Objective Quality Assessment of Image Retargeting by Incorporating Fidelity Measures and Inconsistency Detection

Yichi Zhang, *Student Member, IEEE*, King Ngi Ngan, *Fellow, IEEE*, Lin Ma, *Member, IEEE*, and Hongliang Li, *Senior Member, IEEE*

Abstract—The tremendous growth in mobile devices has resulted in huge generation and usage of digital images. Image quality assessment is thus an important issue for mobile media applications. In this paper, we focus on the quality evaluation of images generated by content-aware image retargeting, in which the reference and the distorted images are of different sizes. Through retargeting, many types of deformation inconsistency lead to shape distortion, deformation artifacts, and content information loss, worsening its perceptual quality. The deformation inconsistency occurs on different levels of the retargeted images. Limited by the accuracy of the alignment between the original and retargeted images, previous methods only focus on pixel-level and patch-level fidelity analyses and fail to detect deformation inconsistency. In this paper, we improve the alignment algorithm and propose a three-level representation of the retargeting process. Based on the analysis of this three-level representation, both fidelity measures and inconsistency detection are combined to determine the final retargeting quality. The proposed algorithm is validated on the public data sets RetargetMe and CUHK. Experimental results demonstrate that inconsistency detection contributes to accurately assessing the image retargeting perceptual quality. This inspires us to investigate more about deformation inconsistency to formulate the objective quality of image retargeting.

Index Terms—Image retargeting, quality assessment, content information loss, shape distortion, deformation artifacts, fidelity measures.

I. INTRODUCTION

IN the last decade, we have witnessed the fast development of mobile devices, which has imposed new demands on convenient image display. Despite the fixed sizes and aspect ratios of media contents, image and video retargeting techniques adapt the original content to different resolutions. Traditional methods such as linear scaling and cropping suffer from shape distortion or content loss. In contrast, content-aware image retargeting methods adjust source images into arbitrary sizes under two objectives: preserve salient contents and avoid annoying artifacts. To assess the performance of retargeting algorithms, many recent works focus on the quality assessment of image retargeting [1]–[8].

Image retargeting algorithms [9], [10] can be roughly categorized into two classes: discrete and continuous. Discrete methods [11]–[14] directly remove (or insert) less important pixels or patches to generate retargeted images. Seam carving [11] iteratively removes eight-connected paths of pixels known as seams. The direct removal of image content may cause deformation artifacts such as jittering or discontinuities and degrade perceptual quality. In contrast, continuous methods [15]–[18] determine an optimal pixel-wise mapping from the original size to the target size. The pixel-wise mapping is determined by either solving a linear system [15] or minimizing energy functions [16]–[18]. The shape of salient content may be distorted due to inconsistent deformation, worsening retargeting quality. Several studies [19], [20] combined a few retargeting operators to alleviate these weaknesses. In spite of their novelties, the above-mentioned methods validate their performance only by small-scale subjective assessment. Hence, a systematic objective image retargeting quality evaluation method is necessary.

Image retargeting brings about deformations, which are crucial in assessing the retargeting quality. Their existence reduces the fidelity of retargeted images and induces various types of inconsistency as shown in Fig. 1. Therefore, previous methods on image retargeting quality assessment align the reference and retargeted images, and define the overall similarity between aligned contents as the retargeting quality [2]–[8]. The formulation of the quality measures is generally based on the analyses of three levels:

- **Region-Level Analysis:** The original and retargeted images are segmented into regions. Region-level analysis examines the deformation of regions' shapes. Region-level inconsistency occurs if some regions are over-squeezed or over-stretched. In Fig. 1(a), the squashing of the girl's face leads to region-level inconsistency.
- **Patch-Level Analysis:** The segments in the region-level analysis are further divided into patches. Patch-level
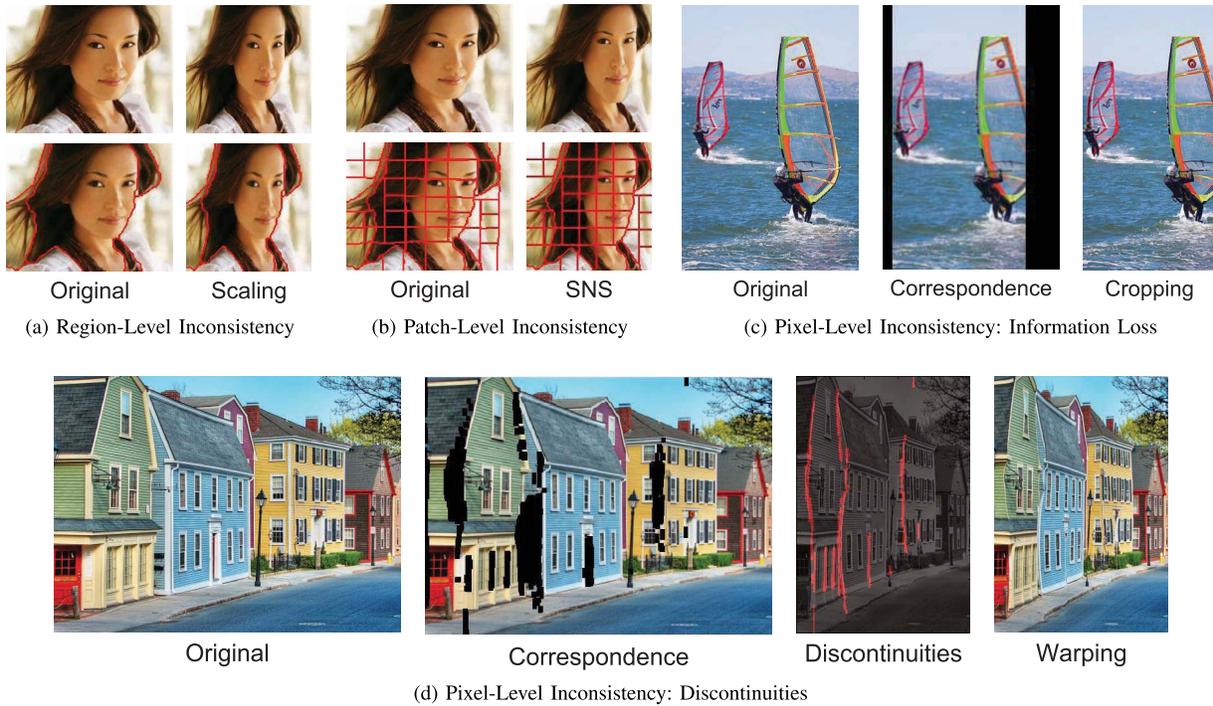
Fig. 1. Retargeted images with deformation inconsistency. (a) Scaling deforms the shape of the face, leading to region-level inconsistency. (b) SNS introduces inconsistent warping, distorting the girl's face, resulting in patch-level inconsistency. (c) The cropped image snips the surfboard, damaging the integrity of salient objects. (d) Non-homogeneous warping removes some content in the original image, resulting in the pixel discontinuities labelled in red.

analysis explores the deformation of patches. Patch-level inconsistency happens if some patches go through deformations differently from the others, leading to a distortion mirror effect. In Fig. 1(b), the retargeted image obtained by scale-and-stretch (SNS) [16] retains the height of the forehead but lengthens the heights of the nose, mouth and jaw, leading to distortion of proportions.

- **Pixel-Level Analysis:** Pixel-level analysis investigates the pixels in the original image that are employed to generate the retargeted image. Pixel-level inconsistency indicates that not all the pixels are preserved through retargeting, causing two possible instances of disagreement: discontinuities and content information loss. In Fig. 1(c), cropping removes part of foreground objects while in Fig. 1(d) the retargeted image obtained by non-homogeneous warping [15] suffers from severe discontinuities. Both images are disliked by human viewers.

However, due to the limited accuracy of image alignment algorithms, existent quality metrics for retargeted images do not fully exploit the perceptual quality degradations on all levels: [2] and [6] rely on pixel-level analysis; [3]–[5] and [7] incorporate both pixel- and patch-level analyses; [8] combines pixel- and region-level analyses. Moreover, these quality metrics only consider an overall similarity or fidelity of the retargeted images but ignore the deformation inconsistency. Hence, we propose a full reference objective quality evaluation method for image retargeting by analyzing the fidelity and detecting the inconsistency of the retargeted image on three different levels. The original and retargeted images are first aligned by flow estimation, which enables us to analyze

the retargeting process on three different levels: region-level analysis that models the shape distortion of segments, patch-level analysis that investigates the shape distortion of patches, as well as pixel-level analysis that studies the deformation artifacts and content information loss. Apart from the fidelity measurements, we also detect three types of deformation inconsistency, which achieves promising performance on benchmark datasets.

Compared with the preliminary conference paper [8], the precision of the flow estimation has been improved by modifying the energy function. In [8], we only attempted to capture the discontinuities of retargeted images, which can be regarded as the pixel-level inconsistency. In this paper we include patch-level and region-level inconsistency detections, which further improve the prediction performance. In the fidelity analysis, the conference paper only combined pixel-level information loss and the region-level aspect ratio change. In this paper we also incorporate the patch-level fidelity analysis. The current work outperforms [8] on both RetargetMe and CUHK datasets.

The rest of the paper is organized as follows. Section II introduces some related works on image retargeting quality assessment. In Section III and IV, we describe the details of the proposed fidelity measures and the detection of inconsistency respectively. Section V presents the experimental results and the final section concludes the paper.

