Reduced-Reference Structural Quality Assessment for Retargeted Images

Chau-Wai Wong, *Student Member, IEEE,* Wenjun Lu, *Member, IEEE,* and Min Wu, *Fellow, IEEE*

***Abstract*—Recent years have witnessed tremendous growth in the generation and consumption of digital images. Monitoring and evaluating image quality is an important issue for online and mobile media applications. Conventional quality assessment work mostly focus on intensity level distortion caused by operations that do not change image aspect ratio/size, such as distortion caused by compression, noise, and blurring. Here, we study the problem of quality assessment for images undergone content-adaptive resizing, also known as retargeting operations. Retargeting is an increasingly popular technique for rendering images to screens with different aspect ratios, and the dominant distortion in question is mostly on the structural and content level rather than on the intensity or signal level. Quality assessment on image structural distortion is not as well studied as the quality assessment on intensity level distortions. In this work, we design a reduced-reference corner-point-matching-based framework to analyze the structural distortion caused by retargeting and propose a set of quality scores and fuse them to achieve positive correlation with human observations. The proposed quality scores by each individually achieve comparable performance with the best known full-reference quality metric; and the fused score has a better performance than this best quality metric.**

***Index Terms*—Image quality assessment, reduced reference, image retargeting.**

I. Introduction

The goal of automatic image quality assessment is to design objective quality metrics that correlate well with human subjectives. Such automated quality metrics help enhance the efficiency and objectivity of human evaluations on image quality degradations that might happen during image distribution. Depending on the amount of available information of the original image, image quality assessment can be classified into full reference (the reference image is fully available), reduced reference (only compact partial information about the reference image is available), and no reference (no access to the reference image is allowed). A comprehensive survey of image quality assessment techniques can be found in [1].

A fundamental way of assessing image quality is by subjective evaluation involving human observers. The mean opinion score (MOS) is one of the commonly used and well regarded subjective measures for image quality assessment. However, involving human observers can be expensive and not scalable for many practical applications. Therefore, the goal of objective image quality assessment is to design computational models such that an estimated quality by the models correlates well with human subjectivity. The simplest and most widely used full reference quality metric is the mean squared error (MSE) and related metric of peak signal-to-noise ratio (PSNR). Although MSE and PSNR are easy to be calculated, their results are not well matched to the perceived visual quality of human beings. Taking advantage of known characteristics of human visual system (HVS), most state-of-the-art quality assessment works have adopted a two-stage structure, namely, local distortion measure and spatial pooling to obtain a final quality score. Some representative local distortion measures include structure similarity (SSIM) index [2], [3], block discrete cosine transform [4] and wavelet-based approaches [5], [6]. These local quality measures are then pooled together to maximize the correlation between objective and subjective image quality ratings [7]–[10].

Most existing image quality assessment work focus on distortion caused by operations such as compression, additive noise, blurring, and contrast/brightness changes. These types of distortions mainly alter the intensity and noise level of the image, but do not change the image size and the content structures within the image. In this case, exact pixel-level correspondences can be established to compute the distortion. However, the pixel-level distortion does not adequately model the perceived change of an image to human beings if the editings to the image are mainly structural, and hence it is intuitive to quantify the distortion caused by structural changes explicitly. We model the *structural distortion* based on the displacement, addition, and removal of features points on structural objects such as contours and shapes, which is different from the perspective taken by the well-known SSIM family metric [2], [3], [11] that considers interdependencies among pixels, regions, and features as "structural" information and often quantifies them through statistics from pixel/feature domain. The structural distortion that we focus on in this work is receiving an increasing amount of attention because of the image retargeting technique that mainly leads to structural distortion has been increasingly popular to resize images for better viewing experience on smaller screen sizes such as mobile phones. Below, we briefly discuss related image retargeting work and the challenges of quality assessment on retargeted images.

Image retargeting is a class of techniques that provide content-adaptive image resizing to facilitate viewing images on screens of different sizes. Image retargeting methods can be roughly classified into discrete versus continuous [12]. Discrete approaches remove or insert unimportant pixels or

Chau-Wai Wong and Min Wu are with the Department of Electrical and Computer Engineering, and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742, USA (email: {cwwong,minwu}@umd.edu).

Wenjun Lu was with the Department of Electrical and 1Computer Engineering, University of Maryland, College Park, MD 20742, USA. He is now with Google Inc. (email: luwj02@gmail.com).

patches from the interior of the images [13]–[15]. One representative retargeting algorithm is the seam carving [13], which adaptively removes "seams", sequences of 8-connected pixels with low energy values defined by some energy function, from an original image. The resulting image keeps the most salient content intact while achieving the desired aspect ratio and size. Although the results are promising, visible discontinuities or aliasing artifacts are often noticeable at large aspect ratio changes. On the other hand, continuous retargeting techniques optimize a mapping from the source media size to the target size without explicit content removal. The key idea of these approaches is to scale visually important feature regions uniformly while allowing arbitrary deformations in unimportant regions of the image. Some representative works include [16]–[18]. A comprehensive comparison of state-of-the-art image retargeting techniques can be found in a recent work by Rubinstein et al. [19]. Most retargeting algorithms try to minimize some objective functions that measure the distortion between the original image and the retargeted ones. In Rubinstein et al.'s [19] work on comparing different retargeting algorithms, the authors compared objective functions such as bidirectional similarity (BDS) [20], bidirectional warping (BDW) [15], and earth mover's distance (EMD) [21] with human subjective ratings. These three objective functions can serve as full-reference metrics for quality assessment on retargeted images. SIFT flow (SIFTflow) [22] can also be regarded as a full-reference quality assessment metric since it generates the hash by calculating the 128-d SIFT descriptor for every pixel of the image.

In this work, we study the problem of quality assessment for retargeted images, under the assumption that only limited side information is available, i.e., a reduced-reference scenario. Given that the original reference image is not available in many real-world applications, reduced-reference quality assessment will be a more feasible alternative, where the compact partial information can be embedded into the header file of a potentially retargeted image and be used to monitor and assess image quality. The problem itself is interesting as it helps us understand what specific features in the image content that actually capture the structural distortion well as perceived by humans. We notice a parallel work [23] which focuses on the quality assessment in the full-reference scenario which is a much stringent scenario than ours, and yet its experimental methodology is not fully compatible with Rubinstein et al.'s work [19] that we heavily base on, hence we leave possible comparisons to further work.

We approach the problem of reduced-reference quality assessment of retargeted images by looking at strong corner points on salient image edges. Corner points can be compactly represented and encoded as a binary map, whereas provide good amount of information on content structure of the image. The challenge lies in given a small set of corner points, i) how to identify structural distortion between the original image and the retargeted image, and ii) how to quantify the structural distortion into scores that are consistent with human subjective ratings. We utilize the shape context work from Belongie et al. [24] to find correspondences between the corner points from the original image and those in the retargeted image. Such correspondences reveal a distortion vector field where each distortion vector captures how a corner point has been moved relative to the other corner points. With the distortion vector field, we are able to capture the structural distortion of the retargeted image using carefully designed metrics: metrics based on parametric affine transform models, metrics derived from saliency-weighted inconsistency between local neighborhoods, and etc. Our proposed quality metrics show comparable performances as and fused metric even outperforms the best known full-reference metrics for the retargeting operation.

There are several contributions of our work: (1) To our best knowledge, this is the first dedicated quality assessment work on image retargeting using compact reduced reference. The focus is on content-level structural distortion as compared to intensity-level distortion. Intensity-level distortion has been extensively studied but structural distortion has received relatively fewer research effort. (2) We made contributions in achieving a more fair comparison comparison between objective quality scores and human subjective ratings motivated by the observation that human rating has inherent fuzziness in deciding between images of similar quality. By grouping subjective scores into clusters and penalizing intra- and inter-cluster discordant pairs with different strengths, we provide an alternative way of comparing scores in a more stable and meaningful fashion comparing to the general approach of Kendall's τ correlation. (3) Our corner point matching-based framework successfully demonstrates the possibility that even with very compact side information extracted from the original image, quality assessment of structural distortion is not only possible, but also elegant—even better than full-reference metric, i.e., the earth mover's distance (EMD). (4) In addition to be able to give quality scores from various perspectives and a meaningful fused score, the proposed framework also provides annotated images showing the displacement vectors of corner points, and reconstructed original images to give observers a rough sense of how the original image may look like. These two types of auxiliary images can be used to assist human evaluation and as well be extended for many other applications such as classifying different retargeting operations.

