Depth-Preserving Warping for Stereo Image Retargeting

Bing Li, *Student Member, IEEE*, Ling-Yu Duan, *Member, IEEE*, Chia-Wen Lin, *Senior Member, IEEE*, Tiejun Huang, *Senior Member, IEEE*, and Wen Gao, *Fellow, IEEE*

Abstract—**The popularity of stereo images and various display devices poses the need of stereo image retargeting techniques. Existing warping-based retargeting methods can well preserve the shape of salient objects in a retargeted stereo image pair. Nevertheless, these methods often incur depth distortion, since they attempt to preserve depth by maintaining the disparity of a set of sparse correspondences, rather than directly controlling the warping. In this paper, by considering how to directly control the warping functions, we propose a warping-based stereo image retargeting approach that can simultaneously preserve the shape of salient objects and the depth of 3D scenes. We first characterize the depth distortion in terms of warping functions to investigate the impact of a warping function on depth distortion. Based on the depth distortion model, we then exploit binocular visual characteristics of stereo images to derive region-based depth-preserving constraints which directly control the warping functions so as to faithfully preserve the depth of 3D scenes. Third, with the region-based depth-preserving constraints, we present a novel warping-based stereo image retargeting framework. Since the depth-preserving constraints are derived regardless of shape preservation, we relax the depth-preserving constraints to fulfill a tradeoff between shape preservation and depth preservation. Finally, we propose a quad-based implementation of the proposed framework. The results demonstrate the efficacy of our method in both depth and shape preservation for stereo image retargeting.**

Index Terms—**Stereo image retargeting, image warping, optimization, depth preservation.**

I. INTRODUCTION

S TEREO images are becoming more and more popular in our daily media consumption as they offer rich and joyful

real-world viewing experience. The popularity of binocular cameras brings about the explosive growth of stereo images. Moreover, with the development of display technologies, stereo images can be viewed on a variety of devices with different display resolutions and aspect ratios, such as cinema screens, televisions, computers, and mobile phones. To fit the stereo image contents into the display of various resolutions and aspect ratios calls for efficient stereo image retargeting methods.

2D Image Retargeting: Recently, content-aware image retargeting schemes have been proposed to adapt 2D images to heterogeneous display devices with various resolutions and aspect ratios. Following [1], we classify the content-aware retargeting schemes into two categories: discrete approaches and continuous approaches. Discrete approaches, mainly including cropping [2], [3] and seam carving [4], [5], treat an image as a discrete entity (i.e., pixels) and retarget the image by iteratively removing or inserting pixels. In contrast, continuous methods [6]–[11] retarget an image via continuous-domain warping. These warping-based methods first divide an image into regions (e.g., a mesh partition). Then the image regions are non-uniformly warped in a continuous manner, where the local warping functions of those regions with high importance (e.g., salient regions) are constrained to preserve the shapes of such important regions. Thanks to the continuous warping manner, these methods do not introduce discontinuity artifacts. Nevertheless, since such kind of methods (e.g., in [7]) allow the shape of a retargeted quad to be arbitrary quadrilateral, some quads would be deformed significantly, causing noticeable distortion on a structural object. To assess the visual qualities of retargeted images, effective objective quality metrics that agree with subjective assessments reasonably well have been proposed [12], [13].

Stereo Image Retargeting: Compared with 2D images, stereo images contain an additional depth dimension. With the additional depth dimension, stereo image retargeting aims at simultaneously maintaining the depth of scenes and preserving shapes of salient objects to provide satisfactory 3D viewing experience. Although many content-aware 2D image retargeting approaches can well preserve the shapes of salient objects via non-uniform image resizing, applying these approaches to independently resize the left and right images of a stereo pair[1] would incur depth distortion. Specifically, in a perceived 3D scene, the depth of a 3D point is related to the disparity (i.e., the location difference) between

---

[1]We focus on stereo images consisting of two images captured from two perspectives.
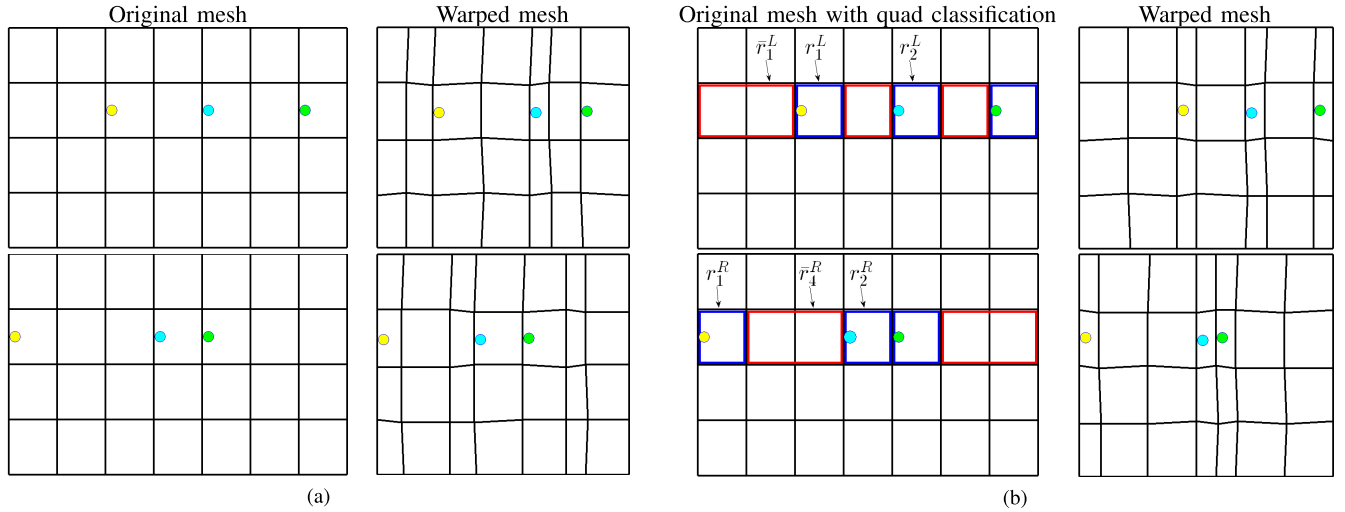
Fig. 1. Illustration of the impact of different depth-preserving constraints on depth preservation. From top to bottom: the mesh of the left image and the mesh of the right image, where every two dots with the same color indicate a correspondence pair. (a) An example of sparse correspondence-based depth-preserving constraints that enforce the disparity of one correspondence pair (indicated by blue dots) to remain unchanged. The disparity of yellow correspondence and green ones, however, are not maintained. (b) After grouping the quads into paired regions r (indicated by red blocks) and non-paired regions $\bar{r}$ (indicated by blue blocks), respectively, the proposed depth-preserving constraints are imposed to control the warping functions to preserve the width of non-paired regions $\bar{r}$, and consistently resize paired regions r and their corresponding regions. The disparity of all correspondences are maintained in warped meshes.

the pair of corresponding 2D points (i.e., the correspondences) in the left and right images. Since the left and right images are slightly different, independently resizing these images in a non-uniform manner would change the relative positions of correspondences in the retargeted pair, causing inconsistent disparity/depth with the original value.

Recently, there have been a few works studying how to retarget a stereo image pair. To preserve the shapes of salient objects while maintaining the depths of 3D scenes, most existing stereo image retargeting methods extend content-aware 2D image retargeting approaches by introducing additional depth-preserving constraints. For example, Niu *et al.* [14] proposed to extend cropping-based retargeting to stereo image retargeting by preserving the aesthetic value of an input photo. However, the method may still partially or completely discard salient objects or useful context information due to cropping. Moreover, when an image is significantly cropped, those objects with negative disparity in the retargeted image may hit the side frame of display, leading to window violation. The methods proposed in [15] and [16] extend seam carving [17] to stereo image pairs by iteratively removing a pair of seams, one per stereo image pair. The removal of each seam pair is subject to the constraint that pixels of a removed seam in one image of a stereo pair shall correspond to those of the seam removed from the other image. Owing to the discrete manner, such seam removal does not change the relative position of remaining correspondence pairs, thereby maintaining the depth well. Nevertheless, these methods often introduce discontinuity artifacts in a retargeted image pair, due to its discrete nature.

In contrast, the methods proposed in [18]–[20] extend warping-based image retargeting to stereo image retargeting by imposing depth-preserving constraints over a sparse set of correspondences in the stereo pair. As illustrated in Fig. 1a, these methods divide each image of a stereo pair into a quad

mesh, and then warp each quad by a local warping function for the quad, such that the image is non-uniformly resized. To achieve depth preservation, these methods resort to depth-preserving constraints, which are inspired by depth remapping techniques (see [21]–[24]). In particular, the depth remapping techniques proposed in [21] and [22] extend image warping to depth remapping by imposing depth-editing constraints which enforce the disparity values of a sparse set of correspondences to be identical to the target disparity. Similarly, the warping-based stereo image retargeting methods [18]–[20] propose to preserve depth by keeping the disparity values of a sparse set of correspondences consistent with their original values as much as possible. They, however, often fail to faithfully maintain depth consistency of 3D scenes, since their depth-preserving constraints work on a set of sparse correspondences rather than directly controlling the warping functions, making them ineffective for depth preservation.

More specifically, the constraints for preserving the disparity of a sparse set of correspondences can only indirectly influence the local warping functions for a few quads, which, however, may not faithfully maintain the disparity of remaining correspondences. For example, as shown in Fig. 1a, the constraints are built to maintain the disparity of the blue correspondence pair. Since its disparity depends on the first five warped quads in the second row of the left image and the first three warped quads in the second row of the right image, the constraints indirectly influence the warping functions of these quads. Yet, such influence may not effectively maintain the disparity of the remaining correspondence pairs. For example, for the yellow correspondence pair, although all its associated warping functions (i.e., the warping functions of the first two quads in the second row) are influenced by the constraints, its disparity are changed in warped pair. Similarly, for the green correspondence pair, some of its associated warping functions

are even not influenced by the constraints, leading to disparity inconsistency.

In this paper, to preserve the scene depth of warped stereo images, we address the issue: *how to explicitly control the warping functions to preserve the depth for quad-based warping*, which, to the best of our knowledge, has not yet been addressed before. From this new perspective, we propose a warping-based stereo image retargeting approach, which achieves high-quality retargeting in both object shape preservation and scene depth preservation. In particular, we characterize the depth/disparity distortion energy in terms of local warping functions, such that we can estimate the depth distortion in a retargeted stereo pair. Based on the depth distortion model, we can then figure out the conditions for eliminating the depth distortion, so as to derive effective depth-preserving constraints for two different kinds of regions, respectively, as shown in Fig. 1b. The region-wise depth-preserving constraints are used to directly control the local warping functions of the quads of each region class to achieve depth preservation. By adequately integrating the proposed depth-preserving constraints with shape-preserving constraints, we propose a novel warping-based stereo image retargeting framework, which can achieve a good tradeoff between depth preservation and shape preservation.

The main contributions of this paper are three-fold: (i) We model and analyze the depth/disparity distortion in terms of warping functions to study the relationship between a warping function and the resulting depth distortion, which was ignored in existing warping-based retargeting methods; (ii) According to the depth distortion model, we propose depth-preserving constraints which directly control local warping functions to achieve depth preservation, without sacrificing shape preservation performance; (iii) We propose a practical quad-based retargeting platform to realize the proposed framework with reasonable computational complexity.

The rest of this paper is organized as follows. Sec. II. formulates the problem of the warping-based stereo image retargeting. Sec. III models the depth distortion in terms of warping functions, and then proposes the depth-preserving constraints accordingly. The proposed warping-based stereo image retargeting framework is detailed in Sec. IV. Sec. V describes the quad-based implementation of the proposed warping-based retargeting framework. Sec. VI reports the experimental results. Finally, we conclude the paper in Sec. VII.