II. RELATED WORKS

Many pioneer works have devoted to large-scale user study on benchmark datasets of retargeted images. Rubinstein *et al.* set up the dataset RetargetMe [1] while Ma *et al.* built the
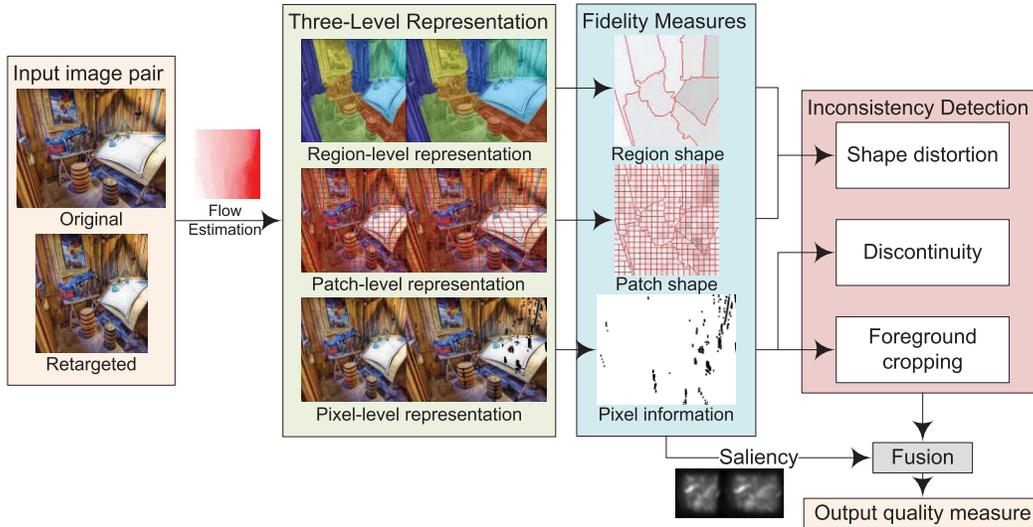
Fig. 2. The image retargeting quality assessment framework. Flow estimation matches original and retageted pixels and constructs the three-level representation of the retargeting process. For each level, we propose a fidelity measure. On the raw measure maps, darker areas indicate lower fidelity. Based on the raw measures, three cases of inconsistent distortions are detected. The fusion of the fidelity measures and the inconsistency detection determines the final quality score.

dataset CUHK [21]. Both studies claimed that humans agree on the perceptual quality of retargeted images. They also mentioned that content loss, deformation artifacts and shape distortion degrade the retargeting quality. Besides, they concluded that the image distance metrics [20], [22]–[26] could hardly predict image retargeting quality. Castillo *et al.* [27] collected the eye-tracking data on a benchmark of retargeted images and found that eye-tracking information improves the quality predicting performance of computational distance metrics. Conducting large-scale user study and using an eye tracker are not always feasible. Hence, recent studies on image retargeting quality assessment concentrate on objective methods inspired by image quality assessment.

Traditional image quality assessment metrics, such as mean square error (MSE), peak-signal-to-noise ratio (PSNR) [28], structural similarity (SSIM) [29], [30] and some recent quality metrics [31], [32] cannot be applied to image retargeting quality assessment because of the difference of sizes between the original and retargeted images. To investigate the retargeting deformation, scale invariant feature transform (SIFT) [33] and SIFT flow [25] have been adopted to align original and retargeted images [2]–[8]. Given the alignment, the retargeting process can be analyzed on different levels. In [2] and [6], the average SSIM indices for matched SIFT keypoints are defined as the retargeting quality. These two methods only examine pixel-level deformation artifacts and content information loss. Hsu *et al.* [7] defined the local variance of SIFT flow vector fields as the perceptual geometric distortion. Liang *et al.* [5] incorporated aesthetic components in the quality measure. In [4], the author applied Delaunay triangulation on SIFT feature points to separate original images and retargeted images into triangles. The shape deformation of triangles is also considered in the final quality measure. Similarly, Zhang *et al.* [3] divides original and retargeted images into patches and define the average aspect ratio similarity as the

retargeting quality. These four methods add patch-level analysis to their objective quality evaluation methods. Liu *et al.* [34] fused four quality factors of both shape deformation and visual content change by using support vector regression (SVR) [35]–[37] for retargeting quality measure. Ma *et al.* [38] employ pairwise rank learning for no-reference retargeted image quality assessment. In [8], we combined region-level analysis of foreground objects, pixel-level deformation artifacts and content information loss to formulate the retargeting quality. By combining analyses on different levels, these methods achieve better performance on predicting the retargeting quality. Still, the above-mentioned methods have limited performance due to the misalignment of SIFT-based alignment algorithms and ignorance of deformation inconsistency in the formulation of the quality measure.

III. IMAGE RETARGETING FIDELITY MEASURES

The overall framework of the proposed method is shown in Fig. 2. Given a pair of original and retargeted images, we estimate the flow vector field by colors. The flow field is visualized by the method in [39] in Fig. 2. The estimated flow vector field bridges the difference of sizes between the original and retargeted images, and the three-level representation of the retargeting process is established based on it: region-level segmentation, patch-level partition, and pixel-level correspondence. Three fidelity measures, region shape fidelity, patch shape fidelity as well as pixel information fidelity, are defined to model the overall similarity between the retargeted image and the original one. From the raw measure maps in Fig. 2, we observe that the area of severe distortion is overwhelmed by well-preserved regions. A weighted sum of all the raw measures easily ignores such severe distortions. Therefore, deformation inconsistency at different levels is also examined, including shape distortion, foreground cropping, and discontinuity. By combining fidelity measures and inconsistency

Fig. 3. Misalignment of SIFT keypoints under radical change of target aspect ratio. The retargeted image on the left is generated by SNS. SIFT keypoints are represented by circles. The radius of the circle and the direction of the shown radius represent the scale and the orientation of the corresponding SIFT keypoint respectively.

detection, we obtain the final quality measure for the retargeted image. In this section, we concentrate on the image alignment and the definition of three fidelity measures. Throughout this paper, we adopt the following rules in the denotations. A symbol with a superscript or subscript o indicates its relation with the original image. In contrast, superscript or subscript r implies the connection with the retargeted image. Those symbols without such denotations are related to both images.

A. Alignment Between Original and Retargeted Images

As original and retargeted images are of different sizes, a correspondence map is essential to align pixel pairs from the original and retargeted images. Recent works on image retargeting quality assessment [2]–[4], [6]–[8] adopt SIFT feature matching or SIFT flow to build up the correspondence. However, radical change of target aspect ratio affects the SIFT descriptors. Besides, some retargeting algorithms generate unforeseen textures whose SIFT descriptors differ greatly from the original ones. These facts defy the assumption that SIFT descriptors remain constant through image retargeting. In Fig. 3, we present an example that SIFT descriptors fail to align feature points correctly. The retargeted image and the original one are presented alongside, and only 40 pairs of matched SIFT keypoints are exhibited to facilitate visualization. The keypoint, labeled in red, at the third (from left to right) pier of Ponte Sant'Angelo in the retargeted image is wrongly aligned to the second pier in the original image. This suggests that previous methods based on SIFT feature matching are sometimes comparing unrelated pixels or patches. To improve the accuracy of alignment, we investigated the characteristics of different retargeting algorithms and discovered that color is the best descriptor to align retargeted images. Despite the change of aspect ratio after retargeting, the correctly aligned pixels still present similar color. Inspired by SIFT flow, we propose a similar feature matching method, color-flow, whose energy function is defined as Eq. 1.

$$E(\mathbf{w}) = \sum_{\mathbf{p}_r} \min(||I_r(\mathbf{p}_r) - I_o(\mathbf{p}_r + \mathbf{w}(\mathbf{p}_r))||_1, t)$$
$$+ \sum_{(\mathbf{p}_r, \mathbf{q}_r) \in \epsilon} \min(\alpha |u(\mathbf{p}_r) - u(\mathbf{q}_r)|, d)$$
$$+ \min(\alpha |v(\mathbf{p}_r) - v(\mathbf{q}_r)|, d) \quad (1)$$

Here I_o and I_r denote the original image and the retargeted image. The variable to optimize is the flow field $\mathbf{w}$, and for each retargeted pixel $\mathbf{p}_r$, $u(\mathbf{p}_r)$ and $v(\mathbf{p}_r)$ are the two components of the flow vector $\mathbf{w}(\mathbf{p}_r)$. Set ϵ consists of all the spatial neighborhoods in the retargeted image. The first term acts as the data term, which forces the color similarity for matched pixels. The parameter t limits the amount of maximum error, and avoids the impact of outlier pixel pairs whose colors differ significantly. The second term is a smoothness term which encourages the continuity of the adjoined pixels' flow vectors. The parameter α controls the continuity of the flow field. Using a larger α will achieve a more consistent flow field. Similar to the parameter t, d avoids the impact of outlier pixel pairs whose flow vector components differ significantly. We remove the small displacement term in the energy function of SIFT flow because a retargeted image is composed of patches across the whole original image and some flow vectors may have considerable magnitudes. This energy function is optimized by belief propagation and the parameter settings are discussed in the experiment results.

