High Quality Disparity Remapping with Two-Stage Warping

Bing Li[1], Chia-Wen Lin[2], Cheng Zheng[1], Shan Liu[3], Junsong Yuan[4], Bernard Ghanem[1], C.-C. Jay Kuo[5]
[1]King Abdullah University of Science and Technology (KAUST), Saudi Arabia
[2] Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan
[3]Tencent Media Lab, Palo Alto, USA
[4]University at Buffalo, NY, USA
[5]University of Southern California, Los Angeles, California, USA

Abstract

A high quality disparity remapping method that preserves 2D shapes and 3D structures, and adjusts disparities of important objects in stereo image pairs is proposed. It is formulated as a constrained optimization problem, whose solution is challenging, since we need to meet multiple requirements of disparity remapping simultaneously. The one-stage optimization process either degrades the quality of important objects or introduces serious distortions in background regions. To address this challenge, we propose a two-stage warping process to solve it. In the first stage, we develop a warping model that finds the optimal warping grids for important objects to fulfill multiple requirements of disparity remapping. In the second stage, we derive another warping model to refine warping results in less important regions by eliminating serious distortions in shape, disparity and 3D structure. The superior performance of the proposed method is demonstrated by experimental results.

1. Introduction

Stereoscopic 3D visual contents have been popular nowadays. With recent hardware development, stereo visual content can be viewed in various environments with different viewing conditions (*e.g.* theaters, computers and VR devices). The disparity (or depth) dimension of stereo images often introduces unsatisfactory viewing experience [9, 29]. For example, a pair of stereo images with a large disparity range may yield a strong depth effect on a large screen, which may exceed the comfort zone of the human visual system. In contrast, stereo images with a small disparity range tend to exhibit a small depth effect, leading to poor 3D viewing experience. Consequently, it is important to develop *disparity remapping* tools that can adjust the disparity to meet various viewing conditions and adjust the disparity according to the viewer's preference.

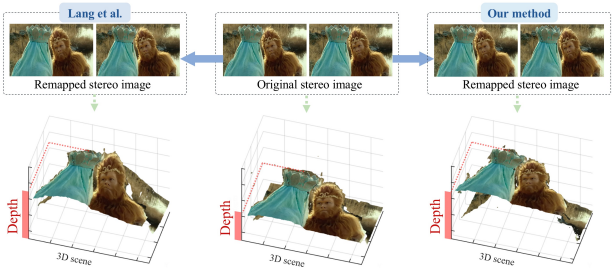


Figure 1: Illustration of the advantage of our method. The disparities of the man and monkey are decreased to enhance their depth strength. The state-of-the-art method [11] fails to preserve the 3D structure of monkey. Our method faithfully adjusts the disparities while preserving the 3D structure well. Original frame is from *The Monkey King 2*.

Being different from 2D images, stereo images offer virtual 3D scenes to viewers through the inherent disparity information. In a 3D scene, an object has not only a spatial shape projected on the x-y plane but also a geometric structure (*i.e.* 3D structure) along the depth direction. This structure information is crucial to human viewing experience. For example, given an object (see Figure 1), the 3D structure determines whether the object is parallel with the display screen or tilted in the 3D scene. To enhance the 3D viewing experience, we aim to design a disparity remapping method that adjusts the disparity, while preserving the 3D structure for important objects.

Lang *et al.* [11] adopt content-aware warping for disparity remapping. Image regions are non-uniformly warped to preserve the shapes of important objects and the disparity is adjusted by constraining the disparity values of a few correspondence pairs to be identical to their target values. Different from classical disparity remapping methods, warping-based methods [11][38] do not need to resort to image completion to fill in holes with synthesized content. On the other hand, they often distort the 3D structure of an object, since

they do not have a proper mechanism to directly preserve this structure. The set of sparse correspondences and/or the spatial shape information cannot characterize the structural information of 3D objects well. As shown in Figure 1, for a 3D object whose surface is parallel with the display screen, its flat surface is bent when the method in [11] is applied.

In this paper, we propose a novel warping-based method for disparity remapping, which adjusts disparity while preserving 2D shape and 3D structure for important objects. We derive a 3D structure-preserving constraints by explicitly taking the 3D structure information into account. With these constraints, we formulate disparity remapping as a constrained optimization problem. That is, we minimize the total distortion of 2D shapes, 3D structures, and disparities of important objects in a stereo image pair.

Besides 3D-structure-preserving constraints, we propose a two-stage warping for disparity remapping, different from existing methods that find the optimal solution for the whole stereo image pair in one stage. In particular, since we have to meet multiple requirements of disparity remapping simultaneously, solving this constrained problem is challenging by itself. The obtained results either degrade the performance in important objects or introduce serious distortions in background regions using only one-stage warping. To address this challenge, we build two warping models and search for the optimal solution at two stages. In the first stage, we develop a warping model that finds the optimal warping grids for important objects to fulfill multiple requirements of disparity remapping. In the second stage, we derive another warping model that is used to refine warping results in less important regions by eliminating serious distortions in shape, disparity and 3D structure. Compared with one-stage warping, our two-stage warping is more effective in finding optimal grids meeting multiple requirements and generates better warping results.

Contributions: There are four major contributions of this work. **(1)** We propose a new energy function to preserve the 3D structures of objects explicitly. **(2)** By decomposing disparity remapping into a two-stage optimization problem, our method can meet multiple requirements of disparity remapping simultaneously, significantly outperforming existing methods that use one-stage optimization and often degrade disparity adjustment performance on important objects. **(3)** To the best of our knowledge, this is the first work that uses the vertex-based warping to refine disparity remapping results of axis-aligned warping locally. With axis-aligned warping as initialization, we speed up the convergence of the algorithm and mitigate distortions. **(4)** It is easy to incorporate additional requirements (e.g., object size adjustment and temporal consistency) in our method, which can facilitate future stereo visual editing tasks that find applications in creating highly realistic 3D structures and special effects.

