

StereoPasting: Interactive Composition in Stereoscopic Images

Ruo-Feng Tong, Yun Zhang, and Ke-Li Cheng

Abstract—We propose “*StereoPasting*,” an efficient method for depth-consistent stereoscopic composition, in which a source 2D image is interactively blended into a target stereoscopic image. As we paint “disparity” on a 2D image, the disparity map of the selected region is gradually produced by edge-aware diffusion, and then blended with that of the target stereoscopic image. By considering constraints of the expected disparities and perspective scaling, the 2D object is warped to generate an image pair, which is then blended into the target image pair to get the composition result. The warping is formulated as an energy minimization, which could be solved in real time. We also present an interactive composition system, in which users can edit the disparity maps of 2D images by strokes, while viewing the composition results instantly. Experiments show that our method is intuitive and efficient for interactive stereoscopic composition. A lot of applications demonstrate the versatility of our method.

Index Terms—StereoPasting, depth consistent, stereoscopic composition, disparity map

1 INTRODUCTION

WITH the ever increasing availability of 3D media such as 3D movie and TV, more and more people prefer to watch 3D content due to its exciting visual experiences and enjoyment. Thus, it becomes more and more important and urgent to create and manipulate stereoscopic images and videos in an intuitive and efficient way. In this paper, we focus on *image composition*, a basic and important operation in image editing, which involves selecting a patch from a source image and pasting it into any desired location of a target image. Recently, much progress [5], [12], [15], [17], [24] has been made for natural composition in 2D images. These methods aim at removing the visible seam along the cloning boundary and adjusting the color appearance of the cloned patch to fit the target scene. Unlike 2D images, stereoscopic images are taken by binocular devices like stereo cameras, and humans feel depth from the disparity in image pairs. Therefore, naive extension of these methods to stereoscopic images, which processes left and right images separately, usually breaks disparity consistency and causes visual discomfort.

Recently, many efforts have been made in stereoscopic media editing, such as manipulating perspective in stereoscopic images [11], StereoBrush [32], 3D Copy&Paste [21], nonlinear disparity mapping [18], content-aware stereoscopic image editing [4], and stereoscopic inpainting [31]. However, relative few works focus on stereoscopic composition, and their methods are still not robust and efficient enough for practical use. Lo et al. proposed 3D Copy&Paste

[21], which is state-of-the-art in stereoscopic image composition. They provided an interactive tool to jointly segment objects from image pairs, and proposed a stereo billboard method to paste the selected object pairs into the target image pairs. Their method is effective to preserve the stereo volume of the copied objects, while dynamically adjusting their scale and orientation in the composition results. Although successful in many examples, 3D Copy&Paste [21] focuses on pasting stereoscopic foregrounds into stereoscopic backgrounds, which cannot handle the case of pasting 2D images into stereoscopic images. In general, stereoscopic images are scarce and hard to process, while 2D images are easier to access and edit. In addition, it is more intuitive and convenient for users to paste 2D images into stereoscopic scenes. Thus, taking 2D images as sources will increase the utility of stereoscopic composition.

To address the problems above, we propose *StereoPasting*, an efficient method for depth-consistent stereoscopic composition, in which a source 2D image is interactively blended into a target stereoscopic image. We first define some terminologies for clarity of this paper: we refer to the object selected from a source 2D image as *2D foreground*, and the target stereoscopic image as *3D background*. For *depth-consistent* composition, we first allow users to edit the disparity map of the 2D image by strokes, which is then blended with the disparity values of the 3D background at the pasting locations so as to avoid depth discontinuity. Then the 2D foreground is warped to generate an image pair by an energy optimization with constraints of the disparities after blending and perspective scaling, which could be solved instantly. Finally, the warped image pair is seamlessly blended into the target image pair to produce visual-pleasing stereopsis in the composition results. Based on *StereoPasting*, we present an interactive composition system, in which users can edit the disparity maps of 2D foregrounds by strokes, while receiving immediate feedback in the stereoscopic composition results.

- The authors are with the State Key Lab of CAD&CG, Institute of Artificial Intelligence, Zhejiang University, Room 401, Cao Guangbiao Building, Yuquan Campus, Hangzhou 310027, China.
E-mail: {trf, zhangyun_zju, chengkeli}@zju.edu.cn.

Manuscript received 5 June 2012; revised 22 Sept. 2012; accepted 19 Nov. 2012; published online 12 Dec. 2012.

Recommended for acceptance by M. Agrawala.

For information on obtaining reprints of this article, please send e-mail to: tcvg@computer.org, and reference IEEECS Log Number TVCG-2012-06-0102. Digital Object Identifier no. 10.1109/TVCG.2012.319.

Contributions. To the best of our knowledge, we are the first to blend 2D foregrounds into 3D backgrounds with the following contributions:

- We propose an optimization-based formulation for depth-consistent stereoscopic composition, which could be solved instantly. To ensure depth consistency, we propose improved solutions to edit the disparity maps of 2D foregrounds by strokes and blend them with the disparity values of 3D backgrounds.
- We present a novel system for interactive stereoscopic composition, and users can dynamically edit the disparity maps of 2D foregrounds according to the instant feedback in the stereoscopic compositions.

2 RELATED WORK

Stereoscopic image and video editing is a challenging task that has received much attention in recent years. Here we briefly review the techniques most related to our work.

2D image composition. For 2D image composition, there exist two main categories: alpha matting and gradient-domain method. The recent progress on matting techniques is reviewed in [30]. For image matting, many methods are based on optimizations, such as Bayesian approach [7], closed-form solution [19], soft scissors [29], and shared sampling [13]. Very recently, He et al. proposed global sampling [14], which is motivated by the search space proposed in PatchMatch [2]. Although having achieved great success, alpha matting may fail to produce good results when there exist large differences between the source and target images or unclear boundaries such as water, smoke. To solve these problems, Pérez et al. proposed Poisson image editing [24], which applies the gradient-domain method to image composition for the first time. Its key idea is to construct a membrane to smoothly interpolate the discrepancies along the cloning boundary. Drag-and-drop [15] improved poisson image editing [24] by calculating an optimized boundary. Farbman et al. proposed mean-value coordinates (MVC) [12] to efficiently construct a membrane by a weighted sum of values along the cloning boundary rather than solving a large linear system. To eliminate smudging and discoloration artifacts in composition, several methods [5], [10], [17], [40] are proposed to combine gradient-domain method and alpha matting for better visual effects. Our system allows users to roughly select objects with unclear boundary, and the warped image pairs are cloned into the target image pairs using the combination of alpha matting and gradient-domain method.

Stereoscopic media creation. Recently, Ward et al. [34] presented an interactive system (Depth Director) for converting traditional 2D video to stereoscopic using existing computer vision algorithms, which allows users to interactively manipulate the depth between objects. Different from [34], Wang et al. [32] proposed *StereoBrush*, a novel work flow for creating stereoscopic content from 2D images/videos, in which users paint strokes on the 2D images/videos, while instantly receiving the corresponding 3D content. Assuming that the depth is piecewise continuous,

a discontinuous warping method is proposed for this conversion. To improve the efficiency in stereoscopic creation, Kim et al. [16] proposed a single image model that represents all regions from stereo views, and stereoscopic image pairs can be easily by disparity adjustment. Oskam et al. [23] presented a solution for interactive stereoscopic applications by dynamically optimizing the camera parameters, showing 3D content with optimal 3D viewing. Zhang et al. [39] produced stereoscopic videos from monocular ones by exploiting the motion parallax, while avoiding recovering the depth map. Wan et al. [28] presented a fast stereoscopic rendering algorithm for the volumetric environments, and further extended the method to produce view-dependent shading and transparent effects. Inspired by *StereoBrush* [32], we blend 2D foregrounds into stereoscopic images, and the stereopsis of the foregrounds can be created by painting strokes. However, *StereoBrush* [32] only considers color similarity in the disparity diffusion, while ignoring unclear edges which cannot be easily distinguished by previous edge-aware methods [1], [32], [36], [37]. Actually unclear edges are very common in natural images, so detection of them is important for many tasks in image editing. In this paper, we allow users to specify unclear edges in the foreground images by strokes, and improve the traditional color propagation framework by introducing user interaction.

