Shape-Preserving Stereo Object Remapping via Object-Consistent Grid Warping

Bing Li, *Member, IEEE*, Chia-Wen Lin, *Fellow, IEEE*, Cheng Zheng, Shan Liu, *Member, IEEE*, Bernard Ghanem, *Member, IEEE*, Wen Gao, *Fellow, IEEE*, and C.-C. Jay Kuo, *Fellow, IEEE*

Abstract—Viewing various stereo images under different viewing conditions has escalated the need for effective object-level remapping techniques. In this paper, we propose a new object spatial mapping scheme, which adjusts the depth and size of the selected object to match user preference and viewing conditions. Existing warping-based methods often distort the shape of important objects or cannot faithfully adjust the depth/size of the selected object due to improper warping such as local rotations. In this paper, by explicitly reducing the transformation freedom degree of warping, we propose an optimization model based on axis-aligned warping for object spatial remapping. The proposed axis-aligned warping based optimization model can simultaneously adjust the depths and sizes of selected objects to their target values without introducing severe shape distortions. Moreover, we propose object consistency constraints to ensure the size/shape of parts inside a selected object to be consistently adjusted. Such constraints improve the size/shape adjustment performance while remaining robust to some extent to incomplete object extraction. Experimental results demonstrate that the proposed method achieves high flexibility and effectiveness in adjusting the size and depth of objects compared with existing methods.

Index Terms—Object remapping, warping, stereoscopic image, image editing, depth adjustment.

I. INTRODUCTION

STEREOSCOPIC 3D content is becoming mainstream in consumer media, as it has been widely adopted by many 3D movies and virtual/augmented-reality (VR/AR) games. The popularity of stereo images/videos calls for convenient visual editing tools, which can manipulate stereo visual content such as object-level depth adjustment, object scaling, image manipulation, and viewing experience enhancement, similar to those for 2D images. Different from 2D images/videos, the 3D viewing experience of stereo images/videos is heavily affected by viewing conditions, owing to the vergence-accommodation conflicts in the human visual system (HVS) [1], [2]. For example, as revealed in [3], [4], stereo visual contents produced for a theater screen would look shallow and rather different on a computer monitor, leading to unsatisfactory viewing experience. Therefore, professional 3D makers use complex softwares (e.g. Ocula [5]) or tools to edit the size and depth for objects of interest in stereo images/videos to fit various viewing conditions. However, this requires 3D movie makers to have expertise on 3D post-production, while the editing tools require high-quality segmentation, inpainting, or camera calibration which are computationally expensive. Therefore, various viewing conditions and viewer preferences posing the need of efficient and convenient editing methods for adjusting the depth/size at the object level.

With the development of interactive image editing tools, viewers can conveniently select and extract objects-of-interest from a stereo image, and then perform object-level editing and manipulation. In this paper, we focus on *object remapping*, which aims to adjust the depth and size of a selected object to their target values in a stereo image pair. Stereo object remapping has been studied for years. Compared with solely adjusting the depth, jointly adjusting the size and depth of an object is more challenging in two aspects. (1) Since both size and depth adjustment would alter the locations of regions in the object, simultaneous depth and size adjustments often severely distort grids, leading to noticeable shape distortions. (2) Since an object usually involves multiple grids, the shapes/sizes of grids associated with an object should be consistently adjusted; otherwise, the object will be deformed.

Among the existing object remapping works, Lei *et al.* [6] was the first to introduce non-uniform warping into object remapping, where the depth and size of a selected object are adjusted by non-uniformly warping the grids associated with the object and its surrounding neighbors. Thanks to the continuity nature of grid warping, this method changes the depth of an object without generating holes. The method, however, tends to inconsistently adjust the depth and size of an object, thereby introducing visually annoying shape distortions as illustrated in Fig. 1. Such shape distortions not only degrade the fidelity of visual content, but also negatively affect depth perception.

To address the above issues, we propose a warping-based object remapping method for faithful object-level size and depth adjustment in a stereo image while avoiding noticeable shape distortions. Specifically, we adopt axis-aligned grid warping to effectively mitigate the shape and depth distortions
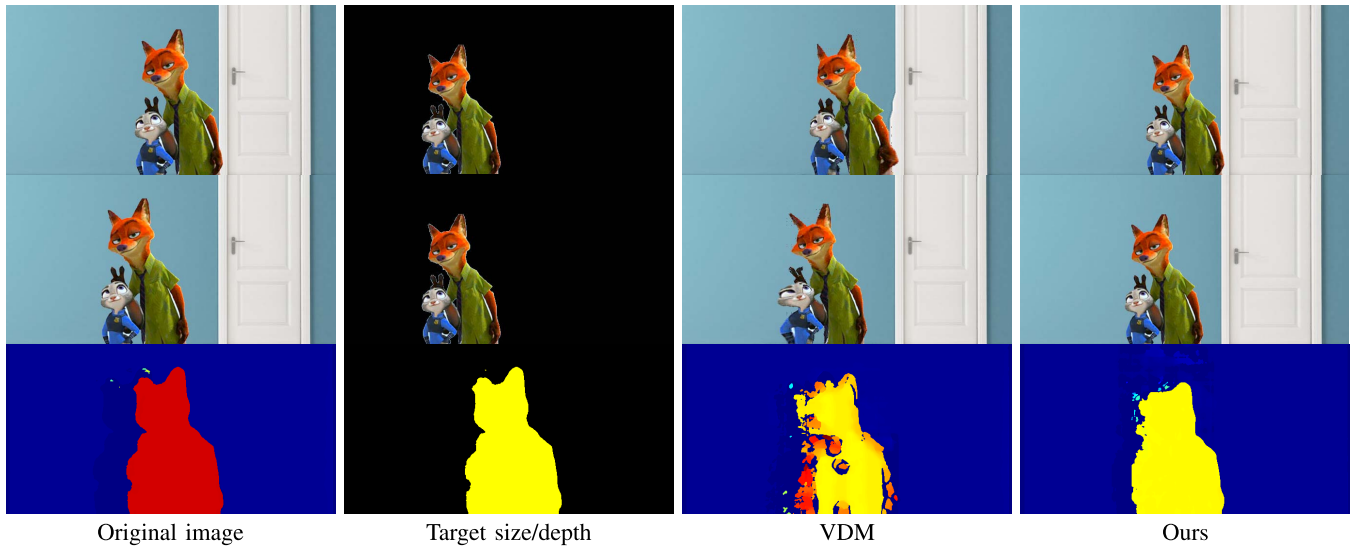
Fig. 1.   Examples of object remapping methods. From top to bottom, the left-view image, right-view image, and depth map. The sizes and depths of the fox and rabbit (from movie *Zootopia*) are decreased to 89% and by 4cm, respectively. VDM [6] introduces noticeable shape distortions on the door in the left-view image and the face of the rabbit in the right-view image. VDM also fails to faithfully adjust the size/depth of the rabbit. Our method faithfully adjusts the sizes and depths of the two objects without introducing noticeable shape distortions.

caused by the high degree of freedom of conventional grid warping schemes [6]. On top of the proposed axis-aligned warping, we build an optimization model and derive efficient constraints for object resizing, depth adjustment, shape preservation, and object consistency. Moreover, our optimization model adjusts the size and depth of a selected object via warping unimportant regions of the entire stereo image, rather than the local neighboring regions surrounding the object suggested by existing methods [6], since the HVS tends to be relatively insensitive to distortions in unimportant regions [7].

Our contributions are listed as follows:

- To the best of our knowledge, we are among the first to introduce axis-aligned warping to object remapping. Thanks to its reduced degree of freedom of grid warping, the proposed axis-aligned warping effectively mitigates shape distortions, while faithfully adjusting the depth and size.
- Our method mitigates the distortions of object boundary regions surrounding selected objects by propagating the distortions to unimportant regions in the entire stereo image. This enables further shape and size adjustment.
- Our object consistency constraints consistently warp grids associated with a selected object, making them robust to imperfect object extraction to some extent.
- Our method offers the flexibility of setting depth adjustment and object resizing separately, making it easier to meet multiple requirements for object remapping simultaneously.

Compared with its preliminary conference version [8], this paper has been significantly expanded in three aspects. First, we clarify why the high-degree-of-freedom grid warping proposed in [6] can significantly distort important/structural objects. Second, new object consistency constraints are derived to ensure grid warping consistency for an object. Third, more experiments and a user study are conducted along with insightful analyses on the evaluations. With these newly added components, the superiority of the proposed method is thoroughly validated.

The rest of our paper is organized as follows. Sec. II surveys related work on object remapping as well as warping-based

editing methods. Then, we formulate the problem of object remapping in Sec. III. The object remapping method is proposed in Sec. IV. We demonstrate experimental results in Sec.V.

II. RELATED WORK

A. Image and Video Warping

Warping has been widely-adopted in many image/video editing areas including image/video retargeting, video stabilization, image distortion correction, image morphing, etc. Image/video retargeting is to adapt a 2D image/video to various sizes or aspect ratios. Many warping-based methods [9]–[14] were proposed for 2D image/video retargeting. To maximize viewing experience, these methods preserve the shapes of important objects while stretching or shrinking less important regions via non-uniformly warping pixels/regions. Since the HVS tends to be relatively insensitive to the shape distortions of unimportant regions, these warping-based methods achieve better perceptual quality than uniform scaling. Inspired by these 2D image and video retargeting methods, warping-based methods have been explored for stereo image/video retargeting. Compared with 2D image/video retargeting, stereo image/video retargeting methods [15]–[21] usually resort to additional depth-preserving constraints to preserve scene depth. Specifically, the depth-preserving constraints proposed in [15]–[17] maintain depth by encouraging the disparity value of a few correspondences in the retargeted image to be equal to their original values. Besides depth preservation, Shao *et al.* [22] further consider visual comfort for stereo image retargeting. Chai *et al.* [23] propose a hybrid method which combines warping and cropping to generate thumbnail for stereo images.

Video stabilization aims to remove undesired camera shake and jitters to improve viewing experience. Warping-based methods [24]–[26] were proposed for stabilizing videos. For example, Liu *et al.* [24] proposed a content-preserving warping for stereo video stabilization. Tang *et al.* [26] devised a 3D spherical warping model for 360° video stabilization.

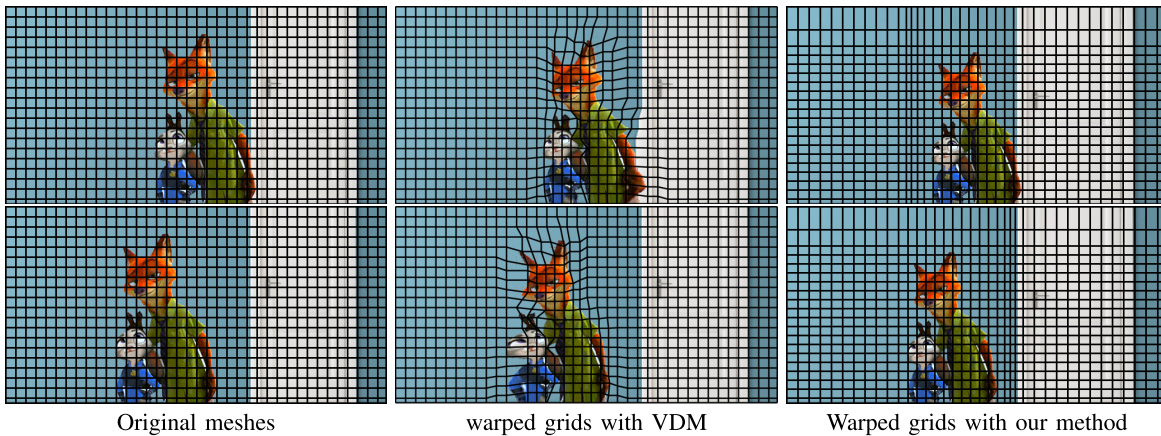| Original meshes | warped grids with VDM | Warped grids with our method |

Fig. 2. Illustration of grid warping of depth remapping methods in Fig. 1. From top to bottom, we show the grids of left-view image and the grids of right-view image. Because VDM adopts a high-degree-freedom grid warping to adjust the size/depth, most grids in the background remain unchanged and only neighboring grids around the objects are shrunk/stretched. In contrast, with the proposed axis-aligned warping, our methods achieve significantly better remapping (see Fig. 1) by warping grids in unimportant regions of the image to have more room for depth/size adjustment.

