

Perspective-Aware Warping for Seamless Stereoscopic Image Cloning

Sheng-Jie Luo¹

I-Chao Shen²

Bing-Yu Chen¹

Wen-Huang Cheng²

Yung-Yu Chuang¹

¹National Taiwan University

²Academia Sinica

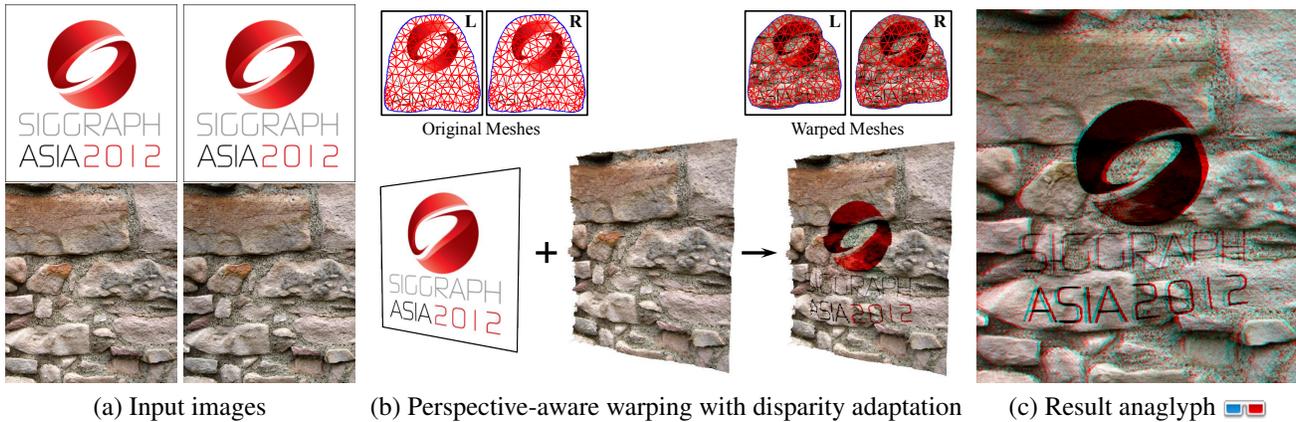


Figure 1: We present a novel technique for seamlessly cloning content from one stereoscopic image pair to another. Given a synthetic 3D SIGGRAPH Asia 2012 logo as the source image pair and a target stereoscopic image pair of a bumpy wall (a), we use perspective-aware warping to adjust the structure of the logo and paste it on to the bumpy wall (c). The perceived depth and projection of the pasted logo (b) are adjusted locally and adaptively to fit onto the bumpy surface. (Note that the resultant left and right images are included in the supplemental materials. It is recommended to watch them with stereoscopic displays for better visual effects.)

Abstract

This paper presents a novel technique for seamless stereoscopic image cloning, which performs both shape adjustment and color blending such that the stereoscopic composite is seamless in both the perceived depth and color appearance. The core of the proposed method is an iterative disparity adaptation process which alternates between two steps: disparity estimation, which re-estimates the disparities in the gradient domain so that the disparities are continuous across the boundary of the cloned region; and perspective-aware warping, which locally re-adjusts the shape and size of the cloned region according to the estimated disparities. This process guarantees not only depth continuity across the boundary but also models local perspective projection in accordance with the disparities, leading to more natural stereoscopic composites. The proposed method allows for easy cloning of objects with intricate silhouettes and vague boundaries because it does not require precise segmentation of the objects. Several challenging cases are demonstrated to show that our method generates more compelling results compared to methods with only global shape adjustment.

Keywords: Seamless cloning, stereoscopic images, image gradients, disparity gradients, Poisson equation, disparity adaptation, perspective-aware warping.

Links: [DL](#) [PDF](#) [WEB](#) [VIDEO](#)

1 Introduction

The success of stereoscopic 3D movies has ignited the rapid development of 3D cameras and displays for consumers. Once users are able to capture and display stereoscopic 3D media easily, the next requirement will be the ability to manipulate these 3D media similar to the way 2D media are manipulated. However, it can be challenging to directly apply 2D media editing tools to stereoscopic 3D media, because the additional information (i.e., depth) in stereoscopic 3D images introduces additional constraints in maintaining a comfortable and enjoyable 3D viewing experience. Naive extensions of existing 2D image editing methods usually fail as they do not take these constraints into account.

This paper focuses on stereoscopic image cloning, that is, selecting a region from a source image and pasting it into a target image. Although successful 2D image cloning methods have been proposed [Pérez et al. 2003; Jia et al. 2006; Farbman et al. 2009; Yang et al. 2009], stereoscopic 3D image cloning has its own challenges: (1) we must adjust disparity values within the cloned region for depth continuity; (2) we must alter the projected shape of the cloned region according to the disparity change to model perspective effects such as foreshortening; (3) we must maintain the coordination between the left and right views for comfortable 3D viewing. Lo *et al.* [2010] proposed a cut-and-paste system for stereoscopic images which uses a segmentation technique to accurately select the object that users intend to clone. Although their system successfully meets some of these challenges, it still has two shortcomings. First, it is difficult to accurately segment out objects with complex silhouettes and objects without obvious boundaries between them and the background. Second, their system models the cloned objects as stereo billboards, which is more effective for approximating objects standing on a ground plane and for flat objects without much depth variation.

To address the above challenges and limitations, we propose a novel technique for seamless stereoscopic image cloning which performs both shape adjustment and color blending such that the stereoscopic

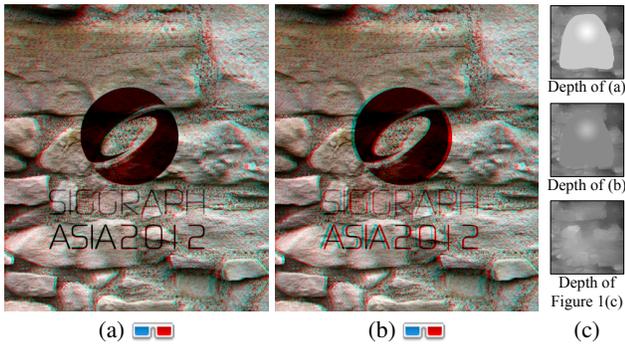


Figure 2: The results produced by two naive methods: (a) *direct-pasting* and (b) *global-adjustment*. (c) Depth map comparisons of the two naive methods and ours, which show that the two naive methods do not perfectly attach the logo to the bumpy surface, while the proposed method (the bottom one) generates a more compelling depth map.

composite is seamless in both the perceived depth and color appearance. In order to achieve this, we propose an iterative disparity adaptation procedure which consists of gradient-domain disparity estimation and perspective-aware warping, so that the depth structure of the cloned region is locally and adaptively adjusted to fit that of the target pasted location. In addition, the color of the cloned region is seamlessly blended with the target image using Poisson blending. Furthermore, to provide an easy-to-use selection mechanism, our method allows users to roughly select a region on only one view of the source stereoscopic image, after which the system automatically locates the corresponding region in the other view. As Poisson blending techniques and segmentation-based copy-and-paste methods for 2D images are suitable for different situations and are complementary to each other, we consider the proposed method complementary to Lo *et al.*'s method [2010] as well. Their method is more suitable for objects that can be easily segmented out and stand on the ground while ours is better suited for objects that are difficult to segment and attach on the ground.

