

Image Retargeting Using Mesh Parametrization

Yanwen Guo, Feng Liu, Jian Shi,
Zhi-Hua Zhou, *Senior Member, IEEE*, and Michael Gleicher

Abstract—Image retargeting aims to adapt images to displays of small sizes and different aspect ratios. Effective retargeting requires emphasizing the important content while retaining surrounding context with minimal visual distortion. In this paper, we present such an effective image retargeting method using saliency-based mesh parametrization. Our method first constructs a mesh image representation that is consistent with the underlying image structures. Such a mesh representation enables easy preservation of image structures during retargeting since it captures underlying image structures. Based on this mesh representation, we formulate the problem of retargeting an image to a desired size as a constrained image mesh parametrization problem that aims at finding a homomorphous target mesh with desired size. Specifically, to emphasize salient objects and minimize visual distortion, we associate image saliency into the image mesh and regard image structure as constraints for mesh parametrization. Through a stretch-based mesh parametrization process we obtain the homomorphous target mesh, which is then used to render the target image by texture mapping. The effectiveness of our algorithm is demonstrated by experiments.

Index Terms—Image retargeting, mesh parametrization, attention model.

1 INTRODUCTION

Mobile devices, such as cellular phones and PDAs, are increasingly common. These devices often have image display functionality. Normally, images have much higher resolutions and different aspect ratios than the small screens of these mobile devices. Those images need to be adapted to fit the target displays. We define the problem of adapting an image to various target screens as image retargeting.

A common solution to image retargeting is to uniformly rescale the original image according to the target screen size. This naive scaling is problematic. The important objects in image maybe become too small to be recognized. Moreover, if the original image has a different aspect ratio than target screen, distortion is introduced into the result. Alternatively, an important image region is cropped and displayed [3], [21], [25]. The disadvantage of cropping is the loss of contextual content which is important to appreciate the image. When there exist multiple important objects that are faraway from each other, cropping inevitably loses some of them in order to keep a necessary resolution of the selected object.

Recent methods achieve focus+context image retargeting by non-uniform warping [12] or segmentation-based image composition [24], [23]. The former suffers from image distortion and the latter is subject to performance of segmentation. The seam carving method achieves image resizing by iteratively carving less noticeable seams [2].

Yanwen Guo, Jian Shi, and Zhi-Hua Zhou are with the National Key Lab for Novel Software Technology, Nanjing University, Nanjing 210093, P. R. China.
Email: ywguo@nju.edu.cn, shilenjian@gmail.com, zhouzh@nju.edu.cn

Feng Liu, and Michael Gleicher are with the Department of Computer Sciences, University of Wisconsin-Madison.

Email: {fliu, gleicher}@cs.wisc.edu

Manuscript received Sep. 1st, 2008; revised Mar. 2nd, 2009.

Although it tries to remove less noticeable seams, it still produces artifacts like breaking objects since image structures have not been explicitly considered. Similarly, the non-homogenous retargeting method [29] tries to achieve seamless retargeting results without resorting to image segmentation. It, however, fails to take image structures into account, and sometimes breaks important object shapes.

Effective image retargeting should emphasize important content while retaining surrounding context with minimal visual distortion. Given the limited target size and different aspect ratios, we necessarily have to introduce distortion into the target image, in order to emphasize important content and meanwhile to retain the context. The challenge is to minimize the visual distortion. In this paper, we propose to represent the image by a mesh that is consistent with the underlying image structures, and then adapt images via mesh transformation. This mesh representation enables us to easily preserve image structures during retargeting. Also adapting images via transforming mesh ensures smooth image transformation. This helps to achieve effective image retargeting. Specifically, in this paper, we define image retargeting as a mesh parametrization problem, which aims to find a homomorphous mesh with the target size. We first build a feature relevant controlling mesh from the source image. To emphasize salient objects, we associate image saliency with the mesh, and accordingly encourage the mesh to morph in such a way that the size of a salient mesh cell (triangle) is reduced less than non-salient ones. To minimize visual distortion, our method encodes preservation of salient objects and image structure as constraints during mesh parametrization. The target mesh is solved using a stretch minimizing parametrization scheme. The target image is finally rendered by texture mapping.

The main contribution of this paper is an effective image retargeting method which has the following benefits.

- Our method emphasizes important image content while retaining the surrounding context with minimal visual distortion. Mesh representation endows salient objects and strong structures with compact representations. Emphasis of salient objects as well as preservation of structures are simultaneously achieved by encoding them as constraints of parametrization during retargeting.
- Image with multiple salient objects can be easily retargeted via mesh parametrization. Retargeted positions of salient objects are automatically determined by the mesh parametrization process, rather than placing objects' positions beforehand as previous methods have done. Furthermore, exaggerated scales of salient objects can be determined automatically or controlled by user with simple parameter setting.

The rest of this paper is organized as follows. We give a brief overview of previous work in Section II. In Section III, we describe the retargeting approach using saliency-based mesh parametrization in detail. We examine our method in Section IV and conclude the paper in the last Section.

2 RELATED WORK

The problem of retargeting images to small screens has been considered by many researchers. Almost all image browsers on small-screen devices provide scaling functionality. A large image is uniformly downsampled to fit the target screens. Important objects in the image are often small and difficult to identify. To account for different aspect ratios, either the distortion caused by aspect ratio change is introduced or black letter box is used to fill in the blank space, thus wasting the precious space.

Many methods identify an important region in the image and crop this region to fit target screen. These methods usually rely on image analysis, such as attention model extraction and object detection, to identify the important region [3], [7], [9], [8], [17]. Eye-tracking systems can also be used to infer the important region [21]. After obtaining the important region, it is cropped out [3], [21], [25]. These cropping-based retargeting methods can effectively preserve important information of the original image, at the expense of losing valuable context information. Moreover, their performance is dependent on the accuracy of important region detection. Alternatively, when multiple important objects are identified, a tour on the image can be created by jumping through each object [10], [16]. While this scheme is effective in some situations, it is infeasible when viewer must examine many images within limited time.

Recently, Liu and Gleicher [12] presented a fish-eye view warping scheme to achieve the goal of emphasizing the important region while retaining the surrounding

context. This method can effectively achieve the focus+context view of the original image. However, it creates noticeable distortion, especially in the contextual region. Setlur et al. [24] proposed to segment the prominent objects and to re-compose them onto resized background. This method can achieve focus+context at the minimal distortion. However, the performance is subject to that of image segmentation. In [13], Gal et al. presented a Laplacian editing based texturing method that allows any image to be warped in a feature-aware manner. One difference between their method and the proposed approach is that they employ regular grid to represent input image, while we use triangular meshes. Using irregular triangular meshes enables a better approximation of structured edges, thus facilitating the preservation of them, especially for slant edges. A similar quad grids-based method is given by Wang et al. [14], which computes the optimal target grids by distributing warping scales non-uniformly among grid lines according to image content. An efficient formulation for the non-linear optimization, which allows interactive image resizing, has also been developed. Most recently, Avidan and Shamir [2] proposed a nice “seam carving” method for resizing images. Their method iteratively carves unnoticeable seams to reduce the image size. However, it often breaks image structures since the carving procedure only relies on low-level saliency information and fails to consider image structures.

Retargeting videos has also been considered [5], [11], [15], [20], [27], [29]. For example, Fan et al. [5] used a visual attention model to find important regions, and provided automatic, semi-automatic and manual modes for users to select and zoom into these important regions while browsing. Liu and Gleicher [15] utilized cinematography rules and designed a video retargeting system to automate pan and scan operation. Both Setlur et al. [23] and Cheng et al. [4] proposed context recomposition for preserving salient information. These methods first segment salient objects based on the detected saliency prior. Then they fill in the holes left after object segmentation. Finally, they compose salient objects back onto the resized background video frames based on a set of aesthetic criteria. Wolf et al. introduced an optimization algorithm for non-homogeneous video retargeting [29]. Considering spatial-temporal constraints, a per-pixel discrete transformation that shrinks less important pixels is utilized to resize video sequences.

