Changing Perspective in Stereoscopic Images

Song-Pei Du, Shi-Min Hu, *Member, IEEE,* and Ralph R. Martin

Abstract—Traditional image editing techniques cannot be directly used to edit stereoscopic ('3D') media, as extra constraints are needed to ensure consistent changes are made to both left and right images. Here, we consider manipulating perspective in stereoscopic pairs. A straightforward approach based on depth recovery is unsatisfactory: instead, we use feature correspondences between stereoscopic image pairs. Given a new, user-specified perspective, we determine correspondence constraints under this perspective, and optimize a 2D warp for each image which preserves straight lines and guarantees proper stereopsis relative to the new camera. Experiments verify that our method generates new stereoscopic views which correspond well to expected projections, for a wide range of specified perspective. Various advanced camera effects, such as dolly zoom and wide angle effects, can also be readily generated for stereoscopic image pairs using our method.

Index Terms—Stereoscopic images, stereopsis, perspective, warping.

1 Introduction

The renaissance of '3D' movies has led to the development of both hardware and software stereoscopic techniques for both professionals and consumers. Availability of 3D content has dramatically increased, with 3D movies shown in cinemas, launches of 3D TV channels, and 3D games on PCs. Consumers can now take 3D photos or video as easily as traditional 2D images, and view them on 3D TVs. While acquisition may be simple [1], [2], editing of stereoscopic media is challenging. Tasks such as editing the disparity [3], changing content [4], or just simply rotating a stereo picture to a new orientation all require non-trivial computations. Existing image editing techniques cannot be simply extended to stero (as we shortly explain) and professional stereoscopic 3D production is costly.

Photographic stereoscopic 3D images comprise an image pair taken using a binocular device. Humans use the *disparity* in such stereoscopic images to determine depth, via the principle of *stereopsis*. The two images in a stereoscopic pair are not arbitrary images, but contain *consistent* image content related by the disparity, which depends on depth. Editing the images independently using human best efforts is likely to break the stereoscopic consistency; even minor discrepancies lead to a poor stereoscopic viewing experience or misleading depth information. For example, an anaglyph image cannot be rotated directly, as this would introduce a vertical disparity, giving an unnatural stereoscopic experience [5]. Several methods have already been proposed for processing stereoscopic images, such as stereoscopic inpainting [6], disparity mapping [3], [7], and 3D copy & paste [4]. However, relatively few tools are available compared to the number of 2D image tools.

Perspective plays an essential role in the appearance of images. In 2D, a simple approach to changing the perspective of a given image is to approximate the entire image as a single planar region and transform it under a homography. This fundamental image processing tool enables applications like image stitching [8]. More sophisticated effects can be provided with the aid of user interaction, and the combination of images using multiple viewpoints. For example, [9] creates images from an artistic perspective based on user-specified line constraints and vanishing points, while [10] creates multi-viewpoint panoramas from several photographs of a scene. However, existing perspective manipulation techniques for 2D images cannot be applied independently to each image in a stereo pair, as additional constraints are needed. A comparable stereo method should take into account two considerations: (i) a stereo image pair contains depth information, and this can be used to help determine the results under new perspective, and (ii) proper stereoscopic consistency should be maintained to ensure a plausible and comfortable 3D viewing experience.

Given two images captured by a binocular device, a direct approach to stereo perspective manipulation would be to compute depth at each pixel from the stereo image pair, reconstruct the scene geometry, and then compute the projection under the new camera configuration. However, this solution would not be perfect, as occlusion would lead to gaps in the output. Furthermore, while the human visual system can readily understand 3D geometry from a stereo image pair, it is much more difficult to algorithmically reconstruct scene geometry accurately, even with the help of user interaction, and a more robust solution

- *S.P. Du and S.M. Hu are with TNList, Department of Computer Science and Technology, Tsinghua University, Beijing, China. E-mail: dusongpei@gmail.com and shimin@tsinghua.edu.cn.*
- *R.R. Martin is with the School of Computer Science & Informatics, Cardiff University, 5 The Parade, Roath, Cardiff, CF24 3AA, UK. E-mail: Ralph.Martin@cs.cardiff.ac.uk*

Fig. 1. Manipulating perspective. Top: one image from a stereoscopic image pair and corresponding monochrome anaglyph. Middle: new view produced by altering the camera viewpoint to increase perceived depth. Bottom: addition of radial camera distortion.

is needed. [3] gives an alternative solution for editing stereoscopic images and videos by consistently warping the pair of images. [11] introduces a 2D to 3D conversion method based on warping, and synthesizing one stereo view from the other, but does not consider perspective change.

Motivated by the above observations, we give a technique for performing *stereoscopic image perspective manipulation*, in which the user specifies stereoscopic camera parameters for the new view. Our key technical contributions are as follows. Firstly, instead of using homographies, we compute a 2D warp for each image using an optimization framework. Instead of dense point correspondences, which are difficult to obtain, we use selected reliably matched feature points. Secondly, the human visual system is sensitive to straight lines, so we are careful to detect straight lines in the stereo image pair, and retain them. Warping is controlled by these robust point and line segment correspondences, avoiding the inaccuracies arising from a dense depth map. Constraints are used to ensure proper stereopsis, as well as to preserve the straight lines. The problem is posed as a quadratic energy minimization problem (as in [3]), and can be efficiently solved via a sparse linear system.

Manipulating perspective of stereoscopic images is a useful tool in its own right, but also has applications as a preprocessing stage for other tasks such as image blending and composition. We later demonstrate the versatility of our approach in applications such as image stitching and advanced camera effects.

2 Related Work

We next summarize prior work on stereoscopic image processing, and warping techniques.

2.1 Stereoscopic Image Processing

The human visual system perceives depth from stereo and other cues; quality of stereoscopic viewing experience is affected by several factors. Numerous works have analyzed shear distortion, keystoning, and lens distortion in stereoscopy, and their effect on visual realism and comfort [5], [12], [13]. A user study [5] showed that large vertical disparity leads to eye strain, suggesting it should be minimised for viewer comfort. Editing the disparity of stereoscopic images is an important tool in stereoscopic content postproduction, and various rules have been devised for use in 3D movie making [14]. Methods have been developed for analyzing disparity as well as for reducing visual distortion to create a more pleasing 3D effect [15], [16]. [17] proposed a general framework to adapt disparities of stereo images and video based on dense depth maps. [3] introduced more general nonlinear disparity mapping operators, and described an effective technique for stereoscopic disparity editing based on image warping. Complementary to the above postproduction editing techniques, [1] gave a method for adjusting stereoscopic live-action video, controlling camera convergence and interaxial separation during real-time stereoscopic rendering, to ensure visual comfort. None of these editing techniques, however, aim to provide a new view of the scene, but rather aim to make the existing image or video more comfortable to watch, by carefully controlling the horizontal disparity.

As well as disparity editing, several 2D image editing operations have been extended to stereoscopic content, such as image inpainting [6], and image copy-and-paste [4]. To change the orientation of copied objects, [4] uses a plane proxy to compute a pair of consistent homographies applied to each image. This works well for scene elements which are almost planar, but hardly applies to whole scenes with varying depths: in such cases it leads to vertical disparity. [18] showed how to resize stereoscopic images to a target resolution in a content-aware manner, adapting depth to the comfort zone of the display. [7] proposed a *viewer-centric* editor for stereoscopic video allowing the user to edit stereo parameters, such as interocular distance, field of view, and viewer location. In contrast, our method is *scene-centric*, and considers the camera setup. [19] surveys techniques for 3D video processing; no current method can manipulate the perspective of a stereoscopic image pair while controlling the disparity for viewer comfort. We solve this

problem by computing a pair of warping functions which approximate the projection according to the new perspective, as well as ensuring that the disparity is horizontal in the output transformed stereo images.