Stereoscopic image and video editing. In stereoscopic image and video editing, recent works focus on disparity editing for human visual comfort, and extending the 2D editing to the stereoscopic domain. The warping-based method is an effective framework in many applications of image and video editing. Very recently, Du et al. [11] proposed a warping-based method to manipulate perspective in stereoscopic images. They formulated the warping as an energy minimization, which preserves correct geometric information such as straight lines and ensures proper stereopsis in new perspectives. Pollock et al. [25] focused on depth perception within the stereoscopic multiuser virtual environments (VEs). Their findings helped us to understand the spatial distortions in multiuser VEs, and provided possible solutions for reducing distortion. Chang et al. [4] proposed a warping-based technique to efficiently resize stereoscopic images by solving a quadratic energy minimization. Similar to [4], [11], our method also formulates the stereoscopic composition as a warping-based optimization with several constraints for visual-pleasing stereopsis in the composition results. Since disparity values are key for human to perceive depth, Lang et al. [18] proposed a series of disparity mapping operators to change the disparity range of stereoscopic images and videos. Motivated by Lang et al. [18], we reconstruct the disparity values of the pasted objects to fit the target stereoscopic scene using the gradient-domain operator. To estimate the magnitude of perceived disparity changes, a perceptual model of disparity is presented in [9]. Stereoscopic inpainting [31] is proposed to jointly complete the missing color and depth for stereo pairs. 3D Copy&Paste [21], a state-of-the-art in stereoscopic image composition, is mostly related to our work. In their work, an interactive tool is proposed to segment objects from image pairs, and they introduced a

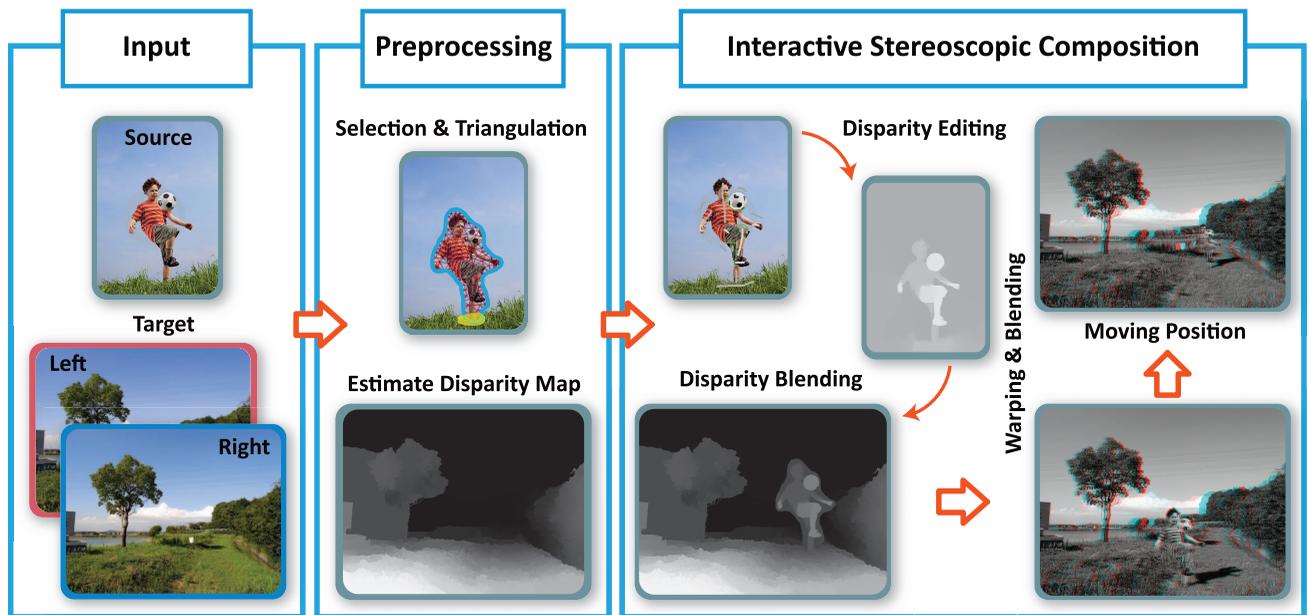


Fig. 1. Pipeline of *StereoPasting*. In the preprocessing step, we first select and triangulate the 2D foreground, then estimate the disparity map of the target scene. After that we edit the disparity map of the 2D foreground by painting strokes and blend it with the 3D background for depth-consistent composition. At last, the 2D foreground is warped and blended into the target image pair to get the composition results.

stereo billboard method to paste objects into stereoscopic scenes. Compared with [21], our work makes the stereoscopic composition more intuitive and general from the following aspects: 1) We take 2D objects as sources, which are much easier to obtain and edit. Further, our method allows pasting objects with unclear boundary such as water, smoke, while [21] can only paste objects which are precisely segmented. 2) We propose an optimization-based method to blend 2D foregrounds into 3D backgrounds. By painting “disparity” on a 2D foreground, users can create the foreground’s stereopsis according to their needs. Actually, Lo et al. [21] cannot work well when taking 2D images as sources, as it is difficult to initialize a correct geometric relationship between objects and their supporting surfaces by painting “disparity” on the 2D images, and this error will be magnified after the surface alignment.

3 OVERVIEW

Fig. 1 shows the pipeline of *StereoPasting*. The input to our approach is a 2D foreground (2D source image) and a 3D background (target stereoscopic image). In addition, our method requires the user-specified strokes with strength for editing the disparity map of the 2D foreground, specifying the pasting locations and depth-consistent boundary. The goal is to paste a 2D foreground into a 3D background. Our method allows users to iteratively paint strokes to edit the disparity maps of 2D foregrounds, while receiving the stereoscopic composition results in real time, which consists of the following key steps.

Preprocessing. In this step, we first select a foreground from the 2D image, and a triangular mesh is constructed on the selected region, which is used for warping in our stereoscopic composition. The triangular mesh is generated by the constrained *Delaunay* triangulation [26] with constraints of a set of uniformly sampled points on the selected

boundary. Then we estimate the dense disparity map of the 3D background, which is important to reconstruct the disparities of the pasted 2D foreground for depth-consistent composition. Assuming that the stereoscopic images are rectified with only parallel disparities, we apply the method proposed by Smith et al. [27] to estimate the disparity map of the 3D background. The middle of Fig. 1 shows the triangular mesh on the selected foreground and the estimated disparity map.

Stroke-based disparity editing. The goal of this step is to allow users to arbitrarily edit the disparity map of the 2D foreground. As users paint “disparity” on the 2D image, the disparity map is constructed progressively.

Depth-consistent disparity blending. This step aims at reconstructing the disparity map of the 2D foreground by blending it with the disparity values of the 3D background. For *depth-consistent* composition, we need to make the disparity values continuous along the depth-consistent boundary, while discontinuous along the depth-inconsistent boundary.

Warping and blending. In this step, the 2D foreground and the corresponding alpha matte are warped to generate image pairs with constraints of the expected disparities and perspective scaling. We formulate the warping as an energy minimization which can be solved instantly. After warping, we blend the foreground pair into the background pair using the combination of gradient-domain method and alpha matting [5], [10], [17], [40], such that the color appearance of the foreground can well fit the background and the composition result is natural and visual-pleasing.

4 STEREOPASTING

4.1 Stroke-Based Disparity Editing

Inspired by *StereoBrush* [32], which interactively converts a 2D image to a stereoscopic image by painting strokes, we

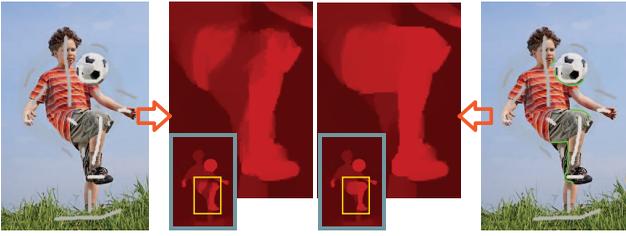


Fig. 2. Stroke-based disparity editing. Left two images: user-specified strokes and the visualization of disparity map produced by *StereoBrush* [32]. Right two images: user-specified strokes (latent boundaries marked by green lines), and the visualization of disparity map produced by our method. The zoom-in views show the comparison of the two methods.

paint “disparity” on a 2D foreground to produce its stereopsis in the composition results. However, *StereoBrush* [32] only considers color similarity, and cannot distinguish “latent” boundary which refers to the unclear edges in 2D images.