Figure 1 shows a sample cloning result and Figure 2 shows the comparisons to two naive methods: *direct-pasting* (Figure 2(a)), which completely ignores the depth discrepancy, and *global-adjustment* (Figure 2(b)), which globally shifts the cloned region according to the average of disparities within it. Although the global-adjustment method (Figure 2(b)) compensates for depth discrepancy by globally shifting the underlying disparities, obvious depth discontinuities exist along the boundary. Neither method produces visually pleasing results because local adjustment should be performed over the cloned region to match the local perspectives and depth structures. In contrast, our perspective-aware warping as well as the iterative disparity adaptation locally and adaptively adjust the depth structure (Figure 1(c)). The depth maps from these methods (Figure 2(c)) show that our method provides the best depth composition. Specifically, the contributions of this paper include:

- a disparity adaptation technique which can be used to alter the composite depth structure such that it resembles the local shape of the cloned region and adapts to the target image;
- a novel perspective-aware warping method for locally adapting the disparities as well as the projected shape and size of the cloned region to those of the target location;
- a stereoscopic image cloning system which allows users to clone image regions to another image without accurate foreground segmentation.

2 Related Work

Image Composition. Image composition has attracted immense research attention for many years, and many techniques have been proposed. Matting-based methods are probably the most popular image cloning techniques [Wang and Cohen 2007]. They extract the foreground object with matting techniques, and then composite it into another image. Although the methods are quite effective, they can produce unnatural composites when colors between the source and target images are too different. Gradient-domain techniques address this problem effectively by manipulating image gradients, e.g., Poisson cloning [Pérez *et al.* 2003]. Jia *et al.* [2006] suggested finding the optimal boundary before blending. Farbman *et al.* [2009] introduced a mean-value interpolation scheme for seamless cloning. By doing away with the need to solve a large sparse matrix derived from the Poisson equation, the blending process can be accelerated dramatically. Yang *et al.* [2009] introduced additional color fidelity terms into the original Poisson cloning method for better visual results. In the past few years, other research [Lalonde *et al.* 2007; Chen *et al.* 2009; Ding and Tong 2010] has combined both gradient-domain and matting-based methods.

Stereoscopic media editing. Recently, research on stereoscopic 3D media editing is becoming more and more popular, and many methods have been proposed. The stereoscopic viewing experience is an important issue when watching stereoscopic content. Much scientific research has been devoted to analyzing what is needed for a comfortable stereoscopic viewing experience [Lambooi *et al.* 2007; Kim *et al.* 2011b]. For stereoscopic 3D film production, there are quite a few practice rules learned from experience in the 3D film industry [Mendiburu 2009]. For specific editing operators, Wang *et al.* [2008a] extended 2D image inpainting to stereoscopic 3D images. Koppal *et al.* [2011] formulated a mathematical framework to enable user-centric manipulation with several stereo parameters. Lang *et al.* [2010] proposed a framework for disparity manipulation. Chang *et al.* [2011] and Basha *et al.* [2011] addressed the stereoscopic media retargeting problem with continuous and discrete approaches, respectively. Kim *et al.* [2011a] generated stereoscopic pairs with per-pixel control over disparity using a lightfield dataset. Lo *et al.* [2010] is most related to our work as both focus on stereoscopic image cloning. Their method can be taken as the 3D counterpart of matting-based 2D image cloning approaches while ours belongs to the gradient-domain methods.

Image warping and deformation. Our method relates to warping-based image manipulation methods, which have been widely used to manipulate image structure. Igarashi *et al.* [2005] and Schaefer *et al.* [2006] proposed methods for deforming 2D shapes according to user-specified constraints while minimizing the distortion of local shapes. In recent years, image warping techniques have also been widely used in many image manipulation applications, such as image and video retargeting [Wang *et al.* 2008b; Shamir and Sorkine 2009], perspective manipulation [Carroll *et al.* 2010] and video stabilization [Liu *et al.* 2009]. In this paper, we propose a novel perspective-aware warping technique to locally adjust the depth structure of the cloned region to adapt the perceived depth and model the shape changes due to local disparity changes.

3 Stereoscopic Image Seamless Cloning

Figure 3 is an overview of our stereoscopic image cloning technique. The inputs are source and target stereoscopic image pairs, (I_S^l, I_S^r) (I_T^l, I_T^r) , as shown in Figure 3(a). They are assumed rectified. The system first estimates the disparity maps for both image pairs and extracts features for the source stereoscopic im-