2.2 Image and Video Warping

Our framework simultaneously warps both images in a stereoscopic pair, using energy constraints to provide the desired perspective and stereopsis. We build on earlier warping methods used to solve other problems in image and video processing, e.g. image resizing [20], video resizing [21], [22], [23], correction of wide-angle lens distortion [24], [25], production of projections for artistic perspective [9], and 3D video stabilization [23]. [23] proposes a content-preserving warp based on sparse correspondences in monocular image sequences. While sharing a similar feature-point-driven warping framework, we solve a different problem and use specifically designed energy terms. Warping techniques have also been used to manipulate disparities of stereoscopic images [3]. [11] uses a discontinuous warping technique for 2D to 3D conversion, computing a projection from a new viewpoint. Their main goal is synthesis of a proper stereo image pair from an input 2D image and user input sketches, while our input is a stereo image pair, and we handle the problem of changing the stereo perspective in a general sense. [26] introduces StereoPasting for interactively compositing multiple stereo images. Active appearance models [27] match statistical models of appearance to images, again based on warping, but using a simple linear model. Our warping method has different goals: preserving proper stereopsis and line segments.

Our method shares a similar framework with previous work on image and video warping: each image warp is described by a quad mesh, and energy minimization is used to determine the warp which best meets a set of conflicting requirements. Constraints are used to compute a warped image which attempts to match a desired projection of automatically detected feature points and lines, to produce the desired stereopsis. Unlike [3] which focused on using disparity mapping operators to provide a comfortable viewing experience, we change the image projection as well as the disparity.

3 Stereoscopic Correspondence

The human visual system can fuse a stereoscopic image pair to generate a 3D viewing experience. This is hard to emulate algorithmically; it is difficult to compute a dense, accurate disparity map from a stereoscopic image pair [28], [29]. Instead, we employ *sparse* correspondences between the images in a stereo pair (I_l, I_r), as these can be detected more robustly. We use two kinds of correspondences: feature points and straight line segments. These can generally be determined with relatively high confidence in a stereo pair. We note that straight lines have high visual significance, and thus should be preserved as such during image editing. We assume that the intrinsic parameters of the stereo camera pair used to capture the input are known.

3.1 Feature Point Correspondences

We first establish sparse point correspondences between the images in the source stereoscopic pair. Finding sparse point correspondences in stereo is discussed in [3], [11]. We use a simple, similar technique: SIFT provides initial feature correspondences, and we follow [30] to remove outliers. An adaptive non-maximal suppression procedure [31] is used to improve the spatial distribution of feature points; we set the minimum suppression radius equal to half of the quad mesh length (see Section 4).

This set of sparse feature points with robust correspondences is then used to compute for later use separate approximate disparity maps D_l and D_r for each image in the stereoscopic pair. [11] computed a dense disparity map based on propagation of sparse scribbles. We just employ a direct interpolation scheme which works well, using a set of Gaussian radial basis functions centered at each feature point. The approximate disparity value $D_l(\mathbf{x})$ at any point $\mathbf{x} \in I_l$ is computed as

$$D_l(\mathbf{x}) = \sum_i \omega_i e^{-\sigma^2|\mathbf{x}-\mathbf{x}_i|^2},$$

where $\{\mathbf{x}_i\}$ is the set of all *feature points* in I_l. At the feature points, $\mathbf{x}_i$, $D_l(\mathbf{x}_i)$ is set to the exact disparity value d_i computed from the feature correspondence. The weights ω_i are determined by solving the following quadratic energy minimization problem using linear least squares,

$$E = \sum_i \|D_l(\mathbf{x}_i) - d_i\|^2 = \sum_i \| \sum_j \omega_j e^{-\sigma^2|\mathbf{x}_i-\mathbf{x}_j|^2} - d_i\|^2$$

The variance σ is set to the inverse of the length of the image diagonal. D_r is computed similarly.

3.2 Straight Line Correspondences

Human perception is sensitive to straight lines. Finding straight line correspondences allows us explicitly preserve them under change of perspective. Instead of finding straight line segments in the two images separately and matching them (as in [32]), we combine the problems of line extraction and line correspondence in a single voting procedure with vote pruning.

In Hough transform space, an infinite line is represented by (ρ, θ), where $\rho \geq 0$ and $\theta \in [0, 2\pi)$. A pair of corresponding lines $(\theta_l, \rho_l, \theta_r, \rho_r)$ can be determined by 4 points, 2 in each image. We employ a voting procedure to find potential line correspondences in a

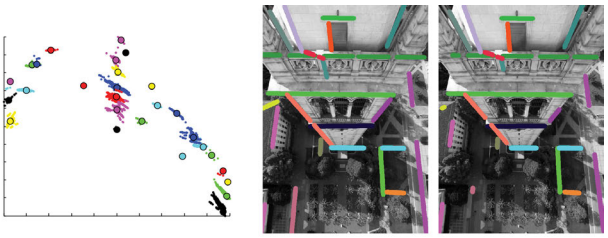

Fig. 2. Left: Votes (projected into $\mathbb{R}^2$), clustered using mean-shift, shown in different colors. Middle, right: Corresponding line segments, indicated by color.

stereoscopic image pair. In practice, to reduce false matches, we choose 3 points per line instead of 2.

In detail, gradient methods [33] allow us to find edge sets E_l and E_r in each image. Each edge $e_l \in E_l$ is sampled in equal-sized steps. We choose all approximately collinear triples of sample points (v_{0l}, v_{1l}, v_{2l}), by measuring the total distance from the three points to their best matching line. For each of v_{0l}, v_{1l}, v_{2l}, we find its best match with some edge point in E_r with the same horizontal height; the horizontal search range can be limited using the approximate disparity D_l. If the best matching triple (v_{0r}, v_{1r}, v_{2r}) is also approximately collinear, (v_{0l}, v_{1l}, v_{2l}), (v_{0r}, v_{1r}, v_{2r}) provide a vote for $(\theta_l, \rho_l, \theta_r, \rho_r)$ in 4-dimensional parameter space. Figure 2 illustrates such a vote distribution, projected into $\mathbb{R}^2$ by (θ_l, ρ_l). We use mean-shift clustering to find cluster centers in this parameter space, following [34]. To improve robustness, we reject any vote for which the angle $|\theta_l - \theta_r| > 20°$, as a probable mismatch. We only seek point pairs $(\mathbf{l}_0, \mathbf{l}_1)$ with a distance between them in the range: $d_{\min} < \|\mathbf{l}_0 - \mathbf{l}_1\| < d_{\max}$ for $d_{\min} = 10$ and $d_{\max} = 100$.

Suppose $(\theta_l, \rho_l, \theta_r, \rho_r)$ is the center of one cluster in parameter space. A straight line is fitted to that cluster using least-squares fitting; its line segments are identified by the votes. Finding corresponding parts of related non-horizontal line segments is trivial: we keep those parts of each segment which agree in vertical height. Alternatively, we may keep the entire line segments in both views, as the mismatch may be due to occlusion or imperfect edge detection or clustering. This approach fails for horizontal lines, as each point on the line no longer has a unique height. To resolve this issue, we assume that depth changes linearly along the line segment. To match two horizontal line segments $s_l \in I_l$ and $s_r \in I_r$, for each pixel in s_l, its best match in s_r is computed and we fit a linear function to depths along s_l using least squares. These new depths give the final correspondence between s_l and s_r. In practice, we use this method for all line segments whose angle with the horizontal axis $< 20°$.