In the rest of sections, we first describe various components of the proposed quality assessment algorithm in Section II with detailed descriptions on the matching part and our proposed quality metrics. We then explore assessment techniques by discussing Kendall's τ and its variant in Section III. Finally, we present the quality assessment results in Section IV and conclude in Section V.

II. Proposed Framework for Retargeted Images

Retargeting algorithms mainly introduce structural changes, therefore, to carry out image quality assessment on these images and with reduced references only, we need to extract compact partial information that can capture the structural information of the image and allow robust analysis of structural changes. The partial information used in this work is a set of corner points extracted from the original image.

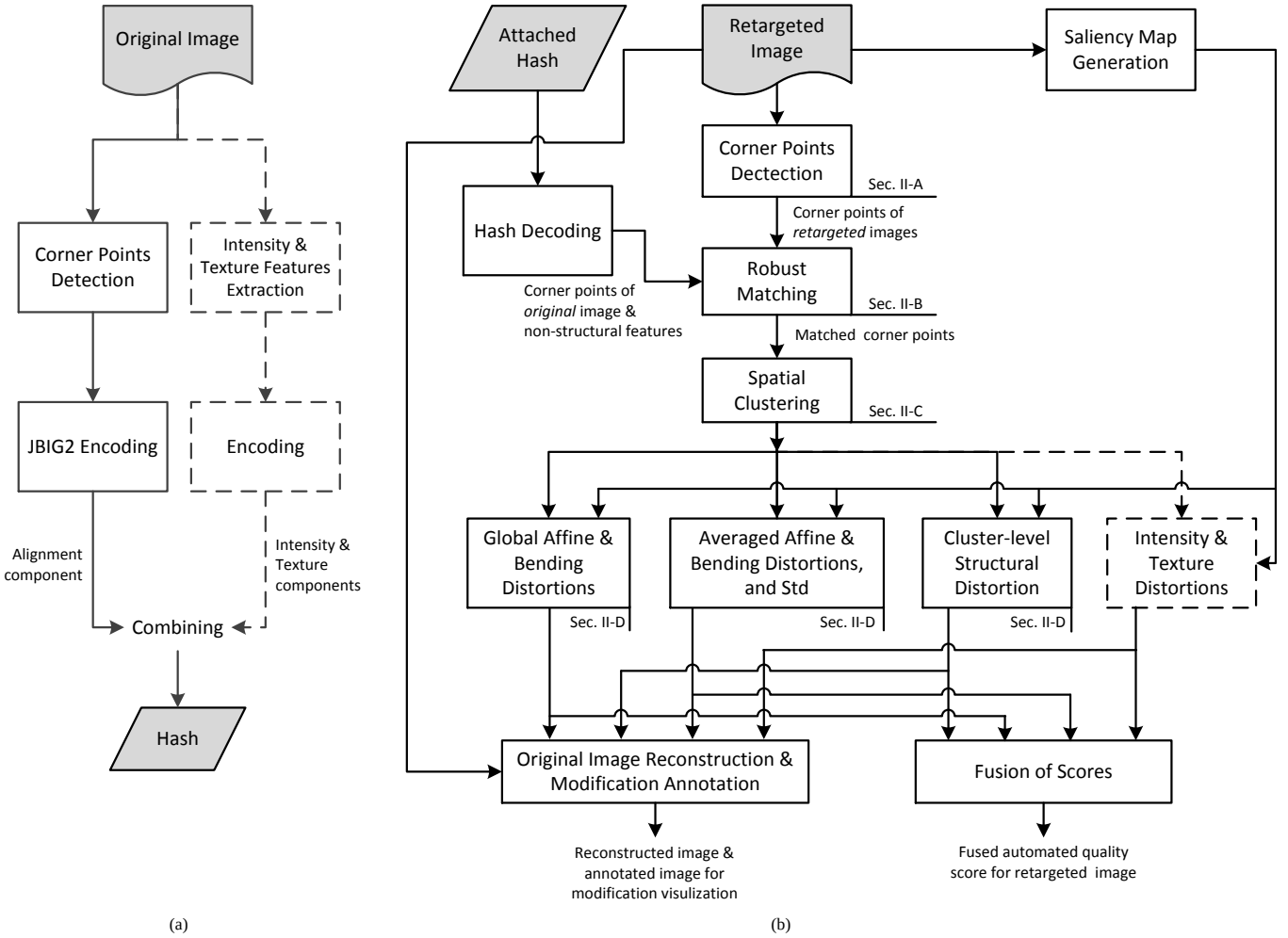


Fig. 1: Reduced-reference quality assessment framework for retargeted images: (a) hash generation; and (b) quality assessment which not only generates various quality scores, but also provides annotated images showing the displacement vectors of matched corner points and reconstructed images to give observers a rough sense of how the original image may look like.

In the retargeted image, the corresponding corner points change locations relative to those in the original image, so we aim to match the corner points of the retargeted image against those encoded in the partial information. Given the highly nonlinear and content-adaptive changes to the image structure, a robust matching between corresponding corner points between original and retargeted images is essential to enable accurate distortion measure. After matching, we compute quality scores, and obtain annotated images showing how different regions of the image are changed from the original reference image. Such annotated images can be used for subjective structural analysis and image quality measure. An reconstructed original image can also be generated to provide the observer a rough sense of how the original image looks like.

The output of this quality assessment framework includes not only quality scores, but also annotated images that describe the structural distortion of the retargeted image in detail. The role of the quality score is similar to that in conventional quality assessment work, i.e., to predict the quality degradation as would be perceived by a human observer. In the case of structural distortion, recent study by Rubinstein et al. [19] found that even human beings have difficulties in judging quality of retargeted images and have large discrepancies on how important each type of distortion is among content loss, symmetry violation, distorted edges/lines, and deformed objects/faces, etc. Therefore, providing a detailed distortion map in addition to quality scores can be very important to assist subjective quality assessment on the retargeted images and can allow for adaptations to the needs for specific applications.

In this paper, we focus our attention on the new and more challenging issue of the structural/geometric distortion, since relatively more matured modules measuring other types of distortion, such as intensity distortion and texture distortion, can be readily integrated into the framework. For examples, the intensity and the contrast components of SSIM [2], which uses block-level mean and variance as features respectively, can complement our proposed structural distortion measurement component towards a general-purpose quality assessment scheme. The overall framework is shown in Fig. 1.

Next, we will discuss each major component in the proposed quality assessment framework.

A. Compact Selection of Partial Information

In order to evaluate structural distortion, we need to select compact partial information to capture the structural information of the reference image. There have been a wide variety of detectors for interest point and corner point in literature, and they can be divided into three categories: contour based, intensity based, and parametric model based methods. A comprehensive survey and comparison of local interest point detectors can be found in [25].

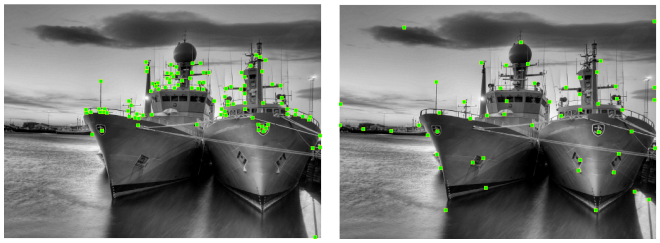
Corner points roughly capture the structural information of the image as they are typically located on dominant image structures, and they are also efficient to compute as compared to more computationally intensive local features such as SIFT. In this work, we try to reduce the use of partial information as much as possible by only recording the positions of corner points. We use the detection method proposed by Harris and Stephens [26], which identifies corner points through the intensity information of the local patch around the points. More specifically, a Harris matrix is computed for each point as

$$\mathbf{A} = \sum_u \sum_v w(u,v) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$

where I_x and I_y are the partial derivative of the image intensity along horizontal and vertical directions, respectively; $w(u,v)$ are the weights assigned to each neighboring pixel. A Gaussian weighting can be applied to provide an isotropic response. A corner point can then be characterized by analyzing the eigenvalues λ_1 and λ_2 of the Harris matrix $\mathbf{A}$, where $\lambda_1 \geq \lambda_2$. If $\lambda_1 \approx 0$ and $\lambda_2 \approx 0$, the pixel (x,y) is not an corner point nor an edge point. If λ_1 has large positive value and $\lambda_2 \approx 0$, then an edge point is found. If both λ_1 and λ_2 have large positive values, the pixel (x,y) is considered a corner point. Efficient detection of corner points may use the function $\det(\mathbf{A}) - k \cdot \text{trace}^2(\mathbf{A})$ in place of computing the eigenvalues.