The estimated flow bridges the different resolutions of the original and retargeted images. The fidelity and inconsistency of the retargeted images can be thus studied on different levels. Before calculating the retargeting fidelity measures and detecting inconsistent distortion, we obtain the saliency map of the original image, denoted as $\mathbf{S}_o$, to represent the importance of each pixel in the image. The determining factor in the choice of the saliency algorithm is whether the algorithm can light up the whole salient region evenly because we want to quantify the amount of distortion at different levels in the inconsistency detection. Most early saliency algorithms [40]–[42] assign high importance values only on the boundary pixels of the salient content while severe inconsistent distortion often occurs in the smooth inner regions. The saliency algorithm applied in this work is the graph-based visual saliency (GBVS) [43] because of its accuracy in predicting human gaze fixation. Besides, GBVS assign consistent importance values in the salient content, without paying too much attention to the boundary pixels. Recent saliency algorithms, such as the region contrast method introduced in [44], assign identical importance values to all the pixels in the same region. Such algorithms may also be employed in our framework. The saliency map of the retargeted image $\mathbf{S}_r$ is derived from the original saliency map $\mathbf{S}_o$ and the estimated flow field $\mathbf{w}$ by assigning $\mathbf{S}_r(\mathbf{p}_r) = \mathbf{S}_o(\mathbf{p}_r + \mathbf{w}(\mathbf{p}_r))$. The saliency map pair $\mathbf{S}_o$ and $\mathbf{S}_r$ acts as the weights of different regions, patches and pixels in the calculation of the retargeting fidelity measures.

B. Region-Level Measure: Region Shape Fidelity

The segmentation of the original and retargeted images contributes to analyzing region-level deformations of the retargeting process. A perfect segmentation result should delineate well the object boundaries and superpixel-based segmentation algorithms suit such requirement. The algorithm [45] tends to over-segment the original image and have irregular segment boundaries. In our conference paper [8], we adopted the method [46], which segments an image by merging

Fig. 4. Image segmentation results. Segments boundaries are labelled in red. The red regions in the images on the right are the corresponding salient regions by using a threshold of 60%.

superpixels. This method has one parameter that controls the merging of pixels and using the same parameter on different images may obtain different number of segments. Therefore in this work, we apply segmentation to original images by aggregating superpixel (SAS) [47], whose only parameter is the number of segments. We denote the set of pixels in the ith segment of the original image as $\mathbb{R}_i^o$. Its counterpart in the retargeted image can be obtained by assigning $\mathbb{R}_i^r = \{\mathbf{p}_r | \mathbf{p}_r + \mathbf{w}(\mathbf{p}_r) \in \mathbb{R}_i^o\}$. Morphological filtering is applied to the warped segmentation to alleviate the impact of possible mismatches. The segmentation of the original and retargeted images establishes the region-level representation of the retargeting process.

A segment consisted of salient pixels is considered important. We sort all the segments in the descending order of the average saliency of their pixel members, $\sum_{\mathbf{p}_o \in \mathbb{R}_i^o} \mathbf{S}_o(\mathbf{p}_o)/|\mathbb{R}_i^o|$. Here $|\cdot|$ stands for the cardinality of the set. Segments which rank in the top and make up 60% of total saliency values are considered important segments to preserve. The saliency threshold of 60% is tested on the RetargetMe and is sufficient to capture all the salient content in an image. Some segmentation results and the corresponding salient regions of images on the dataset RetargetMe are shown in Fig. 4. The set of important segments is designated as $\mathbb{S}_{imp}$. Only the important segments are considered in the region-level analysis. In addition, the pixels that belong to any important segments compose the important pixel set $\mathbb{P}_{sal}$. Finally, the saliency of an important segment i is defined by Eq. 2.

$$s_{seg}^i = \frac{\sum_{\mathbf{p}_o \in \mathbb{R}_i^o} \mathbf{S}_o(\mathbf{p}_o)}{\sum_{\mathbf{p}_o \in \mathbb{R}_j^o, j \in \mathbb{S}_{imp}} \mathbf{S}_o(\mathbf{p}_o)} \quad (2)$$

All the saliency values for important segments can form a vector $\mathbf{s}_{seg}$ and they sum up to 1. By defining such a saliency measure for segments, we regard that region-level analysis of retargeting quality is more relevant with large salient segments.

1) Region Shape Fidelity: A well-preserved retargeted region should share similar shapes with the original one. In the conference paper [8], we proposed the region shape fidelity the aspect ratio change of the segments' bounding boxes. This measure can detect some severe compressed regions, however, in many cases only part of the salient regions has been squeezed and the bounding box does not change significantly. Therefore in this paper we focus on the deformation of the segment boundaries. To investigate shape distortion, we rely on some shape coding techniques. Freeman chain code [48] could represent the contour of a patch or region. Therefore, the histogram of chain code (CCH) [49] is widely adopted in shape coding and recognization of irregular objects such as characters. Inspired by this, we compare the chain code histograms of original and retargeted segments to examine the shape fidelity of regions. CCH can help to detect both the consistent compression and the local distortion of the segments. Throughout this work, we consider the chain codes consisting of 8 directions.

For an important segment $i \in \mathbb{S}_{imp}$ which has not been removed in the retargeted image, we first calculate the convex hulls of the original and retargeted segments. Then, the contours of the convex hulls are extracted, denoted as $\mathbf{c}_i^o$ and $\mathbf{c}_i^r$ respectively. For each pixel on the contour, its direction is coded with integer values $k = 0, 1, \ldots, 7$ in a counterclockwise sense starting from the direction of the positive x-axis. The chain code histogram is a simple discrete function as given in Eq. 3,

$$h(k) = \frac{n_k}{n}, \quad k = 0, 1, \ldots, 7 \quad (3)$$

where n_k is the number of contour pixels whose chain code equals k, and n is the total number of links in the chain code. After obtaining the two chain code historgrams $\mathbf{h}_i^o$ and $\mathbf{h}_i^r$, we calculate the histogram intersection as the region shape similarity $m_{region}^i = \sum_k \min(\mathbf{h}_i^r(k), \mathbf{h}_i^o(k))$. The region shape fidelity measure m_{region} is defined as the weighted average of histogram intersections, defined by Eq. 4.

$$m_{region} = \frac{\sum_{i \in \mathbb{S}_{imp}, |\mathbb{R}_i^r| > 0} s_{seg}^i m_{region}^i}{\sum_{i \in \mathbb{S}_{imp}, |\mathbb{R}_i^r| > 0} s_{seg}^i} \quad (4)$$

A higher value of m_{region} indicates better preservation of the regions' shape. This measurement is sensitive to the squeeze or stretch of important objects.

C. Patch-Level Measure: Patch Shape Fidelity

Given the retargeted segmentation, patch-level partition is established to scrutinize local deformations. For each of the segment in the retargeted image, we partition it by 20×20 squares. The intersection of the segment and squares are defined as local patches. Those with a smaller area are merged with adjacent larger ones to guarantee that all patches share similar area. Repeating this procedure on all segments,

we can derive the patch-level representation of the retargeting process. The ith retargeted patch is denoted as $\mathbb{P}_i^r$. By warping, we can obtain its counterpart in the original image: $\mathbb{P}_i^o = \{\mathbf{p}_o|\mathbf{p}_o = \mathbf{p}_r + \mathbf{w}(\mathbf{p}_r), \mathbf{p}_r \in \mathbb{R}_i^r\}$. The importance of each patch is assessed by the retargeted saliency map, as shown in Eq. 5.

$$s_{patch}^i = \frac{\sum_{\mathbf{p}_r \in \mathbb{R}_i^r} \mathbf{S}_r(\mathbf{p}_r)}{\sum_{\mathbf{p}_r} \mathbf{S}_r(\mathbf{p}_r)} \qquad (5)$$

The importance measures of all patches can form a vector $\mathbf{s}_{patch}$, which sum up to 1.

1) Patch Shape Fidelity: Apart from region-level shape distortion, patch-level shape distortion also determines the retargeting quality. A well-preserved retargeted image should keep the aspect ratios of the patches in the important regions and avoid distortion of proportions. Similar to region shape fidelity, the chain code histograms of original and retargeted patches are compared to examine patch shape fidelity. The only difference is that both important and less important patches are considered in the definition of this measurement. For a patch i, the convex hulls of the original and retargeted patches are computed. By extracting the contours of the convex hulls and counting the occurrence of directions for contour pixels, we obtain two chain code histograms $\mathbf{h}_i^o$ and $\mathbf{h}_i^r$. The patch shape similarity is defined as $m_{patch}^i = \sum_k \min(\mathbf{h}_i^r(k), \mathbf{h}_i^o(k))$. The patch shape fidelity is a weighted sum of patch shape similarity, as defined by Eq. 6.

$$m_{patch} = \sum_i s_{patch}^i m_{patch}^i \qquad (6)$$

The patch shape fidelity describes the preservation of local shapes, which is sensitive to large discontinuities as well as the squeeze or stretch of patches.