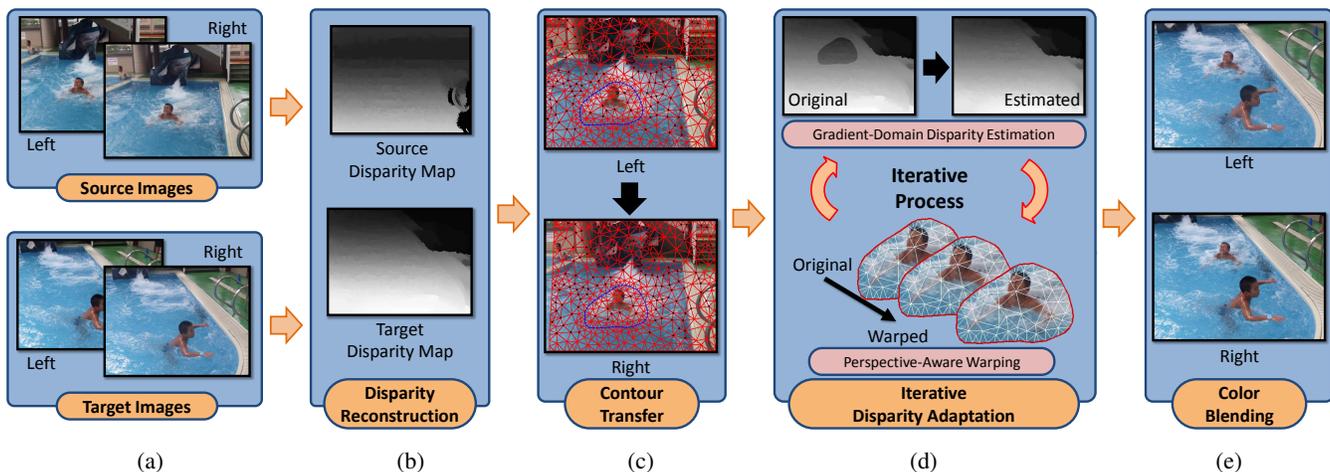


Figure 3: An overview of the proposed method. Given input source and target stereoscopic image pairs (a), their disparity maps (b) are estimated and corresponding feature pairs are extracted for the source image pair for offline preprocessing. To clone a region of the source image pair, users simply draw a contour on one source image to specify it, and the system automatically transfers the contour to the corresponding location with corresponding shape in the other image (c). Consistent meshes are also constructed for the source pair in this step. When pasting the cloned region to the target image pair, the disparities of the cloned region are adapted to fit those of the pasted location on the target image by iteratively performing gradient-domain disparity estimation and perspective-aware warping until convergence (d). Finally, the warped cloned region is seamlessly blended onto the target image using the Poisson blending method (e).

age (Figure 3(b)) in an offline process (Section 3.1). To clone a region of the source image pair, users simply draw a contour on one of its images, and the corresponding contour in the other image is automatically generated. Consistent meshes are constructed for both source images (Figure 3(c)), as described in Section 3.2, after which the user specifies a target location to which to paste the cloned region. To avoid depth seams between the cloned region and the target image, our disparity adaptation algorithm re-estimates the disparities of the cloned region in the gradient domain so that the disparities are continuous across the boundary of the cloned region and the target image. Perspective-aware warping is then adopted to adjust the shape and size of the cloned region according to the estimated disparities. The above estimation and warping processes are performed iteratively until the shape of the cloned region does not change much (Figure 3(d)), as described in Section 3.3. Finally, the warped source patches are blended into the target images using Poisson blending (Section 3.4). Figure 3(e) shows the final result which is seamless for both color and depth structures.

3.1 Preprocessing

Disparity map estimation. Horizontal disparity, the x -coordinate difference between corresponding pixels on the two retinal images, is the central factor in human depth perception. The proposed method is based on the reconstruction of a plausible disparity map from the source disparity map D_S and the target disparity map D_T , which are assumed to be estimated for the left image using the method proposed by Smith *et al.* [2009] (Figure 3(b)).

Feature extraction. The estimated disparity values may be inaccurate in certain areas because of limitations of the particular algorithm. Fortunately, previous research on disparity mapping and stereoscopic image retargeting [Lang *et al.* 2010; Chang *et al.* 2011] has demonstrated that plausible manipulation of depth perception can still be achieved by coupling image warping with sparse but reliable feature correspondences. Thus, our method relies on robust features to avoid the problem potentially caused by imperfect disparity maps. SIFT [Lowe 2004] is used to establish the corresponding feature pairs $\{(f_i^l, f_i^r)\}_{i=1..n_f}$ between the left and right

views for the source image¹; RANSAC with the fundamental matrix as the model is used to filter out outliers. Therefore, the disparity values of these robust features are quite accurate and can be used to guide image warping for contour transfer (Section 3.2) and disparity adaptation (Section 3.3).

3.2 Contour transfer and mesh construction

To specify the cloning region (Ω^l, Ω^r) we require a pair of contours $(\partial\Omega^l, \partial\Omega^r)$ in the source stereoscopic image pair. However, it is redundant (and likely inconsistent) to ask users to specify contours on both the left and right views. Our system requests users to draw the contour on one view only and automatically infers the corresponding contour for the other view. Without loss of generality, we assume that the users draw the contour $\partial\Omega^l$ on the left image and the right contour $\partial\Omega^r$ is generated by the system.

A contour is implemented as a closed polyline in our system. Assume that the left contour $\partial\Omega^l$ consists of n_c vertices $q_i^l \in \mathbb{R}^2$. A naive solution for contour transfer would be to find the corresponding point $q_i^r \in \mathbb{R}^2$ in the right image for each vertex q_i^l by looking up the estimated source disparity map, i.e., $q_i^r = q_i^l - (D_S(q_i^l), 0)$, where $D_S(q_i^l)$ denotes the disparity value of the vertex q_i^l in the disparity map D_S . However, because the user-drawn contour vertices are not necessarily located within salient regions, their disparity values may be inaccurate. Using only such estimated disparities would result in jaggy transferred contours (Figure 4(b)). For more robust contour transfer, we adopt mesh-based warping and reliable feature correspondences to avoid unreliable contour correspondences. We first construct a triangular mesh for the left source image. In addition to contour transfer, the mesh is also used for perspective-aware warping during iterative disparity adaptation (Section 3.3). Since the extracted features in Section 3.1 and user-drawn contour are reliable, it is preferred that the features and contour are parts of the triangular mesh. In addition, we also uniformly sample a set of points \mathbf{B} along the image border and use them as

¹Note that we do not need features for the target stereoscopic image because its disparity values will not be altered in the composite.

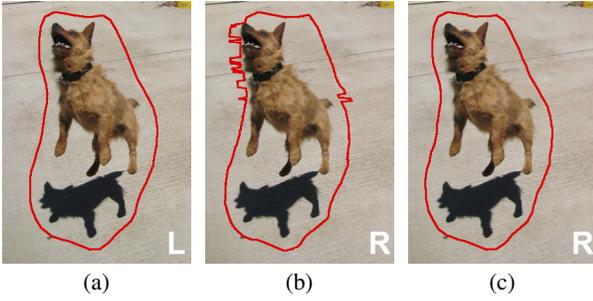


Figure 4: Contour transfer. (a) User-specified left contour. (b) Right contour transferred by finding the correspondence points in the right image using the estimated source disparities. The contour jaggedness is due to imperfect disparity values in the texture-less area. (c) The right contour transferred using our method.