To avoid unexpected diffusion around the latent boundary, we improve the traditional value propagation framework by introducing user interactions. See Fig. 2, we allow users to specify the unclear edges by strokes (marked with *green* lines), and the disparity map u is calculated by minimizing the energy defined in (1). Compared with previous edge-aware methods for value propagation [32], [36], [37], we add a new term to distinguish the latent boundary, see the right part of the multiplication symbol in (3)

$$E(u) = \sum_{p \in I} (w_d(p)(u(p) - d(p))^2 + \nabla \mathbf{u}_p^T \mathbf{w}_p \nabla \mathbf{u}_p), \quad (1)$$

where $d(p)$ refers to the values on the strokes, and $w_d(p)$ specifies the mask of the strokes. $\nabla \mathbf{u}$ denotes the gradient of the disparity map u along the x - and y -axis. \mathbf{w}_p refers to a 2×2 diagonal matrix whose diagonal elements are $w_x(p)$ and $w_y(p)$.

$$w_d(p) = \begin{cases} \infty, & p \in \text{strokes}, \\ 0, & \text{otherwise}, \end{cases} \quad (2)$$

and

$$\begin{bmatrix} w_x(p) \\ w_y(p) \end{bmatrix} = \begin{bmatrix} e^{-\frac{\|\nabla f_x(p)\|^2}{\sigma}} \\ e^{-\frac{\|\nabla f_y(p)\|^2}{\sigma}} \end{bmatrix} \cdot \frac{1}{(1 + L(p))^\beta}. \quad (3)$$

In (3), $w_x(p)$ and $w_y(p)$ are used to measure the similarity of nearby pixels in an edge-aware manner. The left part of the multiplication symbol measures the similarity of nearby pixels. ∇f_x , ∇f_y refer to the gradient values on the x -, y -axis of the source 2D image, and σ is used to control the importance of color similarity in determining the affinity between two pixels. The right part of the multiplication symbol is designed to avoid unexpected diffusion near the latent boundary. With the user-specified latent boundary, $L(p)$ is constructed by feathering the latent boundary by 2-4 pixels, and then normalized to $[0, 1]$ in the feathered area ($L(p) = 0$ for other area). We use β to control the discontinuity near the latent boundary. In this paper, we choose $\sigma = 0.1, \beta = 10$ for all examples, which produces satisfying disparity diffusion. Compared with *StereoBrush*

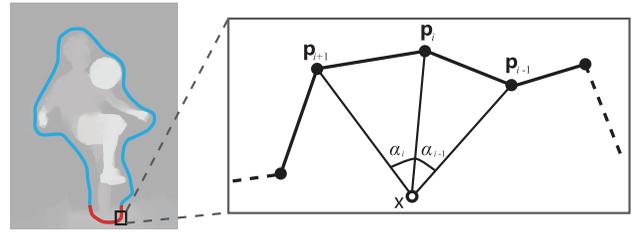


Fig. 3. Definition of *weighted* MVC [12]. Left: the source disparity map and the blending boundary (red and blue parts refer to the depth consistent and inconsistent boundary respectively); Right: angle definition of MVC.

[32] which only considers color similarity in nearby regions (see left two images in Fig. 2), our method can easily distinguish unclear discontinuities by specifying the latent boundary (see the right two images in Fig. 2). We give visualization of the disparity maps produced by our method and *StereoBrush* [32], and the zoom-in views show the advantages of our method.

4.2 Depth-Consistent Disparity Blending

The stroke-based disparity editing only provides relative disparities of the 2D foreground, which should be blended with the disparities of the 3D background for depth-consistent composition. Lang et al. [18] proposed a gradient domain operator which is effective to change human depth perception. Inspired by that, we proposed *weighted* MVC [12] to reconstruct the disparities of 2D foregrounds in the gradient domain. By *weighted*, we mean that the weights of MVC need to be adjusted according to the depth consistency between the 2D foreground and 3D background. The main idea of our method is to avoid interpolation from the depth inconsistent boundary. As shown in Fig. 3, the blending boundary is drawn by red and blue colors, which indicate depth consistent and inconsistent boundary respectively. w_i is calculated using the definition in MVC for vertices in the red boundary, and set to zero for vertices in the blue boundary. In our system, the *weighted* MVC $\{\lambda_i\}$ can be calculated as follows:

$$\lambda_i = \frac{w_i}{\sum_0^{m-1} w_i}, i = 0, \dots, m - 1, \quad (4)$$

where

$$w_i = \begin{cases} \frac{\tan(\alpha_{i-1}/2) + \tan(\alpha_i/2)}{\|p_i - x\|}, & p_i \in \text{continuous}, \\ 0, & \text{otherwise}. \end{cases} \quad (5)$$

Fig. 3 gives the definition of α_i , p_i , and x (refer to [12] for more details). Actually, it is time consuming to calculate the MVC of a selected region. However, the MVC of the pasted region are fixed after having selected the 2D foreground, and therefore need to be calculated only once, which ensures the efficiency of disparity blending.

As shown in Fig. 4, after specifying the source patch, we paint the boundary near the feet with yellow strokes to show that the depth along the boundary of the stroke area is consistent with that of the target scene, while discontinuous in the other parts of the boundary. Fig. 4 shows the disparity blending results using our method and traditional



Fig. 4. Depth-consistent disparity blending. Left to right: source 2D image (yellow region marks the depth-consistent boundary) and the disparity values of 2D foreground; disparity blending result by our method; disparity blending result by traditional MVC [12].

MVC [12], respectively. Obviously, our method can produce satisfying disparities to ensure a *depth-consistent* composition, while the traditional MVC cannot.

4.3 Warping and Blending

Unlike composition in 2D images, which only considers the color appearance of pasted objects, stereoscopic composition is nontrivial due to the added depth dimension. We first need to warp the 2D foreground into an image pair with constraints of the disparity constraint $E_d(\mathbf{v})$ and perspective scaling $E_s(\mathbf{v})$, which adjust the shape and scale of the image pair. To ensure a depth-consistent composition, the depth consistency constraint $E_c(\mathbf{v})$ is added to fix the position of the warped foreground. To warp the 2D foreground, we construct meshes for the foreground image pair. Specifically, given the expected disparity values, we compute the vertices $\mathbf{v} = \{\mathbf{v}_k, k = 1, \dots, N\}$ (N is the number of vertices of the left and right meshes) by minimizing the energy $E(\mathbf{v})$, which is a linear combination of $E_d(\mathbf{v})$, $E_s(\mathbf{v})$ and $E_c(\mathbf{v})$ (see (6)). We elaborate each of the energy terms as follows:

$$\arg \min_{\mathbf{v}} E(\mathbf{v}) = \arg \min_{\mathbf{v}} (w_d E_d(\mathbf{v}) + w_s E_s(\mathbf{v}) + w_c E_c(\mathbf{v})). \quad (6)$$

Disparity constraint $\{E_d(\mathbf{v})\}$. See (7), $E_d(\mathbf{v})$ is used to ensure horizontal disparity consistency and vertical disparity alignment. For each vertex pair \mathbf{v}_k^l and \mathbf{v}_k^r in left and right meshes, their horizontal differences should approximate the disparity values after blending, and the vertical differences of each vertex pair should approach zero to eliminate visual discomfort [4], [18]. \mathbf{d}_k in (7) is defined as a vector that gives the expected vertical and horizontal disparities of each vertex pair

$$E_d(\mathbf{v}) = \sum_{\{\mathbf{v}_k^r, \mathbf{v}_k^l\} \in \mathbf{v}} \|\mathbf{v}_k^r - \mathbf{v}_k^l - \mathbf{d}_k\|^2. \quad (7)$$

Perspective scaling $\{E_s(\mathbf{v})\}$. According to perspective projection, the scale of an object in a stereoscopic scene depends on the perceived depth. In [35], the relationship of perceived depth Z_p and disparity D_p is defined as

$$Z_p = \frac{Kb}{b - D_p}, \quad (8)$$

where p is a point in the 3D space, x_p^l , x_p^r are the projections on the left and right images, and $D_p = x_p^r - x_p^l$. K is the