II. PROBLEM FORMULATION

In this paper, we propose a quad-based warping framework for stereo image retargeting. To formulate the problem of quad-based warping for stereo image retargeting, we first introduce the representation of quad-based warping. The quad-based warping first partitions an image by a quad mesh, and then individually warps each quad using a local warping function, such that the image is non-uniformly resized. To obtain the local warping functions of quads, generally most existing warping-based retargeting methods represent the local

TABLE I
NOTATION

Symbols	Meanings
I^z	left image of a stereo pair if $z = L$; otherwise, right image
D^s, D^d	shape distortion energy and depth distortion energy
q_k^z	the k-th quad in I^z
$\tilde{\mathbf{v}}_k^z$	vertices (also the local warping function) of q_k^z in the warped mesh
$d_i, \tilde{d}_i$	the horizontal disparity values of $\mathbf{c}_i$ in the original pair and retargeted pair
$\mathbf{c}_i = \{c_i^L, c_i^R\}$	the i-th pair of corresponding pixels between I^L and I^R
Δd_i	depth distortion energy of $\mathbf{c}_i$
$x_i^z, \tilde{x}_i^z$	x-coordinates of c_i^z in the original image and retargeted image
$e_k^z = \Phi(\tilde{\mathbf{v}}_k^z)$	the width of q_k^z on horizontal line $\mathbf{E}_i$ in the warped image
w_k^z, h_k^z	width and height of quad q_k^z in the original mesh
$\bar{r}^z, r^z$	non-paired region, paired region
$\bar{\Upsilon}, \Upsilon$	the set of non-paired regions, the set of paired regions
q_{ij}^z	quad located at row i and column j in I^z
W', H'	width and height of a retargeted stereo pair
s_{ij}^z	importance of quad g_{ij}^z
w_j^z, h_i^z	the width of the jth quad row and the height of the ith column in the original quad mesh
$\tilde{w}_j^z \, \tilde{h}_i^z$	the width of the jth quad row and the height of the ith column in the warped quad mesh
P_{ijk}	alignment probability between quad q_{ij}^z and $q_{ik}^{z'}$
D_p	energy function of paired regions

warping functions in terms of the vertices of the warped mesh. Specifically, let I^L and I^R denote the left and right images of a stereo image pair, respectively, q_k^z denotes the k-th quad q_k^z in I^z, $z \in \{L, R\}$. For q_k^z, its local warping function is represented by $\tilde{\mathbf{v}}_k^z$, which consists of four vertices of q_k^z in the warped mesh. Accordingly, the warping functions of the stereo image pair can be represented by the set of all quads' local warping functions, i.e., $\tilde{\mathbf{V}} = \{\mathbf{v}_k^z\}$. Readers are referred to Table I for the list of notations in this paper.

The warping-based stereo image retargeting aims to find the optimal warping functions for individual quads, which can simultaneously minimize the shape distortions of salient objects and keep the scene depth in the retargeted stereo pair faithful to the original one. To this end, we formulate the stereo image retargeting problem as the following optimization problem:

$$\tilde{\mathbf{V}}^* = \arg\min_{\tilde{\mathbf{V}}} D^s \quad s.t \quad D^d = 0 \tag{1}$$

where D^s denotes the shape distortion energy, and D^d the depth distortion energy.

The above problem can be factored into two subproblems: shape preservation of salient objects ($\min D^s$) and depth preservation ($D^d = 0$) for 3D scenes. We propose to solve the two subproblems separately to reduce problem complexity. Furthermore, since the shape preservation subproblem has been successfully addressed using shape-preserving constraints in existing quad-based retargeting methods, we focus on finding effective depth-preserving constraints to solve the depth preservation subproblem. We then propose a method to properly integrate the proposed depth-preserving constraints with existing shape-preserving constraints to solve the stereo image retargeting problem.

To the best of our knowledge, the joint depth and shape preservation problem for stereo image retargeting has not been well addressed. Existing methods [18]–[20] solve the depth preservation subproblem by resorting to depth-preserving

constraints which attempt to preserve as much as possible the disparity values of a sparse set of correspondences in the retargeted stereo pair. Such depth-preserving constraints, however, may fail to faithfully maintain the disparity of the remaining correspondences of the retargeted stereo pair, since these constraints do not have direct control on the warping functions.

Different from the existing methods, our approach aims to figure out how to explicitly constrain the warping functions, such that the depth is faithfully maintained. To this end, we first model D^d in terms of the warping functions $\tilde{V}$ for the stereo image pair (see Sec. III-A for details), so as to reveal the relationship between local warping functions and the depth distortion. With such model $D^d(\tilde{V})$, we derive depth-preserving constraints which explicitly constrain the warping functions (see Sec. III-B), by exploiting the binocular visual characteristics of regions. More specifically, according to the binocular visual characteristics of regions, we divide image regions into two classes, and then derive the depth-preserving constraints for the two region classes by finding the conditions that the local warping functions of each region class should satisfy for keeping $D^d(\tilde{V}) = 0$.

With the derived depth-preserving constraints, we then jointly consider shape preservation and depth preservation. Since depth-preserving constraints are separately derived without taking into account shape preservation, they may contradict shape-preserving constraints. Thus, we propose a set of relaxed disparity-preserving constraints, based on region attributes as will be elaborated later. By integrating the relaxed depth-preserving constraints with shape-preserving constraints, we propose a region-wise warping-based optimization framework for stereo image pair retargeting in Sec. IV. We further propose a quad-based warping platform to efficiently implement the propose framework in Sec. V.

III. REGION-WISE DEPTH-PRESERVING WARPING CONSTRAINTS

In this section, we aim to figure out how to faithfully preserve the depths of 3D scenes in a stereo image pair by explicitly constraining the warping functions. To this end, we first model the depth distortion energy as a function of warping functions such that we can analyze the effect of local warping on depth distortion. We then derive effective depth-preserving constraints for local warping functions by exploiting the binocular visual characteristics of different region classes in a stereo images pair.

A. Depth Distortion Energy

To characterize depth distortion energy, we represent depth in terms of horizontal disparity, since the magnitude of depth is inversely proportional to the associated horizontal disparity value given the same viewing configuration. Furthermore, since the vertical disparity does not impact the depth magnitude and non-zero vertical disparity can be easily eliminated, without loss of generality we assume there is no vertical disparity between each pair of correspondences in the original stereo pair and the retargeted version.
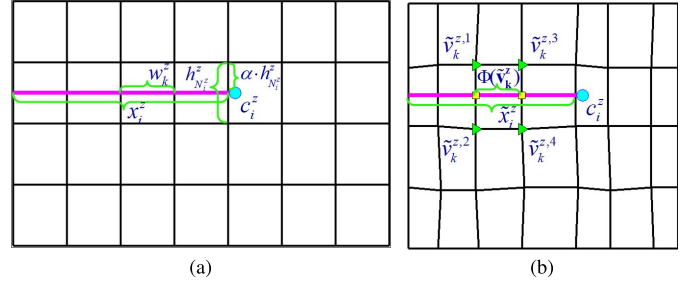

Fig. 2. Example of representing x-coordinates of c_i^z by the local warping functions. The blue dots indicate the correspondence c_i^z, the green triangles in the warped mesh represent the vertices of quad q_k^z, the pink lines represent the horizontal line segment $\mathbf{E}_i^z$ crossing c_i^z, which intersects q_k^z at o_k and o_{k+1} (indicated by yellow squares). (a) Original mesh. (b) Warped mesh.

Let $\mathbf{C} = \{\mathbf{c}_i\}$ denote the set of all correspondence pairs in the stereo pair, $\mathbf{c}_i = \{c_i^L, c_i^R\}$ the i-th pair of the corresponding pixels between the left and right images I^L and I^R. We define the total depth distortion energy D^d of a stereo pair as the sum of depth distortion energy of all correspondence pairs as follows[2]:

$$D^d = \sum_{\mathbf{c}_i \in \mathbf{C}} \Delta d_i = \sum_{\mathbf{c}_i \in \mathbf{C}} |d_i - \tilde{d}_i|$$
$$= \sum_{\mathbf{c}_i \in \mathbf{C}} |(x_i^R - x_i^L) - (\tilde{x}_i^R - \tilde{x}_i^L)| \quad (2)$$

where Δd_i, representing the depth distortion energy of $\mathbf{c}_i$, is defined as the absolute difference between $\mathbf{c}_i$'s horizontal disparity in the retargeted stereo pair and that in the original pair. $d_i = x_i^R - x_i^L$ and $\tilde{d}_i = \tilde{x}_i^R - \tilde{x}_i^L$ respectively denote the horizontal disparity values of $\mathbf{c}_i$ in the original pair and the retargeted pair, x_i^z and $\tilde{x}_i^z$ $z \in \{L, R\}$ indicate the x-coordinates of c_i^z in the original pair and the retargeted pair, respectively.

The above formulation, however, does not explicitly incorporate the effects of each mesh unit's warping function on the depth. Instead, we reformulate the depth distortion energy in terms of local warping functions. First, to characterize the effect of a local warping on the disparity values of associated correspondences, we assume a fine-granular quad mesh, where a quad contains only one single pixel located at the left edge in the quad. Then, we represent x_i^z and $\tilde{x}_i^z$ in terms of quads (see Fig. 2). In particular, as shown in Fig 2, since x_i^z is equal to the length of horizontal line segment $\mathbf{E}_i^z$ which is bounded by c_i^z and the left edge of I^z, we discretize $\mathbf{E}_i^z$ by quads to represent x_i^z in terms of quads. That is, let $q_1^z \ldots q_{N_i^z}^z$ represent the quads that are crossed by $\mathbf{E}_i^z$ and lie at the left side of c_i^z. We represent $x_i^z = \sum_{k=1}^{N_i^z} w_k^z$, where w_k^z is the width of quad q_k^z, and similarly, $\tilde{x}_i^z = \sum_{k=1}^{N_i^z} e_k^z$, where e_k^z is the width of q_k^z on horizontal line $\mathbf{E}_i$. We propose to further represent e_k^z in terms of local warping functions, i.e., $e_k^z = \Phi(\tilde{\mathbf{v}}_k^z)$ (see Appendix A for more details). As a result, the depth distortion energy in (2) can be rephrased as follows:

$$\Delta d_i = \left| \sum_{k=1} w_k^R - \sum_{k=1} w_k^L - \left(\sum_{k=1} \Phi(\tilde{\mathbf{v}}_k^R) - \sum_{k=1} \Phi(\tilde{\mathbf{v}}_k^L) \right) \right| \quad (3)$$

[2]The position of a correspondence can be non-integer, as an image is considered as a sampled continuous function.
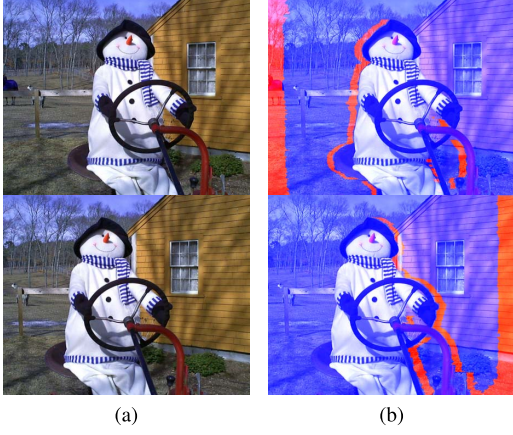
Fig. 3. Region classification results on *Snowman*. From top to bottom: left images and right images, where paired regions and non-paired regions are colored red and blue, respectively. (a) Original images. (b) Classified images.

where

$$\Phi(\tilde{\mathbf{v}}_k^z) = (1-\alpha)\tilde{v}_{k,x}^{z,3} + \alpha \cdot \tilde{v}_{k,x}^{z,4} - ((1-\alpha)\tilde{v}_{k,x}^{z,1} + \alpha \cdot \tilde{v}_{k,x}^{z,2})$$
(4)

where $\tilde{\mathbf{v}}_k^z$ is composed of $\tilde{v}_k^{z,1}, \tilde{v}_k^{z,2}, \tilde{v}_k^{z,3}, \tilde{v}_k^{z,4}$ that denote the left-top, left-bottom, right-top, and right-bottom vertices of q_k^z's retargeted version (see Fig. 2), respectively, $\tilde{v}_{k,x}^{z,j}$ is the x-coordinate of $\tilde{v}_{k,x}^{z,j}$, and α denotes the barycentric coordinate, as elaborated in Appendix A.

To derive the ideal depth-preserving warping functions that yield $D^d = 0$, we set the depth distortion energy of each correspondence to be zero individually, that is

$$\Delta d_1 = \Delta d_2 = \ldots = \Delta d_{N_c} = 0,$$
(5)

where N_c is the total number of correspondences.

Note, we do not need to precisely solve the ideal depth-preserving warping functions, since we just need to figure out the conditions that the local warping functions need to satisfy for depth preservation. Therefore, we reformulate (5) by exploiting the binocular visual characteristics of the stereo images, and then derive the depth-preserving constraints for the local warping fuctions.

B. Region-Based Depth Preservation

As revealed in the literature [25]–[27], different image regions exhibit different binocular visual characteristics due to the different viewing angles of cameras, thereby leading to different levels of impact on the perceived depth, given a viewing configuration. We argue that quads, which are attributed to different region classes with different characteristics, should be imposed with different depth-preserving constraints. Hence, as illustrated in Fig. 3, we propose to group image regions into the following two classes:

Paired Region: A paired region in one image of a stereo pair can find a corresponding region in the other image, both are the projections of a 3D object onto the left and right images. Hence, a correspondence pair consists of a paired region and its correspondence.

Non-Paired Region: A non-paired region in one image cannot find any correspondence in the other image. Such non-paired region comes from a 3D object which is projected onto only one image plane of a stereo image pair.

