

Object-Coherence Warping for Stereoscopic Image Retargeting

Shih-Syun Lin, Chao-Hung Lin, *Member, IEEE*, Shu-Huai Chang, and Tong-Yee Lee, *Senior Member, IEEE*

Abstract—This paper addresses the topic of content-aware stereoscopic image retargeting. The key to this topic is consistently adapting a stereoscopic image to fit displays with various aspect ratios and sizes while preserving visually salient content. Most methods focus on preserving the disparities and shapes of visually salient objects through nonlinear image warping, in which distortions caused by warping are propagated to homogenous and low-significance regions. However, disregarding the consistency of object deformation sometimes results in apparent distortions in both the disparities and shapes of objects. An object-coherence warping scheme is proposed to reduce this unwanted distortion. The basic idea is to utilize the information of matched objects rather than that of matched pixels in warping. Such information implies object correspondences in a stereoscopic image pair, which allows the generation of an object significance map and the consistent preservation of objects. This strategy enables our method to consistently preserve both the disparities and shapes of visually salient objects, leading to good content-aware retargeting. In the experiments, qualitative and quantitative analyses of various stereoscopic images show that our results are better than those generated by related methods in terms of consistency of object preservation.

Index Terms—Mesh warping, optimization, stereoscopic image retargeting.

I. INTRODUCTION

CONTENT-aware retargeting has elicited increasing attention in the field of computer science during the last decade. The basic idea of this technique is to adapt the aspect ratios of media data to fit various display devices while preserving visually salient content [1], [2]. This technique was recently applied to stereoscopic images because of the rapid development of stereoscopic equipment. Similar to image retargeting, seam carving, and mesh warping are recent techniques for stereoscopic images. The main idea of seam carving is to consistently remove seams passing through

Manuscript received July 31, 2013; revised September 18, 2013 and October 15, 2013; accepted October 16, 2013. Date of publication November 18, 2013; date of current version May 2, 2014. This work was supported in part by the Headquarters of University Advancement at the National Cheng Kung University, and in part by the National Science Council, Taiwan, under Grant NSC-100-2628-E-006-031-MY3, Grant NSC-100-2221-E-006-188-MY3, Grant NSC-101-2221-E-006-257-MY2, and Grant NSC-102-2221-E-006-194. This paper was recommended by Associate Editor L. Zhang.

S.-S. Lin, S.-H. Chang, and T.-Y. Lee are with the Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan 701, Taiwan (e-mail: catchyls@hotmail.com; c.shu.huai@gmail.com; tonylee@mail.ncku.edu.tw).

C.-H. Lin is with the Department of Geomatics, National Cheng Kung University, Tainan 701, Taiwan (e-mail: linhung@mail.ncku.edu.tw).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TCSVT.2013.2291282

homogenous regions in a stereoscopic image pair, and that of mesh warping is to optimize a deformation in a stereoscopic image. Although these two approaches perform well in most cases, seam carving may yield jagged edges and mesh warping may produce inconsistent object deformation. The presence of jagged edges and inconsistent object deformation will lead to discontinuous artifacts and incorrect disparities, respectively. An object-coherence warping method inspired by the concept of object-aware retargeting [3] is therefore proposed, in which the visually salient objects in an image pair are deformed consistently and rigidly, leading to consistent content preservation.

The basic idea behind the proposed method is to measure content significance and resize a stereoscopic image by utilizing the information of matched objects rather than that of matched pixels. The information of matched objects in an image pair allows the generation of an object significance map and the consistent preservation of visually salient objects in warping. In the proposed method, the input stereoscopic image is segmented into several objects and the corresponding objects in the left and right images are assigned the same significance value in the preprocessing. During warping, an object with high significance value is forced to undergo as-consistent-as-possible and as-rigid-as-possible deformation using various preservation constraints while propagating distortions through an optimization process.

The goal of preserving disparity of visually salient content in this paper is the same as that in [4] and [5], and the goal of preserving shapes of visually salient objects is the same as that in [3]. However, the proposed method is substantially different from these related methods. First, instead of the pixel-coherence warping in [4], an object-coherence warping is proposed in the current paper to ease the distortions caused by inconsistent object deformation. As shown in Fig. 1, the inconsistent deformation that appears in the disparity and shape of the foreground object is eased by our method. Second, an object-based significance map with consistent partition of a stereoscopic image pair is proposed to preserve the visually salient content. Third, continuous mesh warping is adopted rather than discontinuous one [5]. Instead of resizing each object layer separately, following the mesh-warping-based methods, our method combines all the object saliency values in a significance map and then a continuous mesh warping is performed on the basis of this significance map. Fourth, some consistent deformation constraints are introduced in the optimization to deform visually salient objects not only rigidly as in [3] but also consistently. These differences enable



Fig. 1. Deformation distortion. Left: original stereoscopic image. Middle: retargeting results generated by Chang *et al.* [4]. Disparity and shape distortion occurs on the road (see the blue lines) and car (see the close-up view at the bottom) because of the inconsistent object deformation. Right: our result. Distortion is eased by the proposed method.

our method to yield better retargeting results in terms of deformation consistency and disparity preservation, compared with the related methods.

II. RELATED WORK

Many content-aware retargeting techniques have been proposed in the last decade. Only the methods for stereoscopic images are discussed in this section. Readers can refer to techniques on image retargeting in the survey in [6] and video retargeting in recent works [7]–[10]. The retargeting methods are classified into discrete and continuous based on the categorization suggested in [6]. In the discrete methods, a stereoscopic image is resized by cropping [11] or seam carving [12], [13]. Niu *et al.* [11] propose an aesthetics-based cropping method in which an input image pair is optimally cropped and then rescaled such that the aesthetic value defined according to the principles of stereoscopic photography is maximized. Although this method performs well in most cases, it is inappropriate for the case in which visually salient contents are located near the borders of images. Utsugi *et al.* [12] and Basha *et al.* [13] extend the image seam-carving technique [1] to stereoscopic images. In these methods, a pair of one-pixel-wide seams with minimal significance in the left and right images is iteratively and simultaneously carved to resize the input images to the desired aspect ratio. This technique allows high flexibility in seam removal and can, therefore, be applied to object removal. However, carving seams sometimes produces discontinuous artifacts in the visually salient content, thereby causing visual distortion.

Compared with the methods that discretely remove pixels in homogenous regions or crop the borders of an image, continuous methods that optimize warping using several deformation and smoothness constraints potentially perform better in images containing dense information. In [4] and [5], stereoscopic image retargeting is formulated as a least-squares warping problem. Warping from a source data to a target display is optimized with the aid of a significance map and depth layer information. Therefore, the disparities and shapes of objects in the high-significance layers are preserved, whereas the low-significance regions are squeezed or stretched. However, inconsistent object deformation and discontinuous warping may occur in [4] and [5], respectively. Niu *et al.* [14] provide

various editing operations for stereoscopic images. The idea behind these operations is to warp one of the images in a stereoscopic pair using user-defined warping and then warp the other image to follow the previous warping and to preserve disparities. Wang and Sawchuk [15] develop a disparity manipulation system in which the warping technique is utilized for view synthesis. Lang *et al.* [16] discuss the perceptual aspects of stereo vision and their applications for content manipulation. They provide a set of basic disparity mapping operators and a warping function to achieve desirable disparity distributions. Luo *et al.* [17] and Liu *et al.* [18] further apply the mesh-warping technique to seamless stereoscopic image cloning and authoring, respectively. Although the above-mentioned retargeting methods can yield good results, the lack of consideration of the consistency of object deformation or the continuity of mesh warping may result in apparent distortions in the disparities and shapes of visually salient objects. In this paper, we ease the inconsistency of deformation by utilizing the information of matched objects. In addition, following [2], we adopt a continuous warping scheme rather than a discontinuous one.

