Scene Warping: Layer-based Stereoscopic Image Resizing

Ken-Yi Lee   Cheng-Da Chung   Yung-Yu Chuang*

National Taiwan University

Email:kez@cmlab.csie.ntu.edu.tw, johnny751125@cmlab.csie.ntu.edu.tw, cyy@csie.ntu.edu.tw

Abstract

This paper proposes scene warping, a layer-based stereoscopic image resizing method using image warping. The proposed method decomposes the input stereoscopic image pair into layers according to the depth and color information. A quad mesh is placed onto each layer to guide the image warping for resizing. The warped layers are composited by their depth orders to synthesize the resized stereoscopic image. We formulate an energy function to guide the warping for each layer so that the composited image avoids distortions and holes, maintains good stereoscopic properties and contains as many important pixels as possible in the reduced image space. The proposed method offers the advantages of less discontinuous artifacts, less-distorted objects, correct depth ordering and enhanced stereoscopic quality. Experiments show that our method compares favorably with existing methods.

1. Introduction

Image resizing (image retargeting), adapting images for displays with different sizes and aspect ratios, has received considerable attention recently due to diversity of displays. Traditional scaling and cropping methods easily cause significant distortions or information loss. Content-aware image retargeting methods take into account the saliency distribution of the image and attempt to keep the salient features uncontaminated by hiding distortion within the less noticeable areas. As another trend, recently, stereoscopic and autostereoscopic displays have been deployed in theaters, televisions, computer screens, and even mobile devices. Due to the diversity among resolutions and aspect ratios of stereoscopic displays, similar to 2D images, stereoscopic images need to be retargeted for displaying properly on stereoscopic displays with different specifications.

Recognizing the importance of stereoscopic image retargeting, several stereoscopic image resizing methods have been proposed in the spirit of their 2D counterparts.

Basha *et al.* extended seam carving to stereoscopic image resizing [2]. Although this method produces geometrically consistent results, as a descendant of seam carving, its discrete nature may cause noticeable discontinuity on structural objects. Chang *et al.* extended the warping-based approaches to stereoscopic image resizing [3]. Although their results contain less discontinuity artifacts on structural objects, the stereoscopic quality of the resized images could be reduced because it models the whole image as a rubber sheet and can not create proper occlusions and depth discontinuity, important cues for human depth perception. As a result, depth edges could be made less prominent after resizing, thus reducing stereoscopic quality.

Our method is inspired by scene carving for 2D image resizing [10], which decomposes an image into layers and applies seam carving to synthesize scene consistent retargeting results. It can create proper occlusions and depth discontinuity because of its layered nature. However, since it adopts seam carving, scene carving suffers from the same artifacts of discontinuous structured objects. Our method can be taken as a hybrid of scene carving and warping-based methods [3]. Similar to scene carving, we decompose the input stereoscopic image pair into a set of layers. Since our input is an stereoscopic image, we can take advantage of the disparity map to create layers more easily than scene carving. We adopt a warping-based approach. Each layer is warped by its own mesh deformation and the warped layers are composited together to form the resized image. We formulate an energy function to guide image warping of each layer so that the composited resized image has the following properties: (1) it avoids distortions and holes as much as possible; (2) it maintains good stereoscopic properties and (3) it contains as many important pixels as possible in the reduced image space.

Compared to existing stereoscopic image resizing methods, the proposed method offers the following advantages. (1) It avoids artifacts of discontinuous structured objects commonly encountered by discrete resizing methods. (2) It shares the same advantages with scene carving that objects are protected (*i.e.* not distorted) and their depth orders are correctly maintained. (3) Thanks to its layered nature,

it better preserves the depth edges and creates proper occlusions, both enhancing stereoscopic quality. (4) It applies different deformations on different layers. Thus, it has better chances to hide more distortions into unimportant areas while keeping important areas uncontaminated.

2. Related work

Image retargeting. Shamir and Sorkine categorized the content-aware image retargeting into two main classes: discrete approaches and continuous approaches [16]. Seam carving [1] is a well-known discrete method, which removes a seam with the lowest importance from an image at a time. A seam is a connected path crossing the image from top to bottom or from left to right. Seam carving has been improved by many others [14, 15]. Mansfield *et al*. proposed scene carving [10] to generalize seam carving. With a user-provided relative depth map, the image is decomposed into layers. Seams are removed from the background and foreground objects are re-arranged spatially. Their algorithm has the advantages that the objects are protected (*i.e.* not distorted) and their depth orders are correctly maintained. Warping-based methods, also called continuous approaches, place a quad mesh onto the image and deform the mesh to guide image warps for resizing the image [17, 18]. Wolf *et al*. obtained warping functions by a global optimization that squeezes or stretches homogeneous regions to minimize the resulting distortions [17]. Wang *et al*. [18] proposed to assign spatially varying scaling factors by optimization. They also designed an energy term to preserve edge orientations of the mesh for important areas.

Stereoscopic image retargeting. Basha *et al*. [2] extended seam carving for stereoscopic image retargeting. Their method simultaneously carves a pair of seams, each for a view. By defining occluding and occluded pixels, they guarantee that the removed seam pairs are geometrically consistent. Nevertheless, their method suffers from limitations of seam carving and might cause obvious artifacts on structured objects, especially when the aspect ratio changes intensively. Chang *et al*. proposed a content-aware display adaptation method which simultaneously resizes a stereoscopic image to the target resolution and adapts its depth to the comfort zone of the display while preserving the perceived shapes of prominent objects [3]. Our method is similar to theirs in the aspect that both use image warping for stereoscopic image resizing.