3 RETARGETING USING MESH PARAMETRIZATION

Image retargeting attempts to adapt a source image to display of a smaller size and different aspect ratio than originally intended. We formulate image retargeting as a mesh parametrization problem that emphasizes the important content while retaining the surrounding background with slight distortion. Fig. 1 illustrates the outline of our algorithm. To emphasize salient objects in

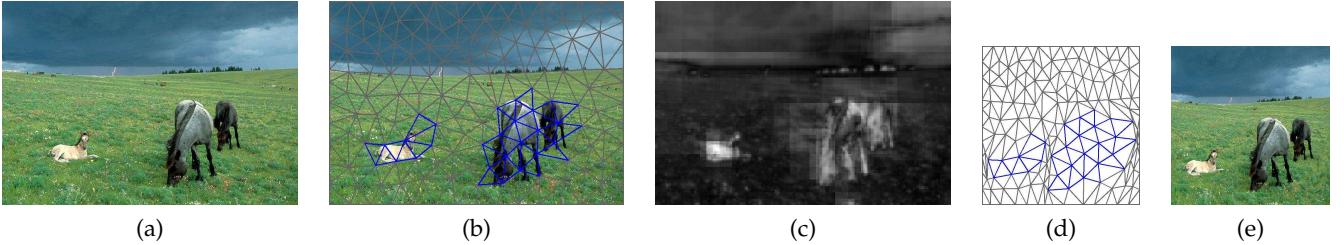


Fig. 1. Algorithm overview. Our algorithm first builds a feature relevant controlling mesh (b) from the input (a). The source mesh is then associated with image saliency information (c) to emphasize the salient objects, marked with blue edges in (b). Afterwards, the target mesh with desired resolution is produced by solving a mesh parametrization problem (d). Standard texture mapping is finally used to render target image as shown in (e).

image, we first calculate a saliency map of the source image (Section 3.1). Then we build a controlling mesh of the input image that is consistent with the underlying image structures, and associate the saliency information with the source mesh (Section 3.2). In this way, retargeting is transformed into a parametrization problem of finding a target mesh with the desired resolution. In order to improve visual quality of the target image, we interpret preservation of saliency and image structure information as constraints of parametrization (Section 3.3). Afterwards, the target mesh is solved by a constrained stretch-based mesh parametrization scheme (Section 3.4). Retargeting result is finally rendered using the standard texture mapping algorithm.

3.1 Importance Computation

The information of important objects in an image is necessary for image retargeting to emphasize those objects. Although semantic image understanding is beyond the state-of-the-art, heuristic rules have been used successfully as an alternative (c.f. [16], [3], [23], [12], [25]). Like previous work, we use the low-level saliency information as well as high-level object information to infer important objects.

Saliency has been used in many applications to detect the important region in an image. It has been extensively explored in the past several years [9], [8], [3], [17]. We use the contrast-based saliency detection method proposed by Ma et al. [17] to compute a normalized saliency value for each image pixel. Based on heuristic rules about human visual perception, this method calculates the visual-feature contrast as saliency.

Human face and body are usually important in photos. We employ the Viola-Jones face detector to detect faces [26]. Besides, We have developed a method that uses the located face to automatically estimate the body. We estimate one rough head-shoulder mask covering human body and then extract human body using an iterative Graph Cut algorithm [28]. Specially, saliency values for those pixels in human body are set 1. In addition, users can also specify important objects that are hard to detect automatically.

3.2 Image Representation

Generating a mesh from an image has been used in some multimedia applications like video compression [1]. The key issue is to generate a mesh that is consistent with the input image structures. Inspired by the previous work [1], we first detect feature points from the input image I_s , and then use Delaunay triangulation [22] algorithm to generate the mesh as follows:

- We first evenly discretize the input image boundary, and use all the points there as part of the feature points.
- We employ an edge detection operator, for example Canny operator, for extracting some feature points. Note that, only part of Canny detected pixels are kept as mesh points using a distance threshold. For keeping uniformity of point density, some auxiliary ones are usually added, as mesh with nearly uniform density normally facilitates its processing.
- Finally, we use the constrained Delaunay triangulation algorithm [22] to generate a feature-consistent mesh \mathcal{M}_s (Fig. 2).

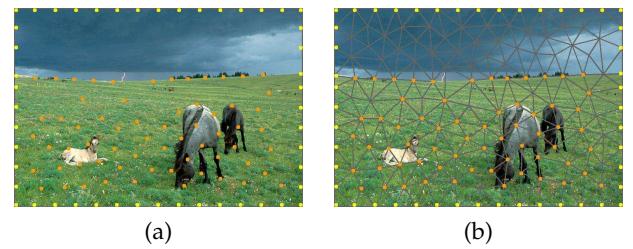


Fig. 2. Mesh generation. (a) Input image with boundary (the yellow dots) and feature points (orange ones) evenly distributed. (b) Delaunay triangulation result. The rest mesh points are added automatically with a distance threshold to maintain mesh uniformity.

For clarity of exposition, we define here some notations of the mesh. Let \mathcal{M}_t be the target mesh to be solved. $\{P_i = (x_i, y_i)|i = 1, \dots, n\}$ and $\{Q_i = (u_i, v_i)|i = 1, \dots, n\}$ represent the points of \mathcal{M}_s and \mathcal{M}_t separately. $\{Q_i|i = 1, \dots, n\}$ are to be solved in \mathcal{M}_t , and Q_i is the counterpart of P_i of \mathcal{M}_s . $\Delta_P = \triangle(P_i, P_j, P_k)$ and $\Delta_Q = \triangle(Q_i, Q_j, Q_k)$ are corresponding triangle pairs of \mathcal{M}_s and \mathcal{M}_t . Due to the same topology of \mathcal{M}_t and \mathcal{M}_s , edge $e(Q_iQ_j)$ in \mathcal{M}_t corresponds to edge $e(P_iP_j)$ in \mathcal{M}_s .

In Fig. 1, (b) is \mathcal{M}_s produced using the above scheme and (d) is \mathcal{M}_t solved by parametrization.

In order to treat image regions discriminatively, we associate the source mesh \mathcal{M}_s with the saliency map of I_s . The salient value S_{Δ_P} of a triangle Δ_P is calculated by averaging saliency values of all pixels in Δ_P . With these associated saliency information, we can perform mesh parametrization in such a way that salient edges shrink less than those less salient ones. In this way, salient objects are emphasized.

3.3 Constraints in Parametrization

Retargeting should completely preserve salient objects that are immune from even slight distortion. Meanwhile, emphasize them to some extent. Furthermore, image structures in the remaining regions, which show in terms of strong edges, are important visual features. They shall be retained as-rigid-as possible. To achieve these goals, we define here several constraints, and incorporate them into mesh parametrization. The constraints include boundary constraint ensuring boundary consistency, saliency constraint emphasizing salient objects and avoiding distorting them, and structure constraint preserving strong edges as-rigid-as possible.

3.3.1 Boundary constraint

The source image boundary shall be mapped onto that of the target screen. Therefore, we require that the target mesh boundary should be adhere to the target screen boundary by defining the following constraints.

First, we map the points of \mathcal{M}_t which correspond to four corner points of \mathcal{M}_s onto the target screen's corners by fixing those corner points. For the other non-corner points on the up and bottom boundaries, the v -coordinates of \mathcal{M}_t 's boundary points are fixed as the corresponding y -coordinates of the screen's up and bottom boundary respectively. While their u -coordinates are to be solved by parametrization. Similarly, for those non-corner points on the left and right boundary, we fix their u -coordinates while solving for their v -coordinates (Fig. 3).

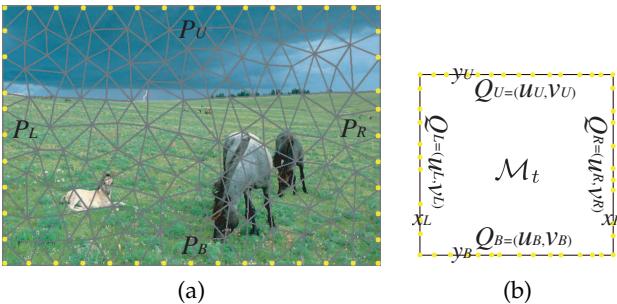


Fig. 3. Boundary constraint. (a) Boundary points P_U , P_B , P_L , and P_R of \mathcal{M}_s should be mapped onto boundary of the target screen. (b) Accordingly, v -coordinates of Q_U and Q_B are set as y_U and y_B separately, and we solve for their u -coordinates. For Q_L and Q_R , opposite case exists.

