Backward Registration-Based Aspect Ratio Similarity for Image Retargeting Quality Assessment

Yabin Zhang, Yuming Fang, Weisi Lin, *Fellow, IEEE*, Xinfeng Zhang, *Member, IEEE*, and Leida Li

***Abstract*— During the past few years, there have been various kinds of content-aware image retargeting operators proposed for image resizing. However, the lack of effective objective retargeting quality assessment metrics limits the further development of image retargeting techniques. Different from traditional image quality assessment (IQA) metrics, the quality degradation during image retargeting is caused by artificial retargeting modifications, and the difficulty for image retargeting quality assessment (IRQA) lies in the alternation of the image resolution and content, which makes it impossible to directly evaluate the quality degradation like traditional IQA. In this paper, we interpret the image retargeting in a unified framework of resampling grid generation and forward resampling. We show that the geometric change estimation is an efficient way to clarify the relationship between the images. We formulate the geometric change estimation as a backward registration problem with Markov random field and provide an effective solution. The geometric change aims to provide the evidence about how the original image is resized into the target image. Under the guidance of the geometric change, we develop a novel aspect ratio similarity (ARS) metric to evaluate the visual quality of retargeted images by exploiting the local block changes with a visual importance pooling strategy. Experimental results on the publicly available MIT RetargetMe and CUHK data sets demonstrate that the proposed ARS can predict more accurate visual quality of retargeted images compared with the state-of-the-art IRQA metrics.**

***Index Terms*— Image retargeting quality assessment, geometric change, backward registration.**

I. Introduction

RECENTLY, the increasing trend of widely used display devices has imposed the demand for image adaptation to different resolutions and aspect ratios. In addition, designers like to edit images into different sizes for different purposes. In the past few years, a number of content-aware image retargeting methods have been developed for image resizing [1]. The image retargeting operators can be broadly categorized into two types: discrete and continuous approaches [2], based on whether they treat images as discrete pixels or continuous signals. The typical discrete approaches mainly include manual Cropping (CR), Seam-Carving (SC) [3], and Shift-Map (SM) [4], while uniform Scaling in one dimension (SCL), non-homogeneous Warping (WARP) [5], Streaming Video (SV) [6] and Scale-and-Stretch (SNS) [7] are representative continuous methods. Both kinds of content-aware approaches remove pixels or warp the image to the targeted resolution according to the visual importance of image content. The purpose of the content-aware image retargeting is to preserve visually important content and structure (i.e. to reduce information loss), and at the same time, to limit the visual distortion in the retargeted image [1], [2], [8]. CR and SCL are two traditional methods based on geometric constraints without considerations of image content. In retargeted images by CR, there is only information loss occurring, while in images produced by SCL, visual distortions due to squeezing or stretching degrade the image quality. Since the size reduction during retargeting is inevitable, most content-aware methods try to remove or shrink the less important content, and thus achieve better overall performance by balancing information loss and visual distortion.

Although there are many image retargeting operators proposed for image resizing, the visual quality evaluation of retargeted images is still a challenging task. To measure the quality of retargeted images, Rubinstein *et al.* [1] conducted a comparative study of different retargeting operators, and also provided 6 objective quality metrics: Bidirectional Similarity (BDS), Bidirectional Warping (BDW), Edge Histogram (EH), Color Layout (CL), SIFT flow and Earth-Mover's Distance (EMD). The BDS is a patch based similarity metric formulated with a bidirectional mapping between two images. For patch in one image, the matched patch is searched in the other image and vice versa. The mean distance of these patches is summarized as the BDS distance. The BDW is one similar metric like BDS, but the mapping between two images is constrained to be monotonic. The EH and CL are two low-level metrics in MPEG-7 standard, where EH uses the histogram of the spatial edge distribution as the image representation and the

Manuscript received September 21, 2015; revised March 15, 2016 and May 16, 2016; accepted June 21, 2016. Date of publication June 28, 2016; date of current version July 21, 2016. This work was supported in part by the National Research Foundation, Singapore, within the Rapid-Rich Object Search Laboratory, Nanyang Technological University through the Interactive Digital Media Strategic Research Programme and in part by Singapore MoE Tier 1 Project under Grant M4011379 and Grant RG141/14. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Stefan Winkler. *(Corresponding author: Weisi Lin.)*

Y. Zhang, W. Lin, and X. Zhang are with the Rapid-Rich Object Search Laboratory, Nanyang Technological University, Singapore 639798 (e-mail: zhan0398@ntu.edu.sg; xfzhang@ntu.edu.sg; wslin@ntu.edu.sg).

Y. Fang is with the School of Information Technology, Jiangxi University of Finance and Economics, Nanchang 330013, China (e-mail: fa0001ng@e.ntu.edu.sg).

L. Li is with the School of Information and Electrical Engineering, China University of Mining and Technology, Xuzhou 221116, China (e-mail: reader1104@163.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2016.2585884

L1 distance of the two histogram as the image distance, while similarly, CL is an image descriptor which captures the spatial distribution of colors in the image. SIFT flow is a recently proposed across-scene image alignment algorithm and EMD is a dissimilarity metric between two distributions defined as the minimal cost to transform one distribution into the other. Among these metrics, EMD and SIFT flow show the best agreement with subjective rankings, and the reason might be that their constrained matching appears to be able to model the retargeting modification on images more accurately.

The early IRQA work [9] proposed an automatic evaluation framework for image retargeting in the principle of user perception based criteria of important content emphasis, global information coverage and visual distortion reduction. In [10], Liu *et al.* exploited global structure and local correspondence to measure the quality of retargeted images. By traversing from coarse scales to fine scales, that metric extracted the global geometric structure and established the local pixel correspondence in a top-down manner. Finally, based on the correspondence, the retargeted image is evaluated with a saliency weighted similarity metric. In [11], Ma *et al.* investigated the performance of different shape descriptors for the quality evaluation of retargeted images. Fang *et al.* [12] proposed an algorithm (IR-SSIM) to generate Structural Similarity (SSIM) map to measure the quality of the preserved structural information in retargeted images. They first utilized the SIFT flow to explore the dense correspondence between the images, and then generated a SSIM map to measure the amount of the preserved structural information in the retargeted image. They exploited both bottom-up and top-down image saliency estimation as the visual importance map for retargeting. The spatial pooling of the visual importance map with the SSIM map provided the overall evaluation of the retargeted image. In [13], Zhang *et al.* investigated the features from three determining factors for objective quality of experience assessment on retargeted images: global structural distortions (G), local region distortions (L) and loss of salient information (S), and proposed the GLS metric by fusion with regression methods. Liu *et al.* [14] also utilized machine learning to fuse different quality estimator including spatial estimator and frequency domain quality estimator to achieve promising performance.

In [8], Hsu *et al.* obtained the dense correspondence between original and retargeted images with SIFT flow technique as well. By measuring the local variance of the SIFT flow field between the images, the perceptual geometric distortions are measured, and by measuring the saliency loss during warping the original saliency map to the target resolutions, the perceptual information loss is measured. An adaptive fusion method is proposed to fuse perceptual geometric distortion and information loss as the overall evaluation of the retargeted image. However, the relationship between the images are still not fully investigated and the correspondence estimation has limitations on finding the retargeting modification.

In traditional Image Quality Assessment (IQA) [15]–[17], the images are generally assumed to be well-aligned and the image difference can be obtained by the direct subtraction, while in IRQA the relationship between the original and retargeted images is more complicated due to various kinds of retargeting modifications. In this paper we propose an efficient geometric change estimation method to find the relationship between the original and retargeted images and also a novel Aspect Ratio Similarity metric (ARS) to predict the visual quality of retargeted images. Compared to related existing works, our major contributions are three-fold:

- Firstly, we give a unified interpretation of image retargeting and show that the geometric change estimation is an efficient way to clarify the relationship between the original and retargeted images.
- Secondly, we formulate the geometric change estimation as a backward registration problem with the MRF and provide a practical and effective solution.
- Thirdly, under the geometric change guidance we develop a novel ARS metric, which is effective and outperforms other existing methods on publicly available datasets.