Accordingly, we refer to a pixel in one image as a *paired pixel* if it belongs to a paired region; otherwise a *non-paired pixel*.

Based on the above region classification, we reformulate the depth distortion energy. We first group quads into the two region classes. First, if the pixel in quad q_k^z is a paired pixel, q_k^z is labeled as a paired quad; otherwise, it is labeled as a non-paired quad. Subsequently, as shown in Fig. 1b, we treat each paired-quad as a paired region r^z which only contains a single paired-quad. We then group horizontally consecutive non-paired quads q_k^z into a non-paired region $\bar{r}^z$. Specifically, given correspondence pair $\mathbf{c}_i$, its quads $q_1^L \ldots q_{N_i^L}^L$ in the left image are grouped into paired regions $r_1^L, \ldots r_{n_i}^L$ and non-paired regions $\bar{r}_1^L, \ldots \bar{r}_{m_i}^L$. Similarly, quads $q_1^R \ldots q_{N_i^R}^R$ in the right image are grouped into non-paired regions $\bar{r}_{m_i+1}^R, \ldots \bar{r}_{n_i'}^R$ and paired regions $r_1^R, \ldots r_{n_i}^R$, where r_k^R denotes the corresponding quad of r_k^L.

Based on the above region partition, we reformulate (3) as follows:

$$\Delta d_i = \Bigg| \sum_{j=1}^{n_i'} \rho_j^z \Big(\sum_{q_k^z \in r_j^z} \Phi(\tilde{\mathbf{v}}_k^z) - \sum_{q_k^z \in r_j^z} w_k^z \Big)$$

$$+ \sum_{j=1}^{n_i} \sum_{q_k^z \in r_j^z} \big(\Phi(\tilde{\mathbf{v}}_k^R) - \Phi(\tilde{\mathbf{v}}_k^L) - (w_k^R - w_k^L) \big) \Bigg|$$
(6)

where $\rho^z = -1$ if $z = L$, and $\rho^z = 1$ if $z = R$.

We then plug Δd_i in (6) into (5), and show by mathematical induction (see Appendix B for detailed derivation) that the depth can be preserved, should the following constraints be imposed on the local warping functions:

$$\begin{cases} \sum_{q_k^z \in \bar{r}_j^z} \Phi(\tilde{\mathbf{v}}_k^z) = \sum_{q_k^z \in \bar{r}_j^z} w_k^z, & \forall \bar{r}_j^z \in \bar{\Upsilon} \\ \Phi(\tilde{\mathbf{v}}_k^L) - w_k^L = \Phi(\tilde{\mathbf{v}}_k^R) - w_k^R, & \forall q_k^z \in r_j^z, \forall r_j^z \in \Upsilon \end{cases}$$
(7)

where $\bar{\Upsilon}$ denotes the set of non-paired regions, and Υ the set of paired regions. For example, as shown in Fig. 1b, intuitively the following constraints can be used: 1) for each non-paired region, constraining the local warping function to preserve the region's width; 2) for each paired region, constraining the local warping function to consistently resize the region and its correspondence along the horizontal direction.

IV. WARPING-BASED FRAMEWORK FOR DEPTH-PRESERVING STEREO IMAGE RETARGETING

To provide satisfactory viewing experience, we need to preserve both the shapes of salient objects and the depths of scenes by adequately integrating depth-preserving constraints with shape-preserving constraints. However, the depth-preserving constraints derived in (7) may contradict shape-preserving constraints, since they are derived by
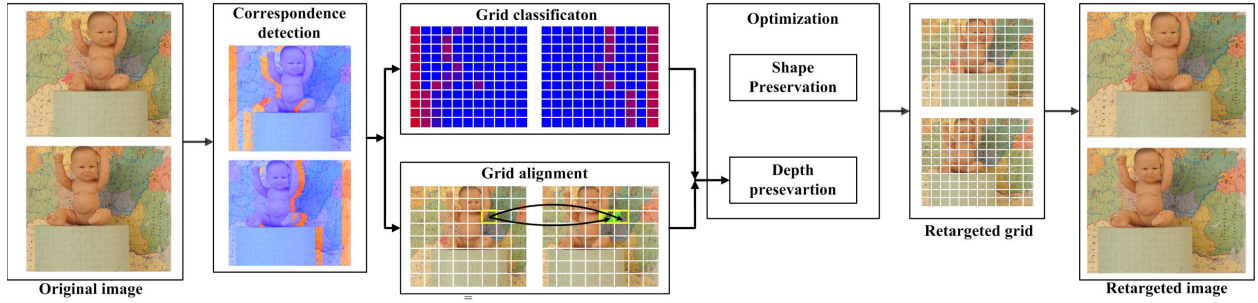
Fig. 4.   Proposed quad-based warping for stereo image retargeting.

enforcing depth distortion energy $D^d = 0$ without considering the shape preservation requirements. On one hand, the depth-preserving constraints may lead to the shape distortions of salient objects. For example, the depth-preserving constraints do not allow non-paired regions to be horizontally stretched or shrunk largely, even if the non-paired regions are often visually unimportant. As a result, the space saved from resizing such non-salient regions may be too few to allow for shape preservation for salient objects, especially when non-paired regions occupy a large area in an image pair. On the other hand, should strong constraints be imposed to preserve the shapes of salient objects, the shape-preserving constraints and depth-preserving constraints may render conflict warping behaviors (e.g., one requests stretching while the other requests shrinking). Consequently, there may be no feasible solution to meet such conflicting constraints.

To reach a good tradeoff between shape preservation and depth preservation, we propose to relax the hard depth-preserving constraints in (7) to allow nonzero depth distortion energy D^d for paired regions. That is, for $q_k^z \in \Upsilon$, the warping mismatch $|\Phi(\tilde{\mathbf{v}}_k^L) - w_k^L - (\Phi(\tilde{\mathbf{v}}_k^R) - w_k^R)|$ can be greater than 0. To do so, we define the following energy function D^p to measure the depth distortion in terms of the warping mismatch:

$$D^p = \sum_{r_j^z \in \Upsilon} \sum_{q_k^z \in r_j^z} \varpi_k (\Phi(\tilde{\mathbf{v}}_k^L) - w_k^L - (\Phi(\tilde{\mathbf{v}}_k^R) - w_k^R))^2 \quad (8)$$

where ϖ_k is the weight used to adjust the value of $|\Phi(\tilde{\mathbf{v}}_k^L) - w_k^L - (\Phi(\tilde{\mathbf{v}}_k^R) - w_k^R)|$.

On top of the shape-preserving quad-based warping schemes in [7]–[11] and [28], by integrating shape-preserving constraints with the relaxed depth-preserving constraints, we propose the following region-based optimization for warping-based stereo image retargeting:

$$\min \quad D^s + \lambda \cdot D^p$$
$$s.t. \quad \sum_{q_k^z \in \bar{r}_j^z} \Phi(\tilde{\mathbf{v}}_k^z) = \sum_{q_k^z \in \bar{r}_j^z} w_k^z, \quad \forall \bar{r}_j^z \in \bar{\Upsilon} \quad (9)$$

where D^s represents the shape-preserving constraints, and λ stands for the weight for depth distortion energy D^d.

Note that the relaxation of depth preservation may cause inconsistent warping of some paired regions and their correspondences, which may further accumulate quad by quad,

thereby yielding significant disparity distortion eventually. In the following section we shall elaborate our quad-based implementation of the depth-preserving stereo image retargeting which can effectively address such error accumulation problem.

V. Quad-Based Implementation of Proposed Retargeting Framework

To achieve high computational efficiency, the quad-based methods proposed in [10] and [11] present a rectangular transformation based warping, which restricts the shapes of retargeted quads to be rectangular. With such simplification, these methods [10], [11] can well preserve the shapes of salient objects, while achieving high computational efficiency as the simplification significantly reduces the number of variables compared to [7]. Based on the quad-based schemes, we propose a low-complexity quad-based implementation of the proposed warping-based retargeting framework as shown in Fig. 4. In the proposed system, each image of the input stereo pair is partitioned by a uniform quad mesh. Our method consists of the following steps. First, we preprocess the stereo pair for building depth-preserving constraints. That is, we establish the correspondences in the input stereo pair based on stereo matching [29]. According to the stereo matching results, we group the quads in the stereo image pair into non-paired regions and paired regions. For each paired region in one image of the image pair, we align it with those regions containing its corresponding pixels in the other image. Second, we formulate a constrained optimization model over quads. In the model, depth-preserving constraints are built to preserve the width of each non-paired region as well as constrain the width of each paired region and its correspondences to be consistently resized, whereas the shapes of salient objects are maintained via shape-preserving constraints. Finally, we employ a convex optimization solver to obtain the optimal local warping functions for individual quads so as to generate the final retargeted image pair.

A. Quad Classification

We first separate paired pixels from non-paired pixels in the stereo image pair, according to the disparity map estimated by stereo matching [29]. We subsequently group quads into two region classes based on the disparity map. Since existing
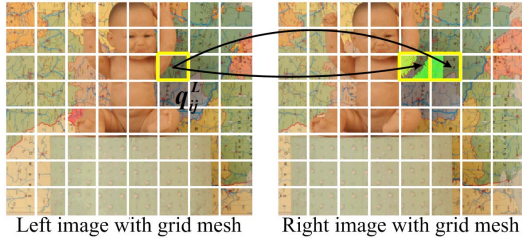
Fig. 5. Illustration of quad alignment. We align q_{ij}^L (marked by yellow) with quads containing corresponding content. Corresponding content of q_{ij}^L in I^R is marked by green.

disparity estimation methods cannot precisely detect all correspondences between the left and right images, we group quads according to the ratios of paired pixels and non-paired pixels in each quad. Specifically, given quad q_{ij}^z, which denotes the quad located at row i and column j in I^z, if the ratio of paired pixels in q_{ij}^z is greater than δ_1, q_{ij}^z is a paired quad r. Similarly, if the ratio of non-paired pixels in q_{ij}^z is greater than δ_2, q_{ij}^z is a non-paired quad $\bar{r}$. δ_1 and δ_2 are empirically set to 20% and 30%, respectively. Note that a quad may be attributed to both a paired quad and a non-paired quad, if the ratio of its paired pixels and non-pared pixels are greater than δ_1 and δ_2, respectively. Finally, we separate each paired-quad as a paired region, and group horizontally consecutive non-paired quads into a single non-paired region.

B. Quad Alignment

Depth-preserving constraints need to find the correspondences of paired quads. That is, given a paired quad $q_{ij}^z \in r_n^z$, we align it with $q_{ik}^{z'}$ that contains the corresponding content of q_{ij}^z. Since q_{ij}^z's corresponding content often appear in multiple quads in $I^{z'}$ (see Fig. 5), we allow a quad to be aligned with multiple quads. Instead of hard alignment, we propose a soft probabilistic quad alignment scheme, where the alignment probability P_{ijk} between q_{ij}^z and $q_{ik}^{z'}$ is defined as the ratio of those pixels in q_{ij}^z that can find their correspondence in $q_{ik}^{z'}$.

C. Shape Preservation

We adopt the shape-preserving constraints proposed in [10], [11], and [28]. The key to these constraints is to restrict the shapes of retargeted quads to be rectangular. That is, all retargeted quads in each row/column are restricted to have the same height/width. Such rectangular transformation has proven to be simple yet effective in preserving the shapes of structural objects. Following [10], [11], and [28], the constraints are expressed in three terms: shape distortion energy, neighboring-quad constraints, and resizing-budget constraints. As a result, the warping functions are also represented as functions of the height of each quad row and the width of each quad column, rather than the locations of vertices, to reduce the computational complexity of the optimization.

Shape Distortion Energy: The shape distortion energy of a quad is simply measured by the change of the quad's aspect ratio, which is the ratio between the height of a quad row and the width of a quad column. Hence, the total shape distortion of an image pair is calculated by

$$D^s = \sum_{z \in \{L,R\}} \sum_{i=1} \sum_{j=1} \| w_j^z \cdot \widetilde{h}_i^z - h_i^z \cdot \widetilde{w}_j^z \|^2 \cdot s_{ij}^z, \quad (10)$$

where w_j^z and h_i^z respectively denote the width of the jth quad row and the height of the ith column in the original quad mesh, and $\widetilde{w}_j^z$ and $\widetilde{h}_i^z$ are the corresponding width and height in the retargeted quad mesh, s_{ij}^z is the importance of quad q_{ij}^z calculated by averaging the importance values of pixels in q_{ij}^z.