Some works [27]–[29] also use content-preserving warping for stereo image stitching.

Recently, warping was introduced into aesthetic-driven stereo image recomposition. Aesthetic-driven stereo image recomposition is to improve the composition of the stereo image, such that the aesthetic quality of the image is enhanced. Islam *et al.* [30] proposed a hybrid method for aesthetic-driven stereo image recomposition. This method [30] first segments a stereo image into foreground and background objects, and then the segmented objects are pasted through warping and inpainting with the guidance of aesthetic rules. Chai *et al.* [31] proposed a novel warping-based image recomposition for stereo images [31]. By explicitly considering the trade-off between information loss and aesthetic quality as well as depth adaption, method [31] generates high-quality recomposed stereo images. To ensure the aesthetic quality of recomposed stereo images, these methods [30], [31] adopt computationally complex algorithms such as inpainting or triangle-mesh-based optimization.

B. Depth Remapping

One class of depth remapping is view synthesis-based methods [32]–[35], which globally adjust the depth of a stereo image by synthesizing new views from the original stereo image pair. For example, the methods in [32], [33], [35] first estimate the disparity map from the original stereo image pair. Then, the depth map of a new viewpoint is estimated based on camera parameters. With the estimated depth map, the new-view image is synthesized from the original stereo images. View synthesis-based methods, however, rely on camera parameters, which may not be readily available. Moreover, these methods often yield disoccluded regions that need inpainting to fill in.

As indicated in [1]–[3], [36], due to the vergence-accommodation conflicts in the HVS, improper depth values could result in an uncomfortable viewing experience. The methods in [37], [38] map the disparity/depth *range* of an entire stereo image/video into the so called "comfort zone." Lang *et al.* [4] is the first to propose remapping the depth range through warping. Yan *et al.* [39] proposed a warping-based method for stereo video, where additional constraints are devised to ensure temporal consistency. In contrast to view synthesis-based methods, the methods proposed in [4], [39]

do not need camera parameters nor inpainting of disoccluded regions.

C. Object Remapping

Different from depth remapping which adjusts the depth range of an entire image, object remapping aims to adjust the depths and sizes of selected objects. One class of object remapping is shifting-based methods [40], [41], which first employ a segmentation tool to extract a selected object, and then shift the object such that its depth is altered to a target value. These methods, however, often generate disoccluded holes that require inpainting.

Recently, warping-based methods were proposed for object remapping in stereo images. Specifically, by dividing the stereo image into grids, the method proposed in [6] adjusts the depth and size of a selected object by non-uniformly warping neighboring unimportant regions surrounding the object based on high-degree-of-freedom warping functions. Without the need to shift objects, warping-based methods do not generate holes. However, they often introduce visually annoying depth/shape distortions in stereo images, as illustrated in Fig. 2. These distortions mainly originate as byproducts of the high-degree-of-freedom warping functions and the limited space for object remapping gained from warping the neighboring unimportant regions surrounding a selected object.

III. PROBLEM FORMULATION

Given an original stereo image pair consisting of a left-view image and a right-view image, our object remapping method aims to generate a high-quality stereo image pair, in which the depths and sizes of selected objects are altered to their desired values.

We tackle the object remapping problem via grid warping. As mentioned earlier, existing grid warping-based remapping methods often cause noticeable shape distortions in remapped stereo images (see Fig. 1 and Fig. 3), when adjusting the depth/size of a selected object. Such shape distortions can be visually annoying, especially when noticeable distortions appear on important objects. Besides introducing annoying visual artifacts, shape distortions may also confuse depth perception, as revealed in the literature [42], [43]. Specifically, the HVS may incorrectly perceive the depth of an object with severe shape distortions, leading to stereo artifacts like
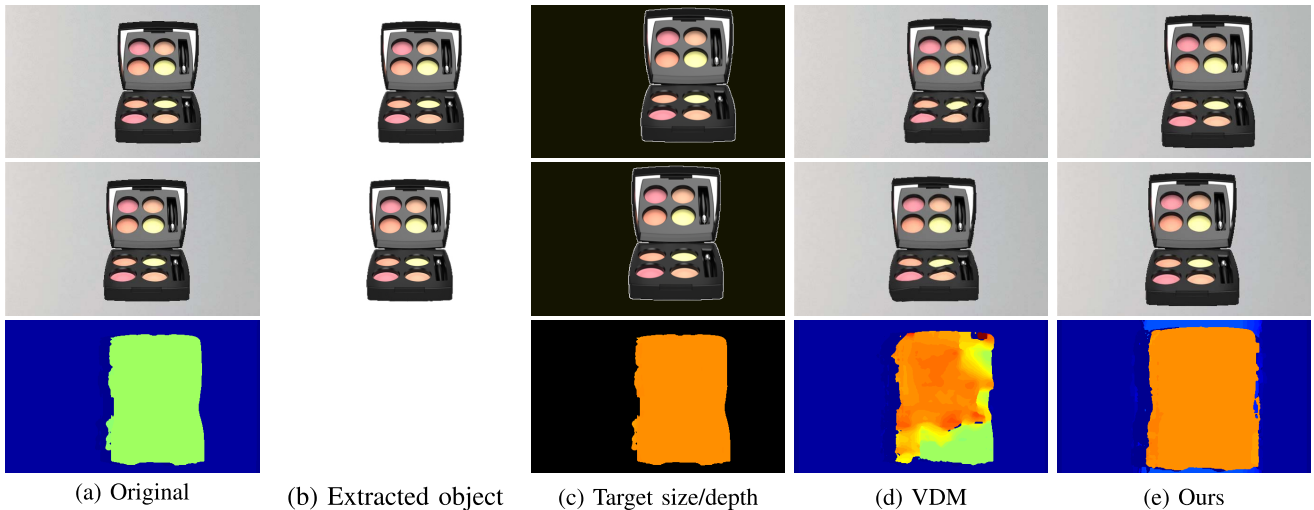
Fig. 3. Illustration of object remapping of (a) the original stereo image to adjust the depth/size of (b) the selected eyeshadow box with an accurate segmentation mask to (c) the target values. (d) VDM distorts both the shape and the depth (disparity) of the eyeshadow box. (e) Our method faithfully preserves the shape of the eyeshadow box while well adjusting its disparity to the target value. From top to bottom, the left-view image, right-view image and disparity map.
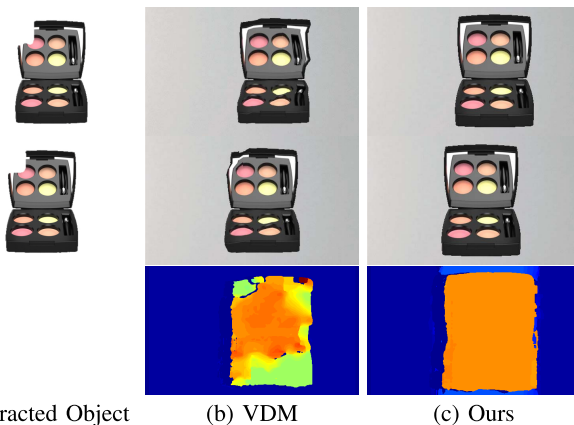

Fig. 4. Illustration of object remapping with inaccurate object extraction, where (a) the top-left part of the object is missing compared to the accurate object extraction in Fig. 3. (b) The method proposed in [6] leads to a distorted depth map due to inconsistent warping between the left-view and right-view images. (c) Despite the inaccurate object extraction, our method does not degrade the quality of depth remapping but achieves comparable performance with that using accurate object extraction in Fig. 3. From top to bottom: the left-view image, right-view image and disparity map.

retinal rivalry. Furthermore, viewers tend to fixate on important objects and perceive their depths to understand the 3D scene of a stereo image pair. If one cannot properly perceive the depths of important objects, he/she would suffer from an uncomfortable viewing experience that may lead to headaches and nausea. Hence, we aim to mitigate shape distortions on important objects in a stereo image pair, while faithfully adjusting their sizes and depths.

We consider an additional requirement that the regions associated with a selected object should be consistently warped to avoid the object shape deformation, which was never taken into account by existing methods to the best of our knowledge. Moreover, when editing an object in an image, the object may not be accurately extracted from the image using an automatic object segmentation tool, as illustrated in Fig. 4. This inaccurate object extraction issue will degrade the visual quality of remapped stereo images as well. In the example shown in Fig. 4, since the top-left region in the eyeshadow box is not well extracted, the method in [6] inconsistently

warps the corresponding regions of the object, leading to a distorted depth map. In contrast, our method can tolerate this inaccurate object extraction while remapping the object.

To meet the above requirements, we formulate object remapping as an optimization problem aiming at (1) preserving the shapes of important objects while avoiding noticeable shape distortions on unimportant objects, (2) faithfully adjusting the depths and sizes of selected objects to their target values, and (3) consistently warping the grids associated with a selected object, while mitigating the negative effect of inaccurate object extraction.

Let $\mathbf{I} = \{I^L, I^R\}$ denote a stereo image pair, where I^z, $z \in \{L, R\}$ represents the z-view image. Our objective is to find the optimal warping functions of grids that minimize the following energy function:

$$E = (E^S + \lambda^D \cdot E^D + \lambda^O \cdot E^O + \lambda^C \cdot E^C) \qquad (1)$$

where E^S is the energy for shape preservation, E^D the depth adjustment energy measured by the deviation from the remapped depths of selected objects to their target values, E^O the object resizing energy measured by the deviation from the remapped sizes of selected objects to their target values, E^C the object consistency energy, and λ^D, λ^O and λ^C the weighting factors for the depth energy, size energy, and object consistency, respectively.

The optimization of object remapping is challenging, since the optimal warping functions need to fulfill multiple requirements simultaneously, which often causes conflicts among the requirements. Recent stereo image/video retargeting works [22], [44] indicate that the HVS can tolerate depth distortions on unimportant regions to some extent. Inspired by these works, our method allows slight depth distortions on unimportant regions to provide enough room to accommodate the adjustments on selected objects. Thus, our method allow grids on unimportant regions to be stretched/shrunk. Moreover, we adopt axis-aligned warping [45], [46] for object remapping, which enforces the shapes of warped grids to be rectangular so as to mitigate shape distortion while performing object size and depth adjustments. On top of the axis-aligned warping, we formulate the shape-preserving
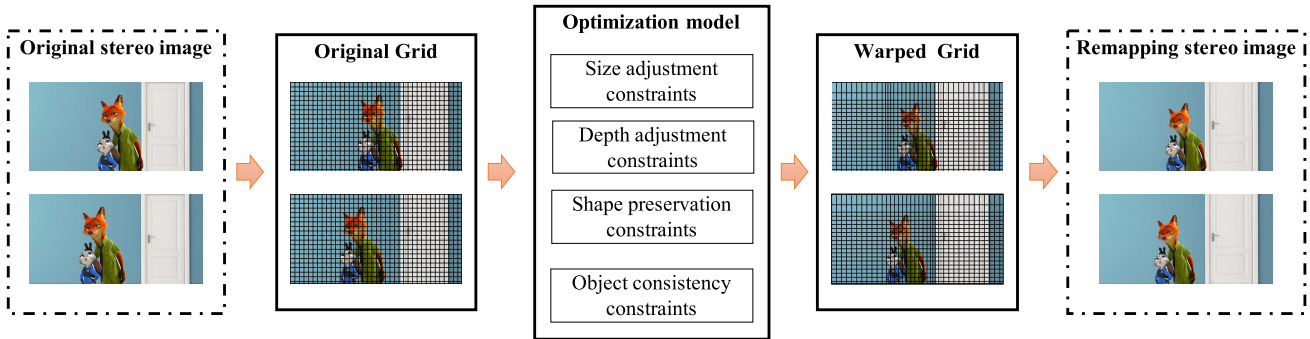
Fig. 5. Proposed object remapping framework. Our method first divides a stereo image pair into grids. Then, an optimization model is derived to find optimal warped grids, where resizing constraints, depth adjustment constraints, shape-preserving constraints, and object consistency constraints are built.

energy, object-resizing energy, depth adjustment energy, and object consistency energy, respectively, in Sec. IV.