3. Method

Given a stereoscopic image pair $\{I_L, I_R\}$ whose dimensions are $w \times h$, the goal of stereoscopic image resizing is to change their dimensions to the desired size $\hat{w} \times \hat{h}$. We first compute a disparity map between two views using the semi-global stereo matching algorithm [5]. Inspired by scene

(a) left view (b) right view

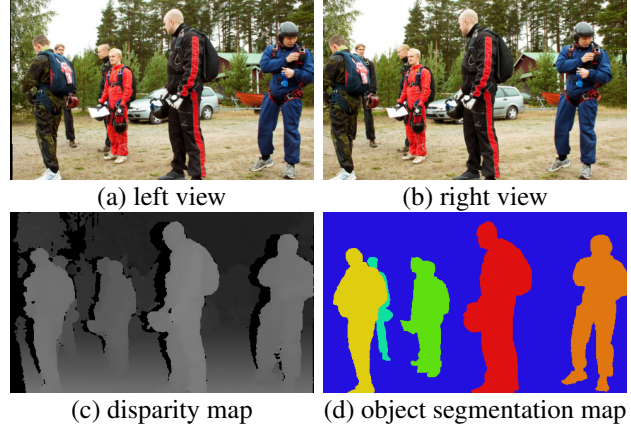
(c) disparity map (d) object segmentation map

Figure 1. The disparity map and the object segmentation map. Given an input stereoscopic image pair ((a) for the left view and (b) for the right view), we compute its disparity map (c) and its object segmentation map (d). Each color in (d) represents an object layer. Notice that only maps for the left view are shown here.

carving [10], we decompose the images into multiple object layers. Each pixel in the input image pair is assigned to one object layer based on the computed disparity map. The corresponding pixels between views should be assigned to the same object layer. Object layers can be obtained automatically or semi-automatically [9] by utilizing color and disparity. In our current implementation, we used a GrabCut system [13] to segment the stereoscopic images with user hints. Pixels are assigned to the background layer if they are not explicitly assigned to any object layer . Through this process, we obtain a set of object layers (including the background layer) $\boldsymbol{S} = \{s_1^L, s_1^R, s_2^L, s_2^R, \ldots, s_N^L, s_N^R\}$, in which the l-th object has two corresponding object layers, s_l^L and s_l^R, for the left and right views, respectively. We assume that object layers in $\boldsymbol{S}$ are sorted by their average depths; and s_1^L and s_1^R are the background layers. We also define a $w \times h$ object segmentation map, $\mathbf{O}_k$ ($k \in \{L, R\}$), in which $\mathbf{O}_k(x, y) = l$, if the pixel (x, y) of I_k belongs to object layer s_l^k. Given a stereoscopic image pair (Figure 1(a) and (b)), we compute the disparity map (Figure 1(c)) and obtain the object segmentation map (Figure 1(d)).

As most continuous methods, we place a quad mesh onto each object layer and compute a new geometry for each mesh to deform the associated object layer. Each object layer s_l^k is associated with a quad mesh of fixed quad size (20×20 in all experiments). Let $\mathbf{V}_l^k = \{v_{i,j}^{k,l}\}$ be the vertex set of the quad mesh for s_l^k, where $v_{i,j}^{k,l}$ denotes the position of the vertex at i-th column and j-th row of the mesh. Let $\hat{\mathbf{V}}_l^k = \{\hat{v}_{i,j}^{k,l}\}$ denote the vertex set of the deformed quad mesh. The goal of stereoscopic image resizing is to find the optimal vertex positions $\hat{v}_{i,j}^{k,l}$ for these deformed quad mesh with respect to some energy function.

3.1. Multi-layer image compositing

Before elaborating how to obtain the optimal vertex positions, we describe the process of compositing the resized image assuming that we have obtained the optimal vertex positions. To render the resized stereoscopic image pair $\{\hat{I}_L, \hat{I}_R\}$ with the desired size $\hat{w} \times \hat{h}$, each object layer s_l^k is first warped by the associated quad mesh $\hat{\mathbf{V}}_l^k$ to obtain the warped object layer $\hat{s}_l^k$. Next, the warped object layers $\{\hat{s}_l^k | 1 \le l \le N\}$ belonging to the same view k are composited together to obtain $\hat{I}_k$ according to their depth orders. Since layers are sorted by depths, we can use Painter's algorithm to composite the final retargeted image $\hat{I}_k$ by rendering in the order of $\hat{s}_1^k$, $\hat{s}_2^k$, ..., $\hat{s}_N^k$. Figure 2 shows the compositing process.

3.2. Problem formulation

The goal of stereoscopic image resizing is to find a stereoscopic image pair $\{\hat{I}_L, \hat{I}_R\}$ with the desired size which (1) has less distortions and holes; (2) maintains good stereoscopic properties and (3) contains as many important pixels as possible. We formulate the stereoscopic image resizing problem as an optimization problem to find a set of optimal vertex positions of deformed quad meshes, $\hat{\mathbf{V}} = \{\hat{\mathbf{V}}_l^k | k \in \{L, R\}, 1 \le l \le N\}$, which minimize the following objective function:

$$\mathbf{E}(\hat{\mathbf{V}}) = \mathbf{E}_Q(\hat{\mathbf{V}}) + \lambda_S \mathbf{E}_S(\hat{\mathbf{V}}) + \lambda_I \mathbf{E}_I(\hat{\mathbf{V}}), \quad (1)$$

where $\mathbf{E}_Q$ is the image quality energy; $\mathbf{E}_S$ is the stereoscopic quality energy; and $\mathbf{E}_I$ is the importance energy. These energy terms correspond the above three requirements respectively.

To avoid folding artifacts and heavily distorted quads, the following constraints are applied on all deformed mesh vertex positions $\hat{v}_{i,j}^{k,l} = (x_{i,j}, y_{i,j})$:

$$
\begin{aligned}
x_{i,j} &< \min(x_{i+1,j-1}, x_{i+1,j}, x_{i+1,j+1}), \\
x_{i,j} &> \max(x_{i-1,j-1}, x_{i-1,j}, x_{i-1,j+1}), \\
y_{i,j} &< \min(y_{i-1,j+1}, y_{i,j+1}, y_{i+1,j+1}), \\
y_{i,j} &> \max(y_{i-1,j-1}, y_{i,j-1}, y_{i+1,j-1}), \quad (2)
\end{aligned}
$$

where i and j index columns and rows, respectively. These are hard consraints and will be strictly enforced in our iterative optimization procedure (Section 4).