D. Pixel-Level Measure: Pixel Information Fidelity

In Fig. 5, we exhibit how the pixel-level representation is established. The estimated flow in the top row is visualized by [39]. At pixel level, each retargeted pixel $\mathbf{p}_r$ has its counterpart $\mathbf{p}_o = \mathbf{p}_r + \mathbf{w}(\mathbf{p}_r)$ in the original image. Repeating this procedure for all the retargeted pixels, we can obtain an original mask whose resolution equals that of the original image as shown in Fig. 5. This mask denotes all the original pixels which are employed to generate the retargeted image. After applying morphological closing to fill the narrow seams or holes in the mask, we define the filtered mask as the correspondence map of the retargeted image, denoted as M_{corr}. This map evaluates the pixel information fidelity and detects deformation inconsistency such as foreground cropping and discontinuity. Here we denote several sets of pixels which is useful in the detection of pixel-level inconsistency. The set of boundary pixels in the retargeted image is denoted as $\mathbb{P}_{bound}^r$. According to the correspondence map, we locate the corresponding set of pixels in the original image $\mathbb{P}_{bound}^o$. A correspondence map may contain many holes because of content removal, as shown in the right image in the second row in Fig. 5. For each connected hole, we apply morphological dilating to obtain the set of original pixels on the boundary of the hole, denoted as $\mathbb{P}_{hole}^o$. According to the estimated
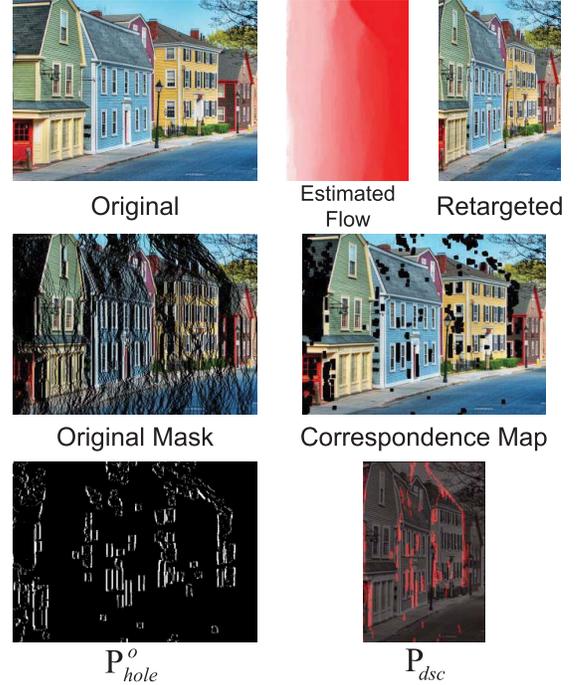
Fig. 5. Pixel-level image representation. The top row displays the original and the retargeted images as well as the estimated flow field. The second row presents the original mask and the correspondence map M_{corr}. The last row shows the boundary pixels of the holes in the correspondence map $\mathbb{P}_{hole}^o$ and the pixels of discontinuity candidates $\mathbb{P}_{dsc}$.

flow field $\mathbf{w}$, its corresponding pixel set in the retargeted image, $\mathbb{P}_{hole}^r$, can be located, and the set of discontinuity candidates can be derived as $\mathbb{P}_{dsc} = \{(\mathbf{p}_o, \mathbf{p}_r)|\mathbf{p}_r \in \mathbb{P}_i^r, \mathbf{p}_o = \mathbf{p}_r + \mathbf{w}(\mathbf{p}_r), \forall i, \exists \mathbf{p} \in \mathbb{P}_i^r \bigcap \mathbb{P}_{hole}^r\}$.

1) Pixel Information Fidelity: To quantify information loss, we should recognize the regions preserved by the retargeted image. The correspondence map M_{corr} is used in the pixel-level analysis of retargeting quality. Saliency is regarded as the information in our work. The ratio between the amount of saliency in the correspondence map and the total saliency is defined as the pixel information fidelity, as shown in Eq 7.

$$m_{pxl} = \frac{\sum_{\mathbf{p}_o \in M_{corr}} S_o(\mathbf{p}_o)}{\sum_{\mathbf{p}_o} S_o(\mathbf{p}_o)} \qquad (7)$$

By doing so, we establish a three-level representation of the retargeting process: region-level segmentation that examines the deformation of regions' shapes, patch-level partition that investigates the deformation of patches and pixel-level correspondence that studies the preservation of information. The three measures on these levels describe the overall fidelity of retargeted images.

IV. IMAGE RETARGETING INCONSISTENCY DETECTION

Apart from the three-level fidelity measures, we observe several anomalies that degrade retargeting quality greatly as shown in Fig. 1. Firstly, shape distortion can be inconsistent at two levels: region-level and patch-level. At region level, some important regions will be squeezed or stretched more severely than others while distortion of proportions occurs at patch

Fig. 6. Retargeted images with different types of shape distortion. For the Sanfrancisco image, the heart-shape structure is squeezed greatly. For the Deck image, discontinuities occupy the salient region and lead to shape distortion. The third retargeted image suffers from the distortion of proportions because the scaling factors for houses differ and result in shape distortion.

level. Secondly, some retargeted images crop important regions to preserve less important ones, contravening the objectives of image retargeting. Last but not least, deformation artifacts due to discontinuities alter the visual semantics of the original image. These two anomalies can be assessed by pixel-level correspondence map. In this section, we detail how to probe the anomalies in the retargeted image.

A. Region- & Patch-Level Inconsistency: Shape Distortion

Recently, some papers pay attention to shape distortion [3], [4], [8]. Still, they are not able to locate it in the retargeted image. As illustrated in [1], there are two causes of shape distortion: over-squeeze or over-stretch of foreground objects, and distortion of proportions in important regions. According to our observation, three instances of deformation may lead to such shape distortion. In Fig. 6, we exhibit three retargeted images, each of which represents a specific deformation instance of shape distortion. We detect such deformation instances by checking whether any of the following rules is violated.

Firstly, some algorithms apply continuous mapping to important regions, leading to over-squeeze or over-stretch. For instance, the heart-shape sculpture in the Sanfrancisco image of Fig. 6, which is an important structure in the original image, is squeezed heavily in the horizontal direction. To detect such deformation inconsistency, we compare the retargeted image with the image scaled by the same retargeting ratio in terms of the region shape similarity of important objects. If there exists some important regions whose region shape similarity is smaller than that of the corresponding scaled image, then the retargeted image is considered inconsistently retargeted

because it does not preserve important regions well. This first rule can be represented as Eq. 8,

$$r_{shape,1} = \begin{cases} 0 & \exists i \in \mathbb{S}_{imp}, \ m^i_{region} < m^i_{region,scl} - \delta \\ 1 & \text{otherwise} \end{cases} \quad (8)$$

where $\delta = 0.05$ is a tolerance parameter to increase the robustness of the rule. Secondly, several algorithms remove some content of important regions, causing over-squeeze or over-stretch. For example, some chairs on the deck in the Deck image of Fig. 6 are removed, inducing the over-squeeze of the foreground structures. To detect such deformation inconsistency, we check whether the important segments in the retargeted image are blemished by discontinuities. This rule can be represented by Eq. 9.

$$r_{shape,2} = \begin{cases} 0 & \exists i \in \mathbb{S}_{imp}, \ (\mathbf{p}_o, \mathbf{p}_r) \in \mathbb{P}_{dsc} : \mathbf{p}_r \in \mathbb{R}^r_i \\ 1 & \text{otherwise} \end{cases} \quad (9)$$

Finally, the third deformation instance is the distortion of proportions in the important regions. In the Brasserie_L_Aficion image of Fig. 6, the houses are scaled by different scaling factors, resulting in a distortion mirror effect. The third rule checks the consistency of deformation inside the same segment. All the local patches are examined whether its warping differs significantly from the segment it belongs to. If the portion of such local patches exceeds a threshold, then we consider the retargeting process incoherent. This rule can be represented by Eq. 10.

$$r_{shape,3} = \begin{cases} 1 & \sum_{i \in \mathbb{L}} s^i_{patch} > 95\% \\ & \mathbb{L} = \{i \,|\, \mathbb{P}^r_i \subset \mathbb{R}^r_j, r_{w,i} \in (0.8 \times r_{w,j}, 1.2), \\ & r_{h,i} \in (0.8 \times r_{h,j}, 1.2)\} \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

where $r_w = w_r/w_o$ and $r_h = h_r/h_o$ represents the width ratio and height ratio for the corresponding patch or segment after retargeting. By constraining the range of possible dimension ratios, the patches that suffer from abnormal deformation could be located. The overall detection rule for global inconsistency is hence $r_{shape} = r_{shape,1} \wedge r_{shape,2} \wedge r_{shape,3}$.

B. Pixel-Level Inconsistency: Foreground Cropping Detection

Image retargeting aims at the protection of salient objects or regions in an image. The cropping of foreground regions is thus undesirable. Still, some images have multiple foreground objects or regions and removal of some content is inevitable. As a result, we only consider the cropping of foreground regions for images labelled as evident foreground objects. If the image contains multiple foreground regions, we skip this detection and set $r_{crop} = 1$.

The detection of the above situation is quite simple. If the intersection of the importance pixel set $\mathbb{P}^o_{sal}$ and the boundary pixel set $\mathbb{P}^o_{bound}$ is not empty, then the retargeting process crops the important regions. The quality is thus worsened. This detection can be defined by Eq. 11.

$$r_{crop} = \begin{cases} 0 & \exists \mathbf{p}_o \in \mathbb{P}^o_{bound} : \mathbf{p}_o \in \mathbb{P}^o_{sal} \\ 1 & \text{otherwise} \end{cases} \quad (11)$$
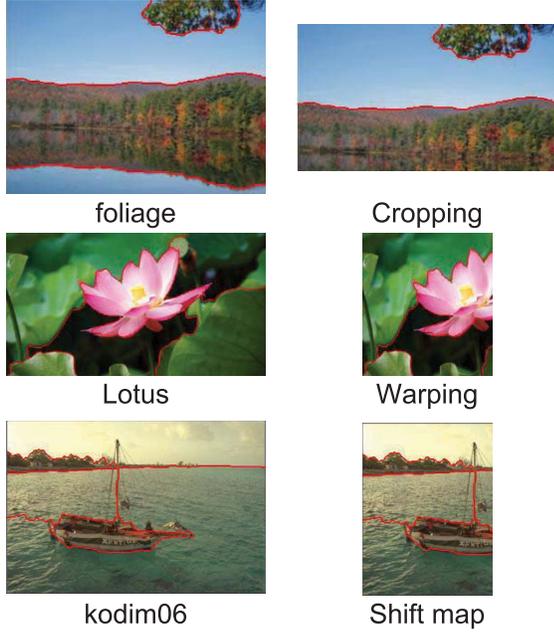
Fig. 7. Retargeted images which crop foreground regions. For each image, the segmentation boundary is labelled in red. The important objects are the forest, the lotus and the fishing boat respectively.