Neighboring Quad Constraints: When the quads associated with the same object undergo significantly inconsistent warping, noticeable object distortion may occur. Therefore, to avoid such spatially-inconsistent shape distortion, neighboring quads are constrained to undergo consistent warping according to the absolute difference of importance between two horizontally/vertically neighboring quads as follows:

$$\begin{cases} \dfrac{1}{\varphi_w(|s_{ij}^z - s_{i,j+1}^z|)} \leq \dfrac{\widetilde{w}_j^z - w_j^z}{\widetilde{w}_{j+1}^z - w_{j+1}^z} \leq \varphi_w(|s_{ij}^z - s_{i,j+1}^z|) \\ \dfrac{1}{\varphi_h(|s_{ij}^z - s_{i+1,j}^z|)} \leq \dfrac{\widetilde{h}_i^z - h_i^z}{\widetilde{h}_{i+1}^z - h_{i+1}^z} \leq \varphi_h(|s_{ij}^z - s_{i+1,j}^z|), \end{cases}$$
$$(11)$$

where $\varphi_w(\cdot) > 1$, $\varphi_h(\cdot) > 1$.

Resizing-Budget Constraints: Suppose the target size is $W' \times H'$, the sizes of individual quads are constrained to meet the following target size budget:

$$\begin{cases} \sum_{i=1}^M \widetilde{h}_i^z = H', \\ \sum_{j=1}^N \widetilde{w}_j^z = W', \\ \widetilde{h}_i^z > 0, \widetilde{w}_j^z > 0. \end{cases} \quad (12)$$

D. Depth Preservation

We build up the depth-preserving constraints on paired region and non-paired region in (9), respectively. In practice, the vertical disparity is eliminated.

To derive the depth-preserving constraints in (9), we first represent $\Phi(\tilde{\mathbf{v}}_{ij}^z)$ in terms of $\widetilde{w}_j^z$ and $\widetilde{h}_i^z$. Nevertheless, the original formulation for $\Phi(\tilde{\mathbf{v}}_{ij}^z)$ was based on the "fine-granular mesh" assumption that each quad contains only a single pixel, which is not the case for a coarse mesh where one quad contains multiple pixels. To extend the "fine-granular mesh" formulation to more general "coarse-granular mesh" cases, we treat the pixels in a quad as a super-pixel that lies in the left edge of the quad. Then we can calculate $\Phi(\tilde{\mathbf{v}}_{ij}^z) = \widetilde{w}_j^z$, since a retargeted quad remains rectangular.

Non-Paired-Region Constraints: For a non-paired region $\bar{r}_n^z$, from (9), we have $\sum_{q_{ij}^z \in \bar{r}_n^z} \Phi(\tilde{\mathbf{v}}_{ij}^z) = \sum_{q_{ij}^z \in \bar{r}_n^z} w_j^z$. Thus, for $\bar{r}_n^z$, the following hard constraint is imposed to preserve the total width of quads belonging to $\bar{r}_n$:

$$\sum_{q_{ij} \in \bar{r}_n} w_j^z = \sum_{q_{ij} \in \bar{r}_n} \widetilde{w}_j^z, \quad \forall \bar{r}_n \in \bar{\Upsilon}, \quad (13)$$

Paired-Region Constraints: We impose the relaxed depth-preserving constraints i.e., the energy function D_p

in (9) to softly control the warping of paired regions, and build additional constraints to avoid disparity distortions caused by the relaxation. By substituting $\Phi(\mathbf{v}_{ij}^z)$ with $\widetilde{w}_j^z$ in (8), we have D_p which softly constrains each paired-region quad $q_{ij}^z \in r_n^z$ and its aligned quad $q_{ik}^{z'}$ to undergo consistent width change:

$$D^p = \sum_{r_n^z \in \Upsilon} \sum_{q_{ij} \in r_n^z} \varpi_{ijk}(\widetilde{w}_j^z - w_j^z - (\widetilde{w}_k^{z'} - w_k^{z'}))^2, \quad (14)$$

where weight ϖ_{ijk} depends on the alignment probability P_{ijk} between q_{ij}^z and $q_{ik}^{z'}$:

$$\varpi_{ijk} = \begin{cases} 1, & if \ P_{ijk} = \arg\max_t P_{ijt}, \\ P_{ijk}, & otherwise, \end{cases} \quad (15)$$

Thus, the higher the alignment probability between q_{ij}^z and $q_{ik}^{z'}$ is, the smaller the value of $|\widetilde{w}_j^z - w_j^L - (\widetilde{w}_k^z - w_k^R)|$ will be.

We build additional constraints to avoid disparity distortions caused by the relaxation. In particular, we choose a few left-top quad vertices, which can find their correspondence in the other image, as control points, and directly enforce the disparity between these control points and their correspondences to remain unchanged. Let $\mathbf{U}_c$ be the set of control points, $c_k^z \in \mathbf{U}_c$ the left-top vertices of $q_{i,m}^z$, and $c_k^{z'}$ the correspondence of c_k^z. The constraints become

$$\frac{1}{\varphi_c} \le \big(\sum_{j=1}^m \widetilde{w}_j^z - \widetilde{x}_k^{z'} - (\sum_{j=1}^m w_j^z - x_k^{z'})\big) \le \varphi_c \quad \forall c_{i,m}^z \in \mathbf{U}_c \tag{16}$$

where $\sum_{j=1}^m w_j^z$ and $\sum_{j=1}^m \widetilde{w}_j^z$ are the x-coordinates of c_k^z in the original quads and retargeted quads, respectively, $\varphi_c > 1$, $x_k^{z'}$ and $\widetilde{x}_k^{z'}$ denote the x-coordinate of $c_k^{z'}$ in the original quads and the retargeted quads, respectively.

Vertical Disparity: Since nonzero vertical disparity would introduce unconformable 3D experience e.g., eye fatigue or strain [30], we constrain the retargeted quads lying at the same row to be of the same height, so as to eliminate nonzero vertical disparity:

$$h_i^z = h_i^{z'} \tag{17}$$

E. Solving the Optimization Model

To find the optimal retargeted quads, we minimize the objective function $D^s + \lambda D^p$, subject to the shape-preserving constraints in (11) and (12), and the depth-preserving constraints in (13), (14), (16), and (17).

The above optimization problem is a quadratic programming problem and can be solved using the active-set method proposed in [31].

VI. Experimental Results

In the experiments, we first validate the effectiveness of the proposed depth-preserving constraints, and then evaluate the performance of the proposed method by comparing it with the state-of-the-art approaches. We finally evaluate the

efficiency of the proposed method. Complete results are available in [32].

Dataset: To make fair comparisons with existing methods, we select stereoscopic image pairs from the following publicly available datasets: the Middlebury stereo dataset [33] and the Flicker datasets used in [15]. In selecting the test image pairs, the following two aspects are taken into account: (a) Depth diversity: a test pair should contain objects with diverse depths in a significant depth range. Such depth diversity can better evaluate the depth preservation performance of a stereo image retargeting method. (b) Content diversity: part of the test stereo pairs should contain enough non-salient regions for evaluating the shape preservation performance of a retargeting method on salient objects, whereas part should contain sufficiently large salient objects to challenge a retargeting method in both shape preservation for salient objects and depth preservation for 3D scenes.

A. Performances of Depth-Preserving Constraints

We evaluate the performances of depth-preserving constraints in two aspects: (i) whether the constraints can faithfully preserve the depth; (ii) whether they can properly constrain the warping to preserve the depth, where the ideal depth-preserving constraints in (7) are used as a reference. For (i), we derive a map and define a metric to assess the depth distortion of a retargeted stereo pair. For (ii), we define two metrics to assess the inconsistency between the ideal depth-preserving constraints and the warping functions of the retargeted stereo pair, and derive a map to indicate regions/pixels of which local warping functions are inconsistent (mismatched) with the ideal depth-preserving constraints. Specifically, for (i), the following map and quality metric are calculated:

(a) Disparity map: The map is used to subjectively evaluate the depth distortion of a retargeted stereo pair by comparing the retargeted disparity map with the original map. The disparity map is obtained by stereo matching [29].

(b) Average disparity distortion (ADD): The metric quantitatively evaluates the total depth distortion of a stereo image pair by calculating the total disparity differences of quad vertices and their correspondences as follows:

$$ADD = \frac{\sum_{v_i^L \in \mathbf{U}_g} |d_i - \tilde{d}_i|}{N_{\mathbf{U}_g}} \tag{18}$$

where $\mathbf{U}_g$ is the set containing those quad vertices which can find a correspondence in the other image, and $N_{\mathbf{U}_g}$ is the size of $\mathbf{U}_g$.

For (ii), we define a map and two quality metrics as follows:

(a) Warping deviation map: The map is used to evaluate the deviation (mismatch) between the warping functions of a retargeted stereo pair and the ideal depth-preserving constraints. In particular, given pixel $p_{xy}^z \in q_k^z$ located in (x, y), if p_{xy}^z is a non-paired one, the warping deviation map $\{\Omega_{xy}^z\}$ records the deviation between the local warping function of q_k^z and the ideal depth-preserving constraints for a non-paired region; otherwise, it records the deviation between the local warping function of q_k^z and the ideal depth-preserving

Left image   Right image   Warping deviation map of left image   Warping deviation map of right image   Disparity map
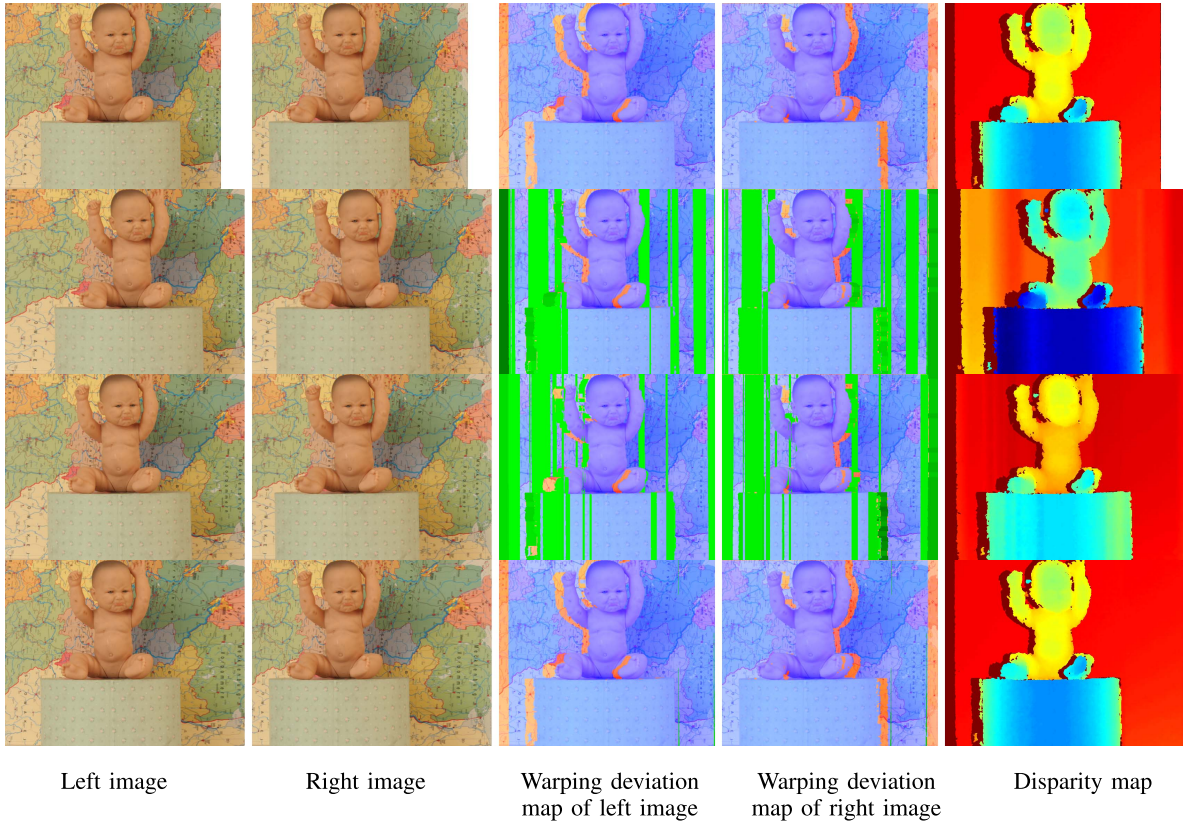
Fig. 6. Retargeting results and the associated disparity maps and warping deviation maps using different depth-preserving constraints on *Baby1* (its width is enlarged by 10%). The first row shows the original images and maps (first row), and the second to fourth rows show the results of VDP+QW, SCDP+QW, and WF+QW (our method), respectively. In the warping deviation map, a paired/non-paired pixel is colored green (darker green indicates higher deviation), if the local warping functions of its associated quads deviate from the ideal depth-preserving constraints; otherwise, a non-paired pixel is colored by red and a paired pixel is colored by blue. In the disparity map, we visualize the pixel disparity by colors ranging from dark red (the lowest absolute disparity) to dark blue (the highest one).

constraints for a paired region as follows:

$$\Omega_{x,y}^z$$
$$= \begin{cases} |\Phi(\tilde{\mathbf{v}}_k^L) - w_k^L - (\Phi(\tilde{\mathbf{v}}_k^R) - w_k^R)|, & p_{xy}^z \text{ is a non-paired pixel} \\ |\sum_{q_{\tilde{k}}^z \in \tilde{r}_j^z} \Phi(\tilde{\mathbf{v}}_{\tilde{k}}^z) - \sum_{q_{\tilde{k}}^z \in \tilde{r}_j^z} w_{\tilde{k}}^z|, & \text{otherwise.} \end{cases}$$

*(b) Average warping deviation over paired regions (*AWD$_P$*):* The metric evaluates the total amount of deviation between the individual local warping functions and the ideal depth-preserving constraints over paired regions as calculated:

$$AWD_P = \frac{\sum_{p_{xy}^z \in \mathbf{U}_P} \Omega_{x,y}^z}{N_{\mathbf{U}_P}} \qquad (19)$$

where $\mathbf{U}_P$ is the set of all paired pixels in the left and right images of a stereo image pair, and $N_{\mathbf{U}_P}$ is the size of $\mathbf{U}_P$.