2. Related Work

Stereo image retargeting has been widely used to resize stereo images to different sizes and aspect ratios. Several retargeting methods were proposed to preserve the depth information for stereo image [3, 12] [15, 39, 1, 14]. The methods in [3, 12, 15, 39] extended the warping-based 2D image retargeting methods [5, 19, 40] to stereo images and added a depth preservation constraint to preserve the depth. As compared with depth-preserving retargeting, disparity remapping adjusts the depth information of stereo images to meet new requirements due to display devices and/or viewer's preference. The additional requirements make the optimization problem more difficult, leading to more severe distortion in 3D object and scene structures. In other words, disparity remapping is a more demanding problem than disparity-preserving retargeting.

Disparity remapping approaches [37, 20, 11, 3, 10, 33, 18] can be roughly categorized into three types: view interpolation methods, shifting-based methods, and warping-based methods. View interpolation methods [22, 43, 42] treat disparity remapping as a view interpolation problem. They rely on accurate camera parameters to calculate a depth map with respect to a virtual camera and, then, interpolate a new view for the virtual camera. Usually, the new view demands post-processing such as image completion to fill in dis-occluded regions (*i.e.* holes), which is computationally demanding. It is also not easy to estimate camera parameters accurately for most stereo images. Recently, some methods [42, 4] proposed to extrapolate high-quality views via neural networks, which, however, require a large labeled training dataset with camera parameters.

Shifting-based methods [26, 31, 13] first select objects by interactive segmentation tools. Then, the selected objects are horizontally shifted so that its disparity is identical to a target value. Object shifting yields dis-occluded regions in edited stereo image pairs and image completion is required to fill holes. Moreover, segmentation errors affect the quality of disparity remapping results.

Warping-based methods modify the disparity by non-uniformly warping image regions. Unlike view interpolation methods which can only adjust the disparity to a virtual camera, warping-based methods can flexibly adjust the disparity to various values. Lang *et al*. [11] were the first to adopt image warping for disparity remapping and to devise disparity constraints for warping. These constraints adjust the disparity by demanding the disparity values of a few corresponding pairs to be identical to target values. Chang *et al*. [3] improved upon this technique by refining the shape-preserving constraints for linear disparity remapping. The disparity constraints in [11] are employed by later work with additional requirements, *e.g*. stereo video remapping [38], visual discomfort reduction [30], and size and depth adjustment [13]. Since warping yields contin-

uous change of visual content, these methods [11, 3, 38] do not yield holes. However, these disparity constraints do not consider the 3D structure information of the stereo image explicitly. This often results in severe 3D structure distortions in edited stereo images. Besides, vertex-based warping usually has difficulty in meeting multiple requirements of disparity remapping in [11, 13, 38]. Differently, we adopt axis-aligned warping and combine it with vertex-based warping in a novel manner to generate high-quality remapping images.

3. Method

Problem Formulation. Given stereo image pair $\{I^L, I^R\}$, disparity remapping is to generate a high-quality stereo image pair $\{\tilde{I}^L, \tilde{I}^R\}$, where their current disparities are adjusted to target ones, while maximizing a viewer's 3D experience. We first partition each of the input stereo pair into grid meshes. Then, the remapped images can be obtained by finding optimal warped meshes that satisfy the requirements of disparity remapping. Therefore, we formulate the disparity remapping as a minimization problem of the following energy function:

$$E = E^G + \alpha E^S + \beta E^D, \tag{1}$$

where E^G is the energy for 3D structure preservation in the left image I^L and the right image I^R, E^S is the energy of spatial shape distortion, E^D is the energy of the difference between the disparity of remapped stereo images and its target value, and α and β are two weighting factors. The derivation of E^G, E^S and E^D will be elaborated later.

Minimizing the total energy function E in Eq. (1) has two challenges. First, how to preserve the 3D structure of objects for disparity remapping? This have not been explored yet, to the best of our knowledge. Second, one possible solution to Eq. (1) is to apply vertex-based warping like existing methods. However, since Eq. (1) aims to meet multiple requirements simultaneously (shape preservation, disparity adjustment, and 3D structure preservation), it is difficult to find a high quality warped grid mesh based on vertex-based warping alone. To address this challenges, we propose a new energy function for 3D structure preserving in Sec. 3. We also propose a novel two-stage optimization algorithm in Sec. 3.3.

3.1. 3D Structure-Preserving Constraint

The 3D-structure distortion energy, E^G, is derived in this subsection and it is a contribution of this paper. To preserve the 3D structures of objects, we first construct the 3D scene of warped images and, then, investigate what warping constraints can preserve the 3D structure of objects. The following constraint is found to be most effective:

C: *Corresponding pixels/regions are consistently warped between the resulting left and right image.*

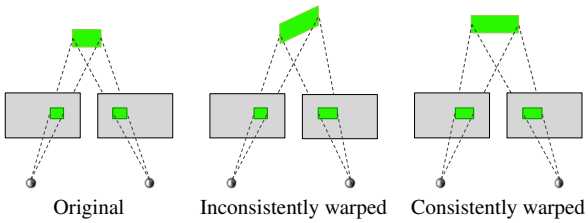


Figure 2: Illustration of the influence of inconsistent warping on the 3D structure.

We show that this constraint is sufficient in preserving the 3D structure in the **supplemental material**. An illustrative example is given in Figure 2. The structure of a green patch is a flat surface in the original 3D scene. Inconsistent warping of the corresponding green patches in the left and right images distorts the flat surface and makes it slanted. In contrast, consistent warping preserves its flat structure.