Most images in the datasets RetargetMe and CUHK contain multiple important objects. Therefore, the number of retargeted images in need of foreground cropping detection is small. Apart from the example we show in Fig. 1(c), there are ten other retargeted images which suffer from foreground cropping. Cropping, non-homogeneous warping and shift map editing are susceptible to foreground cropping. In Fig 7, we show three retargeted images with foreground cropping.

C. Pixel-Level Inconsistency: Discontinuity Detection

Deformation artifacts caused by discontinuities deteriorates retargeting quality greatly. However, due to their sparsity in the retargeted image, previous evaluation methods often ignore them after averaging all the quality measurements. To detect discontinuities in a retargeted image, we rely on the set of discontinuity candidates $\mathbb{P}_{dsc}$. SSIM index is applied to compare the similarity between the local patches around each pixel pair in the set. If the index is smaller than a threshold $\eta = 0.9$, the retargeted pixel in the pixel pair is regarded as a discontinuous pixel. If the affected retargeted pixels make up 5% of the total saliency, then we regard the image blemished by discontinuities. The detection for discontinuity can be defined by Eq. 12.

$$r_{dsc} = \sum_{(\mathbf{p}_o, \mathbf{p}_r) \in \mathbb{P}_{dsc}, SSIM(\mathbf{p}_o, \mathbf{p}_r) < \eta} S_r(\mathbf{p}_r) / \sum_{\mathbf{p}_r} S_r(\mathbf{p}_r) < 5\%$$

$$(12)$$

Discontinuity is prevalent in image retargeting, whose existence is detrimental to objects and structures, as shown in Fig. 8. In our work, the three proposed fidelity measures hardly detect them in most images. The introduction of this detection improves the performance of perceptual quality evaluation substantially.

Fig. 8. Retargeted images with discontinuities. Pixels which lead to discontinuities are labeled in red.

D. Overall Quality

To determine the final quality measure, we combine the fidelity measures and inconsistency detection by Eq. 13. Both fidelity measure Q_{Fidelity} and inconsistency detection $Q_{\text{Inconsistency}}$ ranges from [0.0.5]. Fidelity measure examines how much content has been preserved and how well they are preserved overall while the detection of deformation inconsistency tests against some anomalies. Here we assume that these three common distortions of retargeting, namely shape distortion, foreground cropping and discontinuity detection, are independent of each other. It is intuitive that foreground cropping does not lead to shape distortion nor discontinuities in the retargeted image, and vice versa. Discontinuities are mainly local distortions where far-away original pixels form new patches unseen in the original image. In contrast, shape distortions are global distortions where the original regions or patches are stretched or squeezed inconsistently. Consequently discontinuities and shape distortions are approximately independent. Any occurrence of these three distortions further reduces the retargeting quality. As a result, the inconsistency measure is defined as the linear fusion of the three detection results. The weights for three criteria sum up to one. Since shape distortion and deformation artifacts outweigh content information loss in determining image retargeting quality, we set $w_{crop} = 0.4$, $w_{shape} = w_{dsc} = 0.3$. The detection of any deformation inconsistency brings about similar deduction of quality indices.

$$Q = \underbrace{\frac{1}{4} m_{pxl} \times (m_{region} + m_{patch})}_{Q_{\text{Fidelity}}}$$
$$+ \underbrace{\frac{1}{2}(w_{crop} \times r_{crop} + w_{shape} \times r_{shape} + w_{dsc} \times r_{dsc})}_{Q_{\text{Inconsistency}}}$$

$$(13)$$

V. Experiment Results

A. Datasets

For the experiments, we test our retargeting quality assessment framework on two benchmark datasets: RetargetMe dataset [1] and CUHK dataset [21]. The RetargetMe dataset

TABLE I

COMPARISON OF RETARGETED IMAGE ALIGNMENT ACCURACY

Method	Retargeting Ratio	0.50		0.75		ALL	
	Database	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
SIFT-flow	RetargetMe	23.81	0.8081	32.20	0.9490	29.02	0.8957
	CUHK Session 1	26.60	0.7935	33.53	0.9508	28.13	0.8437
	CUHK Session 2	25.37	0.8254	30.56	0.9526	26.59	0.8554
Color-flow	RetargetMe	37.09	0.9796	38.30	0.9837	37.84	0.9821
	CUHK Session 1	37.44	0.9768	40.10	0.9859	38.29	0.9797
	CUHK Session 2	38.98	0.9831	38.09	0.9865	38.77	0.9839

contains 37 original images. Eight different operators, including cropping (CR), scaling (SCL), seam carving (SC) [11], multi-operator (MOP) [20], shift-map editing (SM) [14], scale-and-stretch (SNS) [16], streaming video (SV) [18] as well as non-homogeneous warping (WARP) [15], are compared based on the linked-paired comparison design. The viewer chooses the one with better quality from two retargeted images alongside. The number of votes for retargeted images is collected as the subjective quality scores. To compare subjective and objective scores, their rankings are assessed by Kendall τ rank coefficient defined by Eq. 14.

$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)} \qquad (14)$$

where n is the length of the ranking and equals 8 in this case. In addition, n_c is the number of concordant pairs while n_d is the number of disconcordant pairs. Here, the concordance examines whether the image with the higher subjective quality in the image pair is also favored by the predicted quality measure. A perfect match of all rankings induces that $\tau = 1$ while a perfect disagreement of all rankings has $\tau = -1$.

The CUHK dataset contains 57 original images. Apart from the eight algorithms studied in RetargetMe dataset, another two operators, optimized seam carving and scaling (SCSL) [19] and energy-based deformation (ENER) [17] are included as candidate operators. The dataset contains 171 retargeted images and the mean opinion scores (MOS) are collected as the subjective quality scores, which resembles traditional subjective image quality assessment. Besides, the standard deviations of opinion scores σ are also calculated for each image. To compare subjective and objective scores, we first fit the objective scores by the nonlinear function in Eq. 15,

$$f(x) = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\beta_2(x-\beta_3)}} \right) + \beta_4 x + \beta_5 \qquad (15)$$

which is inherited from [50]. After nonlinear regression, the fitted objective scores and subjective scores are compared in terms of Pearson linear correlation coefficient (LCC), Spearman rank-order correlation coefficient (SROCC), root mean square error (RMSE) and outlier ratio (OR). Here outlier ratio means the ratio of images that the fitted objective score is not in the interval $[MOS - 2\sigma, MOS + 2\sigma]$. Larger LCC and SROCC indicate higher correlation between subjective and objective scores and hence imply good performance of the objective quality measure. In contrast, a smaller RMSE and OR means better performance.

B. Alignment Accuracy

We tested on two benchmark datasets RetargetMe and CUHK to compare the alignment capabilities of SIFT descriptor and colors. Both datasets consist of retargeted images whose retargeting ratios are either 0.50 or 0.75. For each pair of original and retargeted images, we apply the alignment algorithm to obtain a correspondence map. A deformed image is generated by warping the original image according to the correspondence map. Finally, we compare the deformed image with the retargeted image. If the alignment is perfect, the two images should appear exactly the same. PSNR and SSIM index are utilized to compare the similarity between the two images. For SIFT-flow, we use the default parameter setting. In the implementation of color-flow, we normalize the three color channels before alignment. In this case, color-flow achieves its best performance when the smoothness parameter α falls within the range of [10, 30] for images whose dimensions are smaller than 500. We set $\alpha = 20$ in the experiment. The comparison is presented in Table I.

Besides, we have set up a validation dataset for the flow estimation. The dataset consists of about 140 original images from RetargetMe and CUHK database, more than 800 warped images and the ground truth warping correspondence. We choose three retargeting methods, seam carving [12], ARAP and ASAP [51] to retarget all the original images. Each image is retargeted by these three methods at two retargeting ratios: 0.50 and 0.75. We also warp 8 original images by applying content-aware rotation (CAR) [52]. Each image is rotated by two angles: 5 degrees and 10 degrees. Some original and warped images are shown in Fig. 9. The comparison between SIFT-flow and color-flow on this validation dataset is shown in Tables II. Here MAE is short for the mean absolute error and measured in terms of pixels. We also assess the precision of the alignment result, which is the ratio of pixels whose estimated flow vectors equal the ground truth one. The ground truth flow field for ASAP and ARAP are rounded to the nearest integer because they are continuous methods and the ground truth flow vector may have a sub-pixel precision while our estimated vectors only have pixel-level precision.

