An Objective Quality of Experience (QoE) Assessment Index for Retargeted Images

Jiangyang Zhang, C.-C. Jay Kuo
Ming Hsieh Department of Electrical Engineering
University of Southern California
Los Angeles, CA, USA
jiangyaz@usc.edu, cckuo@sipi.usc.edu

ABSTRACT

Content-aware image resizing (or image retargeting) is a technique that resizes images for optimum display on devices with different resolutions and aspect ratios. Traditional objective quality of experience (QoE) assessment methods are not applicable to retargeted images because the size of a retargeted image is different from its source. In this work, three determining factors for humans visual QoE on retargeted images are analyzed. They are global structural distortion (G), local region distortion (L) and loss of salient information (S). Different features are selected to quantify their respective distortion degrees. Then, an objective quality assessment index, called GLS, is proposed to predict viewers' QoE by fusing selected features into one single quality score. Several regression models used for feature fusion are discussed and compared. Experimental results demonstrate that the proposed GLS quality index has stronger correlation with human QoE than other existing objective metrics in retargeted image quality assessment.

Categories and Subject Descriptors

I.4.7 [**Computation Methodologies**]: Image Processing and Computer Vision—*Feature Measurement*

Keywords

Image retargeting; Quality assessment; QoE assessment; Content-aware image resizing

1. INTRODUCTION

Content-aware image resizing (or image retargeting) is a technique that addresses the increasing demand to display image contents on devices of different resolutions and aspect ratios. Traditional resizing techniques do not meet this requirement since they either discard important information (e.g. cropping) or introduce visual artifacts by oversqueezing the content (e.g. homogenous scaling). The goal of image retargeting is to change the aspect ratio and the resolution of images while preserving its visually important content and avoiding noticeable artifacts.

Several content-aware image retargeting solutions have been proposed in the last 7-8 years. They can be classified into two types: discrete and continuous approaches [19]. A discrete approach resizes an image by removing unimportant pixel regions iteratively [1, 15, 18] while a continuous approach conducts resizing through non-uniform image warping [7, 21, 25]. Most previous work has demonstrated novelty in problem formulation and algorithmic design. However, evaluation on the performance of different retargeting methods remains to be ad hoc as most of them rely on simple visual comparison or small-scale user studies. Clearly, there is a need to develop a better methodology for evaluating all retargeting results in a systematic and quantitative way.

Recently, Rubinstein *et al.* [17] conducted a systematic study on eight state-of-the-art retargeting algorithms through a large scale user study. Besides collecting and analyzing subjective evaluation results, they evaluated the performance of six distances as possible objective measures for retargeted images. However, there exists significant disagreement between their chosen measures and subjective evaluation results. Thus, a better objective QoE assessment index for retargeted images is still in need.

Objective image QoE assessment indices have been extensively studied in the last decade [8, 22]. They can be divided into three categories: full-reference (FR), reduced-reference (RR) and no-reference (NR). However, traditional image QoE assessment indices are not applicable in the context of image retargeting for the following reasons. For FR and RR methods, one underlying assumption is that the size of the original image should be matched with that of the distorted image. Since the original and retargeted images differ significantly in sizes and aspect ratios, this assumption does not hold. Furthermore, a retargeted image should preserve as much important information in its original image as possible. Thus, referring to the original image is an indispensable part in evaluating a retargeted result which rules out NR methods.

In this paper, we attempt to address the QoE assessment issue and propose a novel objective index that accounts for three major determining factors for humans visual perception on retargeted images. These factors include: the global structural distortion (G), the local region distortion (L) and the loss of salient information (S). Various features are chosen to quantify their respective distortion degrees. Then, an objective quality assessment index, called GLS, is developed to predict viewers' QoE by fusing these features into one

MM'14, November 3–7, 2014, Orlando, Florida, USA.
Copyright 2014 ACM 978-1-4503-3063-3/14/11 ...$15.00.
http://dx.doi.org/10.1145/2647868.2654922.

quality score. In developing the GLS quality index, we compare several regression models in fusing multiple features. It is shown by experimental results that the proposed GLS index has stronger correlation with human QoE than other existing objective indices in retargeted image quality assessment. The effectiveness of new extracted features is the basis for the impressive performance gain of the proposed GLS index as compared with other existing QoE indices.

The rest of this work is organized as follows. Related previous work on image retargeting, distorted image quality assessment, and retargeting image QoE assessment is reviewed in Section 2. Three major distortion types for image retargeting are analyzed in Section 3. Then, the GLS assessment index for retargeted images is proposed in Section 4. Experimental results and related discussion are presented in Section 5. Finally, concluding remarks are given and future research directions are pointed out in Section 6.

2. REVIEW OF RELATED WORK

2.1 Image Retargeting

Content-aware image resizing methods can be classified into discrete and continuous approaches [19]. For the discrete approach, image resizing is achieved by identifying and discarding unimportant image contents. The cropping-based method [2] identifies the most prominent components in an image with saliency-based measures and cuts out a rectangular region as the desired retargeting result. This method clearly fails when there are two salient objects located at the two ends of an image. The seam carving method [1] resizes an image by iteratively removing paths of pixels with the least amount of saliency. This method may lead to local distortion of a salient object such as a human fact and yield an unpleasant result. Realizing that no single retargeting operator could perform well on all images, Rubinstein, Shamir and Avidan proposed the multi-operator method [18] that combines three different operators; namely, scaling, cropping and seam carving, to achieve a more robust result across a wide range of images. For the continuous approach, image retargeting is formulated as a global optimization problem [25, 21] where the salient image regions ought to be well preserved while non-salient regions are allowed to be squeezed or stretched.