Based on constraint **C**, a 3D-structure-preserving cost function, E^G, can be defined. We first define a way to measure warping inconsistency for a pair of corresponding grids. Specifically, given a grid g_k^L in left image I^L, $g_{k'}^R$ is its corresponding grid in right image I^R. The warping inconsistency function between g_k^L and $g_{k'}^R$, $E^G(g_k^L, g_{k'}^R)$, can be measured by the vector difference:

$$E^G(g_k^L, g_{k'}^R) = \sum_{(i,j)\in \mathbf{e}_k} \| \overrightarrow{\tilde{v}_i^L \tilde{v}_j^L} - \overrightarrow{\tilde{v}_i^R \tilde{v}_j^R} \|^2, \tag{2}$$

where $\tilde{v}^z (z \in \{L, R\})$ is a vertex of the warped version of g^z, $\overrightarrow{\tilde{v}_i^z \tilde{v}_j^z}$ is an edge vector from $\tilde{v}_i^z$ to $\tilde{v}_j^z$ in the warped grid, and $\mathbf{e}_k$ is the set of grid edges of g_k^z in the counter-clockwise direction. Then, the 3D structure-preserving energy E^G for a stereo image pair can be defined by summing up warping inconsistency values of all corresponding grids as:

$$E^G = \sum_k E^G(g_k^L, g_{k'}^R) \cdot \lambda_k, \tag{3}$$

where λ_k is the weight used to control the warping inconsistency between g_k^L and $g_{k'}^R$. We set λ_k to the average grid importance of g_k^L and $g_{k'}^R$, where grid importance are defined in Sec. 3.3.

3.2. Shape preservation and Disparity Adjustment

Shape Preservation. For shape preservation, we cannot employ E^S of axis-aligned warping in Eq. (6), since vertex-based warping allows a warped grid to be an arbitrary quadrilateral. We define E^S as the sum of the grid deformation energy in background regions:

$$E^S = \sum_z \sum_{g_k^z \in \mathbf{g}_u} E^S(g_k^z) \cdot \delta_k^z \tag{4}$$

where $\mathbf{g}_u$ be a set that contains grids in background regions of the stereo image, $E^S(g_k^z)$ is the shape distortion of g_k^z.

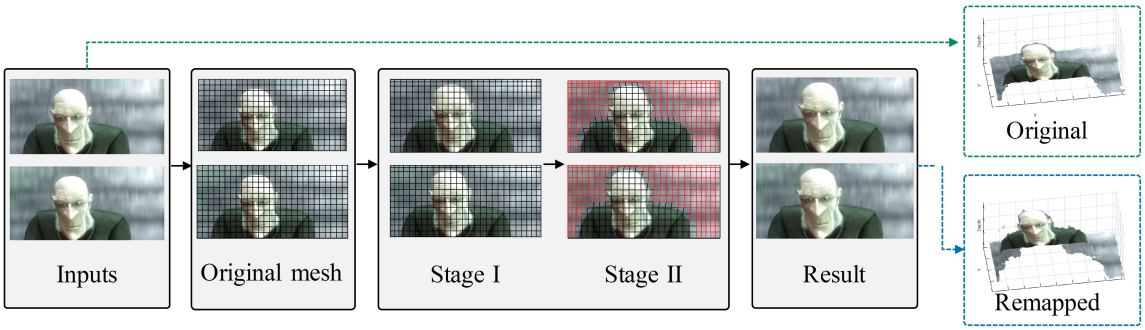


Figure 3: Illustration of the proposed two-stage warping. Stage I finds optimal warped grids for important object, and Stage II then locally adjusts grids on less important regions, where the grids on less important regions are marked in red.

Disparity Adjustment. To meet the disparity requirement, we adjust the horizontal disparity using a few correspondence pairs. The disparity adjustment energy demands the disparity of each correspondence pair in the warped stereo image to be as close as possible to the target value. In particular, the disparities of the correspondence pairs in important regions should be mapped more faithfully. This is different from existing methods [3, 11] that treat all correspondence pairs equally. Let $\tilde{\mathbf{f}}_c = \{\tilde{f}_c^L, \tilde{f}_c^R\}$ denote the c-th pair of corresponding points in the warped left and right image, where f_c^z is a point in I^z, $z \in \{L, R\}\}$. We express E^D as

$$E^D = \sum_c E^D(\tilde{\mathbf{x}}_c, \tilde{d}_c) = \sum_c \|\tilde{x}_c^L - \tilde{x}_c^R - \tilde{d}_c\|^2 \cdot \delta_c^z \quad (5)$$

where $\tilde{d}_c$ is the target disparity value of $\mathbf{f}_c$, $E^D(\tilde{\mathbf{x}}_c, \tilde{d}_c)$ is the disparity adjustment energy of $\tilde{\mathbf{f}}_c$, $\tilde{x}_c^z$ is the x-coordinate of $\tilde{f}_c^z$ in the warped image $\tilde{I}^z$ and $\mathbf{x}_c = \{\tilde{x}_c^L, \tilde{x}_c^R\}$. The value, $\tilde{x}_c^z$, can be represented by grid width $\tilde{w}_k^z$. Nonzero vertical disparities often introduce 3D fatigue or eye strain to viewers [21, 23]. To eliminate the vertical disparity, we constrain warped grids lying at the same row to have the same height between left and right images.

3.3. Two-Stage Warping Algorithm

Existing methods [11][38][13] resort to vertex-based warping for finding optimal warped meshes. However, directly employing vertex-based warping has two *shortcomings*. First, its optimization problem is non-convex. Second, since vertex-warped allows the shape of a warped grid to be an arbitrary quadrilateral, its high degree of freedom can introduce unsatisfactory foldover (*i.e.* self-intersection of grids) or shape distortion of structured objects [25] (see Figure 4). To address this issue, complex or even non-convex constraints can be used [5, 7, 36, 28]. However, it is difficult to optimize all warped vertices, coupled with complex constraints, to meet all energy constraints simultaneously, thus degrading the performance of disparity remapping.

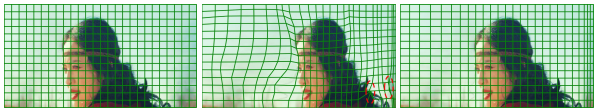


Figure 4: Warped meshes by vertex-based warping and axis-aligned warping, respectively. Vertex-based warping introduces foldovers (marked in red dash ellipse), whereas axis-aligned warping sacrifices the flexibility of local transformation. Original frame is from movie Chronicles, ©CFGC.