3.3. Image quality energy

We evaluate image quality from two aspects: *image distortion* and *image incompleteness*. The first one measures how layers (images) are distorted by the mesh deformation and the second one counts how many pixels are left uncovered (holes) in the final composited image. Thus, we define the image quality energy $\mathbf{E}_Q$ as

$$\mathbf{E}_Q(\hat{\mathbf{V}}) = \mathbf{E}_F(\hat{\mathbf{V}}) + \lambda_C \mathbf{E}_C(\hat{\mathbf{V}}), \quad (3)$$

I_L $\hat{I}_L$

O_L $\hat{O}_L$

s_1^L $\hat{s}_1^L$

s_4^L $\hat{s}_4^L$

s_5^L $\hat{s}_5^L$
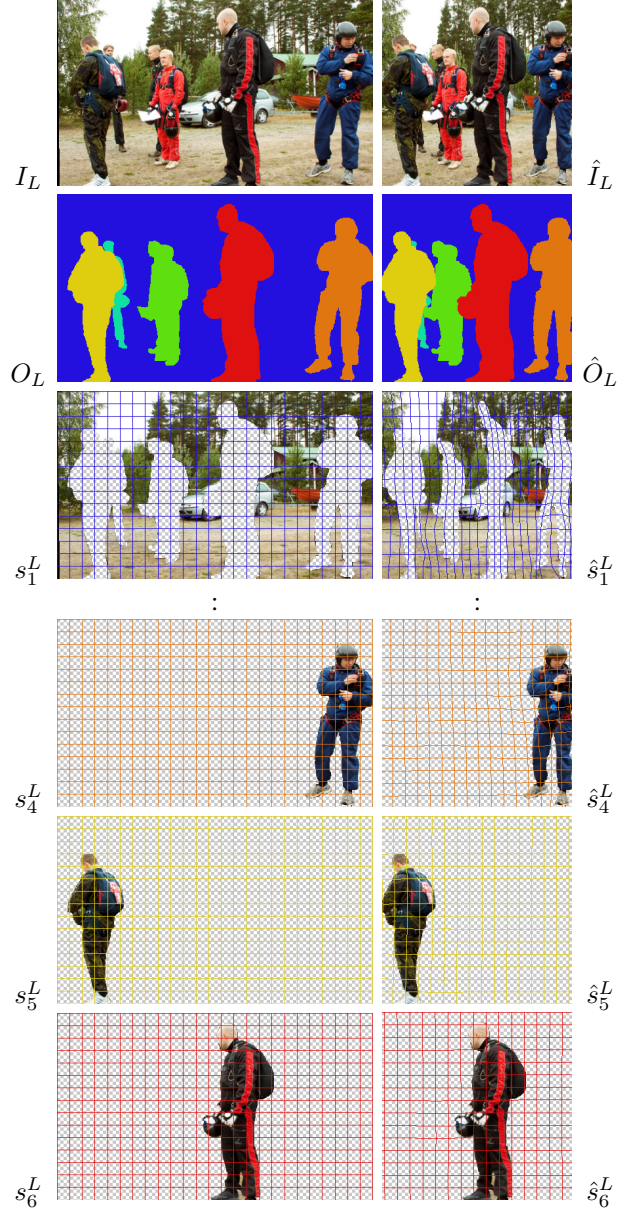
s_6^L $\hat{s}_6^L$

Figure 2. The composting process. In this example, the width is reduced by 40%. For saving space, two object layers are not displayed.

where $\mathbf{E}_F$ is the image distortion energy, and $\mathbf{E}_C$ is the image incompleteness energy. The total image distortion energy is the sum of quad distortion energy terms of all layers and all views:

$$
\mathbf{E}_F(\hat{\mathbf{V}}) = \sum_{\hat{\mathbf{V}}_l^k \in \hat{\mathbf{V}}} \sum_{\hat{q} \in \hat{\mathbf{V}}_l^k} \mathbf{W}_Q^k(q)(\mathbf{E}_R(\hat{q}) + \lambda_E \mathbf{E}_E(\hat{q}) + \lambda_O \mathbf{E}_O(\hat{q})),
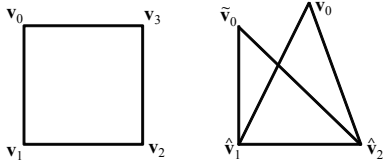$$

$$(4)$$

Figure 3. Shape deformation measurement.

where $\hat{q} = (\hat{v}_{i,j}^{k,l}, \hat{v}_{i,j+1}^{k,l}, \hat{v}_{i+1,j+1}^{k,l}, \hat{v}_{i+1,j}^{k,l})$ represents a quad in a warped mesh; $\mathbf{E}_R$ is the similarity energy; $\mathbf{E}_E$ is the size energy; $\mathbf{E}_O$ is the line bending energy; $\mathbf{W}_Q^k$ is an image saliency map which can be computed from I_k using any saliency detection algorithm and $\mathbf{W}_Q^k(q)$ is the average saliency value of the quad q in the original mesh.