In order to keep partial information as compact as possible for the reduced-reference settings, we need to control the number of corner points that are selected. Towards this goal, we favor the corner points to appear on major object or structure in the image, rather than capturing small details such as texture on the background. To suppress noise and retain only corner points on salient objects, we perform smoothing on the image before doing the corner point detection. A small amount of blurring helps remove noisy corner points that might appear on smaller edges which may not be useful and stable for evaluating the global structure distortion. Also, in order to avoid a large amount of corner points in a small neighborhood, local maximum suppression is carried out to retain only the dominant corner in a local patch. The larger the window is used for the local maximum suppression, the more spread out we can expect from the detected corner points. An example of corner point detection and selection is shown in Fig. 2, where we can see that a larger maximum suppression window causes the detected corner points to spread out more.

After corner points are detected, we need to compactly encode their positions. By considering the corner map as a binary image, where corner points are represented by value 1 and all other pixels in the image have value 0, we can use the JBIG2 encoding [27] to compactly represent such a binary image. JBIG2 is an international standard that provides a very good compression ratio for binary images and serves as a good candidate for our application. The size of JBIG2 compressed corner map with respect to different number of corner points is shown in TABLE I. For comparison, we also show the size of the corner map encoded by PNG format. We can see that JBIG2 compression provides more than 40% savings in representing the partial information of corner points than simple PNG encoding. The simulation results for this paper are based on the setting that each original image uses 120 corner points as partial information.



(a) (b)

Fig. 2: Corner points detected at different maximum suppression window sizes N: (a) $N = 3$, and (b) $N = 15$. (Figures are best viewed in color.)

TABLE I: Compact encoding of corner points

# of corner points	50	80	100	120	150	200
JBIG2 size (byte)	225	276	310	339	391	469
PNG size (byte)	380	475	541	598	682	852

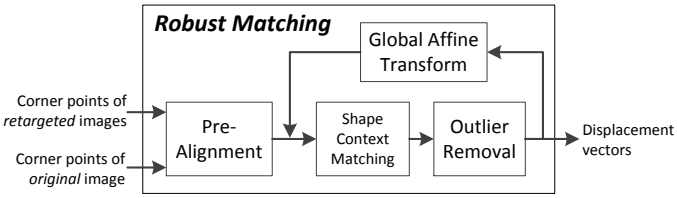


Fig. 3: Robust matching using distance map, shape context, and global affine transform.

B. Robust Matching

The robust matching module aims at creating correspondences between corner points in the original image and those in the retargeted image, and then output displacement vectors that can be exploited for determining structural distortion. As shown in Fig. 3, the module consists of a pre-alignment stage and an iterative refinement stage, and the latter includes shape context matching, outlier removal, and global affine transform. Details of the robust matching module are described as follows.

1) Pre-alignment: This stage provides a provisional alignment compensating the translational motion between the retargeted image and its original version by minimizing their Chamfer distance [28] on distance map. Fig. 4 illustrates

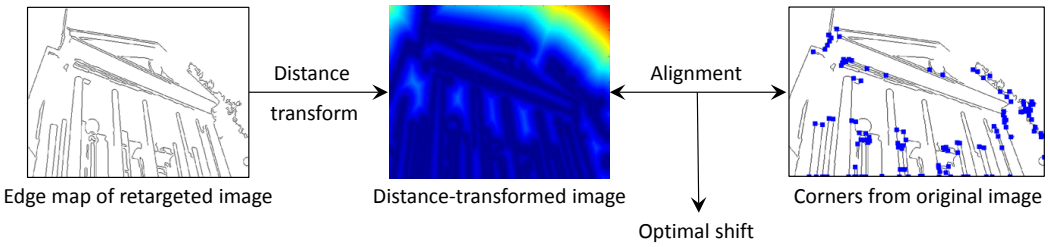


Fig. 4: Pre-alignment: align corner points from original image to the retargeted image by minimizing Chamfer distance. (Edge maps inverted.)

the details of the pre-alignment stage. Given the retargeted image, we first compute its edge map and perform the distance transform [29] on the edge map. The distance transform, which generates an image in which value of each pixel represents its distance to the closest edge point, can be calculated by

$$dt_E(\mathbf{p}) = \min_{\mathbf{p_e} \in E} \|\mathbf{p} - \mathbf{p_e}\|_2$$

where E is the edge map of the retargeted image, dt_E is the distance-transformed image, $\mathbf{p_e}$ is a point in E, and $\mathbf{p}$ is a point in dt_E. To find the optimal alignment, we overlay the corner map obtained from the original image onto the distance-transformed image by examining different shift vectors. For each shift vector $\mathbf{s} = (x, y)$, we compute the Chamfer distance defined as

$$d(\mathbf{s}) = \frac{1}{N} \sum_{\mathbf{p_c}} dt_E(\mathbf{p_c} + \mathbf{s})$$

where $\mathbf{p_c}$ represents a corner point position, and N is the total number of corner points encoded as partial information. A smaller distance indicates that most of the corner points are located closer to the edges in the retargeted image, and therefore, a better alignment. The optimal pre-alignment is then computed as the shift $\mathbf{s}^*$ that gives the minimum Chamfer distance.

2) Shape context matching: As retargeting operation can change the internal structure of the image by scaling, carving, and warping, the task after pre-alignment is to find correspondence between the points in the original and retargeted images. This task is similar to that of shape matching, where a shape is matched with another similar but deformed one. Shape context proposed by Belongie et al. [24] is a useful technique for shape matching, and we use it here to find correspondence between corner points.

Shape context is a local descriptor associated with each shape point to describe the coarse distribution of the rest of corner points with respect to the current point. Fig. 5 shows two sets of points represent the same character "A" but has slightly different shapes, and Fig. 5c shows a partition of polar-angular bins. For each point, its shape context is essentially a histogram counting the number of other points that fall into each of the bins shown in Fig. 5c. Different bins capture shape information in various orientations and distance relative to the current point.

For our problem, we compute the shape context for each corner point, and finding correspondence is then equivalent to finding for each corner point in one set, a corner point in the other set that has the most similar shape context. The distance between two shape contexts can be computed using the χ^2 test statistic:

Fig. 5: Illustration helps to demonstrate the shape context computation. (Figure from Belongie et al. [24])

$$C(p, q) = \frac{1}{2} \sum_{i=1}^{K} \frac{[h_p(i) - h_q(i)]^2}{h_p(i) + h_q(i)}$$

where $h_p(i)$ and $h_q(i)$ denote the K-bin normalized shape context histogram at points p and q, respectively. Given the shape context distance between all pairs of corner points p_i in the original image and corner points q_j in the retargeted image, we need to find a one-to-one correspondence to minimize the total cost of matching $C_s(p, q) = \sum_i C(p_i, q_{\pi(i)})$, where π is a permutation representing the one-to-one correspondence. We further regularize with term $\|p - q\|_2$ so that $C_{total}(p, q) = \alpha C_s(p, q) + (1 - \alpha)\|p - q\|_2$. This matching problem can be solved using the Hungarian method [30] in $O(N^3)$ time.

3) Outlier removal: A small set of neighboring points on a rigid body usually have similar displacements, and we use this criterion to perform an additional step of outlier removal to refine the matching results. For each corner point p_i in the original image, we compute a displacement vector $v_i = q_{\pi(i)} - p_i$, which captures the displacement of its corresponding point in the retargeted image. We then compare v_i with every other displacement vector v_j of its neighboring point p_j that falls into a small neighborhood of p_i. If the difference between v_i and the average of $\{v_j, j \neq i\}$ is larger than a certain threshold, we consider $(p_i, q_{\pi(i)})$ a false match. The outlier removal leads to a more robust matching and thus paves the way to a meaningful distortion evaluation.

4) Affine compensation: The pre-alignment, shape context matching, and outlier removal perform well for those cases that retargeted images are not significantly scaled or rotated. However, for such more difficult cases as significant scaling/rotation and shearing, a correct correspondence may be

misclassified as an outlier and thus removed. Hence, there is a need to compensate geometric distortions, possibly before shape context matching in order to obtain a set of more trustful correspondences. We propose to transform the original image by a set of global affine parameters estimated from the displacement vectors and align with the retargeted image. Once the warping is done, the shape context matching can be carried out again to match the affine-transformed version of original image with retargeted image. The steps iterate as depicted in Fig. 3, until a set of stable correspondences is established.