The remainder of this paper is organized as follows. Section II introduces the proposed backward registration based methodology. We presents the unified interpretation for retargeting, the backward registration problem formulation and its solution. Section III presents the details about ARS metric. Experimental results shown in Section IV demonstrate that our metric outperforms other existing metrics. Finally, Section V concludes the paper and discusses the future work.

II. The Proposed Backward Registration Based Methodology

The relationship between the original and retargeted images is complicated due to different kinds of artificial modifications. To develop effective IRQA metrics, it is important to discover the undergone image retargeting modification to clarify the relationship between the original and retargeted images.

The proposed IRQA framework is shown in Fig. 1. Given the original and retargeted images, it is necessary to find the undergone geometric change before the quality evaluation. We solve the backward registration problem to reveal the geometric change during image retargeting. We exploit the geometric change to establish the block correspondence between the original and retargeted images. For the quality evaluation, we focus on how the local blocks are modified in the retargeted image and design a novel ARS metric to assess the local block changes by considering the information loss and visual distortion. By pooling it with the visual importance map, we can predict the overall visual quality of the retargeted image effectively.

A. Motivation

1) To Evaluate the Retargeted Image: The visual quality evaluation of retargeted images is a semantic high-level task. As the analysis in [1], [8], and [12], the objective of image retargeting is to preserve the important image content while avoid introducing the visual distortion. However, due to the complicated relationship between the original and retargeted images, it is difficult to measure the information loss and visual distortion directly.

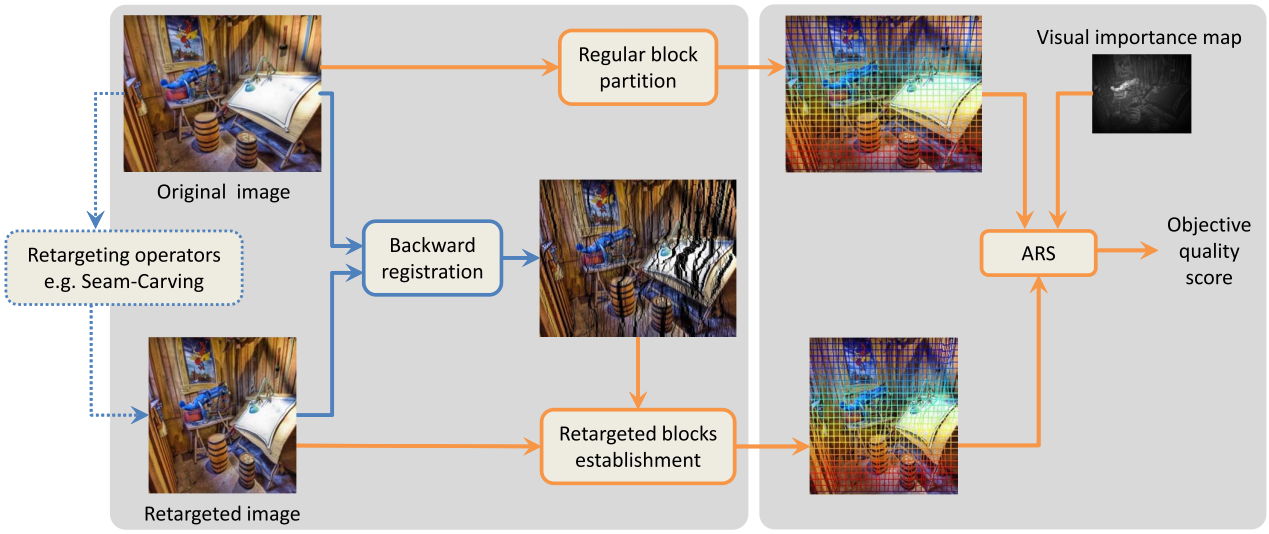


Fig. 1. The framework of the proposed method. The backward registration (blue solid path) aims to reveal the geometric change during the image retargeting and under the guidance of the estimated geometric change, ARS metric (orange solid path) obtains the objective quality score.

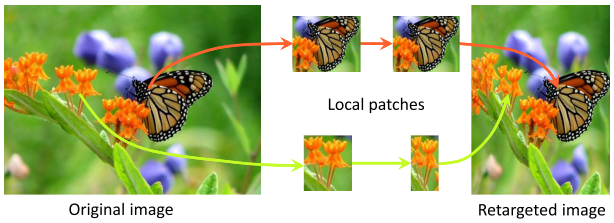


Fig. 2. Illustration of the visual quality change. To assess the quality of the retargeted image, we are eager to know how the image content is modified.

Similar to the study [1], it is crucial to capture the retargeting modification that the image has undergone without specifying the retargeting operator. As shown in Fig. 2, it is helpful for the image retargeting quality assessment to acquire the details about how the butterfly and flowers are changed. We suggest that before the quality evaluation it is necessary to approximate the retargeting modification, so that the image content in the original and retargeted images can be linked to find the corresponding quality degradations.

2) IQA vs IRQA: In the traditional IQA, the reference and distorted images are normally assumed to be well-aligned and the pixel-to-pixel relationship is self-evident. The common image distortions like Gaussian noise, blur, and JPEG compression artifacts are mostly the *intensity* change, and the substantial difference can be obtained by the direct subtraction of two images. In IRQA, it is unlikely to obtain the undergone artificial retargeting modification on images simply by the subtraction. Before the quality evaluation, it becomes necessary to find the relationship between the original and retargeted image in distinctly different resolutions.

When we investigate the retargeted image generation process, we classify the artificial modifications into *intensity* change and *geometric* change.[1] The *intensity* change mostly happens in the continuous retargeting operators like SCL and WARP when the subpixel interpolation is involved. The subpixel sampling may create the intensity fluctuations such as blockiness and aliasing artifacts [2]. The *geometric* change refers to the artificial modification on the image grid to achieve overall resolution alternation. Note, the *geometric* change is dissimilar to the change projected in the real world, like the *geometric* deformation caused by viewpoint disparity [18].

In Fig. 3, we investigate the influence of the *intensity* change and *geometric* change on the process of the retargeted image generation. It includes four groups of retargeted images, the regenerated images, and error maps. The regenerated images are built from original images with the estimated *geometric* change, so there may be the intensity errors between the retargeted and regenerated images as shown in Fig. 3(c). The SSIM [17] quality scores for regenerated images are 1.0000 (CR), 0.9006 (SCL), 0.9366 (SC) and 0.8993 (WARP). It shows that the regenerated images in Fig. 3(b) are generally in high quality compared to reference images from Fig. 3(a). As we can see, with the estimated geometric change it is feasible to regenerate the nearly identical retargeted image from the original image. It indicates that the geometric change plays an important role in the generation of the retargeted image, and the estimation of the geometric change is an efficient way to discover the relationship between the given images. This is the reason that we adopt the geometric change as the guidance for the image retargeting quality assessment in Section III.

B. A Unified Interpretation of Image Retargeting

To estimate the geometric change during the image retargeting, it is necessary to have in-depth understanding of different retargeting operators. Here we formulate different kinds of retargeting operators in a unified interpretation framework.

We use four representative retargeting operators to introduce the unified interpretation of image retargeting, as shown

[1]The geometric change describes how the retargeted image is resampled from the original image with the resampling grid (see details in Section II-B).