For the similarity energy $\mathbf{E}_R$, we encourage each quad to undergo a similarity transformation [4] and use a quadratic energy term to measure how far the deformation of a quad is from a similarity transformation [7]. More specifically, as shown in Figure 3, by picking any three vertices of a quad in the counter-clockwise direction, taking $\mathbf{v}_0, \mathbf{v}_1$, and $\mathbf{v}_2$ as an example, we can define $\mathbf{v}_0$ by $\mathbf{v}_1$ and $\mathbf{v}_2$ as

$$\mathbf{v}_0 = \mathbf{v}_1 + \mathbf{R}_{90} \, \overrightarrow{\mathbf{v}_1\mathbf{v}_2}, \mathbf{R}_{90} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad (5)$$

After deformation, given vertex positions $\hat{\mathbf{v}}_1$ and $\hat{\mathbf{v}}_2$, we can obtain the expected position of $\mathbf{v}_0$ after deformation as

$$\tilde{\mathbf{v}}_0 = \hat{\mathbf{v}}_1 + \mathbf{R}_{90} \, \overrightarrow{\hat{\mathbf{v}}_1\hat{\mathbf{v}}_2} \, . \quad (6)$$

Ideally, if the quad undergoes a similarity transformation, the expected position $\tilde{\mathbf{v}}_0$ should be identical to $\hat{\mathbf{v}}_0$, the position of $\mathbf{v}_0$ after deformation. Thus, $\mathbf{E}_R(\hat{q})$ can be calculated by summing $(\tilde{\mathbf{v}}_0 - \hat{\mathbf{v}}_0)^2$ for all combinations of three vertices in a quad q.

Inspired by Wang *et al*. [18] and Niu *et al*. [12], it is also important to maintain the sizes and orientations of salient regions. To maintain the original quad size, the energy term $\mathbf{E}_E$ is added to measure the edge length differences. Another energy term $\mathbf{E}_O$ is added to maintain the orientation by measuring the degree of line bending.

$$\mathbf{E}_E(\hat{q}) = (x_{i+1,j}-x_{i,j}-S)^2 + (y_{i,j+1}-y_{i,j}-S)^2 \quad (7)$$
$$+ (x_{i+1,j+1}-x_{i,j+1}-S)^2 + (y_{i+1,j+1}-y_{i+1,j}-S)^2,$$

$$\mathbf{E}_O(\hat{q}) = (x_{i,j+1}-x_{i,j})^2 + (y_{i+1,j}-y_{i,j})^2 \quad (8)$$
$$+ (x_{i+1,j+1}-x_{i+1,j})^2 + (y_{i+1,j+1}-y_{i,j+1})^2,$$

where $(x_{i,j}, y_{i,j})$ is the position of the vertex $\hat{v}_{i,j}^{k,l}$, and S is the width of the original quad.

The image incompleteness term $\mathbf{E}_C$ measures the incompleteness of the final resized stereoscopic image. There could exist holes in the final composited image if some pixels are not covered by any resized object layer. We would

like to reduce holes (uncovered pixels) as much as possible for better visual quality. Thus, the image incompleteness term $\mathbf{E}_C$ is defined as the number of uncovered pixels in the resized images, $\{\hat{I}_L, \hat{I}_R\}$. To count the number of uncover pixels, we obtain the resized object segmentation map by the following steps. First, for each object layer, s_l^k, we define a $w \times h$ mask M_l^k, where

$$M_l^k(x,y) = \begin{cases} l & \text{if } \mathbf{O}_k(x,y) = l, \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

Then we warp each M_l^k by $\hat{V}_l^k$, and denote the warped mask as $\hat{M}_l^k$. With our multi-layer image compositing method (Section 3.1), we can composite all these warped masks $\{\hat{M}_l^k | 1 \le l \le N\}$ to form a resized object segmentation map $\hat{\mathbf{O}}_k$ of size $\hat{w} \times \hat{h}$. Figure 2 shows an example for the resized object segmentation map, $\hat{\mathbf{O}}_L$. With $\hat{\mathbf{O}}_k$, $\mathbf{E}_C$ is defined as:

$$\mathbf{E}_C(\hat{\mathbf{V}}) = \mathbf{Z}(\hat{\mathbf{O}}_L) + \mathbf{Z}(\hat{\mathbf{O}}_R), \quad (10)$$

where the operator $\mathbf{Z}(\cdot)$ counts the number of zero-pixels in the input image.

3.4. Stereoscopic quality energy

In order to maintain good stereoscopic properties, we use two criteria. The first is to maintain the original disparity values as much as possible, and the second is to ensure that there is no vertical offset between the corresponding points across views. From the disparity map, we can obtain a set of corresponding points $\mathbf{F} = \{(p_i^L, p_i^R)\}$, in which p_i^L and p_i^R are a pair of corresponding points between the left and right views. After the mesh deformation defined by $\hat{\mathbf{V}}$, we have a set of warped corresponding points $\hat{\mathbf{F}} = \{(\hat{p}_i^L, \hat{p}_i^R)\}$. To preserve good stereoscopic quality, we require (1) the disparity value between warped corresponding points $\hat{p}_i^L$ and $\hat{p}_i^R$ should be the same as the original disparity between p_i^L and p_i^R and (2) their vertical offset should be zero:

$$\mathbf{E}_S(\hat{\mathbf{V}}) = \sum_{(\hat{p}_i^L, \hat{p}_i^R) \in \hat{\mathbf{F}}} \mathbf{W}_S(p_i^L)(\mathbf{E}_D(\hat{p}_i^L, \hat{p}_i^R) + \lambda_V \mathbf{E}_V(\hat{p}_i^L, \hat{p}_i^R)),$$

$$(11)$$

where $\mathbf{E}_D$ measures disparity consistency and $\mathbf{E}_V$ ensures zero vertical drift:

$$\mathbf{E}_D(\hat{p}_i^L, \hat{p}_i^R) = ((\hat{p}_i^R[x] - \hat{p}_i^L[x]) - (p_i^R[x] - p_i^L[x]))^2 \quad (12)$$

$$\mathbf{E}_V(\hat{p}_i^L, \hat{p}_i^R) = (\hat{p}_i^R[y] - \hat{p}_i^L[y])^2, \quad (13)$$

where the operator $[x]$ extracts the x-component of the input 2-D vector; $[y]$ extracts the y-coordinate; and $\mathbf{W}_S$ is a stereoscopic saliency map to encourage more salient regions to have better stereoscopic quality. We will discuss the stereoscopic saliency map in Section 4.