From both tables, we notice that our alignment algorithm outperforms SIFT-flow on both retargeted and rotated images. The image reconstructed by color-flow looks more like the warped images subjectively. Besides, color-flow has less registration errors and higher accuracy. SIFT-flow degrades significantly when the retargeting ratio is 0.50 or when the rotation angle is 10 degrees. The warped images suffer from severe distortion under both cases. Warped images contain many
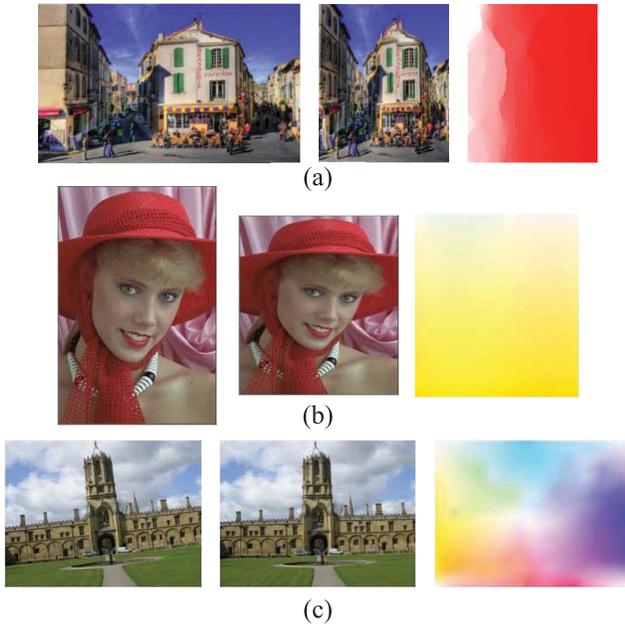
Fig. 9. Images from the validation dataset. From left to right: original image, warped image, correspondence flow vectors. From top to bottom: (a) 0.5 times width retargeting by SC; (b) 0.75 times height retargeting by ASAP; (c) 5 degrees CAR. The flow field is visualized by [39].

TABLE II

VALIDATION OF THE FLOW FIELD ESTIMATION

		SIFT-Flow		Color-Flow	
Retargeting	ratio	MAE	PREC	MAE	PREC
SC	0.50	6.28	0.46	4.35	0.56
	0.75	1.40	0.71	0.90	0.75
ARAP	0.50	5.78	0.21	3.51	0.48
	0.75	1.24	0.53	0.80	0.68
ASAP	0.50	7.01	0.07	4.25	0.29
	0.75	1.44	0.32	1.14	0.48
Rotation	degree	MAE	PREC	MAE	PREC
CAR	5°	0.54	0.65	0.51	0.71
	10°	0.98	0.41	0.65	0.61

unforeseen textures whose SIFT descriptors differ greatly from the original patches. Therefore, the performance of SIFT flow is very poor. Even though the amount of unforeseen textures is reduced when the retargeting ratio is 0.75 or when the rotation angle is 5 degrees, our method is still superior.

The basic hypothesis that the SIFT descriptors of the matched pixels may not alter does not hold for image retargeting. Two matched local patches may differ significantly in terms of SIFT descriptors because of severe distortion. Although color information is not a good matching descriptor for many applications, in the context of image retargeting, two matched pixels should still share the same color. Thats why we make use of color information to guide the flow estimation. The second term in the flow energy functions guarantees the continuity of the flow vectors; therefore the descriptors reflect both local and global color similarity, leading to higher estimation accuracy. This experiment shows that colors align retargeted images much more accurately than SIFT descriptors.

TABLE III

THE IMPACT OF THE SALIENCY THRESHOLD FOR SALIENT SEGMENTS

Threshold	0.50	0.55	0.60	0.65	0.70
mean of τ	0.596	0.600	0.625	0.613	0.613
std of τ	0.227	0.228	0.231	0.236	0.236

TABLE IV

THE IMPACT OF PARAMETERS FOR THE INCONSISTENCY DETECTION

w_{crop}	w_{dsc}	w_{shape}	average τ coefficients
1	0	0	0.204
0	1	0	0.329
0	0	1	0.415
0.1	0.1	0.8	0.557
0.8	0.1	0.1	0.577
0.1	0.8	0.1	0.507
0.5	0.2	0.3	0.584
0.4	0.3	0.3	0.577

C. Parameters Settings for the Overall Quality

In the inconsistency detection, we define the salient segments as the segments which occupy 60% of total saliency values. The threshold has some impact on the number of salient segments and hence influence the result of shape distortion detection. In this section, we first validate the choice of this threshold value. This threshold is sampled in the range of $[0.50, 0.70]$ with a step size of 0.05. We apply our framework on the RetargetMe dataset using different thresholds and evaluate the average τ coefficients on the 37 images. The comparison is shown in Table III. As shown in the table, the choice of the saliency threshold has some impact on the quality prediction performance. When the threshold is set within $[0.50, 0.70]$, the performance of the overall quality is robust to the change of the saliency threshold. The number of salient segments by using different thresholds in this range differs by no more than two. Therefore, its impact on the shape distortion detection is not significant.

There are three parameters for the inconsistency detection in Eq. 13. In this subsection, we discuss how different parameters impact on the inconsistency detection. The experiment is tested on the RetargetMe dataset and we compare the quality prediction performance of the inconsistency detection $Q_{\text{Inconsistency}}$ under different parameter settings in terms of average Kendall τ coefficient. The performance under different parameters is listed in Table IV. Among the three inconsistency detections, foreground cropping has the least τ coefficients because its occurrence is quite low. Retargeted images with discontinuity are sometimes favored by human viewers, such as the ones shown in Fig. 11. Shape inconsistency has the best performance among the three because it occurs quite often in retargeted images. As listed in the table, the performance is quite robust to the change of parameters as long as the setting are not too extreme, for example setting one of the three parameters within in the range of $[0, 0.1]$. For most parameter settings, the average τ coefficients are greater than 0.50. The inconsistency detection is hence an essential component that classifies the retargeted images into different quality groups. Together with the fidelity measures, the inconsistency detection makes a good quality predictor.

TABLE V

COMPARISON OF DIFFERENT IMAGE RETARGETING QUALITY ASSESSMENT METHODS ON RETARGETME DATASET

Quality Evaluation Method	Attribute						Total		
	Lines/Edges	Faces/People	Foreground Objects	Texture	Geometric Structures	Symmetry	Mean	std	p-value
SIFT flow	0.097	0.252	0.218	0.161	0.085	0.071	0.145	0.262	0.031
EMD	0.220	0.262	0.226	0.107	0.237	0.500	0.251	0.272	1e-5
IRQA [6]	0.097	0.290	0.293	0.161	0.053	0.150	0.164	0.263	0.028
IR-SSIM	0.309	0.452	0.377	0.321	0.313	0.333	0.363	0.271	1e-3
Liang [5]	0.351	0.271	0.381	0.304	0.415	0.548	0.399	-	-
PGDIL [7]	0.431	0.390	0.389	0.286	0.438	0.523	0.415	0.296	6e-10
ARS [3]	0.463	0.519	0.444	0.330	0.505	0.464	0.452	0.283	1e-11
Region-Based [8]	0.477	0.571	0.519	0.462	0.448	0.456	0.503	0.183	9e-14
Q_{Fidelity}	0.071	0	0.046	0.201	0.064	0.102	0.076	0.320	0.1064
$Q_{\text{Inconsistency}}$	0.563	0.578	0.581	0.515	0.573	0.556	0.577	0.133	8e-40
Proposed	**0.632**	**0.605**	**0.597**	**0.588**	**0.591**	**0.651**	**0.625**	**0.231**	**2e-21**

Our proposed fidelity measure is the multiplication of the pixel information fidelity m_{pxl} and the shape fidelity $m_{region}+m_{patch}$. By applying this fidelity measure in the overall quality, the mean and standard deviation of the τ coefficients for the overall quality is $(0.625, 0.231)$ on the RetargetMe dataset. We also test the performance of the overall quality by modifying the fidelity measure as a linear fusion of the three measures $Q'_{\text{Fidelity}} = w_{pxl} \times m_{pxl} + w_{region} \times m_{region} + w_{patch} \times m_{patch}$. When the trio of parameter $(w_{pxl}, w_{region}, w_{patch})$ equals $(0.2, 0.15, 0.15)$, the mean and standard deviation of the resultant τ coefficients is $(0.627, 0.221)$. By setting the parameters as $(0.17, 0.17, 0.17)$, the mean and standard deviation of the resultant τ coefficients is $(0.633, 0.201)$. The use of weighted sum or multiplication in the fidelity measure does not differ significantly on the performance. We choose multiplication of fidelity measures mainly because the measures for region shape fidelity and patch shape fidelity are always biased towards the retargeting method of cropping, which happens to achieve high retargeting quality on the RetargetMe dataset. By multiplication of information fidelity and shape fidelity, the measure is less biased towards cropping and apply to more general datasets where cropped images are not so preferred. In the following subsection, the performance of the whole framework is tested on two benchmark datasets.

D. Performance on RetargetMe Dataset

To demonstrate the performance of our evaluation method, we compare it with two computational image distances SIFT flow [25] and EMD [26], and six quality measures based on image alignment: IRQA [6], IR-SSIM [2], Liang [5], PGDIL [7], ARS [3] and the region-based method [8]. Among them, IRQA and IR-SSIM focus on pixel-level fidelity; PGDIL and ARS combine pixel-level and patch-level fidelity analyses; Liang's method and the region-based method combine pixel-level and region-level fidelity analyses. We also compare the respective performance of the fidelity measures Q_{Fidelity} and the inconsistency detection $Q_{\text{Inconsistency}}$. In Table V, we show the mean and standard deviation of the τ coefficients on the 37 image sets. The performance on images of different attributes is also listed in the table. In addition, a χ^2 test against the null hypothesis that the observed coefficients are randomly sampled from the τ distribution is examined on each retargeting method. A small p-value indicates that we can reject the hypothesis and the predicted ranks are not
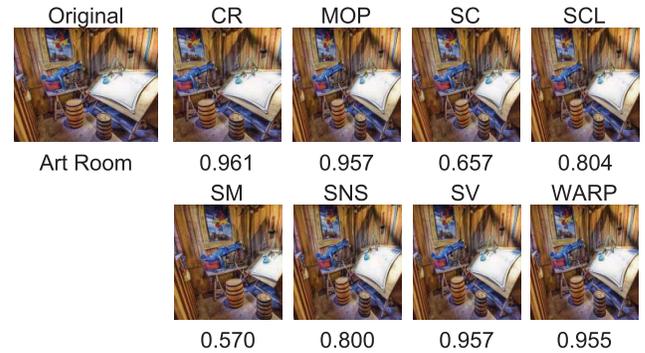
Fig. 10. Predicted quality of the retargeted images on the 'ArtRoom' image set.

randomly chosen. For a fair comparison, we directly copy the performance statistics provided by the original papers with some missing data left blank.