Suppose that \mathcal{M}'_s 's up, bottom, left, and right boundary points are $Q_{U,i|i=1,..n_U}$, $Q_{B,i|i=1,..n_B}$, $Q_{L,i|i=1,..n_L}$, and $Q_{R,i|i=1,..n_R}$ respectively. Correspondingly, target screen's up and bottom y -coordinates are y_U and y_B . Its left and right x -coordinates are x_L and x_R . We express the above boundary constraints as:

$$\begin{aligned} F_B = & \sum_{i=1}^{n_U} |v_{U,i} - y_U| + \sum_{i=1}^{n_B} |v_{B,i} - y_B| + \\ & \sum_{i=1}^{n_L} |u_{L,i} - x_L| + \sum_{i=1}^{n_R} |u_{R,i} - x_R| = 0. \end{aligned} \quad (1)$$

3.3.2 Saliency constraint

We aim to emphasize salient objects in an image and minimize their distortion. Since automatically obtaining accurate object boundary is difficult, we use a conservative strategy. Specifically, we cluster triangles with saliency above a given threshold S_μ as an object. Generally, we try to include non-object area rather than miss a triangle that is a part of the object by using a small threshold. An object is defined as follows:

$$\{O_{sj} = \bigcup_{i=1}^{n_j} (\Delta_i > S_\mu) | j = 1, \dots, J\}, \quad (2)$$

where O_{sj} is a salient object in the source image with n_j triangles. J is the number of salient objects. Through extensive experiments, S_μ is set to 0.6 empirically.

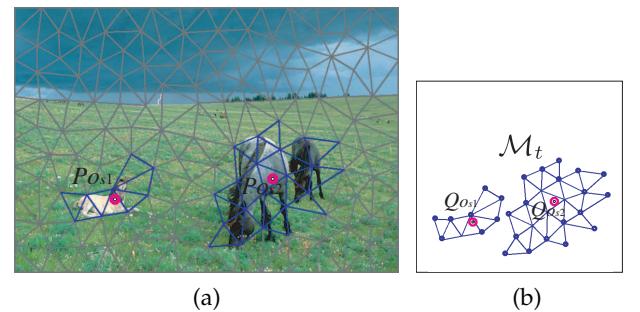


Fig. 4. Saliency constraint. (a) The input image contains two salient objects. (b) Each salient object undergoes rigidly scale and translation transform. Its retargeted part is determined by mass point's position, together with a scale factor encoded in polar coordinate transformation.

We define the following constraint to avoid distortion of objects. Specifically, only a scaling transformation is allowed for every salient object O_{sj} . To achieve this goal, a mass-center point $P_{O_{sj}}$ is first calculated by averaging all points in O_{sj} of \mathcal{M}_s . Afterwards, the polar coordinates (r_i, θ_i) of P_i in O_{sj} is obtained by taking $P_{O_{sj}}$ as the pole. During retargeting, O_{sj} undergoes rigid transformation with a scale factor $z_{O_{sj}}$. Hence the transformed polar coordinates of P_i equal $(z_{O_{sj}} \cdot r_i, \theta_i)$. If given $Q_{O_{sj}}$, the target position of $P_{O_{sj}}$ after retargeting, we can convert the polar coordinates into the target Euclidean coordinates of Q_i , the corresponding target point of P_i .

(Fig. 4). Assume that the formula of transforming polar coordinates to Euclidean coordinates is f , we then have:

$$Q_i = f(Q_{O_{sj}}, z_{O_{sj}} \cdot r_i, \theta_i), \quad (3)$$

We can represent the above saliency constraint as:

$$F_{SA} = \sum_{j=1}^J \left(\sum_{i=1}^{n_j} |Q_i - f(Q_{O_{sj}}, z_{O_{sj}} \cdot r_i, \theta_i)| \right) = 0, \quad (4)$$

where retargeted part of O_{sj} has n_j points $Q_{|i|=1..n_j}$.

To emphasize salient objects, relative scales of salient objects should be exaggerated in contrast to other regions so that salient ones have higher resolution. We calculate the scale factor $z_{O_{sj}}$ as follows. For a single salient object in image, $z_{O_{sj}}$ can be determined by restricting the retargeted salient object within the target screen. If several salient objects exist, total area of retargeted objects should not exceed that of the target screen. Furthermore, the summed width and height of retargeted objects should not exceed those of the target screen as well.

In parametrization, $Q_{O_{sj}}$ is the only unknown parameter for salient object O_{sj} to be solved by parametrization. All points of retargeted salient object are encoded by it in (3). Once $Q_{O_{sj}}$ is known, they can be figured out using (3).

3.3.3 Structure constraint

Strong edges are important visual features. They are vital clues for understanding image content, and should be maintained as-rigid-as possible. As salient objects are preserved completely by using the above saliency constraint, we only address here strong edge segments in the rest regions. We detect those strong edge segments using Hough transform. The segments detected are as follows:

$$\{L_k = (\bigcup_{i=1}^{m_k} P_{k,i}) | k = 1, \dots, K\}. \quad (5)$$

where L_k is the segment passing through points $P_{k,i|i=1..m_k}$ of \mathcal{M}_s as illustrated in Fig. 5 (a). K is the number of segments. As Hough transform may produce too many trivial segments, we filter out those segments by requiring that the length of each segment exceeds a given threshold. The default value of this threshold is set to 0.1 fold of the minimum of w_s and h_s . Furthermore, only one segment is reserved if several segments are close enough. In implementation, Hough transform is first exerted and the resulting segments are discretized into points before Delaunay triangulation for ensuring that mesh edges adhere to strong edges.

We require that $Q_{k,i|i=1..m_k}$, target points of $P_{k,i|i=1..m_k}$, still pass through a line. We call it linearity constraint. Assume the retargeted lines are:

$$\{RL_k = (\bigcup_{i=1}^{m_k} Q_{k,i}) | k = 1, \dots, K\}. \quad (6)$$

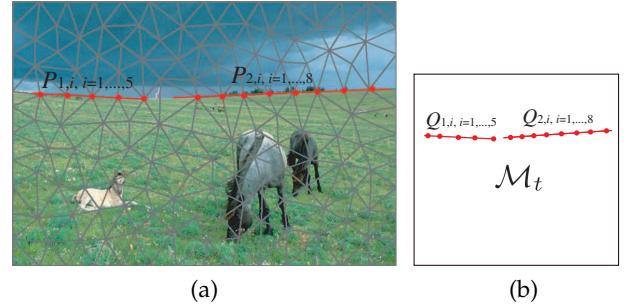


Fig. 5. Structure constraint. (a) Two feature point sets $P_{1,i, i=1..5}$ and $P_{2,i, i=1..8}$ adhere to strong edges detected in input image. (b) Their retargeted sets $Q_{1,i, i=1..5}$ and $Q_{2,i, i=1..8}$ should pass through lines as well. We interpret this as soft constraint.

RL_k is the retargeted line of L_k by the expression $y = a_k \cdot x + b_k$. $Q_{k,i}$ satisfies $v_{k,i} = a_k \cdot u_{k,i} + b_k$. Here a_k and b_k can be easily represented by two points in RL_k .

In practice, however, some line segments have to violate the linearity constraint in order to emphasize important objects in the input image. So we define the following soft energy constraint and encourage meet of the above linearity constraint as much as possible:

$$E_{ST} = \sum_{k=1}^K \left(\sum_{i=1}^{m_k} (v_{k,i} - a_k \cdot u_{k,i} + b_k)^2 \right). \quad (7)$$

We have defined several constraints. In following, we show how to integrate them into parametrization as soft or hard constraints during parametrization.

3.4 Constrained Mesh Parametrization

With the source mesh \mathcal{M}_s , we formulate retargeting as a parametrization problem that tries to find a target mesh \mathcal{M}_t . \mathcal{M}_t has the same topology as \mathcal{M}_s and the target screen size. Mesh parametrization has been extensively studied in computer graphics. Its original intention is to establish the correspondence between unordered 3D mesh and the mesh on parametric domain, e.g., plane or sphere. Most methods typically seek solutions by minimizing the metric deformation involved in parametrization. Till now, many different metric criteria have been brought forward. Among them, stretch-based methods usually work well when reducing the global mesh distortion [18], [30]. The basic observation is that from perspective of discrete computational geometry, the mesh can be fully determined by the lengths of all its edges. We extend traditional 3D-to-2D mesh parametrization to 2D-to-2D, and solve \mathcal{M}_t using one constrained stretch-based parametrization scheme.