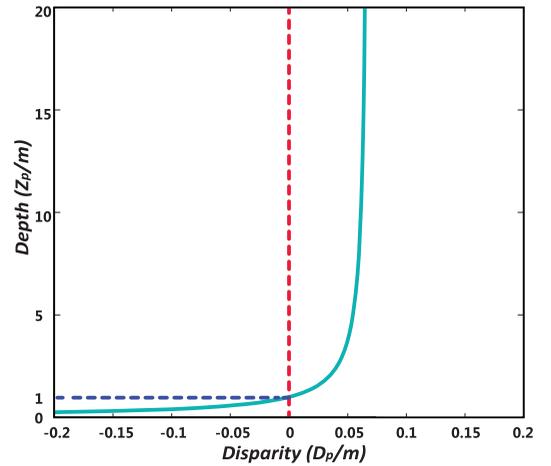


Fig. 5. Relationship of disparity and depth. We set $b = 0.068$ and $K = 1.0$ (blue dotted line) in (8) (red dotted line is a watershed for depth changes).

distance from the observer's eyes to the display plane, and b refers to the distance between the observer's eyes. We set $b = 0.068$ and $K = 1.0$ in this paper. Equation (8) shows that the perceived depth is nonlinear to the disparity. As shown in Fig. 5, when the disparity is negative, D_p increases very fast while Z_p increases a little. However, as the disparity reaches positive, D_p increases very slowly even when Z_p has a large increment. This is in accord with the human visual system, in which the depth perception is more sensitive to nearby objects [18]. Therefore, we focus on pasting objects into the sensitive depth perception region for better 3D viewing in this paper.

Given the relation between disparity and perceived depth in (8), we calculate the perceived depth $\{Z_{v_i}\}$ and $\{Z_{v_j}\}$ for all the vertices of the left and right meshes before and after disparity blending, and the scale factor S_{v_i} is defined by $Z_{v_i}/Z_{v_i}^o$. For each edge $e_{i,j} = \langle \mathbf{v}_i^o, \mathbf{v}_j^o \rangle$ in the original mesh, the expected scaling factor $S_{e_{i,j}}$ is given by $(S_{v_i} + S_{v_j})/2$. Inspired by Wang et al. [33], we define $E_s(\mathbf{v})$ as the distortion energy of the triangular mesh for perspective scaling by the expected scales. See (9), the distortion energy is the sum of distortion in the left and right meshes

$$E_s(\mathbf{v}) = \sum_{\langle \mathbf{v}_i, \mathbf{v}_j \rangle \in e^l} \|(\mathbf{v}_i - \mathbf{v}_j) - S_{e_{i,j}}(\mathbf{v}_i^o - \mathbf{v}_j^o)\|^2 + \sum_{\langle \mathbf{v}_i, \mathbf{v}_j \rangle \in e^r} \|(\mathbf{v}_i - \mathbf{v}_j) - S_{e_{i,j}}(\mathbf{v}_i^o - \mathbf{v}_j^o)\|^2. \quad (9)$$

Depth consistency $\{E_c(\mathbf{v})\}$. For the energy terms above, $E_d(\mathbf{v})$ is used to ensure correct disparities between the left and right meshes, and $E_s(\mathbf{v})$ is used to adjust the shape and scale of the left and right meshes. However, they are not enough to produce satisfying results. Therefore, we need to add another term $E_c(\mathbf{v})$ to fix the position of the warped foreground so as to keep depth consistent after warping. This is achieved by fixing the centroid of the depth-consistent boundary S in the left mesh before and after warping. See (10), \mathbf{p}_c is the centroid of S before warping and \mathbf{v}_m^l refers to all the vertices of S after warping. For depth-consistent composition, the disparity values of the warped 2D foreground need to be continuous with that of the 3D

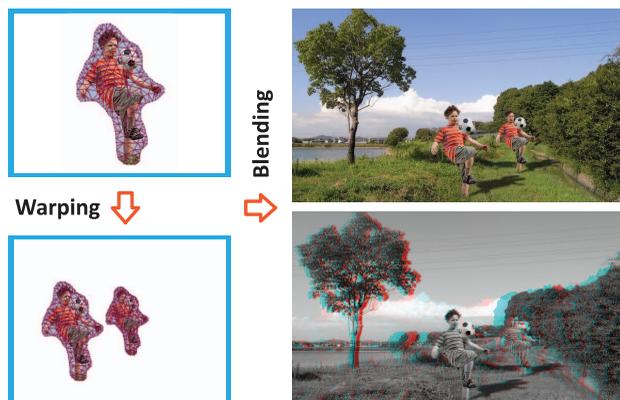


Fig. 6. Warping and blending. Left: the original mesh and the warped left meshes of the 2D foreground at different locations. Right: the left image of the cloning result and the corresponding *red&cyan* anaglyph.

background along the depth-consistent boundary, which ensures that the boy is standing on the grassland with visual-comfortable stereopsis (see Fig. 6)

$$E_c(\mathbf{v}) = \left\| \left(\frac{1}{M} \sum_{m \in S} \mathbf{v}_m^l \right) - \mathbf{p}_c \right\|^2. \quad (10)$$

Optimization. The three energy terms above are necessary to make satisfying compositions, and the analysis of these terms is as follows: $E_d(\mathbf{v})$ is used to constrain the relationship of corresponding vertices in the left and right meshes; $E_s(\mathbf{v})$ is used to control the shape and scale of the meshes; and $E_c(\mathbf{v})$ is used to determine the position of the warped foreground for depth-consistent composition. From the derivation in the supplemental materials, which can be found on the Computer Society Digital Library at <http://doi.ieeecomputersociety.org/10.1109/TVCG.2012.319>, we know that the warping energy $E(\mathbf{v})$ defined in (6) can be easily minimized by solving a sparse linear system ($Ax = b$, A is symmetric and positive definite (SPD)). In experiments, we found that A should become singular by leaving out any of the three energy terms, and the solution would be uncontrollable with less constraints, thus making the composition results unacceptable. See the second row of Fig. 7, the left and right images are composition results by leaving out $E_c(\mathbf{v})$ and $E_d(\mathbf{v})$, respectively. We can see that without $E_c(\mathbf{v})$, the 2D foreground can be properly warped but the absolute position is not determined, which makes the composition depth-inconsistent. Without $E_d(\mathbf{v})$, the relationship between the left and right meshes is not considered, thus the warped foreground is invisible in the right image of the composition result.

We have tried many different weights for the energy minimization, and the results show some differences. However, the human eye is not sensitive enough to differences in the results when the weights are set in a certain range. In particular, from experiments we find that the stereopsis of the composition result is satisfying when $w_d/w_s > 0.2$, and w_c is assigned any value (usually near the range of w_d and w_s) except 0. In general, we should place adequate emphasis on $E_d(\mathbf{v})$ to produce visual-pleasing stereopsis of the foreground, and too small values of w_d could make the blended foreground flat, which is the

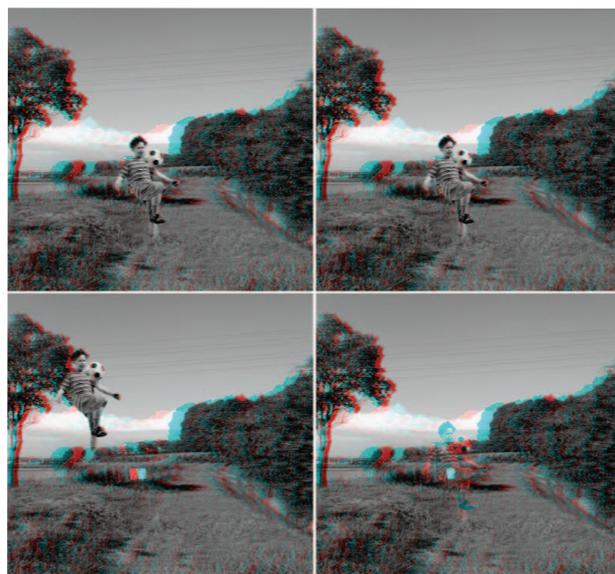


Fig. 7. Composition results by different weights and leaving out some energy terms. First row: gray scale *red&cyan* anaglyph results by setting ($w_d = w_s = w_c = 1$) and ($w_d = 10, w_s = 20, w_c = 1$), respectively. Second row: gray scale *red&cyan* anaglyph results by leaving out $E_c(\mathbf{v})$ and $E_d(\mathbf{v})$, respectively.

cardboarding effect in stereo [21]. As Fig. 7 illustrates, even when using very different weights the results are satisfying for human eyes. In our experiments, we simply set them to equal weights ($w_d = w_s = w_c = 1$) to achieve satisfying results.