2.2 Distorted Image QoE Assessment

Objective image QoE assessment, which studies the degree of quality degradation due to distortions such as additive noise, blurring and compression, is a hot research topic in recent years [8]. Early work evaluates degraded image quality using the pixelwise distortion measure such as the mean squared errors (MSE) and the peak signal-to-noise ratio (PSNR). Although the MSE and PSNR values are simple to calculate, they do not correlate with human subjective visual experience well. To overcome this problem, other quality indices have been proposed to account for characteristics of the human visual system (HVS). Examples include contrast sensitivity function (CSF) masking [3], just noticeable difference (JND) threshold [24], structure similarity (SSIM) [23], feature similarity (FSIM) [27], etc. One common assumption of the aforementioned full-reference image quality indices is that the size and the aspect ratio of the source image and its distorted one are the same. Since the size of a retargeted image is different from its source, these indices are not applicable to the QoE assessment of retargeted images.

2.3 Retargeted Image QoE Assessment

The first comparison study on retargeted image quality was presented in [17]. In this work, the authors conducted a large scale user study to compare the performance of eight representative state-of-the-art image retargeting methods. In addition to performance evaluations based on user responses, six objective image distance measures were evaluated and compared with actual human perception. Since none of these six measures was in well alignment with human ranking, there is a need to search for other retargeted image quality indices that can offer better agreement. A similar study was conducted by Ma *et al.* [12], in which a different image retargeting database was built and evaluated by human viewers. Several existing objective QoE measures were evaluated. It was concluded that a better quality index could be obtained by combining the shape distortion measure and the content information loss. Although in-depth performance analysis on QoE index design was conducted in [12] and [17], none of them offers a satisfactory retargeted image QoE index. Recently, Liu *et al.* [10] proposed an objective QoE assessment method for image retargeting using a top-down approach. Although the method is capable of measuring both local and geometric distortions using the SSIM index, it does not account for information completeness which will be shown to play an important role in the quality assessment of retargeted images.

In this work, we will start with identifying three dominant distortion types for image retargeting; namely, the global structural distortion (G), the local region distortion (L) and the loss of salient information (S). Then, features are extracted to characterize the severity of these distortions. Finally, we propose a GLS quality index that adopts a machine learning methodology to fuse all distortion features.

3. IMAGE RETARGETING DISTORTION ANALYSIS

In this section, we identify three main distortion types of image retargeting – global structural distortion, local region distortion and loss of salient information. The first two distortion types introduce visual unpleasing artifacts to retargeted results such as over-squeezing the object shapes (global) or breaking the prominent lines (local). The third type does not necessarily introduce visually noticeable artifacts, yet the retargeted result fails to preserve all salient information in the original image. Understanding the characteristics of these major distortion types would lay out a basis for the GLS quality index design, which will be elaborated in Section 4.

3.1 Global Structural Distortion

Global structural distortion occurs when an image is over-squeezed or over-stretched after retargeting, leading to unpleasing shape deformation of prominent objects. This distortion is especially noticeable when the salient object is improperly deformed and/or different parts of a salient object are deformed unproportionally, leading to inconsistency as compared with the original image. This type of distortion produces artifacts at the global scale.

Both discrete and continuous image retargeting methods could potentially produce global structural distortion. For

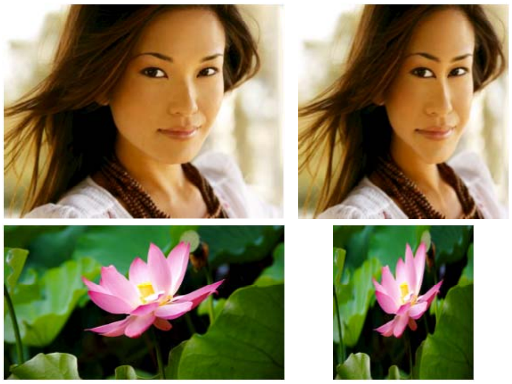


Figure 1: Two examples of global structural distortion. Upper row: the original image of *face* (left) and the retargeted result by [1] (right). Lower row: the original image of *lotus* (left) and the retargeted result by scaling (right).

example, as shown in Fig. 1, the global structure of the prominent object in the *face* image is heavily distorted because the relative positions of eyes, nose and mouth are misaligned after retargeting. However, the shape of each individual face component (eyes, nose, mouth, etc.) are kept intact. In other words, there is no distortion in local regions. For image *lotus*, there is also heavy global structural distortion as the prominent object (flower) is over-squeezed after retargeting.

3.2 Local Region Distortion

After retargeting, some local regions in the image may be heavily distorted, especially near those regions with prominent edges. Discrete retargeting methods, such as seam carving [1], may introduce broken lines when the removed region overlaps with a prominent edge region. On the other hand, continuous retargeting methods may result in heavy edge bending when the underlying mesh behind the edge region undergoes significant warping. The local region distortions become less noticeable at regions with homogeneous textures, (e.g. sky, surface, wall, etc.) and irregular textures (e.g. trees, grass, sand, etc.) [26]. We show an example of heavy local region distortion using the seam carving method in Fig. 2. Prominent edges are heavily bended after retargeting since many of the removed seams passed through the edge of pencils.

3.3 Loss of Salient Information

Besides reducing visual artifacts, a good retargeted image should be able to include all important content in its original image as much as possible. Loss of salient information is a distortion type commonly introduced by discrete operators such as cropping. When the salient object is too large and/or spans across the whole image, cropping will inevitably discard some important information. For example, as shown in Fig. 3(a), there are four different buildings in the original image, each of them with similar visual importance. To retarget this image to half of its original width, a simple cropping method will inevitably discard some salient information, leaving only two of the buildings in the retargeted result. A better retargeting method for this image should be able to remove redundant information from each building but preserve all four buildings as shown in Fig. 3(b) (b). This type of distortion is less observable for continuous retargeting methods, since different regions of an image are scaled disproportionately without discarding pixels in them.