4 Perspective Manipulation

When editing a stereoscopic image pair, depth-based methods as in [4], [6] can produce results with physically accurate depth information. However, in practice it is challenging to reconstruct a dense depth map from a stereoscopic image pair [28]. Instead, we use warping to deform the stereoscopic image pair [3], building on 2D image warping techniques [22]. Such methods allow the use of *sparse* correspondences across the stereoscopic image pair; these can be more readily and reliably found than a dense depth map. Although the results may introduce distortion, unlike depth-based methods, the results are typically sufficiently visually plausible [19]. To perform warping, each image in the input stereoscopic pair is covered by a quad mesh whose nodes are manipulated to edit the image content.

Staring from a stereoscopic image pair (I_l, I_r), the user first specifies the new perspective in terms of stereoscopic camera parameters: lens parameters and distortions, stereo camera separation and convergence angle, and position and orientation of the stereo camera. We wish to find two warping functions $\mathbf{w}_l$ and $\mathbf{w}_r$ which map the input image pair into an output pair corresponding to the new perspective view determined by these parameters. Each warp $\mathbf{w} = (w_x, w_y)$ maps input image coordinates $\mathbf{x} = (x, y)$ to output coordinates $(w_x(\mathbf{x}), w_y(\mathbf{x}))$. The warp is based on a discrete uniform grid, the overall warp being computed by bilinear interpolation of discrete warps at grid nodes. The new positions of the grid nodes are set to minimize a quadratic energy $E = E(\mathbf{G})$ where $\mathbf{G}$ is the set of all grid nodes. E takes into account perspective point and homography constraints for plausible stereopsis, and piecewise smoothness and straight line constraints for visual quality.

4.1 Perspective Point Constraints

The feature correspondences established in Section 3.1 provide robust geometric information about the 3D scene. When changing perspective, we use these constraints to determine a suitable warp. Corresponding feature points in the input stereoscopic pair come from P_l and P_r. The 3D position of the underlying 3D point can be computed using the given camera parameters. It may then be projected using the target cameras in accordance with the new perspective to find its new position. Doing so for all such correspondences gives $\tilde{P}_l$ and $\tilde{P}_r$. An energy term E_p is used whose effect is to help ensure these feature points appear in the positions required by the new projection:

$$E_p = \sum_{(\mathbf{x}_i, \tilde{\mathbf{x}}_i) \in (P_l, \tilde{P}_l)} \|\mathbf{w}_l(\mathbf{x}_i) - \tilde{\mathbf{x}}_i\|^2 + \sum_{(\mathbf{x}_i, \tilde{\mathbf{x}}_i) \in (P_r, \tilde{P}_r)} \|\mathbf{w}_r(\mathbf{x}_i) - \tilde{\mathbf{x}}_i\|^2$$

where $\tilde{\mathbf{x}}_i$ is the location of $\mathbf{x}_i$ under the new perspective. As we have reliable disparity and depth value at feature points, we force the warpi to place

each feature point $\mathbf{x}_i$ at the location $\tilde{\mathbf{x}}_i$ ([3] constrains disparity). This constraint at feature points is equivalent to the depth-image-based rendering and point constraints for a single image in [23], which helps to guarantee proper transformation for the new perspective. Feature points are only accidentally at grid nodes; to convert the constraints to grid nodes, we use bilinear interpolation based on the four grid nodes surrounding each feature point. We achieve camera convergence by image-shift instead of toe-in, as it produces less distortion. The image shift is computed from the convergence angle as in [5]. This model leads to a new perspective with horizontal disparity, which our constraints attempt to enforce at feature points. Due to the piecewise smoothness of the warp, the disparity of the output stereo image is almost horizontal everywhere.

4.2 Homography Constraints

Homography constraints are used to constrain the local behaviour of the warping function; the perspective constraint energy E_p only controls the projection at feature points. At other points in the left image, the *local* neighborhood of $\mathbf{x}$ in I_l can be regarded as the projection of a 3D planar proxy (a first order approximation of depth at the corresponding 3D point of $\mathbf{x}$), so the warping function $\mathbf{w}_l$ should locally be a homography $H_l(\mathbf{x})$ at $\mathbf{x}$. (We consider the issue of depth steps at occlusions later.)

We only have an approximate disparity value $D_l(\mathbf{x})$ at $\mathbf{x}$ to determine $H_l(\mathbf{x})$. It would not be appropriate to constrain the warped position $\mathbf{w}_l(\mathbf{x})$ to appear exactly at the position predicted by $H_l(\mathbf{x})$ locally, as an inaccurate strong constraint might break the epipolar geometry. Instead we employ a first order constraint to guarantee a local homography in a weak sense. As $\mathbf{w}_l(\mathbf{x}+\triangle\mathbf{x}) \approx \mathbf{w}_l(\mathbf{x}) + J(\mathbf{w}_l(\mathbf{x}))\triangle\mathbf{x}$, we constrain the Jacobian of $\mathbf{w}_l$ to be equal to that of H_l. Thus, at feature points, we have accurate new projection positions, so the warped points are forced to these positions. At other points, the trend of the warping function should locally agree with the homography computed by the approximate depth map. Doing the same for both images gives an overall energy term:

$$\begin{aligned} E_h &= \sum_{\mathbf{x}} \|J(\mathbf{w}_l(\mathbf{x})) - J(H_l(\mathbf{x}))\|^2 \\ &+ \sum_{\mathbf{x}} \|J(\mathbf{w}_r(\mathbf{x})) - J(H_r(\mathbf{x}))\|^2. \end{aligned}$$

In practice, we use the discrete version of the Jacobian at grid nodes. For a grid node $\mathbf{x}_{i,j} = (u_{i,j}, v_{i,j}) \in I_l$, let the warped point be $\mathbf{w}_l(\mathbf{x}_{i,j}) = (u'_{i,j}, v'_{i,j})$. Then the Jacobian of the warping function is

$$J(\mathbf{w}_l(\mathbf{x}_{i,j})) = \begin{bmatrix} u'_{i+1,j} - u'_{i,j} & u'_{i,j+1} - u'_{i,j} \\ v'_{i+1,j} - v'_{i,j} & v'_{i,j+1} - v'_{i,j} \end{bmatrix}.$$

To compute the local Jacobian of $H_l(\mathbf{x}_{i,j})$, the local neighborhood of $\mathbf{x}_{i,j}$ is treated as a planar region with disparity $D_l(\mathbf{x}_{i,j})$ at $\mathbf{x}_{i,j}$, $D_l(\mathbf{x}_{i+1,j})$ at $\mathbf{x}_{i+1,j}$ and $D_l(\mathbf{x}_{i,j+1})$ at $\mathbf{x}_{i,j+1}$. We directly compute H_l for $\mathbf{x}_{i,j}$ and its two neighbors $\mathbf{x}_{i,j+1}$, $\mathbf{x}_{i+1,j}$, by calculating their 3D corresponding positions using D_l and the projections from the new perspective. Then $J(H_l(\mathbf{x}_{i,j}))$ can be computed in a similar way to $J(\mathbf{w}_l(\mathbf{x}_{i,j}))$.

At occluding edges, depth and disparity are discontinuous, and the warping function is not locally a homography. The sums in E_h should thus include all grid nodes *except* for cells containing occluding edges. Approximate depths can be estimated from the disparity maps D_l and D_r and occluding edges can be detected as in [11].

4.3 Piecewise Smooth Warping

Where depth and disparity are continuous, the warping function should be smooth, and in such cases we encourage the Hessian of the warping function to be zero, with the constraints $E_s = \sum \left\| \frac{\partial J}{\partial \mathbf{x}} \right\|^2$. The local Hessian matrix is approximated by finite differences on grid nodes [9].