(c) Average warping deviation over non-paired regions *(*AWD$_{NP}$*):* The metric evaluates the total amount of deviation between the individual local warping functions and the ideal depth-preserving constraints over non-paired regions as calculated by

$$AWD_{NP} = \frac{\sum_{p_{xy}^z \in \mathbf{U}_{NP}} \Omega_{x,y}^z}{N_{\mathbf{U}_{NP}}} \qquad (20)$$

where $\mathbf{U}_{NP}$ is the set of all non-paired pixels in a stereo image pair, and $N_{\mathbf{U}_{NP}}$ is the size of $\mathbf{U}_{NP}$.

We compare the performances of the proposed warping function-based depth-preserving constraints (denoted by WF) and two other heuristic depth-preserving constraints: the vertical depth-preserving (VDP) constraints and the sparse correspondences-based depth-preserving (SCDP) constraints. The VDP constraints are to maintain the vertical disparity, regardless of the horizontal disparity, whereas the SCDP constraints are to preserve as much as possible the disparity value of a set of sparse correspondences [18], [20]. We conduct three baseline experiments that apply the three types of constraints (VDP, SCDP, and WF) to the quad-based warping method, namely, VDP+QW, SCDP+QW, and WF+QW (our method), respectively.

Fig. 6 and Fig. 8 show the subjective and objective comparison results for test image pair *Baby1*, respectively. As shown in Fig. 8, WF+QW (our method) preserves the depth very well in terms of the *ADD*, *AWD$_P$*, and *AWD$_{NP}$* metrics (all the three distortions are low), whereas VDP+QW and SCDP+QW severely distort the depth, despite the image width is enlarged by only 10%. This is because the VDP and SCDP constraints do not have direct control on the warping functions, and hence may not properly constrain the local warpings, thereby leading to large deviations between the local warping functions and
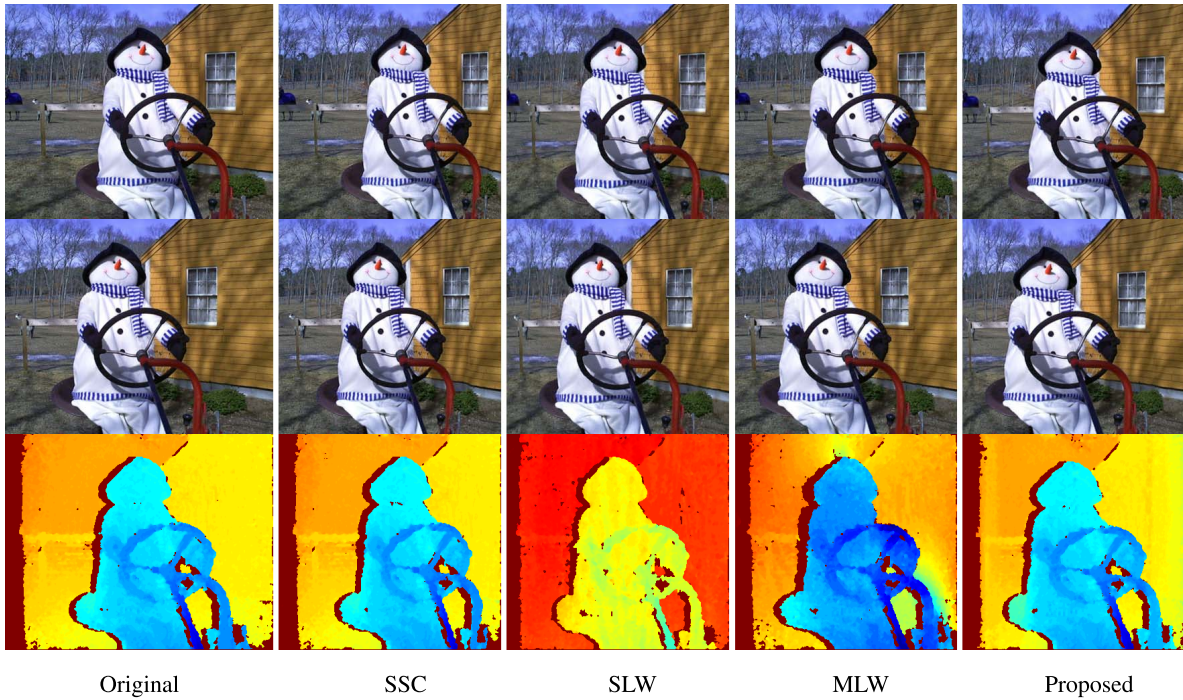
Fig. 7. Performance comparison for *Snowman*. From top to bottom: the left image, the right image, and the disparity map. We visualize the pixel disparity by colors ranging from dark red (the lowest absolute disparity) to dark blue (the highest one).
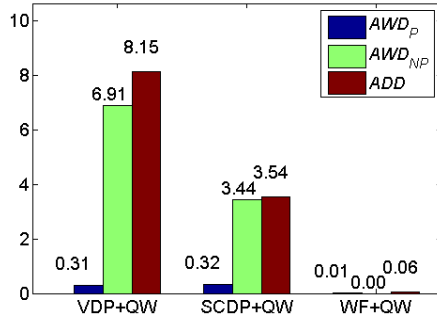

Fig. 8. Comparison of depth distortion in terms of ADD, AWD_P, and AWD_{NP} metrics with different depth-preserving constraints on *Baby1* whose width is enlarged by 10%.

the ideal depth-preserving warping constraints. For example, Fig. 8 shows that the AWD_{NP} values of VDP+QW and SCDP+QW are 6.91 and 3.44, respectively, both indicating large deviation of local warping functions from their ideal depth-preserving constraints. The warping deviation maps of VDP+QW and SCDP+QW in Fig. 6 also illustrate that, in some paired and non-paired regions, the two methods lead to significant deviation between their local warping functions and the ideal depth-preserving constraints. In contrast, the local warping functions of our method agree with the depth-preserving constraints quite well, thereby achieving good depth preservation performance.

B. Comparisons of Stereo Retargeting Methods

Shape and Depth Preservation: We subsequently compare our method with three state-of-the-art stereo image retargeting schemes including Stereo Seam Carving (SSC) [15], Single-Layer Warping (SLW) [18], and Multi-Layer Warping (MLW) [20] in terms of shape preservation

for salient objects and depth preservation for 3D scenes. Following the resizing setting in [15], the test image pairs' widths are shrunk by 20% fors *Aloe* and by 17% for the rest.

We first compare our method with two warping-based schemes: SLW and MLW. Figs. 7, 9, 10, 11 illustrate the retargeting results on test image pairs *Snowman*, *Man*, *People* and *Aloe*, respectively. Overall, SLW, MLW, and our method all perform well in shape preservation for salient objects, thanks to the continuous manner of warping-based retargeting. On the other hand, our method significantly outperforms SLW and MLW in depth preservation, since SLW and MLW tend to preserve as much as possible the disparity values of a set of sparse correspondences (i.e., the SCDP constraint), but often cannot do a good job in depth preservation due to the lack of direct control on local warping functions of quads. For example, as illustrated in Figs. 7, 9 and 10, SLW changes the disparity maps significantly compared to the original maps in most regions of *Snowman*, *Man* and *People*, whereas MLW distorts the disparity maps of *Snowman*, *Man* and *Aloe* (see Figs. 7, 9, and 11). Moreover, MLW and SLW tend to produce inconsistent disparity changes in different regions. For example, for *Snowman*, MLW decreases the disparity for background while increasing the disparity for foreground objects. Besides, SLW introduces holes in the disparity map of *Aloe*, meaning that there exists depth discontinuity in an object. In contrast, since the proposed depth-preserving constraints directly control the warping functions, our method preserves the depths of 3D scenes very well for all the test images pairs.

In addition, we also compare our method with SSC [15] which is a representative discrete retargeting method. As illustrated in Figs. 7, 9, 10, and 11, although SSC can do a good job in depth preservation, it yields noticeable
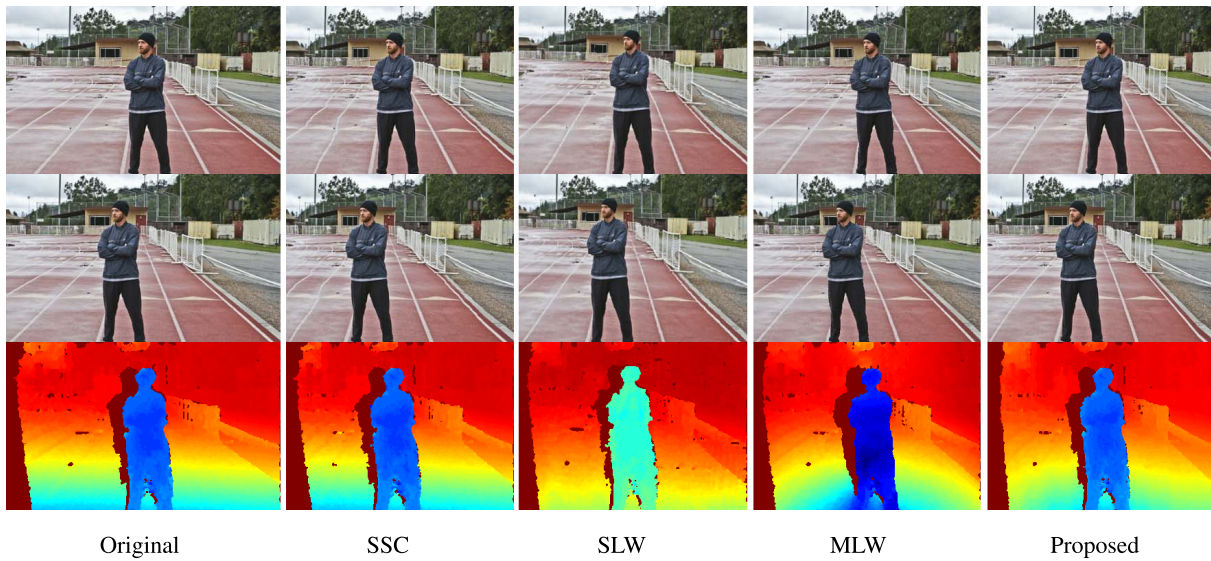
Fig. 9. Performance comparison for *Man*. From top to bottom: the left image, the right image, and the disparity map. We visualize the pixel disparity by colors ranging from dark red (the lowest absolute disparity) to dark blue (the highest one).
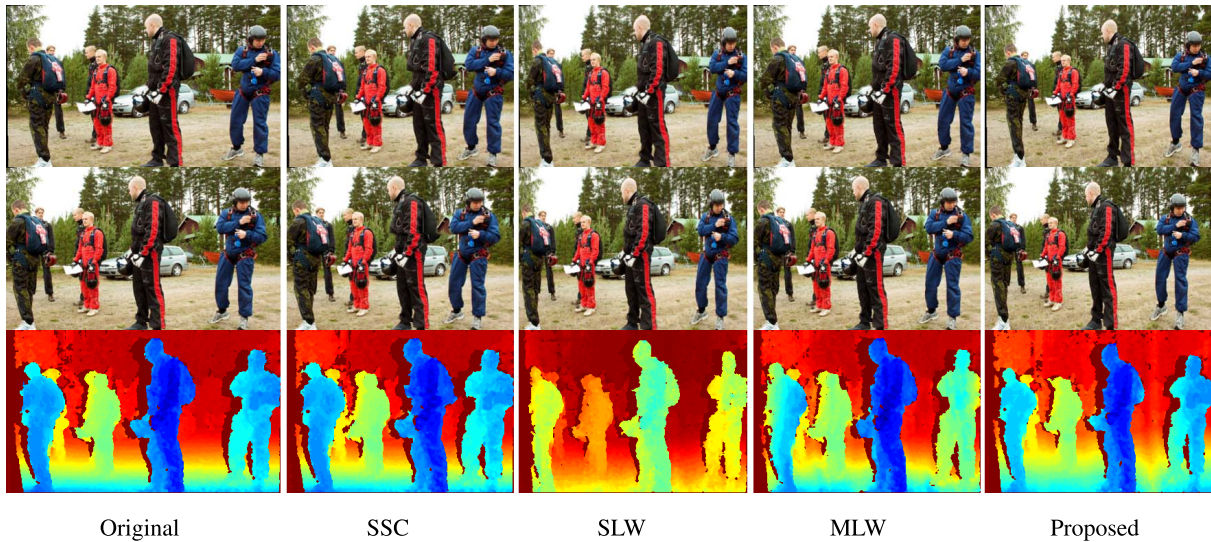

Fig. 10. Performance comparison for *People*. From top to bottom: the left image, the right image, and the disparity map. We visualize the pixel disparity by colors ranging from dark red (the lowest absolute disparity) to dark blue (the highest one).

deformations on object shapes. For example, SSC severely distorts the shapes of the flowerpot and aloe leaves in Fig. 11, the red tube of the steering wheel in Fig. 7, and the upper part of the man wearing a black sports jacket in Fig. 10. The shape distortions are caused by SSC's discrete pixel-wise processing which trims a region in an unsmooth manner, thereby incurring significantly inconsistent changes in neighboring regions around an object. Such shape deformations usually cause uncomfortable 3D viewing experience. In contrast, our method achieves good depth preservation for 3D scenes without introducing severe shape distortions.