3.4.1 Computation of edge lengths of target mesh

\mathcal{M}_t is completely characterized by the lengths of its edges. As edges of salient objects are determined by their scale factors, we describe the method of computing the ideal lengths of remaining edges of \mathcal{M}_t . Such ideal

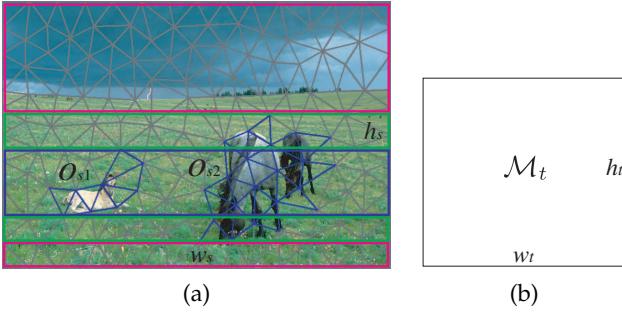


Fig. 6. Scale factor for each edge is determined by resolution variance of retargeting, as well as its position with respect to salient objects.

lengths will serve as the objective stretches in parametrization. Reasonable length setting normally facilitates the parametrization process, leading to topology valid target mesh.

Intuitively, edges of \mathcal{M}_t are transformed from their counterparts of \mathcal{M}_s with respect to resolution variation of retargeting. Transforming scale of each edge is also subject to exaggerated scales of salient objects. We address such scale in x and y directions separately. To determine transforming scale in x direction, the mesh is divided according to salient objects' positions as shown in Fig. 6 (a). We first compute a virtual scale for each mesh point, then compute the transforming scale for one edge by averaging scales of its two points.

Let $w_s \times h_s$ and $w_t \times h_t$ represent resolutions of the source image and target screen separately. It is obvious that if a point $P_i = (x_i, y_i)$ lies in the red rectangle in Fig. 6 (a), its ideal scale should be w_t/w_s . Otherwise, if it lies in the green or blue rectangle, the scale is also related to the scale factors of salient objects. Assume that segment lengths formed by the line $y = y_i$ intersecting with salient objects are $\{l_{O_{sj}} | j = 1, \dots, J\}$. The ideal scale factor of P_i is

$$(w_t - \sum_{j=1}^J z_{O_{sj}} * l_{O_{sj}}) / (w_s - \sum_{j=1}^J l_{O_{sj}}). \quad (8)$$

where $z_{O_{sj}}$ is the scale factor of O_{sj} . For instance the mesh of Fig. 6 (a), if P_i lies in the red rectangle, $l_{O_{s1}} = l_{O_{s2}} = 0$. If it lies in the green rectangle, $l_{O_{s1}} = 0$ and $l_{O_{s2}}$ is positive. Otherwise both $l_{O_{s1}}$ and $l_{O_{s2}}$ are positive.

We can similarly compute the virtual scale for P_i in y direction. For each edge $e(P_i P_j)$, its scale factor is calculated by averaging the scale factors of P_i and P_j in x and y directions separately. Let (s_{xij}, s_{yij}) be $e(P_i P_j)$'s scale factor.

Assume that x and y directional lengths of edge $e(P_i P_j)$ are (l_{xij}, l_{yij}) . Taking the above $e(P_i P_j)$'s scale factor into account, the reasonable length configuration of edge $e(Q_i Q_j)$ in \mathcal{M}_t is:

$$l_{ij} = l_{e(Q_i Q_j)} = \sqrt{(s_{xij} * l_{xij})^2 + (s_{yij} * l_{yij})^2}. \quad (9)$$

3.4.2 Mesh parametrization

\mathcal{M}_t can be inferred from its edge lengths by minimizing the following energy function:

$$E_l = \sum_{(Q_i Q_j) \in \text{edges}} (||Q_i - Q_j||^2 - l_{ij}^2)^2 / l_{ij}^2. \quad (10)$$

For salient object O_{sj} , stretching terms about edges within its retargeted part are precluded from (10). Only the terms about edges connecting their boundary points with outer points take effect. Boundary points are encoded by retargeted pole of O_{sj} as described in subsection 3.3.2.

Taking symmetry of the above equation into account, energy gradients on point Q_i are:

$$\frac{\partial E_l}{\partial u_i} = 8 \sum_{(Q_i Q_j) \in \text{edges}} (||Q_i - Q_j||^2 - l_{ij}^2) \cdot (u_i - u_j) / l_{ij}^2 \quad (11)$$

$$\frac{\partial E_l}{\partial v_i} = 8 \sum_{(Q_i Q_j) \in \text{edges}} (||Q_i - Q_j||^2 - l_{ij}^2) \cdot (v_i - v_j) / l_{ij}^2 \quad (12)$$

Obviously, the above equations can be easily solved. But when \mathcal{M}_t is excessively dense, directly solving the above equations may cause adjacent triangles of \mathcal{M}_t to flip over, leading to invalid topology. This is caused by inverting the orientation of the triangle points. To tackle this issue, we revise the energy function by penalizing reversion of triangle orientation with *sign* function [18].

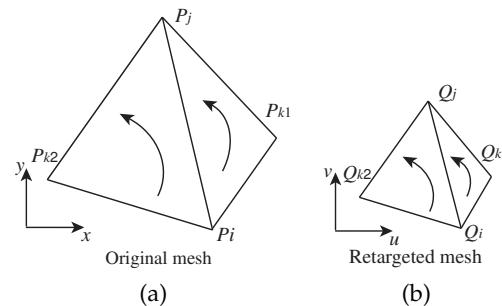


Fig. 7. Orientations of the triangles (b) in retargeted mesh should keep consistent with orientations of their corresponding ones (a) in original mesh.

Assume that $\triangle_{Q1} = \triangle(Q_i Q_{k1} Q_j)$, $\triangle_{Q2} = \triangle(Q_i Q_j Q_{k2})$ are two adjacent triangles incident upon edge $e(Q_i Q_j)$. Their corresponding triangles in \mathcal{M}_s of the source image are $\triangle_{P1} = \triangle(P_i P_{k1} P_j)$, and $\triangle_{P2} = \triangle(P_i P_j P_{k2})$ (Fig. 7). For each pair of corresponding triangles, the orientations of points should be equal. To achieve this, we define:

$$w_{ij} = \text{sign} \min(\det(\overrightarrow{Q_i Q_{k1}}, \overrightarrow{Q_j Q_{k1}}) \cdot \det(\overrightarrow{P_i P_{k1}}, \overrightarrow{P_j P_{k1}}), \det(\overrightarrow{Q_i Q_{k2}}, \overrightarrow{Q_j Q_{k2}}) \cdot \det(\overrightarrow{P_i P_{k2}}, \overrightarrow{P_j P_{k2}})). \quad (13)$$

The new energy function is defined as follows accordingly:

$$E_l = \sum_{(Q_i Q_j) \in \text{edges}} (w_{ij} \cdot ||Q_i - Q_j||^2 - l_{ij}^2)^2 / l_{ij}^2, \quad (14)$$