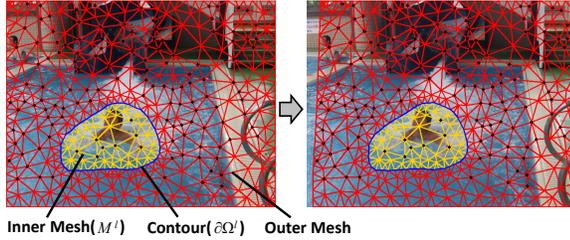


Figure 5: The mesh structure. The blue polyline in the left image represents the user-specified contour. Constrained Delaunay triangulation is used to construct the mesh for the left source image as shown on the left. A consistent mesh is constructed for the right image as shown on the right. Subsequent processing requires only the inner meshes (indicated by yellow).

the vertices of the mesh. Specifically, to construct the mesh of the left image, we perform a constrained Delaunay triangulation with $\{f_i^l | i = 1..n_f\} \cup \{q_i^l | i = 1..n_c\} \cup \mathbf{B}$ as vertex constraints and line segments $\{q_1^l q_2^l, q_2^l q_3^l, \dots, q_{n_c}^l q_1^l\}$ as edge constraints. The left of Figure 5 shows an example of the constructed mesh. Let \mathbf{U}^l be the set of all vertices of the left mesh, which contains features, contour vertices, border vertices, and vertices added during triangulation. To transfer the left mesh to the right image, we seek vertex set \mathbf{U}^r satisfying the following constraints:

(1) *Feature correspondence.* Given a reliable feature correspondence (f_i^l, f_i^r) , the corresponding vertex in the right mesh ($v_i^r \in \mathbf{U}^r$) of the feature f_i^l should closely approach f_i^r . This term is defined as $E_f = \sum_{i=1}^{n_f} (v_i^r - f_i^r)^2$.

(2) *Vertical alignment.* Because the image pair is rectified, each vertex’s y -coordinate in the right mesh should be close to that of the corresponding vertex in the left mesh. This term is defined as $E_v = \sum_{v_i^l \in \mathbf{U}^l} (v_i^r[y] - v_i^l[y])^2$, where the operator $[y]$ extracts the y -coordinate of a vertex. We don’t enforce it as a hard constraint because the human is somewhat tolerant to vertical drift.

(3) *Triangle shape distortion.* The triangles should not be distorted too much from the left mesh to the right mesh; this can be enforced by requiring each triangle in the right mesh to resemble its counterpart in the left mesh. For a single triangle with vertices (v_i^l, v_j^l, v_k^l) in the left mesh, vertex v_i^l can be represented by the other two vertices as $v_i^l = F_l(v_j^l, v_k^l)$, where F_l means the transformed coordinate of v_i^l in the local frame formed by v_j^l and v_k^l as defined in [Igarashi et al. 2005]. We formulate the term as

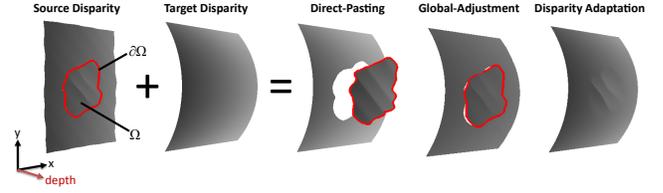


Figure 6: Two naive methods and our disparity adaptation. With direct-pasting, the pasted region “floats” on the target image. Although global-adjustment globally adjusts disparities, depth discontinuities along the boundary cause unnatural shape interpretation. With our approach, the region is seamlessly pasted onto the target image.

$$E_s = \sum_{(v_i^l, v_j^l, v_k^l) \in \mathbf{T}^l} \|v_i^r - F_l(v_j^r, v_k^r)\|^2, \text{ where } \mathbf{T}^l \text{ is the set of triangles in the left mesh.}$$

By minimizing an energy function combining these three constraints, $E = w_f E_f + w_v E_v + w_s E_s$, we obtain the vertices of the right mesh \mathbf{U}^r . We used $w_f = 10$, $w_v = 10$, and $w_s = 1$.

After finding the vertices \mathbf{U}^r , the same connectivity as the left mesh is applied to complete the mesh structure for the right image. The set of corresponding vertices of the left mesh’s contour vertices $\{q_i^l | i = 1..n_c\}$ in \mathbf{U}^r specify the transferred contour for the right image. Figure 4(c) shows that the transferred contour using the proposed approach is much better than the naive method (Figure 4(b)). Figure 5 shows the constructed mesh on the left image and the transferred result on the right image for the example shown in Figure 3. For our cloning application, only the inner meshes of the left and right images, $(\mathbf{M}^l, \mathbf{M}^r)$, are needed in following processes. Hence, the cloning regions and contours can be inferred from the inner meshes. Let $\Omega(\mathbf{M})$ be the region operator which returns the region defined by mesh \mathbf{M} . We have $\Omega^l = \Omega(\mathbf{M}^l)$ and $\Omega^r = \Omega(\mathbf{M}^r)$. Similarly, let $\partial\Omega(\mathbf{M})$ be the contour operator which returns the contour defined by \mathbf{M} ; thus we have $\partial\Omega^l = \partial\Omega(\mathbf{M}^l)$ and $\partial\Omega^r = \partial\Omega(\mathbf{M}^r)$.

3.3 Iterative disparity adaptation

Given the source disparity map D_S , target disparity map D_T , and meshes \mathbf{M}^l and \mathbf{M}^r specifying the cloned region in the source image, the goal of *iterative disparity adaptation* is a proper disparity map and mesh structure which adapt to both the shape and perspective of the target image while closely approximating the original local shape of the cloned region.

The disparities of adjacent pixels along the boundary of the cloned region may be discontinuous if composited directly or incorporated using a simple global adjustment. Thus, for seamless cloning, *gradient-domain disparity estimation* is used to re-estimate the disparities within the source cloned region in the gradient domain to remove the disparity discontinuities along the boundary while maintaining the local shape of the cloned region (Section 3.3.1). Figure 6 compares the two naive methods with ours.

An important characteristic of perspective projection is *foreshortening*: objects become smaller as their distances from the observer increase. In other words, the projected size of an object depends on its disparity (depth). This provides a strong cue for the position (depth) of an object. When there are large changes in disparity due to gradient-domain disparity estimation or when there is a significant discrepancy between the perspectives of the source and target images, the 2D shape of the cloned region must be adjusted according to the adapted disparity values (depths) to match such

changes or discrepancies. We propose *perspective-aware warping* to accomplish this (Section 3.3.2).

The disparity adaptation process is iterative. In the first step, given the meshes, we re-estimate the disparities within the cloned region in the gradient domain. In the second step, we adjust the meshes according to the adapted disparities (depths) using perspective-aware warping, after which the locations and shapes (both contours and interior structures) of the meshes are updated, potentially leading to disparity adjustment within the cloned region. Adjusted disparities can also lead to shape changes in the cloned objects. Therefore, perspective-aware warping and target disparity adjustment are mutually dependent; this dependency is addressed using an iterative method which alternately adjusts the shapes and disparities until convergence.