The concept of object-aware retargeting has been introduced in the image-retargeting approaches [3], [19], [20]. Niu *et al.* [19] and Zhang *et al.* [20] proposed to extract and preserve foreground objects in warping based on the fact that users are more interested in foreground objects. These two methods can preserve the shapes of foreground objects well. However, the problem of over-constraining may make the methods unsuitable for images/videos that contain visually salient content in the background. To solve this problem, Lin *et al.* [3] proposed a patch-based warping method to force visually salient objects to undergo as-rigid-as-possible deformation, while low significance contents are warped as close as possible to linear rescaling. In this paper, an object-coherence warping is proposed to deform visually salient objects not only rigidly but also consistently. In this way, both the disparities and shapes of visually salient objects can be well preserved.

III. STEREOSCOPIC IMAGE RETARGETING

The proposed content-aware retargeting is performed in two steps: object-based significance map generation and object-coherence mesh warping. In the preprocessing, the left and right images are consistently segmented into several objects and significance measurement is then performed for the segmented object pairs to generate a significance map. In this step, context-aware saliency estimation [21] is adopted to estimate the saliency value of each pixel in the left and right images. Afterward, each segmented object pair is assigned an average saliency value to address the problem of inconsistent object deformation. In the retargeting, two grid meshes are created to cover the left and right images, and the proposed object-coherence warping is performed to force the grids within an object to undergo as-consistent-as-possible and as-rigid-as-possible deformation during resizing. We describe the significance map generation and the object-coherence warping in Sections III-A and III-B, respectively, followed by the optimization solver, which is described in Section III-C.

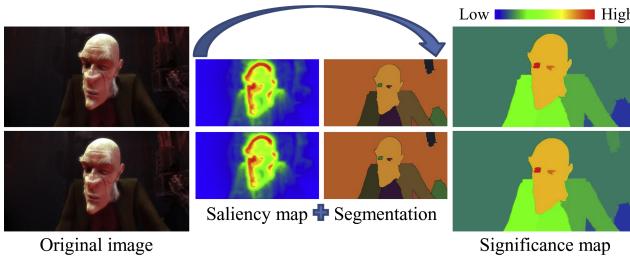


Fig. 2. Significance map generation. The significance map is designed by combining the estimated pixel saliences and segmented objects to reduce object inconsistencies in both shape and disparity deformation. The left and right images are shown at the top and bottom, the segmented objects are represented by colors, and the significance values are presented by colors ranging from blue (the lowest significance) to red (the highest significance).

A. Significance Map Generation

The significance map plays an important role in the content-aware retargeting algorithms, and many salient detection methods have been proposed [21]–[24]. A pixel with high saliency value is generally considered a significant pixel. Besides, the pixel significance is measured by considering the content in both the left and right images instead of generating significance maps individually [4], [13]. However, inconsistent deformation may arise when such a pixel-based measurement is utilized in retargeting. Our aim is to consistently preserve the disparities and shapes of high-significance objects rather than those of high-significance pixels. We, therefore, propose an object-based significance measurement. Given a stereoscopic image, the left and right images are simultaneously segmented into several homogenous objects using spatiotemporal segmentation [25]. Each object is assigned the average of saliency values of pixels within that object. Pixel saliency is estimated using the saliency-detection approach [21].

The left and right images are viewed as overlaid images in the segmentation; thus, the images are partitioned into several two-slide volumetric objects through a hierarchical graph-based segmentation technique. Segmentation quality is then improved using the optical flow technique to establish the connections between the left and right images in the partition. However, the problem of over-segmentation may occur. To address this problem, the commonly used merge process is performed with the information of pixel color and object disparity. Specifically, the adjacent objects with similar color and disparity are merged.

Once the stereoscopic image is segmented and the pixel saliency is obtained, the significance value of an object is calculated by averaging the saliency values of pixels within the object. Each mesh grid is assigned the significance value of the object that occupies this grid. With the aid of significance map, the grids occupied by an object have the same significance value; thus, this object in the stereoscopic image can be consistently deformed in warping. Fig. 2 provides an example of stereoscopic image segmentation and significance map generation. The objects, including the foreground and background objects, in the left and right images are consistently extracted, and the object significance values are calculated by combining the information of segmented objects and pixel saliency values. This significance map can enable our method to consistently

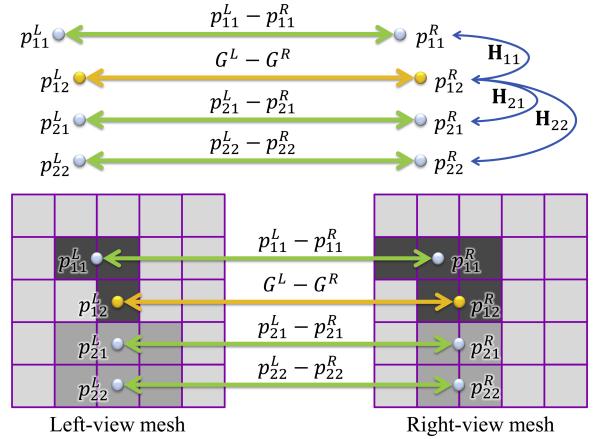


Fig. 3. Illustration of disparity preserving term. G^L and G^R are the representative pixels of the left and right images, respectively; p_{ij}^L and p_{ij}^R are the representative pixels of the j th row of the object o_i in the left and right images, respectively, and H_{ij} is the similarity transformation between the vectors $(p_{ij}^L - p_{ij}^R)$ and $(G^L - G^R)$.

preserve the disparities and shapes of objects during warping. Note that generating perfect segmentation is difficult, even when a state-of-the-art segmentation algorithm is used. Fortunately, the proposed approach can ease the problem of imperfect object segmentation by propagating distortions to nearby regions. This characteristic is discussed in Section IV-A.

B. Object-Coherence Warping

A uniform grid mesh pair $(\mathbf{M}^L, \mathbf{M}^R)$ is created to cover the input stereoscopic image pair $(\mathbf{I}^L, \mathbf{I}^R)$. Each grid mesh contains a vertex set $\mathbf{V} = \{v_1, \dots, v_{n_v}\}$, an edge set $\mathbf{E} = \{e_1, \dots, e_{n_e}\}$, and a grid set $\mathbf{Q} = \{q_1, \dots, q_{n_q}\}$. Here, n_v , n_e , and n_q represent the number of vertices, edges, and grids, respectively. In addition, a set of objects $\mathbf{O} = \{o_1, \dots, o_{n_o}\}$ and their significance values $\mathbf{S} = \{s_1, \dots, s_{n_o}\}$ obtained in the preprocessing are employed in the warping, where n_o represents the number of segmented objects. Two energy terms, namely, disparity preservation and shape preservation, are defined with an optimization solver to preserve both the disparity information and spatial shapes. These two energy terms are described as follows.