IV. PROPOSED OBJECT REMAPPING FRAMEWORK

Our method consists of four steps. First, we extract selected objects via an automatic or interactive object segmentation tool, and then specify the target depths and sizes for the selected objects. Second, we divide the left-view and right-view images of a stereo image pair into grid meshes. Third, we formulate an optimization model over the grid meshes. In the model, size/depth adjustment constraints are derived to adjust the sizes/depths of the selected objects while preserving that of the remaining non-selected objects. We also impose shape-preserving constraints and object consistency constraints to avoid noticeable shape distortions in unimportant regions while reducing the negative effects caused by inaccurate object extraction. Finally, we solve the proposed optimization problem to obtain optimal warped grids for the stereo image, and then use the warped grids to generate the edited stereo image pair with depth-remapped objects.

In the following subsections, we first discuss what kind of grid transformation is suitable for object remapping. We then elaborate on our object remapping model including constraints for resizing, depth adjustment, shape preservation, and object consistency.

A. Grid Warping for Object Remapping

The key challenge of object remapping in a stereo image pair is to simultaneously fulfill multiple requirements including shape preservation, resizing and depth adjustment for selected objects, which would usually compete for limited resources in the image. For example, resizing constraints can easily conflict with shape-preserving constraints and depth-adjustment constraints. When such conflicts occur, existing methods (e.g., [6]) usually tend to deform the shapes of objects or cannot correctly adjust the depths/sizes of selected objects to their target values.

We here investigate why non-uniform warping adopted by existing methods (e.g., [6]) causes conflicts among the constraints for object remapping. Specifically, non-uniform warping adopted by method [6], referred to as "vertex warping", allows a grid to be an arbitrary quadrilateral. Such vertex warping attempts to adjust the size of an object by directly imposing constraints on the vertices. That is, the lengths of grid edges between two neighboring vertices are constrained to be identical to the target ones. However, vertex warping needs to constrain four vertices of individual grids, each containing two variables (i.e., the x and y coordinates of a vertex), thereby
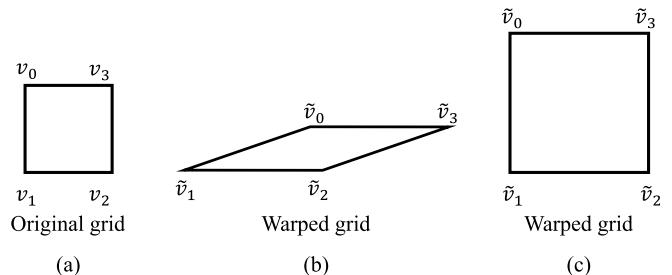

Fig. 6. Illustration of enlarging (a) the original grid by 60% to obtain the warped grids by (b) vertex warping and (c) axis-aligned warping, respectively. In (b), although each edge between two neighboring vertices is increased by 60%, the grid height wrongly decreased to 52% while the grid width increases to 300%. In contrast, the grid width and height are properly increased by 60% in (c).

constituting a large solution space. As a result, owing to such a high degree of transformation freedom with vertex warping, the resizing constraints together with the depth adjustment constraints often stretch and rotate grid edges, leading to shape distortions of objects as illustrated in Fig. 6. In addition, the size (height and width) of an object involving warped grids often cannot be effectively adjusted, although the edge lengths in the warped grids are constrained to be identical to their target lengths (see Fig. 6).

To address this problem, we propose to reduce the degree of freedom of grid warping, such that our object remapping model can meet multiple requirements simultaneously. We hence adopt axis-aligned warping [45], [46], which has been proven to be effective and efficient in 2D image/video retargeting [11], [45], especially for 2D images containing structured objects. Axis-aligned warping restricts the shape of warped grids to be rectangular, which explicitly eliminates the rotation of grid edges. Thus, with a largely reduced degree of freedom of grid warping, axis-aligned warping is much easier to control compared to vertex warping. Moreover, since rotation is eliminated, axis-aligned warping can properly adjust the size of objects while preserving the structure of objects.

B. Shape Preservation

Traditional object remapping schemes warp the grids covering a selected object, while preserving the locations of grids in the unselected regions (except for the neighbors surrounding the selected object). However, since unselected regions usually occupy a large portion in a stereo image, preserving the locations of grids on these regions significantly limits the zoom for size/depth adjustments of selected objects, making the adjustments unsatisfactory. To address the

problem, our method allows all grids to be warped, such that more resources can be utilized for object remapping. Nevertheless, significant warping of grids covering important content may introduce visually annoying shape distortions or largely alter object positions. Moreover, object-resizing and depth-adjustment constraints may also distort the shapes of important objects. To address the two issues, we build two sets of shape preservation constraints corresponding to the following energy:

$$E^S = E_1^S + E_2^S \qquad (2)$$

where E_1^S is the set of the constraints for preserving the shape of important objects, and E_2^S for avoiding large shifting of object position.

To preserve the shapes of important objects, our method resorts to adaptive warping of grids by their importance, inspired by shape-preserving constraints developed in image/video retargeting methods [11], [14], [17], [19], [46]. We define the shape preservation energy as the weighted sum of all grids' distortions, each being measured by the discrepancy between its original aspect ratio and that of its warped version, since the shape of a warped grid is constrained to be rectangular. As a result, we simply find the optimal set of heights and widths of warped grids rather than the grids' vertices, since grids at the same row/column have the same height/width. Given a stereo image pair containing left-view image I^L and right-view image I^R, the shape preservation energy is defined as follows:

$$E_1^S = \sum_{z \in \{L,R\}} \sum_{i=1} \sum_{j=1} (w_{i,j}^z \cdot \tilde{h}_{i,j}^z - h_{i,j}^z \cdot \tilde{w}_{i,j}^z)^2 \cdot s_{i,j}^z, \qquad (3)$$

where $w_{i,j}^z$ and $h_{i,j}^z$ ($z \in L, R$) respectively denote the original width and height of grid $g_{i,j}^z$, $\tilde{w}_{i,j}^z$ and $\tilde{h}_{i,j}^z$ are the width and height of the warped grid, respectively, and $s_{i,j}^z$ represents the importance of grid $g_{i,j}^z$.

Since our method allows all grids to be warped, the grids in the left/right boundary regions may be seriously shrunk/stretched. This can globally alter object locations largely along the horizontal direction. To avoid this object-shift artifact, we constrain the left/right-boundary grids not to significantly deviate from their original x-coordinates by defining the following boundary deviation energy:

$$E_2^S = \sum_{g_k^z \in \mathbf{B}} (x_k^z - \tilde{x}_k^z)^2 \cdot s_{i,j}^z \qquad (4)$$

where x_k^z and $\tilde{x}_k^z$ are the x-coordinates of the top-left vertex in the original grid and its warped version, respectively. $\mathbf{B}$ is the set containing the left/right-boundary grids of the stereo image pair.

1) Importance Map: Given a grid $g_{i,j}^z$, it importance $s_{i,j}^z$ is computed by averaging importance values of pixels belonging to $g_{i,j}^z$, where the importance of each pixel is measured by importance map. We calculate the importance map as a weighted summation of the object map, disparity-based saliency map, and image-based saliency map. Specifically, the image-based saliency map and disparity-based saliency map are calculated by employing the methods proposed in [4], [47], [48]. For the object map, we assume the grids involved in the selected objects are the most important ones. Therefore, we build an object map $\mathbf{o}$ from the selected objects. That is,

if a pixel $p_{i,j}$ belongs to the selected objects, its corresponding value $\mathbf{o}_{i,j} = 1$ in the object map; otherwise $\mathbf{o}_{i,j} = 0$.

C. Object Resizing

We derive constraints to correctly resize selected objects to their target sizes. Thanks to the axis-aligned warping, our method requires simpler constraints for object resizing. In particular, since the warped grids remain rectangular, we can simply constrain the heights of warped grids to be as close as the target values as possible. Since the aspect ratio of a warped grid is constrained to be same as the original value, the grid width would proportionally increase (or decrease) with the grid height. Hence, we propose the following constraints to encourage grids covering the selected objects to undergo the same size change κ:

$$E^O = \sum_{z \in \{L,R\}} \sum_{g_{i,j}^z \in \theta} (\tilde{h}_{i,j}^z - \kappa \cdot h_{i,j}^z)^2 \cdot s_{i,j} \qquad (5)$$

where θ is the set containing grids which cover the selected objects, $s_{i,j}$ indicates the importance level of grid $g_{i,j}$, and κ is the scaling factor of the selected objects. Note, we can also build the object resizing constraints in terms of grid width instead.

Suppose both I^L and I^R are divided by $N_r \times N_c$ grids, where N_r and N_c denote the row and column number of the grid division, respectively. The number of variables involved in our resizing constraints is not greater than N_r, whereas that for resizing involved in vertex warping-based methods is $2 \times N_r \times N_c$.

1) Specifying Object Size: In the method proposed in [6], a selected object is restricted to be resized along with its depth change in the remapped stereo image. However, we argue that it is not necessary to resize an object along with its depth change, as long as the resizing does not conflict with other cues such as horizontal disparity cues and perspective cues. For example, for stereo images containing few perspective cues (see Fig. 3 for example), the object size can be decoupled from its depth. In such cases, independently adjusting the depth and size of an object (e.g., maintaining the depth of an object while shrinking it) usually does not cause visual conflict or hurt viewing experience.

Therefore, compared to existing methods, our method allows to set depth change and size change separately. Consequently, the target size of an object can be specified in several manners. For example, our method allows one to specify the target object size according to his/her preference or demand. Moreover, the target size of an object can also be determined along with its depth change as suggested in [6].

D. Depth Adjustment

We impose two constraints for depth adjustment. The first is to adjust the horizontal disparities of a warped stereo image to the target values. The second is to eliminate nonzero vertical disparities for the warped stereo image.

1) Target Horizontal Disparity: Given an object in a 3D scene, as illustrated in Fig. 7, its depth with respect to the display screen increases with its horizontal disparity between the left-view and right-view images. Hence, we adjust the depth via controlling the horizontal disparities of selected control points. To this end, we first formulate the relationship between the target depth of an object and the corresponding horizontal disparity on the display screen. Given an object and
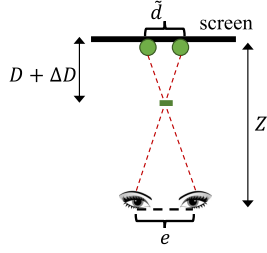
Fig. 7.   Relationship between depth and disparity.

the viewing condition, let D denote the object's original depth in the 3D scene constructed from the input stereo image pair, and $D + \Delta D$ be the target depth, where ΔD is the amount of depth adjustment. According to [49], we can calculate the target horizontal disparity from the target depth as follows:

$$\tilde{d} = -\frac{(D + \Delta D) \cdot e}{Z - D - \Delta D}, \tag{6}$$

where Z is the viewing distance from the viewer to the display screen, and e is the interocular distance between the pupils of the viewer's eyes.