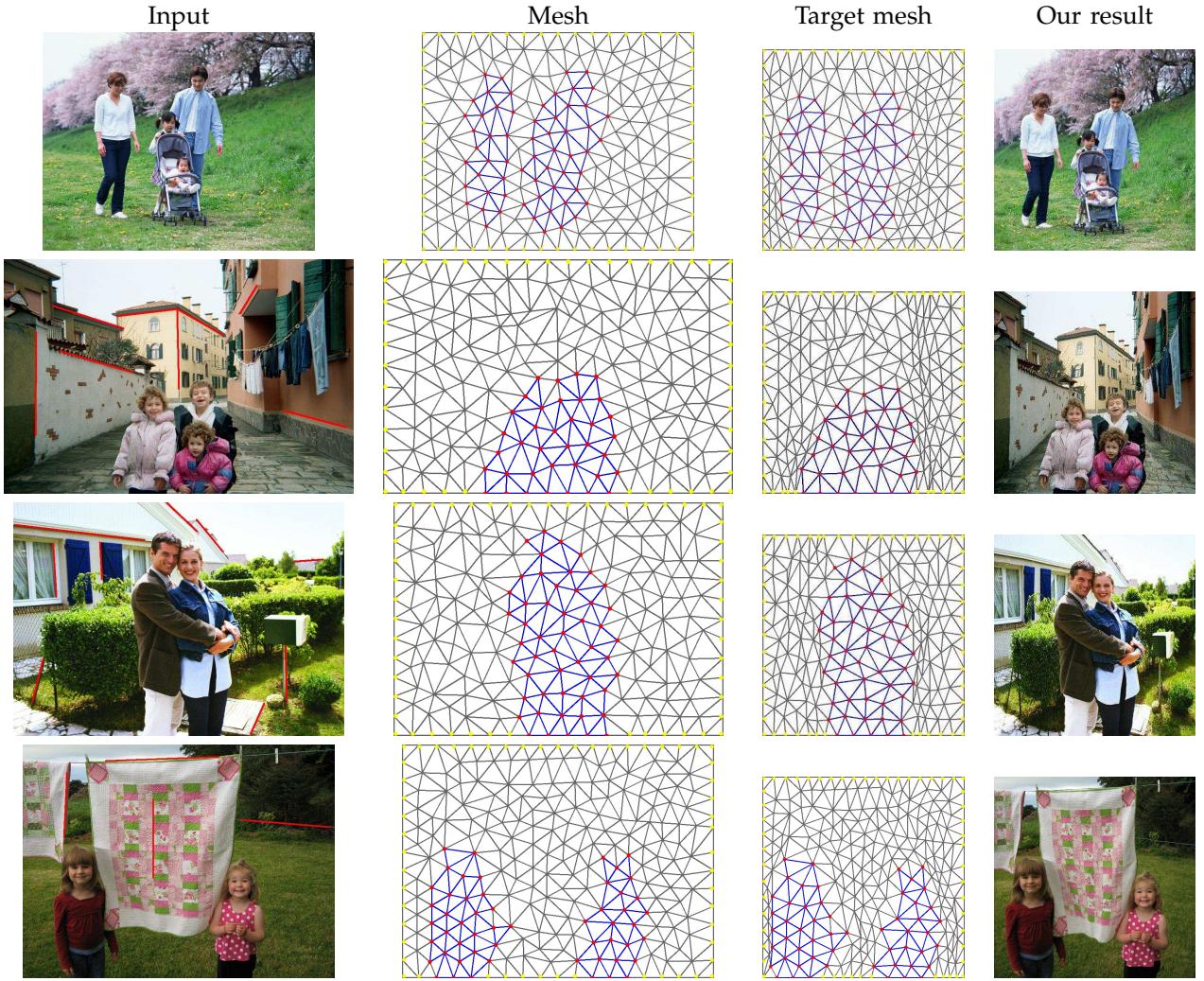


Fig. 8. Our retargeting results. The first column shows the input images with lines detected. The middle two columns are the meshes resulting from triangulation and the target meshes. In the last column, we show our retargeting results.

where coefficient w_{ij} penalizes the triangle in \mathcal{M}_t that flips over its orientation. If so, w_{ij} is chosen as -1, otherwise +1. This scheme guarantees a valid target mesh.

Overall, taking into account constraints in subsection III.3, the constrained mesh parametrization is finally defined as,

$$\operatorname{argmin}_{Q_i, i=1, \dots, n} (E_l + \lambda \cdot E_{ST}), \quad s.t. \quad F_B, F_{SA} = 0. \quad (15)$$

Boundary constraint and saliency constraint are hard constraints in the above equation, while structure constraint is viewed as soft constraint. λ is one weight factor balancing the influence of parametrization energy and structure constraint. Through extensive experiments, it is set to 0.5 for all our results in Section 4.

The energy of (15) is minimized using the multi-dimensional Newton's method. For each iteration of Newton's method, a multi-grid solver [19] is used to solve the sparse linear equations. In practice, the initial solutions of the equation are set to the positions of mesh points resulting from scaling the input image. For the

spare mesh density as shown in Fig. 1(b), the equation converges to the final solution very fast.

Points of \mathcal{M}_t in less salient region and retargeted poles of salient objects are generated through minimization of (15). Afterwards, points on retargeted salient objects are produced using equation (3). Once the target mesh \mathcal{M}_t is obtained, we render the resulting image using a standard texture mapping algorithm [6]. Such process in general can be efficiently implemented, since texture mapping is a basic function supported by modern graphics processing units (GPUs) and graphics hardware on cellular phones.

4 EXPERIMENTS

We experimented with our algorithm on a variety of images. Some representative results are shown in Figs. 8, 9, 10, 11, and 15.

Fig. 8 demonstrates some results together with the intermediate meshes. The second and third columns give the meshes generated by triangulation, and the target meshes produced by our algorithm respectively. For

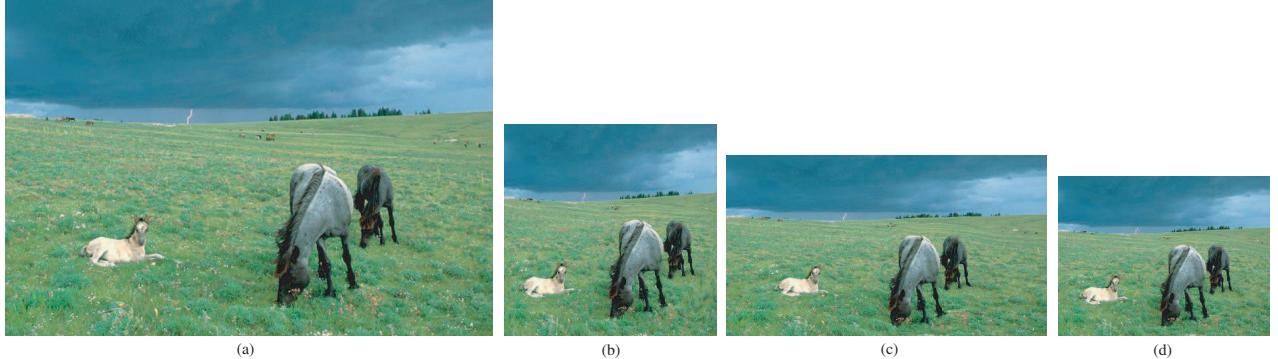


Fig. 9. Results for horses in grassland with different aspect ratios. (a) Original image with 743×512 resolution. (b), (c), and (d) Our results of sizes 320×320 , 480×270 , and 320×240 , and in aspect ratios of $1 : 1$, $16 : 9$, and $4 : 3$ separately.

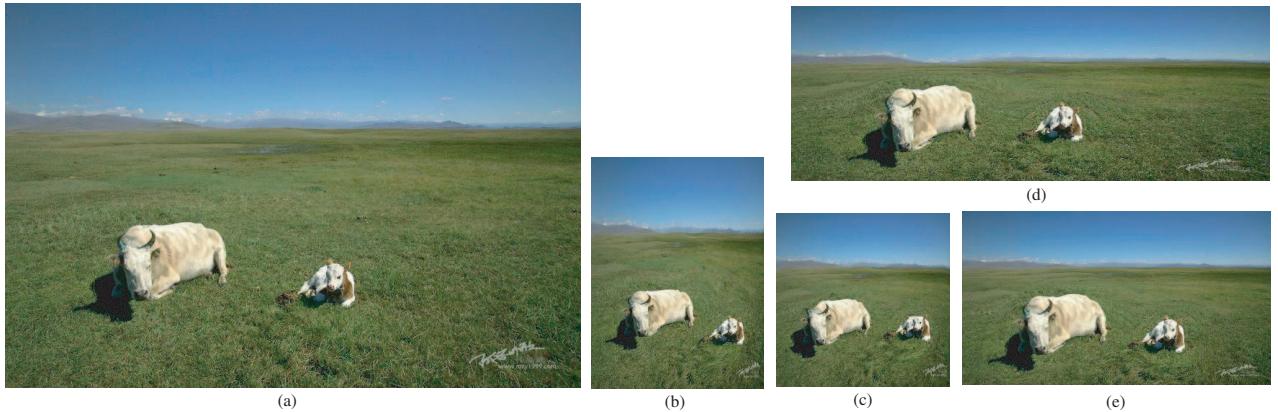


Fig. 10. Results for the cow and goat in grassland with different aspect ratios. (a) Original image with 800×534 resolution. (b), (c), (d), and (e) Our results of sizes 240×320 , 240×240 , and 664×240 and 428×240 . They are in aspect ratios of $3 : 4$, $1 : 1$, $25 : 9$, and $16 : 9$ respectively.



Fig. 11. Example of retargeting salient objects with different target scales. From left to right: input image of size 910×720 , and resulting images of size 320×320 and with the scales of 0.25 , 0.35 and 0.5 of salient people compared with the original.

images with background structures, lines detected and used as structure constraints are shown in the first column. Although such lines are sparse, they exert crucial effect in preserving background structures. For the input image in the fourth row, retargeting positions of the two girls are automatically computed by our algorithm.