Disparity preservation energy. Our warping method aims to obtain a pair of deformed grid meshes $(\tilde{\mathbf{M}}^L, \tilde{\mathbf{M}}^R)$ in which the high-significance objects are deformed consistently for disparity preservation. Two energy terms, namely disparity preserving and deformation consistency, are defined for this purpose. The disparity preserving term is to preserve the original disparity values as much as possible and to avoid vertically shifting effects happened between the corresponding pixels in the left and right images. In the related work, the disparity energy is generally defined as the difference between the disparity values of the corresponding pixels in the original and deformed images. Although the disparity can be preserved in this manner, disregarding the consistency of disparity values of the pixels within an object may result in unnatural stereoscopic content. To address this problem, the concept of object-based preservation is introduced to this term. Specifically, the disparity preserving energy is formulated by

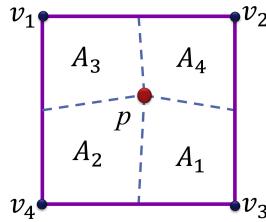


Fig. 4. Illustration of barycentric coordinate. A_1, A_2, A_3, A_4 are the areas of quadrangles corresponding to vertices v_1, v_2, v_3, v_4 , respectively, which are used to calculate the barycentric coordinate of p .

measuring the consistency of the local deformation of objects in the left and right images as

$$\Psi_{Dp}(\mathbf{I}^L, \mathbf{I}^R) = \sum_{i=1}^{n_o} s_i \times \sum_{j=1}^{n_{row}} \|(\tilde{p}_{ij}^L - \tilde{p}_{ij}^R) - \mathbf{H}_{ij}(\tilde{G}^L - \tilde{G}^R)\|^2 \quad (1)$$

where s_i is the significance value of object o_i ; n_{row} represents the number of rows in the object o_i ; and \tilde{p}_{ij}^L and \tilde{p}_{ij}^R are the representative pixels of the j th row of the object o_i in the deformed left and right images, respectively (see Fig. 3). The pixel close to the center of each row within an object is selected as the representative pixel of that row. \tilde{G}^L and \tilde{G}^R are the representative pixels of the deformed left and right images, respectively. The vector formed by these two pixels, that is, $(\tilde{G}^L - \tilde{G}^R)$, serves as the deformation pivot of the vectors $(\tilde{p}_{ij}^L - \tilde{p}_{ij}^R)$ for the consideration of disparity preservation. In the implementation, the nonzero vector $(p_{ij}^L - p_{ij}^R)$ closest to the vector formed by the centers of the left and right images is selected as the pivot vector. \mathbf{H}_{ij} is the similarity transformation between the vectors $(p_{ij}^L - p_{ij}^R)$ and $(G^L - G^R)$ in the original images. Therefore, this energy measures the changes in the geometric relations of the corresponding rows of an object during warping. In other words, this energy based on the concept of consistent deformation of local object can achieve the goals of disparity preservation and vertically shifting effect alleviation.

The points p_{ij} and G lie within the grids of mesh, denoted by $q(p_{ij}) = \{v_m\}_{m=1}^4$ and $q(G) = \{v_n\}_{n=1}^4$. Therefore, p_{ij} and G can be rewritten as a linear combination of the grid's vertices using the barycentric coordinate given by

$$p = b_1 v_1 + b_2 v_2 + b_3 v_3 + b_4 v_4 \quad (2)$$

where (b_1, b_2, b_3, b_4) are the coefficients of barycentric coordinate of p corresponding to the vertices (v_1, v_2, v_3, v_4) . The barycentric coordinate is known as an area coordinate. The coefficients (b_1, b_2, b_3, b_4) are proportional to the corresponding areas (A_1, A_2, A_3, A_4) (see Fig. 4); that is, $b_k = A_k / (A_1 + A_2 + A_3 + A_4)$, $k = 1 \dots 4$. By using the barycentric coordinate, we can reformulate (1) as

$$\begin{aligned} \Psi_{Dp}(\mathbf{M}^L, \mathbf{M}^R) = & \sum_{i=1}^{n_o} s_i \times \sum_{j=1}^{n_{row}} \left\| \left(\sum_{m=1}^4 b_m^L \tilde{v}_m^L - \sum_{m=1}^4 b_m^R \tilde{v}_m^R \right) \right. \\ & \left. - \mathbf{H}_{ij} \left(\sum_{n=1}^4 b_n^L \tilde{v}_n^L - \sum_{n=1}^4 b_n^R \tilde{v}_n^R \right) \right\|^2 \end{aligned} \quad (3)$$

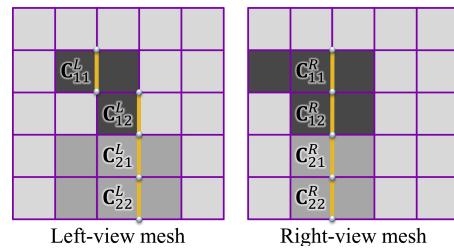


Fig. 5. Illustration of deformation consistency energy term. C^L and C^R are the row representative edges, which are used as pivots for consistent row deformation.

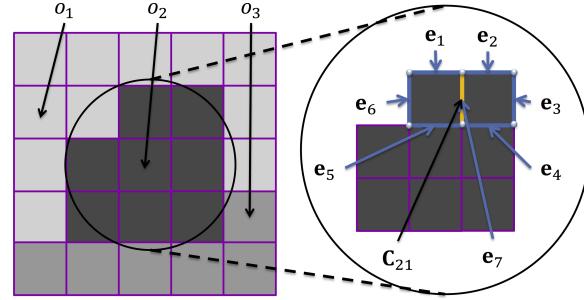


Fig. 6. Illustration of shape deformation energy. C_{21} is the pivot of the edges $\{e_1 \dots e_7\}$ in the first row of object o_2 .

where $\{b_m^L\}_{m=1}^4$, $\{b_m^R\}_{m=1}^4$, $\{b_n^L\}_{n=1}^4$, and $\{b_n^R\}_{n=1}^4$ are the barycentric coordinates of the points p_{ij}^L , p_{ij}^R , G^L , and G^R , respectively, with respect to the grids $\{v_m^L\}_{m=1}^4$, $\{v_m^R\}_{m=1}^4$, $\{v_n^L\}_{n=1}^4$, and $\{v_n^R\}_{n=1}^4$. By using this formulation, we can combine this energy term with the other terms and then solve the optimization for the vertices $(\tilde{\mathbf{V}}^L, \tilde{\mathbf{V}}^R)$ in the deformed grid meshes rather than the pixels.