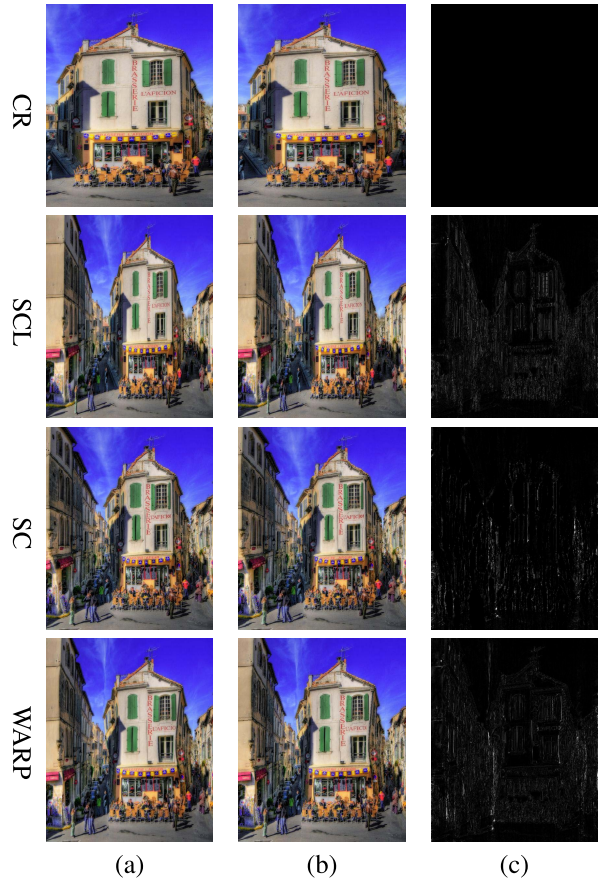


Fig. 3. The influence of the *geometric* change and *intensity* change. (a) Images retargeted by CR, SCL, SC, and WARP. (b) The regenerated retargeted image from original image according to the estimated *geometric* change. (c) The absolute error map between the (a) and (b) images, where the SSIM scores for the regenerated images are 1.0000 (CR), 0.9006 (SCL), 0.9366 (SC) and 0.8993 (WARP).

in Fig. 4. In each model, the top part is the regular retargeted image grid (red nodes), while the bottom parts are the regular original image grid (gray nodes) and the resampling grid in different shapes (blue nodes). The image retargeting is interpreted as the resampling grid generation and forward sampling. The resampling grids are generated in different shapes according to the retargeting operator and the green arrows denote the forward sampling from original image by directly pixel copy or interpolation at subpixel locations, to construct retargeted images in regular grids.

The CR operator is to manually choose an optimal rectangle window and directly copy the image content within the window. Instead of the fixed cropping window, the content-aware discrete methods like SC generalize CR by determining pixels for removal in a seam-wise or pixel-wise way. The reserved pixels are collected together to make up the retargeted image in regular grids. The SCL operator resizes the image in one direction uniformly. It inflates the resampling grid of target resolution onto the original image size to cover all the image content and resamples the image based on the new grid via the interpolation methods like bi-cubic interpolation. The content-aware continuous methods like WARP can be regarded as extensions of SCL, which manipulate the resampling grid

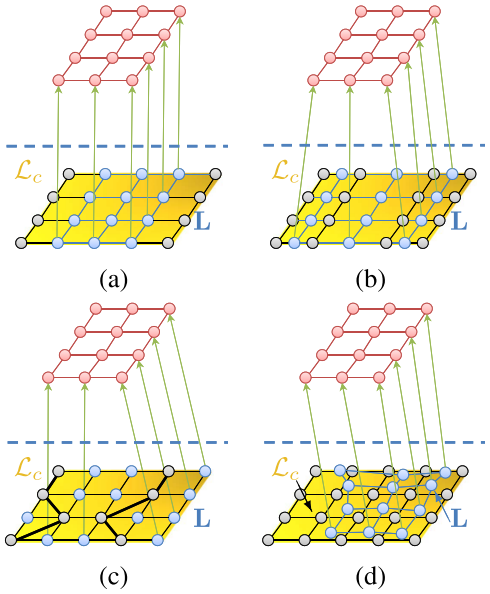


Fig. 4. Grid models. We give the grid models for four typical retargeting operators: (a) Cropping (CR) model; (b) Scaling (SCL) model; (c) Seam-Carving (SC) model (the bold black lines are the removed seams); (d) Non-homogeneous Warping (WARP) model. Note that the blue grid **L** is the resampling grid and the yellow region $\mathcal{L}_c$ is the continuous resampling space in the Section II-C.

consistently in a more sophisticated way by certain energy minimization process.

With these grid models, we are able to interpret the image retargeting with two steps: resampling grid generation and forward resampling. The resampling grids for discrete methods are special cases of the grids for the continuous methods at the sub-pixel level, thus we can adopt the subpixel level resampling grid to model different retargeting operators. Accordingly, if the resampling grid coincides with the original image grid, a.k.a. at pixel level, the forward resampling process is the direct pixel copy; otherwise, the retargeted image pixels are resampled at the sub-pixel location with the adopted interpolation method.

Obviously, we can conclude that the retargeted image is almost determined once the resampling grid is generated. The resampling grid represents the geometric change of the image during the retargeting process and is the strong evidence about the retargeting modification details. As we can observe, the part with dense resampling grid reserves more image content while the zero-density parts can remove the corresponding image region. The sophisticated retargeting operators usually try to preserve visually important content through dense sampling and shrink or even remove other visual unimportant content through sparse or zero sampling. On the other hand, some retargeting operators may destroy the regularity of the resampling grid in the important regions like foreground objects, people, and faces, which would cause the obvious annoying visual artifacts.

C. Backward Registration Problem Formulation

According to the unified interpretation above, we need to estimate the resampling grid to reveal the geometric change

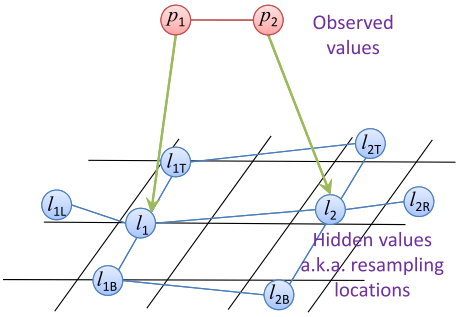


Fig. 5. Markov Random Field (MRF) model for backward registration. The red node p corresponds to the observed pixel in the retargeted image. The blue node l is the resampling location (x, y) from the resampling grid, and the subscripts $\{L, T, B, R\}$ denote the four-connected neighborhood. The aim is to estimate the resampling grid, which is the estimation of the resampling location for each pixel from the retargeted image.

for IRQA. Compared to the forward resampling, to reveal the resampling grid is a reverse problem, and here we formulate it as a *backward registration* problem with Markov Random Field (MRF). Given the original and retargeted image pair $I = \{I_{org}, I_{ret}\}$, the retargeted image grid is chosen as the graph for labelling. We denote $\mathcal{P}$ as the set of pixels in I_{ret}, $\mathbf{L}$ as the resampling grid, and $\mathcal{L}_c$ as the continuous resampling space corresponding to the I_{org}, as shown in Fig. 4. In Fig. 5, assume $p \in \mathcal{P}$ is the observed pixel from the retargeted image and in grid $\mathbf{L}$, each node $l_p = (x_p, y_p) \in \mathcal{L}_c$ is the unknown resampling location for pixel p, where the coordinates are $(x_p, y_p) \in \mathbb{R}$.

Without loss of generality, there are two assumptions about the resampling grid: the uniqueness assumption and ordering assumption. The uniqueness assumption is that each pixel in I_{ret} has one unique resampling point in I_{org}, which is a one-to-one mapping between the pixels in I_{ret} and the resampling grid $\mathbf{L}$. The uniqueness assumption is used to exclude the forward registration as we will discuss in Section II-E. The resampling grid $\mathbf{L}$ following the ordering assumption should be monotonic without any fold, which is reasonable for qualified image retargeting operators, otherwise the image resampled based on the resampling grid would be incomplete. The grid nodes $(x_i, y_j) \in \mathbf{L}$ satisfy the constraints *s.t.* $x_i < x_{i+1}$ and $y_j < y_{j+1}$, where $i, j \in \mathbb{N}$. It indicates the ordering of the resampling location for retargeted image should be same to the pixel ordering in the original image.

We model the resampling grid with the MRF and the backward registration problem can be formulated as the labelling problem for retargeted image pixels in the energy minimization framework as Eq. (1).

$$E = E_d + \lambda E_s \tag{1}$$

The data term E_d describes the likelihood that pixel p is resampled from the location l_p in the original image based on the similarity distance of them

$$E_d = \sum_p d_p(f(p), f(l_p))) \tag{2}$$

where d_p is the distance between the feature $f(p)$ for p and the feature $f(l_p)$ at location (x_p, y_p) may be involved with subpixel interpolations when $(x_p, y_p) \notin \mathbb{N}$.