3.3.1 Gradient-domain disparity estimation

For seamless cloning, the disparity values must be adjusted to remove any disparity discrepancy along the boundary of the cloned region. Lang *et al.* [2010] indicate that disparity gradients significantly influence human depth perception, and that it is possible to alter the perceived depth structure by manipulating these disparity gradients. Inspired by their results, we adopt a gradient-domain approach for adapting disparity values to the shape of the target image. By using the desired disparity gradients as the guidance field G and enforcing the disparity values along the boundary in the new source disparity map to be the same as the target disparity map, we obtain the adapted source disparity map \hat{D}_S by minimizing the following weighted Poisson equation:

$$\Psi_g(\hat{D}_S) = \iint_{\Omega^l} \omega(D_S) \left| \nabla \hat{D}_S - G \right|^2 \text{ with } \hat{D}_S|_{\partial\Omega^l} = D_T|_{\partial\Omega^l}, \quad (1)$$

where $\omega(D_S) \propto 1/|\nabla D_S|$ is the weighting function defined over the pasted region, which is used to penalize the pixels in smoother depth regions more because there is no need to maintain the gradient in depth-discontinuous regions (i.e., the boundaries of foreground and background objects). The guidance field is obtained as $G = \nabla D_S$ for general cases, which keeps the shape of the pasted object. Alternatively, to attach the cloned region onto the target image, the field can be calculated as $G = \max(\nabla D_S, \nabla D_T)$. Note that the new disparity map \hat{D}_S is defined only within the cloned region Ω^l . In addition, the guidance field may be warped and region Ω^l may be deformed if mesh \mathbf{M}^l has been deformed by perspective-aware warping (Section 3.3.2).

3.3.2 Perspective-aware warping

As mentioned above, the projective size of an object depends on its depth under perspective projection. To better match the perspective and adapted depth obtained from Section 3.3.1, the size and shape of the cloned region must be adjusted according to the disparity values. We first sketch the relationship between the projective size and the disparity value and then describe how to warp the meshes according to the disparity values.

The relationship between scale and disparity. The scale of an object on an image depends on its depth, which is related to the disparity. We seek to derive the relationship between the scale and disparity so that we can scale the object according to its disparity. First, we derive the relationship between depth and disparity. Here, we assume a parallel stereoscopic camera setting, which is a common configuration for consumer stereoscopic cameras (Figure 7). From Figure 7(a), the image disparity is $d = x_L - x_R$, where x_L and x_R are the x -coordinates of a pair of corresponding points. Note that x_R is negative in this particular example. According to the similar triangle theorem, we have $(b-d)/b = z/(z+f)$,

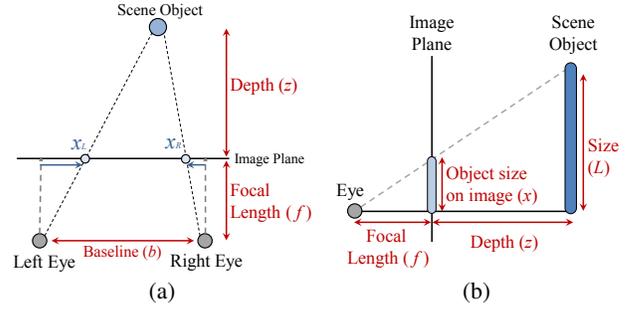


Figure 7: Stereoscopic camera configuration for inferring the relationship between the projective size and the disparity.

where b is the baseline distance between two cameras; f is the focal length and z is the depth of the scene point. From this equation, we have the relationship between the depth z and the disparity d as $z = (b-d)f/d$. Next, we relate the disparity value with the scale by the depth. From Figure 7(b), we know that the projective size x is related to object size L as $x = Lf/(z+f)$. Substituting $z = (b-d)f/d$ in this equation, we have the relationship between the projective size and disparity as $x = (L/b)d$. Note that although focal length plays a role during the derivation, it disappears in the end. Thus it is not a required parameter.

Perspective-aware warping. For the i -th vertex of the left mesh \mathbf{M}^l , we obtain the associated scale factor s_i according to the ratio between the current and original disparity values, i.e., $s_i = \hat{x}_i/x_i = \hat{D}_S(\hat{v}_i^l)/D_S(v_i^l)$, where \hat{v}_i^l and v_i^l are the current (after deforming) and original (obtained in Section 3.2) vertex positions, respectively; D_S and \hat{D}_S are the original (obtained in Section 3.1) and updated (calculated in Section 3.3.1) disparity maps, respectively. The scale factors describe how the meshes should be deformed so that their sizes and shapes match the updated depths. Taking Figure 6 as an example, to match the target shape, the source patch must move back to attach to the target shape. Thus the disparity values decrease and the scale factors are smaller than 1. In this case, the cloned region should shrink in size to adapt to the target image. In more complex examples where vertices have different scaling factors, the shape of the mesh may be distorted. In addition to the scale constraints, stereoscopic constraints are used to make sure the deformation maintains good stereoscopic properties. The goal of perspective-aware warping is to deform the meshes to satisfy both scale and stereoscopic constraints. We design an energy function with the following four terms to deform the meshes by finding their new vertex positions \hat{v}_i^l and \hat{v}_i^r .

Disparity-dependent scaling. For each edge (v_i^l, v_j^l) of the left mesh \mathbf{M}^l , given the associated scale factors, s_i and s_j , assuming the scale factor varies linearly along the edge, the scale factor for the edge is $l_{ij} = \frac{1}{2}(s_i + s_j)$. The same scale factor is applied to the corresponding edge in the right mesh. Hence we have the disparity-dependent scaling term

$$\Phi_s(\hat{v}_i^l, \hat{v}_i^r) = \sum_{(v_i^l, v_j^l) \in \mathbf{E}^l} \left(\|\hat{v}_i^l - \hat{v}_j^l\| - l_{ij} \|v_i^l - v_j^l\| \right)^2 + \sum_{(v_i^r, v_j^r) \in \mathbf{E}^r} \left(\|\hat{v}_i^r - \hat{v}_j^r\| - l_{ij} \|v_i^r - v_j^r\| \right)^2, \quad (2)$$

where \mathbf{E}^l and \mathbf{E}^r are the sets of edges of \mathbf{M}^l and \mathbf{M}^r , respectively. This energy term promotes meshes with the desired edge lengths derived from the disparity values.