C. Spatial Clustering

The main idea of state-of-the-art retargeting operations is to resize different parts of the image differently through scaling, warping, or carving. Salient objects or regions will be scaled more uniformly so that the aspect ratio of the object is preserved and minimum distortion is introduced. For less important regions, the resizing operation will have fewer constraints since a large distortion can be tolerated. This main spirit of retargeting motivates us to group corner point correspondences into clusters so that evaluation for structural distortion can be carried out differently for various regions of an image. With clusters having homogeneous displacement vector fields, more stringent pruning can also be carried out within clusters to achieve a good matching.

Spatial clustering, by which similar spatial objects are grouped, is an important component of spatial data mining [31]. Spatial clustering techniques can be classified into four categories: partitioning method, hierarchical method, density-based method and grid-based method. A detailed survey of spatial clustering techniques can be found in [32]. In this work, we use the K-medoids algorithm to cluster corner points. K-medoids is similar to K-means but instead of using the average of cluster elements as a center, K-medoids uses the most central data point in the cluster as its center. This makes K-medoids clustering more robust to noise and outlier data than K-means clustering. We use an efficient K-medoids clustering method called Clustering Large Application using RANdomized Search (CLARANS) [33] for its efficiency and good quality of clustering. The basic idea of CLARANS is to perform a randomized search for a node with the minimum cost in a graph. The node here represents a selection of K-medoids, and the graph is composed of such nodes and any two neighboring nodes differ by only one medoid. The use of randomized search provides the advantage of efficiency and the benefit of not confining the search to only a localized area. As both K-means and K-medoids are partition-based clustering, they have the limitation of requiring to specify the number of clusters K at the beginning. To alleviate such a constraint, we select a relatively large K to start with, then merge neighboring clusters if their displacement vectors are similar, and split a cluster if the displacement vectors within are diverse. An example of the spatial clustering is shown in Fig. 6, where Figs. 6(a) and 6(b) show the original and retargeted images and Fig. 6(c) shows the corner point correspondences and their clustering result. Corner points of different clusters are labeled with different colors.

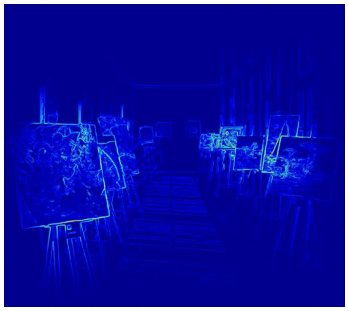


Fig. 7: Saliency map of the resized image Fig. 6c.

D. Distortion Metrics

Now with the correspondences obtained between corner points of an original and its retargeted version, we can measure the structural distortion due to the retargeting operation from several perspectives.

1) Cluster-level structural distortion: This distortion is calculated by weighing the distortion of each local cluster with its saliency. Saliency is a subjective measure that captures what human observers consider as important in an image. Examining saliency is an important step in sophisticated image retargeting methods, and different saliency metrics have been used in the literature, such as gradient magnitudes [16] and discontinuity of neighbors if a pixel is removed [14]. In this work, we use the saliency measure proposed by Wang et al. [17], which combines the gradient magnitude and the saliency map by Itti et al. [34]. More specifically, the gradient information captures structural areas, and the saliency map captures attractive areas that have different color, intensity and orientation properties than their surroundings. The combined saliency map is evaluated as $W = W_\alpha \times W_\beta$, where $W_\alpha = \sqrt{(\frac{\partial}{\partial x} I)^2 + (\frac{\partial}{\partial y} I)^2}$, and W_β is the saliency map by Itti et al. An example of the saliency evaluation is shown in Fig. 7.

We have now obtained the clusters $C_1, \cdots, C_k$ of corner points that indicate how different parts of the image have been transformed during the retargeting operation, and the saliency map that shows how important different areas of the image are. Combining the two, we can compute some quality score that measures structural distortion through how consistent those related regions are translated. For each cluster C_i, we select l neighbor clusters $C_{i1}, \cdots, C_{il}$ that are within certain distance threshold to the current cluster. The distortion contributed by two neighboring clusters is computed as

$$d_{ij}(C_i, C_{ij}) = \|v_i - v_{ij}\| \cdot d_s e^{-\alpha \cdot d_e}.$$

Here v_i and v_{ij} are the displacement vectors of the medoids in cluster C_i and C_{ij}, respectively; d_e is the distance between the two clusters; $d_s = \frac{S_{ij}}{\max(S_i + S_j, S_{ij})}$ measures how likely the two clusters cover the same object, where S_i and S_j are the saliency values of regions covered by C_i and C_{ij}, respectively, and S_{ij} is the saliency of regions between C_i and C_{ij}. If two regions of high saliency are connected by other regions of high saliency, these two regions are likely to cover the same object or the same group of connected objects. If they are connected

(a) (b) (c)

Fig. 6: Illustration for spatial clustering: (a) original image, (b) its resized image by a retargeting algorithm called Streaming Video [18], (c) resized image overlaid with spatial clustered displacement vectors

by regions of low saliency, they may cover different objects, and therefore the weight assigned to the discrepancy penalty between their displacement vectors will be lower. The quality score that measures such cluster-level structural distortion is computed as

$$\text{GSS} = \sum_{i=1}^{K} \sum_{j=1}^{l} d_{ij}(C_i, C_{ij}).$$

2) Global affine and related global distortions: Given the corner point correspondences, we are able to estimate a global affine transform $\mathbf{A}$ that transforms the set of corner points in the original image to the set of corner points in the retargeted image. The singular value decomposition of the 2-by-2 matrix $\mathbf{A}$ is $\mathbf{A} = \mathbf{O}_{\theta_1}\mathbf{D}\mathbf{O}_{\theta_2}$, where $\mathbf{D}$ is a diagonal matrix with singular values λ_1 and λ_2, suggesting that the transform basically scales along the direction θ_1 with factor λ_1 and scales along the orthogonal direction θ_2 with factor λ_2. If λ_1 and λ_2 are close to each other, the global transform roughly preserves the aspect ratio of the image; if λ_1 is much larger than λ_2, the aspect ratio changes significantly and large perceptual distortion is expected. We define the global affine cost as

$$\text{GAffine} = \log(\lambda_1/\lambda_2).$$

We propose to incorporate the second measure called global bending energy, which complements the global affine cost. It is computed from a parametric *thin-plate spline mapping* $f : \mathbb{R}^2 \to \mathbb{R}^2$ [35], [36] estimated from the point correspondences obtained in Section II-C using the thin-plate spline model [37], [38] which is useful to measures how twisted the estimated mapping f is. The mapping is a superposition of the global affine transform and a set of affine-free *principal warps*, and only the principal warps contribute to the bending energy defined as

$$\text{GBending} = \iint \left\{ \left(\frac{\partial^2 f}{\partial x^2}\right)^2 + 2\left(\frac{\partial^2 f}{\partial x \partial y}\right)^2 + \left(\frac{\partial^2 f}{\partial y^2}\right)^2 \right\} dxdy.$$

In this work, we use the global bending energy to measure the remaining global structural distortion orthogonal to global affine distortion. A smaller structural distortion leads to less twisted mapping and thus smaller bending energy.

3) Averaged affine and related averaged distortions: The affine transform captures linear distortions and is incapable of capturing higher-order nonlinear warpings and detailed distortions. We propose to use local affine model to capture locally specific distortions. We calculate affine distortions for all clusters individually, weigh them according to the saliency map, and denote the result as an averaged affine (AAfine) distortion.

In analogous to global bending serving as a complement measure of the global affine, we propose to explore the remaining distortion after compensating the local affine transform. Here, we develop a metric of averaged bending (ABending) distortion, which is the weighted local bending distoriton of all clusters. The weights again are obtained from the saliency map.

In addition to AAfine and ABending metrics derived from a parametric viewpoint, we also try to quantify distortion using a nonparametric approach through a weighted average of cluster-level standard deviations (AStd) of displacement vectors. It is evaluated by first calculating the standard deviation of affine-compensated displacement vectors for each cluster, and then averaging the standard deviations using the weights obtained from the saliency map. This is motivated by the observation that a good retargeting operation tends to incur less distortion within a local cluster.

III. Performance Evaluation Methodology

A. Dataset and Subjective Reference Scores

We carry out experimental work using the RetargetMe dataset [39] containing 37 distinct color images, and each of these images has 8 different versions generated by 8 retargeting algorithms which are briefly reviewed in Appendix A.