2) Horizontal Disparity Adjustment: To fulfill the target depth requirement, we adjust the horizontal disparities of a few correspondences. We refer to these correspondences as control point pairs. Let $\mathbf{f}_k = \{f_k^L, f_k^R\}$ be the k-th control point pair in the left and right images, where f_k^L and f_k^R correspond to each other between the left and right images of the stereo image pair. We constrain the disparity values of all control points in the warped stereo image pair to match the target disparity values by using the following energy function of disparity deviation:

$$E^D = \sum_k E_d(\tilde{\mathbf{f}}_k, \tilde{d}_k) \cdot \tau_k, \tag{7}$$

where $\tilde{d}_k$ is the target disparity for control point pair $\mathbf{f}_k$, τ_k is the weight factor, which is set larger for those control points that are inside the selected objects or near the left/right boundary of the input stereo image. $E^D(\mathbf{f}_k, \tilde{d}_k)$ is the deviation from the disparity of $\mathbf{f}_k$ in the warped stereo image to its target value:

$$E^D(\mathbf{f}_k, \tilde{d}_k) = (\tilde{x}_k^R - \tilde{x}_k^L - \tilde{d}_k)^2, \tag{8}$$

where $\tilde{x}_k^z$ ($z \in L, R$) is the x-coordinate of control point $\mathbf{f}_k$ in warped image $\tilde{I}^z$.

We further represent $\tilde{x}_k^z$ in Eq. 8 in terms of warped grids' width, such that $E^D(\tilde{\mathbf{x}}_k, \tilde{d}_k)$ adjusts the warped grids to match the target disparity values. Assuming that f_k^z lies in grid $g_{i,n+1}$, the corresponding x-coordinate of f_k^z in the warped image is determined as:

$$\tilde{x}_k^z = \sum_{j=1}^{n} \tilde{w}_{i,j}^z + \alpha_k^z \cdot \tilde{w}_{i,n+1}^z, \tag{9}$$

where $w_{i,j}^z$ and $\tilde{w}_{i,j}^z$ are the width of grid $g_{i,j}^z$ and its warped version, respectively, α_k^z is a constant value calculated from the original grid width and x-coordinate x_k^z of control point f_k in original image $\tilde{I}^z$ as follows:

$$\alpha_k^z = \frac{x_k^z - \sum_{j=1}^{n} w_{i,j}^z}{w_{i,n+1}^z} \tag{10}$$

3) Vertical Disparity Elimination: Object resizing in Sec. IV-C may change the heights of grids and therefore introduce the y-coordinate differences of correspondences between the left-view and right-view images, thereby leading to nonzero vertical disparity. However, nonzero vertical disparities can often cause an unconformable viewing experience, e.g., eye fatigue or diplopia. To eliminate nonzero vertical disparities for grid warping, we constrain the grids in the same row of the left/right-view images to have the same height:

$$\tilde{h}_{i,j}^L = \tilde{h}_{i,j}^R \tag{11}$$

E. Object Consistency Constraints

To ensure the parts of a selected object are consistently scaled, all grids involved in the selected object (termed "object grids") need to be consistently warped. Moreover, the selected object may be imprecisely segmented, where some parts of the object are not extracted. As a result, these non-extracted parts may be inconsistently warped with the correctly extracted parts of the object, thereby leading to noticeable object shape distortion. Hence, our method also aims to tolerate imprecise object extraction to some extent. To this end, we build two sets of object consistency constraints.

The first set of object consistency constraints is designed to encourage neighboring grids involved in a selected object to be warped consistently. Specifically, given an object grid $g_{i,j}^z$, we constrain $g_{i,j}^z$ and its neighboring grids to undergo consistent warping along the vertical and horizontal directions, respectively:

$$E^C = E_h^C + E_w^C \tag{12}$$

where

$$E_h^C = \sum_{z \in \{L,R\}} \sum_{g_{i,j}^z \in \theta, g_{i+1,j}^z \in \theta} (\tilde{h}_{i,j}^z - \tilde{h}_{i+1,j}^z)^2$$

$$E_w^C = \sum_{z \in \{L,R\}} \sum_{g_{i,j}^z \in \theta, g_{i,j+1}^z \in \mathbf{v}} (\tilde{w}_{i,j}^z - \tilde{w}_{i,j+1}^z)^2, \tag{13}$$

where θ is the set containing grids which cover the selected objects.

As mentioned above, for object grid $g_{i,j}^z$ that is not extracted due to imprecise segmentation, the inconsistency between the warping on this grid and those on the correctly extracted object grids usually results in object shape and depth deformation. To mitigate the deformation, we devise the following set of object consistency constraints to constrain the warping discrepancies between $g_{i,j}^z$ and its neighboring grids according to the distance from the object center to the center of $g_{i,j}^z$. In other words, the farther from the center of the selected object to the center of $g_{i,j}^z$, the higher the warping difference is allowed:

$$E_h^C = \sum_{z \in \{L,R\}} \sum_{g_{i,j}^z \notin \theta} \sum_m (\tilde{h}_{i,j}^z - \tilde{h}_{i+m,j}^z)^2 \cdot \mu_{i,j}^{z,h}$$

$$E_w^C = \sum_{z \in \{L,R\}} \sum_{g_{i,j}^z \notin \theta} \sum_m (\tilde{w}_{i,j}^z - \tilde{w}_{i,j+m}^z)^2 \cdot \mu_{i,j}^{z,w}, \tag{14}$$

where $\tilde{h}_{i+m}^z$ is the height of warped grid $g_{i+m,j}^z$, $\tilde{w}_{j+m}^z$ is the width of warped grid $g_{i,j+m}^z$, $g_{i+m,j}^z$ and $g_{i,j+m}^z$ are the neighbor grids of $g_{i,j}^z$, $\mu_{i,j}^{z,h}$ and $\mu_{i,j}^{z,w}$ are weight factors to

control the warping difference between neighboring girds. Inspired by a data propagation algorithm for colorization [50], we determine $\mu_{i,j}^{z,h}$ and $\mu_{i,j}^{z,w}$ by the following Gaussian smooth function:

$$
\begin{aligned}
\mu_{i,j}^{z,h} &= \kappa_h \cdot \frac{1}{2\pi\sigma} e^{-\frac{(y_{i,j}^z - y_o)^2}{\sigma^2}}, \\
\mu_{i,j}^{z,w} &= \kappa_w \cdot \frac{1}{2\pi\sigma} e^{-\frac{(x_{i,j}^z - x_o)^2}{\sigma^2}}
\end{aligned} \tag{15}
$$

where σ denotes the standard deviation of Gaussian function, (x_o, y_o) is the center of the selected object, $(x_{i,j}^z, y_{i,j}^z)$ is the center of grid $g_{i,j}^z$, and κ_h and κ_w is a linear scaling function of original grid height $h_{i,j}^z$ and grid width $w_{i,j}^z$, respectively.

The object consistency constraints can help tolerate imprecise object extraction to some extent. As illustrated in Fig. 4, thanks to the proposed object consistency constraints, the top-left part of the box, that is unextracted in the segmentation mask, undergoes consistent warping with neighboring object grids correctly extracted in the segmentation mask, thereby effectively mitigating the shape and depth distortions caused by inconsistent warping with VDM [6].

F. Optimization

To find the optimal warped grids for object remapping, we minimize the total objective functions in Eq. (1) subject to a set of boundary constraints, where the terms in Eq. (1) are defined in Eq. (3), (4), (5), (8), (13), and (14), respectively, and the boundary constraints follow that in [11], [14]. We also set the lower bound of $\tilde{h}_{i,j}^z$ and $\tilde{w}_{i,j}^z$ to be $\tilde{h}_{i,j}^z \geq 1$ and $\tilde{w}_{i,j}^z \geq 1$ to prevent them from being negative. Such optimization model is kind of quadratic programming, since all the objective terms are quadratic ones and the boundary constraints are linear equalities. We can employ the interior points, trust-region-reflective or active-set algorithms to solve the quadratic programming problem [51]. We adopt the active-set algorithm in this paper.

V. EXPERIMENTAL RESULTS

Dataset: We evaluate our method on the Flickr dataset [52], which has been widely used by recent stereo image editing methods (e.g., [17], [18], [22]). We also collect several stereo images from commercial 3D movies and the dataset used in [30], [31] to evaluate the effectiveness of our method on stereo movie images. The test stereo images are selected by considering two factors. (1) We choose those types and scenes of stereo images widely appearing in commercial 3D moves. We hence pick stereo images containing human in indoor or outdoor scenes, involving various representative shot types including close-up, close, and full ones. (2) We select challenging test images for warping-based object remapping methods. We collect stereo images containing at least an object which occupies a large area of the whole images at various depths. Such large objects are challenging for adjusting the depth and size while avoiding noticeable distortions.

Evaluation Criteria: We evaluate the performance of object remapping in terms of shape preservation, size adjustment, and depth adjustment. To evaluate the performance of depth adjustment, we first employ the disparity estimation algorithm proposed in [53] to estimate the disparity map for a stereo image pair. Then, we calculate the depth map of the stereo

image from its corresponding disparity map as follows:

$$
D = \frac{e \cdot Z}{e - d} \tag{16}
$$

where the viewing distance $Z = 500$ mm and the interocular distance (IOD) $e = 60$ mm in our experiments.

The objects selected can be extracted by interactive object segmentation tools or manual annotation tools. In this work, we employ GrabCut [54], which has been widely used in image editing [17], [55] for interactive object extraction. To reduce the manual effort needed for this process (i.e., marking the foreground/background), we allow the object to be imprecisely extracted to some extent like in Fig. 15. We also evaluate the impact of such imprecise segmentation on depth remapping performance with our method. We compare our method with VDM in [6], the most recent warping-based object remapping method, and the shifting-based method (SDM) in [41].

A. Comparisons

Fig. 8 shows the comparison on *Thor* that contains an indoor scene involving a person and structural background. We first evaluate the performance of our method with increasing depth of the object.[1] The object size is also adjusted along with its depth change in the target stereo image, following the experimental setting of VDM. That is, the depth of the man is increased by 3.5cm (i.e. $\Delta D = 3.5$cm), and its size is increased by 6% (i.e. $\kappa = 1 + 6\% = 1.06\%$).

Our results show that, with depth remapping, SDM cannot adjust the size of the man to the target value. Moreover, since SDM is shift-based, it introduces holes in the remapped image, which need to be subsequently completed using image inpainting algorithms. In contrast, VDM outperforms SDM in size adjustment, since VDM takes object resizing into account. However, VDM distorts the head of the man because the warping function employed by VDM is of high freedom degree, which rotates grids of the head when imposing the resizing and depth adjustment constraints. Moreover, due to such rotations, VDM does not faithfully adjust the size of the man to the target value. In contrast, thanks to the significantly reduced degree of freedom of grid warping with axis-aligned warping, our method can faithfully alter the sizes and depths of objects as desired without introducing visually noticeable shape distortions.

