiesydney, roger_harris and turbguy. The first author was partially supported by CyberLink Technical Elite Fellowship. This work was partly supported by NSC grants 99-2628-E-002-015 and 99-2622-E-002-026-CC2.

References

[1] Z. Wartell, L. F. Hodges, and W. Ribarsky, "A geometric comparison of algorithms for fusion control in stereoscopic HTDs," *IEEE Transactions on Visualization and Computer Graphics*, vol. 8, no. 2, pp. 129–143, 2002.
[2] F. L. Kooi and A. Toet, "Visual comfort of binocular and 3D displays," *Displays*, vol. 25, no. 2-3, pp. 99 – 108, 2004.
[3] M. Lambooij, W. A. IJsselsteijn, M. Fortuin, and I. Heynderickx, "Visual discomfort and visual fatigue of stereoscopic displays: A review," *Journal of Imaging Science and Technology*, vol. 53, no. 3, pp. 1–14, 2009.
[4] Y.-S. Wang, C.-L. Tai, O. Sorkine, and T.-Y. Lee, "Optimized scale-and-stretch for image resizing," *ACM Trans. Graph.*, vol. 27, no. 5, article 118, 2008.
[5] Y.-S. Wang, H. Fu, O. Sorkine, T.-Y. Lee, and H.-P. Seidel, "Motion-aware temporal coherence for video resizing," *ACM Trans. Graph.*, vol. 28, no. 5, article 127, 2009.
[6] F. Liu, M. Gleicher, H. Jin, and A. Agarwala, "Content-preserving warps for 3d video stabilization," *ACM Trans. Graph.*, vol. 28, no. 3, article 44, 2009.
[7] R. Carroll, A. Agarwala, and M. Agrawala, "Image warps for artistic perspective manipulation," *ACM Trans. Graph.*, vol. 29, no. 4, article 127, 2010.
[8] R. Carroll, M. Agrawal, and A. Agarwala, "Optimizing content-preserving projections for wide-angle images," *ACM Trans. Graph.*, vol. 28, no. 3, article 43, 2009.
[9] C. Wheatstone, "On some remarkable, and hitherto unobserved, phenomena of binocular vision," *Philosophical Transactions of the Royal Society of London*, vol. 128, pp. 371–394, 1838.
[10] C. Wang and A. A. Sawchukg, "Disparity manipulation for stereo images and video," in *Proceedings of Stereoscopic Displays and Applications XIX*, 2008.
[11] S. Koppal, C. L. Zitnick, M. Cohen, S. B. Kang, B. Ressler, and A. Colburn, "A viewer-centric editor for stereoscopic cinema," *IEEE Computer Graphics and Applications*, 2010, to appear.
[12] A. Shamir and O. Sorkine, "Visual media retargeting," in *SIGGRAPH Asia 2009 Course Notes*, 2009.
[13] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," *ACM Trans. Graph.*, vol. 26, no. 3, article 118, 2007.
[14] M. Rubinstein, A. Shamir, and S. Avidan, "Improved seam carving for video retargeting," *ACM Trans. Graph.*, vol. 27, no. 3, article 16, 2008.
[15] Y. Guo, F. Liu, J. Shi, Z.-H. Zhou, and M. Gleicher, "Image retargeting using mesh parametrization," *IEEE Transactions on Multimedia*, vol. 11, no. 5, pp. 856–867, 2009.
[16] S. Lee, D. Lischinski, Y. Yu, G. xin Zhang, M. ming Cheng, S. min Hu, and R. R. Martin, "A shape-preserving approach to image resizing," *Computer Graphics Forum*, vol. 28, no. 7, pp. 1897 – 1906, 2009.
[17] W.-J. Kim, S.-D. Kim, J. Kim, and N. Hur, "Resizing of stereoscopic images for display adaptation," in *Proceedings of Stereoscopic Displays and Applications XX*, 2009.
[18] M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic, and M. Gross, "Nonlinear disparity mapping for stereoscopic 3D," *ACM Trans. Graph.*, vol. 29, no. 4, article 75, 2010.
[19] M. Lambooij, W. A. IJsselsteijn, and I. Heynderickx, "Visual discomfort in stereoscopic displays: A review," *Proceedings of SPIE*, vol. 6490, 2007.
[20] B. Mendiburu, *3D Movie Making: Stereoscopic Digital Cinema from Script to Screen*. Focal Press, 2009.
[21] G. Sun and N. Holliman, "Evaluating methods for controlling depth perception in stereoscopic cinematography," in *Proceedings of SPIE*, vol. 7237, no. 1, 2009.
[22] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proceedings of Neural Information Processing Systems*, 2006.
[23] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, vol. 47, pp. 7–42, 2002.
[24] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, 2004.
[25] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, 2004.
[26] M. Brown, R. Szeliski, and S. Winder, "Multi-image matching using multi-scale oriented patches," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, 2005, pp. 510–517.
[27] Y. Zhang, G. Jiang, M. Yu, and K. Chen, "Stereoscopic visual attention model for 3D video," in *Proceedings of Multimedia Modeling*, 2010, pp. 314–324.

Chen-Han Chang received the B.S. degree and M.S. degree in Computer Science and Information Engineering from National Taiwan University in 2008 and 2010, respectively. He is currently a Ph.D. student of the Graduate Institute of Networking and Multimedia at National Taiwan University. His research interests include image processing and computer vision.

Chia-Kai Liang received the B.S. degree in electrical engineering and Ph.D. degree in communication engineering in 2004 and 2009 respectively, both from National Taiwan University. In 2009 he was a research intern in Nokia Research Center Palo Alto. In 2010 he joined Refocus Imaging Inc. He received CSVT Transactions Best Paper Award in 2008. His research interests include computational photography, image processing, computer vision, and multimedia hardware architecture. He is a member of Phi Tau Phi and ACM SIGGRAPH.

Yung-Yu Chuang received his B.S. and M.S. from National Taiwan University in 1993 and 1995 respectively, Ph.D. from University of Washington at Seattle in 2004, all in Computer Science. He is currently an associate professor in the Department of Computer Science and Information Engineering, National Taiwan University. His research interests include computational photography, multimedia, computer vision and rendering. He is a member of the IEEE and a member of the ACM.