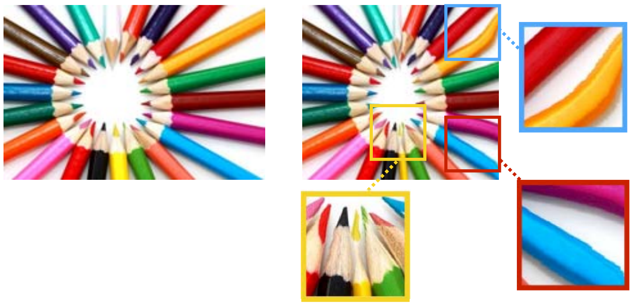
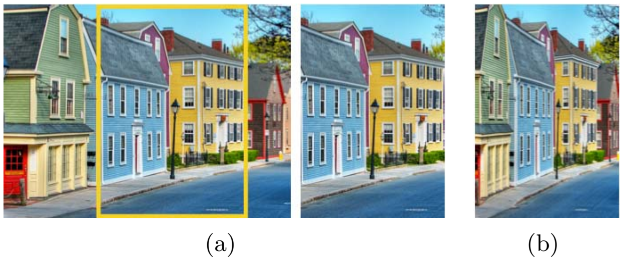


Figure 2: Illustration of local region distortion. Left: the original image of *pencil*. Right: the retargeted result using the seam carving method in [1] with three zoomed-in local regions.



(a) (b)

Figure 3: Illustration of loss of salient information: (a) the original image of *Marble* and its cropping result, where the yellow box shows the optimum cropping result. (b) a better retargeted result obtained by [18].

The objective quality indices examined in [10, 17] primarily focus on measuring distortions (global or/and local). However, they do not pay much attention to the importance of information completeness. In contrast, we attempt to find good features to characterize all three distortion types and combine them into one single score for quantitative evaluation of retargeted results in this work.

4. GLS QUALITY INDEX FOR RETARGETED IMAGES

In this section, we will introduce the proposed GLS quality index for retargeted images. Given the original image, I, and its retargeted results $\hat{I}_i$ (i=1,2,3,...), we would like to compute an objective quality score S_i for each result to achieve the following two objectives:

1. the relative rank of retargeted results is consistent with the subjective ranking;
2. the predicted quality score matches with subjective quality scores.

We will first introduce the overall framework in Section 4.1 and, then, describe each stage in detail in Sections 4.2-4.5.

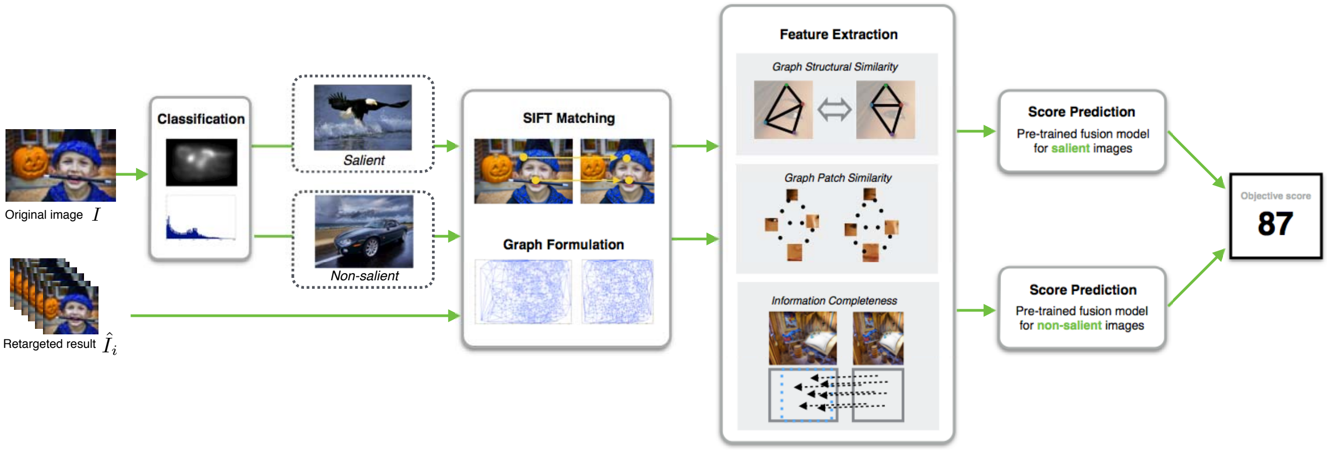


Figure 4: The system framework in computing the proposed GLS quality index for retargeted images.

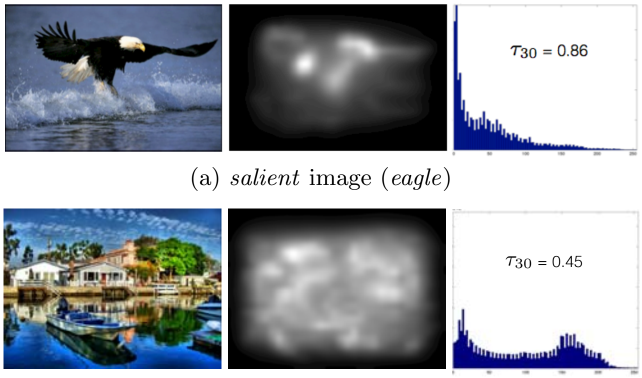
4.1 Overview of System Framework

The challenge of objective image QoE assessment lies in formulating effective features and fusing them into a single number to predict the quality score. In the proposed GLS scheme, we first conduct saliency analysis and SIFT feature mapping to determine whether I is a *salient* or *non-salient* image and build the mapping correspondence between I and $\hat{I}_i$. For each retargeted result $\hat{I}_i$, we extract features to quantify the three types of distortions as mentioned in Section 3. A pre-trained machine learning model is used to fuse all features into one single quality score as the final result. The machine learning model is trained using subjective evaluation results of existing image retargeting databases [12, 17]. Fig. 4 shows the overall system framework in computing the proposed GLS quality index.