4.4 Line Constraints

Humans are sensitive to straight lines, and to retain plausibility, straight lines should remain straight after reprojection and warping. Given the line segment correspondences established earlier, and the camera parameters, we can determine each line in 3D space, and hence its projections in the two new perspective images. The following energy term E_{ls} is used to constrain the orientation of the line segments:

$$E_{ls} = \sum_{s \in L_l \cup L_r} \int_s \|[\sin\tilde{\theta}_s, \cos\tilde{\theta}_s] \cdot \begin{bmatrix} \cos\theta(\mathbf{w}(s)) \\ -\sin\theta(\mathbf{w}(s)) \end{bmatrix} \|^2 ds,$$

where $\theta(\mathbf{w}(s))$ is the orientation of the warped line segment along $\mathbf{w}(s)$, $\mathbf{w}$ is $\mathbf{w}_l$ or $\mathbf{w}_r$ as appropriate to s, and $\tilde{\theta}_s$ is the target orientation of the segment s in the new perspective. To discretize this calculation, we parameterize $\mathbf{w}(s)$ and sample it in steps of the same size as the mesh grid length, $\theta(\mathbf{w}(s))$ at a sample point is calculated based on this point and one endpoint of $\mathbf{w}(s)$, and E_{ls} is computed by summing at these sample points.

$$E_{ls} = \sum_{s \in L_l \cup L_r} \sum_i \|[\sin\tilde{\theta}_s, \cos\tilde{\theta}_s] \cdot [\mathbf{w}(\mathbf{x}_i) - \mathbf{w}(\mathbf{x}_0)]^{\mathrm{T}}\|^2 ds.$$

4.5 Energy Optimization

To find the warp, we minimize the total energy, a weighted combination of all the above energy terms:

$$E = k_p E_p + k_h E_h + k_s E_s + k_{ls} E_{ls}.$$

E_p is the most important constraint so should have a larger weight. Results are relatively insensitive to choice of parameters; all figures in the paper used

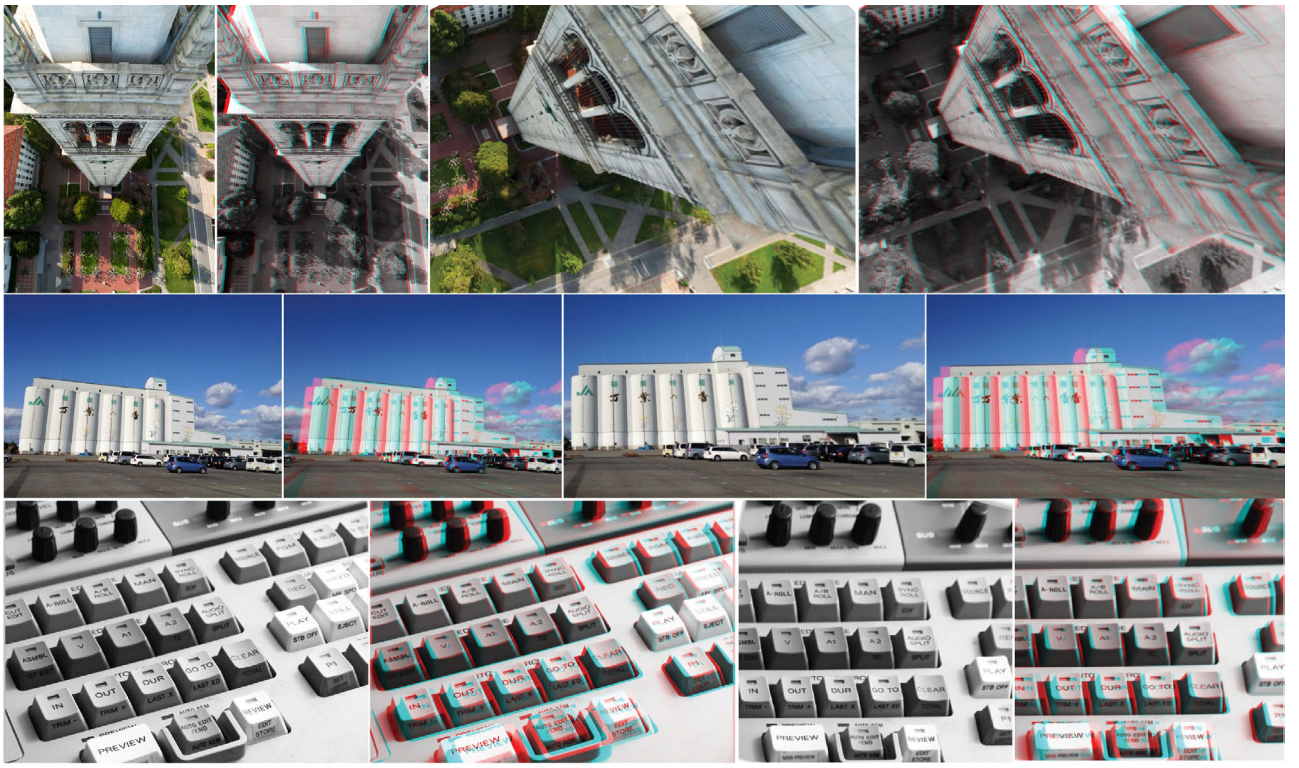

Fig. 3. Stereo images with manipulated perspective showing various orientation changes. Left to right: input left image, input anaglyph image, output left image with new perspective, output anaglyph image.

fixed weights of $k_p = 2, k_h = 1, k_s = 1, k_{ls} = 1$. The energy has been carefully designed to be a quadratic function in the coordinates of the grid mesh, so that it can be efficiently minimized via a sparse linear system.

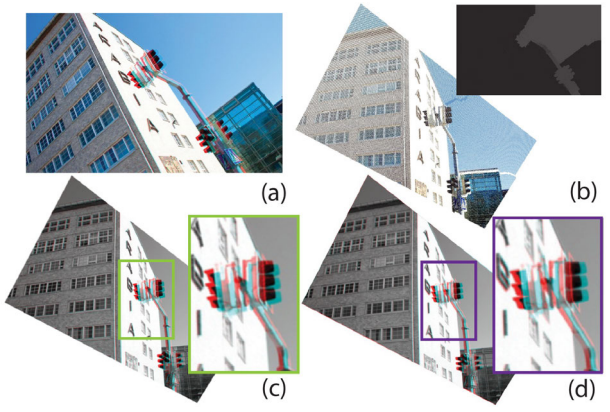

Fig. 4. Alternative approaches to perspective change. (a): input anaglyph image. (b)–(d): results by: direct rendering of the reconstructed point cloud, transformation by two homographies (note the large vertical disparity at the traffic light, and our method.

5 Results and Discussions

We have tested our algorithm on a variety of stereoscopic image pairs, as exemplified in Figure 3. A grid spacing of between 10 and 20 pixels sufficed for input images up to 1000^2 pixels, producing visually pleasing results in under 10s on a PC with a 3.0GHz quad core CPU and 4GB RAM. Our method produces plausible results with almost purely horizontal disparity, providing a comfortable viewing experience; an approach using two homographies following [4] does not—see Figures 4 and 5.

We have also compared our method with a naive depth-based approach, using the technique in [35] to compute a dense disparity map, and rendering the reconstructed 3D point cloud from the new perspective. However, computing a dense disparity map is neither efficient nor robust, and as the perspective changes, the depth-based method leads to gaps, which need extra non-trivial inpainting operations [6] to fill—see Figure 4. Warping avoids such gaps, and still produces visually acceptable results.