The subjective evaluation on the quality of the retargeted images [19] were carried out through Amazon's Mechanical Turk crowdsourcing service. The subjective scores are calculated as follows. A human observer is presented with two selected retargeted versions of an original image and asked about his/her preference in terms of which one looks better. The questions are asked over all $\binom{8}{2} = 28$ possible pairs of retargeted versions, and the number of votes received by each retargeted version is considered as human's rating. Such

experiment is repeated on 210 participants and leads to the collective subjective rating for the retargeting methods. More details can be found in [19].

B. Assess Quality Metrics' Subjective Correlation

1) Kendall's τ correlation: To assess the the proposed quality metrics for the reduced-reference quality evaluation, we examine to what extent they correlate with human's ratings. The previous RetargetMe work [19] adopted the conventional version of Kendall's τ correlation between the subjective ratings and objective scores:

$$\tau = \frac{n_c - n_d}{n_c + n_d}$$

where n_c is the number of concordant pairs and n_d is the number of discordant pairs over all pairs of entries in the two rankings. We can see that τ ranges from -1 to 1, where a value of $+1/-1$ indicates perfect agreement/disagreement between the subjective votes and automated scores, and a value of 0 means no agreement at all. Furthermore, to verify that the correlation values are significantly different from 0, a χ^2 hypothesis test is carried out against the null hypothesis that the observed correlations have zero mean, suggesting that the subjective and objective scores are uncorrelated.

2) Limitations of Kendall's τ: Rank correlation such as Kendall's τ and Spearsman's ρ are known for their robustness against outliers when compared to the commonly used Pearson product-moment correlation coefficient because ranking effectively equalizes the distance between consecutively ordered samples and omits their absolute differences. While this robustness makes the measurements less susceptible to outliers, when Kendall's τ is employed to quantify the correlation between subject votes and automated scores, its "robustness" may not be adequately capture the inconsistency between subject votes and automated scores. More specifically, we see two potential limitations of the conventional version of Kendall's τ for our quality assessment application:

i) an enhanced sensitivity to those subjective ratings bearing similar amounts of votes, and
ii) a reduced sensitivity to penalizing the discordant pairs.

The first aspect concerns that visually similar retargeted images are easy to introduce mismatches between rank vectors, which results in a smaller correlation coefficient. Our visual check on the RetargetMe database along with the subjectives votes reveals that observers of subjective evaluation usually have difficulty in differentiating among similar visually pleasing retargeted images. As shown in Fig. 8, images generated by various retargeting algorithms can be grouped into several clusters (using such a classification algorithm as the nearest neighbor approach) according to the votes that they received. Fig. 8(a) shows that within Cluster 2, a slightly higher number of observers believe that SC is better than SM. It is possible that if the subjective test is carried out on another set of observers, they may conclude SM looks slightly better than SC. This motivates us to consider an improvement on Kendall's τ to better account for those ratings receiving similar votes and not to penalize their rank differences when calculating the final correlation coefficient.

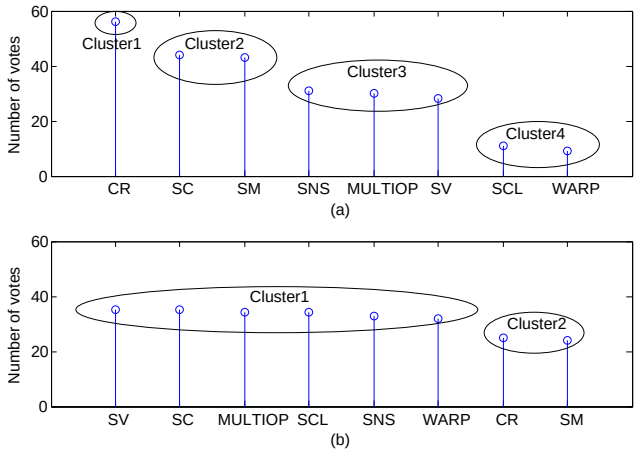


Fig. 8: Subjective votes for eight retargeting methods applied to (a) Image#19, and (b) Image#36.

The second aspect concerns that the discordant pairs are not penalized enough, which results in a larger correlation coefficient. In the example shown in Fig. 8(a), if an automated method favors SCL over SC, this would be substantially different from the subjective results and yet Kendall's τ does not adequately account for such significance mismatch. Hence, it is desirable to be able to take into account the level of discordance.

3) Clustered Kendall's τ: To address the two issues mentioned above, we propose a clustered Kendall's τ that first group the subjective votes into clusters and then apply Kendall's τ with adaptive penalty levels—lower for intra-cluster discordant pairs, and higher for inter-cluster discordant pairs. More specifically:

- If the pair is concordant, then increase the concordant counter n_c by 1, i.e., $n_c \leftarrow n_c + 1$ (same as in conventional Kendall's τ).
- If the pair is discordant, then depending whether two elements of subjective votes fall into the same cluster or different clusters, adaptively determine the penalty levels for this pair:
  - if the two elements are within the same cluster, we may either penalize slightly or do not penalize at all, according to a parameter p_0, i.e., $n_d \leftarrow n_d + p_0$, $0 \leq p_0 < 1$;
  - if the two elements belong to different clusters, we penalize the discordant pair based on the difference between clusters' indices Δi, i.e., $n_d \leftarrow n_d + g(\Delta i)$, where g is a monotonically nondecreasing function.

TABLE II shows an example of clustered Kendall's τ, and compare them with the conventional Kendall's τ. Here, we choose $p_0 = 0.5$ and $g(1) = g(2) = 1, g(3) = g(4) = ... = 2$. Note that the difference in the setting of the conventional Kendall's τ: n_d is increased by 1 for each discordant pair.

It is easy to verify that value of correlation also ranges from -1 to $+1$ for the clustered Kendall's τ, and that when a concordant pair becomes discordant, the resulting τ value will drop. Note that the "no-agreement" value now depends on the selections of penalty levels, and are may not be 0. This also

TABLE II: Comparison between the conventional version of Kendall's τ and proposed clustered Kendall's τ

	Kendall's τ		Clustered Kendall's τ	
	Concordant (reward level)	Discordant (penalty level)	Concordant (reward level)	Discordant (penalty level)
Intra Cluster	1	1	1	0.5
Inter Cluster	1	1	1	$1, 1, 2, 2, ...$

TABLE III: Correlation with subjective votes with and without reference images (Kendall's τ)

	Metric	Subjects Having Reference			Subjects Having No Reference		
		Mean	Std	p-value	Mean	Std	p-value
Prior art	SIFTflow	**0.145**	0.26	0.001	**0.170**	0.25	0.000
	EMD	**0.251**	0.27	0.000	**0.199**	0.31	0.000
	BDS	0.083	0.27	0.040	**0.135**	0.28	0.002
	BDW	0.046	0.18	0.164	0.098	0.24	0.019
	EH	0.004	0.33	0.468	−0.002	0.33	0.516
	CL	−0.068	0.30	0.923	−0.093	0.33	0.975
	RAND	−0.031	0.28	0.742	0.006	0.28	0.451
Proposed	GAffine	**0.263**	0.29	0.000	**0.249**	0.33	0.000
	GBending	**0.170**	0.26	0.000	0.095	0.25	0.023
	AAffine	**0.156**	0.34	0.000	**0.143**	0.36	0.001
	ABending	**0.112**	0.25	0.009	0.095	0.24	0.023
	AStd	**0.193**	0.22	0.000	0.060	0.30	0.104
	GSS	0.097	0.23	0.021	0.087	0.26	0.034

TABLE IV: Correlation with subjective votes with and without reference images (Clustered Kendall's τ)

	Metric	Subjects Having Reference		Subjects Having No Reference	
		Mean	Std	Mean	Std
Prior art	SIFTflow	**0.120**	0.29	**0.171**	0.27
	EMD	**0.244**	0.31	**0.210**	0.33
	BDS	0.053	0.30	**0.131**	0.31
	BDW	−0.008	0.20	0.091	0.26
	EH	−0.039	0.35	−0.020	0.34
	CL	−0.105	0.33	−0.111	0.37
	RAND	−0.087	0.30	−0.010	0.31
Proposed	GAffine	**0.264**	0.35	**0.249**	0.37
	GBending	**0.152**	0.31	0.087	0.27
	AAffine	**0.134**	0.39	**0.134**	0.40
	ABending	0.081	0.29	0.082	0.28
	AStd	**0.160**	0.26	0.054	0.32
	GSS	0.062	0.27	0.067	0.28

implies that specific correlation values given by Kendall's τ and clustered Kendall's τ are not directly comparable.