Figs. 9 and 10 show several examples where there exist multiple important objects. Original images are adapted to different sizes and aspect ratios of popular mobile devices, e.g. 1:1, 3:4, 4:3, 16:9, even 25:9. As can be seen, our method effectively emphasizes all the important objects by non-uniformly shrinking contextual

regions. The distortion in these results is merely visible. Using our approach, target scales of salient objects can be controlled by the user. Fig. 11 gives such an example. The retargeting scales of salient objects vary according to user input scale parameters.

To demonstrate the effectiveness of our algorithm, we also compare to the results of many representative methods as shown in Fig. 15. The original images in the first, third and fifth rows are of sizes 1024×768 , 685×512 and 800×586 separately. All resulting images are of size 320×320 except the uniform scaling results. Obviously, we can see that uniform scaling usually makes the faces

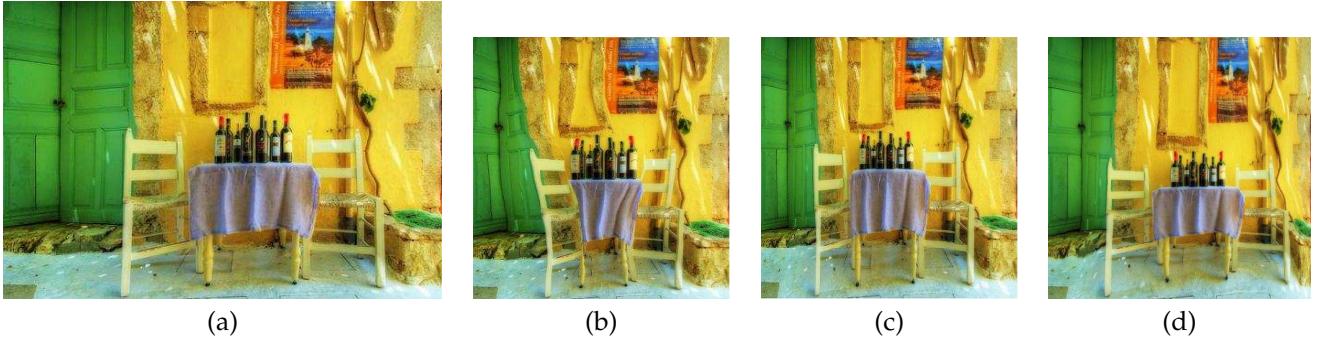


Fig. 12. Comparison between our approach and quad grids-based methods. (a) Input image. (b) Gal et al's result [13]. (c) Wang et al's result [14]. (d) Our result. Our result is comparable to their results. (a), (b), and (c) are fetched from the paper [14].

too small to identify. While automatic cropping can effectively emphasizes the important content [25], the context is lost. Results of fisheye warping (last column in odd rows) can achieve the goal of focus+context. However, fisheye warping may introduce noticeable distortions to background structures. Moreover, it only works well for image with single important object. The results in the first column of even rows are produced by segmentation-based composition [23]. For image with similar foreground and background color distributions, accurate segmentation poses a big challenge. Moreover, inpainting for background with complex structure is difficult. The seam carving method [2] may also destroy persons in the images (second column of even rows). We then integrate saliency and face detection into seam carving algorithm such that humans can be preserved completely (third column of even rows). But objects in background are destroyed now (see the artifacts highlighted by the red rectangles).

The last column of even rows in Fig. 15 gives our retargeting results. In our results, the persons are given high resolution and free from distortion. Since our method preserves image structures as much as possible, little visual distortion is introduced to the background of images. For example, the structure of the Golden Gate Bridge is well preserved.

We further compare our result with those produced by quad grids-based methods [13], [14]. From Fig. 12, it is obvious that our result is comparable to their results.

Computational complexity of our approach mainly depends on the size of the mesh, namely the number of mesh points and edges. The mesh size again depends on the source image size and mesh density. The major computation is spent on minimizing the energy in Equation (15). Normally, a source image is divided into a spare mesh, with several hundred mesh points. For example, the mesh of Fig. 1(b) has around 150 points. The number of variables of equation (15) for this example is around 120. Such equation can be solved efficiently using the multidimensional Newton's method. Given the target mesh, rendering the target image using the standard texture mapping method is trivial especially when such operation is accelerated by a common PC or

cellular phone graphics chip.

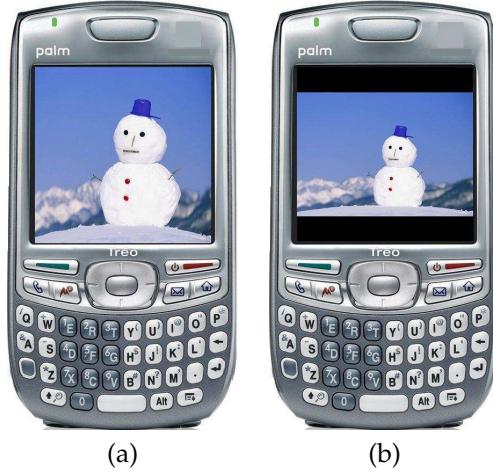


Fig. 13. Example of the image display for user study. Participants were shown two versions of each image on the image panels of a virtual Palm Treo. The input image was also given as reference.

4.1 User Study

To further evaluate the effectiveness of our approach, we carried out one user study. The objective is to determine whether the results of our method are preferred by users to those of other methods including uniform scaling, automatic cropping [25], fisheye warping [12], and seam carving [2].

TABLE 1
Statistical data of user study.

	Images	Participants	Mean (%) of wins	Std. dev.	p
Cropping	25	28	18.5 (73.86%)	2.05	0.0001
Scaling	25	23	19.7 (78.8%)	1.61	0.0001
Fisheye warp	25	22	16.27 (65.1%)	3.91	0.0041
Seam carving	25	24	15.3 (61.2%)	2.58	0.0126