3.5. Importance energy

With only the image quality term and the stereoscopic quality term, the optimal solution would be cropping as cropping perfectly preserves stereoscopic constrains and introduces neither distortions nor holes. Cropping however is not a preferred solution as it could remove important content. In addition to cropping, layer occlusions could also introduce important content loss as some pixels could become occluded and not shown in the resized image. We would like to reduce content loss as much as possible. For example, content loss due to layer occlusions could be reduced if the layers are repositioned so that they occlude less important areas instead of important ones. We add the importance energy term to ensure that the resized image keeps as much important content as possible.

We assume that each object layer has an importance map, $\mathbf{W}_I^{k,l}$. There are many ways for obtaining such importance maps. For example, objects in the front are often more important than the ones in the back. Thus, we could use a layer's depth order as its importance. It can also be provided by the users or set as the same as the image's saliency map $\mathbf{W}_Q^k$. To measure importance loss, $\mathbf{E}_I$, we first obtain the visibility map [11], which describes whether a pixel in the original image is visible in the resized image. A pixel becomes invisible usually due to occlusions. The importance loss $\mathbf{E}_I$ can then be determined by summing the importance values of all unseen pixels. To determine whether a pixel (x, y) of I_k is visible in the resized image, the pixel is warped from (x, y) to $(\hat{x}, \hat{y})$ by $\hat{\mathbf{V}}_{\mathbf{O}_k(x,y)}^k$, and its visibility can be determined as:

$$A^k(x,y) = \begin{cases} 1, & \text{if } 1 \le \hat{x} \le \hat{m}, 1 \le \hat{y} \le \hat{n} \\ & \text{and } \hat{\mathbf{O}}_k(\hat{x}, \hat{y}) = \mathbf{O}_k(x,y), \\ 0, & \text{otherwise.} \end{cases} \quad (14)$$

With the visibility map A^k, $\mathbf{E}_I$ is defined as:

$$\mathbf{E}_I(\hat{\mathbf{V}}) = \sum_{k \in \{L,R\}} \sum_{(x,y) \in I_k} (1 - A^k(x,y)) \times \mathbf{W}_I^k(x,y) \quad (15)$$

4. Implementation details

Iterative optimization. The major challenge for optimizing $\mathbf{E}(\hat{\mathbf{V}})$ is the calculation of the terms $\mathbf{E}_I$ and $\mathbf{E}_C$. They cannot be parameterized and can only be evaluated by counting pixels or summing importance values in the composited images. Thus, for optimization, we adopt a coarse-to-fine strategy and iteratively update one mesh vertex position at a time by searching for the local minimum within a small neighborhood of the current solution.

To find $\hat{\mathbf{V}}$ which minimizes $\mathbf{E}$, the input images are first scaled down to the coarsest level. We use five levels with a scaling factor of 2 for all results. At the coarsest level, we take uniform scaling as the initial guess. After obtaining the

optimal solution at a coarser level, the resultant $\hat{\mathbf{V}}$ is scaled up into the next finer level and used as the initial guess for that level. At each level, we first fix $\hat{\mathbf{V}}^R$ and update $\hat{\mathbf{V}}^L$. Then, we fix $\hat{\mathbf{V}}^L$ and update $\hat{\mathbf{V}}^R$. We alternatively update between views for T_1 iterations. When updating $\hat{\mathbf{V}}^k$, all N object layers, $\hat{\mathbf{V}}_l^k \in \hat{\mathbf{V}}^k$, are optimized one by one by fixing all other layers. This process is repeated for T_2 iterations. In the current implementation, both T_1 and T_2 are 4.

To find the optimal $\hat{\mathbf{V}}_l^k$ when fixing all other object layers, we iterate through every vertex $\hat{v}_{i,j}^{k,l} \in \hat{\mathbf{V}}_l^k$, and evaluate $\mathbf{E}$ in a small neighborhood around the current solution $\hat{v}_{i,j}^{k,l}$ using a local search. More specifically, we take a uniform grid of samples around the small neighborhood of the current solution and search for the minimum of $\mathbf{E}$ at these samples. The grid of samples can be written as $\hat{v}_{i,j}^{k,l} + (\delta x, \delta y)$ where $\hat{v}_{i,j}^{k,l}$ is the current solution and $(\delta x, \delta y) \in \{(t_x P, t_y P) | t_x, t_y \in \mathbb{Z} \text{ and } -K \le t_x, t_y \le K\}$ That is, we take $(2K+1) \times (2K+1)$ samples with the sampling interval P in each dimension. In our implementation, we used $K = 6$ and $P = 0.25$. We evaluate $\mathbf{E}$ at these samples and update $\hat{v}_{i,j}^{k,l}$ as $\hat{v}_{i,j}^{k,l} + (\delta x, \delta y)$ with the minimum energy among these samples. Although we have to evaluate $\mathbf{E}$ at each sample, fortunately, only one vertex is updated at a time and the updates of $\mathbf{E}_I$ and $\mathbf{E}_C$ are local. They can be implemented efficiently using incremental calculation.

Importance maps. There are three types weighting maps in the energy, each accounting for image quality importance ($\mathbf{W}_Q$), stereo quality importance ($\mathbf{W}_S$), and content importance ($\mathbf{W}_I$). A reasonable choice for $\mathbf{W}_Q$ and $\mathbf{W}_S$ is the image's saliency map. $\mathbf{W}_I$ can be supplied by the users or by the saliency map. In practice, we found that foreground objects usually have higher importance. Thus, the object segmentation map or the estimated disparity map are also reasonable choices for $\mathbf{W}_I$. This observation is the same as the stereoscopic saliency map used by Lang *et al.* [8]. In the current implementation, we use the same map for $\mathbf{W}_I^k$, $\mathbf{W}_Q^k$, and $\mathbf{W}_S^k$. The map is defined as

$$\mathbf{W}^k(x,y) = \begin{cases} 1, & \mathbf{O}_k(x,y) > 1 \\ 0.01, & \text{otherwise.} \end{cases} \quad (16)$$

5. Results

We used the following datasets to test the propose method: *Aloe* from the Middlebury stereo dataset [6], *People*, *Snowman*, and *Man* from *Flicker*[1]. Two methods were compared: a seam carving based approach (ICCV'11) [2] and a warping based approach (TMM) [3]. For all results in the paper, our method took around five minutes. The process can potentially be sped up using parallel processing with GPUs or multithreading.