To further address the problem of inconsistent deformation, deformation consistency energy is introduced in the optimization. This term is formulated by measuring the distance between the row representative edges of a deformed object in the left and right images as

$$\Psi_{Dc}(\mathbf{M}^L, \mathbf{M}^R) = \sum_{i=1}^{n_o} \sum_{j=1}^{n_{row}} \left\| \tilde{\mathbf{C}}_{ij}^L - \tilde{\mathbf{C}}_{ij}^R \right\|^2 \quad (4)$$

where $\tilde{\mathbf{C}}_{ij}^L$ and $\tilde{\mathbf{C}}_{ij}^R$ represent the deformed row representative edges of the j th row of the object o_i in the left and right grid meshes, respectively (see Fig. 5). Similarly, the edge closest to the center of a row in an object is suitable to represent that row; thus, this edge is selected as the row representative edge. In this manner, the local object is represented by its representative edge and the difference between the corresponding representative edges can measure the local deformation of the corresponding objects in the left and right images.

The total disparity preservation energy is obtained by summing up the individual energy terms as

$$\Psi_{DP} = \alpha \times \Psi_{Dp} + (1 - \alpha) \times \Psi_{Dc} \quad (5)$$

where α is the weighting factor of these two disparity preservation terms. In the implementation, the default value of α is set to 0.5.

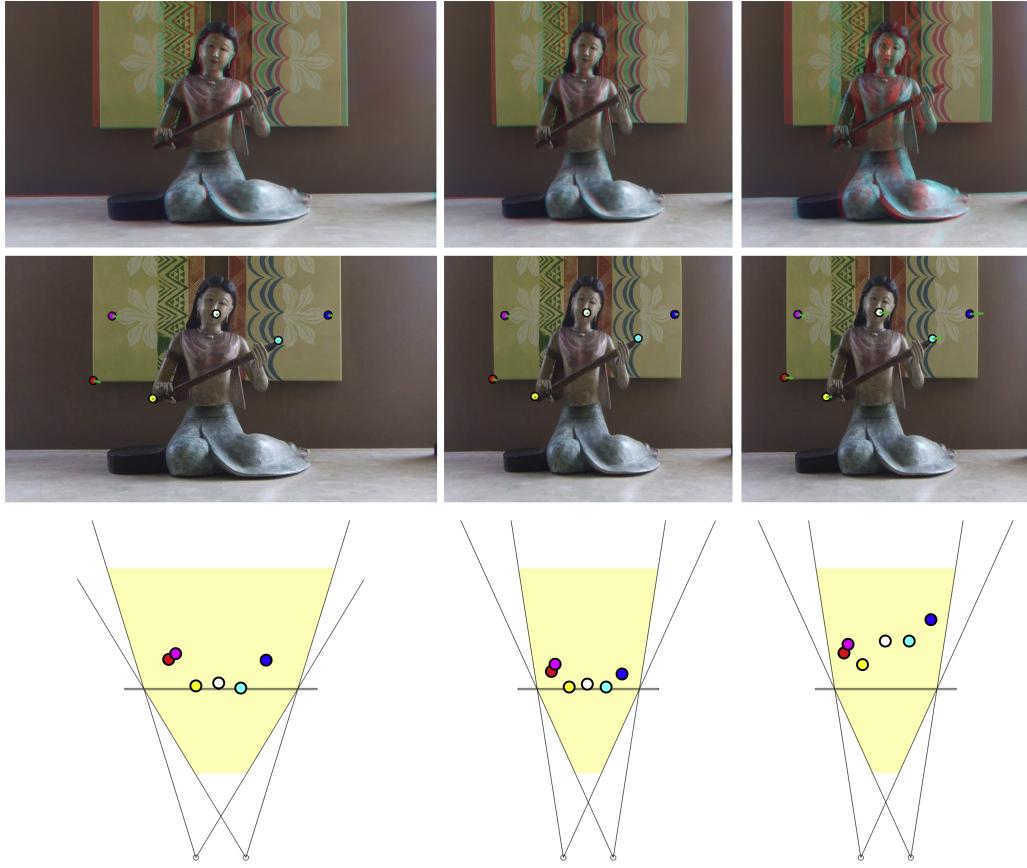


Fig. 7. Comparison of retargeting using our warping method (middle) with and (right) without disparity energy in the optimization. Top row: stereoscopic images. Middle row: left image with the manually selected feature points. Bottom row: depth distribution of the selected feature points.

Shape preservation energy. To achieve the goal of deforming high-significance objects as rigidly as possible during warping, two energy terms, namely shape deformation and line bending, are defined. In [3], the shape deformation term is defined by measuring the rigidity of objects as

$$\Psi_{Sd}(\mathbf{M}) = \sum_{o_i \in \mathbf{O}} (s_i \times \sum_{\mathbf{e}_j \in \mathbf{E}(o_i)} \|\tilde{\mathbf{e}}_j - \mathbf{T}_{ij} \tilde{\mathbf{R}}_i\|^2) \quad (6)$$

where s_i is the significance value of object o_i ; $\tilde{\mathbf{e}}_j$ represents an edge in the object o_i ; $\tilde{\mathbf{R}}_i$ is the representative edge of object o_i ; and \mathbf{T}_{ij} is the similarity transformation between the edge \mathbf{e}_j and the representative edge $\tilde{\mathbf{R}}_i$ in the original mesh. This definition works well for a single image. However, the lack of consideration of vertically shifting effects makes this definition inappropriate for stereoscopic images. By considering both the left and right images and alleviating the vertically shifting effect, we reformulate (6) as

$$\Psi_{Sd}(\mathbf{M}^L, \mathbf{M}^R) = \sum_{o_i \in \mathbf{O}^L \cup \mathbf{O}^R} s_i \times \sum_{j=1}^{n_{row}} \sum_{k=1}^{n_{edge}} \|\tilde{\mathbf{e}}_{ijk} - \mathbf{T}_{ijk} \tilde{\mathbf{C}}_{ij}\|^2 \quad (7)$$

where \mathbf{O}^L and \mathbf{O}^R are the object sets in the left and right images; n_{row} represents the number of rows in the object o_i ; and n_{edge} denotes the number of edge in the j th row of object o_i (see Fig. 6). This energy measures the changes in the geometric relations of the edges within an object's row in both the left and right images during warping. Through consistent

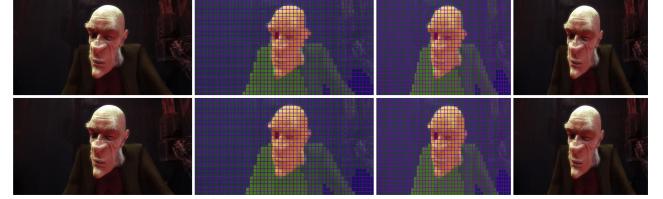


Fig. 8. Consistent shape deformation. From left to right: original stereoscopic image pair, grid mesh, and deformed mesh, and retargeting result. The grid and deformed meshes are visualized by significance values.

row deformation, this energy can address the problems of inconsistent spatial object deformation and vertical shifting.