IV. Performances of Automated Scores

A. Average Performances in Terms of Kendall's τ

We compute the Kendall's τ for the proposed quality metrics and compare to metrics listed in [19]. In TABLE III, we show the results by the conventional version of Kendall's τ which is comparable to those results in [19]; in TABLE IV, we show the results by clustered Kendall's τ to assess the correlation from a perspective which accounts for penalties in such a way that it is less sensitive to intra-cluster discordance and emphasizes more for inter-cluster discordance. Note that due to difference in the two versions of Kendall's τ as discussed in the last section, the correlation values are meant to be compared within each tables but not to be compared between tables. Correlation values greater than 0.1 are emphasized using boldface.

The RetargetMe database provides two scenarios for a human observer to answer his/her preference are tested: with and without the original image. The case of subjects having reference was conceptually similar to the way that full-reference and reduced-reference metrics work, and the case of without reference was considered more as an exploration. We thus mainly focus on the "with reference" case, and yet for the sake of completeness, we also include the results comparing with subjective ranking when subjects have no reference. To help analyze the results, we review various metrics that were compared in the RetargetMe paper [19] in Appendix B.

TABLES III and IV consistently reveal that most elements of our proposed quality metrics using partial reference from corner point locations can achieve good performances, and some are as good as or better than such top performing full-reference metrics as EMD and SIFTflow. We see that the global affine cost is the best in the sense that it is highly likely to give a score that is consistent with human's ratings.

TABLE V: Comparison of Lengths of Raw Hashes of Various Metrics

Metric	Length of Raw Hash
BDS	N (original image is needed)
BDW	N (original image is needed)
EMD	$\frac{N}{n^2} \times$ 5-tuple[a]
SIFTflow	$N \times 128$-tuple
EH	16×5-tuple
CL	64×12-tuple
Proposed	$\sim 120 \times 2$-tuple

[a] N represents the total number of pixels of an test image. RetargetMe database [39] has images of sizes range from 300×392 to 813×1024 so that $N \sim 10^5$. And n is an image downscaling factor which is set to 8 in [21].

TABLE V summarizes the lengths of raw hashes for the aforementioned metrics. Amongst the metrics, BDS and BDW require the original image to be available and thus fall into the category of full-reference quality assessment; The hash length of SIFTflow is proportional to the number of pixels that can be as large as 10^7 for a moderate-resolution image; EMD uses about 10^3 to 10^6 five-parameter tuples, as compared to the proposed method of only about 120 two-parameter tuples. EH and CL have short hash length but do not explicitly account for

structural distortion. This comparison shows that our proposed corner-point-based hash along with the corresponding alignment method is highly compact and has a length of about two to five orders of magnitude shorter than other well-performing metrics in the prior art (such as EMD and SIFTflow).

B. Structural Distortion Analysis

In this part, we demonstrate how well the proposed quality assessment framework works through examples. Not only does the framework generate *quality scores*, but also it provides *a detailed annotated image* and *a reconstructed image* that can assist human observers to make their own decisions on the quality of the image, and allow for flexible adaptation to the needs of different application scenarios.

In Fig. 9, we show images retargeted by three different operations: cropping (CR), seam-carving (SC), and streaming video (SV). We choose these retargeting methods for comparison since they capture three major types of retargeting effects. Cropping retains the aspect ratio of salient objects, and is found to be preferred by many human observers, especially when an original image is not available for comparison [19]. Seam carving resizes an image by removing seams of low energy; it often change the aspect ratios of salient objects when the resizing factor is large, and incur non-uniform global distortion. Streaming video tries to scale various parts of the image differently so that the edge discontinuity as observed in seam carving can be reduced.

1) Image annotated with displacement field: In Fig. 9, we show the point correspondences between the retargeted images and the original image by displaying the *raw* displacement vector fields and the *affine-compensated* displacement vector fields. The correspondences reveal different types of retargeting effects. For the image having undergone cropping shown in Figs. 9a and 9d, we can see as expected that the matched corner points have no displacement at all. For seam carving results shown in Figs. 9b and 9e, the displace vectors are all horizontal, which confirms that vertical seams have been removed from the image by the seam carving operations. Also, several regions have different displacement sizes, which indicates the different amount of carving at different regions. For example, Fig. 9e shows that even when the effect of global affine transform is removed, some affine-compensated displacement vectors are still strong and varies across regions. For image having undergone streaming video based retargeting shown in Figs. 9c and 9f, we can see the retargeting has an overall scaling effect, which scales down the salient region (the lower part of the image) and stretches the nonsalient region (the upper part of the image). After compensating the global affine effect, the bottom part of the image matches nearly perfectly with almost no residual displacement vectors left.

With the assistance of these annotated image results, a user can have a balanced view on the image quality—from the user's high-level visual perception as well as the assistance by the two proposed types of objective displacement vector fields showing how corner points are moved. All these are based on very compact reference information containing encoded a small number of corner point positions of the original image.

TABLE VI: Automated quality scores for images retargeted by CR, SC, and SV methods

Metric	Retargeting operations		
	CR	SC	SV
GAffine	0.01	0.33	0.10
GBending	0.00	0.00	0.00
ABending	0.00	0.27	0.57
AAfine	0.00	0.21	0.08
AStd	1.21	6.00	3.16
GSS	0.00	0.00	0.01
Fused score	0.04	0.38	0.16

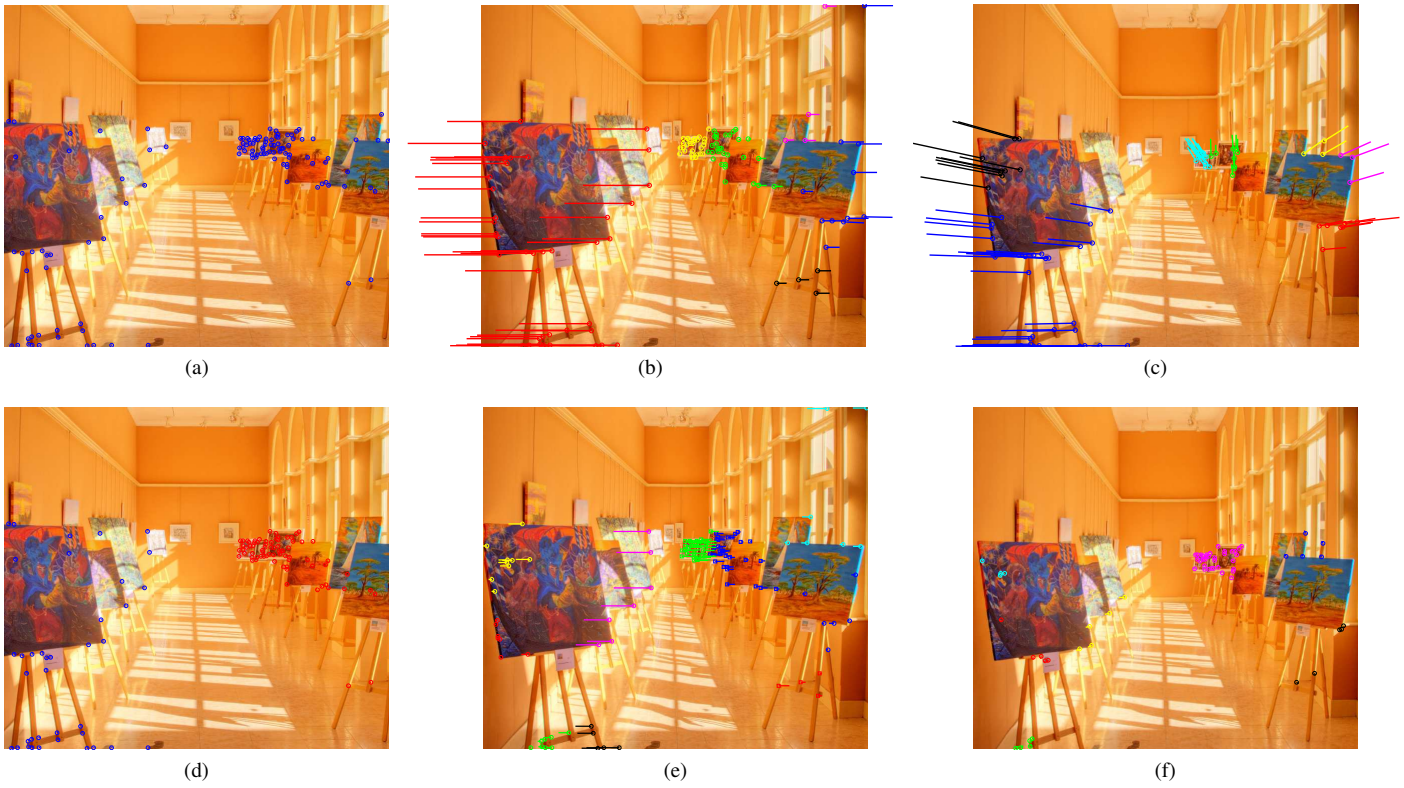
2) Automated quality scores: The automated quality scores computed on three retargeted images in Fig. 9 are shown in TABLE VI. We can see that cropping causes the least amount of error, which is expected because no structural distortion is ever introduced for the remaining content. Seam carving has the highest affine distortions in both global and averaged scales, which adequately reflects the visible structural distortions introduced by this retargeting technique. Seam carving also has larger averaged standard deviation of displacement vectors (AStd), which reveals how uniform in length the affine-compensated vectors of every cluster are. Overall, we see that the proposed metrics correlate well with subjective expectations.