We further use an optimization to solve the linear system instantly, which makes the warping efficient for interactive composition. From the derivation in the online supplemental materials, we find that A is SPD and is fixed after specifying the pasted region and depth continuous boundary, while b needs to be updated when moving the pasted 2D foreground or editing the disparity map by strokes. Therefore, we can first decompose A using the concise sparse *Cholesky* factorization [8], and then apply the back-substitution to solve the linear system. In terms of time complexity, the factorization is $O(N^3)$, while the back-substitution is only $O(N)$. Since A needs to be decomposed only once, the linear system can be solved instantly.

Having obtained the deformed left and right triangular meshes, the warped foreground pair and the corresponding alpha matte pair can be efficiently produced by texture mapping. After that the foreground pair is blended into the target image pair at the depth-consistent locations, using the combination of gradient-domain method and alpha matting [5], [10], [17], [40]. Fig. 6 shows the warped left meshes and the *red&cyan* anaglyph results when placing the foreground at different locations. To enhance the sense of reality, we interactively add shadows to the 2D foregrounds by specifying the direction and distance of the light source. Then the shadows can be produced by projections according to the shape of the foreground objects, and blended into the stereoscopic background using our method. In general, the 3D background always has large depth variations for shadow blending. However, our method can nonuniformly warp the 2D shadows to be stereoscopic. As shown in the right

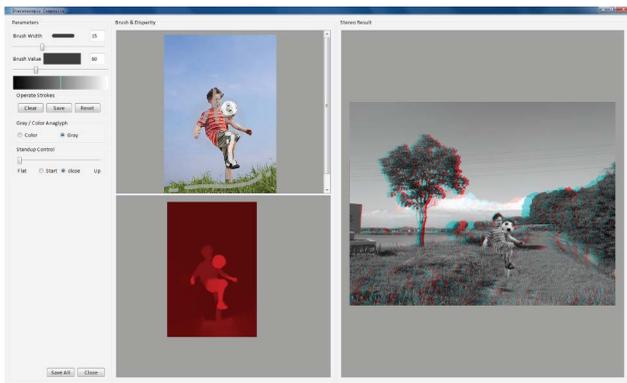


Fig. 8. A screenshot of the interactive stereoscopic composition system.

of Fig. 6, the final results are more visual-pleasing and realistic by adding the shadows.

4.4 Interactive Stereoscopic Composition

We propose a novel system for interactive stereoscopic composition, in which users can freely edit the disparity maps of 2D foregrounds by strokes, while viewing the stereoscopic composition instantly. Fig. 8 shows the screenshot of our interactive composition system. After loading a 2D foreground and a 3D background as input, we can interactively edit the disparity map of the 2D foreground by painting strokes. As users draw sparse scribbles with specified widths and values on the 2D foreground, the disparity map is calculated and visualized on the left-bottom side, then the warped image pair is blended into the target stereoscopic image instantly using our method. In our system, users can interactively edit the disparity maps of 2D foregrounds according to instant feedback in the stereoscopic composition results. The accompanying video demo will further demonstrate the advantages of our composition system.

5 RESULTS AND APPLICATIONS

In this section, we will show a variety of depth-consistent stereoscopic composition results, and several applications that benefit from our approach. Then, the performance of our method for the examples in this paper is given. In this paper, the composition results are shown by *red&cyan* anaglyph images, so users need to wear *red&cyan* stereo glasses for convincing 3D viewing experiences.

5.1 Interactive Composition

Fig. 9 shows three examples (from top to bottom), in which the disparity map of the 2D foreground has to be blended with that of the 3D background, so as to ensure depth consistency. In the first example, a car is blended into a road with large depth variations, and this case cannot be well handled by the stereo billboard method in [21]. Using our method, the disparities of the car can be reconstructed to adapt to the road, then a visual-pleasing composition is produced after warping and blending. For the second example, we blend a cartoon character into a street. After painting “disparity” on the 2D foreground and marking the depth-consistent region (see top-right corner of the left image), the object is warped according to the specified

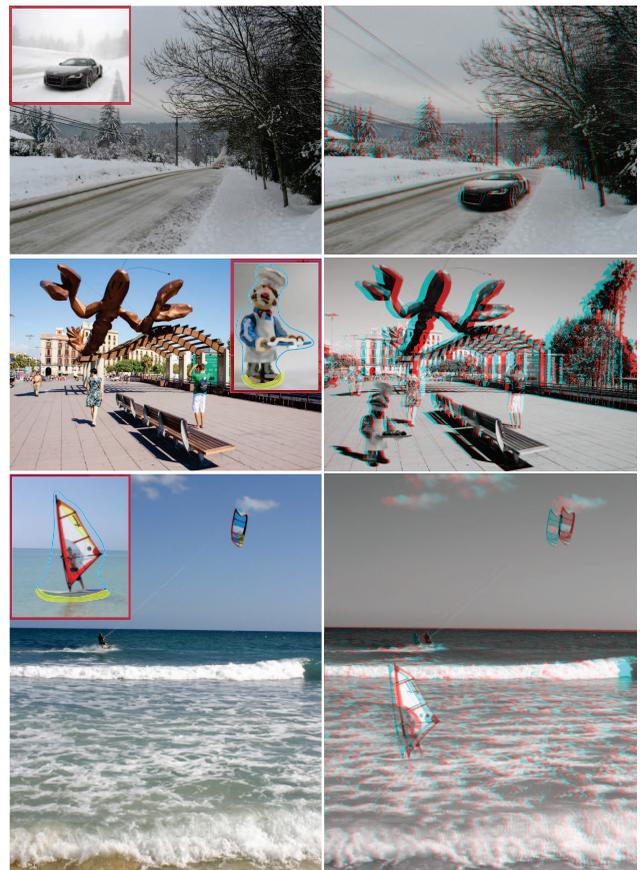


Fig. 9. Composition results by disparity blending. Left: left image of the 3D background (source 2D images and user interactions are shown on the top side). Right: *red&cyan* anaglyph results.

disparity map, and pasted into the stereoscopic scene using our method. The composition result shows a convincing stereopsis of the cartoon, and the added shadow further enhances visual effects. In the last example, we simply mark yellow strokes on the sailboat to ensure a depth-consistent composition. After the warping, the warped sailboat pair is blended into the target image pair using image cloning with the hybrid boundary [5], such that the color appearance of the sailboat can fit the target scene very well.

Unlike the examples in Fig. 9, there are no regions in the target scenes to determine the disparities of the 2D foregrounds in Fig. 10, for example, birds and planes are flying in the air. In this case, we need to reconstruct the foreground’s disparity map by specifying values in a certain range which is evaluated from the 3D background, thus we do not need to recover the target disparity map in these examples. As we change the disparity values by dragging a slider, the perceived depth of the pasted objects can be adjusted. See the examples in Fig. 10, the two birds are placed near and far from the people, and the jet planes are located to different distances from the girl by specifying different disparities. For the second example in Fig. 10, it is difficult to segment the jet plane which has unclear boundary, and this case cannot be handled by Lo et al. [21]. However, in our approach, we can first warp the roughly selected jet plane, and then blend the warped foreground pair with the air using the gradient-domain method [12]. Finally, the jet plane can

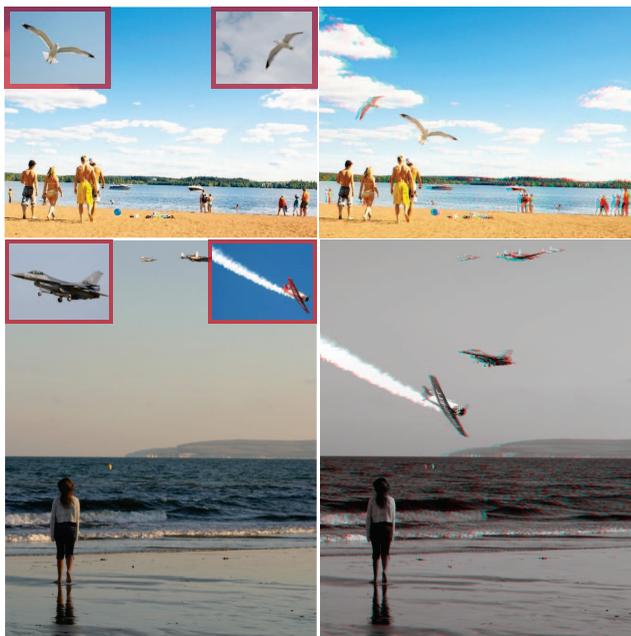


Fig. 10. Interactive composition results with user-specified disparities. Left: left image of the 3D background (source 2D images are shown on the top side). Right: *red&cyan* anaglyph results.

well fit the color style of the air, showing a convincing stereopsis in the composition.