5.1 Applications

Changing perspective enables many other editing operations on stereoscopic images. Here we demonstrate four uses of perspective manipulation.

5.1.1 Disparity Mapping

Our method can be used to change the perceived depth of a stereoscopic image pair. Similar results can also be generated using the technique in [3].

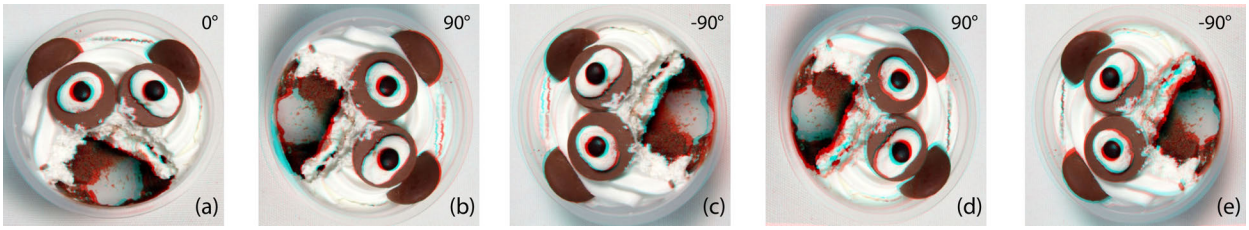

Fig. 5. Camera rotation along the z axis. (a): input anaglyph image. (b,c): results produced by our method. (d,e): results of transformation by two homographies.

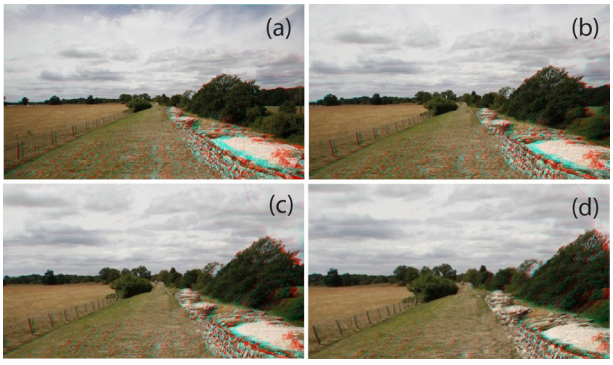

Fig. 6. Dolly zooming: (a) input anaglyph, (b-d) later shots, moving the camera away while zooming in.

By changing the relative position of the stereoscopic cameras, the disparity is changed in a scene-centric manner, rather than relying on a user designed disparity filter: our approach generates physically realistic results based on the camera parameters.

5.1.2 Wide-angle Effect

A wide-angle effect is often intentionally used for artistic effect in photography. We can generate wide-angle-like stereoscopic images by adding lens distortion to each camera. In this case, we omit the E_{ls} term since lines should no longer be straight after projection. However, lens radial distortion introduces vertical disparity [5], so we adapt the energy E_p. For a pair of feature points $\mathbf{x}_l \in P_l$ and $\mathbf{x}_r \in P_r$, with new positions $\tilde{\mathbf{x}}_l \in \tilde{P}_l$ and $\tilde{\mathbf{x}}_r \in \tilde{P}_r$, we constrain the vertical position of warped feature points $w_l(\mathbf{x}_l)$ and $w_r(\mathbf{x}_r)$ to the average vertical position of $\tilde{\mathbf{x}}_l$ and $\tilde{\mathbf{x}}_r$: we redefine E_p to be

$$\begin{aligned} E'_p &= \sum \left(w_l(\mathbf{x}_l)[x] - \tilde{\mathbf{x}}_l[x]\right)^2 \\ &+ \sum \left(w_r(\mathbf{x}_r)[x] - \tilde{\mathbf{x}}_r[x]\right)^2 \\ &+ \sum \left(w_l(\mathbf{x}_l)[y] - \frac{\tilde{\mathbf{x}}_l[y] + \tilde{\mathbf{x}}_r[y]}{2}\right)^2 \\ &+ \sum \left(w_r(\mathbf{x})[y] - \frac{\tilde{\mathbf{x}}_l[y] + \tilde{\mathbf{x}}_r[y]}{2}\right)^2. \end{aligned}$$

This leads to a fisheye-like effect while constraining horizontal disparity, as shown in Figure 1. Note that for zero camera distortion, $E'_p \equiv E_p$.

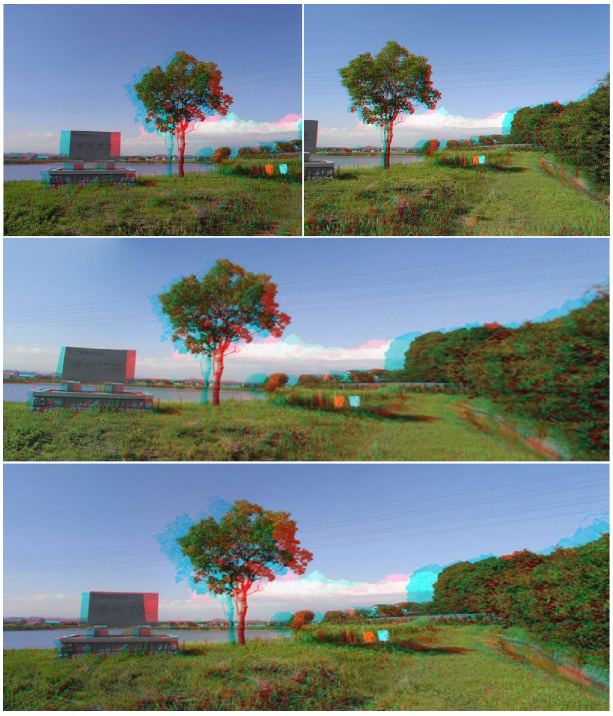

Fig. 7. Stereoscopic stitching example. Top: input two stereo pairs. Middle: panorama generated using our method. Bottom: panorama generated by stitching the left images and the right images separately; note the disparity near the trees (too large and including vertical disparity), and the foreground grass (vertical disparity), leading to poor 3D viewing.

5.1.3 Dolly Zoom

A dolly zoom effect can be produced by zooming out while moving towards the subject. We can emulate this to produce an interesting dynamic effect from a still stereoscopic image pair—see Figures 6(a-d). The camera moves forwards along the z-axis, while the focal length of the stereo cameras is adjusted to zoom out simultaneously.

5.1.4 Stereoscopic Stitching

Stitching is commonly used to merge multiple images, e.g. to increase the image resolution or angle of view. Assume we have two stereo image pairs (I_{l0}, I_{r0}) and (I_{l1}, I_{r1}) captured from two different perspectives (so four monocular viewpoints). A panorama could be