Our insight is that we can combine vertex-based warping with axis-aligned warping to complement their shortcomings for disparity remapping, which have not yet been explored by existing methods. Axis-aligned warping constrains all warped grids to be rectangular, which has proven to be robust and effective for content-aware retargeting [25, 17] in two aspects. First, axis-aligned warping reduces the degree of freedom for transformation and eliminates foldovers without imposing additional complex constraints. Second, its optimization problem is convex and is of low computational complexity. However, its *shortcoming* lies in that simplification in transformation tends to sacrifice the flexibility of local transformation, yielding severe distortions of 3D structure and disparity in less important regions for disparity remapping.

We argue that axis-aligned warping can more easily fulfill all the constraints imposed on the warping of important objects, thanks to its foldover-free transformation and convex optimization. On the other hand, due to the simplification in transformation, axis-aligned warping tends to introduce distortions in the less important regions. We can exploit the flexibility of vertex-based warping to refine warped grids in these regions to mitigate the distortions.

The major challenge is that we cannot simply combine vertex-based warping and axis-aligned warping into a single-step optimization, since they would conflict with each other. In particular, axis-aligned warping strictly constrains all warped grids at the same row/column to have the

same height/width (see Figure 4). Such hard constraints do not allow vertex-based warping to locally rotate/scale each grid edge. To tackle the challenge, we propose a two-stage optimization algorithm which finds optimal meshes in two stages (see Figure 3). In the first stage, it adopts an axis-aligned warping to optimize all grids of the meshes to ensure that they fulfill all requirements of disparity remapping for those important objects identified by an importance map following [11]. In the second stage, it uses a vertex-based warping to locally refine warped meshes on less important regions to alleviate the distortions on these regions.

Stage I: Axis-aligned Warping

The goal of this stage is to obtain warped grid meshes by optimizing the grid width and height with a computationally efficient axis-aligned warping. Given grid $g^z_{i,j}$, let w^z_k and h^z_k be its original grid width and height, respectively, and $\tilde{w}^z_k$ and $\tilde{h}^z_k$ be the warped version. To find warped meshes that minimize E in Eq. (1), we formulate energy functions E^S, E^D and E^G in terms of $\tilde{w}^z_k$ and $\tilde{h}^z_k$.

To preserve the shapes of important objects, the shape preservation energy, E^S, is defined as the weighted sum of all grid distortions [16, 25]. Given grid g^z_k, its distortion energy is the difference between the aspect ratio of its warped version and that of original one since the shape of a warped grid is rectangular. Hence, E^S can be expressed as

$$E^S = \sum_{z\in\{L,R\}} \sum_{g^z_k} \|w^z_k \cdot \tilde{h}^z_k - \tilde{w}^z_k \cdot h^z_k\|^2 \cdot \delta^z_k, \quad (6)$$

where δ^z_k is grid importance of g^z_k. We calculate δ^z_k by summing up pixel importance in g^z_k, where pixel importance is defined by importance map. The importance map is defined as the weighted sum of the image saliency map [41] and the disparity saliency map[8], following[11].

To preserve 3D structure, we represent E^G in Eq. (3) in terms of $\tilde{w}^z_k$ and $\tilde{h}^z_k$. In particular, since the shape of a warped grid is rectangular, we have $\overrightarrow{\tilde{v}^z_i \tilde{v}^z_j} = [\tilde{w}^z_k, 0]$ in Eq. (2), if $\overrightarrow{\tilde{v}^z_i \tilde{v}^z_j}$ is a bottom/top edge of grid g^z_k. Otherwise, we get $\overrightarrow{\tilde{v}^z_i \tilde{v}^z_j} = [0, \tilde{h}^z_k]$. For each grid g^L_k, we determine its corresponding grids in I^R according to the disparity map.

To find the first-stage warped meshes, we minimize E in Eq. (1) that contains terms in Eqs. (3), (6) and (5), subject to the boundary constraints as proposed in [25]. This is a convex quadratic programming problem, and we can find a global optimal solution for E using the active-set algorithm [24] for instance.

Stage II: Vertex Warping

In this stage, we adopt a vertex-based warping to refine warped grid meshes obtained in the first stage. Since the stage I has already well optimized the grids on important objects, we retain the vertex coordinates of these warped grids, and thereby only optimize grids in regions which do not contain important objects. We refer to these regions as *background regions*, and $\tilde{\mathbf{V}}_u = \{\tilde{v}_i\}$ denotes the set of warped grids' vertexes on these regions. Here, vertex-based warping is to find the optimal positions for vertexes in $\tilde{\mathbf{V}}_u$.

To begin with, we need to define E^S, E^D and E^G in Eq. (1) corresponding to shape preservation, disparity adjustment and 3D structure preservation for $\tilde{\mathbf{V}}_u$.

Given g^z_k, $E^S(g^z_k)$ is measured by the deviation of its warping from the similarity transformation. Specifically, $E^S(g^z_k)$ is defined as the difference between each warped edge and its similarity transformation version, by following [36, 12]:

$$E^S(g^z_k) = \sum_{(i,j)\in\mathbf{e}_k} \| (\tilde{v}^z_i - \tilde{v}^z_j) - s^z_k(v^z_i - v^z_j) \|^2, \quad (7)$$

where $\mathbf{e}_k$ is the set containing the four edges of g^z_k, $s^z_k = \frac{\sum_{(i,j)\in\mathbf{e}_k} (\tilde{v}^z_i - \tilde{v}^z_j)^T (v^z_i - v^z_j)}{\sum_{(i,j)\in\mathbf{e}_k} \|v^z_i - v^z_j\|^2}$.