5.2 Applications

By intuitive human interactions, our method can efficiently blend 2D foregrounds into stereoscopic backgrounds. This is also a basic and important operation for several tasks in stereoscopic image editing, which will be described as follows:

Stereoscopic image editing. To edit stereoscopic images, users usually need to specify regions of interest from image pairs. For example in Fig. 12, we first need to select the boats from the image pair, and then perform copy-and-paste. In general, it is labor-intensive and impractical to consistently select regions from left and right images, respectively. Thus, we propose *consistent selection* to allow users to select an interested region in the left image, and the selection is propagated to the right image automatically. This consistent propagation is proceeded in the following two steps:

1. *Triangulation for the left image.* The left triangular mesh is constructed by the constrained *Delaunay* triangulation [26] with the following constraints: a) Sampled points on the selected boundary. b) The feature points. We extract sparse but robust *SIFT* features [22] to establish reliable correspondences between the left and right images. The matched features are used as important constraints for consistent mesh propagation. In general, *SIFT* features are much denser in some parts with distinguished features, which is not helpful for triangulation and warping. We apply an adaptive non-maximal suppression [3] to filter these densely distributed features. c) Sampled points along the image border. The left part of Fig. 11 shows the triangular mesh of the left image.

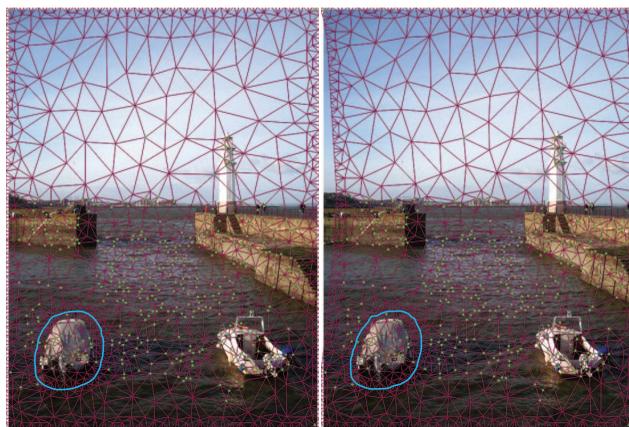


Fig. 11. Consistent selection. Left: the left triangular mesh generated by constrained *Delaunay* triangulation [26]. Right: the right triangular mesh which is propagated from the left mesh (the left and right closed boundaries in cyan show the consistent selection).

2. *Triangular mesh propagation.* We propagate the mesh on the left image to the right image with the following constraints: a) *Feature consistency.* The feature vertices on the right mesh should closely approximate the corresponding *SIFT* features in the right image. b) *Vertical alignment.* Assuming that our stereoscopic images are visual-comfortable, the vertical disparities of the corresponding vertices in the left and right meshes should approach zero to avoid visual discomfort [4], [18]. c) *As similar as possible.* There should be less distortion between the left and right meshes.

Inspired by Zhang et al. [38], the shape distortion is defined by a quadratic total conformal energy [20]. The linear combination of the constraints above can be formulated as a quadratic energy, and the vertices of the right mesh can be efficiently calculated by solving a linear system. After propagation, the object on the right image can be selected in a consistent manner (see right part of Fig. 11). With the *consistent selection*, our composition method can be extended to some stereoscopic image editing tasks. Fig. 12 shows an example of stereoscopic copy-and-paste. After selecting the boat pairs by our *consistent selection*, the boats can be freely moved and blended with the stereo background using our method in Section 4. Results in Fig. 12 show that the scale and disparity of the boats can well fit the target scene when moving them to different locations.



Fig. 12. Stereoscopic image editing. Left: left image of the source/result (top/bottom) stereoscopic image. Middle to Right: *red&cyan* anaglyph of source and result images.



Fig. 13. Stereoscopic painting. Top: user paintings on the left image. Bottom: corresponding red&cyan anaglyph painting.

Stereoscopic painting. Different from the painting on 2D images, it is nontrivial to paint figures on the stereoscopic image. First, we need to specify canvases in the stereoscopic

image pair using the *consistent selection*, and our painting is restricted in the canvas area. Then, users can paint figures on the canvas of the left image, and the stereoscopic painting is produced by our stereoscopic composition, according to the disparities in the canvas region. In this process, users can view the stereoscopic painting instantly as they paint 2D figures. See Fig. 13, we can write 2D “hi, TVCG” on the left image by painting characters with different colors, while viewing the stereoscopic results simultaneously.

Sketch2Stereo. To compose images, we always need to find suitable candidates from the Internet by text-based search, which is time consuming and labor intensive. *Sketch2Photo* [5] provided a tool to retrieve desirable images from the Internet by providing simple sketches and text labels, however they can only compose 2D images. We present *Sketch2Stereo* to produce realistic stereoscopic images by providing sketches and the corresponding text labels. We first obtain 2D foregrounds using the tool in *Sketch2Photo* [5]. To retrieve stereoscopic backgrounds, we provide keywords, which contain “stereo pair” or “stereo parallel.” Then, each object can be easily added to the stereoscopic scene using our stereoscopic composition. See Fig. 14, we compose three interesting stereoscopic images (photograph, playing frisbee, and football), which can provide more interesting and exciting visual experiences than *Sketch2Photo* does.