Disparity consistency. This constraint forces the disparities of all vertices \mathbf{V}^l within the cloned region to be consistent with the es-

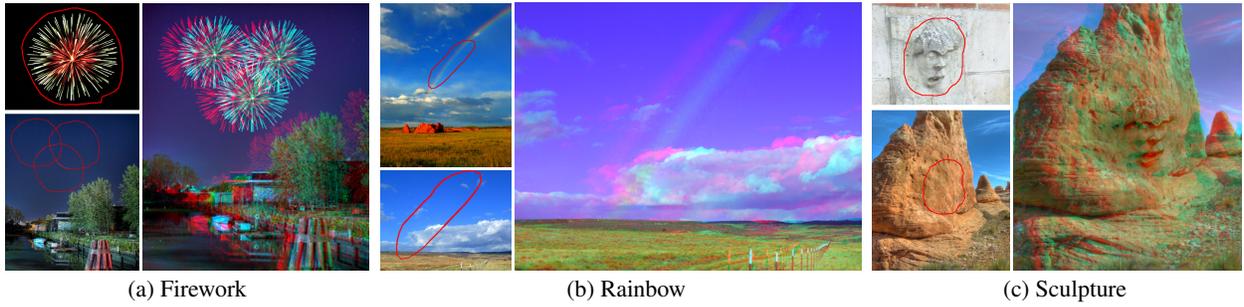


Figure 8: A number of cloning results using our method.  Firework has intricate silhouettes, rainbow contains transparency, and sculpture has vague boundaries. These cases are challenging for segmentation-based approaches.

estimated disparities. For each vertex pair $(v_i^l, v_i^r) \in \mathbf{V}$, their x -coordinate difference should approximate the current estimated disparity. Let d_i be the target disparity for the i -th feature pair according to the current disparity map, i.e., $d_i = \hat{D}_S(\hat{v}_i^l)$. We have the disparity consistency term:

$$\Phi_d(\hat{v}_i^l, \hat{v}_i^r) = \sum_{(\hat{v}_i^l, \hat{v}_i^r) \in \hat{\mathbf{V}}} \left((\hat{v}_i^l[x] - \hat{v}_i^r[x] - d_i) \right)^2, \quad (3)$$

where operator $[x]$ extracts the x -component of a 2D vector.

Vertical alignment. To avoid unwanted vertical parallax, this constraint restricts the left and right meshes to preserve vertical alignment after warping. We have

$$\Phi_v(\hat{v}_i^l, \hat{v}_i^r) = \sum_{(\hat{v}_i^l, \hat{v}_i^r) \in \hat{\mathbf{V}}} \left(\hat{v}_i^l[y] - \hat{v}_i^r[y] \right)^2, \quad (4)$$

where operator $[y]$ extracts the y -component of a 2D vector. It is not enforced as a hard constraint because the human vision system is somewhat tolerant to vertical drift; we would like to take advantage of this flexibility.

Position fixation. The above three constraints restrict only the shapes of the meshes but not their positions. Thus, we add a position fixation term so that the location of the cloned region is close to the user-specified position despite changed shape and size. We choose to fix the left mesh center. Let c^l be the center of the initial left mesh, i.e., $c^l = \frac{1}{|\hat{\mathbf{V}}^l|} \sum_{v_i^l \in \mathbf{V}^l} v_i^l$. The position fixation term is

$$\Phi_p(\hat{v}_i^l, \hat{v}_i^r) = \left\| c^l - \frac{1}{|\hat{\mathbf{V}}^l|} \sum_{\hat{v}_i^l \in \hat{\mathbf{V}}^l} \hat{v}_i^l \right\|^2. \quad (5)$$

The energy function is a weighted sum of the above four terms, $\Phi = w_s \Phi_s + w_d \Phi_d + w_v \Phi_v + w_p \Phi_p$, where $w_s = 50$, $w_d = 50$, $w_v = 100$, and $w_p = 1$ for all results in this paper. The energy function is non-linear because of the disparity-dependent scaling term. We optimized it using the steepest-descent method. For the initial guess, the meshes are globally shifted to the average disparities of the pasted location.

3.4 Color blending

Iterative disparity adaptation outputs the adapted left and right meshes $\hat{\mathbf{M}}^l$ and $\hat{\mathbf{M}}^r$ for the source stereoscopic image pair. Let $W(I, \mathbf{M}, \hat{\mathbf{M}})$ be the warping operator which returns the warped image patch by mapping the original mesh \mathbf{M} to target mesh $\hat{\mathbf{M}}$.

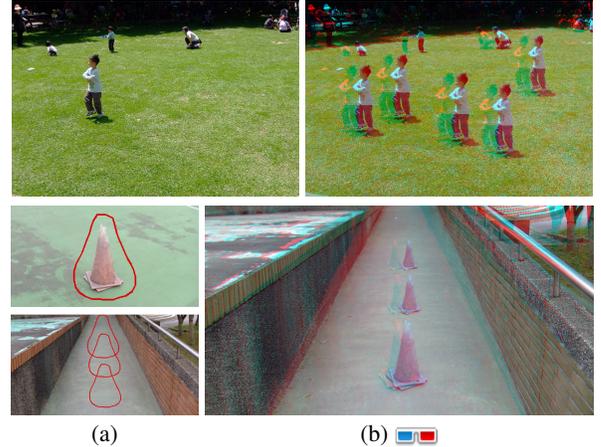


Figure 9: Automatic resizing. (a) The left view of input stereoscopic images. (b) The cloned objects pasted at several places. The scales of the cloned objects are automatically inferred using perspective-aware warping.

We obtain the warped cloned region pair $\hat{P}^l = W(I_S^l, \mathbf{M}^l, \hat{\mathbf{M}}^l)$ and $\hat{P}^r = W(I_S^r, \mathbf{M}^r, \hat{\mathbf{M}}^r)$, where \mathbf{M}^l and \mathbf{M}^r are the initial meshes obtained in Section 3.2. The system then seamlessly blends the color channels of \hat{P}^l and \hat{P}^r with the target images I_T^l and I_T^r , respectively using Poisson blending [Pérez et al. 2003] to obtain the final composite.

4 Results

In this section, we demonstrate several stereoscopic image cloning results using the proposed method. The input stereoscopic images are either collected from Flickr or captured by a Fujifilm FinePix W3 camera. The results are presented as red-cyan anaglyph images in the paper; the uncompressed left and right images are included in the supplemental materials. We encourage readers to view them with stereoscopic displays for better visual effects.