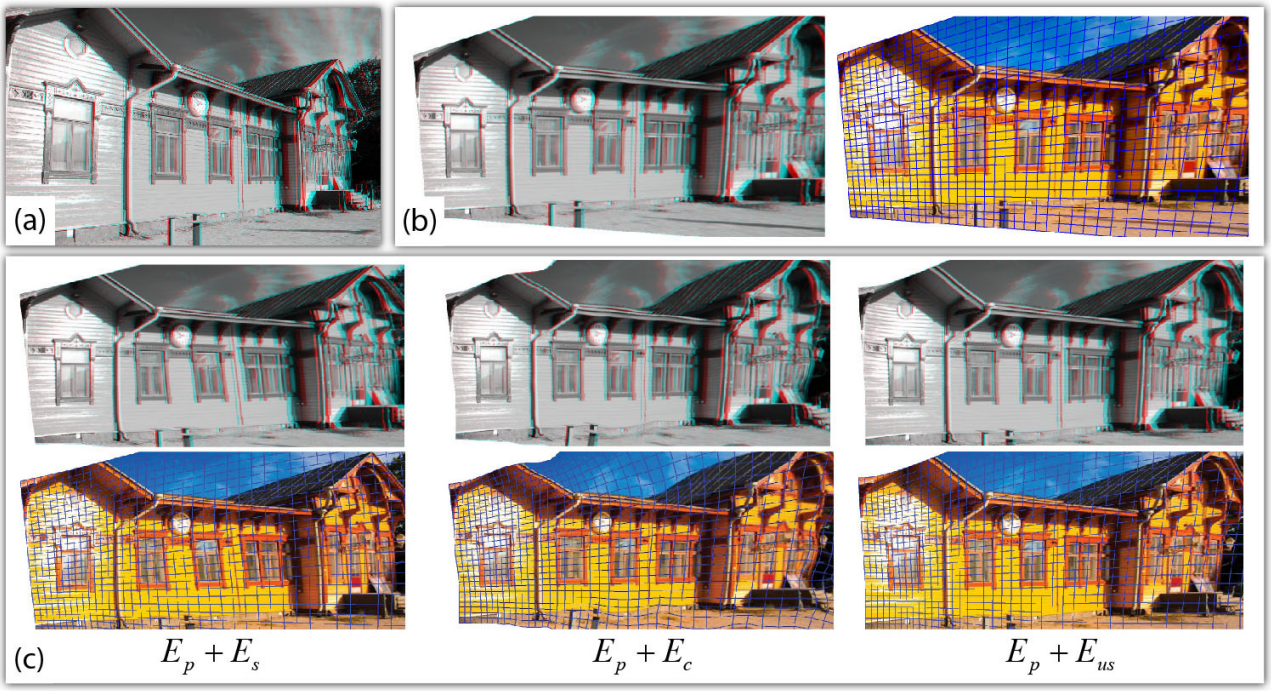

Fig. 8. Comparison of alternative energy terms. (a) Input anaglyph. (b) Our result and superimposed warping grid. (c) Results using $E_p + E_s$, $E_p + E_c$, and $E_p + E_{us}$.

naively constructed by stitching the two left images I_{l0}, I_{l1} and two right images I_{r0}, I_{r1} separately, resulting in left and right panorama images. However, this will probably break the stereopsis, since in the pinhole camera model, this is equivalent to applying different homographies individually to the two images in each stereo pair (see Section 5), causing incorrect disparity.

Instead, we proceed as follows: we change the perspective of one stereo image (I_{l0}, I_{r0}) by changing the stereo camera oreintation, obtaining (I'_{l0}, I'_{r0}), such that (I'_{l0}, I'_{r0}) and (I_{l1}, I_{r1}) can be directly stitched (through translation) to construct a rotational stereoscopic panorama. This will lead to better results with little vertical disparity (see Figure 7). In practice, the change of the stereo camera orientation can be approximated as the average rotation of the two views, computed from the feature correspondences between I_{l0} and I_{l1}, I_{r0} and I_{r1} separately.

5.2 Homography Constraint

Unlike previous warping based stereoscopic image or video editing techniques as in [3], [11], [23], a homography constraint E_h is introduced to guarantee proper perspective change under warping. [3] is mainly concerned with disparity changes while preserving image content, and thus uses saliency constraints to minimize visual distortion, enforcing the local Jacobian to be an identity matrix in salient regions. [23] uses a similarity transformation energy term to ensure the warp is content-preserving. These approaches are inappropriate for perspective changes. We illustrate the benefit of our homography constraint energy term E_h and compare its use with two other energy terms: a uniform scaling constraint E_{us} expressed as $\frac{\partial w_x}{\partial x} = \frac{\partial w_x}{y}, \frac{\partial w_x}{y} = \frac{\partial w_y}{x} = 0$ and a conformal constraint E_c expressed as $\frac{\partial w_x}{\partial x} = \frac{\partial w_x}{y}, \frac{\partial w_x}{y} = \frac{\partial w_y}{x}$.

Fig.8(c) shows the results of changing the perspective of the input stereo image shown in Fig.8(a) using different energy terms: $E_p + E_s$, $E_p + E_c$ and $E_p + E_{us}$. E_s, E_c and E_{us} are all 2D image content based energies. The warping function performs 2D interpolation of feature point warps, without any consideration of the underlying disparity and the changed perspective. As our experiments in Fig.8 show, these alternative energies lead to noticeable distortion in the image. Our result is shown in Fig.8(b), and is less distorted.

5.3 Depth Discontinuities

Discontinuity of disparity is common in stereoscopic images, and special attention must be paid to such regions since the warping function is not smooth. Ideally, when changing the perspective of an image, new content should be created (where new parts of the scene become exposed) and occluded regions should be hidden, as shown in Fig. 9(a). This is a difficult underconstrained problem, one solution to which is the costly process of stereo inpainting [6].

The alternative used in our energy terms E_h and E_s is to allow discontinuous warps [11]. As described in Sec.4, E_h and E_s should not be summed over discontinuities (in practice, we use a small but non-

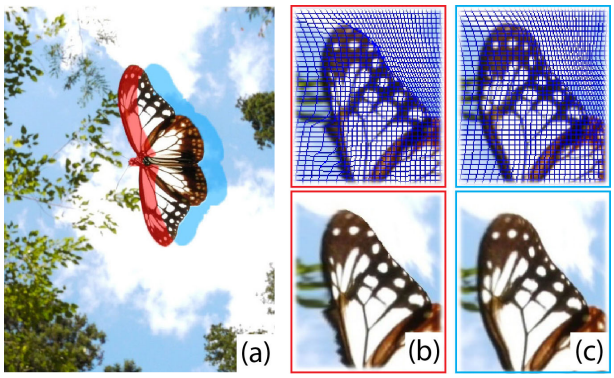

Fig. 9. Warping near occluding regions. (a) Under large camera translation, new content is needed (red region) and occluded regions should be hidden (blue). (b) Warping without using E_h. (c) Our result.

zero weight $w_\epsilon = 0.001$ to prevent reduced rank in the coefficient matrix).

Near occluding edges, warping fills holes and hides overlapping regions by locally stretching or shrinking the image content. As noted in [11], a discontinuous disparity map can produce significant over-stretching and over-shrinking near occluding regions, [11] introduced a content-and-disparity aware hole filling technique which allocates distortion preferentially to the background. In our case, the homography energy term E_h is computed from a interpolated smooth disparity map, and helps to locate the distortion near occluding regions. This produces acceptable results without noticeable artifacts. A comparison is shown in Fig.9(b) (without using E_h) and Fig.9(c) (using E_h).

5.4 Camera Parameters

Our method needs the stereo camera parameters. Determining camera parameters from multi-view images has been carefully studied [36], but it is difficult to determine accurate values. However, we have found that approximately knownf camera parameters suffice to provide visually acceptable results. Fig. 10 demonstrates the robustness of our methods.

5.5 Validation

Viewing comfort for stereo images and video is affected by various factors [37], and there is no general objective measurement of stereo visual quality. However, large vertical disparity and excessive horizontal disparity are two factors widely-accepted to increase visual discomfort. We compute an approximate projection under a new perspective for stereo images, and consequently the range of horizontal disparity is highly dependent on the user-specified perspective change. Table 1 shows the average vertical disparity (in pixels) at feature points before and after perspective manipulation for examples in this paper.

TABLE 1
Average vertical disparity for examples in this paper, before and after manipulation.