Subjective Paired Comparisons: We conduct subjective evaluations on a 1680×1050 ViewSonic VX2286wm 3D 22-inch monitor with NVIDIA active shuttered glasses and an NVIDIA GeForce 3D Vision Solution. The width and height of the display screen are 0.474 m and 0.296 m, respectively. We select a moderate display size to ensure

that the perceptual depth stays in the comfort zones for all participants. The viewing conditions follow the guidelines of ITU-R BT.2021 [34]–[36], that suggest a viewing distance of 4 times the height of the display screen (1.1 m in our experiments).

We invite 24 subjects (10 males and 14 females) with ages ranging from 21 to 28 to participate in the subjective user study, in which two of them have specialized experiences in the 3D perception field. All subjects do not have any prior knowledge about the experiment hypothesis.

We perform subjective paired comparisons in a way similar to that of several 2D image and video retargeting methods [37]. In particular, an original image pair together with its two retargeted versions generated by different retargeting methods are displayed side-by-side on the screen, where the original image pair is placed in the center, whereas the retargeted versions are placed in a random order so that the subjects
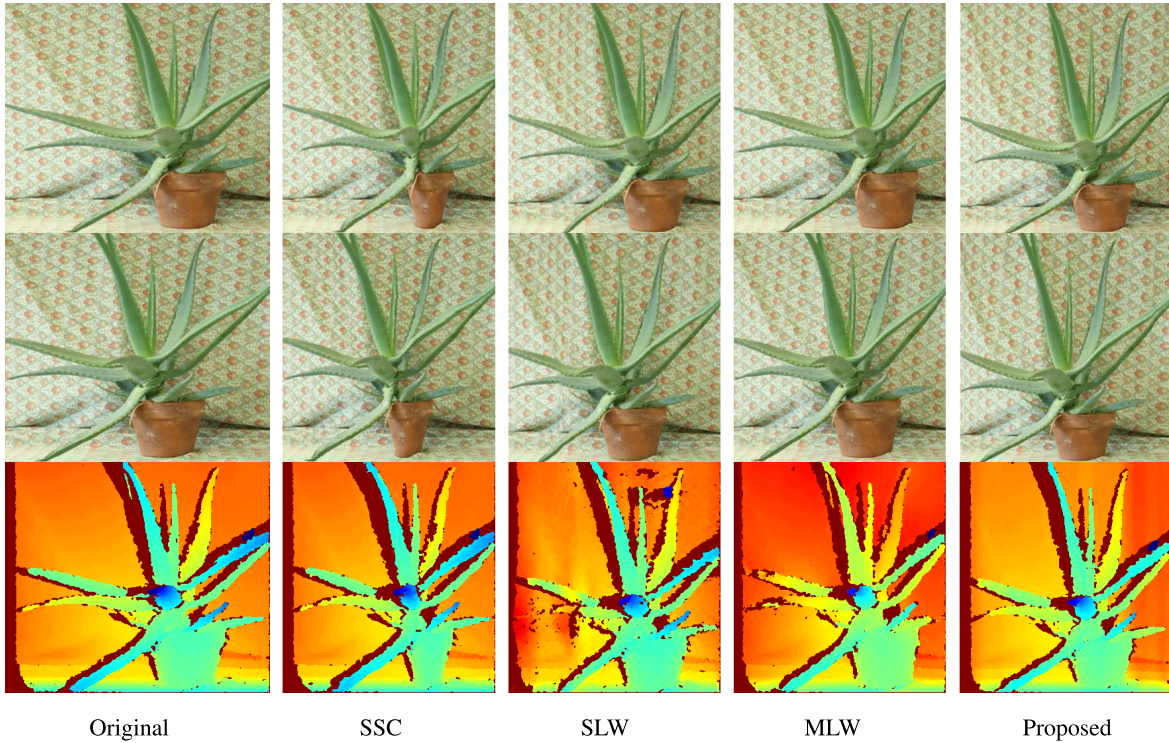
| Original | SSC | SLW | MLW | Proposed |

Fig. 11. Performance comparison for *Aloe*. From top to bottom: the left image, the right image, and the disparity map. We visualize the pixel disparity by colors ranging from dark red (the lowest absolute disparity) to dark blue (the highest one).

TABLE II

WINNING FREQUENCY MATRIX OF SUBJECTIVE PAIRED COMPARISONS
FOR DEPTH PRESERVATION QUALITY

	Ours	US	SLW	MLW	SSC	Total
Ours	-	90(93.8%)	76(79.2%)	75(78.1%)	59(61.5%)	**300**(78.1%)
US	6(6.3%)	-	17(17.7%)	10(10.4%)	8(8.3%)	41(10.7%)
SLW	20(20.8%)	79(82.3%)	-	29(30.2%)	32(33.3%)	160(41.7%)
MLW	21(21.9%)	86(89.6%)	67(69.8%)	-	33(34.4%)	207(53.9%)
SSC	37(38.5%)	88(91.7%)	64(66.7%)	63(65.6%)	-	252(65.6%)

TABLE III

WINNING FREQUENCY MATRIX OF SUBJECTIVE PAIRED COMPARISONS
FOR SHAPE PRESERVATION QUALITY

	Ours	US	SLW	MLW	SSC	Total
Ours	-	85(88.5%)	80(83.3%)	48(50.0%)	81(84.4%)	**294**(76.6%)
US	11(11.5%)	-	6(6.3%)	7(7.3%)	30(31.3%)	54(14.1%)
SLW	16(16.7%)	90(93.8%)	-	22(22.9%)	54(56.3%)	182(47.4%)
MLW	48(50.0%)	89(92.7%)	74(77.1%)	-	80(83.3%)	291(75.8%)
SSC	15(15.6%)	66(68.8%)	42(43.8%)	16(16.7%)	-	139(36.2%)

TABLE IV

WINNING FREQUENCY MATRIX OF SUBJECTIVE PAIRED COMPARISONS
FOR THE OVERALL RETARGETING QUALITY IN DEPTH
AND SHAPE PRESERVATION

	Ours	US	SLW	MLW	SSC	Total
Ours	-	84(87.5%)	78(81.3%)	55(57.3%)	81(84.4%)	**298**(77.6%)
US	12(12.5%)	-	9(9.4%)	2(2.1%)	28(29.2%)	51(13.3%)
SLW	18(18.8%)	87(90.6%)	-	28(29.2%)	50(52.1%)	183(47.7%)
MLW	41(42.7%)	94(97.9%)	68(70.8%)	-	55(57.3%)	258(67.2%)
SSC	15(15.6%)	68(70.8%)	46(47.9%)	41(42.7%)	-	170(44.3%)

TABLE V

RUNTIME COST

Method	SLW	MLW	Ours
Time	(4.93+(3+0.086))s	300s	(4.32+0.380)s

have no idea about the associated methods. In the subjective test, each subject is requested to choose one retargeted version which he/she prefers according to each of the following three criteria: (1) the quality in shape preservation, (2) the quality in depth preservation, and (3) the overall quality incorporating both shape and depth preservation.

We compare our method with four representative methods: Uniform Scaling (US), SSC, SLW and MLW, using four stereo image pairs: *Snowman*, *Man*, *Aloe* and *People*. For a stereo pair, each subject spends about 10 minutes to perform 10 paired comparisons between five compared methods. Following the suggestion of ITU-R BT.2021 [34], [35], each subject would take a break when the test duration reaches 20 minutes.

We collect $24 \times 40 = 960$ paired comparison results, in which each method is compared $4 \times 24 \times 4 = 384$ times. All results are listed in Tables II, III, IV, where entry *n* of row A and column B indicates that method A receives *n* preference votes compared to method B. As listed in Table II, in terms of depth preservation, our method outperforms the other four methods in 78.1% of the paired comparisons on average, including 93.8%, 79.2%, 78.1%, and 61.5% of the

votes in preference for our method to US, SLW, MLW, and SSC, respectively. Besides, Table III shows that, in terms of shape preservation, our method receives 76.6% preference votes on average, where our method is preferred to US in 88.5%, to SLW in 83.3%, to MLW in 50.0%, and to SSC in 84.4% of the preference votes, respectively. In terms of the
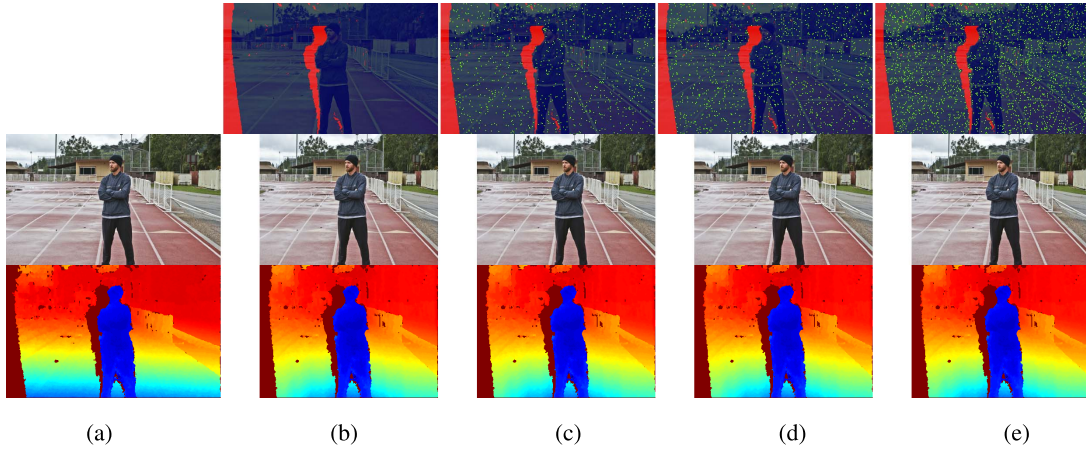
| (a) | (b) | (c) | (d) | (e) |

Fig. 12. Results for *Man* where some randomly selected paired pixels are misclassified as non-paired ones. From top to bottom: the region classification map, the left image, and the resulting disparity map. From left to right: (a) the original images, (b) the results based on the original disparity map,(c)-(e) the results based on the noisy disparity maps, with different percentages of misclassified pixels. The region classification map in (b) is used as the ground-truth, where a paired pixel and non-paired one are colored blue and red, respectively. In the top row of (c)-(e), those paired pixels that are misclassified as non-paired ones are colored green.

overall retargeting quality, Table IV shows that our method outperforms the other four methods in 77.6% of the preference votes on average, where the breakdowns are 87.5%, 81.3%, 57.3%, and 84.4% of the votes in preference for our method to US, SLW, MLW, and SSC, respectively.

C. Run-Time Complexity

We implement our method on a PC with a 2.26 GHz Duo CPU and 4 GB RAM to compare the run-time costs of our method, SLW and MLW. As shown in Table V, both our method and SLW consume significantly lower time costs compared to MLW. For our method and SLW, we divide the time cost into two parts: the preprocessing before the optimization and the optimization itself. For an 800×600 stereo image pair, our method consumes on average 4.32s in disparity estimation of preprocessing, whereas SLW consumes about 4.93s, mainly in SIFT feature extraction and matching.