To avoid skewed artifacts in the grid meshes, the line-bending term proposed in [2] is integrated in the optimization. The skewed artifacts are measured by the bending of grid lines. A grid is denoted by $q : \{v_a, v_b, v_c, v_d\}$ and contains two horizontal edges (v_a, v_b) and (v_d, v_c) , and two vertical edges (v_a, v_d) and (v_b, v_c) . This term is defined by measuring the distance of the y component between the vertices of the deformed horizontal edges and the distance of the x component between the vertices of the deformed vertical edges as

$$\begin{aligned} \Psi_{Lb}(\mathbf{M}^L, \mathbf{M}^R) = & \sum_{q \in \mathbf{Q}^L \cup \mathbf{Q}^R} (\|\tilde{v}_{ay} - \tilde{v}_{by}\|^2 + \|\tilde{v}_{dy} - \tilde{v}_{cy}\|^2 \\ & + \|\tilde{v}_{ax} - \tilde{v}_{dx}\|^2 + \|\tilde{v}_{bx} - \tilde{v}_{cx}\|^2). \end{aligned} \quad (8)$$

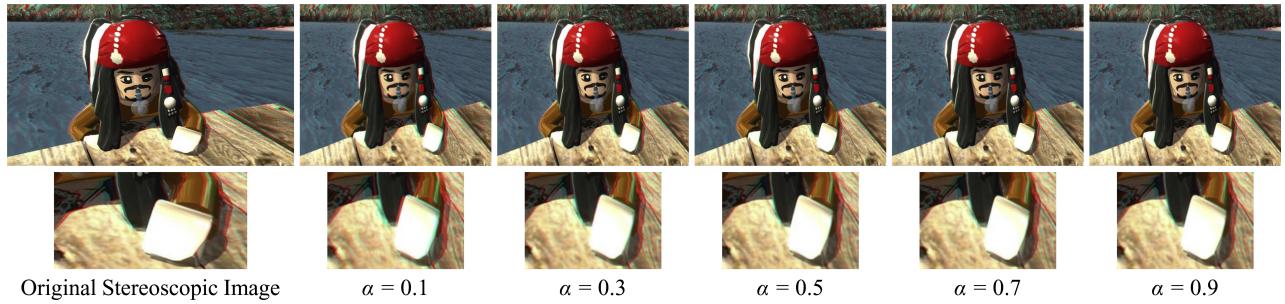


Fig. 9. Retargeting using different values of parameter α .

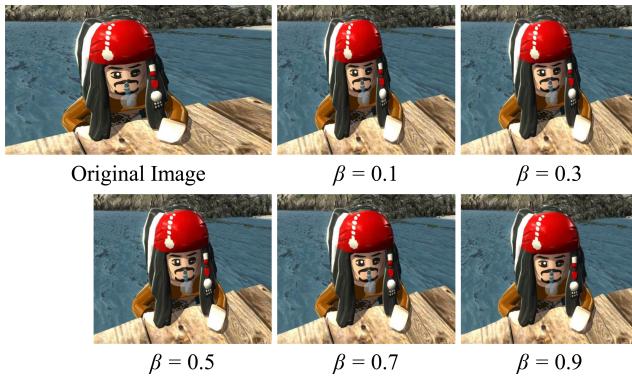


Fig. 10. Retargeting using different values of parameter β .

The total shape preservation energy is obtained by summing up the individual energy terms with their weights

$$\Psi_{SP} = \beta \times \Psi_{SD} + (1 - \beta) \times \Psi_{LB} \quad (9)$$

where β is the weighting factor that balances the contributions of the shape deformation term and the line bending term. In the experiments, β is set to 0.5 to take the abovementioned considerations into account.

C. Minimization of Energy Function

By combining the disparity and shape energies, the final optimization for the mesh warping is formulated as

$$\arg \min_{\tilde{\mathbf{V}}^L, \tilde{\mathbf{V}}^R} (\Psi_{SP} + \Psi_{DP}) \quad (10)$$

subject to the constraints of the positions of the boundary vertices. The top-left vertex position of the stereoscopic image is fixed in the implementation and all the boundary vertices are constrained to slide along their respective boundary lines. In the optimization, a least-squares linear system $\mathbf{A}[\tilde{\mathbf{V}}^L \tilde{\mathbf{V}}^R]^T = \mathbf{b}$ with a sparse design matrix \mathbf{A} can be obtained from (10). This system is solved for the deformed grid meshes $\tilde{\mathbf{V}}^L$ and $\tilde{\mathbf{V}}^R$ as well as for the retargeting images $\tilde{\mathbf{I}}^L$ and $\tilde{\mathbf{I}}^R$ by using an iterative solver called conjugate gradient method. The iterative process is terminated when the movements of vertices are smaller than 0.5 pixels.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Properties of Proposed Object-Coherence Warping

The proposed warping method has several properties that demonstrate its potential for content-aware retargeting of stereoscopic images. First, with the aid of disparity preservation energy, the disparity of stereoscopic image can be preserved as much as possible. A comparison between the retargetings with and without the disparity energy is shown in Fig. 7. The shape of the foreground object is preserved using the shape-preservation energy; however, the disparity values and their depth order are incorrect. By contrast, when both the disparity and shape preservation energies are used, not only the shapes but also the disparity values are preserved.

Second, our approach is capable of consistently preserving visually salient objects. With the aid of the object-based significance map, the grids occupied by an object have the same significance value. Therefore, the high-significance object pairs in the left and right images can be deformed consistently, and their shapes and disparities can be preserved. As demonstrated in Fig. 8, the foreground object in the left and right images is assigned a significance value. Therefore, this object is consistently deformed, and the shapes are well preserved in warping.

Third, the proposed warping scheme can ease the difficulty caused by imperfect segmentation. Accurate segmentation is difficult and oversegmentation with unfavorable object boundaries may occur. Fortunately, the scheme of distortion propagation in mesh optimization can address the problem of imperfect segmentation in warping. In addition, in the case of retargeting with extreme oversegmentation, the smallest object is a pixel, which means that the retargeting result is similar to that of pixel-based retargeting.

B. Retargeting Evaluation

Our algorithm is tested on a desktop PC with 2.66 GHz CPU and 4 GB memory. The computation cost of mesh warping depends on the number of nonzero entries (denoted as K) in the design matrix \mathbf{A} and the number of iterations (denoted as T) in the optimization solver. The time complexity of our mesh warping is $O(K\sqrt{T})$, which is the same with the mesh-warping-based method [3] that use conjugate gradient method as the optimization solver. For a stereoscopic image of 800×600 resolution, the average computation time for mesh warping is 0.18 s and that for preprocessing is



Fig. 11. Comparison of shape preservation with the seam-carving-based method (SSC) [13], mesh-warping-based method (SMW) [4], linear scaling (LS), and the proposed method. The left image and stereoscopic image are shown at the top and bottom, respectively.

6.76 s, including 0.65 s for segmentation and 6.1 s for saliency detection, which is similar to those of [4] that takes 6.56 s for preprocessing, including 4.93 s for feature matching and 1.63 s for saliency detection.