3) Reconstructed original image: To further assist the observer, it is beneficial to convert a retargeted image back to the grid of the original image to reconstruct the original image. A raw displacement vector field generated by the point correspondences can assist this reconstruction. As a proof-of-concept, we show a simple approach that compensates the distortion due to the equivalent global affine transform. The reconstructed images are shown in Fig. 10.

In summary, the raw displacement field and affine-compensated displacement field, which are estimated from the corresponding corner points between the retargeted image and the original image, can not only be used to compute scores for automatic quality evaluation, but more importantly, they provide an important tool to assist human evaluation when the original image is not available. The compact side information of the corner points provides rich information on how the image has been retargeted. Since people generally have different preferences on various kinds of distortions, such as edge discontinuity, violated symmetry, and content loss, such displacement fields allow users to make their own judgements instead of relying on a "one-size-fit-all" set of quality scores. This is an important advantage of our proposed framework.

C. Fusion of Multi-Facet Scores via Learning Theory

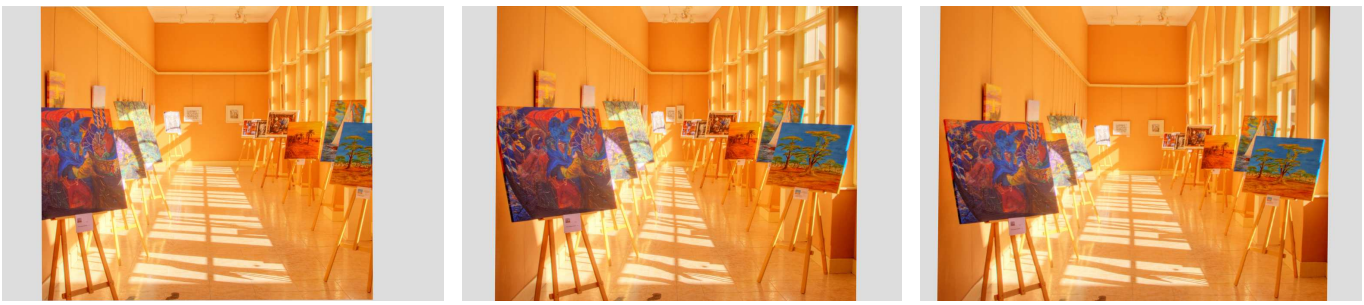
As we have seen, our proposed metrics have very good correlations with the human ratings and are comparable to the current state-of-the-art full-reference metrics. Each of these proposed scores provides some aspect of structural quality assessment of an retargeted image. In some applications,



(a) (b) (c)

(d) (e) (f)

Fig. 9: Annotated images: *raw* displacement vector fields (top row) and *affine-compensated* displacement vector fields overlaid on images undergone retargeting operations (bottom row) (a,d) Cropping (CR), (b,e) Seam Carving (SC), and (c,f) Streaming Video (SV), respectively. (Best viewed on screen with scaling ratio 400% or more.)



(a) (b) (c)

Fig. 10: Images reconstructed from retargeted images by various operations: (a) CR, (b) SC, and (c) SV. The original image is available in Fig. 6a. Missing regions due to retargeting operations are shaded.

it is desirable to have a single score that can provide as comprehensive assessment as possible. We now examine the fusion of elements of the metrics developed.

We consider a fusion strategy by a weighted average of various metric elements and formulate as a learning problem of weight parameters: the objective function for maximization is Kendall's τ correlation coefficient, and the variable to be learned is the weight vector $\mathbf{w} \in \mathbb{R}^6$.

We propose using leave-one-out cross-validation to learn the weights for the proposed metrics. We apply our prior knowledge to exclude those images that have substantially wrong corner-point matching or matching are not on the main objects. This gives us 23 images, of which 18 randomly selected images form a training set, and 5 images form a testing set. As the optimization problem is nonconvex, a pattern search via the Hooke-Jeeves optimization algorithm [40] is used.

Fig. 11 shows the weight vectors whose validation results give positive correlations, and the average weights are shown in thick curve. As the learned vectors are fairly consistent across validation runs, we adopt the average vector and test it on five testing images. The testing results show very good Kendall's of $(0.21, 0.57, 0.43, 0.79, 0.43)$ suggesting good performance of the fusion. To put this in perspective, we recall

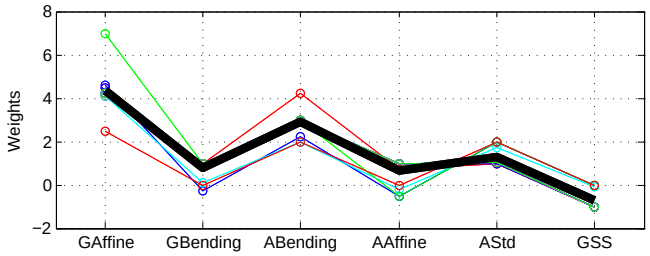


Fig. 11: Weight vectors learned using leave-one-out cross-validation, and the thicker one is their average.

that the best average Kendall's τ value of our proposed metrics is 0.263 given by GAffine, and the best average Kendall's τ of full-reference metrics is 0.251 by EMD. In the last row of TABLE VI, we calculate the fused scores for the qualities of retargeted images.

V. Conclusion

In this paper, we have studied the problem of reduced-reference quality assessment for retargeted images by measuring structural quality degradation. We have proposed to compactly encode strong corner points as partial information and compute correspondences of corner points to estimate detailed displacement vector fields. Novel structural quality metrics are proposed, and experiments show that the proposed metrics have significant positive statistical correlation with human subjective evaluations and a fusion of the metrics outperforms the best known full-reference quality metric for retargeted images. Furthermore, the framework also provides annotated images and reconstructed images which are useful tools assisting users to evaluate image quality based on their own preferences, instead of having to rely on a single quality score by a conventional quality assessment metric. The proposed structural metrics can complement and be integrated with metrics that assess intensity and texture qualities to form a more general quality assessment scheme. As future work, it will be interesting to see how the estimated distortion map can be used to classify different retargeting methods, and how a more precise matching can be achieved with a limited number of corner points.

Appendix A
Brief Review of Various Retargeting Algorithms

A survey of a number of retargeting algorithms can be found in the supplemental sheet of [19]. Here we provide a brief review. Specifically, scaling (SCL) [19] applies non-uniform scaling with bicubic interpolation to resize the image. Cropping (CR) [19] manually extracts a subimage from the original image. Seam carving (SC) [14] resizes the image by adaptively removing "seams", sequences of 8-connected pixels with low values defined by some energy function from the original image; and such an energy function can reflect gradient magnitude, entropy, visual saliency, or eye-gaze movement. Nonhomogeneous warping (WARP) [16] shrink less important regions more severely according to an importance map. Scale-and-Stretch (SNS) [17] iteratively computes optimal local scaling factors for each local region guided by a significance map and updating a warped image that matches these scaling factors as closely as possible. Multi-operator (MULTIOP) [15] uses a combination of seam carving, scaling and cropping to resize the image. Shift-maps (SM) [41] operates at pixel level but is able to remove pixels belonging to objects altogether with a global optimization. Streaming Video (SV) [18] uses mainly per-pixel warp with a set of automatic and interactive quality criteria to obtain an image of required resolution.