Besides, our method also consumes a lower time cost in optimization compared to SLW. Similar to the 2D video retargeting method in [28], our optimization model is formulated as a convex optimization problem, which can be efficiently solved by an active set algorithm. In contrast, SLW tackles the optimization problem by matrix factorization and back substitution, in which matrix factorization is computationally expensive especially for a large matrix (i.e., involving a large number of variables). For a quad size of 20×20, our method consumes about 0.380s, whereas SLW spends about 2s in matrix factorization and 0.0086s in back substitution.

Compared to our method and SLW, MLW's time complexity is significantly higher, since, besides disparity estimation, MLW needs to segment each image into objects, which is computationally expensive. In addition, the iterative optimization used in MLW consumes a high time cost as well.

D. Discussions

Note that, the accuracy of disparity estimation affects the accuracy of region classification, thereby influencing the

retargeting performance of our method. Nevertheless, our quad classification based method can tolerate disparity estimation errors to some extent, thanks to the soft quad classification and the quad-based warping used in the method.

To show the error-tolerance capability of our method, we intentionally change the disparity values for a randomly selected set of non-paired/paired pixels (0%-7% of the total non-paired/paired pixels), such that part of these non-paired/paired pixels are misclassified. As shown in Fig. 12, the retargeting results using the noisy disparity map are still very close to those using the original disparity map. This shows that our method can tolerate small disparity estimation errors since the quad classification is based on the ratios of paired/non-paired pixels. With such ratio-based classification, even if some paired/non-paired pixels are misclassified due to disparity estimation errors, a quad containing misclassified pixels can still be correctly classified when the ratio of misclassified pixels in the quad is not significant.

We then intentionally increase the amount of noise injected to the disparity maps, leading to the misclassification of quite a few quads, to further evaluate the error-tolerance capability of our method. As shown in Fig. 13(c) and 13(d), when a moderate percentage (0%-6%) of quads are misclassified, our method still achieves comparable performances with that using the original disparity map because our quad-based warping scheme can partially compensate for quad classification errors. In particular, our quad-based warping with the neighboring quad constraints tends to warp a quad in a manner similar to those of the quad's neighbors with the same class. It is very likely that, for most quads, a significant number of their neighboring quads belong to the paired/non-paired class. In such cases, although some quads are misclassified, thanks to the neighboring quad constraints, these quads are warped in a way similar to their neighbors with the same class. As a result, these misclassified quads may be properly warped, thereby tolerating quad classification errors to some extent. However, as illustrated in Fig. 13(e), when too many quads are misclassified (e.g., 21% misclassified paired quads in this example), our method would incur depth distortions,
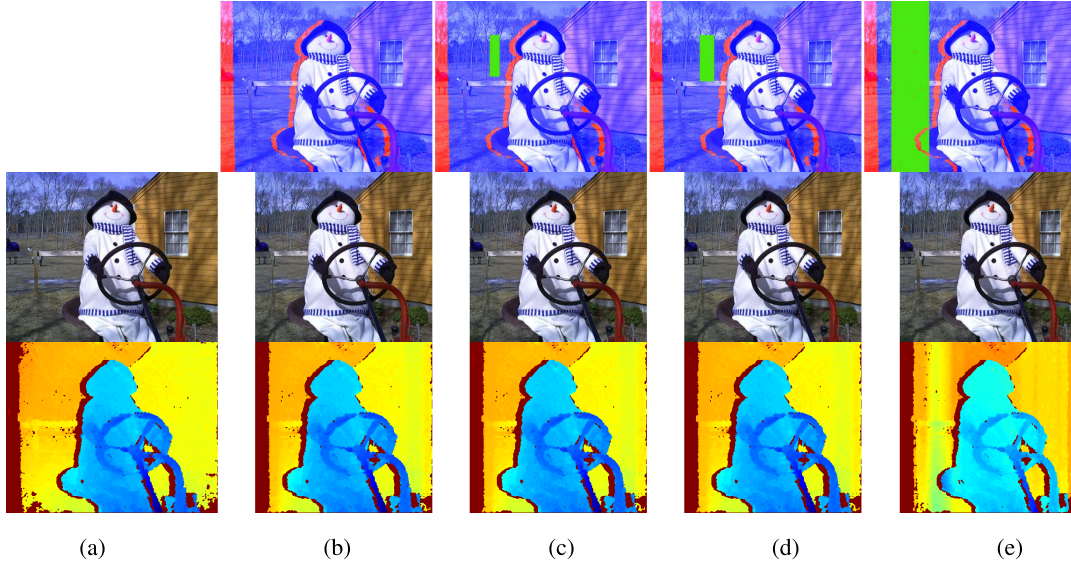
Fig. 13.   Results for *Snowman* where some paired regions are misclassified as non-paired ones. From top to bottom: the region classification map, the left image and the resulting disparity map. From left to right: (a) the original images, and (b) the results based on the original disparity map, (c)-(e) the results based on noisy disparity maps. The region classification map in (b) is treated as the ground-truth, where paired pixels and non-paired ones are colored blue and red, respectively. In the top row of (c)-(e), those paired pixels that are misclassified as non-paired ones are colored green.

since the neighbors of misclassified quads are also largely misclassified.

VII. CONCLUSION

In this paper, we investigated how to preserve scene depth by directly controlling the local warping functions. Based on the binocular visual characteristics of two different region classes, we derived effective region-wise depth-preserving constraints to control the local warping functions of the two region classes directly. With the derived depth-preserving constraints, we have proposed a warping-based stereo retargeting framework to simultaneously achieve both shape preservation and depth preservation while retargeting a stereo image pair. We have also proposed a quad-based implementation of the proposed retargeting framework with a reasonable computational complexity. Our experimental results demonstrate that our method achieves promising subjective and objective visual qualities in both shape preservation and depth preservation for various test stereo image pairs.

APPENDIX A
DETAILED DERIVATION OF (3)

We here elaborate how to represent x-coordinate $\tilde{x}_i$ of correspondence $\mathbf{c}_i$ in a warped image in terms of the local warping functions of quads for (3). Given $\mathbf{c}_i$, let $q_1^z \ldots q_{N_i^z}^z$ be the quads that are crossed by $\mathbf{E}_i^z$ and lie to the left of c_i^z, where $\mathbf{E}_i^z$ intersects the left edge of q_k^z at o_k in the warped mesh (see Fig.2). Hence, $\tilde{x}_i^z$ is discretized by the width $|e_k^z|$ of line segment $o_k o_{k+1}$, i.e., $\tilde{x}_i^z = \sum_{k=1}^{N_i^z} |e_k^z|$.

Furthermore, we represent $|e_k^z|$ by the local warping functions of q_k^z, using the barycentric coordinates:

$$|e_k^z| = \Phi(\tilde{\mathbf{v}}_k^z) = (1 - \beta_k)\tilde{v}_{k,x}^{z,3} + \beta_k \cdot \tilde{v}_{k,x}^{z,4}$$
$$- ((1 - \beta_{k+1})\tilde{v}_{k,x}^{z,1} + \beta_{k+1} \cdot \tilde{v}_{k,x}^{z,2}) \quad (21)$$

where $\tilde{\mathbf{v}}_k^z$ consists of $\tilde{v}_k^{z,1}, \tilde{v}_k^{z,2}, \tilde{v}_k^{z,3}, \tilde{v}_k^{z,4}$ which represent the left-top, left-bottom, right-top, right-bottom vertices of the retargeted version of q_k^z, respectively, $v_{k,x}^{z,1}, v_{k,x}^{z,2}, v_{k,x}^{z,3}$ and $v_{k,x}^{z,4}$ represent the x-coordinates of $v_k^{z,1}, v_k^{z,2}, v_k^{z,3}$ and $v_k^{z,4}$, respectively, β_k the barycentric x-coordinate of o_k.

We present a simplified calculation of β_k, based on the domain knowledge of warping-based image retargeting. Almost all existing warping-based image retargeting methods enforce the angle between the original edge and the warped edge to be as small as possible. In such case, the top and bottom edges of a warped quad are parallel or approximately parallel, making the barycentric coordinates of o_k approximate that of c_i. Therefore, we assume $\beta_k = \alpha$, where α is the barycentric x-coordinates of c_i^z. Note, α is a known value, since it just depends on the mesh partition. Moreover, pixels whose quad lying in the same row have the same barycentric coordinates. Hence, we can easily determine α_k after mesh partition. For example, if we use a uniform mesh with 1×1 quad size and all pixels lie at the middle of the left edge of a quad, $\beta_1 = \beta_2 \ldots = \alpha = 1/2$.

APPENDIX B
DETAILED DERIVATION OF DEPTH-PRESERVING CONSTRAINTS IN (7)

We prove that, given $\forall \mathbf{c}_i \in \mathbf{C}$, $d_i = 0$ if $\mathbf{V}$ satisfies

$$\begin{cases} \sum_{q_k^z \in r_j^z} \Phi(\tilde{\mathbf{v}}_k^z) = \sum_{q_k^z \in r_j^z} w_k^z, & \forall r_k^z \in \Upsilon_o \\ \Phi(\tilde{\mathbf{v}}_k^L) - w_k^L = \Phi(\tilde{\mathbf{v}}_k^R) - w_k^R, & \forall q_k^L \in r_j^L, q_k^R \in r_j^R. \end{cases} \quad (22)$$

Proof: We use mathematical induction to prove the above. Note that the depth-preserving constraints for a quad only depend on the region class that the quad is attributed to. Therefore, we just need to prove that for $\mathbf{c}_i$, no matter how

many quads lie to the left of c_i^L or c_i^R, and no matter which region class these quads belong to, if these quads satisfy the constraints related to their region class, $\Delta d_i = 0$.

We first prove the simplest case. That is, given $\mathbf{c}_i$, no region lies to the left of c_i^R, while only a non-paired region $\bar{r}_1^L$ lies to the left of c_i^L and is crossed by the horizontal line $\mathbf{E}_i$. For $\mathbf{c}_i$, Δd_i is then calculated by

$$\Delta d_i = \left| -\left(\sum_{q_k^L \in \bar{r}^L} \Phi(\tilde{\mathbf{v}}_1^L) - \sum_{q_k^L \in \bar{r}^L} w_k^L \right) \right| \quad \bar{r}_1^L \in \Upsilon. \quad (23)$$

If $\sum_{q_k^L \in \bar{r}_1} \Phi(\tilde{\mathbf{v}}_1^L) = \sum_{q_k^L \in \bar{r}_1} w_k^L$, $\Delta d_1 = 0$. Therefore, the result is true for the simplest case.

Assume the result is true for the case : N quads $q_1^L \ldots q_N^L$ lie to the left of c_i^L, which are grouped into paired regions $r_1^L, \ldots r_n^L$ and non-paired region $\bar{r}_1^L, \ldots \bar{r}_m^L$; N' quads $q_1^R \ldots q_{N'}^R$ lie to the left of c_i^R, which are grouped into $r_1^R, \ldots r_n^R$ and $\bar{r}_{m+1}^R, \ldots \bar{r}_{n'}^R$. Thus, we have

$$\Phi(\tilde{\mathbf{v}}_k^R) - w_k^R - ((\Phi(\tilde{\mathbf{v}}_k^L) - w_k^L)) = 0, \quad r_j^z = r_1^z, \ldots r_n^z$$
$$\sum_{q_k^z \in r_j^z} \Phi(\tilde{\mathbf{v}}_k^z) - \sum_{q_k^z \in r_j^z} w_k^z = 0, \quad \bar{r}_j^z = \bar{r}_1^L, \ldots \bar{r}_m^L,$$
$$\bar{r}_{m+1}^R, \ldots \bar{r}_{n'}^R \quad (24)$$

Then we need to prove that the result also holds for the cases that more than N (or N') quads lie to the left of c_i^L (or c_i^R). Due to the space limit, we prove for two cases.

In the first case, $N + 1$ quads $q_1^L \ldots q_{N+1}^L$ lie to the left of c_i^L, where q_{N+1}^L are grouped into a paired region r_{n+1}^L; $N' + 1$ quads $q_1^R \ldots q_{N'+1}^R$ lie to the left of c_i^R, where $q_{N'+1}^R$ are grouped into r_{n+1}^R. In this case, the depth distortion Δd_i of $\mathbf{c}$ is calculated by

$$\Delta d_i = \left| \sum_{j=1}^{n+1} \sum_{q_k^z \in r_j^z} \left(\Phi(\tilde{\mathbf{v}}_k^R) - w_k^R - (\Phi(\tilde{\mathbf{v}}_k^L) - w_k^L) \right) \right.$$
$$\left. + \sum_{j=1}^{n'} \rho_j^z \left(\sum_{q_k^z \in r_j^z} \Phi(\tilde{\mathbf{v}}_k^z) - \sum_{q_k^z \in r_j^z} w_k^z \right) \right| \quad (25)$$

Plugging (24) into (25), we have

$$\Delta d_i = \left| \left(\sum_{q_k^z \in r_{n+1}^z} \left(\Phi(\tilde{\mathbf{v}}_k^R) - w_k^R - (\Phi(\tilde{\mathbf{v}}_k^L) - w_k^L) \right) \right) \right|. \quad (26)$$

Since $\Delta d_i = 0$, if $\Phi(\tilde{\mathbf{v}}_k^R) - w_k^R = (\Phi(\tilde{\mathbf{v}}_k^L) - w_k^L)$ for $q_k^z \in r_{n+1}^z$, the result holds for the first case.