The parameters α and β are the main parameters in our method. Different parameter values are tested to show the sensitivity of the retargeting results to the parameters. In our method, α is the weighting factor for balancing the contributions of the disparity preserving term and the deformation consistency term. A large value can preserve disparity values and a small value can address the problem of inconsistent deformation (see Fig. 9). In the implementation, α is set to 0.5. β is the weighting factor for the shape deformation term and the line-bending term. From the experimental result shown in Fig. 10, a large value can force high-significance objects to become rigid in warping and a small value can alleviate skewed artifacts. In the experiments, β is set to 0.5 to take the abovementioned considerations into account. Note that the parameters α and β are tunable. $\alpha = 0.5$ and $\beta = 0.5$ are the default values in our retargeting system and this selection is used in all the experiments.

To give a fair comparison, most images used in related works are tested in the experiments. In addition, some tested images are selected from the datasets CMU/VASC [26] and NVIDIA [27]. Several representative cases where images have evident foreground objects and structure lines are shown in Figs. 11 and 12. All the experimental materials are available on the following website: <http://graphics.csie.ncku.edu.tw/Stereoscopic-Image-Retargeting>. The experimental results are automatically generated based on the default parameters, that is, grid resolution is 20 pixels \times 20 pixels and $\alpha = \beta = 0.5$. Our method is compared with the recent seam-carving-based method (SSC) [13] and mesh-warping-based method (SMW) [4]. In addition, to further validate the preservation of disparity, our method is compared with the recent single-view mesh warping method (MW) [3]. The resolution of the grid mesh in the proposed method is the same as that in the MW and SMW methods to ensure objectivity in the comparison. The key to successful content-aware retargeting of stereoscopic images is the preservation of shape and disparity. We, therefore, conduct the com-

parisons of shape preservation (Fig. 11) and disparity preservation (Fig. 12). The results shown in Fig. 11 indicate that SSC generates discontinuous artifacts in visually salient objects because of the removal of seams, particularly for images containing dense information. The results shown in Figs. 11 and 12 indicate that the mesh warping methods MW and SMW have the advantage of absorbing distortion by homogeneous regions. However, considering only spatial content preservation in MW leads to unnatural stereoscopic content and the lack of considering consistent object warping in SMW sometimes results in apparent deformation on the structure lines, which are sensitive to human vision. By contrast, our method efficiently eases the inconsistent deformation and preserves both the shapes and disparities of objects, leading to pleasing content-aware retargeting. For more comparisons and experimental results, please refer to the accompanying and supplemental documents.

In addition to the qualitative analysis, we also conduct a quantitative analysis. The quantitative analysis is conducted by utilizing the correlation coefficient, which represents the statistical relationship between two datasets. Given two sets of feature point pairs $X_i : \{p_k^L - p_k^R\}_{k=1}^{n_f}$ and $X_j : \{\tilde{p}_k^L - \tilde{p}_k^R\}_{k=1}^{n_f}$ in the original stereoscopic image and the retargeted image, the correlation coefficient is defined as $\text{Corr}(X_i, X_j) = \text{Cov}(X_i, X_j) / \sigma_{X_i} \sigma_{X_j}$, where $\text{Cov}(X_i, X_j)$ represents the covariance between X_i and X_j , σ_X is the standard deviation of dataset X , and n_f is the number of feature pairs. In addition, vertical shifting is estimated by measuring the vertical distances of feature points; that is, $Y_d = \sum_i^n |\tilde{p}_{i_y} - p_{i_y}| / n$, where n is the number of feature points. In this experiment, our method is compared with the related mesh-warping-based methods, including MW [3] and SMW [4]. Several feature points in the original stereoscopic image and retargeting results are selected manually. The correlation coefficients and the vertical distances of the feature points are calculated. The results are shown in Table I. This analysis indicates that the results of MW are comparable to that of SMW for images containing dense information because MW exhibits better performance in terms of shape preservation, and SMW has better performance in terms of disparity preservation and vertical shifting alleviation. Moreover, our results are closest to the original stereoscopic images (average $\text{Corr} = 0.981$) and

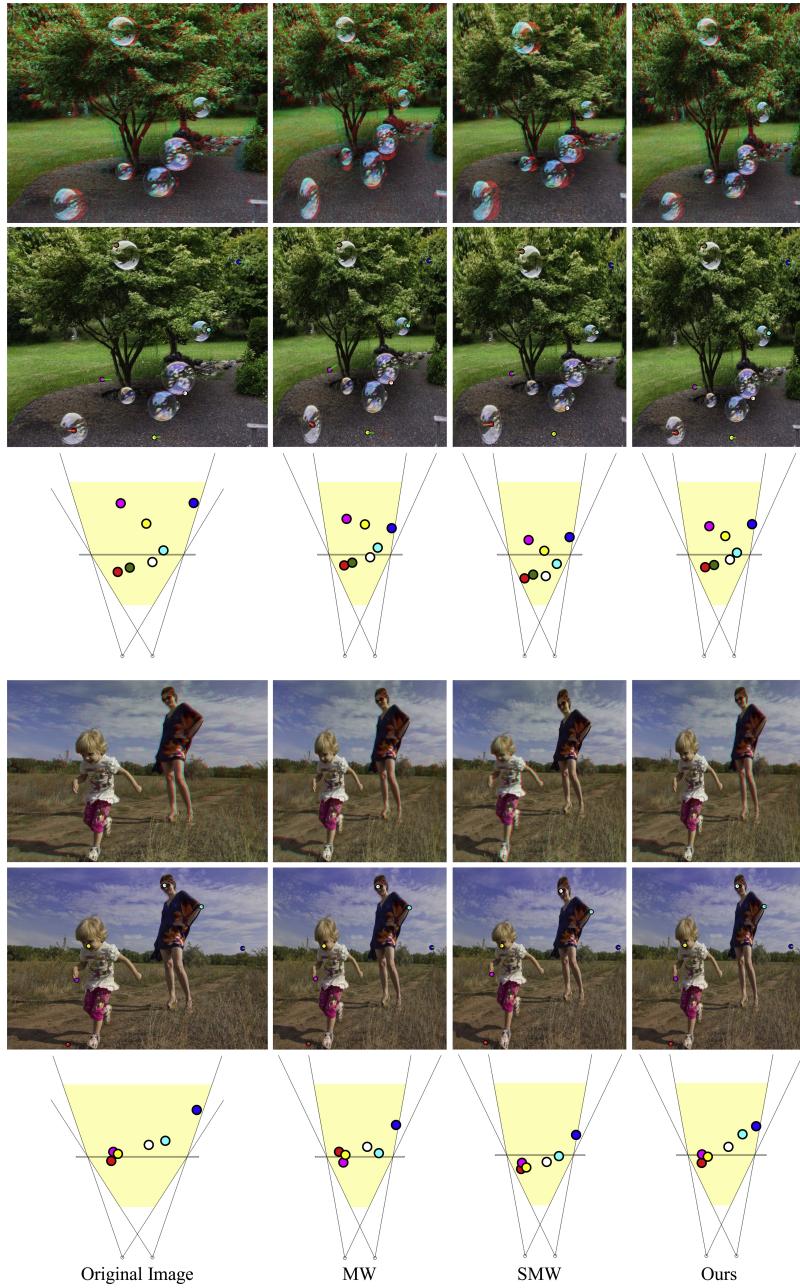


Fig. 12. Comparison of disparity preservation with MW [3], SMW [4], and the proposed method. First row: stereoscopic images. Second row: left images with the selected feature points (marked by colors). Third row: depth distribution of the selected feature points.

have good performance on the alleviation of vertical shifting (average $Y_d = 0.699$), compared with the results of SMW (average $\text{Corr} = 0.880$; average $Y_d = 1.283$) and MW (average $\text{Corr} = 0.931$; $Y_d = 1.891$). Basing on the qualitative and quantitative analyses, we conclude that our method is superior to the related methods in terms of content preservation for the images that contain evident foreground objects and structure lines.