Figure 8 shows several cloning examples for images for which users may have difficulty precisely specifying the cloned regions because of complex silhouettes (Figure 8(a)), transparency (Figure 8(b)), or vague boundaries (Figure 8(c)). It would be difficult to achieve good results using Lo *et al.*'s segmentation-based method [2010]. In contrast, our method does not need to precisely segment out the objects, and is effective for these cases. Figure 9 shows that our method automatically adjusts the sizes of the pasted objects using perspective-aware warping.

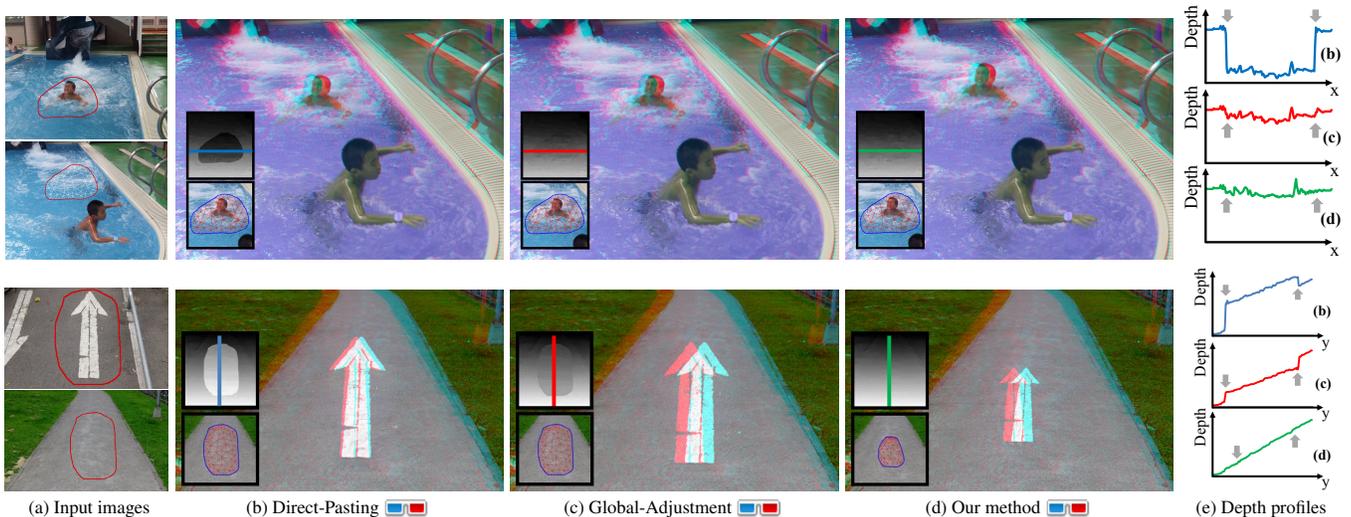


Figure 10: Comparisons with direct-pasting and global-adjustment. Our results look better by taking into account both depth continuity and foreshortening. On the left of each result (b)~(d), we show the composited disparity map (upper) and the mesh (bottom). The depth profiles (e) obtained from the same scanline for these methods show that our method yields the best depth composition. Arrows in (e) indicate composite boundaries. Note the visible depth discontinuities along the boundaries for direct-pasting and global-adjustment.

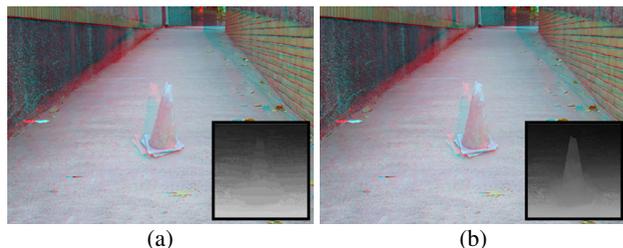


Figure 11: Depth estimation using traditional Poisson equation (a) and our weighted version (Eq.(1)) (b). Both the composite image and the composited disparity map (inset) are shown.

We compare our method with two naive approaches: direct-pasting and global-adjustment of the cloned region. Figures 1 and 2 show the comparison of pasting the SIGGRAPH Asia 2012 logo onto a bumpy surface. Our approach is able to locally adjust the cloned logo such that it is attached to the surface. In contrast, direct-pasting and global-adjustment cannot perfectly adapt to the depth variation of the bumpy surface. Figure 10 shows other cases for comparison. Figure 10(b) show the results of direct-pasting; neither the disparity nor the perspective is adjusted. In the upper row, the cloned swimmer and water region appear far behind the target water surface, while in the lower row, the arrow floats on the road and its perspective does not match the road’s. The global-adjustment method corrects the depth discontinuity somewhat (Figure 10(c)), but only adjusts the disparities of the cloned regions globally and does not take into account perspective foreshortening. Thus the global method still produces noticeable depth discontinuousness. Our results are better because local disparities are adapted (Figure 10(d)). The depth profiles show the superiority of our method over the others (Figure 10(e)).

Figure 11 compares the results generated with the traditional Poisson equation ($\omega(D_S) = 1$ in Eq.(1)) and our weighted version (Eq.(1)). The cloned region contains a cone and a part of the ground. The disparity estimation based on the traditional equation forces the disparity gradients to resemble the guidance gradient field at every point. When pasted into the target image, the

Example	Resolution of source image	# Δ of meshes (entire/inner)	Contour Transfer	Disparity adaptation (3 iterations)
Figure 1	309×319	408/236	0.004	8.385
Figure 8(b)	477×600	626/88	0.001	0.187
Figure 8(c)	472×637	737/172	0.011	1.326
Figure 9(Upper)	904×467	929/73	0.01	0.115
Figure 10(Upper)	837×654	1095/146	0.011	0.104
Figure 10(Lower)	950×513	1113/250	0.011	6.77

Table 1: Statistics for several examples. Time measured in seconds.



Figure 12: A failure case. A large perspective change leads to an unnatural composite. View interpolation is required in this case.

cone leans backwards toward the ground (Figure 11(a)), because the source ground pulls the cone towards the ground to match the less tilted target ground. With our weighted Poisson equation, the disparity gradients need not resemble the guidance gradient field where there are large disparity gradients; this allows the cone to stand up as in the source image (Figure 11(b)).

Performance. Table 1 shows statistics for several examples including image resolution, number of triangles of the entire/inner meshes, and processing time for contour transfer and iterative disparity adaptation (3 iterations are sufficient for convergence). All results were generated on a desktop PC with an Intel Core 2 Duo 3.2GHz CPU and 4GB RAM.