The smoothness term E_s is corresponding to the ordering assumption in the modelling assumption

$$E_s = \sum_{(p,q)\in\mathcal{N}} V(l_p, l_q) \tag{3}$$

where $\mathcal{N}$ is the local neighborhood in the four-connected retargeted image grid. $V(l_p, l_q)$ measures the cost of assigning two resampling locations l_p and l_q to the neighboring pixels.

The smoothness term is used to achieve the smooth monotonic resampling grid, while the well-designed data term contributes to the accurate backward registration results.

D. Backward Registration Problem Approximation

There are several algorithms such as graph-cut and loopy belief propagation to solve the energy minimization problem [19]. To solve the problem, we first consider the label space for the resampling grid. When $\mathcal{L}_c$ is the continuous resampling space, the space can be very large (say $10^5 - 10^8$) even when we quantize it at the quarter subpixel precision. It means that the complexity of the data term D_s is extremely high. Moreover, the subpixel precision $f(l_p)$ in the I_{org} is always related to the interpolation, which is also time-consuming.

A practical way is to approximate the model by replacing the continuous space $\mathcal{L}_c$ with the pixel-level discrete space $\mathcal{L}_d$. With the consideration that the subpixel error of the resampling grid is quite small compared to the space $\mathcal{L}$, the tradeoff between the subpixel accuracy and fast approximation is worthy as long as the designed IRQA metric is robust to the backward registration errors.

Let $p \in \mathcal{P}$ be the pixel from the retargeted image and in grid $\mathbf{L}$ each node $l_p = (x_p, y_p) \in \mathcal{L}_d$ be the hidden resampling location for pixel p, where the coordinates are $(x_p, y_p) \in \mathbb{N}$.

1) Objective Energy Minimization: The objective function Eq. (1) is defined as:

$$E_d = \sum_p \left\| f(p) - f(l_p) \right\|_1 \tag{4}$$

$$E_s = \sum_{(p,q)\in\epsilon} min(\alpha \left| x_p - x_q \right|, d) + min(\alpha \left| y_p - y_q \right|, d) \tag{5}$$

where the truncated L1-norm distance metric is used in the smoothness term to make the objective function robust to outliers and large discontinuities with d as the threshold. The ϵ is the four-connected neighborhood in the retargeted image grid. We adopt energy function minimization implementation in [20], which is a dual-layer loopy belief propagation based algorithm and utilize a coarse-to-fine scheme to speed up the optimization. The implementation parameters are the default setting $\alpha = 2$, $d = 40$, $\lambda = 1$, $nIterations = 60$ and the adaptive pyramid level $nlevels = \lceil log_2(MAX(W, H)/10) \rceil$ for the original image in $W \times H$ resolution, where $\lceil \cdot \rceil$ is the ceiling operation.

2) A Hybrid Descriptor: The smoothness term mainly guarantees a smooth monotonic resampling grid, while the feature descriptor choice in the data term is important to obtain accurate backward registration results. To get an accurate backward registration results, we develop a hybrid feature

descriptor by using CIE-Lab intensity component, dense SIFT descriptor and the relative position component

$$f_p = [\mathrm{Lab}_{1\times 3}\ \mathrm{SIFT}_{1\times 128}\ \mathrm{Pos}_{1\times 2}] \tag{6}$$

The intensity component is the pixel intensity in CIE-Lab color space. Since the intensity change due to subpixel interpolation is minor, the intensity value is reliable for the likelihood estimation. However, it is not distinctive enough when there are similar pixels in the neighborhood. Recently the dense SIFT descriptor has shown promising performance for matching across scene [20], where the SIFT descriptor is the histogram of the gradient information in the local neighborhood. With the dense SIFT component, the included local neighbor information helps to formulate a distinctive descriptor. With these two components, the feature descriptor might still be ineffective at the smooth or textureless regions. Thus, we also adopt the relative positions of the pixels as the third component. For each image, the pixel coordinates are normalized into the range of $[-1, 1]$. The relative position component complies with the ordering assumption for the resampling grid and has the tendency to inflate the retargeted image grid onto the original size.

E. Forward, Backward, or Bi-Directional Registration

The backward registration results are shown in Fig. 6. In each group, the first row is the retargeted image and the second row is visualized backward registration result by distributing retargeted image pixels onto their resampling locations. We can find that the revealed resampling grids are quite appealing for different retargeting operators. The accuracy evaluation for the backward registration solution is shown in Section IV-B.

The contrary way is the forward registration and obviously we can formulate the model with the original image grid as the grid graph and the retargeted image as the resampling space. The objective function and solution can still be the same. One may wonder whether the forward registration can reveal the geometric change similarly, although it is obviously inconsistent with our unified resampling interpretation and the results are not the resampling grid as well.

In Fig. 7 we provide a comparison sample to show the drawbacks of the forward registration. To better visualize the difference, we choose the CR operator with the consideration of its apparent cropping window. The information loss measurement is perceptual intuitive and also suitable as the measurement for comparison. We create a toy visual importance map as Fig. 7(c) with horizontally different colors. In Fig. 7(e), it is the visualized forward registration results by reconstruction with the retargeted image pixels. There are three parts to be noted: the red rectangle is the chosen cropping window, the gray regions are the pixels which have no appropriate matches in the retargeted image, and the region within the green box represents the pixels that find the wrong matches in the retargeted image. In this case, the data term cannot find any suitable match. The remaining effective smoothness term for these pixels is reversed and we can see the matches in the yellow dash box are mirrored compared to the correct matches on their right side.

To measure the information loss based on the forward registration results, one feasible way is to warp the visual importance map to the retargeted image size and measure importance loss amount compared to the original one. In Fig. 7, (d) and (f) are two warped visual importance maps in the scan-order from top-left to right-bottom, and top-right to left-bottom, respectively. There are obvious holes in the warped importance maps. The reason is that the mismatches for the removed part override the correct matches due to the many-to-one mapping. This will seriously disturb the information loss measurement. In contrast, the backward registration satisfied with the uniqueness assumption can avoid the unnecessary matches. Similarly, in the bi-directional registration, there is the same problem in the forward part and only the backward part is useful. In Fig. 7(g) and (h) the backward registration results reveal the cropping window correctly and based on the estimation we can measure information loss in a reliable way.

III. Aspect Ratio Similarity (ARS) Metric

As discussed in the previous sections, the geometric change can effectively clarify the relationship between the original and retargeted images. The well-estimated geometric change captures the detail information about how images are retargeted. If we partition the original image into regular blocks, with its guidance we can establish the retargeted blocks in the retargeted images and then estimate whether they are well-preserved, removed or distorted. To effectively evaluate the visual quality of the retargeted image, here we design a novel Aspect Ratio Similarity (ARS) metric by measuring the local block changes between the original and retargeted images.

The purpose of image retargeting is actually the Aspect Ratio (AR) alternation for the entire image. The global AR change is achieved with the local AR changes collaboratively across the image. The AR changes of the local blocks indicate how they are preserved in the retargeted image, based on which we can measure their information loss and visual distortions. Therefore, the visual quality of retargeted images can be assessed by measuring the local AR changes.

As shown in Fig. 1, the original image is divided into $N \times N$ regular blocks and their corresponding retargeted blocks are established in retargeted image under the geometric change guidance. More results are shown as the third row in Fig. 6. We calculate the maximal width w_{ret} and maximal height h_{ret} of each retargeted block. The height and width change ratios can be denoted as $r_w = w_{ret}/N$ and $r_h = h_{ret}/N$, and the mean ratio $u_r = (r_w + r_h)/2$ denotes the absolute block size changes. The similarity score of the block pair is formulated as follows:

$$S = \left[\frac{2 \cdot r_w \cdot r_h + C}{r_w^2 + r_h^2 + C}\right] \cdot \left[e^{-\alpha(\mu_r - 1)^2}\right] \tag{7}$$

where C is a small positive constant to increase the stability by avoiding the division by zero and the α is the parameter larger than zero to balance the information loss and visual distortions by adjusting the information loss penalty degree.