For E^D, we minimize the disparity adjustment energy given in Eq. (5) for correspondence pairs in background regions. Accordingly, the coordinates of $\mathbf{f}_k$ is also represented in terms of $\tilde{\mathbf{V}}_u$ via barycentric coordinates. Similarly, for E^G, we employ the 3D-structure preservation energy in Eq. (2) and minimize the energy at grids belonging to background regions only. By following [34, 38], we also use the line bending energy, to avoid serious grid transformation (e.g., foldover, structural object distortions)

We minimize Eq. (1) to obtain optimal positions for $\tilde{\mathbf{V}}_u$. This optimization problem is a non-convex one. It is solved as an iterative least squares problem. Note that we use the optimized meshes obtained in the first stage as the initial guess, which speeds up convergence. Moreover, compared with optimization of vertices for all grids [11, 38], the number of parameters in our vertex-based warping is significantly less, since only grid vertices in background regions are optimized. Thus, our optimization coverages faster and is more efficient than that of [11, 38].

4. Extending to Stereo Video

We can extend our method to stereo video with temporal constraints. By adding temporal constraints into the stage I and II, respectively, we can ensure temporal coherence for remapped stereo video. In particular, we first align corresponding grids among frames via motion estimation algorithm (e.g. [2, 32]), for the left/right video of a stereo video. Then, we build temporal constraints to constrain aligned grids to be consistently warped among frames at the stage I and II. More specifically, the temporal constraint of the stage I is built for axis-aligned warping, which encourages aligned grids to have same width/height. Similarly, the temporal constraint of the stage II is built in term of vertex-based warping, which encourages the warped edges of aligned grids to undergo consistent transformation.

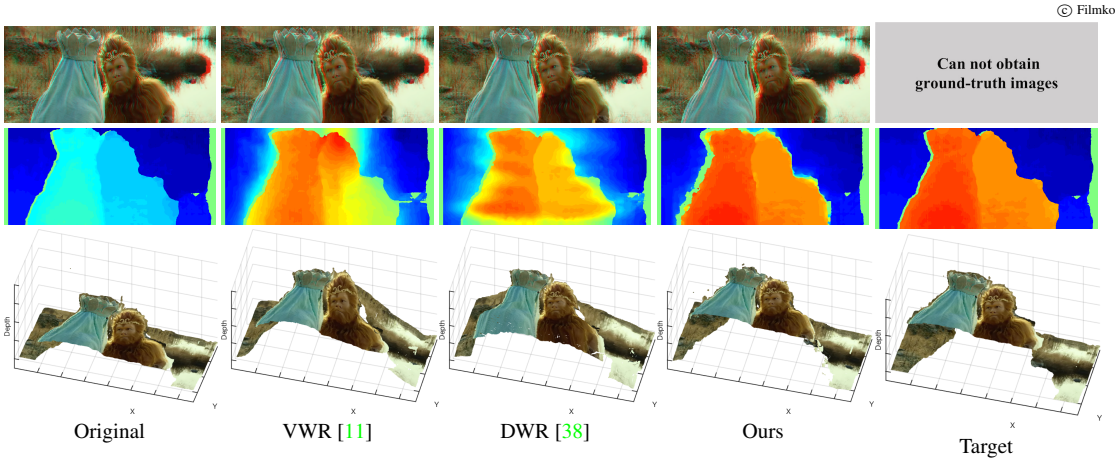


Figure 5: Comparison results on a **stereo image** from *The Monkey King 2*. From top to bottom: red-cyan anaglyph image of the stereo pair, disparity map, and 3D scene reconstructed from the the stereo pair.

5. Experiments

Datasets. We test our method on images with various 3D scene structures collected from the *Flicker* [1] and *Middlebury* datasets [27], both of which are widely adopted for evaluating stereo image editing methods. We also collect stereo images from 3D movies. These images contain objects with irregular shapes and different disparity, posing great challenges to disparity remapping.

We first compare our method with VWR [11] and DWR [38] which are state-of-the-art in warping-based methods. We then compare our method with the size and disparity remapping method (SDR) [13], which jointly adjusts the size and disparity of an object via vertex-based warping. This aims to show the benefit of extending our method with an additional requirement (*i.e*. size adjustment). Please refer to **supplemental materials** for more results.

Disparity remapping on stereo image. A good disparity remapping result faithfully preserves both 2D shapes and 3D structures for important objects, while mitigating the deviations of its disparity values from its target values. To evaluate the performance of disparity adjustment, we build a ground-truth disparity map[1] by manually modifying the disparity values in the original disparity map to their desired target values. The disparity map is calculated by the algorithm in [8], where the darkest red and darkest blue indicates the smallest disparity and the largest disparity, respectively. We reconstruct ground-truth 3D scenes in a similar way, to evaluate the performance of 3D-structure preservation.

As indicated in [11][13], locally adjusting the disparity of a/some objects is much more complex than globally adjusting the disparity of the whole image. We hence set target disparity values by employing the local disparity adjusting function in [11], to challenge our method. Fig. 5 shows the disparity remapping results on a stereo image which contains a man and a monkey. We decrease the disparity of the man and monkey by 280% and 220% respectively, to increase their depth strength. Compared with the ground-truth disparity map in Figure 5, both VWR[11] and DWR[38] fail to adjust the disparity of the man or the monkey to target value. Furthermore, DWR distorts the disparity of other regions (e.g. background). Compared with VWR and DWR, our method achieves best performance in disparity adjustment, due to our two-stage warping. As to 3D structure preservation, both VWR and DVW distort the structure of the monkey's body in the 3D scene. The structure of the man's shoulder is also distorted by DWR. In contrast, even though the 3D structure of the man and monkey is complex, our method well preserves the 3D structure of the whole image, thanks to our 3D-structure preserving energy (see supplemental materials for more results on various stereo images).

Disparity remapping on stereo video. We extend our method to stereo video and test our method on a ***stereo video*** containing a close-up face of a boy. We decrease the disparity of the boy's face by 210%. All methods well preserve spatial shape and temporal coherence (see Fig. VI in supplemental materials). However, VWR and DWR cannot adjust the disparity of the face to the target value. For 3D-structure preservation, VWR and DWR distort the 3D structure of the face which is dissimilar to that in ground-truth 3D scene. In contrast, our method not only preserves the 3D structure of the face well, but also faithfully adjusts its disparity to the target value.