From the table, we observe that most measures based on image alignment outperform computational image distances. By combining different levels of fidelity analysis, PGDIL [7], ARS [3], Liang's [5] and the region-based method [8] achieve better performance than IRQA [6] and IR-SSIM [2]. Although Liang's method [5] also incorporates the region-level analysis by measuring the similarity of area and color histogram of salient objects, the salient regions extracted by [44] are not always aligned after retargeting. This degrades its performance to some extent. Our proposed fidelity measures alone have mediocre performance because they do not detect severe distortions in the raw measures. The retargeted images with a small part of severe distortions are overwhelmed by well-preserved regions, and their fidelity measures are higher than those of the retargeted images that distribute distortions evenly. This is the common problem for many earlier works like IRQA [6] and IR-SSIM [2]. The improvement of our method mainly attributes to the detection of shape distortion, foreground cropping and discontinuities. The detection of severe distortions attempts to classify retargeted images into different quality groups and contributes to the performance of the whole framework.

Besides, we show the predicted quality of the retargeted images on one of the image set in Fig. 10. Among the 8 algorithms, SC and SM suffer from discontinuities as well as shape distortions. Therefore, they have a rather low quality.

TABLE VI

THE VIOLATION RATE OF THREE CRITERIA FOR
DIFFERENT ALGORITHMS (%)

Operator	r_{crop}	r_{dsc}	r_{shape}	None
CR	10.81	0	0	89.19
MOP	0	13.51	8.11	78.38
SC	0	94.59	64.86	5.41
SCL	0	0	1	0
SM	2.7	43.24	29.73	48.65
SNS	0	24.32	75.68	21.62
SV	0	5.41	21.62	72.97
WARP	8.11	51.35	56.76	24.32

TABLE VII

COMPARISON OF DIFFERENT IMAGE RETARGETING QUALITY
ASSESSMENT METHODS ON CUHK DATASET

Measure	LCC	SROCC	RMSE	OR
SIFT flow	0.3141	0.2899	12.817	0.1462
EMD	0.2760	0.2904	12.977	0.1696
IRQA [6]	0.4374	0.4662	12.141	0.1520
GLS [4]	0.4622	0.4760	10.932	0.1345
Liang [5]	0.4428	0.4669	12.105	0.1813
PGDIL [7]	0.5403	0.5409	11.361	0.1520
ARS [3]	0.6835	0.6693	9.855	0.0702
Region-Based [8]	0.7397	0.7519	9.129	0.0760
Q_{Fidelity}	0.3526	0.4252	12.634	0.211
$Q_{\text{Inconsistency}}$	0.7537	0.7601	8.873	0.0760
Proposed	**0.7974**	**0.7984**	**8.378**	**0.0643**

Besides, SCL and SNS only suffer from shape distortion. In contrast, the other 4 retargeted images are free of severe distortions. Consequently they have the best quality.

We also analyze the common types of anomalies for the 8 algorithms. In Table VI, each item shows the rate of occurrence for the corresponding anomaly among all 37 image sets. From the table, we can notice that CR, MOP and SV usually produce high-quality retargeted images because they rarely generate images with deformation inconsistency. Shift map is prone to discontinuities and ranks fourth on the number of retargeted images with no deformation inconsistency. The other four algorithms are not always robust and prone to distortions. SC is susceptible to discontinuities while SNS suffers from distortion of proportions on many images. Non-homogeneous warping is constructed on a linear system, whose robustness is poor. Scaling always produces shape distortion. This analysis coincides with the study in [1] and inspires that we should pay more attention to the detection of inconsistency in formulating the retargeting quality.

E. Performance on CUHK Dataset

On CUHK dataset, we also compare with 8 other retargeting quality assessment frameworks. IR-SSIM is replaced by GLS [4] because the latter one is especially suited to this dataset by incorporating patch-level fidelity analysis. The LCC, SROCC, RMSE and OR between the fitted objective scores and MOS are listed in Table VII. The authors in [5] did not evaluate their method on the CUHK dataset. Therefore, we test its performance by running its open-source code on this dataset. Again, more levels of fidelity analysis promise higher performance in quality prediction. Liang's method [5], which does not align the salient content in the original and
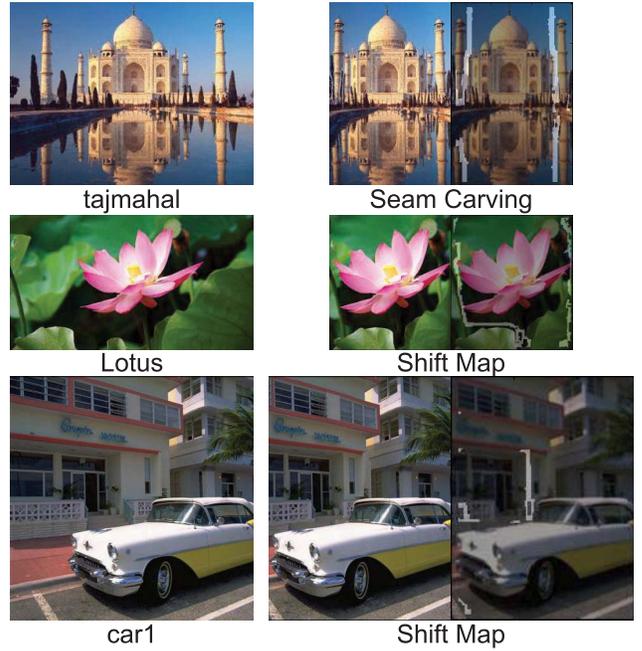
Fig. 11. Retargeted images with discontinuities which do not provoke visual artifacts. The white regions on the retargeted images denote discontinuities.

the retargeted images, performs inferiorly to the other methods with multiple-level fidelity analysis. Our method outperforms the rest because it can recognize the inconsistency in a retargeted image. The detection of deformation inconsistency acts as a quality classifier. With more instances of deformation inconsistency, the objective quality of retargeted images drops quickly. For images which suffer from the same number of inconsistency instances, the fidelity measure is capable of choosing the one with better quality. Overall, the combination of fidelity measures and inconsistency detection achieves great performance on predicting the quality of image retargeting.

F. Limitations

Despite such improvement, the proposed method has its limitations. We assume that inconsistency on different levels degrades retargeting quality. For most images, this assumption holds because such inconsistency leads to visual artifacts and shape distortion. However, as shown in Fig. 11, there are a few retargeted images with strong discontinuities favored by many viewers. Discontinuities in these retargeted images resemble real-world scenes and deceive the viewers. Besides, human viewers have distinct tolerance of deformation inconsistency on different images. The same amount of shape distortion or discontinuities may have disparate impact on the retargeting quality. The semantics of the images need to be studied to solve this problem, which is currently hard to implement.

VI. CONCLUSION

Image retargeting adapts original content to arbitrary sizes under two goals: preserve the important content and avoid introducing annoying inconsistency. Therefore, to formulate the retargeting quality, both the retargeting fidelity and deformation inconsistency should be considered. In this

paper, we propose a retargeting quality assessment framework which could model fidelity and detect inconsistency on different levels. To achieve this goal, we improve the alignment algorithm between the original and retargeted images. To obtain the final quality measure, we assess the retargeting fidelity and detect deformation inconsistency on three different levels: region-level segmentation, patch-level partition and pixel-level correspondence. On both RetargetMe and CUHK dataset, we achieved the best performance. This inspires us to concentrate more on the detection of retargeting inconsistency in the quality assessment of image retargeting.

References

[1] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir, "A comparative study of image retargeting," *ACM Trans. Graph.*, vol. 29, no. 6, p. 160, 2010.

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

[3] Y. Zhang, Y. Fang, W. Lin, X. Zhang, and L. Li, "Backward registration-based aspect ratio similarity for image retargeting quality assessment," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4286–4297, Sep. 2016.

[4] J. Zhang and C.-C. J. Kuo, "An objective quality of experience (QoE) assessment index for retargeted images," in *Proc. 22nd ACM Int. Conf. Multimedia*, 2014, pp. 257–266.

[5] Y. Liang, Y.-J. Liu, and D. Gutierrez, "Objective quality prediction of image retargeting algorithms," *IEEE Trans. Vis. Comput. Graphics*, vol. 23, no. 2, pp. 1099–1110, Feb. 2016.

[6] Y.-J. Liu, X. Luo, Y.-M. Xuan, W.-F. Chen, and X.-L. Fu, "Image retargeting quality assessment," *Comput. Graph. Forum*, vol. 30, no. 2, pp. 583–592, Apr. 2011.

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

[8] Y. Zhang and K. N. Ngan, "Region-based image retargeting quality assessment," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 1757–1761.

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

[10] F. Banterle *et al.*, "Multidimensional image retargeting," in *Proc. SIGGRAPH Asia Courses*, 2011, p. 15.

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

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

[13] M. Grundmann, V. Kwatra, M. Han, and I. Essa, "Discontinuous seam-carving for video retargeting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, pp. 569–576.