Appendix B
Brief Review of Various Quality Assessment Metrics

We also provide a brief review of a number of quality assessment metrics. Specifically, Bidirectional similarity (BDS) is an image similarity measure proposed in [20]. For each patch in one of the two images, a well-matched patch is sought in the other image, and the distance of two images is defined as the mean distance in color space between corresponding patches. [19] Bidirectional warping (BDW) [15] is a similar metric with the exception that the mapping between the two images is constrained to be monotonic, i.e., the resulting mapping will maintain the relative positions of patches in the images. [19] SIFT flow (SIFTflow) [22] adopts the same idea as that of the optical flow, and replaces pixel intensity feature with a 128-d SIFT descriptor. The SIFT descriptor is calculated for every pixel of the image, and SIFT flow is therefore not a compact hash. The SIFTflow distance between two images is given by the minimum value of an energy function achieved by a best estimated SIFTflow vector field which is mainly smooth but also preserves spatial discontinuities. Earth mover's distance (EMD) measures the dissimilarity of two distributions by computing the minimum cost required to transform one distribution into the other. The RetargetMe work [19], following Pele and Werman [21], construct the distributions for a 5-parameter tuple using the spatial coordinates and L*, a*, and b* components of downsampled images in the L*a*b* colorspace. Two lower-level distance metrics are also examined. Edge histogram descriptor (EH) [42] captures the spatial distribution of edges by concatenating the 5-bin histograms (vertical, horizontal, 45° diagonal, 135° diagonal, and isotropic) of 16 subimages into an 80-bin descriptor. L1 distance between the descriptors of two images is calculated to quantify the dissimilarity. Color layout descriptor (CL) [43] is formed by stacking low frequency DCT coefficients of YUV channels together. The distance of two images is defined as a weighted L2 distance of the descriptors. The RetargetMe work also includes a random metric (RAND) for reference purpose: for any pair of images, RAND simply returns a distance value uniformly drawn from $(0, 1)$.

References

[1] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment*. Morgan & Claypool Publishers, Mar. 2006.

[2] ——, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, Mar. 2002.

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

[4] A. B. Watson, "DCTune: A technique for visual optimization of DCT quantization matrices for individual images," in *SID Digest of Technical Papers*, vol. XXIV, 1993, pp. 946–949.

[5] P. C. Teo and D. J. Heeger, "Perceptual image distortion," in *Proceedings of IEEE International Conference on Image Processing*, vol. 2, Nov. 1994, pp. 982–986.

[6] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor, "Visibility of wavelet quantization noise," *IEEE Transactions on Image Processing*, vol. 6, no. 8, pp. 1164–1175, 1997.

[7] Z. Wang and X. Shang, "Spatial pooling strategies for perceptual image quality assessment," in *Proceedings of IEEE International Conference on Image Processing*, Oct. 2006, pp. 2945–2948.

[8] E. C. Larson and D. M. Chandler, "Unveiling relationships between regions of interest and image fidelity metrics," in *Proceedings of SPIE Visual Communications and Image Processing*, 2008, p. 68222A.

[9] E. C. Larson, C. Vu, and D. M. Chandler, "Can visual fixation patterns improve image fidelity assessment?" in *Proceedings of IEEE International Conference on Image Processing*, Oct. 2008, pp. 2572 –2575.

[10] U. Engelke, V. X. Nguyen, and H.-J. Zepernick, "Regional attention to structural degradations for perceptual image quality metric design," in *Proceedings of IEEE International Conference on Acoustic, Speech, and Signal Processing*, Apr. 2008, pp. 869 –872.

[11] A. Rehman and Z. Wang, "Reduced-reference image quality assessment by structural similarity estimation," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3378–3389, Aug. 2012.

[12] A. Sharmir and O. Sorkine, *Visual Media Retargeting*. ACM SIGGRAPH Asia Courses, 2009.

[13] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," *ACM Transactions on Graphics*, vol. 26, no. 3, pp. 10:1–10:9, Jul. 2007.

[14] M. Rubinstein, A. Shamir, and S. Avidan, "Improved seam carving for video retargeting," *ACM Transactions on Graphics*, vol. 27, no. 3, pp. 16:1–16:9, Aug. 2008.

[15] ——, "Multi-operator media retargeting," *ACM Transactions on Graphics*, vol. 28, no. 3, pp. 23:1–23:11, Jul. 2009.

[16] L. Wolf, M. Guttmann, and D. Cohen-Or, "Non-homogeneous content-driven video retargeting," in *Proceedings of IEEE International Conference on Computer Vision*, Oct. 2007, pp. 1–6.

[17] Y.-S. Wang, C.-L. Tai, O. Sorkine, and T.-Y. Lee, "Optimized scale-and-stretch for image resizing," *ACM Transactions on Graphics*, vol. 27, pp. 118:1–118:8, Dec. 2008.

[18] P. Krähenbühl, M. Lang, A. Hornung, and M. Gross, "A system for retargeting of streaming video," *ACM Transactions on Graphics*, vol. 28, pp. 126:1–126:10, Dec. 2009.

[19] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir, "A comparative study of image retargeting," *ACM Transactions on Graphics*, vol. 29, no. 6, pp. 160:1–160:10, Dec. 2010.

[20] D. Simakov, Y. Caspi, E. Shechtman, and M. Irani, "Summarizing visual data using bidirectional similarity," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Anchorage, AK, Jun. 2008, pp. 1–8.

[21] O. Pele and M. Werman, "Fast and robust earth mover's distances," in *Proceedings of IEEE International Conference on Computer Vision*, Sep. 2009, pp. 460–467.

[22] C. Liu, J. Yuen, and A. Torralba, "SIFT Flow: Dense correspondence across scenes and its applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 978–994, May 2011.

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

[24] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509–522, 2002.

[25] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of interest point detectors," *International Journal of Computer Vision*, vol. 37, pp. 151–172, Jun. 2000.

[26] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proceedings of the 4th Alvey Vision Conference*, 1988, pp. 147–151.

[27] F. Ono, W. Rucklidge, R. Arps, and C. Constantinescu, "JBIG2-the ultimate bi-level image coding standard," in *Proceedings of IEEE International Conference on Image Processing*, vol. 1, Sep. 2000, pp. 140–143.

[28] H. G. Barrow, J. M. Tenenbaum, R. C. Bolles, and H. C. Wolf, "Parametric correspondence and chamfer matching: Two new techniques for image matching," in *Proceedings of the 5th International Joint Conferences on Artificial Intelligence*, 1977, pp. 659–663.

[29] A. Rosenfeld and J. L. Pfaltz, "Sequential operations in digital picture processing," *Journal of the ACM*, vol. 13, no. 4, pp. 471–494, Oct. 1966.

[30] C. Papadimitriou and K. Stieglitz, *Combinatorial Optimization: Algorithms and Complexity*. Prentice Hall, 1982.

[31] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 2000.

[32] J. Han, M. Kamber, and A. K. H. Tung, *Geographic Data Mining and Knowledge Discovery*. Taylor and Francis, 2001, ch. Spatial Clustering Methods in Data Mining: A Survey, pp. 1–29.

[33] R. Ng and J. Han, "CLARANS: A method for clustering objects for spatial data mining," *IEEE Transactions on Knowledge and Data Engineering*, vol. 14, no. 5, pp. 1003–1016, Sep. 2002.

[34] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[35] F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 567–585, Jun. 1989.

[36] F. L. Boostein, *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge Univ. Press, 1991.

[37] J. Duchon, "Splines minimizing rotation-invariant semi-norms in Sobolev spaces," in *Constructive Theory of Functions of Several Variables*, ser. Lecture Notes in Mathematics. Springer Berlin / Heidelberg, 1977, vol. 571, pp. 85–100.

[38] J. Meinguet, "Multivariate interpolation at arbitrary points made simple," *The Journal of Applied Mathematics and Physics (ZAMP)*, vol. 30, pp. 292–304, 1979.

[39] M. Rubinstein. (2013, Jul.) RetargetMe database. MIT Computer Science and Artificial Intelligence Laboratory. [Online]. Available: http://people.csail.mit.edu/mrub/retargetme/index.html

[40] R. Hooke and T. A. Jeeves, "Direct search solution of numerical and statistical problems," *Journal of the ACM*, vol. 8, no. 2, pp. 212–229, Apr. 1961.

[41] Y. Pritch, E. Kav-Venaki, and S. Peleg, "Shift-map image editing," in *Proceedings of IEEE International Conference on Computer Vision*, Sep. 2009, pp. 151–158.

[42] B. Manjunath, J.-R. Ohm, V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 6, pp. 703–715, Jun. 2001.

[43] E. Kasutani and A. Yamada, "The MPEG-7 color layout descriptor: A compact image feature description for high-speed image/video segment retrieval," in *Proceedings of IEEE International Conference on Image Processing*, vol. 1, Thessaloniki, Oct. 2001, pp. 674–677.