Fig No.	1	3(a)	3(b)	3(c)	4	5	10
Before	0.2	0.5	0.4	0.5	0.3	0.4	1.2
After	0.2	0.3	0.3	0.3	0.3	0.5	0.3

We also used a subjective experiment to validate our method—a user study was conducted to compare the visual quality of the original and manipulated stereo images. Ten participants were asked to give a integer score from -5 (original image is better) to 5 (manipulated is better), both for red-cyan stereo, and active-shutter stereo. The results in Table.2 show that the perceived quality of the manipulated stereo images was comparable to that of the originals.

TABLE 2
Average evaluation score

Fig No.	1	3(a)	3(b)	3(c)	4	5	10
Avg Score	0.8	-0.9	1.3	0.5	1.0	-0.5	-0.5

5.6 Limitations

Our method inevitably has limitations, the main one being that we can not handle large camera translations well, as these lead to large distortions. Warping cannot produce new scenery required under the new perspective where it is missing from the input, nor can it hide information where it is occluded. A further problem arises when there is a large orientation change at occluding edges, which produces locally noticeable distortions if image content there has strong structures. Figure 11 shows two such examples.

6 Conclusion

Perspective manipulation is a useful tool for image editing, and allows the user to apply various effects after image capture. In this paper, we have presented a practical tool for manipulating perspective in a stereoscopic image pair. We pose the problem based on warping each image in the stereoscopic image pair, using energy terms to constrain them to maintain correct geometric information and visual plausibility under the new perspective. The resulting quadratic minimization problem can be efficiently solved via a sparse linear system.

The 3D information in a stereoscopic image pair is too limited to enable us to make large changes to the projection, especially when moving the camera by a significant distance, as occlusions in the original may result in gaps. There is potential to improve the results by image inpainting. Other extensions to this work could consider multiple stereoscopic images and 3D

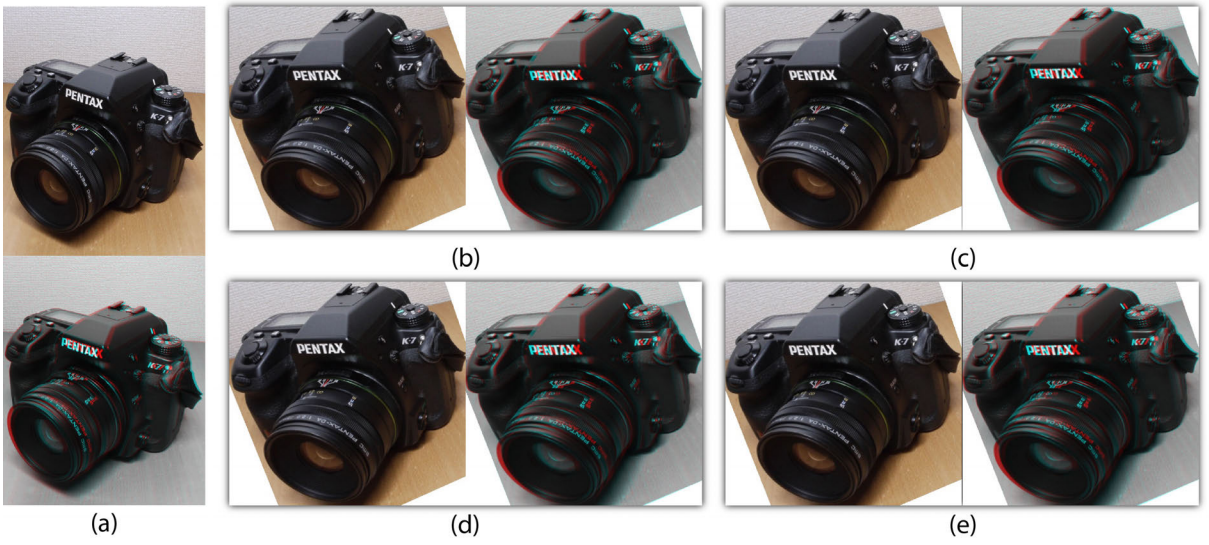

Fig. 10. Effects of errors in determined camera parameters. Perspective of input image (a) is changed under user-specified camera parameters, resulting in (b). (c): Focal length is changed by 10%. (d): Optical center is shifted by 10%. (e): Camera separation is increased by 10%.

movies, when ensuring temporal coherence will be a challenge, as well as choosing a suitable camera path for the new perspective views.

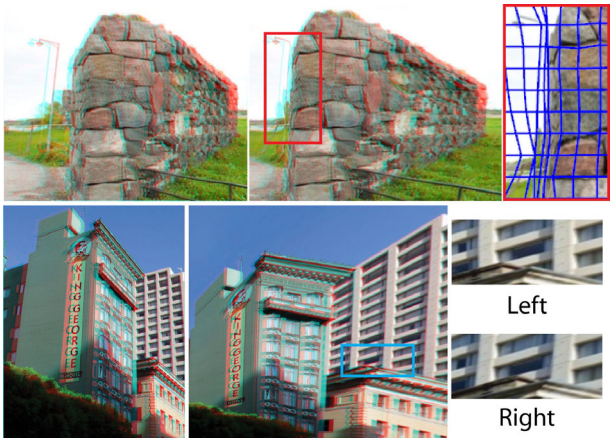

Fig. 11. Examples with distortions. Left: input anaglyph. Right: output anaglyph and close-up.

ACKNOWLEDGEMENT

This work was supported by the National Basic Research Projects of China (Project 2011CB302205), the Natural Science Foundation of China (Projects 61272226 and 61120106007), the National High Technology Research and Development Program of China (Project 2012AA011801), and an EPSRC travel grant. Shi-Min Hu is the corresponding author.

REFERENCES

[1] T. Oskam, A. Hornung, H. Bowles, K. Mitchell, and M. Gross, "Oscam - optimized stereoscopic camera control for interactive 3d," *ACM Trans. Graph.*, vol. 30, pp. 189:1–189:8, 2011. [Online]. Available: http://doi.acm.org/10.1145/2070781.2024223

[2] S. Heinzle, P. Greisen, D. Gallup, C. Chen, D. Saner, A. Smolic, A. Burg, W. Matusik, and M. Gross, "Computational stereo camera system with programmable control loop," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 94:1–94:10, Aug. 2011. [Online]. Available: http://doi.acm.org/10.1145/2010324.1964989

[3] M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic, and M. Gross, "Nonlinear disparity mapping for stereoscopic 3d," *ACM Trans. Graph.*, vol. 29, pp. 75:1–75:10, 2010. [Online]. Available: http://doi.acm.org/10.1145/1778765.1778812

[4] W.-Y. Lo, J. van Baar, C. Knaus, M. Zwicker, and M. Gross, "Stereoscopic 3d copy & paste," *ACM Trans. Graph.*, vol. 29, pp. 147:1–147:10, 2010. [Online]. Available: http://doi.acm.org/10.1145/1882261.1866173

[5] A. Woods, T. Docherty, and R. Koch, "Image distortions in stereoscopic video systems," in *Proceedings of the SPIE 1915*, 1993.

[6] L. Wang, H. Jin, R. Yang, and M. Gong, "Stereoscopic inpainting: Joint color and depth completion from stereo images," in *Proceedings of IEEE CVPR*, 2008, pp. 1 –8.

[7] S. Koppal, C. Zitnick, M. Cohen, S. B. Kang, B. Ressler, and A. Colburn, "A viewer-centric editor for 3d movies," *IEEE Computer Graphics and Applications*, vol. 31, no. 1, pp. 20 –35, 2011.