[14] Y. Pritch, E. Kav-Venaki, and S. Peleg, "Shift-map image editing," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 151–158.

[15] L. Wolf, M. Guttmann, and D. Cohen-Or, "Non-homogeneous content-driven video-retargeting," in *Proc. IEEE 11th Int. Conf. Comput. Vis. (ICCV)*, Oct. 2007, pp. 1–6.

[16] Y.-S. Wang, C.-L. Tai, O. Sorkine, and T.-Y. Lee, "Optimized scale-and-stretch for image resizing," *ACM Trans. Graph.*, vol. 27, no. 5, p. 118, Dec. 2008.

[17] Z. Karni, D. Freedman, and C. Gotsman, "Energy-based image deformation," *Comput. Graph. Forum*, vol. 28, no. 5, pp. 1257–1268, 2009.

[18] P. Krähenbühl, M. Lang, A. Hornung, and M. Gross, "A system for retargeting of streaming video," *ACM Trans. Graph.*, vol. 28, no. 5, p. 126, 2009.

[19] W. Dong, N. Zhou, J.-C. Paul, and X. Zhang, "Optimized image resizing using seam carving and scaling," *ACM Trans. Graph.*, vol. 28, no. 5, p. 125, 2009.

[20] M. Rubinstein, A. Shamir, and S. Avidan, "Multi-operator media retargeting," *ACM Trans. Graph.*, vol. 28, no. 3, p. 23, 2009.

[21] L. Ma, W. Lin, C. Deng, and K. N. Ngan, "Image retargeting quality assessment: A study of subjective scores and objective metrics," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 6, pp. 626–639, Oct. 2012.

[22] D. Simakov, Y. Caspi, E. Shechtman, and M. Irani, "Summarizing visual data using bidirectional similarity," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2008, pp. 1–8.

[23] B. S. Manjunath, J. R. Ohm, V. V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 703–715, Jun. 2001.

[24] E. Kasutani and A. Yamada, "The MPEG-7 color layout descriptor: A compact image feature description for high-speed image/video segment retrieval," in *Proc. IEEE Int. Conf. Image Process.*, vol. 1. Oct. 2001, pp. 674–677.

[25] C. Liu, J. Yuen, A. Torralba, J. Sivic, and W. T. Freeman, "SIFT flow: Dense correspondence across different scenes," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 28–42.

[26] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *Int. J. Comput. Vis.*, vol. 40, no. 2, pp. 99–121, Nov. 2000.

[27] S. Castillo, T. Judd, and D. Gutierrez, "Using eye-tracking to assess different image retargeting methods," in *Proc. ACM SIGGRAPH Symp. Appl. Perception Graph. Vis.*, 2011, pp. 7–14.

[28] W. Lin and C.-C. Jay Kuo, "Perceptual visual quality metrics: A survey," *J. Vis. Commun. Image Represent.*, vol. 22, no. 4, pp. 297–312, 2011.

[29] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[30] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. IEEE 37th Asilomar Conf. Signals, Syst. Comput.*, vol. 2. Nov. 2003, pp. 1398–1402.

[31] W. Zhang, C. Qu, L. Ma, J. Guan, and R. Huang, "Learning structure of stereoscopic image for no-reference quality assessment with convolutional neural network," *Pattern Recognit.*, vol. 59, pp. 176–187, Nov. 2016.

[32] J. Guan, W. Zhang, J. Gu, and H. Ren, "No-reference blur assessment based on edge modeling," *J. Vis. Commun. Image Represent.*, vol. 29, pp. 1–7, May 2015.

[33] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[34] A. Liu, W. Lin, H. Chen, and P. Zhang, "Image retargeting quality assessment based on support vector regression," *Signal Process., Image Commun.*, vol. 39, pp. 444–456, Nov. 2015.

[35] B. Gu and V. S. Sheng, "A robust regularization path algorithm for (ν)-support vector classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 5, pp. 1241–1248, May 2017.

[36] B. Gu, X. Sun, and V. S. Sheng, "Structural minimax probability machine," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 7, pp. 1646–1656, Jul. 2017.

[37] B. Gu, V. S. Sheng, K. Y. Tay, W. Romano, and S. Li, "Incremental support vector learning for ordinal regression," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 7, pp. 1403–1416, Jul. 2015.

[38] L. Ma, L. Xu, Y. Zhang, Y. Yan, and K. N. Ngan, "No-reference retargeted image quality assessment based on pairwise rank learning," *IEEE Trans. Multimedia*, vol. 18, no. 11, pp. 2228–2237, Nov. 2016.

[39] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski, "A database and evaluation methodology for optical flow," *Int. J. Comput. Vis.*, vol. 92, no. 1, pp. 1–31, 2011.

[40] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[41] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2007, pp. 1–8.

[42] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 1915–1926, Oct. 2012.

[43] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proc. Adv. Neural Inf. Process. Syst.*, 2006, pp. 545–552.

[44] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu, "Global contrast based salient region detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 569–582, Mar. 2015.

[45] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.*, vol. 59, no. 2, pp. 167–181, Sep. 2004.

[46] A. Y. Yang, J. Wright, Y. Ma, and S. S. Sastry, "Unsupervised segmentation of natural images via lossy data compression," *Comput. Vis. Image Understand.*, vol. 110, no. 2, pp. 212–225, 2008.

[47] Z. Li, X.-M. Wu, and S.-F. Chang, "Segmentation using superpixels: A bipartite graph partitioning approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 789–796.

[48] H. Freeman, "Computer processing of line-drawing images," *ACM Comput. Surv.*, vol. 6, no. 1, pp. 57–97, 1974.

[49] J. Iivarinen and A. J. E. Visa, "Shape recognition of irregular objects," *Proc. SPIE*, vol. 2904, pp. 25–32, Oct. 1996.

[50] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.

[51] D. Panozzo, O. Weber, and O. Sorkine, "Robust image retargeting via axis-aligned deformation," *Comput. Graph. Forum*, vol. 31, no. 2pt1, pp. 229–236, May 2012.

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

Yichi Zhang (S'15) received the B.S. degree in information engineering from Shanghai Jiao Tong University, Shanghai, China, in 2013. He is currently pursuing the Ph.D. degree with the Department of Electronic Engineering, The Chinese University of Hong Kong.

His research interests include image/video processing, image quality assessment, and visual saliency.

King Ngi Ngan (F'00) received the Ph.D. degree in electrical engineering from Loughborough University, U.K. He was a Chair Professor with The Chinese University of Hong Kong, the Nanyang Technological University, Singapore, and the University of Western Australia, Australia. He is currently a Chair Professor under the National Thousand Talents Program with the University of Electronic Science and Technology, Chengdu, China. He also holds honorary and visiting professorships of numerous universities in China, Australia, and South East Asia.

He has authored extensively including three authored books, seven edited volumes, over 400 refereed technical papers, and edited nine special issues in journals. He also holds 15 patents in the areas of image/video coding and communications. He chaired and co-chaired a number of prestigious international conferences on image and video processing including the 2010 IEEE International Conference on Image Processing, and served on the advisory and technical committees of numerous professional organizations. He served as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the *Journal on Visual Communications and Image Representation*, the *EURASIP Journal of Signal Processing: Image Communication*, and the *Journal of Applied Signal Processing*.

Dr. Ngan is a fellow of the IET, U.K., and the IEAust, Australia, and an IEEE Distinguished Lecturer from 2006 to 2007.
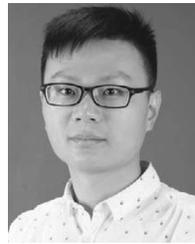
Lin Ma (M'13) received the B.E. and M.E. degrees in computer science from the Harbin Institute of Technology, Harbin, China, in 2006 and 2008, respectively, and the Ph.D. degree from the Department of Electronic Engineering, The Chinese University of Hong Kong, in 2013. He was a Researcher with the Huawei Noah'Ark Laboratory, Hong Kong, from 2013 to 2016. He is currently a Senior Researcher with the Tencent AI Laboratory, Shenzhen, China. His current research interests lie in the areas of deep learning and multimodal learning, specifically for image and language, image/video understanding, and quality assessment.

Dr. Ma received the Best Paper Award from the Pacific-Rim Conference on Multimedia in 2008. He was a recipient of the Microsoft Research Asia Fellowship in 2011. He was a finalist in HKIS Young Scientist Award in engineering science in 2012.

Hongliang Li (SM'12) received the Ph.D. degree in electronics and information engineering from Xian Jiaotong University, China, in 2005. From 2005 to 2006, he joined the Visual Signal Processing and Communication Laboratory, The Chinese University of Hong Kong, as a Research Associate, where he was a Post-Doctoral Fellow from 2006 to 2008. He is currently a Professor with the School of Electronic Engineering, University of Electronic Science and Technology of China. His research interests include image segmentation, object detection, image and video coding, visual attention, and multimedia communication system. He has authored or co-authored numerous technical articles in well-known international journals and conferences. He has co-edited a Springer book *Video Segmentation and Its Applications*. He was involved in many professional activities. He is a member of the Editorial Board of the *Journal on Visual Communications and Image Representation*, and the Area Editor of *Signal Processing: Image Communication*, Elsevier Science. He served as a Technical Program Co-Chair of VCIP2016 and ISPACS 2009, the General Co-Chair of the ISPACS 2010, a Publicity Co-Chair of the IEEE VCIP 2013, a Local Chair of the IEEE ICME 2014, and the TPC members in a number of international conferences, including the ICME 2013, ICME 2012, ISCAS 2013, PCM 2007, PCM 2009, and VCIP 2010.