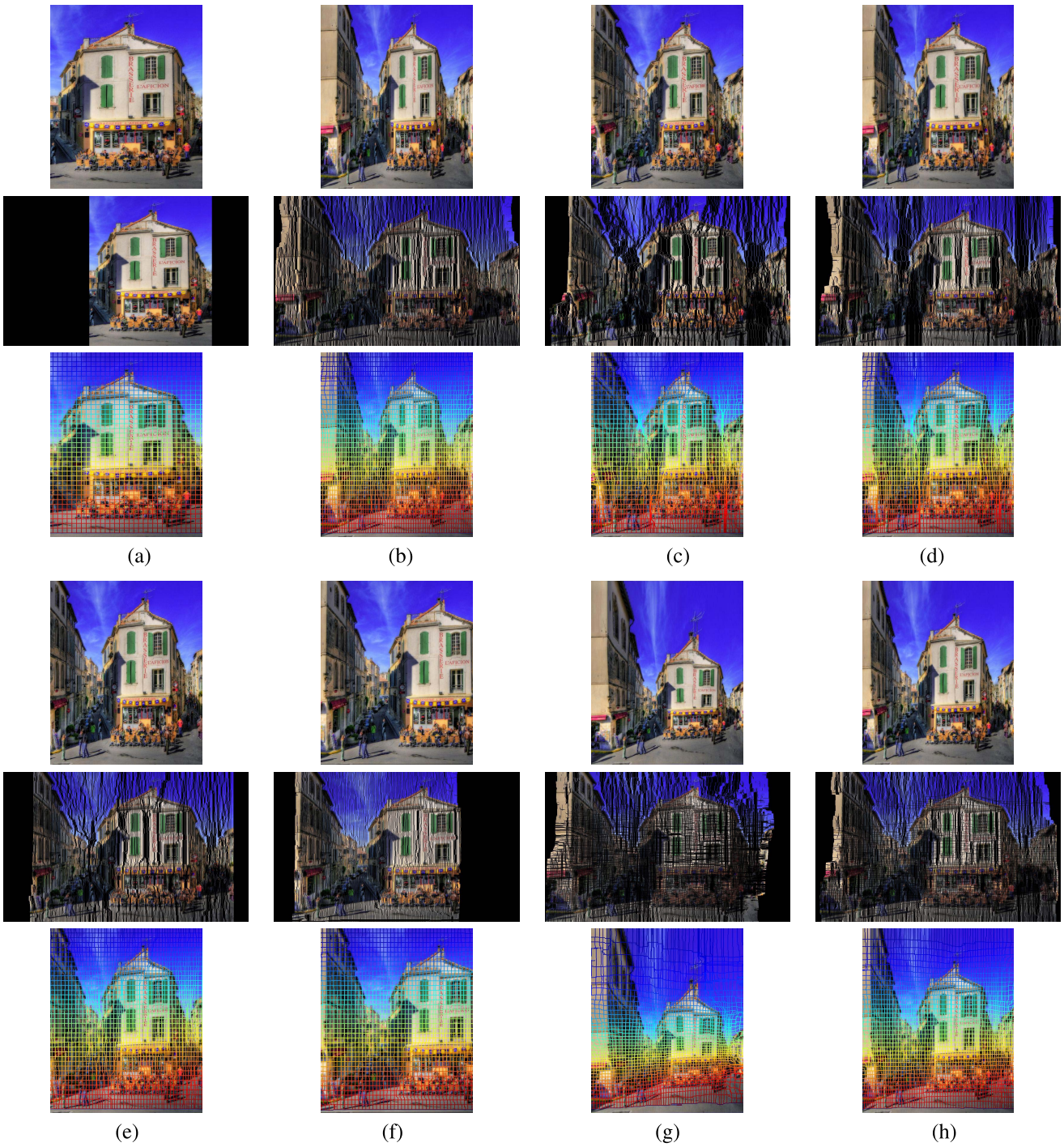


Fig. 6. Backward registration results. We provide the results for images retargeted by eight retargeting operators. In each group, the first row is the retargeted image; the second row is the backward registration result, which reveals the geometric change; in the third row under the guidance of the geometric change, we establish the retargeted blocks corresponding to the regular blocks in the original image. (a) CR. (b) SCL. (c) SC. (d) WARP. (e) MOP. (f) SM. (g) SNS. (h) SV.

In Eq. (7), the first Aspect Ratio (AR) term measures the aspect ratio similarity between the block pairs. Since the aspect ratio loses the absolute size change information, in the second Absolute Size (AS) term we utilize the Gaussian function of the mean ratio of u_r to take into account the absolute size change influence. When S is close to 1, the block content in the original image is generally preserved in high quality, while when S is close or equal to zero, it indicates that the retargeted block is suffering from serious information loss or visual distortions.

To predict the visual quality of the whole retargeted image, we utilize the saliency detection method [21] specifically designed for image retargeting to calculate the saliency map as our visual importance map. The visual quality score ARS for the retargeted image is defined as Eq. (8) by pooling the similarity score of each block with the corresponding

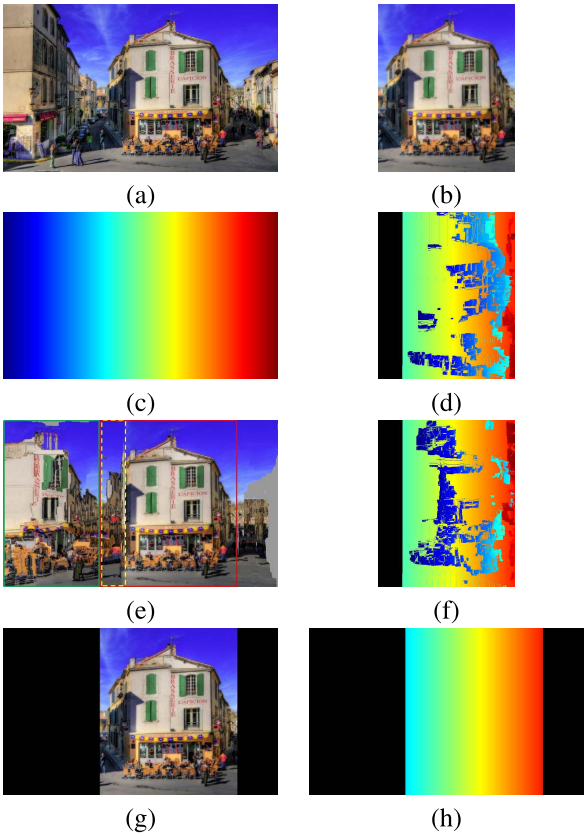


Fig. 7. A comparison of *forward* and *backward* registration with the information loss measurement. (a) The original image. (b) The retargeted image by CR with 50% width reduction. (c) A toy visual importance map. (e) The forward registration results. (d) and (f) are two warped visual importance maps based on forward registration results. (g) The backward registration results. (h) The reserved visual importance map based on backward registration results.

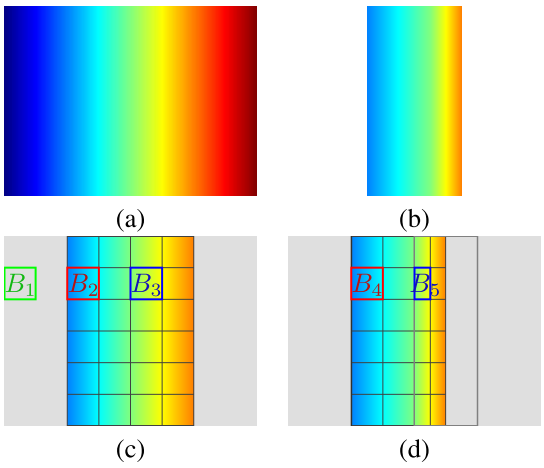


Fig. 8. A toy example for ARS metric. (a) Original image (512 × 384). (b) Retargeted image (192 × 384). We crop out the middle half of the original image as (c) and then the scale the right half by 50% width reduction as (d). The original image is divided into 64 × 64 regular non-overlapped blocks as (c) and the corresponding retargeted blocks are shown in (d). The block B_1 is totally removed, B_2 is well-preserved as B_4, and B_3 is scaled to B_5.