4.2 Saliency-based Classification

The first step is to determine whether the source image, I, is a *salient* or *non-salient* image. If an image contains one salient object which does not cover the entire image, it is considered as salient. Otherwise, if all contents in I have equal visual importance or its salient object is too large and fills up the entire image, I is viewed as non-salient. This classification step is commonly known as data grouping in handling large-scale databases. The main purpose of image grouping is to separate images of different characteristics into multiple disjoint groups so that we can train different prediction models for them separately. This grouping process allows us to design a more accurate prediction model since there is a stronger correlation between the training and test images.

There are many algorithms proposed for saliency computation and, without loss of generality, we simply choose one and adopt the GBVS method [5] here. The salient image classification problem is conducted based on analyzing the histogram of the obtained saliency map, which takes a value ranging from 0 to 255. The saliency value "0" means the lowest saliency level (i.e., no saliency) and "255" means the highest saliency level (i.e. strongest saliency). As shown in Fig. 5, the saliency histogram of a typical salient image usually consisting of a steep peak followed by a quickly descending tail as shown in Fig. 5(a). On the other hand, for a typical non-salient image, the histogram usually has a low-rising peak and a slowly decaying tail as shown in Fig. 5(b). The canal boat image in Fig. 5(b) is classified to a non-salient one since its salient region is too large.



(a) *salient* image (*eagle*)

(b) *non-salient* image (*canal boat*)

Figure 5: Classification based on the saliency map histogram analysis for two representative images: the original image I (left), the saliency map (middle) and its histogram (right).

We define the percentage of pixels below brightness level x as

$$\tau_x = \frac{\sum_x h(x)}{N}, \quad x = 1, 2, \cdots, 256,$$

where $h(x)$ is the number of pixels at bin x in the histogram and N is the total number of pixels in image I. In our experiments, we adopt the following simple rule to decide whether an image is a salient one or not. If $\tau_{30} > \delta$ with threshold $\delta = 0.70$, it is a salient image. Otherwise, it is a non-salient one.

4.3 SIFT Mapping and Mesh Formulation

The next step is compute the mapping correspondence between SIFT features [11] of the original image, I, and its retargeted results, $\hat{I}_i$. This correspondence will help serve as the basis for feature extraction as elaborated in Section 4.4.

We first extract the SIFT features from I and match them with those of each retargeted image $\hat{I}_i$. We discard all SIFT features that are not successfully matched so that we have an equal number of SIFT features for each image pair, I and $\hat{I}_i$, at the end of the matching process. Then, we formulate two graphs for each image pair $(I, \hat{I}_i)$. Each vertex in the graph represents one matched feature in the original image. The graph formulation is completed by connecting all neighboring vertices using delaunay triangulation as shown in Fig. 6.

As a result, we associate each image pair, I and $\hat{I}_i$, with two graphs denoted by $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ and $\hat{\mathbf{G}}_\mathbf{i} = (\hat{\mathbf{V}}_\mathbf{i}, \hat{\mathbf{E}}_\mathbf{i})$, where $|\mathbf{V}| = |\hat{\mathbf{V}}_\mathbf{i}|$ and, for each vertex $\mathbf{v} \in \mathbf{V}$, there is a unique mapping $m(\mathbf{v}) = \hat{\mathbf{v}}_\mathbf{i}$, where $\hat{\mathbf{v}}_\mathbf{i} \in \hat{\mathbf{V}}_\mathbf{i}$. With such a mapping in place, we have converted the problem of measuring the distance between I and $\hat{I}_i$ to the problem of computing the graph similarity between $\mathbf{G}$ and $\hat{\mathbf{G}}_\mathbf{i}$.

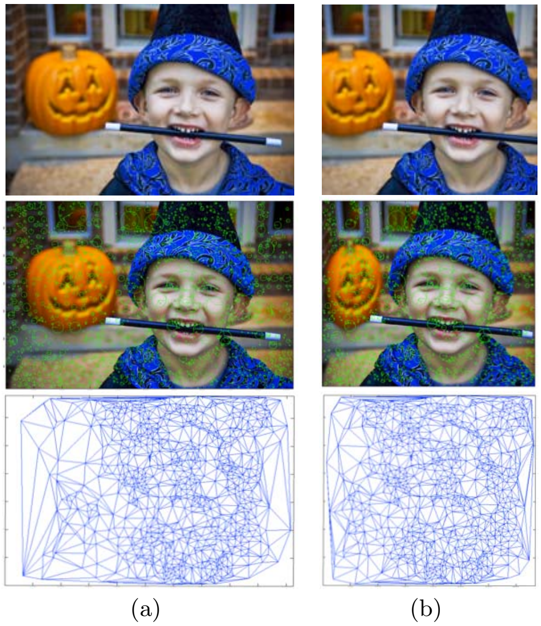


Figure 6: SIFT feature mapping and graph formulation between the original image and its retargeted result: (a) original image I (top), matched SIFT features (middle) and its formulated graph $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ (bottom), (b): retargeted result $\hat{I}_i$ (top), matched SIFT features (middle) and its formulated graph $\hat{\mathbf{G}}_\mathbf{i} = (\hat{\mathbf{V}}_\mathbf{i}, \hat{\mathbf{E}}_\mathbf{i})$ (bottom).