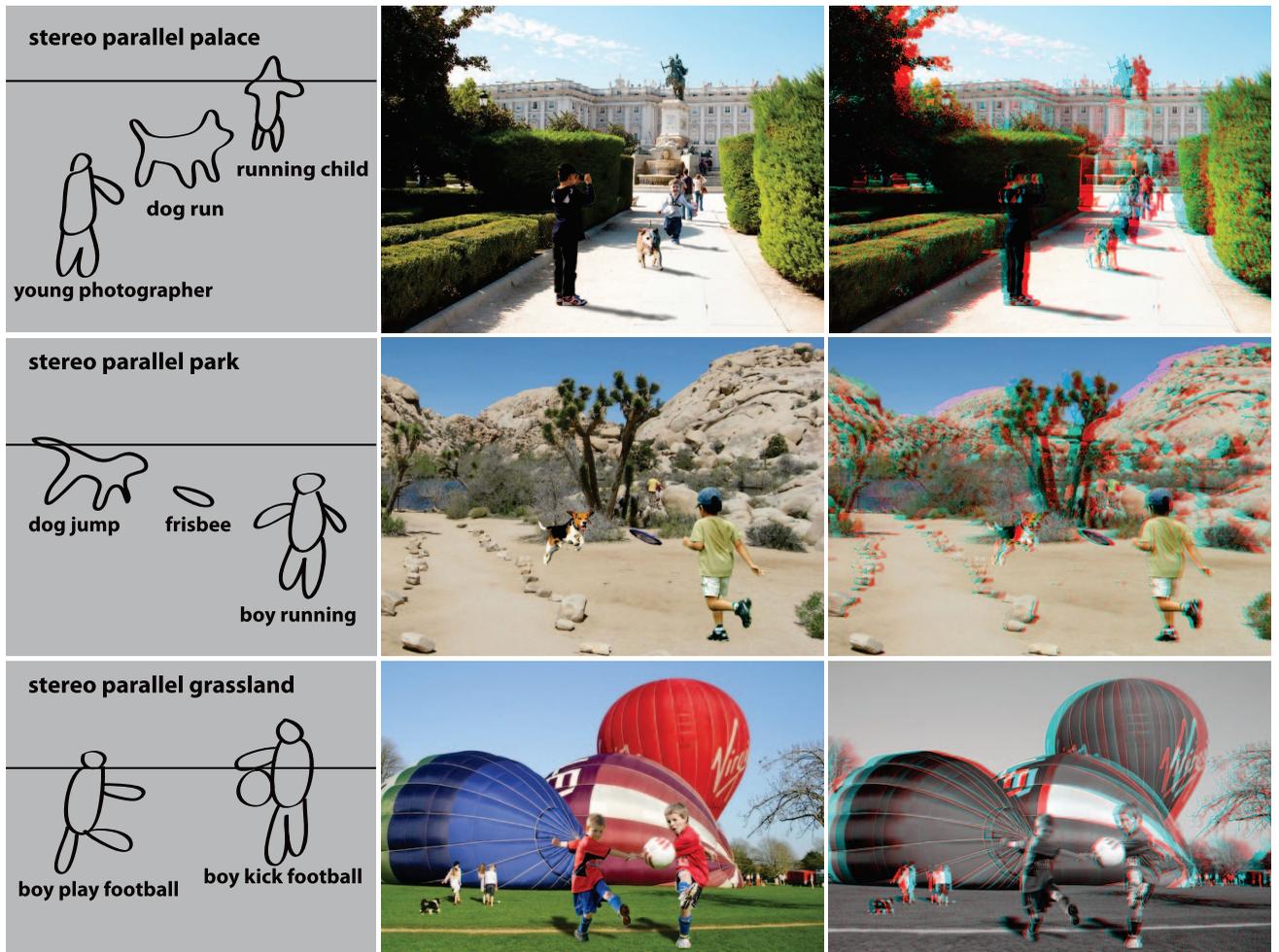


Fig. 14. Results of Sketch2Stereo. Left to right: user drawn sketches, left images, and the red&cyan anaglyph of our Sketch2Stereo results.

TABLE 1
Performance of Examples in This Paper

Examples	time (s)	Examples	time (s)
Fig. 1 - Boy play	62s	Fig. 9 - Snow car	40s
Fig. 9 - Street	61s	Fig. 9 - Sailboat	45s
Fig. 10 - Birds/Planes	59s	Fig. 12 - Boats	58s
Fig. 13 - <i>hi,TVCG</i>	56s	Fig. 14 - Photograph	61s
Fig. 14 - Frisbee	58s	Fig. 14 - Football	62s

The left and right two columns show the examples and total time to produce user-satisfied results.

5.3 Performance

We report the performance of our method on an Intel Core 2 Duo E8400 3-GHz computer with 4G RAM for the examples in this paper. Using our interactive system, even non-professional users can easily compose visual-comfortable and exciting 3D composition results. Take Fig. 1 for example (the source 2D image is 300*500). As we paint one stroke on the 2D foreground, the disparity diffusion costs 200-300 ms, and the disparity blending costs only 5 ms when the MVC are pre-computed. The warping costs 100 ms for the first time (initialize and decompose Matrix A , see the optimization in Section 4.3) while only 25 ms after that. Such efficiency allows users to interactively edit stereoscopic images while viewing the results instantly. We ask several college students with no experiences in image processing for user study. They learn to use our system for several minutes before the experiments. Table 1 shows the average time used by 3-5 students to produce visual-comfortable and convincing results using our system. For each example, it takes around 60 s to produce satisfying results (most time is spent on disparity editing of 2D foregrounds), which shows that our method could be applicable to practical use. We further demonstrate the advantages of our method in the accompanying video demo, which vividly shows our interactive composition and applications.

5.4 Limitations

Similar to some image editing methods, our approach fails in some cases. 1) When there is large perspective differences or occlusion between foregrounds and the objects of 3D background, our method does not work well. See Fig. 15, the head of the airplane should be occluded by the mountain because of its depth information. However, when the disparities of the 3D background are not correctly recovered, it always fails to deal with the occlusion, which makes the composition result unnatural and visual-uncomfortable. 2) Our method usually depends on the quality of the target disparity map, however, existing methods are still not robust enough to calculate reliable and precise disparity map. 3) The disparity map produced by strokes is relatively rough, so when the foreground objects are very close to the camera, our method can not produce precise disparity map to reflect the detailed geometry structure.

6 CONCLUSION

In this paper, we have proposed an efficient approach for depth-consistent stereoscopic composition, in which a 2D



Fig. 15. Failure case of our method. Left: source 2D foreground and 3D background. Right: the *red&cyan* anaglyph result. The incorrect occlusion between the airplane and the mountain makes the composition result unnatural.

foreground is interactively blended into a 3D background. As users iteratively paint “disparity” on the 2D foreground, its disparity map is gradually produced and then blended with that of the 3D background to keep depth consistent. Then the 2D foreground is warped to generate an image pair by an energy minimization constrained by the expected disparities and perspective scaling, which could be solved in real time. The final results are obtained by blending the warped image pair into the background image pair. Experiments show that our method can largely reduce users’ labor and skill in stereoscopic composition, and even nonprofessional users can produce results with exciting 3D viewing. We have also presented a novel system for interactive stereoscopic composition, in which users can freely edit the disparity maps of 2D foregrounds, while receiving the immediate feedback of stereoscopic composition in real time.

In the future, we can improve our method in the following aspects: 1) We will try to solve the problem of occlusion and large perspective differences between 2D foregrounds and 3D backgrounds, which is very common in image editing. In addition, more factors such as geometric structure, image saliency [6] can be taken into account to further improve our stereoscopic composition. 2) Our method can be further extended to other important operations in stereoscopic image editing, such as resizing, completion, and retargeting. 3) We will try to extend our method to stereoscopic videos, and there will be more challenges such as the temporal coherence in disparities, color and warping.

ACKNOWLEDGMENTS

The authors would like to thank all anonymous reviewers for their valuable comments. This work was supported by the National Basic Research Program (No. 2011CB302205) of China, National High-Tech Research and Development Program (No. 2013AA013903) of China, and the National Natural Science Foundation (No. 61170141) of China.

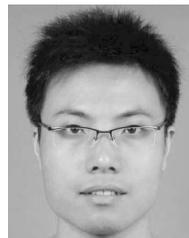
REFERENCES

- [1] X. An and F. Pellacini, “AppProp: All-Pairs Appearance-Space Edit Propagation,” *ACM Trans. Graphics*, vol. 27, no. 3, pp. 40:1-40:9, 2008.
- [2] C. Barnes, E. Shechtman, A. Finkelstein, and D.B. Goldman, “Patchmatch: A Randomized Correspondence Algorithm for Structural Image Editing,” *ACM Trans. Graphics*, vol. 28, no. 3, pp. 24:1-24:11, 2009.