There are a few adjustable parameters in our method, such as λ_S, λ_I, λ_C, and λ_V. In all results, λ_C is set to

---

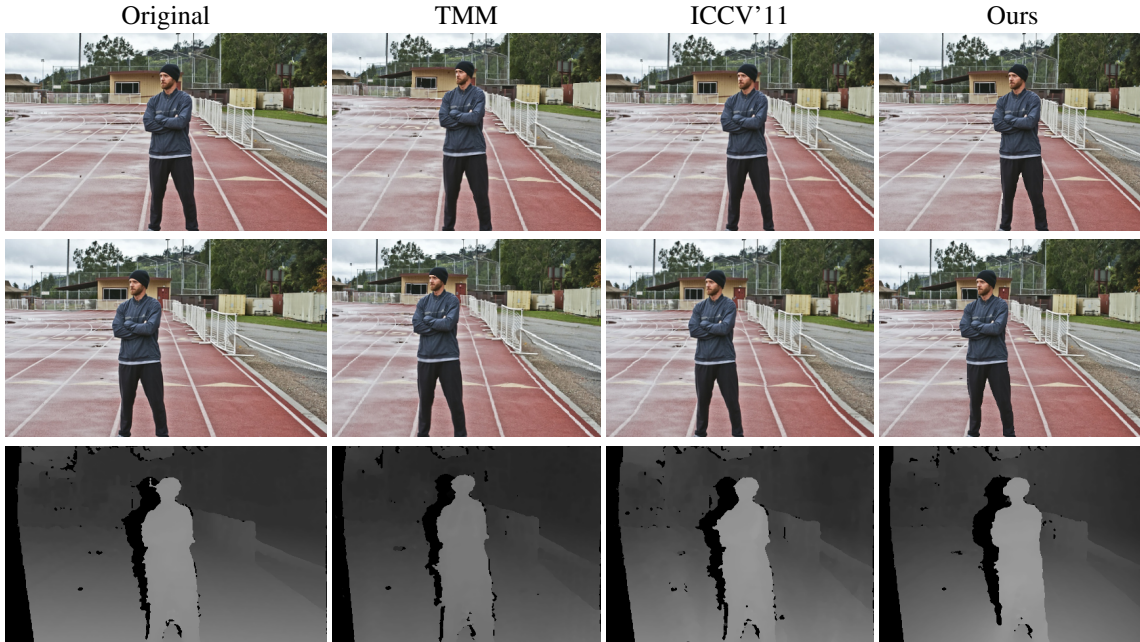[1]Downloaded from the website http://www.eng.tau.ac.il/ talib/Data_SC.html

Figure 4. **Man Dataset.** Reducing width by 17%. (From top to bottom: left view, right view, estimated disparity map)
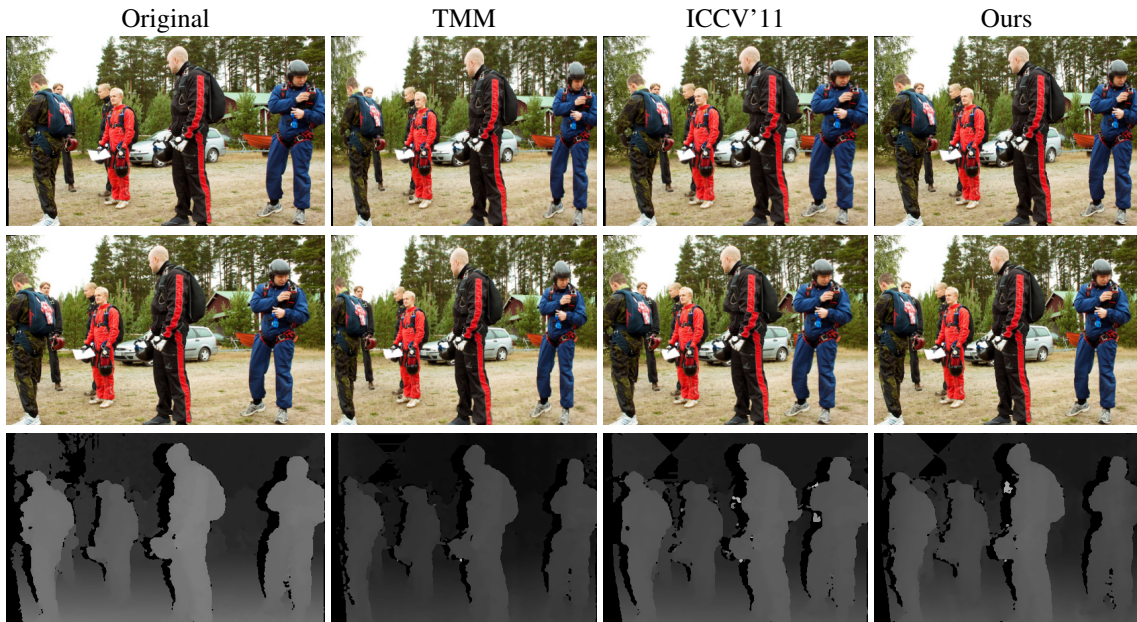

Figure 5. **People Dataset.** Reducing width by 17%. (From top to bottom: left view, right view, estimated disparity map)

a large value of 10^4 because image incompleteness is not preferred in most cases. λ_V controls the degree of vertical offsets between different views, which is crucial for stereoscopic vision, and is also set to a large value of 10^3. As for λ_S and λ_I, we leave them as control options. We start with small values and intuitively adjust them depending on what we prefer: a larger λ_S for better stereoscopic quality, and a larger λ_I for preserving important content better.