Fig. 9 compares the performance of VDM, SDM, and our method on an outdoor stereo image, which contains a man and structured objects including trees and buildings. We increase the depth of the man and adjust his size according to the depth change, following the experimental setting of VDM. The result shows that SDM cannot alter the size of the man to the target value, while VDM introduces severe shape and depth distortions on the object. This is because VDM only allows a few regions surrounding the man to be shrunk/stretched, which does not provide enough room for adjusting the size and depth of the man to the target value under the resizing and depth adjustment constraints. Consequently, the man's shoulder and head are distorted. In contrast, our method allows all unimportant regions to be warped to consider the size and depth adjustment constraints more than VDM. Therefore, our

---

[1] Increasing the depth is to make the object closer to the viewer, and vice versa.
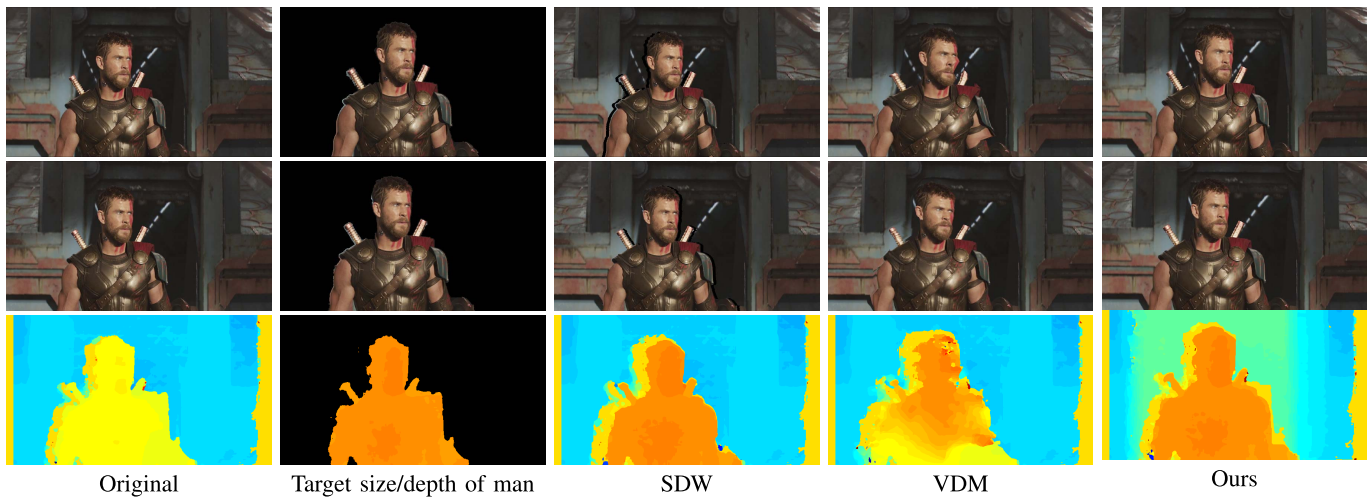
Fig. 8. Comparison of three object remapping methods on *Thor* in movie *Thor*, where the size and depth of the man is increased. From top to bottom: the left-view image, the right-view image, and the depth map. The 2nd column shows the selected object where its size is adjusted to a target size in the 1st and 2nd row, and its ground-truth depth is showed in the 3rd row.
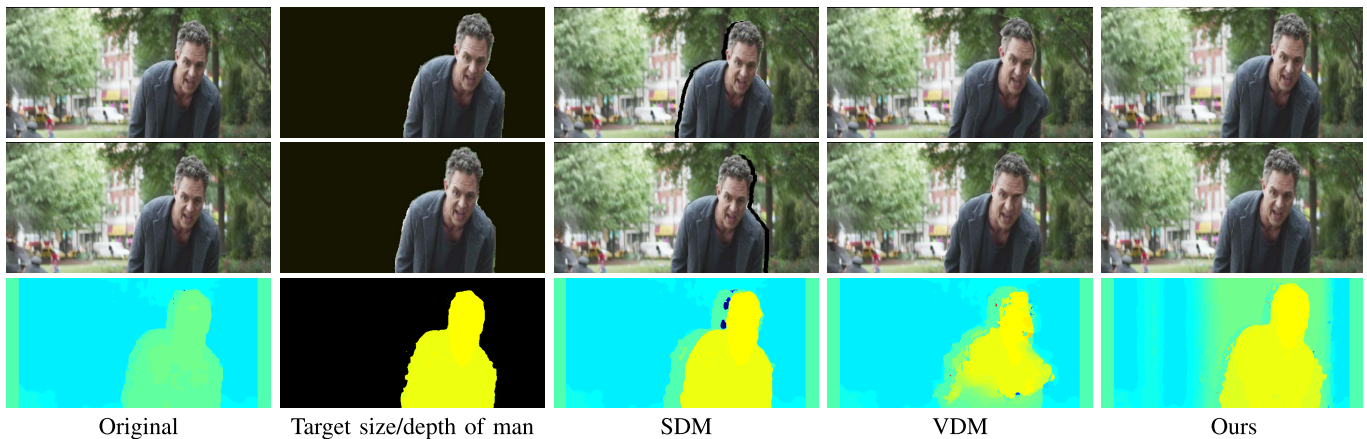

Fig. 9. Comparison of three object remapping methods on *man* in movie *The Avengers*, where the size and depth of the man is increased. From top to bottom: the left-view image, right view image, the disparity map. The 2nd column shows the selected object where its size is adjusted to a target size in the 1st and 2nd row, and its ground-truth depth is showed in the 3rd row.

method faithfully adjusts the depth and size of the selected object while preserving the shape of important objects well.

Fig. 10 compares our method on *People* taken from the Flickr dataset [56], which is challenging for depth remapping. *People* contains multiple salient objects and structured objects (e.g., the car and house). Such visual content can be easily distorted by improper warping. In the experiment, the depth of the selected man is increased by 3cm, and its size remains unchanged. The result shows that VDM [6] causes both depth and shape distortions on the selected person. This is because VDM enforces the object size to increase with its depth, but its warping manner cannot provide enough space to accommodate the object size increase. In contrast, our method can successfully address the problem.

Fig. 11 shows the comparison results on stereo image *Woman*. *Woman* is a close-up portrait, which is a common type of photography. Both the size and depth of the woman is increased simultaneously, following the experimental setting of VDM. In a close-up portrait, the face involves many grids that need to be consistently warped to well preserve the face shape. VDM distorts the face and cannot faithfully adjust the depth due to inconsistent warping of face grids. Again, our

method significantly outperforms VDM in shape preservation and depth/size adjustment.

Fig. 12 shows the comparison results on stereo image *Men*. *Men* is an long-shot stereo image, where the men fully appear in the scene. However, the long shot makes the selected objects occupy much less area than that in Figs. 3, 8, 9, 10 and 11. Both VDM and our method perform well in shape preservation. For the depth adjustment of the men, our method outperforms VDM, since VDM inconsistently adjusts the depth of regions in the men. For size adjustment, our method is slightly better than VDM. Fig. 13 shows the remapping results on stereo image *Penguins* from a cartoon movie. We decrease the depth and size of the penguin at the deeper depth. The selected object is relatively small, compared with that in other Figures. Fig. 13 shows that both VDM and our method well adjusts the size and depth of the penguin while faithfully preserving its shape.

B. Evaluation of Robustness to Imprecise Object Extraction

We show that our method can tolerate imprecise object extraction to some extent. Fig. 14 illustrates the remapping results of VDM and our method with inaccurate object
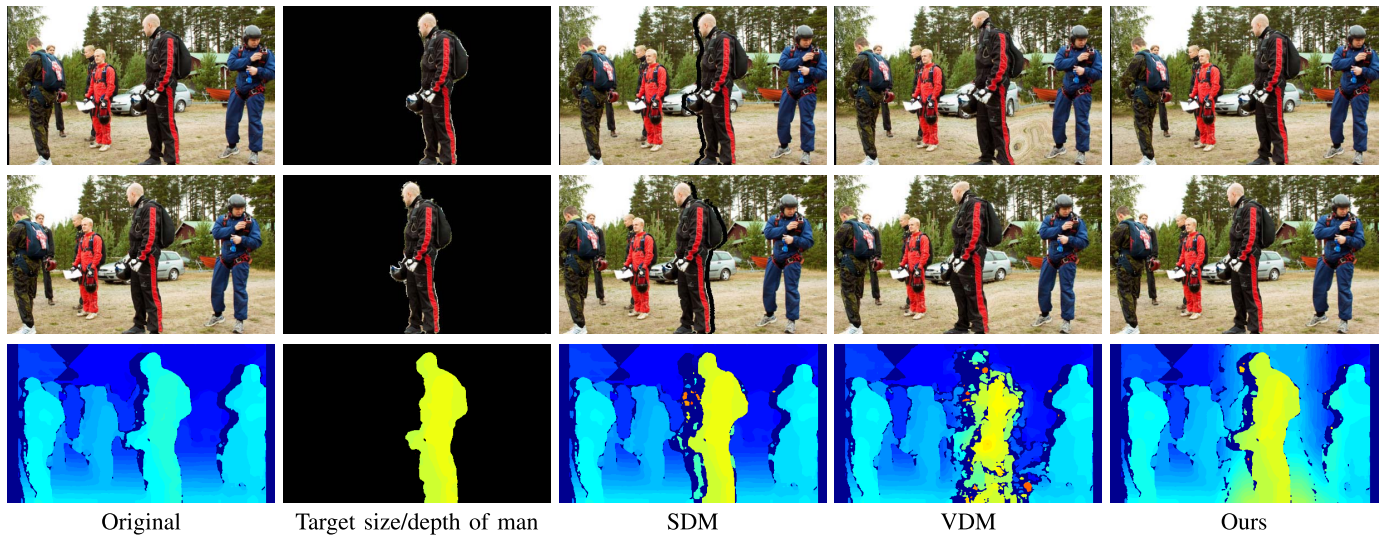
Fig. 10. Comparison of SDM [41], VDM [6] and our method on *People*, where only the depth of the man is increased. From top to bottom: the left-view image, right-view image, and disparity map. The 2nd column shows the selected object where its size is adjusted to a target size in the 1st and 2nd row, and its ground-truth depth is showed in the 3rd row.

Fig. 11. Comparison of our method and VDM [6] on *Woman* in movie *Guardians of the Galaxy*, where the size and depth of the woman is increased. From top to bottom: the left-view image, right-view image, and the disparity map. The 2nd column shows the selected object where its size is adjusted to a target size in the 1st and 2nd row, and its ground-truth depth is showed in the 3rd row.
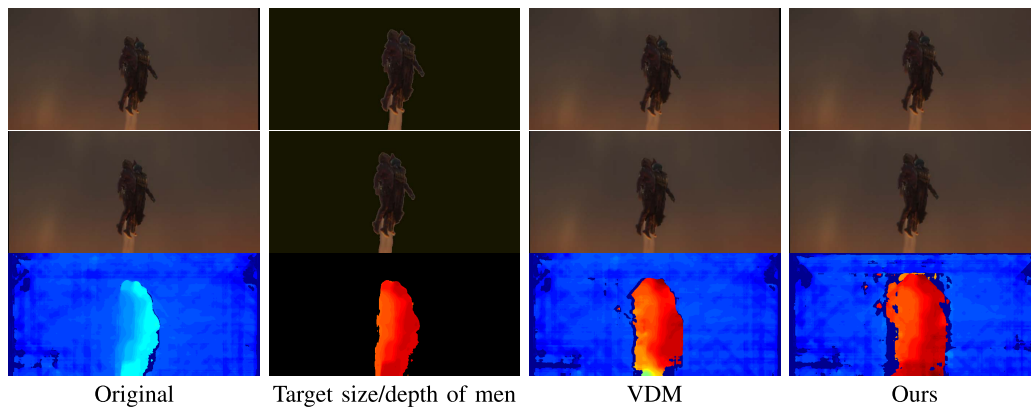