Limitations. Our method shares with Lo *et al.*’s method [2010] the limitation for extremely large perspective differences between the source and target images. In these cases, it is often not plausible to approximate the large 3D perspective changes by 2D image warp-

ing only (Figure 12). This problem could be addressed to some degree using depth-image-based rendering or perspective manipulation [Carroll et al. 2010]. The proposed approach relies heavily on disparity maps and features. Fortunately, existing stereo and feature matching techniques mostly lead to visually plausible results. However, our method still suffers from inaccurate disparity maps and insufficient features.

5 Conclusion

Stereoscopic image cloning is challenging because both the depths and colors of the cloned regions must be seamlessly blended with the target stereoscopic images. We have presented a novel gradient-domain technique to reduce depth discontinuity and maintain global depth structure so that the cloning results are visually plausible and natural. The proposed method has the following advantages: First, our method allows for easy cloning of objects with intricate silhouettes and vague boundaries because it does not require precise segmentation of the objects. Second, the proposed iterative disparity adaptation guarantees not only depth continuity across the boundary but also models perspective foreshortening. Third, the depth structure within the pasted region is better preserved instead of being highly approximated.

Two research directions warrant further exploration. First, view interpolation or depth-image-based rendering might be necessary for handling large perspective differences between the source and target images. Second, we would like to enhance the user controllability for stereoscopic image cloning. For example, as users might want to control the depths of the pasted objects, the disparity values should be adjusted to satisfy both the users' control and plausibility.

Acknowledgments: We thank the anonymous reviewers for their detailed comments. We also thank the following Flickr users for the use of their photos: Wayne Karberg (turbguy) (Fig.8(b) source and target, Fig.8(c) target), -ytf- (Fig.8(a) target), Patrick McDonald (clayspur) (Fig.8(c) source), pinboke_planet (Fig.10(upper) source and target), and Dan Ridley-Ellis (Dan) (Fig.1 target, Fig.8(a) source). This research was supported in part by the National Science Council of Taiwan under grants NSC100-2622-E-002-016-CC2, NSC101-2628-E-002-031-MY3, and NSC101-2221-E-001-016.

References

BASHA, T., MOSES, Y., AND AVIDAN, S. 2011. Geometrically consistent stereo seam carving. In *Proc. IEEE ICCV '11*, 1816–1823.

CARROLL, R., AGARWALA, A., AND AGRAWALA, M. 2010. Image warps for artistic perspective manipulation. *ACM TOG* 29, 4, 127:1–127:9.

CHANG, C.-H., LIANG, C.-K., AND CHUANG, Y.-Y. 2011. Content-aware display adaptation and interactive editing for stereoscopic images. *IEEE TMM* 13, 4, 589–601.

CHEN, T., CHENG, M.-M., TAN, P., SHAMIR, A., AND HU, S.-M. 2009. Sketch2Photo: internet image montage. *ACM TOG* 28, 5, 124:1–124:10.

DING, M., AND TONG, R.-F. 2010. Content-aware copying and pasting in images. *The Visual Computer* 26, 6-8, 721–729.

FARBMAN, Z., HOFFER, G., LIPMAN, Y., COHEN-OR, D., AND LISCHINSKI, D. 2009. Coordinates for instant image cloning. *ACM TOG* 28, 3, 67:1–67:9.

IGARASHI, T., MOSCOVICH, T., AND HUGHES, J. F. 2005. As-rigid-as-possible shape manipulation. *ACM TOG* 24, 3, 1134–1141.

JIA, J., SUN, J., TANG, C.-K., AND SHUM, H.-Y. 2006. Drag-and-drop pasting. *ACM TOG* 25, 3, 631–637.

KIM, C., HORNUNG, A., HEINZLE, S., MATUSIK, W., AND GROSS, M. 2011. Multi-perspective stereoscopy from light fields. *ACM TOG* 30, 6, 190:1–190:10.

KIM, J., HOFFMAN, D. M., AND BANKS, M. S. 2011. The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of Vision* 11, 8, 1–29.

KOPPAL, S., ZITNICK, C. L., COHEN, M., KANG, S. B., RESSLER, B., AND COLBURN, A. 2011. A viewer-centric editor for 3D movies. *IEEE CG&A* 31, 1, 20–35.

LALONDE, J.-F., HOIEM, D., EFROS, A. A., ROTHER, C., WINN, J., AND CRIMINISI, A. 2007. Photo clip art. *ACM TOG* 26, 3, 3:1–3:10.

LAMBOOIJ, M. T. M., IJSSELSTEIJN, W. A., AND HEYNDE-ICKX, I. 2007. Visual discomfort in stereoscopic displays: a review. In *Proc. SPIE*, vol. 6490, 64900I.

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

LIU, F., GLEICHER, M., JIN, H., AND AGARWALA, A. 2009. Content-preserving warps for 3D video stabilization. *ACM TOG* 28, 3, 44:1–44:9.

LO, W.-Y., VAN BAAR, J., KNAUS, C., ZWICKER, M., AND GROSS, M. 2010. Stereoscopic 3D copy & paste. *ACM TOG* 29, 6, 147:1–147:10.

LOWE, D. G. 2004. Distinctive image features from scale-invariant keypoints. *IJCV* 60, 2, 91–110.

MENDIBURU, B. 2009. *3D Movie Making: Stereoscopic Digital Cinema from Script to Screen*. Focal Press.

PÉREZ, P., GANGNET, M., AND BLAKE, A. 2003. Poisson image editing. *ACM TOG* 22, 3, 313–318.

SCHAEFER, S., MCPHAIL, T., AND WARREN, J. 2006. Image deformation using moving least squares. *ACM TOG* 25, 3, 533–540.

SHAMIR, A., AND SORKINE, O. 2009. Visual media retargeting. In *ACM SIGGRAPH Asia '09 Courses*, 11:1–11:13.

SMITH, B., ZHANG, L., AND JIN, H. 2009. Stereo matching with nonparametric smoothness priors in feature space. In *Proc. IEEE CVPR '09*, 485–492.

WANG, J., AND COHEN, M. F. 2007. Simultaneous matting and compositing. In *Proc. IEEE CVPR '07*, 1–8.

WANG, L., JIN, H., YANG, R., AND GONG, M. 2008. Stereoscopic inpainting: Joint color and depth completion from stereo images. In *Proc. IEEE CVPR '08*.

WANG, Y.-S., TAI, C.-L., SORKINE, O., AND LEE, T.-Y. 2008. Optimized scale-and-stretch for image resizing. *ACM TOG* 27, 5, 118:1–118:8.

YANG, W., ZHENG, J., CAI, J., RAHARDJA, S., AND CHEN, C. W. 2009. Natural and seamless image composition with color control. *IEEE TIP* 18, 11, 2584–2592.