A good retargeted stereoscopic result has less distortions and information loss in the left and right views, while preserving original disparity values and depth discontinuities (sharp edges in the disparity map). Based on these criteria, we can compare our method with other approaches. Figure 4 compares our method with ICCV'11 and TMM on *Man*. Note that the white lines in our results remain straight while they become jiggle in ICCV'11. It is the common ar-
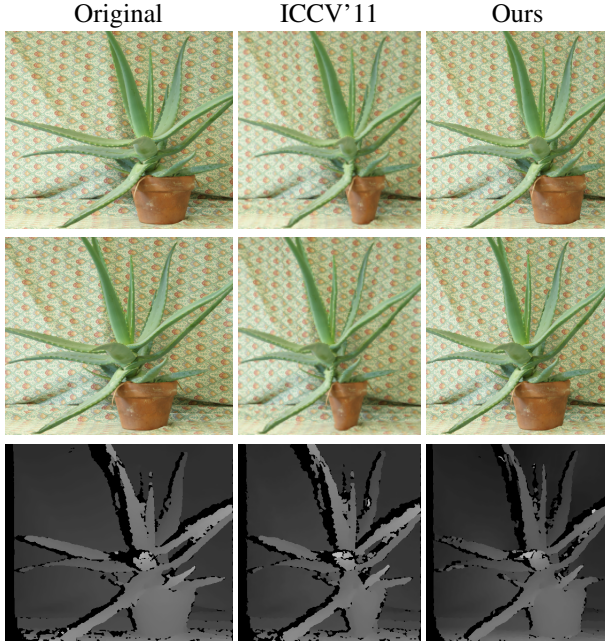
Figure 6. **Aloe Dataset.** Reducing width by 20%. (From top to bottom: left view, right view, estimated disparity map)

Figure 7. **Snowman Dataset.** Reducing width by 17%. (From top to bottom: left view, right view, estimated disparity map)

Figure 8. **Snowman Dataset.** Reducing width by 40%. (From top to bottom: left view, right view, estimated disparity map)

tifacts inherited from discrete methods. TMM distorts the shape of the man, making the head smaller and the lower body fatter. It is because that the warping-based approach uses a single mesh for the whole image. The warping has to make a trade-off between the requirements of the foreground and the background. Our result preserves the shape of man and the lines in the background. Note that, as scene carving, our method could change the relative positions of objects and rearrange objects' spatial relationships. Nevertheless, it still yields a geometrically consistent interpretation of the scene as shown in the disparity map estimated from the resultant left and right images. In Figure 5, again, TMM distorts the shapes of people. ICCV'11 has similar discontinuity artifacts. For example, the boat after the car becomes broken in ICCV'11 results.

Figure 6 compares our method to ICCV'11 on *Aloe*. The discreet nature of ICCV'11 creates very strange shape for the foreground plant and pot. Our method preserves the shapes much better. Figure 7 and Figure 8 compare our method with TMM on *Snowman* with two different size changes. For moderate size change (17% in Figure 7), TMM can produce reasonable results. However, for more intensive size change (40% in Figure 8), TMM introduces shape distortion. In addition, for this case, the disparity range of TMM result is highly compressed and the stereoscopic quality is greatly reduced.

As defined by Basha *et al*. [2], for maintaining geometric consistency, the resized image should have a similar disparity map to the original disparity map. The sharp edges in the disparity map are especially important for good stereo ex-
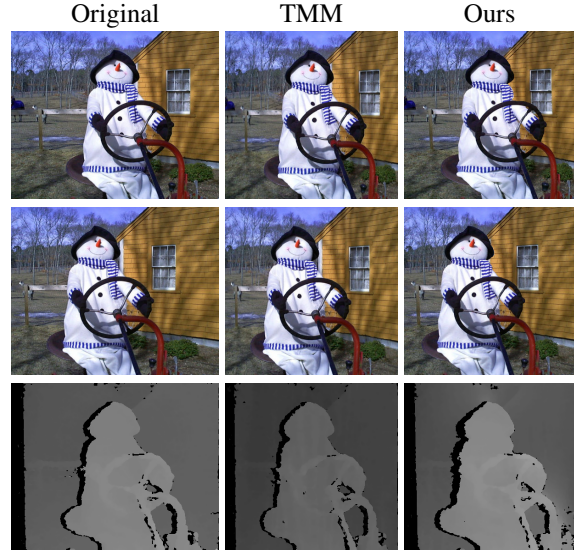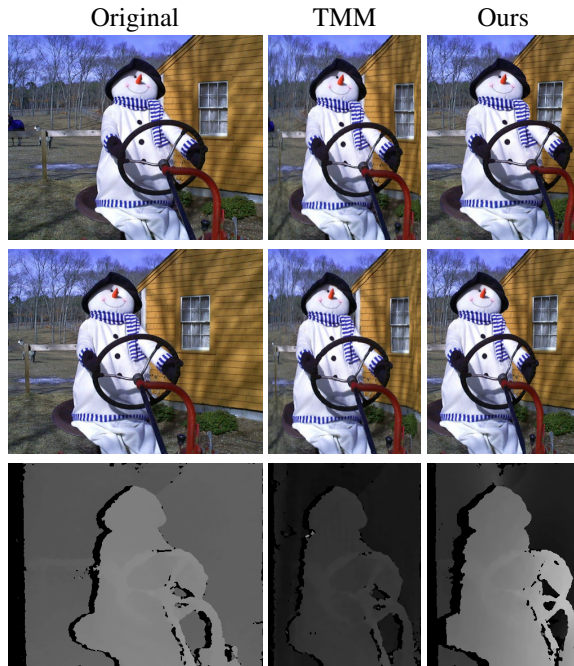
periences. In general, our method has a better disparity map than TMM in terms of geometric consistency. Compared to ICCV'11, our results have less structure discontinuity in appearance. When viewing in 3D with stereoscopic displays, TMM has "rubber sheet" artifacts in depth and worse 3D perception since it does not handle occlusions well. On the other hand, The structure discontinuity artifacts of ICCV'11 often look annoying in appearance when viewing in 3D.
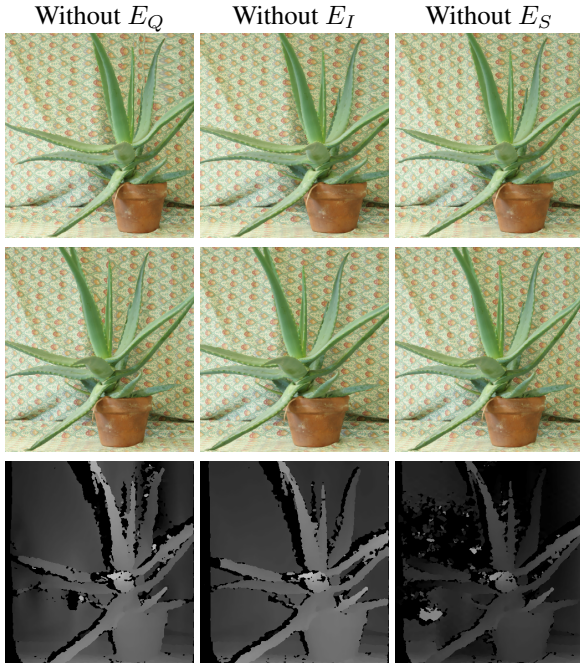
Without E_Q Without E_I Without E_S