Fig. 12. Comparison of VDM [6] and our method on *Men* in movie *Guardians of the Galaxy*, where the size and depth of the men is increased. The 2nd column shows the selected object where its size is adjusted to a target size in the 1st and 2nd row, and its ground-truth depth is showed in the 3rd row.
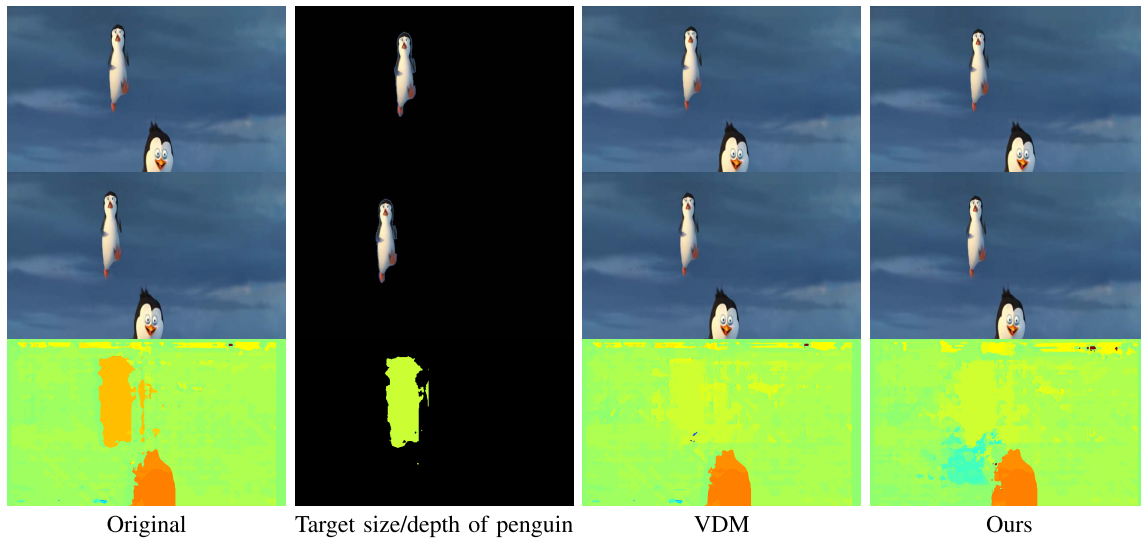
Fig. 13. Comparison of VDM [6] and our method on *Penguins* in movie *Madagascar*, where the size and depth of the left penguin is decreased. The 2nd column shows the selected object where its size is adjusted to a target size in the 1st and 2nd row, and its ground-truth depth is showed in the 3rd row.
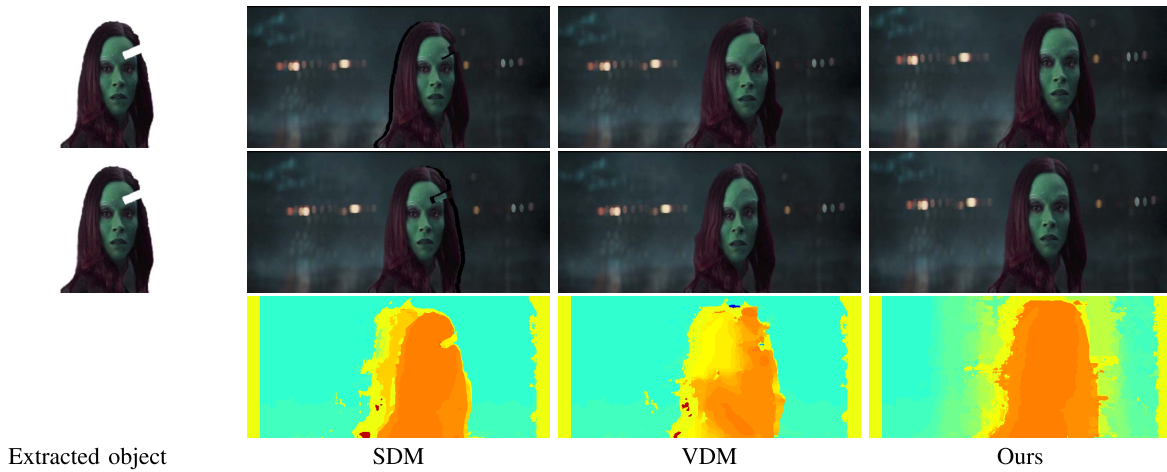

Fig. 14. Comparison of SDM [41], VDM [6] and our method on *Woman*, where the size and depth of the woman is increased. Compared with Fig. 11, a portion of the woman' face region is not extracted. From top to bottom: the left-view image, right-view image, and disparity map. The original stereo image and the target size/depth are shown in Fig. 11.
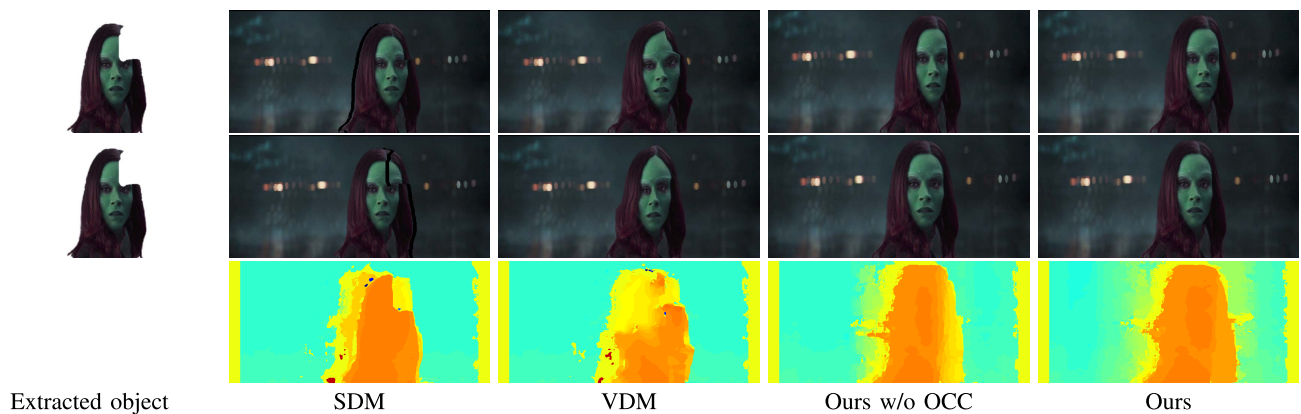

Fig. 15. Comparison of SDM [41], VDM [6] and our method on *Woman*, where the size and depth of the woman is increased. Compared with Fig. 14, a larger portion of the woman's face region is not extracted. From top to bottom: the left-view image, right-view image, and disparity map. The original stereo image pair and the target size/depth are shown in Fig. 11.
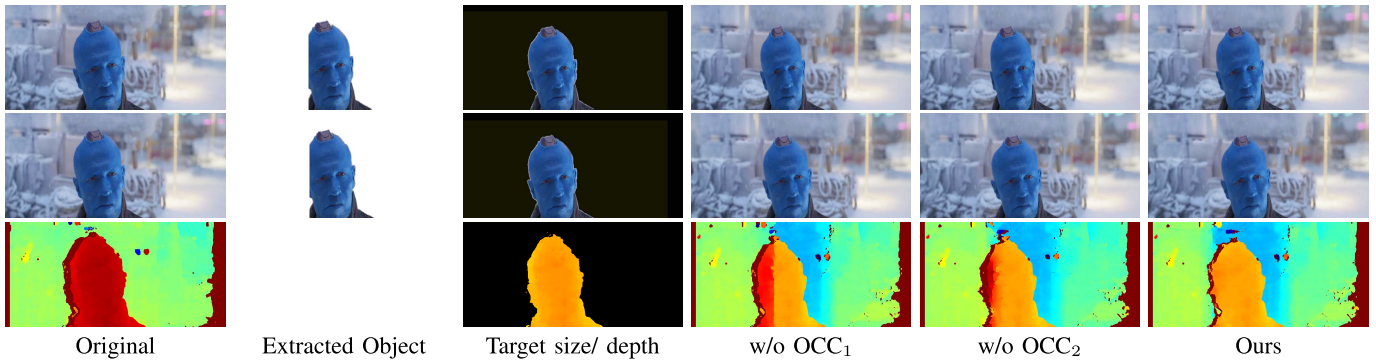
Fig. 16. Comparison of "w/o OCC$_1$", "w/o OCC$_2$" and our method on *Yondu* in movie *Guardians of the Galaxy*, where the size and disparity of the man is decreased. The left region of the man' head is not extracted, which is to test the effectiveness of the object consistency constraints. From top to bottom: the left-view image, right-view image, and disparity map. The 2nd column shows the selected object, where its size is adjusted to a target size in the 1st and 2nd row, and its ground-truth depth is shown in the 3rd row.

extraction, where a small region of the woman is missing, in comparison with the remapping results with accurate object extraction in Fig. 11. The result shows that SDM breaks the top-right region of the woman's face and fails to faithfully adjust the depth of the woman's left region, since SDM's performance highly relies on the accuracy of image segmentation. Compared to Fig. 11 and while there is no broken region, VDM significantly distorts the shape and depth of the woman's face, since it tends to inconsistently warp the extracted and unextracted face grids. In contrast, our method achieves comparable performance with the case of accurate object extraction. This is primarily due to the object consistency constraints. In particular, although part of the face is not extracted, its neighboring face region is selected. With object consistency constraints, neighboring grids in/around the face are constrained to be warped consistently, thus properly warping the unextracted face.

Compared with Fig. 14, Fig. 15 shows the remapping results with a much poorly extracted object, where a large portion of the face is not extracted. Owing to more noise in object extraction, the remapping results of VDM in Fig. 15 are worse than that in Fig. 14. Our method shrinks the face of the woman a bit in the left image of Fig. 15, while our method well preserves the face in Fig. 14 and 11. The disparity adjustment performance of our method is also degrade a little, compared with Fig. 14 and 11. Nevertheless, our method still generates much better remapping results in terms of shape preservation, depth adjustment, and size adjustment, Compared with VDM.

C. Validating Effectiveness of Object Consistency Constraints

We conduct experiments to validate the effectiveness of the object consistency constraints. Let OCC$_1$ and OCC$_2$ denote the Object Consistency Constraints (OCCs) in (13) and (14), respectively. We build two baselines, where the first baseline named "w/o OCC$_1$" removes OCC$_1$ from our method, and the second named "w/o OCC$_2$" removes OCC$_2$.

Fig. 16 compares the results on a challenging testing set, where the imperfectly extracted object is challenging for disparity remapping methods, since the left part of the man's head is missing. As shown in Fig. 16, our method achieves promising remapping results. Although the left part of the man's head is missing, the part is constrained to undergo similar transformations with the selected head regions by OCC$_2$, thereby achieving promising shape preservation and depth/size adjustment. In contrast, "w/o OCC$_2$" introduces noticeable

shape distortions in the left head of the man, since OCC$_2$ is removed. Similarly, "w/o OCC$_1$" inconsistently adjusts the depth of regions in the man's head and distorts the man's head. In contrast, our method outperforms "w/o OCC$_1$", since OCC$_1$ constrains the grids in the selected regions to be consistently warped.

We build the third baseline named "w/o OCC" that removes both OCC$_1$ and OCC$_2$. Fig. 15 shows that "w/o OCC" inconsistently warps the un-extracted parts and extracted ones of the face, leading to the distortions on the woman's face (see the left-view image). In contrast, thanks to the proposed OCCs, our method consistently warps the face of the woman, thereby achieving better shape preservation than "w/o OCC".

D. User Study

We conduct a subjective user study on a 24-inch ASUS 3D monitor with NVIDIA active shuttered glasses and an NVIDIA GeForce 3D Vision Solution. The resolution of the monitor screen is 1920×1080 According to [57] and the suggestions in ITU-R BT.2021 [58], [59], we set the viewing distance from a subject to the monitor to be 1.256 m. We invite 21 subjects involving 15 males and 6 females to participate in the study. The subjects' age ranges from 21 to 33. All subjects are not familiar with object remapping and have no prior knowledge about the experiment. We adopt the paired comparison method [60], which is widely used to subjectively evaluate the performance of image/video processing methods (e.g., [9], [61]).