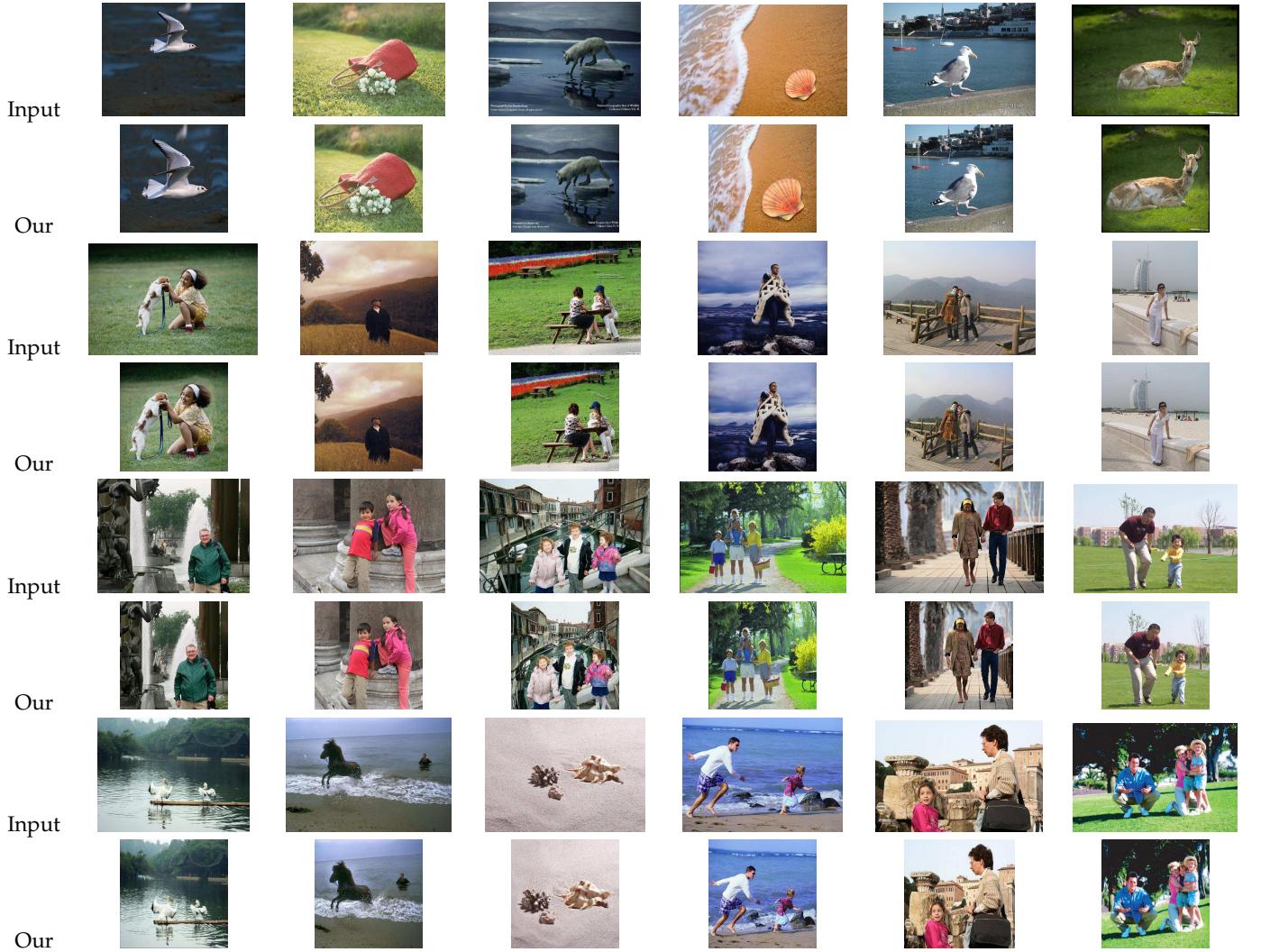


Fig. 14. Images (odd rows) and our retargeting results (even rows) used in user study. Another image is shown in Fig. 13.

25 representative images, together with the results of the tested algorithms, were used in our experiments. The images have wide diversity of types, e.g. images with simple and complex background, people and multiple salient objects. All the target images are of size 320×320 that fits a Palm screen resolution. The input high resolution images and our retargeting results are shown in Fig. 14. The user study consisted of four parallel parts, in which we compare the results of our method with those of scaling, cropping, fisheye warping, as well as seam carving separately. For fair comparison, seam carving is executed by incorporating saliency and face detection.

The study was conducted off-line and on-line simultaneously. For the off-line study, a participant was shown a pair of retargeted results of each input image, displayed on the image panels of a virtual Palm Treo as shown in Fig. 13. Each pair consisted of one result of our method and one of the other method. Whether our result was on the left or right was randomized. Specifically, the input image was also displayed for comparison. The participant was asked to select one from the pair they

preferred, and to write the answer on the answer sheet. On-line study was conducted similarly. The link of the study was opened to our computer science department and one consumer electronics company during the study. Subjects voluntarily responded to advertisements posted to the mailing lists and were not compensated for their time. Totally 97 different participants participated in our user study. They were in ages of 20 to 35, and were mainly undergraduate and graduate students of our department, as well as the IT employees. To our knowledge, they remained unknown about our project.

The statistical data of user study are shown in Table 1. For each part of the study, we count the total number of times that our result was preferred by participants. Overall, participants selected our results over cropped images 18.5 out of 25 times (73.86%) and over scaled images 19.7 (78.7%). p values of the t test given in the last column of the Table show that the comparisons were statistically significant. We further checked the testing images, and found that for images with a simple background, users normally chose cropping against ours.

While for the image with obvious structures in the background, they usually enjoyed ours as we preserved more image structures and avoided distortion as much as possible. This phenomenon indicates that visual features in surrounding context still have dominant effects on viewing experience in addition to the important objects.

Participants selected our results over results of fisheye-view warping 65.1% of times, over those of seam carving 61.2% of times. Actually, both fish-eye warping and seam carving achieve “focus+context” image retargeting with different schemes. Participants finally reported that for most of the results, they cannot distinguish our effect from the effects of fish-eye warping and seam carving at first glance. Our results visually resemble those of the two methods, especially when fitting images with simple background to a Palm screen of size 320×320 . However, for images with complex background structures, fish-eye warping tends to shrink objects near background too much. In addition, seam carving inevitably destroys some background structures. While our results can preserve the original background structures as-rigid-as possible, and meanwhile emphasize the important objects to some extent. As a result, users normally prefer our results when referring to the input images.

4.2 Limitations

Our method retargets images with salient objects by a process of saliency-based mesh parametrization. Salient object extraction relies on image saliency and face detection. Since 2D image is prone to variations of body poses, face expression, and illumination, the robust human face and body detection is still challenging. This makes accurate salient object extraction difficult. In our test examples, our method fails to capture some important human bodies with varied face poses or illumination. It is one drawback of our approach.

In addition, we emphasize salient object with an emphasis scale which can be determined automatically or set by the user. Emphasis of relative scale of salient object however will inevitably distort its nearby objects. We try to relieve such artifact with the structure constraint, which is imposed by detected structured segments distributed over background. Nevertheless, if background contains fewer detected segments, the structure constraint is insufficient. One example can be seen in the Universal Studios near the Globe example in Fig. 15. As the left part of the image contains fewer structured segments, structure constraint can not perfectly deal with the deformation of this area during retargeting. The people standing near the Globe in background are distorted as a result. This is another limitation of our current implementation. In fact, such image has many other structured curves, e.g. the contour of the Globe. In future, we intend to integrate preservation of more complex curved structures into our parametrization formulation, for better relieving obvious structure deformation. Furthermore, now we mainly evaluate the effectiveness

of our structure constraint for images with a variety amount of structures, ranging from simple background to complex background such as some examples shown in Fig. 8 and Fig. 14. We do not fully evaluate it for overly complex images. It will be helpful to further evaluate our approach on more images with more complex background.

5 CONCLUSIONS AND FUTURE WORK

We have proposed an effective image retargeting method based on a mesh image representation. For the first time, we formulate image retargeting as a mesh parametrization problem that aims to find a homomorphous mesh with the desired size of the target display. Our method achieves emphasizing important objects while retaining the surrounding context. Our scheme of using mesh representation for image adaptation enables easy preservation of images structures. Thus little visual distortion is introduced into the target image.

Representing images with mesh provides a feasible way for image retargeting with the solid foundation from mesh parametrization. Under mesh representation, salient objects have a compact structure using triangles, which makes emphasis of them more convenient and controllable. However, for computation efficiency, we usually triangulate the image with sparse mesh density. The object shape is not precise. To relieve this problem, we include more triangles into the object rather than missing some parts of the object. In this way, we guarantee least distortion to the object. Extra valuable screen space however is allocated to some non-important regions that are taken as parts of the object. This is a limitation of our current implementation. To account for this, we will explore the use of multi-resolution meshes for representation of images in future. That is, salient object is precisely approximated by dense triangles, while background is endowed with coarse ones for efficiency.

Our approach tries to adapt images with little distortion. Nevertheless, image layout in terms of semantic and topological relations between objects is not addressed. Such aspects are important to the visual perception. For instance the positional relations of players in football match, as well as their locations relative to the ball are important. In future, we shall take into consideration such relations, and investigate the solutions to the problem. Moreover, extending our current method to video is important.

ACKNOWLEDGMENT

We would like to thank the reviewers for their constructive comments which helped improving this manuscript. Tongwei Ren and Xiang Wang provided the matlab code of Seam carving. This work was supported by the National Natural Science Foundation of China (Grants 60703084, 60723003, 60635030, and 60721002), NSF grant IIS-0416284, Jiangsu Science Foundation (BK2008018) and Jiangsu 333 High-Level Talent Cultivation Program.