Figure 9. Visual contributions of E_Q, E_I and E_S. (From top to bottom: left view, right view, estimated disparity map)

There are three major terms, E_Q, E_S and E_I, in our formulation (Equation 1). If E_Q is removed, the image could be distorted and have holes for preserving more important content (E_I) and maintaining better stereo correspondences (E_S). If E_S is removed, the stereo correspondences might not be maintained well. The perceived depths could be distorted, or even worse the viewer cannot fuse images to have 3D perception. If E_I is removed, the method tends to crop the images to the target size as it perfectly preserves image and stereo qualities. This, however, might remove important contents. Figure 9 evaluates visual contributions of these terms. The results show that the content will be more distorted if E_Q is removed; the content will be cropped if E_I is removed; and the estimated disparity map could deviate much from the original disparity map if E_S is removed.

6. Conclusions

This paper proposes scene warping, a layer-based stereoscopic image resizing method using image warping. The input stereoscopic image is decomposed into layers. Each layer is warped by its own mesh deformation and the warped layers are composited together to form the resized images. The energy function of scene warping enforces the resized image with less distortions and holes, good stereoscopic properties and important content. Compared to existing methods, scene warping offers the advantages of less discontinuous artifacts, less-distorted objects, correct depth ordering and enhanced stereoscopic quality. Our method suffers from the common limitations shared by warping-

based methods. When parts of the input image are crowded with important objects, our method could crop or occlude important objects. In the future, we would like to explore methods to relax this restriction.

References

[1] S. Avidan and A. Shamir. Seam carving for content-aware image resizing. *ACM Trans. Graph.*, 26(3):10, 2007.

[2] T. Basha, Y. Moses, and S. Avidan. Geometric consistent stereo seam carving. In *Proceedings of ICCV*, 2011.

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

[4] R. Gal, O. Sorkine, and D. Cohen-Or. Feature-aware texturing. In *Proceedings of Eurographics Symposium on Rendering*, pages 297–303, 2006.

[5] H. Hirschmüller. Accurate and efficient stereo processing by semi-global matching and mutual information. In *Proceedings of CVRP*, pages 807–814, 2005.

[6] H. Hirschmüller. Evaluation of cost functions for stereo matching. In *Proceedings of CVPR*, 2007.

[7] T. Igarashi and T. M. J. F. Hughes. As-rigid-as-possible shape manipulation. *ACM Trans. Graph.*, 24(3):1134–1141, 2005.

[8] 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, 2010.

[9] W.-Y. Lo, J. van Baar, C. Knaus, M. Zwicker, and M. Gross. Stereoscopic 3D copy & paste. *ACM Trans. Graph.*, 29(6):147, 2010.

[10] A. Mansfield, P. Gehler, L. V. Gool, and C. Rother. Scene carving: Scene consistent image retargeting. In *Proceedings of ECCV*, 2010.

[11] A. Mansfield, P. Gehler, L. Van Gool, and C. Rother. Visibility maps for improving seam carving. In *Media Retargeting Workshop, European Conference on Computer Vision (ECCV)*, 2010.

[12] Y. Niu, F. Liu, X. Li, and M. Gleicher. Warp propagation for video resizing. In *Proceedings of CVPR*, pages 537–544, 2010.

[13] C. Rother, V. Kolmogorov, and A. Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.*, 23(3):309–314, 2004.

[14] M. Rubinstein, A. Shamir, and S. Avidan. Improved seam carving for video retargeting. *ACM Trans. Graph.*, 27(3):16, 2008.

[15] M. Rubinstein, A. Shamir, and S. Avidan. Multi-operator media retargeting. *ACM Trans. Graph.*, 28(3):23, 2009.

[16] A. Shamir and O. Sorkine. Visual media retargeting. In *ACM SIGGRAPH Asia 2009 Course Notes*, 2009.

[17] L. Wolf, M. Guttmann, and D. Cohen-Or. Non-homogeneous content-driven video retargeting. In *Proceedings of ICCV*, 2007.

[18] O. S. Yu-Shuen Wang, Chiew-Lan Tai and T.-Y. Lee. Optimized scale-and-stretch for image resizing. *ACM Trans. Graph.*, 27(5):118, 2008.