Size and disparity remapping. Figure 6 shows the *size and disparity* adjustment results on stereo image. To compare with [13], we incorporate size adjustment constraints into our method. Following [13], the size of the man is in-

[1]The ground-truth disparity map is only used to indicate the disparity values for objects. In the disparity map of a remapped stereo image, the locations of objects can be different from those in the ground-truth.

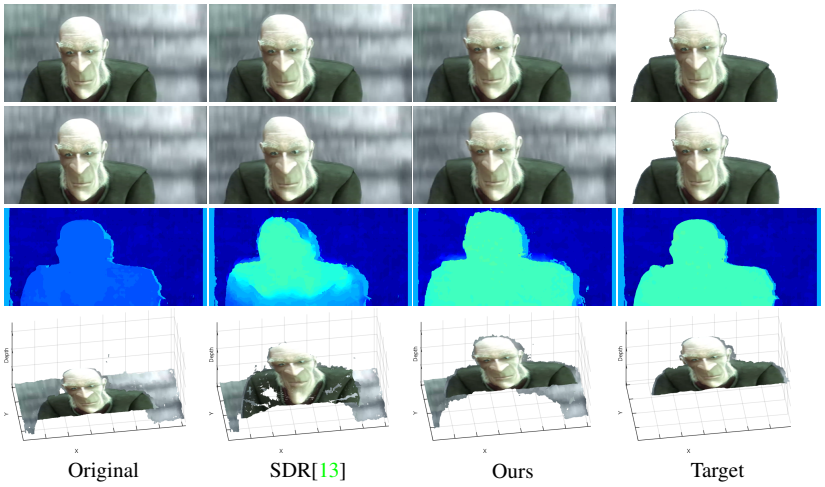


Figure 6: Comparison results of **size and disparity remapping** on a stereo image from movie *Elephants dream*. From top to bottom: left image, right image, disparity map, and 3D scene reconstructed from the left and right images.

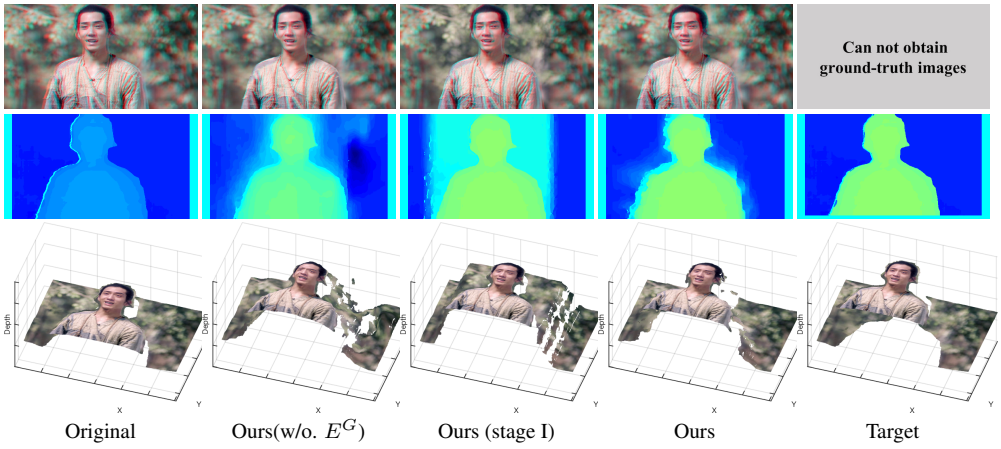


Figure 7: Ablation study of the performance of E^G and two-stage warping on a frame of movie *Monster Hunt*. From top to bottom: red-cyan anaglyph image of the stereo pair, disparity map, and 3D scene reconstructed from the left and right images.

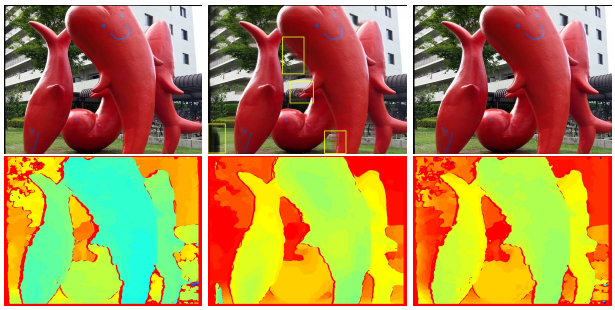
creased by 8%, while his disparity is accordingly increased by 300%. Due to the additional requirement of size adjustment, the task of disparity remapping is much more complex. As a result, SDR inconsistently warps the head of the man between the left and right image, leading that the 3D structure of the man is distorted. In contrast, our method faithfully preserve the man's 3D structure, thanks to our 3D structure preserving constraints. Moreover, our method also adjusts the size and disparity of the man more accurately than SDR. The results validate that our 3D structure-preserving constraints and two-stage optimization are beneficial to disparity remapping with addition requirements like size adjustment.

Quantitative evaluation. We measure shape preservation performance by IR-SSIM [6] which is a metric measuring structural similarity between a non-uniformly resized image and its original image. A higher IR-SSIM value indicates better shape preservation. The average IR-SSIM values of VWR, DWR and our method are high in Table 1, indicating all methods preserve the shape well.

Since 3D structure is non-rigid, it is difficult to measure 3D structure preservation performance. Instead, we examined whether the disparity values of grids belonging to the same foreground/background object are adjusted consistently. This is measured by Kendall's correlation coefficient between the remapped and the target disparity values against the same object. The higher the coefficient is, the better the 3D-structure preservation is. For four stereo images (see Figure IV, V, VI and VII in supplemental materials), Table 1 shows our method significantly outperforms

Table 1: Quantitative evaluation results in terms of shape preservation, 3D structure preservation, and disparity distortion (↑: Higher is better; ↓: Lower is better).

	VWR	DWR	Ours
Shape preservation ↑	0.975	0.980	**0.983**
3D structure preservation ↑	0.654	0.672	**0.876**
Disparity distortion ↓	3.42	3.61	**0.95**



Original SMA [42] Ours

Figure 8: Comparisons of our method and a state-of the-art view interpolation method [42] on a stereo image with camera parameters. The stereo image is originally used in [42]. From top to bottom: left image of the stereo pair, disparity map.