The subjective user study is conducted on 15 stereo images. To each subject, we present an original stereo image at the top, while putting a pseudo-ground-truth stereo image and two remapped versions generated by different remapping methods at the bottom, where the two compared methods are placed in a random order. Since 3D movie frames do not provide ground-truth stereo images for object remapping, we build pseudo-ground-truth stereo image like images in the second column in Fig. 8, where only the selected object with the target size and depth is presented. We ask each subject the following question: *which remapped stereo image has better quality?* The quality criteria include: (1) the faithfulness of the remapped shape of the selected object compared with the ground-truth; (2) the faithfulness od the remapped size and depth of the selected object compared with the pseudo-ground-truth.
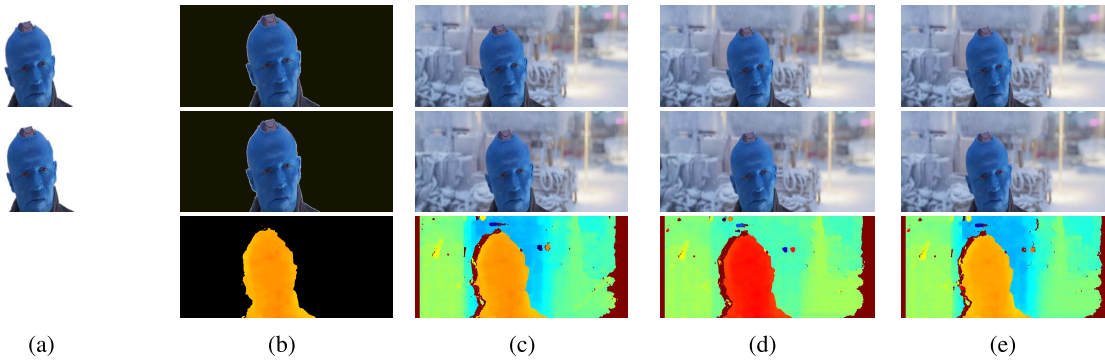
Fig. 17. Remapping results of assigning various values to weighting parameters λ^D, λ^O and λ^C. The size and disparity of the man is decreased. Different from Fig. 16, the man's head regions are accurately extracted. (a) Extracted object, (b) Target/size/depth, (c) $\lambda^O = 100$, $\lambda^D = 3000$, $\lambda^C = 500$, (d) $\lambda^O = 10^4$, $\lambda^D = 300$, $\lambda^C = 500$, (e) $\lambda^O = 10^4$, $\lambda^D = 3000$, $\lambda^C = 500$. From top to bottom: the left-view image, right-view image, and disparity map. The 2nd column shows the selected object, where its size is adjusted to a target size in the 1st and 2nd row, and its ground-truth depth is showed in the 3rd row.

We compare our method with VDM, which is the state-of-the-art warping-based object remapping method and is most related to our method. Our method receives 70.5% preference votes, while VDM receives 29.5%, showing the significant perceptual quality improvement achieved by our method over VDM. In addition, the study indicates that most subjects are more sensitive to noticeable shape distortions than depth distortion in the background.

E. Computational Complexity and Time Cost

Since our optimization model is quadratic programming, the time complexity of our method depends on the number of variables. Thanks to axis-aligned warping, the variable number is reduced to $2 \times (N_c + N_r)$, where N_r and N_c are the row and column numbers of the grid division on I^z.

We implemented our method on a PC equipped with Core i9 2.30 GHz CPU and 16 GB RAM. The optimization of our method consumes 0.125s for a stereo image with 30×30 grid division.

F. Extension to Aesthetic-Driven Stereo Image Recomposition

Our framework is complementary to other warping-based stereo image editing and can be extended to stereo image editing tasks such as stereo image retargeting and aesthetic-driven stereo image recomposition. Aesthetic-driven image recomposition is to recompose images with the guidance of aesthetic principles, such that the aesthetic quality of resulting stereo image are enhanced. We can extend our method to aesthetic-driven stereo image recomposition by incorporating additional constraints to enhance the aesthetic quality of stereo image editing. The aesthetic enhancement constraints are built upon aesthetic principles such as the rule of thirds and visual balance following [31].

We compare our method with the most recent aesthetic-driven stereo image recomposition method (ADR) [31].[2] Fig. 18 show the comparison results of our method with ADR on a stereo image containing multiple objects. ADR well improves the aesthetic quality of the stereo image. However, ADR discards the left-boundary regions of the image, leading to information loss. In contrast, our method

generates high-quality results without information loss. In addition, ADR consumes significantly higher time cost than our method. The optimization of ADR takes more than 20s, since ADR adopts triangular-mesh-based optimization which is computationally expensive. In contrast, the optimization of our method takes only 0.16s.

G. Parameter Setting

Parameters λ^D, λ^O and λ^C in (1) are adjustable, which mainly affects the remapping performance of our method. Specifically, λ^C, λ^D, and λ^C control the strength of the size adjustment, depth adjustment, and object consistency constrains, respectively. Sec. V-C has validated the effectiveness of the object consistency constraints. We hence empirically set $\lambda^C = 500$ in our experiments. We further analyze the influence of λ^D and λ^O on the remapping results. Fig. 17 shows the remapping results using different values of λ^D and λ^O. When a small value λ^O is set, the size of object would not be adequately adjusted, as shown in Fig. 17(d). Similarly, using a small value of λ^D would lead to ineffective depth adjustment of the selected object, as shown in Fig. 17(e). Therefore, we empirically set a large value to λ^O and λ^D say, $\lambda^O = 10^4$ and $\lambda^D = 3000$.

VI. DISCUSSIONS

Assuming the target depth/size value is proper and would not yield visual discomfort, we focus on adjusting the depth and size of an object to their target values. Should the target depth/size value be beyond our assumption, the remapped stereo images by our method may yield visual discomfort.

A few works have explored what kinds of depth/size cause visual discomfort for stereoscopic content. For example, it was revealed in [3], [4] that if the depth value is out of the stereoscopic comfort zone, the remapped stereo image would negatively affect 3D perception and generate visual discomfort. Similarly, as shown in [62], an improper object size may cause visual discomfort. Based on these findings, our method can address visual discomfort issues. For example, if a viewer sets an improper target depth/size value, we can refine the value into a comfortable one based on the results suggested in [3], [62], [63]. On the other hand, given a stereo image containing an object whose depth/size causes uncomfortable experience, we can remap this image to fix the problem by setting a proper target depth/size value.

---

[2]For a fair comparison, our method adjusts the depth/size of objects according to the target depth/size value of ADR.
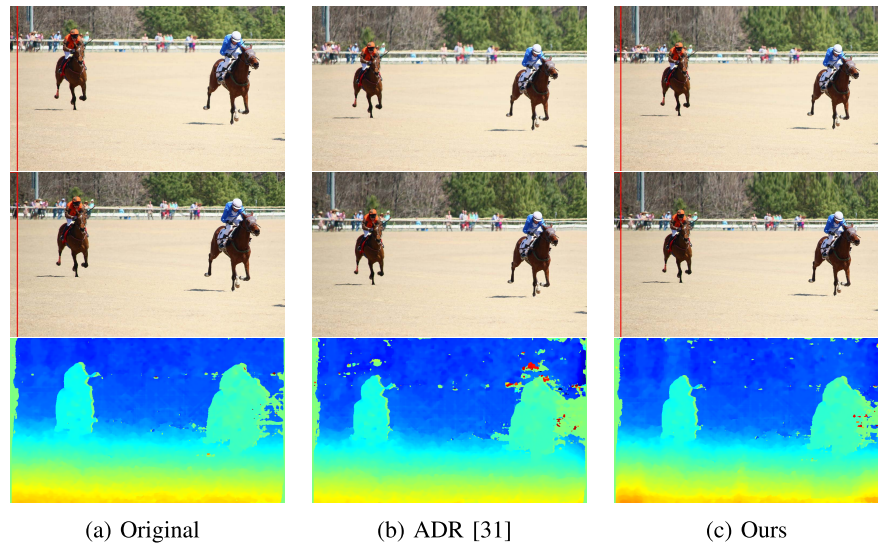
(a) Original   (b) ADR [31]   (c) Ours

Fig. 18. Results of extending our method to aesthetic-driven stereo image recomposition. From top to bottom: the left-view image, right-view image, and disparity map. (a) The left boundary regions marked by red blocks in the original stereo pair are discarded by ADR [31], leading to information loss; (b) Our method well recomposes stereo images without information loss.

Compared with conventional grid warping schemes, axis-aligned grid warping has less flexibility of transformation freedom. Due to this limitation, our method would inevitably introduce depth distortion while adjusting the depth/size of an object. To address this issue perceptually, since we observe that viewers are much more sensitive to shape distortions in important regions than depth distortions in less important ones, our method resorts to dispersing depth distortion in less important background regions. In this way, the overall viewing experience with our method is better than that with VDM especially for stereo images containing large selected objects. In contrast, due to its high degree of transformation freedom, VDM performs comparatively with our method for those stereo images with small-size objects and small disparity/size adjustment (see Fig. 13),, VDM well adjusts the size and depth of the small object while maintaining depth of the background regions.

Similar to state-of-the-art warping-based methods, our method cannot always perform well on all kinds of stereo images. In particular, since our warping scheme requires enough unimportant regions to adjust the disparity/depth, our method may fail if important objects occupy almost the whole image, making unimportant regions not enough. These failure cases are intractable for warping-based methods. That is, due to the lack of unimportant regions, important objects would be stretched/shrunk for size/depth adjustment, causing noticeable distortions.

VII. CONCLUSION

In this paper, we proposed a novel approach for object remapping that faithfully adjusts the depths and sizes of selected objects, while preserving object shapes. To achieve high-quality spatial remapping, we have proposed an optimization model along with effective object resizing constraints, depth adjustment constraints and shape-preserving constraints. Specifically, for spatial remapping, we devised an axis-aligned grid warping scheme to reduce the degree of freedom of warping, and object consistency constraints to encourage consistent warping of those object grids, both

effectively mitigating object shape distortions while faithfully adjusting the depths/sizes of selected objects to desired values. Moreover, the proposed method can tolerate imprecise object extraction to some extent, which significantly reduces the image editing cost. Experimental results show that our method achieves significant visual quality improvement over existing methods for various test image pairs.

REFERENCES

[1] 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, no. 3, pp. 1–30, Mar. 2008.

[2] 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 *Proc. SPIE*, vol. 7863, Feb. 2011, Art. no. 78630P.

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

[4] 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. 3, pp. 75:1–75:10, 2010.

[5] T. Foundry. (2020). *Ocular*. [Online]. Available: https://www.foundry.com/products/Ocula

[6] J. Lei *et al.*, "Shape-preserving object depth control for stereoscopic images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 12, pp. 3333–3344, Dec. 2018.

[7] K. He, H. Chang, and J. Sun, "Content-aware rotation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 553–560.

[8] 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 Process.*, Sep. 2019, pp. 4305–4309.

[9] Y. Niu, W.-C. Feng, and F. Liu, "Enabling warping on stereoscopic images," *ACM Trans. Graph.*, vol. 31, no. 6, pp. 1–7, Nov. 2012.

[10] Y.-S. Wang, J.-H. Hsiao, O. Sorkine, and T.-Y. Lee, "Scalable and coherent video resizing with per-frame optimization," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 1–8, Jul. 2011.

[11] 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.

[12] 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.

[13] S.-S. Lin, C.-H. Lin, I-C. Yeh, S.-H. Chang, C.-K. Yeh, and T.-Y. Lee, "Content-aware video retargeting using object-preserving warping," *IEEE Trans. Vis. Comput. Graphics*, vol. 19, no. 10, pp. 1677–1686, Oct. 2013.

[14] Y.-S. Wang, C.-L. Tai, O. Sorkine, and T.-Y. Lee, "Optimized scale-and-stretch for image resizing," *ACM Trans. Graph.*, vol. 27, no. 5, pp. 1–8, Dec. 2008.

[15] 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.

[16] 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.

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

[18] B. Li, L.-Y. Duan, C.-W. Lin, T. Huang, and W. Gao, "Depth-preserving warping for stereo image retargeting," *IEEE Trans. Image Process.*, vol. 24, no. 9, pp. 2811–2826, Sep. 2015.

[19] 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.

[20] S.-S. Lin, C.-H. Lin, Y.-H. Kuo, and T.-Y. Lee, "Consistent volumetric warping using floating boundaries for stereoscopic video retargeting," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 5, pp. 801–813, May 2016.

[21] B. Li, C.-W. Lin, S. Liu, T. Huang, W. Gao, and C.-C.-J. Kuo, "Perceptual temporal incoherence-guided stereo video retargeting," *IEEE Trans. Image Process.*, vol. 29, pp. 5767–5782, 2020.