TABLE I
BENCHMARK DATASETS FOR IRQA EVALUATION

Dataset	MIT	CUHK
Source Image No.	37	57
Retargeted Image No.	296	171
Retargeted Image -25% No.	184	46
Retargeted Image -50% No.	112	125
Retargeting Operator No.	8	10
Subject No.	210	64
Subjective Score Type	Pair-wise	MOS

importance weight.

$$ARS = \sum_m \sum_n S_{mn} \cdot V_{mn} \Big/ \sum_m \sum_n V_{mn} \tag{8}$$

where S_{mn} represents similarity score S for block (m, n), V_{mn} denotes the sum of the visual importance value within the block (m, n), where the visual importance map is partitioned regularly in the same way as the original image; m and n are the block indices in the original image.

To obtain more intuitive understanding of ARS metric, we provide one toy example to show the quality measurement for different types of block changes in Fig. 8. The original toy image (a) is retargeted to image (b) via the combination of CR and SCL. We investigate the typical well-preserved, totally removed, and horizontal uniformly scaled blocks. The block B_4 is well-preserved from the original image, and thus both AR and AS terms are equal to 1. As we all know, if one image is uniformly down-scaled in both width and height, there should be little visual artifact. When we uniformly down-scale the local blocks, AR term is always equal to 1 and AS term takes account of the overall size changes for information loss. When it comes to the extreme case, e.g. totally removed block B_1, AR term is still equal to 1 while AS term reduces to the minimal $e^{-\alpha}$. For the horizontal scaling case from B_3 (64 × 64) to B_5 (64 × 32), AR term becomes to 0.8 and AS term reduces to $e^{-0.06\alpha}$. We can see that AR and AS terms are used to evaluate the information loss and visual distortion for each block collaboratively.

IV. EXPERIMENTAL RESULTS

A. Image Retargeting Datasets

In the experiments, we adopt two widely used public datasets: MIT RetargetMe dataset [22] and CUHK dataset [23], [24] to evaluate the performance of the proposed objective quality metric, in terms of the correlation between the objective scores and the subjective scores provided in the datasets. The detailed information of these datasets are summarized in Table I.

In MIT dataset, there are 37 source images and their retargeted images are generated by eight retargeting operators including CR, SCL, SC [3], MO [25], SM [4], SNS [7], SV [6], and WARP [5] with 25% (23P) or 50% (14P) of width or height reduction. There are six major image attributes provided for better insights: *Line/Edge, Face/People, Foreground Objects, Texture, Geometric Structures* and *Symmetry*, and each image may own more than one attribute. The subjective tests are conducted in the way of the paired comparison, where subjects choose the better one from two retargeted images

shown in juxtaposition. The dataset provides the corresponding numbers of times that the retargeted image is favored over others as the subjective scores. The correlations between the objective and subjective scores are measured by the Kendall Rank Correlation Coefficient (KRCC) [1]:

$$\mathrm{KRCC} = \frac{n_c - n_d}{0.5n(n-1)} \tag{9}$$

where n is the length of the ranking (here $n = 8$), n_c is the number of concordant pairs and n_d is the number of discordant pairs from all the pairs.

The CUHK dataset contains 171 retargeted images generated from 57 natural source images. Compared to MIT dataset, two other retargeting operators, optimized seam-carving and scale (SCSC) [26] and energy-based deformation (ENER) [27], are included as well. Each original image can be retargeted by either 25% or 50% of the width or height reduction. Different from the pair-wise comparison in MIT dataset, the subjective tests employ 5 category quality scales as "Bad", "Poor","Fair", "Good", and "Excellent", and generate the Mean Opinion Score (MOS) for each retargeted image, which serves for the correlation evaluation between subjective MOSs and the objective scores like the traditional subjective testing in IQA [15], [17].

In CUHK dataset, four commonly used evaluation metrics are employed to evaluate the relationship of the objective quality metric scores and the provided MOSs. The first one is the Pearson Linear Correlation Coefficient (LCC) between MOS and the objective scores after nonlinear regression. The nonlinear regression is the mapping function suggested by Sheikh *et al.* [28]:

$$f(x) = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\beta_2 (x - \beta_3)}} \right) + \beta_4 x + \beta_5 \tag{10}$$

The second metric is the Spearman Rank-order Correlation Coefficient (SRCC), which can measure the prediction monotonicity of the objective IRQA metric scores. The third metric is the Root Mean Squared Error (RMSE) between MOS and the objective scores after nonlinear regression. The fourth metric, the Outlier Ratio (OR) [29] is the ratio of the false objective score number to the total score number. The false score means that the score lies outside the interval [MOS-2σ, MOS+2σ] after the nonlinear regression, where σ is the corresponding standard deviation. As we can see, larger LCC and SRCC values indicate that the objective metric scores correlate better with the subjective MOSs. In other words, the performance of the objective metric is better. The other RMSE and OR metrics correspond to the correlation errors, thus the smaller values indicate the better objective quality metrics.

B. Accuracy Evaluation of Backward Registration

1) Dataset With the Groundtruth: To evaluate the accuracy of the geometric change estimation by backward registration, it is necessary to provide a dataset with the retargeting modification as the ground-truth information. However, due to limited number of retargeting operators that provide the source codes, here we only choose the Seam-Carving [30] and Scale-and-Stretch (SNS)[2] [7] to build the validation datasets as shown in Fig. 9. We use source images from both MIT and CUHK dataset. In MIT part, we use the 37 source images from MIT dataset [22] and retarget the images into similar target resolutions. In CUHK part, we choose the 57 source images and build all the images by retargeting each image with 50% width reduction. In Fig. 9(c), the black regions represent the removed seams by SC. With the binary map we can reconstruct the identical retargeted image by removing the pixels in black regions. In Fig. 9(e) it is the approximated ground-truth map for SNS. Since the ground-truth information for the continuous image retargeting operators is the resampling grid, we brutally transform the resampling grid into the binary map by approximating each resampling point to its nearest-neighbor pixel for the visualization.

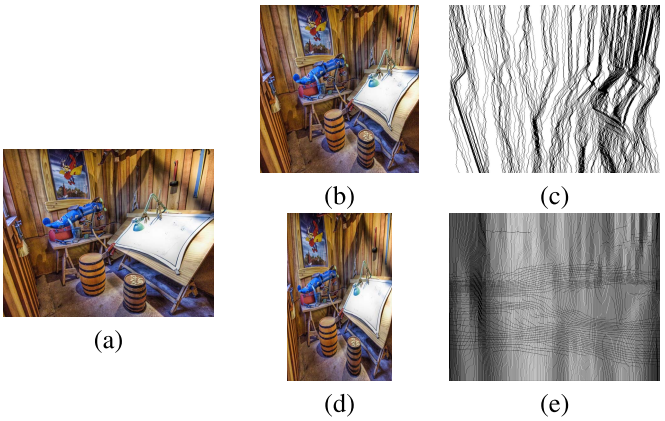


Fig. 9. One sample from the validation datasets. (a) The original image. (b) The image retargeted by SC with 25% width reduction. (c) The ground-truth for (b): a binary map indicating the removed pixels (black regions are the removed seams). (d) The image retargeted by SNS with 50% width reduction. (e) The approximated ground-truth map for (d).

2) Backward Registration Accuracy: The performance evaluation of the backward registration results on the validation datasets is shown in Table II. The performances of different kinds of descriptors including SIFT, Lab, SIFT+Lab and the proposed descriptor are evaluated on SC and SNS sets. According to the unified resampling framework in Section II-B, the backward registration accuracy can be effectively evaluated by measuring the similarity between ground-truth and the estimated resampling grids. Here we adopt the Mean Absolute Error (MAE) as the main accuracy measure:

$$\mathrm{MAE} = \frac{1}{N} \sum_{i}^{N} \left\| (x_i, y_i) - (x_i', y_i') \right\|_1 \tag{11}$$

where (x_i, y_i) is the ground-truth resampling point and (x_i', y_i') is the estimated resampling point; N is the total number of pixels in the retargeted image.

The discrete SC can be regarded as a special case of the continuous retargeting operators. For images retargeted by SC, the accuracy can be measured by whether we have

[2]To obtain the ground-truth information, we implement the SNS in MATLAB and the code is public available at yabinzhangjohn.github.io.