REFERENCES

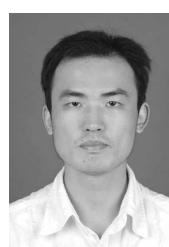
- [1] Y. Altunbasak and A. Tekalp. Closed-form connectivity-preserving solutions for motion compensation using 2-d meshes. *IEEE Trans. Image Processing*, 6(9):1255–1269, 1997.
- [2] S. Avidan and A. Shamir. Seam carving for content-aware image resizing. *ACM Trans. on Graphics*, 26(3):267–276, 2007.
- [3] L.-Q. Chen, X. Xie, X. Fan, W.-Y. Ma, H.-J. Zhang, and H.-Q. Zhou. A visual attention model for adapting images on small displays. *Multimedia Systems*, 9(4):353–364, 2003.
- [4] W.-H. Cheng, C.-W. Wang, and J.-L. Wu. Video adaptation for small display based on content recomposition. *IEEE Trans. Circuits and Systems for Video Technology*, 17(1):43–58, 2007.
- [5] X. Fan, X. Xie, H.-Q. Zhou, and W.-Y. Ma. Looking into video frames on small displays. In *ACM Multimedia*, pages 247–250, 2003.
- [6] D. Hearn and M. P. Baker. *Computer graphics with OpenGL*. Pearson Prentice Hall, 2004.
- [7] Y. Hu, D. Rajan, and L.-T. Chia. Robust subspace analysis for detecting visual attention regions in images. In *ACM Multimedia*, pages 716–724, 2005.
- [8] L. Itti and C. Koch. Computational modeling of visual attention. *Nature Reviews Neuroscience*, 2(3):194–203, March 2001.
- [9] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, 1998.
- [10] S. Jiang, H. Liu, Z. Zhao, Q. Huang, and W. Gao. Generating video sequence from photo image for mobile screens by content analysis. In *IEEE International Conference on Multimedia and Expo*, pages 1475–4178, 2007.
- [11] Y. Li, Y.-F. Ma, and H.-J. Zhang. Salient region detection and tracking in video. In *IEEE International Conference on Multimedia and Expo*, pages 269–272, 2003.
- [12] F. Liu and M. Gleicher. Automatic image retargeting with fisheye-view warping. In *ACM UIST*, pages 153–162, 2005.
- [13] R. Gal, O. Sorkine, and D. Cohen-Or. Feature-aware texturing. In *Proceedings of Eurographics Symposium on Rendering*, pages 297–303, 2006.
- [14] Y.-Sh. Wang, C.-L. Tai, O. Sorkine, and T.-Y. Lee. Optimized Scale-and-Stretch for Image Resizing. In *ACM Trans. Graph. (ACM Siggraph Asia)*, 27(5), 2008.
- [15] F. Liu and M. Gleicher. Video retargeting: Automating pan and scan. In *ACM Multimedia*, pages 241–250, 2006.
- [16] H. Liu, X. Xie, W.-Y. Ma, and H.-J. Zhang. Automatic browsing of large pictures on mobile devices. *ACM Multimedia*, pages 148–155, 2003.
- [17] Y.-F. Ma and H.-J. Zhang. Contrast-based image attention analysis by using fuzzy growing. In *ACM Multimedia*, pages 374–381, 2003.
- [18] J. Maillet, H. Yahia, and A. Verroust. Interactive texture mapping. In *ACM Siggraph*, pages 27–34, 1993.
- [19] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery. *Numerical recipes in C: The art of scientific computing*. Cambridge University Press, 1992.
- [20] Y. Rui, L. He, A. Gupta, and Q. Liu. Building an intelligent camera management system. In *ACM Multimedia*, pages 2–11, 2001.
- [21] A. Santella, M. Agrawala, D. DeCarlo, D. Salesin, and M. Cohen. Gaze-based interaction for semi-automatic photo cropping. In *Proc. of CHI*, 2006.
- [22] R. Seidel. Constrained delaunay triangulations and voronoi diagrams with obstacles. *Technical Report 260, Inst. for Information Processing, Graz, Austria*, 1988.
- [23] V. Setlur, T. Lechner, M. Nienhaus, and B. Gooch. Retargeting images and video for preserving information saliency. *IEEE Computer Graphics and Applications*, 27(5):80–88, 2007.
- [24] V. Setlur, S. Takagi, R. Raskar, M. Gleicher, and B. Gooch. Automatic Image Retargeting. *Proceedings of 4th International Conference on Mobile and Ubiquitous Multimedia*, 2005.
- [25] B. Suh, H. Ling, B. Bederson, and D. Jacobs. Automatic thumbnail cropping and its effectiveness. In *ACM UIST*, pages 95–104, 2003.
- [26] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *IEEE Conf. CVPR*, pages 511–518, 2001.
- [27] J. Wang, M. Reinders, R. Lagendijk, J. Lindenbergh, and M. Kankanhalli. Video content representation on tiny devices. In *IEEE International Conference on Multimedia and Expo*, pages 1711–1714, 2004.
- [28] J. Wang, Y. Ying, Y. Guo, and Q.-S. Peng. Automatic foreground extraction of head shoulder images. In *Computer Graphics International (Springer LNCS 4035)*, pages 385–396, 2006.
- [29] L. Wolf, M. Guttmann, and D. Cohen-Or. Non-homogeneous content-driven video-retargeting. In *IEEE International Conference on Computer Vision*, pages 1–6, 2007.
- [30] K. Zhou, J. Snyder, B. Guo, and H.-Y. Shum. Iso-charts: stretch-driven mesh parameterization using spectral analysis. In *ACM Symp. Geometry Processing*, pages 45–54, 2004.



Yanwen Guo received his Ph.D degree in applied mathematics from State Key Lab of CAD&CG, Zhejiang University in 2006. He is currently an associate professor at the National Laboratory of Novel Software Technology, Department of Computer Science and Technology, Nanjing University. His main research interests include image and video processing, geometry processing, and face related applications. He worked as a visiting researcher in Department of Computer Science and Engineering, the Chinese University of Hong Kong in 2006 and a visiting researcher in Department of Computer Science, the University of Hong Kong in 2008.



Feng Liu received the B.S. and M.S. degrees from Zhejiang University, Hangzhou, China, in 2001 and 2004, respectively, both in computer science. He is currently a Ph. D. candidate in the Department of Computer Sciences at the University of Wisconsin-Madison, USA. His research interests are in the areas of multimedia, computer vision and graphics.



Jian Shi received the BSc degree in computer science from Nanjing University, China, in 2007. He is now a master student in the Department of Computer Science & Technology at Nanjing University. His research interest is in image and video processing.



Zhi-Hua Zhou (S'00-M'01-SM'06) received the BSc, MSc and PhD degrees in computer science from Nanjing University, China, in 1996, 1998 and 2000, respectively, all with the highest honors.

He joined the Department of Computer Science & Technology at Nanjing University as an assistant professor in 2001, and is currently Cheung Kong Professor and Director of the LAMDA group. His research interests are in artificial intelligence, machine learning, data mining, pattern recognition, information retrieval, evolutionary computation, and neural computation. In these areas he has published over 70 papers in leading international journals or conference proceedings.

Dr. Zhou has won various awards/honors including the National Science & Technology Award for Young Scholars of China (2006), the Award of National Science Fund for Distinguished Young Scholars of China (2003), the National Excellent Doctoral Dissertation Award of China (2003), the Microsoft Young Professorship Award (2006), etc. He is an Associate Editor of *IEEE Transactions on Knowledge and Data Engineering*, Associate Editor-in-Chief of *Chinese Science Bulletin*, and on the editorial boards of *Artificial Intelligence in Medicine*, *Intelligent Data Analysis*, *Science in China*, etc. He is/was a PAKDD Steering Committee member, Program Committee Chair/Co-Chair of PAKDD'07, PRICAI'08 and ACML'09, Vice Chair or Area Chair of IEEE ICDM'06, IEEE ICDM'08, SIAM DM'09, ACM CIKM'09, etc., Program Committee member of various international conferences including AAAI, ICML, ECML, ACM SIGKDD, IEEE ICDM, ACM Multimedia, etc., and General Chair/Co-Chair or Program Committee Chair/Co-Chair of a dozen of native conferences. He is a senior member of China Computer Federation (CCF), the Vice Chair of the CCF Artificial Intelligence & Pattern Recognition Society, an Executive Committee member of Chinese Association of Artificial Intelligence (CAAI), the Chair of the CAAI Machine Learning Society, and the Chair of the IEEE Computer Society Nanjing Chapter. He is a member of AAAI and ACM, and a senior member of IEEE and IEEE Computer Society.



Michael Gleicher is an Associate Professor in the Department of Computer Sciences at the University of Wisconsin, Madison. Prof. Gleicher is founder and leader of the Department's Computer Graphics group. Prof. Gleicher's current research falls into three categories: character animation; automated multimedia processing and production; and visualization and analysis tools for life sciences applications. Prior to joining the university, Prof. Gleicher was a researcher at The Autodesk Vision Technology Center and at

Apple Computer's Advanced Technology Group. He earned his Ph. D. in Computer Science from Carnegie Mellon University, and holds a B.S.E. in Electrical Engineering from Duke University.

Fig. 15. Comparisons between our approach and previous methods. All results are produced to fit a 320×320 PDA screen.