In this paper, an object-aware retargeting method is proposed. Although object-aware retargeting is not a novel concept, it remains the goal of recent retargeting studies [5]. In [5], the input stereoscopic image is decomposed into several layers according to depth and color information. The content in each layer is resized separately through mesh warping.

This method can address the problem of inconsistent object deformation. However, separately warping layers may generate discontinuous artifacts. By contrast, our approach combines all the object saliency values in a significance map and continuous mesh warping is applied to the images based on this significance map. Therefore, similar to the general mesh-warping-based methods [2], [4], the proposed method does not suffer from discontinuous warping.

C. Retargeting With cropping

Researchers recently focus on stereoscopic image cropping [11]. The basic idea is to optimally and consistently crop the input image pair and then uniformly scale the cropped images to fit the desired aspect ratio. Combining cropping

TABLE I

QUANTITATIVE ANALYSIS. THE CORRELATION COEFFICIENTS $Corr$ AND VERTICAL DISTANCES Y_d OF SELECTED FEATURE POINTS IN ORIGINAL STEREOGRAPHIC IMAGES AND RETARGETED IMAGES GENERATED BY MW [3], SMW [4], AND PROPOSED METHOD ARE SHOWN IN THIS TABLE. AVG. AND STDEV. REPRESENT AVERAGE AND STANDARD DEVIATION OF $Corr$ AND Y_d

Method											Avg.	Stdev.
$Corr$	MW	0.992	0.810	0.843	0.834	0.990	0.997	0.994	0.990	0.918	0.937	0.931
	SMW	0.957	0.804	0.683	0.798	0.986	0.999	0.995	0.791	0.796	0.987	0.880
	Ours	0.999	0.994	0.923	0.942	0.999	0.999	0.997	0.999	0.974	0.997	0.981
Y_d	MW	1.186	8.944	1.052	0.765	1.740	0.701	0.949	0.669	0.487	2.415	1.891
	SMW	1.466	1.370	1.203	0.918	3.384	0.458	0.674	0.512	2.033	0.815	1.283
	Ours	0.271	0.923	1.098	0.706	0.590	0.595	0.328	0.319	0.713	1.448	0.699



Fig. 13. Combining cropping with the proposed method. Left: original stereoscopic image. Middle: retargeting using the proposed method with cropping. Right: retargeting using the proposed method only.

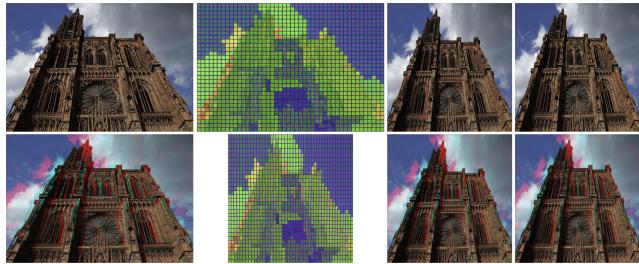


Fig. 14. Over-constrained retargeting. From left to right: original stereoscopic image (filled with a foreground object), grid mesh (top) and deformed grid mesh (bottom) visualized by the significance value, our result, and linear rescaling.

with mesh warping or seam carving in retargeting has been proven to improve the retargeting quality significantly [28], [29]. Integrating cropping into the proposed scheme is easy. Similar to [11], optimal cropping is performed first and the proposed warping approach is applied to further resize the cropped results to fit the desired aspect ratios. Fig. 13 shows the experiment that integrates cropping into the proposed method. The retargeting quality is improved, especially the aspect ratio of the foreground object.

D. Limitation

Similar to previous mesh-warping-based methods, our method tends to overconstrain or underconstrain the content when the image is filled with global structure lines (see Fig. 14) or when the segmented objects have similar

significance values. The results are similar to linear rescaling because all the structure lines or objects are deformed rigidly. One possible solution to this problem is to provide a user interface for users to specify the important content to preserve. Another alternative is to integrate linear rescaling in the retargeting scheme. When over- or underconstraining occurs, linear rescaling is performed instead of mesh warping.

V. CONCLUSION

This paper introduces a novel object-coherence warping method for content-aware stereoscopic image retargeting. The main idea is to utilize the object correspondences in the left and right images in retargeting. The object correspondences allow the generation of an object-based significance map and the consistent preservation of objects during warping. The input stereoscopic image pair is segmented into several object pairs and then each pair is assigned a significance value for consistent object warping. In the optimization of mesh warping, the disparity preservation constraints efficiently preserve the disparity values and force visually salient objects to undergo as-consistent-as-possible deformation, and the shape preservation constraints preserve the shapes of visually salient objects. Furthermore, the optimization with a significance map propagates the distortions to the homogeneous and low-significant regions. These properties efficiently ease the distortions in the disparities and shapes of objects caused by inconsistent warping. The comparisons, including the qualitative and quantitative analyses, show the superiority of the proposed method over the related methods in terms of consistency of object deformation. In the future, we plan to extend our retargeting scheme to stereo videos and solve this problem in 3-D volumetric space. In addition, we plan to integrate cropping or seam carving in our scheme to improve the retargeting quality.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions. The authors would also like to thank Prof. Y.-Y. Chuang and Dr. C.-H. Chang for providing their source codes.