[22] F. Shao, W. Lin, W. Lin, Q. Jiang, and G. Jiang, "QoE-guided warping for stereoscopic image retargeting," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4790–4805, Oct. 2017.

[23] X. Chai, F. Shao, Q. Jiang, and Y.-S. Ho, "MSTGAR: Multioperator-based stereoscopic thumbnail generation with arbitrary resolution," *IEEE Trans. Multimedia*, vol. 22, no. 5, pp. 1208–1219, May 2020.

[24] F. Liu, M. Gleicher, H. Jin, and A. Agarwala, "Content-preserving warps for 3d video stabilization," in *Proc. ACM SIGGRAPH*, 2009, pp. 44:1–44:9.

[25] F. Liu, Y. Niu, and H. Jin, "Joint subspace stabilization for stereoscopic video," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 73–80.

[26] C. Tang, O. Wang, F. Liu, and P. Tan, "Joint stabilization and direction of 360° videos," *ACM Trans. Graph.*, vol. 38, no. 2, pp. 1–13, Apr. 2019.

[27] W. Yan, C. Hou, J. Lei, Y. Fang, Z. Gu, and N. Ling, "Stereoscopic image stitching based on a hybrid warping model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 9, pp. 1934–1946, Sep. 2017.

[28] X. Fan, J. Lei, Y. Fang, Q. Huang, N. Ling, and C. Hou, "Stereoscopic image stitching via disparity-constrained warping and blending," *IEEE Trans. Multimedia*, vol. 22, no. 3, pp. 655–665, Mar. 2020.

[29] F. Zhang and F. Liu, "Casual stereoscopic panorama stitching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 2002–2010.

[30] M. B. Islam, L.-K. Wong, K.-L. Low, and C.-O. Wong, "Aesthetics-driven stereoscopic 3-D image recomposition with depth adaptation," *IEEE Trans. Multimedia*, vol. 20, no. 11, pp. 2964–2979, Nov. 2018.

[31] X. Chai, F. Shao, Q. Jiang, and Y.-S. Ho, "Roundness-preserving warping for aesthetic enhancement-based stereoscopic image editing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 4, pp. 1463–1477, Apr. 2021.

[32] P. Ndjiki-Nya *et al.*, "Depth image-based rendering with advanced texture synthesis for 3-D video," *IEEE Trans. Multimedia*, vol. 13, no. 3, pp. 453–465, Jun. 2011.

[33] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 600–608, Aug. 2004.

[34] S. Rogmans, J. Lu, P. Bekaert, and G. Lafruit, "Real-time stereo-based view synthesis algorithms: A unified framework and evaluation on commodity GPUs," *Signal Process., Image Commun.*, vol. 24, nos. 1–2, pp. 49–64, Jan. 2009.

[35] F. Devernay and S. Duchêne, "New view synthesis for stereo cinema by hybrid disparity remapping," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 5–8.

[36] F. Shao, Z. Li, Q. Jiang, G. Jiang, M. Yu, and Z. Peng, "Visual discomfort relaxation for stereoscopic 3D images by adjusting zero-disparity plane for projection," *Displays*, vol. 39, pp. 125–132, Oct. 2015.

[37] F. Shao, Y. Fei, R. Fu, G. Jiang, and Y.-S. Ho, "Simultaneous object size and depth adjustment for stereoscopic 3D images," *Inf. Sci.*, vol. 481, pp. 280–291, May 2019.

[38] T. Oskam, A. Hornung, H. Bowles, K. Mitchell, and M. Gross, "OSCAM—Optimized stereoscopic camera control for interactive 3D," *ACM Trans. Graph.*, vol. 30, no. 6, pp. 189:1–189:8, 2011.

[39] 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, Mar. 2013.

[40] 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*, vol. 16, no. 2, pp. 326–336, Feb. 2014.

[41] 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, Nov. 2013.

[42] N. Sugie, "Neural models of brightness perception and retinal rivalry in binocular vision," *Biol. Cybern.*, vol. 43, no. 1, pp. 13–21, 1982.

[43] R. Blake, D. H. Westendorf, and R. Overton, "What is suppressed during binocular rivalry?" *Perception*, vol. 9, no. 2, pp. 223–231, Apr. 1980.

[44] 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.*, Jun. 2018, pp. 6517–6525.

[45] D. Panozzo, O. Weber, and O. Sorkine, "Robust image retargeting via axis-aligned deformation," *Comput. Graph. Forum*, vol. 31, no. 2, pp. 229–236, May 2012.

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

[47] 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.*, Dec. 2015, pp. 1404–1412.

[48] Y. Fang, J. Wang, Y. Yuan, J. Lei, W. Lin, and P. L. Callet, "Saliency-based stereoscopic image retargeting," *Inf. Sci.*, vol. 372, pp. 347–358, Dec. 2016.

[49] R. Cormack and R. Fox, "The computation of disparity and depth in stereograms," *Perception Psychophys.*, vol. 38, no. 4, pp. 375–380, Jul. 1985.

[50] A. Levin, D. Lischinski, and Y. Weiss, "Colorization using optimization," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 689–694, 2004.

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

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

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

[54] C. Rother, V. Kolmogorov, and A. Blake, "GrabCut: Interactive foreground extraction using iterated graph cuts," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 309–314, 2004.

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

[56] 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.

[57] Q. Jiang, F. Shao, W. Gao, Z. Chen, G. Jiang, and Y.-S. Ho, "Unified no-reference quality assessment of singly and multiply distorted stereoscopic images," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1866–1881, Apr. 2019.

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

[59] 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.

[60] H. A. David, *The Method of Paired Comparisons*. London, U.K.: Griffin, 1963.

[61] 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.

[62] H. Sohn, Y. J. Jung, S.-i. Lee, and Y. M. Ro, "Predicting visual discomfort using object size and disparity information in stereoscopic images," *IEEE Trans. Broadcast.*, vol. 59, no. 1, pp. 28–37, Mar. 2013.

[63] M. Urvoy, M. Barkowsky, and P. L. Callet, "How visual fatigue and discomfort impact 3D-TV quality of experience: A comprehensive review of technological, psychophysical, and psychological factors," *Ann. Telecommun.-Annales des télécommunications*, vol. 68, nos. 11–12, pp. 641–655, Dec. 2013.

Bing Li (Member, IEEE) received the B.S. degree in computer science from Jinan University, Guangzhou, China, in 2009, and the Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2016. She worked as a Postdoctoral Fellow with the University of Southern California, USA, in 2016. She is currently a Postdoctoral Fellow with the King Abdullah University of Science and Technology. Her research interests include image/video processing, computer vision, and machine learning.
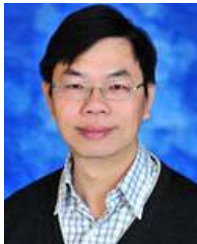
Chia-Wen Lin (Fellow, IEEE) received the Ph.D. degree from National Tsing Hua University (NTHU), Taiwan, in 2000. He is currently a Professor with the Department of Electrical Engineering and the Institute of Communications Engineering, NTHU. His research interests include image and video processing, computer vision, and video networking. From 2013 to 2015, he was a Steering Committee Member of IEEE TRANSACTIONS ON MULTIMEDIA. Since 2021, he has been serving as a Fellow Evaluating Committee Member. He received the best paper awards from VCIP 2010 and 2015. He served as the General Co-Chair for VCIP 2018. He is the Chair of ICME Steering Committee. He served as the TPC Co-Chair for ICIP 2019 and ICME 2010. He was an Associate Editor of IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE TRANSACTIONS ON MULTIMEDIA, and IEEE MULTIMEDIA. From 2018 to 2019, he was a Distinguished Lecturer of the IEEE Circuits and Systems Society.

Cheng Zheng received the B.S. degree in computer science from the Beijing University of Posts and Telecommunications, China, in 2009, and the M.S. degree from the National Space Science Center, Chinese Academy of Sciences, Beijing, China, in 2012. He was formerly a Senior Software Engineer at NetEase Games. He is currently a Software Engineer Consultant with the King Abdullah University of Science and Technology. His research interests include image processing and computer vision.

Shan Liu (Member, IEEE) received the B.Eng. degree in electronic engineering from Tsinghua University and the M.S. and Ph.D. degrees in electrical engineering from the University of Southern California. She was formerly the Director of the Media Technology Division, MediaTek, USA. She was formerly with MERL and Sony. She is currently a Tencent Distinguished Scientist and the General Manager of the Tencent Media Lab. She holds more than 200 granted U.S. and global patents and has published more than 100 journal articles and conference papers. Her research interests include audio-visual, volumetric, immersive and emerging media compression, intelligence, and transport and systems. She has been actively contributing to international standards since the last decade. She has numerous technical proposals adopted into various standards, such as VVC, HEVC, OMAF, DASH, MMT, and PCC. Technologies and products developed by her and under her leadership have served hundreds of millions of users. From 2014 to 2015, she was with the Committee of Industrial Relationship of the IEEE Signal Processing Society. From 2016 to 2017, she was the VP of Industrial Relations and Development of the Asia-Pacific Signal and Information Processing Association and was named an APSIPA Industrial Distinguished Leader in 2018. She received the Best AE Award in 2019 and 2020, respectively. Since 2019, she has been serving as the Vice Chair for IEEE Data Compression Standards Committee. She served as a Co-Editor for H.265/HEVC SCC and H.266/VVC. Since 2018, she has been on the Editorial Board of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY. She served or serves as a guest editor for a few T-CSVT special issues and special sections.
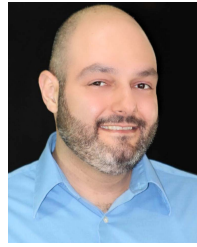
Bernard Ghanem (Member, IEEE) received the bachelor's degree from the American University of Beirut (AUB) in 2005, and the M.S./Ph.D. degree from the University of Illinois at Urbana-Champaign (UIUC) in 2010. He is currently an Associate Professor with the CEMSE Division, the Theme Leader with the Visual Computing Center (VCC), and the Interim Lead of the AI Initiative, KAUST. He has coauthored more than 120 articles in his field and three issued patents. His research interests lie in computer vision and machine learning, with emphasis on topics in video understanding, 3D recognition, and theoretical foundations of deep learning. His work has received several awards and honors, including the four best paper awards for workshops in CVPR 2013 and 2019 and ECCV 2018 and 2020, the Google Faculty Research Award in 2015 (1st in MENA for Machine Perception), and the Abdul Hameed Shoman Arab Researchers Award for Big Data and Machine Learning in 2020. He has served as the Area Chair (AC) for CVPR, ICCV, ICLR, and AAAI. He serves as an Associate Editor for IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (TPAMI).

Wen Gao (Fellow, IEEE) received the Ph.D. degree in electronics engineering from the University of Tokyo, Japan, in 1991. From 1991 to 1995, he was a Professor with the Harbin Institute of Technology. From 1996 to 2006, he was a Professor with the Institute of Computing Technology, Chinese Academy of Sciences. He is currently a Professor in computer science with Peking University. He has published extensively, including five books and over 600 technical articles in refereed journals and conference proceedings in the areas of image processing, video coding and communication, computer vision, multimedia retrieval, multimodal interface, and bioinformatics. He is in the Board of Directors for Digital Media Project and the Advisory Board of IEEE Computing.

C.-C. Jay Kuo (Fellow, IEEE) received the B.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, in 1980, and the M.S. and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1985 and 1987, respectively. He is currently the William M. Hogue Professor in electrical and computer engineering, a Distinguished Professor of electrical and computer engineering and computer science, and the Director of the USC Multimedia Communication Laboratory (MCL), University of Southern California. He is a fellow of the National Academy of Inventors (NAI), the American Association for the Advancement of Science (AAAS), and the International Society for Optical Engineers (SPIE).