TABLE II
PERFORMANCE OF BACKWARD REGISTRATION ON VALIDATION DATASET

Set	Descriptor	MIT part				CUHK part			
		Rc	Pr	R_o	*MAE*	Rc	Pr	R_o	*MAE*
SC	SIFT	0.7831	0.8023	0.034	14.231	0.6277	0.6769	0.077	38.009
	Lab	0.8045	0.8175	0.022	11.383	0.6767	0.6996	0.034	20.914
	SIFT+Lab	0.8164	0.8322	0.024	7.977	0.7075	0.7367	0.040	15.333
	Our descriptor	**0.8257**	**0.8336**	**0.014**	**4.228**	**0.7304**	**0.7410**	**0.014**	**6.425**
SNS	SIFT	0.6306	0.6524	-0.043	25.856	0.4838	0.5279	-0.090	60.271
	Lab	**0.7785**	**0.7806**	-0.003	8.816	**0.6542**	**0.6621**	-0.013	22.605
	SIFT+Lab	0.6905	0.6952	-0.008	7.058	0.5695	0.5825	-0.023	16.192
	Our descriptor	0.6947	0.6949	**-0.001**	**3.723**	0.5806	0.5812	**-0.001**	**7.888**

TABLE III
PERFORMANCE OF DIFFERENT METRICS ON MIT DATASET

Metric	mean KRCC in each subset						Total			
	Line Edge	*Faces People*	*Foreground Objects*	*Texture*	*Geometric Structure*	*Symmetry*	mean KRCC	std KRCC	LCC	p-val
BDS	0.040	0.190	0.167	0.060	-0.004	-0.012	0.083	0.268	0.134	0.017
EH	0.043	-0.076	-0.079	-0.060	0.103	0.298	0.004	0.334	-0.033	0.641
SIFT flow	0.097	0.252	0.218	0.161	0.085	0.071	0.145	0.262	0.227	0.031
EMD	0.220	0.262	0.226	0.107	0.237	0.500	0.251	0.272	0.274	1e-5
CSim	0.097	0.290	0.293	0.161	0.053	0.150	0.164	0.263	0.242	0.028
IR-SSIM	0.309	0.452	0.377	0.321	0.313	0.333	0.363	0.271	0.439	1e-3
PGDIL	0.431	0.390	0.389	0.286	0.438	**0.523**	0.415	0.296	0.468	6e-10
Proposed ARS	**0.463**	**0.519**	**0.444**	**0.330**	**0.505**	0.464	**0.452**	**0.283**	**0.567**	**1e-11**

find the removed pixels correctly, so we utilize the Recall (Rc), Precision (Pr), and the overlapped ratio R_o as the additional accuracy measurement. Here, R_o refers to the percentage of pixels in retargeted image with same resampling locations as other pixels, and in the ideal situation R_o should be equal to zero. We extend these metrics to evaluate the accuracy on images retargeted by SNS. To make it feasible to measure Rc, Pr and R_o, we take the brutally approximated binary maps instead of the ground-truth resampling grids for SNS set.

In Table II the experimental results show that our descriptor achieves the best overall performance. The MAE results show that our descriptor can estimate the resampling grid more accurately. For the Rc, Pr, and R_o on SC validation set, the results show the hybrid descriptor is more advantageous. On SNS validation set, the Lab descriptor shows better Recall and Precision performance. The reason is that the Lab descriptor, without neighborhood information, makes the backward registration a too greedy search strategy for the warping cases with serious deformed resampling grids. As we can see, the MAE for Lab descriptor is still quite large. Since the images in CUHK part are all retargeted with 50% width reductions, where the discontinuities and deformations in retargeted images are more serious, the overall performance is relatively lower compared to MIT part. Although the single SIFT descriptor or Lab descriptor is capable to reveal the resampling grid, the proposed hybrid descriptor shows significantly better performance. The overall performance of the backward registration is satisfactory, which is consistent with the results in Fig. 6, and with its guidance different IRQA metrics can be further developed.

C. Performance Evaluation of ARS on the MIT Dataset

To demonstrate the effectiveness of the proposed ARS metric, we present rank correlation results on the benchmark MIT dataset RetargetMe [22] in Table III. The proposed ARS is compared with BDS [31], EH [32], SIFT flow [20], EMD [33], CSim [10], IR-SSIM [12] and PGDIL [8]. We have adopted the same saliency detection method [21] as the IR-SSIM and PGDIL.

We give the mean and standard deviation values of the rank correlations as well as p-value and linear correlation-coefficient (LCC) in Table III. From this table, we can see that the proposed ARS can obtain statistically better performance than the state-of-the-art methods and in image subsets labelled by different attributes except *Symmetry*, the proposed metric achieves the best correlation results among the compared methods. The relative lower performance in *Symmetry* attribute subset shows the limitation of ARS in the measurement of the global and semantic structures distortions.

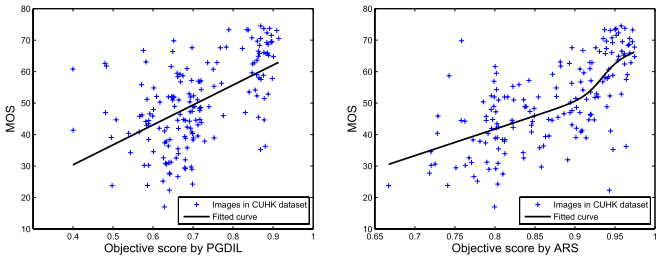
D. Performance Evaluation of ARS on the CUHK Dataset

In Table IV, we compare the performance of the proposed ARS with with BDS [31], EH [32], SIFT flow [20], EMD [33], CSim [10], GLS [13] and PGDIL [8] on CUHK dataset. The performance is measured by LCC, SRCC, RMSE and OR. As shown in Table IV, the proposed ARS performs consistently better than other metrics significantly.

In Fig. 10, we show the scatter plots of subjective scores against objective scores predicted by PGDIL and ARS on CUHK dataset. The curves in Fig. 10 were fitted by the

TABLE IV
PERFORMANCE OF DIFFERENT METRICS ON CUHK DATASET

Metric	LCC	SRCC	RMSE	OR
BDS	0.2896	0.2887	12.922	0.2164
EH	0.3422	0.3288	12.686	0.2047
SIFT flow	0.3141	0.2899	12.817	0.1462
EMD	0.2760	0.2904	12.977	0.1696
CSim	0.4374	0.4662	12.141	0.1520
GLS	0.4622	0.4760	10.932	0.1345
PGDIL	0.5403	0.5409	11.361	0.1520
Proposed ARS	**0.6835**	**0.6693**	**9.855**	**0.0702**



(a) (b)

Fig. 10. Scatter plots of subjective MOS against objective scores obtained by model prediction on CUHK dataset. (a) The PGDIL. (b) Proposed ARS.

nonlinear regression using Eq. (10). As we can see, objective scores predicted by ARS are more correlated with subjective scores compared to the state-of-the-art PGDIL scores.

V. CONCLUSION AND FUTURE WORK

To better understand the image retargeting, we have first provided a unified interpretation of image retargeting with the resampling grid generation and forward resampling. We shown that the geometric change estimation is an efficient way to clarify the relationship between the original and retargeted images. We formulated the geometric change estimation as a backward registration problem via MRF and provided an effective and practical solution. Under the guidance of the geometric change, we have developed an effective ARS metric by exploiting the local block changes to evaluate the visual quality of retargeted images. Compared to other existing IRQA metrics on the publicly available MIT and CUHK datasets, the proposed ARS yields statistically better results in the prediction accuracy.

The proposed ARS metric shows the success by measuring the local block changes under the geometric change guidance. Even though ARS metric is designed based on the local low-level features and it is infeasible to explicitly evaluate the specific attribute degradations such as the broken line and violated symmetry, it still outperforms other existing metrics. Some recent studies have already investigated the influence of specific attributes like *Line*, *Symmetry* and *Aesthetics* both for image retargeting and its quality assessment [6], [34]–[37]. In the future work, we will explore additional detection techniques for attributes like *Line/Edge* and *Symmetry*. With the geometric change we can establish the attribute correspondence between the original and retargeted images. The explicitly evaluation of corresponding attribute quality degradation may help develop better IRQA metrics.