4.4 Extraction of Features

There are two key issues in objective image QoE assessment: 1) extraction and representation of appropriate features, and 2) pooling of features into one single number to represent quality score. We will address the first issue in this section and focus on the issue of feature fusion in Section 4.5.

4.4.1 Graph Structure Similarity

The graph structure similarity feature measures the amount of global structural distortion in the retargeted image. To compute this feature, we make use of the results in Section 4.3, where the problem of comparing image pair, I and $\hat{I}_i$, is reformulated as comparing the graph similarity between $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ and $\hat{\mathbf{G}}_\mathbf{i} = (\hat{\mathbf{V}}_\mathbf{i}, \hat{\mathbf{E}}_\mathbf{i})$.

If there is little global structural distortion during the retargeting process, the relative positions of $\hat{\mathbf{V}}_\mathbf{i}$ should be close to those of $\mathbf{V}$ and the shape of each triangle in $\mathbf{G}$ should be similar to the corresponding matched triangle in $\hat{\mathbf{G}}_\mathbf{i}$. As a result, the global structural distortion in $\hat{I}_i$ can be measured using shape deformation of each mesh triangle.

To measure the shape deformation of each triangle, we make use of the log-polar spatial representation scheme [16], which is computationally more efficient than methods such as RANSAC. It encodes the relative positions and orientations between each pair of nodes in the graph. Fig. 7 shows an example of 5-bit (32 regions) log-polar representations and its corresponding spatial and orientation codes. In our case, we use a spatial code with 8 bits, where the first 3 bits represent the relative orientation angle (quantized into 8 sectors) and the remaining 5 bits represents the relative distance (quantized into 32 levels).

The shape deformation between two triangles is measured using the modified inconsistency sum method [16]. That is, to compare the log-polar codes of nodes in the triangle, we compute the distance between two triangles T_k and $\hat{T}_k$ as

$$d_k = \sum_{i=1}^{3} C_{i,k} \otimes \hat{C}_{i,k},$$

where k and i are indices for triangles and its three nodes, respectively, and $C_{i,k}$ (i=1,2,3) are the codes for node i in triangle T_k, and $\otimes$ denotes the XOR operator. For the test images used in our experiments, the typical range for k is between 200 and 1000. If a triangle in $\mathbf{G}$ is perfectly matched with the corresponding triangle in $\hat{\mathbf{G}}_\mathbf{i}$, then $d_k = 0$. Then, we add up the distances for all triangle pairs and obtain the distance between two graphs, $\mathbf{G}$ and $\hat{\mathbf{G}}_\mathbf{i}$, as

$$f_1 = \sum_k d_k.$$

4.4.2 Graph Patch Similarity

To measure the degree of local region distortion in the retargeted result, we consider a feature called the graph patch similarity. If there is prominent local region distortion such as broken edges or edge bending, the patch difference should be significant.

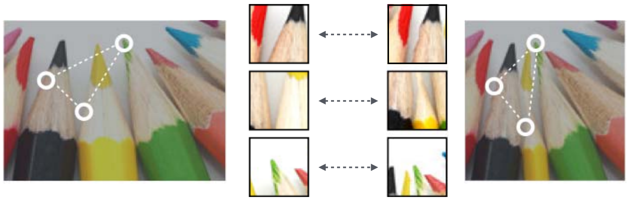
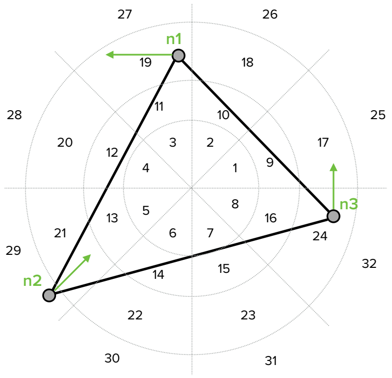


Figure 8: The degree of local region distortion is measured by using the graph patch similarity, where the Euclidean distance of local patches of matched graph nodes are computed and summed up over the entire graph.

For each image pair I and $\hat{I}_i$, we compare the similarity of local patches of size $N \times N$ centered around each node in graphs $\mathbf{G}$ and $\hat{\mathbf{G}}_\mathbf{i}$, where N is chosen to be 15 in our



Node	Position	Orientation
n1	19	π
n2	29	π/4
n3	24	π/2

Figure 7: Log-polar spatial representation scheme [16]. This example shows the 5-bit (32 regions) log-polar representation of a triangle: the position and orientation codes of each triangle node.

experiment. We use $p_{i,k}$ and $\hat{p}_{i,k}$ to denote patches centered at the node with index i of the triangle with index k in these two graphs, respectively. Then, the graph patch similarity feature can be computed as

$$f_2 = \sum_{k} \sum_{i=1}^{3} d(p_{i,k}, \hat{p}_{i,k}),$$

where $d(p_{i,k}, \hat{p}_{i,k})$ represents the Euclidean distance between patches $p_{i,k}$ and $\hat{p}_{i,k}$. We show an example of three matched features and the corresponding local patches for comparison in Fig. 8.

4.4.3 *Information Completeness*

The features of information completeness characterizes how well the retargeted image, $\hat{I}_i$, preserves the important content of the original image, I. When I is a salient image and contains a prominent object, preserving this salient object and its surrounding region is important at the expense of regions of lower saliency. However, all contents that possess similar importance for non-salient images, they should be preserved in a more uniform manner.