[8] R. Szeliski, "Image alignment and stitching: a tutorial," *Found. Trends. Comput. Graph. Vis.*, vol. 2, pp. 1–104, 2006. [Online]. Available: http://dl.acm.org/citation.cfm?id=1295184.1295185

[9] R. Carroll, A. Agarwala, and M. Agrawala, "Image warps for artistic perspective manipulation," *ACM Trans. Graph.*, vol. 29, pp. 127:1–127:9, 2010. [Online]. Available: http://doi.acm.org/10.1145/1778765.1778864

[10] A. Agarwala, M. Agrawala, M. Cohen, D. Salesin, and R. Szeliski, "Photographing long scenes with multi-viewpoint panoramas," *ACM Trans. Graph.*, vol. 25, pp. 853–861, 2006. [Online]. Available: http://doi.acm.org/10.1145/1141911.1141966

[11] O. Wang, M. Lang, M. Frei, A. Hornung, A. Smolic, and M. Gross, "Stereobrush: interactive 2d to 3d conversion using discontinuous warps," in *Proceedings of the Eighth Eurographics*

Symposium on Sketch-Based Interfaces and Modeling (SBIM '11), ACM, New York, NY, USA, 2011, pp. 47–54.

[12] L. B. Stelmach, W. J. Tam, F. Speranza, R. Renaud, and T. Martin, "Improving the visual comfort of stereoscopic images," in *Proceedings of SPIE 5006*, 2003.

[13] L. Meesters, W. IJsselsteijn, and P. Seuntiens, "A survey of perceptual evaluations and requirements of three-dimensional tv," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 3, pp. 381 – 391, 2004.

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

[15] H. J. Kim, J. W. Choi, A.-J. Chang, , and K. Y. Yu, "Reconstruction of stereoscopic imagery for visual comfort," in *Proceedings of SPIE 6803*, 2008.

[16] H. Pan, C. Yuan, , and S. Daly, "3d video disparity scaling for preference and prevention of discomfort," in *Proceedings of SPIE 7863*, 2011.

[17] C. Wang and A. A. Sawchuk, "Disparity manipulation for stereo images and video," in *Proceedings of SPIE 6803*, 2008.

[18] C.-H. Chang, C.-K. Liang, and Y.-Y. Chuang, "Content-aware display adaptation and interactive editing for stereoscopic images," *IEEE Transactions on Multimedia*, vol. 13, no. 4, pp. 589–601, 2011.

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

[20] Y.-S. Wang, C.-L. Tai, O. Sorkine, and T.-Y. Lee, "Optimized scale-and-stretch for image resizing," *ACM Trans. Graph.*, vol. 27, pp. 118:1–118:8, 2008. [Online]. Available: http://doi.acm.org/10.1145/1409060.1409071

[21] Y.-S. Wang, H. Fu, O. Sorkine, T.-Y. Lee, and H.-P. Seidel, "Motion-aware temporal coherence for video resizing," *ACM Trans. Graph.*, vol. 28, pp. 127:1–127:10, 2009. [Online]. Available: http://doi.acm.org/10.1145/1618452.1618473

[22] P. Krähenbühl, M. Lang, A. Hornung, and M. Gross, "A system for retargeting of streaming video," *ACM Trans. Graph.*, vol. 28, pp. 126:1–126:10, 2009. [Online]. Available: http://doi.acm.org/10.1145/1618452.1618472

[23] F. Liu, M. Gleicher, H. Jin, and A. Agarwala, "Content-preserving warps for 3d video stabilization," *ACM Trans. Graph.*, vol. 28, pp. 44:1–44:9, 2009. [Online]. Available: http://doi.acm.org/10.1145/1531326.1531350

[24] R. Carroll, M. Agrawal, and A. Agarwala, "Optimizing content-preserving projections for wide-angle images," *ACM Trans. Graph.*, vol. 28, pp. 43:1–43:9, 2009. [Online]. Available: http://doi.acm.org/10.1145/1531326.1531349

[25] J. Wei, C.-F. Li, S.-M. Hu, R. R. Martin, and C.-L. Tai, "Fisheye video correction," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. PrePrints, 2011.

[26] R. feng Tong, Y. Zhang, and K.-L. Cheng, "Stereopasting: Interactive composition in stereoscopic images," *IEEE Transactions on Visualization and Computer Graphics*, accepted, 2012.

[27] T. Cootes, G. Edwards, and C. Taylor, "Active appearance models," *IEEE TPAMI*, vol. 23, no. 6, pp. 681–685, 2001.

[28] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *IJCV*, vol. 47, pp. 7–42, 2002. [Online]. Available: http://dl.acm.org/citation.cfm?id=598429.598475

[29] http://vision.middlebury.edu/stereo/eval/.

[30] T. Sattler, B. Leibe, and L. Kobbelt, "Scramsac: Improving ransac's efficiency with a spatial consistency filter," in *Proceedings of IEEE ICCV*, 2009, pp. 2090 –2097.

[31] M. Brown, R. Szeliski, and S. Winder, "Multi-image matching using multi-scale oriented patches," in *Proceedings of IEEE CVPR*, vol. 1, 2005, pp. 510 – 517.

[32] Z.-N. Li, "Stereo correspondence based on line matching in hough space using dynamic programming," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 24, no. 1, pp. 144 –152, jan 1994.

[33] P. Bhat, C. L. Zitnick, M. Cohen, and B. Curless, "Gradientshop: A gradient-domain optimization framework for image and video filtering," *ACM Trans. Graph.*, vol. 29, pp. 10:1–10:14, 2010. [Online]. Available: http://doi.acm.org/10.1145/1731047.1731048

[34] N. J. Mitra, L. J. Guibas, and M. Pauly, "Partial and approximate symmetry detection for 3d geometry," *ACM Trans. Graph.*, vol. 25, pp. 560–568, 2006. [Online]. Available: http://doi.acm.org/10.1145/1141911.1141924

[35] B. M. Smith, L. Zhang, and H. Jin, "Stereo matching with nonparametric smoothness priors in feature space," in *Proceedings of IEEE CVPR*, 2009, pp. 485–492.

[36] C. T. Loop and Z. Zhang, "Computing Rectifying Homographies for Stereo Vision," in *Computer Vision and Pattern Recognition*, vol. 1, 1999, pp. 1125–1131.

[37] M. Lambooij, W. IJsselsteijn, and M. Fortuin, "Visual discomfort and visual fatigue of stereoscopic displays: A review," *Journal of Imaging Technology and Science*, vol. 53, pp. 1–14, 2009.

Song-Pei Du received his BS degree in computer science from Tsinghua University in 2009. He is currently a PhD candidate in the Department of Computer Science and Technology, Tsinghua University. His research interests include computer graphics, geometric modeling and image processing.

Shi-Min Hu received the PhD degree from Zhejiang University in 1996. He is currently a professor in the Department of Computer Science and Technology at Tsinghua University, Beijing. His research interests include digital geometry processing, video processing, rendering, computer animation, and computer-aided geometric design. He is associate Editor-in-Chief of The Visual Computer(Springer), and on the editorial boards of Computer-Aided Design and Computer & Graphics (Elsevier). He is a member of the IEEE and ACM.

Ralph Martin is currently a Professor at Cardiff University. He obtained his PhD degree in 1983 from Cambridge University. He has published more than 200 papers and 12 books, covering such topics as solid and surface modeling, intelligent sketch input, geometric reasoning, reverse engineering, and various aspects of computer graphics. He is a Fellow of: the Learned Society of Wales, the Institute of Mathematics and its Applications, and the British Computer Society. He is on the editorial boards of Computer Aided Design, Computer Aided Geometric Design, Geometric Models, the International Journal of Shape Modeling, CAD and Applications, and the International Journal of CADCAM.