VWR and DWR in 3D-structure preservation.

We evaluate the effectiveness of disparity adjustment by measuring the average disparity distortion. The average disparity distortion calculates the average difference between the disparity values of remapped stereo images and the ground-truth. For four testing stereo images (see supplemental materials), Table 1 shows disparity distortion of our method is significantly lower than that of VWR and DWR.

Ablation study. We first evaluate the influence of our proposed 3D structure-preserving energy E^G. As shown in the first and second columns of Figure 7, the 3D structure of the man is significantly distorted if E^G is removed. Moreover, the average disparity distortion of our method is 0.63, however, is increased to 2.21 when E^G is removed.

We also evaluate the effectiveness of the proposed two-stage warping model. Figure 7 shows the disparity remapping results obtained in two stages, respectively. We can observe that the first-stage (axis-aligned warping) preserves both the 2D shape and 3D structure of *Ant-Man* well, due to the smoothness and robustness offered by axis-aligned warping. However, the first stage also introduces severe distortions in the background. With the second-stage refinement, the distortions in the background are effectively mitigated, due to the flexibility of vertex warping.

User study. A subjective user study was conducted on an ASUS 3D 24-inch monitor using the NVIDIA active shutter glasses. We invited 31 subjects to participate in the subjective evaluation of disparity remapping results. All of them had normal stereo perception. We compared our method with VWR and DWR in a side-by-side pairwise comparison manner [35, 23] on five stereo images. The participants had no prior knowledge about these disparity remapping methods. We placed the ground-truth stereo image in the middle, while randomly placing the two to-be-compared disparity remapped pairs in the left and right sides. We then asked each subject to answer the following question: *Which edited stereo image is of better visual quality, according to the ground truth one ?* In total, 73.5% of subjects preferred our method to VWR, and 71.6% of them preferred our method to DWR, thus, validating that our method offers better 3D viewing experience for disparity remapping.

Comparison with view interpolation method. We compare our method with Stereo Magnification Approach (SMA) [42] which is a state-of-the-art in view interpolation. Different from SMA, our method does not requires camera parameters and training data. As shown in Figure 8, however, SMA [42] introduces holes in the left boundary of the images, which provides unsatisfactory viewing experiences. Moreover, SMA also generates blur artifacts into regions around the red fishes (see regions marked by yellow blocks in Figure 8). The average shape preservation of SMA score (0.88) is lower than that of our method (0.94), showing our results are of better quality.

Computational complexity. Given a 20×30 grid division, the run-time cost of our optimization is 0.3s, while that of [11] is 1.47s averagely on a laptop with a 2.26 GHz Duo CPU.

6. Conclusion

We propose a novel approach for disparity remapping, where besides the desired disparity adjustment, both 3D-structure and 2D shape preservation are considered explicitly. To this end, we propose 3D structure-preserving constraint and integrate the constraints into an energy cost function to formulate a constrained optimization problem. We have also proposed a two-stage warping algorithm consisting of axis-aligned warping followed by vertex-warping based refinement to solve the optimization problem. Experimental results show that our approach can simultaneously preserve 2D shape and 3D structure of important objects well, while accurately adjusting the disparity map to a target one, without introducing noticeable distortions.

Acknowledgments. This work was supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research through the Visual Computing Center (VCC) funding.

References

[1] T. D. Basha, Y. Moses, and S. Avidan. Stereo seam carving a geometrically consistent approach. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35:2513–2525, 2013. 2, 6

[2] T. Brox and J. Malik. Large displacement optical flow: descriptor matching in variational motion estimation. *IEEE Trans. Pattern Anal. Mach. Intelli.*, 33(3):500–513, 2011. 5

[3] C.-H. Chang, C.-K. Liang, and Y.-Y. Chuang. Content-aware display adaptation and interactive editing for stereoscopic images. *IEEE Trans. Multimedia*, 13:589–601, 2011. 2, 3, 4

[4] Inchang Choi, Orazio Gallo, Alejandro Troccoli, Min H Kim, and Jan Kautz. Extreme view synthesis. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 7781–7790, 2019. 2

[5] Y.-Y. Chuang and C.-H. Chang. A line-structure-preserving approach to image resizing. In *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recongnit.*, pages 1075–1082, 2012. 2, 4

[6] 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.*, pages 95–105, 2014. 7

[7] K. He, H. Chang, and J. Sun. Content-aware rotation. In *Proc. IEEE Int. Conf. Comput. Vis.*, pages 553–560, 2013. 4

[8] H. Hirschmüller. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.*, pages 328–341, 2008. 5, 6

[9] D. M Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *J. Vis.*, 8:33, 2008. 1

[10] M. B. Islam, L.-K. Wong, K.-L. Low, and C.-O. Wong. Aesthetics-driven stereoscopic 3d image recomposition with depth adaptation. *IEEE Trans. Multimedia*, 2018. 2

[11] M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic, and M. Gross. Nonlinear disparity mapping for stereoscopic 3d. *ACM Trans. Graph.*, 29(4):75:1–75:10, 2010. 1, 2, 3, 4, 5, 6, 8

[12] K.-Y. Lee, C.-D. Chung, and Y.-Y. Chuang. Scene warping: Layer-based stereoscopic image resizing. In *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recongnit.*, 2012. 2, 5

[13] J. Lei, B. Peng, C. Zhang, X. Mei, X. Cao, X. Fan, and X. Li. Shape-preserving object depth control for stereoscopic images. *IEEE Trans. Circuits Syst. Video Technol.*, 28:3333–3344, 2018. 2, 3, 4, 6, 7

[14] J. Lei, M. Wu, C. Zhang, F. Wu, N. Ling, and C. Hou. Depth-preserving stereo image retargeting based on pixel fusion. *IEEE Trans. Multimedia*, 19:1442–1453, 2017. 2