First, we need to determine a region in I that should be present in the retargeted image $\hat{I}_i$. This region, denoted by $\hat{P}_i$, is called the impact area of $\hat{I}_i$. To determine the impact region $\hat{P}_i$, we reverse-map all matched SIFT nodes in $\hat{I}_i$ back to I as shown in Fig. 9, and find a tight rectangular bounding box, which is denoted as $\hat{P}_i$. Based on the saliency map, we classify pixels in I into three regions: critical, important and ordinary as shown in Fig. 10. For a good retargeted result, its impact area $\hat{P}_i$ should contain as much critical and important regions as possible.

Then, to quantify the information completeness, we can define a feature that measures the amount of saliency value covered by the impact region. It is written as

$$f_3 = \alpha \cdot \frac{P_c}{N_c} + (1-\alpha) \cdot \frac{P_i}{N_i},$$

where N_i and N_c are the total numbers of pixels of important and critical regions and P_i and P_c are the total numbers of important and critical region pixels inside the impact area $\hat{P}_i$, respectively. The weighting parameter α is empirically chosen to be 0.70. The value of f_3 ranges from 0 to 1. If all the important and critical regions are encircled by the impact area, we have $f_3 = 1.0$.

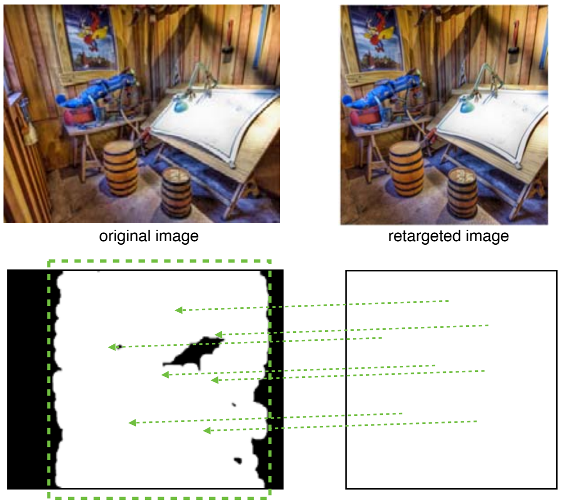


Figure 9: Illustration of the reverse mapping of the matched SIFT nodes in the retargeted image to the source image to compute the impact area $\hat{P}_i$ in the source image. The white region in the bottom left indicates a region where the underlying information can be found in the retargeted image while the black region means this part of information is lost after retargeting.

4.5 Feature Fusion and Model Selection

For effective image quality prediction, not only is the feature selection important but also the mechanism to fuse all features into one single quality score. There is no straightforward solution to feature fusion since the contribution of each feature to the final quality score may be different and is difficult to determine. A few basic pooling methods can be employed, including simple summation, multiplication and linear combinations of features. However, all these methods implicitly make assumptions on the relative importance of each feature, and there is lack of convincing ground for the assumptions.

In the proposed GLS quality index, we take advantage of the subjective human evaluation results and employ the machine learning technique to find a mapping function between the features discussed in Section 4.4 and the final quality score.

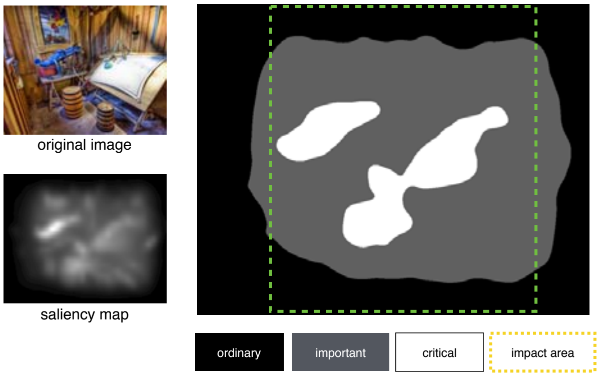


Figure 10: Illustration of the information completeness feature computation: the original image (left top), the saliency map (left bottom) and the segmentation of the saliency map into the critical region (white), the important region (gray) and the ordinary region (black), and the impact area $\hat{P}_i$ encircled by the green dash line.

In addition to features f_1, f_2 and f_3, we consider three more auxiliary features. They are:

- f_4: the total number of matched SIFT node pairs between I and $\hat{I}_i$;
- f_5: the average matching strength of all matching SIFT node pairs;
- f_6: the average matching strength of top 50 matched SIFT node pairs.

These three features measure how good the matching is between I and $\hat{I}_i$. The matching strength represents the distance between each matching feature pair. For each retargeted result, we extract the six features $\{f_1, f_2, ..., f_6\}$ from the given image and normalize them to the range of $[0, 1]$.

To determine the optimal fusion rule, we conduct experiments with the following eight fusion methods and compare their performance in our current application context:

1. Direct feature addition (*add*)
2. Direct feature multiplication (*multi*)
3. Linear regression (*lin*)
4. Logistic regression (*log*)
5. Logistic regression with L_1 penalty (*log-L1*)
6. Support vector regression with linear kernel (*svr-lin*)
7. Support vector regression with polynomial kernel (*svr-pol*)
8. Support vector regression with RBF kernel (*svr-rbf*)

Note that the last six of the above eight fusion methods are based on machine learning.

In the training phase, each candidate model is presented with a training set $\{f_p, y_p\}$ and the model parameters are estimated. The training sets are obtained from subjective evaluation results from existing public datasets, where f_p are the feature descriptors and y_p corresponds to the subjective score. We utilize the cross-validation scheme for each candidate model and choose the optimal model whose objective scores have the highest correlation with human subjective evaluation results. During the test phase, the trained optimal model is presented with the feature descriptors of the test image, and it predicts the estimated objective quality score.

5. EXPERIMENTAL RESULTS

5.1 Datasets

For the experiments, we make use of two public databases for image retargeting: the RetargetMe database [17] and the CUHK database [12].