REFERENCES

- [1] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," *ACM Trans. Graph.*, vol. 26, no. 3, Jul. 2007.
- [2] Y.-S. Wang, C.-L. Tai, O. Sorkine, and T.-Y. Lee, "Optimized scale-and-stretch for image resizing," *ACM Trans. Graph.*, vol. 27, no. 5, pp. 118:1–118:8, 2008.
- [3] S.-S. Lin, I.-C. Yeh, C.-H. Lin, and T.-Y. Lee, "Patch-based image warping for content-aware retargeting," *IEEE Trans. Multi.*, vol. 15, no. 2, pp. 359–368, Feb. 2013.
- [4] C.-H. Chang, C.-K. Liang, and Y.-Y. Chuang, "Content-aware display adaptation and interactive editing for stereoscopic images," *IEEE Trans. Multi.*, vol. 13, no. 4, pp. 589–601, Aug. 2011.
- [5] K.-Y. Lee, C.-D. Chung, and Y.-Y. Chuang, "Scene warping: Layer-based stereoscopic image resizing," in *Proc. IEEE Comput. Vision Pattern Recognit.*, Jun. 2012, pp. 49–56.
- [6] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir, "A comparative study of image retargeting," *ACM Trans. Graph.*, vol. 29, no. 6, pp. 160:1–160:10, 2010.
- [7] C.-K. Chiang, S.-F. Wang, Y.-L. Chen, and S.-H. Lai, "Fast JND-based video carving with GPU acceleration for real-time video retargeting," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 11, pp. 1588–1597, Nov. 2009.
- [8] Z. Yuan, T. Lu, Y. Huang, D. Wu, and H. Yu, "Addressing visual consistency in video retargeting: A refined homogeneous approach," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 6, pp. 890–903, Jun. 2012.
- [9] B. Yan, K. Sun, and L. Liu, "Matching-area-based seam carving for video retargeting," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 2, pp. 302–310, Feb. 2013.
- [10] S.-S. Lin, C.-H. Lin, I.-C. Yeh, S.-H. Chang, C.-K. Yeh, and T.-Y. Lee, "Content-aware video retargeting using object-preserving warping," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 10, pp. 1677–1686, Oct. 2013.
- [11] Y. Niu, F. Liu, W. C. Feng, and H. Jin, "Aesthetics-based stereoscopic photo cropping for heterogeneous displays," *IEEE Trans. Multi.*, vol. 14, no. 3, pp. 783–796, Jun. 2012.
- [12] K. Utsugi, T. Shibahara, T. Koike, K. Takahashi, and T. Naemura, "Seam carving for stereo images," in *Proc. 3DTV-CON*, 2010, pp. 1–4.
- [13] T. Basha, Y. Moses, and S. Avidan, "Geometrically consistent stereo seam carving," in *Proc. IEEE Int. Conf. Comput. Vision*, Nov. 2011, pp. 1816–1823.
- [14] Y. Niu, W.-C. Feng, and F. Liu, "Enabling warping on stereoscopic images," *ACM Trans. Graph.*, vol. 31, no. 6, pp. 183:1–183:7, 2012.
- [15] C. Wang and A. A. Sawchuk, "Disparity manipulation for stereo images and video," in *Proc. SPIE*, vol. 6803, no. 68031E, 2008.
- [16] M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic, and M. Gross, "Nonlinear disparity mapping for stereoscopic 3D," *ACM Trans. Graph.*, vol. 29, no. 4, pp. 75:1–75:10, Jul. 2010.
- [17] S.-J. Luo, I.-C. Shen, B.-Y. Chen, W.-H. Cheng, and Y.-Y. Chuang, "Perspective-aware warping for seamless stereoscopic image cloning," *ACM Trans. Graph.*, vol. 31, no. 6, pp. 182:1–182:8, 2012.
- [18] F. Liu, Y. Niu, and H. Jin, "Casual stereoscopic photo authoring," *IEEE Trans. Multi.*, vol. 15, no. 1, pp. 129–140, Jan. 2013.
- [19] Y. Niu, F. Liu, X. Li, and M. Gleicher, "Warp propagation for video resizing," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2010, pp. 537–544.
- [20] G.-X. Zhang, M.-M. Cheng, S.-M. Hu, and R. R. Martin, "A shape-preserving approach to image resizing," *Comput. Graph. Forum*, vol. 28, no. 7, pp. 1897–1906, Oct. 2009.
- [21] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 1915–1926, Oct. 2012.
- [22] J.-S. Kim, J.-Y. Sim, and C.-S. Kim, "Multiscale saliency detection using random walk with restart," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 2, pp. 198–210, Feb. 2014.
- [23] J. Li, Y. Tian, T. Huang, and W. Gao, "Multi-task rank learning for visual saliency estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 5, pp. 623–636, May 2011.
- [24] W. Kim, C. Jung, and C. Kim, "Spatiotemporal saliency detection and its applications in static and dynamic scenes," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 4, pp. 446–456, Apr. 2011.
- [25] M. Grundmann, V. Kwatra, M. Han, and I. Essa, "Efficient hierarchical graph-based video segmentation," in *Proc. IEEE Comput. Vision Pattern Recognit.*, Jun. 2010, pp. 2141–2148.
- [26] CMU/VASC Image Database: Stereo Image [Online]. Available: <http://vasc.ri.cmu.edu/idb/html/stereo/>
- [27] NVIDIA Image Database: Stereo Image [Online]. Available: <http://photos.3dvisionlive.com/NVIDIA/>
- [28] Y.-S. Wang, H.-C. Lin, O. Sorkine, and T.-Y. Lee, "Motion-based video retargeting with optimized crop-and-warp," *ACM Trans. Graph.*, vol. 29, no. 4, pp. 90:1–90:9, 2010.
- [29] M. Rubinstein, A. Shamir, and S. Avidan, "Multi-operator media retargeting," *ACM Trans. Graph.*, vol. 28, no. 3, pp. 23:1–23:11, 2009.



Shih-Syun Lin

Shih-Syun Lin received the B.S. degree in applied mathematics from Providence University, Taichung, Taiwan, in 2007 and the M.S. degree from the Graduate Institute of Educational Measurement and Statistics, National Taichung University, Taichung, in 2010. He is currently pursuing the Ph.D. degree at the Department of Computer Science and Information Engineering, National Cheng-Kung University, Tainan, Taiwan.

His research interests include video retargeting, mesh deformation, pattern recognition, and computer graphics.



Chao-Hung Lin

(M'13) received the M.S. and Ph.D. degrees in computer science and information engineering from National Cheng-Kung University, Tainan, Taiwan, in 1998 and 2004, respectively.

Since 2006 he has been a Faculty Member with the Department of Geomatics, National Cheng-Kung University, where he is currently an Associate Professor. He leads the Digital Geometry Processing Laboratory (<http://dgl.geomatics.ncku.edu.tw>) and co-leads the Computer Graphics Laboratory, National Cheng-Kung University (<http://graphics.csie.ncku.edu.tw>). His research interests include remote sensing, point cloud processing, digital map generation, information visualization, and computer graphics.

Dr. Lin served as an Editorial Board Member of *International Journal of Computer Science and Artificial Intelligence*.



Shu-Huai Chang

Shu-Huai Chang received the B.S. degree in computer science and information engineering from the National Taiwan University of Science and Technology, Taipei, Taiwan, in 2011 and the M.S. degree from the Department of Computer Science and Information Engineering, National Cheng-Kung University, Tainan, Taiwan, in 2013.

He is with International Games System Co., Ltd., Taipei, Taiwan. His research interests include computer graphics.



Tong-Yee Lee

(SM'10) received the Ph.D. degree in computer engineering from Washington State University, Pullman, WA, USA, in 1995.

He is a Distinguished Professor with the Department of Computer Science and Information Engineering, National Cheng-Kung University, Tainan, Taiwan. He leads the Computer Graphics Group, Visual System Laboratory, National Cheng-Kung University (<http://graphics.csie.ncku.edu.tw/>). His research interests include computer graphics, nonphotorealistic rendering, medical visualization, virtual reality, and media resizing.