REFERENCES

[1] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir, "A comparative study of image retargeting," *ACM Trans. Graph.*, vol. 29, no. 5, pp. 160:1–160:10, 2010.
[2] A. Shamir and O. Sorkine, "Visual media retargeting," in *Proc. SIGGRAPH ASIA*, Yokohama, Japan, Dec. 2009, Art. no. 11.
[3] M. Rubinstein, A. Shamir, and S. Avidan, "Improved seam carving for video retargeting," *ACM Trans. Graph.*, vol. 27, no. 3, 2008, Art. no. 16.
[4] Y. Pritch, E. Kav-Venaki, and S. Peleg, "Shift-map image editing," in *Proc. IEEE 12th Int. Conf. Comput. Vis., (ICCV)*, Kyoto, Japan, Sep./Oct. 2009, pp. 151–158.
[5] L. Wolf, M. Guttmann, and D. Cohen-Or, "Non-homogeneous content-driven video-retargeting," in *Proc. IEEE 11th Int. Conf. Comput. Vis. (ICCV)*, Rio de Janeiro, Brazil, Oct. 2007, pp. 1–6.
[6] P. Krähenbühl, M. Lang, A. Hornung, and M. H. Gross, "A system for retargeting of streaming video," *ACM Trans. Graph.*, vol. 28, no. 5, 2009, Art. no. 126.
[7] 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.
[8] 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.
[9] T. Ren and G. Wu, "Automatic image retargeting evaluation based on user perception," in *Proc. Int. Conf. Image Process. (ICIP)*, Hong Kong, Sep. 2010, pp. 1569–1572.
[10] 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.
[11] L. Ma, L. Xu, H. Zeng, K. N. Ngan, and C. Deng, "How does the shape descriptor measure the perceptual quality of the retargeting image?" in *Proc. IEEE Int. Conf. Multimedia Expo Workshops*, Chengdu, China, Jul. 2014, pp. 1–6.
[12] Y. Fang, K. Zeng, Z. Wang, W. Lin, Z. Fang, and C. 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.
[13] J. Zhang and C.-C. J. Kuo, "An objective quality of experience (QoE) assessment index for retargeted images," in *Proc. ACM Int. Conf. Multimedia (MM)*, Orlando, FL, USA, Nov. 2014, pp. 257–266.
[14] 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.
[15] W. Lin and C.-C. J. Kuo, "Perceptual visual quality metrics: A survey," *J. Vis. Commun. Image Represent.*, vol. 22, no. 4, pp. 297–312, 2011.
[16] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, Jan. 2009.
[17] 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.
[18] R. Szeliski, "Image alignment and stitching: A tutorial," *Found. Trends Comput. Graph. Vis.*, vol. 2, no. 1, pp. 1–104, 2006.
[19] R. Szeliski *et al.*, "A comparative study of energy minimization methods for Markov random fields with smoothness-based priors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 6, pp. 1068–1080, Jun. 2008.
[20] C. Liu, J. Yuen, and A. Torralba, "SIFT flow: Dense correspondence across scenes and its applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 978–994, May 2011.
[21] Y. Fang, Z. Chen, W. Lin, and C.-W. Lin, "Saliency detection in the compressed domain for adaptive image retargeting," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 3888–3901, Sep. 2012.
[22] *RetargetMe Benchmark*, acessed on (Jul. 07, 2016). [Online]. Available: http://people.csail.mit.edu/mrub/retargetme
[23] 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.

[24] *Image Retargeting Subjective Quality Database*, acessed on (Jul. 07, 2016). [Online]. Available: http://ivp.ee.cuhk.edu.hk/projects/demo/retargeting/index.html
[25] M. Rubinstein, A. Shamir, and S. Avidan, "Multi-operator media retargeting," *ACM Trans. Graph.*, vol. 28, no. 3, 2009, Art. no. 23.
[26] 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, 2009, Art. no. 125.
[27] Z. Karni, D. Freedman, and C. Gotsman, "Energy-based image deformation," *Comput. Graph. Forum*, vol. 28, no. 5, pp. 1257–1268, 2009.
[28] 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.
[29] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Process., Image Commun.*, vol. 19, no. 2, pp. 121–132, 2004.
[30] *Seam Carving code*. acessed on (Jul. 07, 2016). [Online]. Available: http://people.csail.mit.edu/mrub/code/seam_carving-1.0.zip
[31] D. Simakov, Y. Caspi, E. Shechtman, and M. Irani, "Summarizing visual data using bidirectional similarity," in *Proc. IEEE Comput. Soc. Conf. (CVPR)*, Anchorage, AK, USA, Jun. 2008, pp. 1–8.
[32] 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.
[33] O. Pele and M. Werman, "Fast and robust earth mover's distances," in *Proc. IEEE 12th Int. Conf. Comput. Vis. (ICCV)*, Kyoto, Japan, Sep./Oct. 2009, pp. 460–467.
[34] H. Wu, Y.-S. Wang, K.-C. Feng, T.-S. Wong, T.-Y. Lee, and P.-A. Heng, "Resizing by symmetry-summarization," *ACM Trans. Graph.*, vol. 29, no. 6, 2010, Art. no. 159.
[35] F.-L. Zhang, M. Wang, and S.-M. Hu, "Aesthetic image enhancement by dependence-aware object recomposition," *IEEE Trans. Multimedia*, vol. 15, no. 7, pp. 1480–1490, Nov. 2013.
[36] C.-H. Chang and Y.-Y. Chuang, "A line-structure-preserving approach to image resizing," in *Proc. IEEE CVPR*, Providence, RI, USA, Jun. 2012, pp. 1075–1082.
[37] Y. Liang, Y.-J. Liu, and D. Gutierrez, "Objective quality prediction of image retargeting algorithms," *IEEE Trans. Vis. Comput. Graphics*, vol. PP, no. 99, Jan. 2016.

Yabin Zhang received the B.E. degree in electronic information engineering from the Honors School, Harbin Institute of Technology in 2013. He is currently pursuing the Ph.D. degree from the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His research interests include video coding, image/video processing, and computer vision.

Yuming Fang received the Ph.D. degree from Nanyang Technological University, Singapore, the M.S. degree from the Beijing University of Technology, China, and the B.E. degree from Sichuan University, China. He is an Associate Professor with the School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, China. He was a Visiting Researcher with the IRCCyN Laboratory, PolyTech' Nantes, University of Nantes, France, National Tsinghua University, Taiwan, and the University of Waterloo, Canada. His research interests include visual attention modeling, visual quality assessment, image retargeting, computer vision, and 3D image/video processing.

Weisi Lin (M'92–SM'98–F'16) received the B.Sc. degree in electronics, the M.Sc. degree in digital signal processing from Sun Yat-Sen University, Guangzhou, China, in 1982 and 1985, respectively, and the Ph.D. degree in computer vision from Kings College, London University, London, U.K., in 1993. He was involved in teaching and research with Sun Yat-Sen University, Shantou University, Shantou, China, Bath University, Bath, U.K., the National University of Singapore, the Institute of Microelectronics, Singapore, and the Institute for Infocomm Research, Singapore. He was the Laboratory Head of Visual Processing Institute for Infocomm Research. He is currently an Associate Professor with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His current research interests include image processing, perceptual modeling, video compression, and multimedia communication.

Xinfeng Zhang (M'16) received the B.S. degree in computer science from the Hebei University of Technology, Tianjin, China, in 2007, and the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2014. He is currently a Research Fellow with Nanyang Technological University, Singapore. His research interests include image and video processing, and image and video compression.

Leida Li received the B.S. and Ph.D. degrees from Xidian University, Xi'an, China, in 2004 and 2009, respectively. In 2008, he was a Visiting Ph.D. Student with the Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, Taiwan. From 2014 to 2015, he was a Visiting Research Fellow with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. He is currently a Professor with the School of Information and Electrical Engineering, China University of Mining and Technology, China. His research interests include multimedia quality assessment, information hiding, and image forensics.