The RetargetMe database contains 80 images, each with eight retargeted results obtained by eight methods: Nonhomogeneous Warping (WARP) [25], Seam-Carving (SC) [1], Scale-and-Stretch (SNS) [21], Multi-Operator (MULTI) [18], Shift-Map (SM) [15], Streaming Video (SV) [7], Cropping (CR) and Homogeneous Scaling (SCL). The subjective evaluation results on 37 images for all eight retargeting methods are provided in this database ($37 \times 8 = 296$ results). The evaluation was conducted with 210 human participants and scores were computed using pairwise comparison method, in which participants were shown two retargeted images side-by-side and were asked to choose their preferred ones.

The CUHK database contains 171 retargeted results from 57 image sources. In addition to the eight retargeting methods studied by [17], this database includes results from two more targeting methods; namely, the optimized seam carving and scale method [4] and the energy-based deformation method [6]. Unlike the pair-wise comparison scheme used in [17], the subjective evaluation in this study employed the 5-category discrete scale (”Bad”, ”Poor”, ”Fair”, ”Good” and ”Excellent”) to obtain the mean opinion scores (MOS) of viewers for each retargeted result.

5.2 Test Methodology

For both databases, we employ the 10-fold cross-validation method to evaluate the performance of the proposed GLS quality index. That is, the data is equally divided into ten parts: one chunk is used for testing and the remaining nine parts are used for training. The experiment is repeated with each of the ten chunk used for testing. The averaged accuracy of the test based on all ten chunks is taken as the final performance measure.

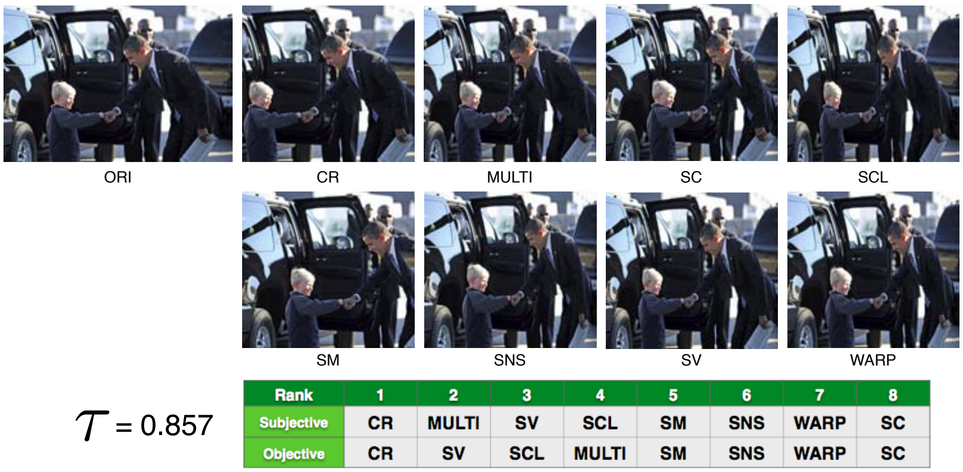
5.3 Comparison of Feature Fusion Methods

For performance evaluation, we consider the following five metrics: 1) the Kendall rank coefficient τ, 2) the Pearson linear correlation coefficient r, 3) the Spearman rank order correlation coefficient ρ, 4) the root mean square error ($RMSE$) between subjective and objective quality scores, and 5) the outliers ratio (OR).

For a perfect match between the objective and subjective scores (or rank), we have

- $\tau = 1.0$
- $r = 1.0$
- $\rho = 1.0$
- $RMSE = 0$
- $OR = 0$

Table 1 shows the Kendall rank coefficient of the eight different fusion methods as discussed in Section 4.5 for the RetargetMe database. Since the subjective evaluation for



$\mathcal{T}$ = 0.857

Rank	1	2	3	4	5	6	7	8
Subjective	CR	MULTI	SV	SCL	SM	SNS	WARP	SC
Objective	CR	SV	SCL	MULTI	SM	SNS	WARP	SC

Figure 11: An example where the proposed GLS index is strongly correlated with the subjective rank (with Kendall rank coefficient $\tau = 0.857$). The original and eight retargeted images of the *Obama* image and the corresponding subjective rank [17] and the objective rank computed using the GLS index are shown.

this database is based on the pairwise comparison, it is difficult to determine the mean opinion score (MOS) and we evaluate the performance of these fusion methods with the Kendall rank coefficient only. We see from Table 1 that the machine-learning-based fusion methods perform significantly better than simple feature addition or multiplication. In particular, the logistic regression fusion method outperforms all others. Furthermore, adding the L_1 penalty further improves the Kendall rank coefficient from 0.355 to 0.382.

Table 1: Kendall rank coefficient τ of different fusion models for RetargetMe database [17]

	add	*multi*	*lin*	*log*	*log-L1*	*svr-lin*	*svr-pol*	*svr-rbf*
τ	0.058	0.095	0.301	0.355	**0.382**	0.307	0.308	0.306

Table 2 compares the performance of different fusion methods for the CUHK database. Since the mean opinion scores (MOS) are provided in this database, we can conduct performance comparison using multiple metrics. As shown in Table 2, the logistic regression (yet without L_1 penalty) outperforms all other fusion methods under almost all performance metrics except for SROCC, where the linear regression is slightly better than the logistic regression.