[15] B. Li, L.-Y. Duan, C.-W. Lin, T. Huang, and W. Gao. Depth-preserving warping for stereo image retargeting. *IEEE Trans. Image Process.*, 24(9):2811–2826, 2015. 2

[16] 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.*, 23(4):1615–1628, 2014. 5

[17] B. Li, C.-W. Lin, B. Shi, T. Huang, W. Gao, and C.-C. J. Kuo. Depth-aware stereo video retargeting. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 6517–6525, 2018. 4

[18] B. Li, C.-W. Lin, C. Zheng, S. Liu, and C.-C. J. Kuo. Stereo depth mapping via axis-aligned warping. In *Proc. IEEE Int. Conf. Image Processing*, pages 4305–4309. IEEE, 2019. 2

[19] S.-S. Lin, I.-C. Yeh, C.-H. Lin, and T.-Y. Lee. Patch-based image warping for content-aware retargeting. *IEEE Tran. Multimedia*, 15(2):359–368, 2013. 2

[20] Belen Masia, Gordon Wetzstein, Carlos Aliaga, Ramesh Raskar, and Diego Gutierrez. Display adaptive 3d content remapping. *Computers & Graphics*, 2013. 2

[21] Bernard Mendiburu. *3D Movie Making: Stereoscopic Digital Cinema from Script to Screen*. Taylor & Francis, 2009. 4

[22] P. Ndjiki-Nya, M. Koppel, D. Doshkov, H. Lakshman, P. Merkle, K. Muller, and T. Wiegand. Depth image-based rendering with advanced texture synthesis for 3-d video. *IEEE Trans. Multimedia*, 13(3):453–465, 2011. 2

[23] Y. Niu, W.-C. Feng, and F. Liu. Enabling warping on stereoscopic images. *ACM Trans. Graph.*, 2012. 4, 8

[24] J. Nocedal and S. Wright. *Numerical Optimization*. Springer, 2nd edition, 2006. 5

[25] D. Panozzo, O. Weber, and O.Sorkine. Robust image retargeting via axis-aligned deformation. *EUROGRAPHICS*, 2012. 4, 5

[26] H. Park, H. Lee, and S. Sull. Efficient viewer-centric depth adjustment based on virtual fronto-parallel planar projection in stereo 3d images. *IEEE Trans. Multimedia*, 16(2):326–336, 2014. 2

[27] D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nešić, X. Wang, and P. Westling. High-resolution stereo datasets with subpixel-accurate ground truth. In *German Conf. Pattern Recognit.*, pages 31–42, 2014. 6

[28] F. Shao, Y. Fei, Q. Jiang, X. Meng, and Y.-S. Ho. Building stereoscopic zoomer via global and local warping optimization. *IEEE Trans. Computational Imaging*, 6:1622–1635, 2020. 4

[29] T. Shibata, J. Kim, D. M. Hoffman, and M. S. Banks. Visual discomfort with stereo displays: effects of viewing distance and direction of vergence-accommodation conflict. In *Stereoscopic Displays and Applications XXII*, volume 7863, 2011. 1

[30] H. Sohn, Y. J. Jung, S.-I. Lee, F. Speranza, and Y. M. Ro. Visual comfort amelioration technique for stereoscopic images: Disparity remapping to mitigate global and local discomfort causes. *Trans. Circuits Syst. Video Technol.*, 24(5):745–758, 2014. 2

[31] H. Emrah Tasli and A. Aydin Alatan. User assisted disparity remapping for stereo images. *Signal Process.: Image Commun.*, 28:1374–1389, 2013. 2

[32] Zachary Teed and Jia Deng. Raft: Recurrent all-pairs field transforms for optical flow. In *ECCV*, 2020. 5

[33] O Wang, M Lang, M Frei, A Hornung, A Smolic, and M Gross. StereoBrush: interactive 2D to 3D conversion using discontinuous warps. In *Proceedings of the Eighth Eurographics Symposium on Sketch-Based Interfaces and Modeling*, pages 47–54, 2011. 2

[34] Y.-S. Wang, H. Fu, O. Sorkine, T.-Y. Lee, and H.-P. Seidel. Motion-aware temporal coherence for video resizing. *ACM Trans. Graph.*, 2009. 5

[35] Y.-S. Wang, H.-C. Lin, O. Sorkine, and T.-Y. Lee. Motion-based video retargeting with optimized crop-and-warp. *ACM Trans. Graph.*, 2010. 8

[36] Y.-S. Wang, C.-L. Tai, O. Sorkine, and T.-Y. Lee. Optimized scale-and-stretch for image resizing. *ACM Trans. Graph.*, 2008. 4, 5

[37] C. Wu, C. Li, Y. C. Lai, C. Cheng, and L. Chen. Disparity remapping by nonlinear perceptual discrimination. In *Proc. Int. Conf. 3D Syst. Appl.*, 2011. 2

[38] T. Yan, R. W. H. Lau, Y. Xu, and L. Huang. Depth mapping for stereoscopic videos. *Int. J. Comput. Vis.*, 102:293–307, 2013. 1, 2, 3, 4, 5, 6

[39] J. W. Yoo, S. Yea, and I. K. Park. Content-driven retargeting of stereoscopic images. *IEEE Signal Process. Lett.*, pages 519–522, 2013. 2

[40] G.-X. Zhang, M.-M. Cheng, S.-M. Hu, and R. R. Martin. A shape-preserving approach to image resizing. *Comput. Graph. Forum*, pages 1897–1906, 2009. 2

[41] J. Zhang, S. Sclaroff, Z. Lin, X. Shen, B. Price, and R. Mech. Minimum barrier salient object detection at 80 fps. In *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recongnit.*, 2015. 5

[42] Tinghui Zhou, Richard Tucker, John Flynn, Graham Fyffe, and Noah Snavely. Stereo magnification: Learning view synthesis using multiplane images. In *SIGGRAPH*, 2018. 2, 8

[43] C L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High-quality video view interpolation using a layered representation. In *ACM transactions on graphics (TOG)*, volume 23, pages 600–608. ACM, 2004. 2