The second case is that quads $q_1^L \ldots q_N^L$ lie to the left of c_i^L, and $q_1^R \ldots q_{N'}^R \ldots q_{N'+a}^R$ lie to the left of c_i^R, where $q_{N'+1}^R \ldots q_{N'+a}^R$ are grouped into $\bar{r}_{n'+1}^R$. Similarly, according to (24), the depth distortion Δd_i of $\mathbf{c}_i$ can be calculated by

$$\Delta d_i = \left| \left(\sum_{q_k^z \in \bar{r}_{n'+1}^R} \Phi(\tilde{\mathbf{v}}_k^z) - \sum_{q_k^z \in \bar{r}_{n'+1}^R} w_k^z \right) \right| \quad (27)$$

$\Delta d_i = 0$, if $\sum_{q_k^L \in \bar{r}_{n'+1}^R} \Phi(\tilde{\mathbf{v}}_k^z) = \sum_{q_k^L \in \bar{r}_{n'+1}^R} w_k^z$.

Therefore, depth preservation is achieved for all correspondences.

REFERENCES

[1] A. Shamir and O. Sorkine, "Visual media retargeting," in *Proc. ACM SIGGRAPH ASIA Courses*, 2009, Art. ID 11.

[2] H. Liu, X. Xie, W.-Y. Ma, and H.-J. Zhang, "Automatic browsing of large pictures on mobile devices," in *Proc. 11th ACM Int. Conf. Multimedia*, 2003, pp. 148–155.

[3] A. Santella, M. Agrawala, D. DeCarlo, D. Salesin, and M. Cohen, "Gaze-based interaction for semi-automatic photo cropping," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2006, pp. 771–780.

[4] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," *ACM Trans. Graph.*, vol. 26, no. 3, Jul. 2007, Art. ID 10.

[5] A. Mansfield, P. Gehler, L. Van Gool, and C. Rother, "Scene carving: Scene consistent image retargeting," in *Proc. 11th Eur. Conf. Comput. Vis.*, 2010, pp. 143–156.

[6] 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, 2009.

[7] 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, 2008, Art. ID 118.

[8] S.-S. Lin, I.-C. Yeh, C.-H. Lin, and T.-Y. Lee, "Patch-based image warping for content-aware retargeting," *IEEE Trans. Multimedia*, vol. 15, no. 2, pp. 359–368, Feb. 2013.

[9] Y.-Y. Chuang and C.-H. Chang, "A line-structure-preserving approach to image resizing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recongnit.*, Jun. 2012, pp. 1075–1082.

[10] B. Li, L.-Y. Duan, J. Wang, J. Chen, R. Ji, and W. Gao, "Grid-based retargeting with transformation consistency smoothing," in *Proc. 17th Int. Multimedia Modeling Conf.*, 2011, pp. 12–24.

[11] B. Li, Y. Chen, J. Wang, L.-Y. Duan, and W. Gao, "Fast retargeting with adaptive grid optimization," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2011, pp. 1–4.

[12] C.-C. Hsu, C.-W. Lin, Y. Fang, and W. Lin, "Objective quality assessment for image retargeting based on perceptual geometric distortion and information loss," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 3, pp. 377–389, Jun. 2014.

[13] Y. Fang, K. Zeng, Z. Wang, W. Lin, Z. Fang, and C.-W. Lin, "Objective quality assessment for image retargeting based on structural similarity," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 4, no. 1, pp. 95–105, Mar. 2014.

[14] Y. Niu, F. Liu, W.-C. Feng, and H. Jin, "Aesthetics-based stereoscopic photo cropping for heterogeneous displays," *IEEE Trans. Multimedia*, vol. 14, no. 3, pp. 783–796, Jun. 2012.

[15] T. D. Basha, Y. Moses, and S. Avidan, "Stereo seam carving a geometrically consistent approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 10, pp. 2513–2525, Oct. 2013.

[16] K. Utsugi, T. Shibahara, T. Koike, K. Takahashi, and T. Naemura, "Seam carving for stereo images," in *Proc. 3DTV-Conf.*, 2010, pp. 1–4.

[17] M. Rubinstein, A. Shamir, and S. Avidan, "Improved seam carving for video retargeting," *ACM Trans. Graph.*, vol. 27, no. 3, 2008, Art. ID 16.

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

[19] J. W. Yoo, S. Yea, and I. K. Park, "Content-driven retargeting of stereoscopic images," *IEEE Signal Process. Lett.*, vol. 20, no. 5, pp. 519–522, May 2013.

[20] K.-Y. Lee, C.-D. Chung, and Y.-Y. Chuang, "Scene warping: Layer-based stereoscopic image resizing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recongnit.*, Jun. 2012, pp. 49–56.

[21] 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, 2010, Art. ID 75.

[22] T. Yan, R. W. H. Lau, Y. Xu, and L. Huang, "Depth mapping for stereoscopic videos," *Int. J. Comput. Vis.*, vol. 102, nos. 1–3, pp. 293–307, 2013.

[23] C. Wu, C.-T. Li, Y.-C. Lai, C.-C. Cheng, and L.-G. Chen, "Disparity remapping by nonlinear perceptual discrimination," in *Proc. 3rd Int. Conf. 3D Syst. Appl.*, 2011, pp. 343–346.

[24] F. Devernay and S. Duchêne, "New view synthesis for stereo cinema by hybrid disparity remapping," in *Proc. 17th IEEE ICIP*, Sep. 2010, pp. 5–8.

[25] F. Shao, W. Lin, S. Gu, G. Jiang, and T. Srikanthan, "Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1940–1953, May 2013.

[26] B. Mendiburu, *3D Movie Making: Stereoscopic Digital Cinema From Script to Screen*. New York, NY, USA: Taylor & Francis, 2009.

[27] B. Li, L. Duan, C.-W. Lin, and W. Gao, "Region-based depth-preserving stereoscopic image retargeting," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2014, pp. 2903–2907.

[28] B. Li, L.-Y. Duan, J. Wang, R. Ji, C.-W. Lin, and W. Gao, "Spatiotemporal grid flow for video retargeting," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1615–1628, Apr. 2014.

[29] H. Hirschmüller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, Feb. 2008.

[30] S.-S. Lin, C.-H. Lin, S.-H. Chang, and T.-Y. Lee, "Object-coherence warping for stereoscopic image retargeting," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 5, pp. 759–768, May 2014.

[31] J. Nocedal and S. Wright, *Numerical Optimization*, 2nd ed. New York, NY, USA: Springer-Verlag, 2006.

[32] (2014). *The Project of Depth-Preserving Warping for Stereo Image Retargeting*. [Online]. Available: http://www.ee.nthu.edu.tw/cwlin/retargeting_stereo/retargeting.html

[33] H. Hirschmüller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recongnit.*, Jun. 2007, pp. 1–8.

[34] Y.-H. Lin and J.-L. Wu, "Quality assessment of stereoscopic 3D image compression by binocular integration behaviors," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1527–1542, Apr. 2014.

[35] *Subjective Methods for the Assessment of Stereoscopic 3DTV Systems*, document ITU-R BT.2021, May 2012.

[36] D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, "Vergence–accommodation conflicts hinder visual performance and cause visual fatigue," *J. Vis.*, vol. 8, p. 33, Mar. 2008.

[37] H. E. Tasli and A. A. Alatan, "User assisted disparity remapping for stereo images," *Signal Process., Image Commun.*, vol. 28, no. 10, pp. 1374–1389, 2013.

Bing Li (S'14) received the B.S. degree in computer science from Jinan University, Guangzhou, China, in 2009. She is currently pursuing the Ph.D. degree with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. Her research interests is image/video processing.

Ling-Yu Duan (M'06) received the M.Sc. degree in automation from the University of Science and Technology of China, Hefei, China, in 1999, the M.Sc. degree in computer science from the National University of Singapore, Singapore, in 2002, and the Ph.D. degree in information technology from The University of Newcastle, Australia, in 2007. He was a Research Scientist with the Institute for Infocomm Research, Singapore, from 2003 to 2008. Since 2008, he has been with Peking University, Beijing, China, where he is currently an Associate Professor with the School of Electrical Engineering and Computer Science. He is leading the group of visual search with the Institute of Digital Media, Peking University. Since 2012, he has been the Deputy Director of the Rapid-Rich Object Search Laboratory, a joint laboratory between Nanyang Technological University, Singapore, and Peking University, China, with a vision to create the largest collection of structured domain object database in Asia and to develop rapid and rich object mobile search. He has authored over 90 publications. His interests are in the areas of visual search and augmented reality, multimedia content analysis, and mobile media computing.
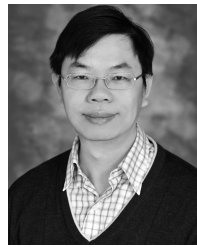
Chia-Wen Lin (S'94–M'00–SM'04) received the Ph.D. degree in electrical engineering from National Tsing Hua University (NTHU), Hsinchu, Taiwan, in 2000. He was with Information and Communications Research Laboratories, Industrial Technology Research Institute, Hsinchu, from 1992 to 2000. He was with the Department of Computer Science and Information Engineering, National Chung Cheng University, Taiwan, from 2000 to 2007. He is currently a Professor with the Department of Electrical Engineering and the Institute of Communications Engineering, NTHU. He is also an Adjunct Professor with the Department of Computer Science and Information Engineering, Asia University, Taichung, Taiwan. His research interests include image/video processing and video networking. He has served as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE MULTIMEDIA, and the *Journal of Visual Communication and Image Representation*. He is also an Area Editor of *EURASIP Signal Processing: Image Communication*. He also serves as a Steering Committee Member of the IEEE TRANSACTIONS ON MULTIMEDIA. He is currently the Chair of the Multimedia Systems and Applications Technical Committee of the IEEE Circuits and Systems Society. He served as the Technical Program Co-Chair of the IEEE International Conference on Multimedia and Expo in 2010, and the Special Session Co-Chair of the IEEE International Conference on Multimedia and Expo in 2009. He is also a Distinguished Lecturer of Asia-Pacific Signal and Information Processing Association. His paper won the Young Investigator Award presented by Visual Communications and Image Processing in 2005. He received the Young Investigator Award presented by the National Science Council of Taiwan in 2006.

Tiejun Huang (M'01–SM'12) He received the bachelor's and master's degrees in computer science from the Wuhan University of Technology, in 1992 and 1995, respectively, and the Ph.D. degree in pattern recognition and intelligent system from the Huazhong (Central China) University of Science and Technology, in 1998. He is currently a Professor of the School of Electronic Engineering and Computer Science, the Chair of Department of Computer Science, and the Director of the Institute for Digital Media Technology with Peking University. His research areas include video coding and image understanding. He is a member of the Board of the Chinese Institute of Electronics, the Board of Directors for Digital Media Project, and the Advisory Board of IEEE Computing. He received the National Science Fund for Distinguished Young Scholars of China in 2014.

Wen Gao (M'92–SM'05–F'09) received the Ph.D. degree in electronics engineering from the University of Tokyo, Japan, in 1991. He was a Professor with the Harbin Institute of Technology, from 1991 to 1995, and the Institute of Computing Technology, Chinese Academy of Sciences, from 1996 to 2006. He is currently a Professor of Computer Science with Peking University. He has authored extensively, including five books and over 600 technical articles in refereed journals and conference proceedings in image processing, video coding and communication, computer vision, multimedia retrieval, multimodal interface, and bioinformatics. He is a member of the Chinese Academy of Engineering, and a fellow of the Association for Computing Machinery (ACM). He was a Chair of a number of prestigious international conferences on multimedia and video signal processing, such as the IEEE International Conference on Multimedia and Expo in 2007, the ACM International Conference on Multimedia in 2009, and the IEEE International Symposium on Circuits and Systems in 2013, and also served on the Advisory and Technical Committee of numerous professional organizations. He served or serves on the Editorial Board of several journals, such as the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE TRANSACTIONS ON AUTONOMOUS MENTAL DEVELOPMENT, the *EURASIP Journal of Image Communications*, the *Journal of Visual Communication*, and *Image Representation*.