- [3] M. Brown, R. Szeliski, and S.A.J. Winder, "Multi-Image Matching Using Multi-Scale Oriented Patches," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 510-517, 2005.
- [4] C.-H. Chang, C.-K. Liang, and Y.-Y. Chuang, "Content-Aware Display Adaptation and Interactive Editing for Stereoscopic Images," *IEEE Trans. Multimedia*, vol. 13, no. 4, pp. 589-601, Aug. 2011.
- [5] T. Chen, M.-M. Cheng, P. Tan, A. Shamir, and S.-M. Hu, "Sketch2photo: Internet Image Montage," *ACM Trans. Graphics*, vol. 28, no. 5, pp. 124:1-124:10, 2009.
- [6] M.-M. Cheng, G.-X. Zhang, N.J. Mitra, X. Huang, and S.-M. Hu, "Global Contrast Based Salient Region Detection," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 409-416, 2011.
- [7] Y.-Y. Chuang, B. Curless, D. Salesin, and R. Szeliski, "A Bayesian Approach to Digital Matting," *Proc. Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 264-271, 2001.
- [8] T.A. Davis, "Algorithm 849: A Concise Sparse Cholesky Factorization Package," *ACM Trans. Math. Software*, vol. 31, no. 4, pp. 587-591, Dec. 2005.
- [9] P. Diddy, T. Ritschel, E. Eisemann, K. Myszkowski, and H.-P. Seidel, "A Perceptual Model for Disparity," *ACM Trans. Graphics*, vol. 30, no. 4, p. 96, 2011.
- [10] M. Ding and R. Tong, "Content-Aware Copying and Pasting in Images," *The Visual Computer*, vol. 26, nos. 6-8, pp. 721-729, 2010.
- [11] S.-P. Du, S.-M. Hu, and R.R. Martin, "Manipulating Perspective in Stereoscopic Images," Technical Report TR-120516, Tsinghua Univ., 2012.
- [12] Z. Farbman, G. Hoffer, Y. Lipman, D. Cohen-Or, and D. Lischinski, "Coordinates for Instant Image Cloning," *ACM Trans. Graphics*, vol. 28, no. 3, pp. 67:1-67:9, 2009.
- [13] E.S.L. Gastal and M.M. Oliveira, "Shared Sampling for Real-Time Alpha Matting," *Computer Graphics Forum*, vol. 29, no. 2, pp. 575-584, 2010.
- [14] K. He, C. Rhemann, C. Rother, X. Tang, and J. Sun, "A Global Sampling Method for Alpha Matting," *Proc. Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 2049-2056, 2011.
- [15] J. Jia, J. Sun, C.-K. Tang, and H.-Y. Shum, "Drag-and-Drop Pasting," *ACM Trans. Graphics*, vol. 25, no. 3, pp. 631-637, 2006.
- [16] Y. Kim, H. Jung, S. Choi, J. Lee, and J. Noh, "A Single Image Representation Model for Efficient Stereoscopic Image Creation," *Computer Graphics Forum*, vol. 30, no. 7, pp. 2067-2076, 2011.
- [17] J.-F. Lalonde, D. Hoiem, A.A. Efros, C. Rother, J.M. Winn, and A. Criminisi, "Photo Clip Art," *ACM Trans. Graphics*, vol. 26, no. 3, p. 3, 2007.
- [18] M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic, and M.H. Gross, "Nonlinear Disparity Mapping for Stereoscopic three-dimensional," *ACM Trans. Graphics*, vol. 29, no. 4, pp. 75:1-75:10, 2010.
- [19] A. Levin, D. Lischinski, and Y. Weiss, "A Closed-Form Solution to Natural Image Matting," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 228-242, Feb. 2008.
- [20] B. Lévy, S. Petitjean, N. Ray, and J. Maillot, "Least Squares Conformal Maps for Automatic Texture Atlas Generation," *ACM Trans. Graphics*, vol. 21, no. 3, pp. 362-371, 2002.
- [21] W.-Y. Lo, J. van Baar, C. Knaus, M. Zwicker, and M.H. Gross, "Stereoscopic Three-Dimensional Copy & Paste," *ACM Trans. Graphics*, vol. 29, no. 6, pp. 147:1-147:10, 2010.
- [22] D.G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int'l J. Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.
- [23] T. Oskam, A. Hornung, H. Bowles, K. Mitchell, and M.H. Gross, "OSCAM-Optimized Stereoscopic Camera Control for Interactive Three-Dimensional," *ACM Trans. Graphics*, vol. 30, no. 6, pp. 189:1-189:8, 2011.
- [24] P. Pérez, M. Gangnet, and A. Blake, "Poisson Image Editing," *ACM Trans. Graphics*, vol. 22, no. 3, pp. 313-318, 2003.
- [25] B. Pollock, M. Burton, J.W. Kelly, S. Gilbert, and E. Winer, "The Right View from the Wrong Location: Depth Perception in Stereoscopic Multi-User Virtual Environments," *IEEE Trans. Visualisation and Computer Graphics*, vol. 18, no. 4, pp. 581-588, Apr. 2012.
- [26] J.R. Shewchuk, "Delaunay Refinement Algorithms for Triangular Mesh Generation," *Computational Geometry*, vol. 22, nos. 1-3, pp. 21-74, 2002.
- [27] B.M. Smith, L. Zhang, and H. Jin, "Stereo Matching with Nonparametric Smoothness Priors in Feature Space," *Proc. IEEE Computer Vision and Pattern Recognition (CVPR '09)*, pp. 485-492, 2009.
- [28] M. Wan, N. Zhang, H. Qu, and A.E. Kaufman, "Interactive Stereoscopic Rendering of Volumetric Environments," *IEEE Trans. Visualization and Computer Graphics*, vol. 10, no. 1, pp. 15-28, Jan./Feb. 2004.
- [29] J. Wang, M. Agrawala, and M.F. Cohen, "Soft Scissors: An Interactive Tool for Realtime High Quality Matting," *ACM Trans. Graphics*, vol. 26, no. 3, p. 9, 2007.
- [30] J. Wang and M.F. Cohen, "Image and Video Matting: A Survey," *Foundations and Trends in Computer Graphics and Vision*, vol. 3, no. 2, pp. 97-175, 2007.
- [31] L. Wang, H. Jin, R. Yang, and M. Gong, "Stereoscopic Inpainting: Joint Color and Depth Completion from Stereo Images," *Proc. IEEE Computer Vision and Pattern Recognition (CVPR '08)*, 2008.
- [32] O. Wang, M. Lang, M. Frei, A. Hornung, A. Smolic, and M.H. Gross, "Stereobrush: Interactive Two-Dimensional to Three-Dimensional Conversion Using Discontinuous Warps," *Proc. Int'l Symp. Sketch-Based Interfaces and Modeling (SBIM '11)*, pp. 47-54, 2011.
- [33] Y.-S. Wang, C.-L. Tai, O. Sorkine, and T.-Y. Lee, "Optimized Scale-and-Stretch for Image Resizing," *ACM Trans. Graphics*, vol. 27, no. 5, p. 118, 2008.
- [34] B. Ward, S.B. Kang, and E.P. Bennett, "Depth Director: A System for Adding Depth to Movies," *IEEE Computer Graphics and Applications*, vol. 31, no. 1, pp. 36-48, Jan./Feb. 2011.
- [35] A. Woods, T. Docherty, and R. Koch, "Image Distortions in Stereoscopic Video Systems," *Stereoscopic Displays And Applications*, vol. 1915, pp. 36-48, 1993.
- [36] C. Xiao, Y. Nie, and F. Tang, "Efficient Edit Propagation Using Hierarchical Data Structure," *IEEE Trans. Visualization and Computer Graphics*, vol. 17, no. 8, pp. 1135-1147, Aug. 2011.
- [37] K. Xu, Y. Li, T. Ju, S.-M. Hu, and T.-Q. Liu, "Efficient Affinity-Based Edit Propagation Using k-d Tree," *ACM Trans. Graphics*, vol. 28, no. 5, pp. 118:1-118:6, 2009.
- [38] G.-X. Zhang, M.-M. Cheng, S.-M. Hu, and R.R. Martin, "A Shape-Preserving Approach to Image Resizing," *Computer Graphics Forum*, vol. 28, no. 7, pp. 1897-1906, 2009.
- [39] G. Zhang, W. Hua, X. Qin, T.-T. Wong, and H. Bao, "Stereoscopic Video Synthesis from a Monocular Video," *IEEE Trans. Visualisation and Computer Graphics*, vol. 13, no. 4, pp. 686-696, July/Aug. 2007.
- [40] Y. Zhang and R. Tong, "Environment-Sensitive Cloning in Images," *The Visual Computer*, vol. 27, nos. 6-8, pp. 739-748, 2011.



Ruo-Feng Tong received the BS degree from Fudan University, China, in 1991, and the PhD degree from Zhejiang University, China, in 1996. He is a professor in the Department of Computer Science, Zhejiang University, China. His research interests include image and video processing, computer graphics, and computer animation.



Yun Zhang received the BS and MS degrees from Hangzhou Dianzi University, China, in 2006 and 2009, respectively. He is currently working toward the PhD degree in the Department of Computer Science, Zhejiang University, China. His research interests include image and video processing and computer graphics.



Ke-Li Cheng received the BS and MS degrees from Chongqing University, China, in 2007 and 2010, respectively. He is currently working toward the PhD degree in the Department of Computer Science, Zhejiang University, China. His research interests include image and video processing and computer graphics.