5.4 Comparison of Objective Quality Indices

Table 3: Performance comparison of five objective image QoE indices for RetargetMe database [17]

	BDS[20]	EH[13]	SIFT-Flow[9]	EMD[14]	GLS
τ	0.083	0.004	0.145	0.251	**0.382**

In this section, we compare the proposed GLS quality index with four other objective QoE indices for retargeted images. They are:

- Bidirectional Similarity (BDS) [20]
- Edge Histogram (EH) [13]
- SIFT-Flow [9]
- Earth Mover Distance (EMD) [14]

Table 3 and Table 4 compare the performance among all five QoE indices for the RetargetMe and the CUHK databases, respectively. We see from the experimental results that the proposed GLS index performs better than all existing QoE indices by a significant margin in all four performance metrics.

5.5 Discussion

The proposed GLS index outperforms all other existing QoE indices for two main reasons. First, the GLS index design is based upon three dominant distortion types for image retargeting as discussed in Section 3. The other quality indices consider only one or two of these distortion types but none of them consider all three together. For example, EH [13], SIFT-Flow [9] and EMD [14] do capture the global structural distortion and the local region distortion of retargeted images, but fail to consider the information completeness factor. The BDS index [20] measures information completeness in a bidirectional way, but it fails to consider either global or local distortions that occurred in the retargeting result fully.

The second reason that explains the good performance of the proposed GLS index is that the machine-learning technique is adopted to fuse features effectively to yield one final quality score. Although our feature design takes into consideration all three distortion types, determining relative weights of multiple features still remains a challenge. In the GLS index, we address this challenge by training a machine learning model that learns from existing subjective evaluation results and intelligently determines the optimal feature weights for each specific image. The more subjective evaluation results we have for the fusion model training, the better the predicted objective score for each retargeted result.

We offer further insights into the performance of the GLS quality index by examining two examples. We show the evaluation of eight retargeted results for image *Obama* in Fig. 11. This image contains two salient objects: President Obama and the boy. As shown in the subjective result,

Table 2: LCC, SROCC, RMSE and OR of different fusion models for CUHK database [12]

	lin	*log*	*log-L1*	*svr-lin*	*svr-pol*	*svr-rbf*
r	0.4402	**0.4622**	0.3961	0.3656	0.3711	0.3658
ρ	**0.4939**	0.4760	0.4002	0.4038	0.3821	0.3961
$RMSE$	12.204	**10.932**	14.026	12.894	13.259	13.212
OR	0.2046	**0.1345**	0.2163	0.2339	0.2022	0.2267

Table 4: Comparison with other objective image QoE metrics for CUHK database [12]

	BDS[20]	EH[13]	SIFT-Flow[9]	EMD[14]	GLS
r	0.2896	0.3422	0.3141	0.2760	**0.4622**
ρ	0.2887	0.3288	0.2899	0.2904	**0.4760**
$RMSE$	12.922	12.686	12.817	12.977	**10.932**
OR	0.2164	0.2047	0.1462	0.1696	**0.1345**

cropping performs the best since the two salient objects can perfectly fit into one cropping window and very little salient information is lost. On the other hand, seam carving [1] and warping [25] perform the worst as they introduce heavy global structural distortion (on President Obama) and local region distortion (the document in President's hand).

As shown in the rank order table of Fig. 11, the objective rank computed with the proposed GLS index correlates well with the subjective rank in [17]. The Kendall rank coefficient is equal to $\tau = 0.857$. The best one and the poorest four image retargeting methods identified by the GLS index are identical with subjective evaluation results. The only slight difference lies in the methods that are ranked from 2 to 4. The GLS index favored the result of the streaming video method [7] while the subjective evaluation ranks the multi-operator method [18] as the second.

However, there are individual cases where the proposed GLS index does not agree well with the subjective evaluation results. We show the evaluation results of the *Buddha* image in Fig. 12. For this case, there is a large discrepancy between the subjective and objective evaluation results with Kendall rank coefficient $\tau = -0.357$. The GLS index gives the highest preference to cropping, seam-carving [1] and warping [25]. However, these three are among the worst performing methods according to subjective evaluation results, thereby leading to a negative Kendall rank coefficient value. The mismatch between the objective and subjective ranks may be explained by the shortage of training data. In the training data set, there is no image similar to the Buddha image which contains the face of a human statue as opposed to an authentic human face. If similar cases are available in the training data or face-detection is added to the saliency detection module, we may expect the proposed machine-learning-based GLS index to learn from these cases and provide more accurate prediction.

6. CONCLUSION AND FUTURE WORK

In this paper, we proposed a novel objective quality of experience (QoE) index, called the GLS index, to evaluate image retargeting results. We first identified three key factors related to human perception on the quality of retargeted images. They are global structural distortion, local region distortion and loss of salient information. Using this knowledge as guidance, we found effective features that capture these distortion types and utilized machine learning to fuse all features into one single quality score. One major advantage of applying the machine learning tool is that the feature weights can be determined automatically. It was shown by experimental results that the proposed GLS index outperforms four other existing objective indices by a significant margin in all performance metrics of consideration.

Since the performance of the machine learning method will be improved with more training data, a larger database with more subjective evaluation results for image retargeting is desired. The prediction of objective scores will be greatly improved with a model trained with more complete data set that contains all distortion types. In addition, this work is mainly focused on evaluating retargeting results of images. Objective QoE assessment for retargeted video will be an important extension of the current work.

7. REFERENCES

[1] S. Avidan and A. Shamir. Seam carving for content-aware image resizing. In *ACM Transactions on graphics (TOG)*, volume 26, page 10. ACM, 2007.

[2] L.-Q. Chen, X. Xie, X. Fan, W.-Y. Ma, H.-J. Zhang, and H.-Q. Zhou. A visual attention model for adapting images on small displays. *Multimedia systems*, 9(4):353–364, 2003.

[3] B. Chitprasert and K. Rao. Human visual weighted progressive image transmission. *Communications, IEEE Transactions on*, 38(7):1040–1044, 1990.

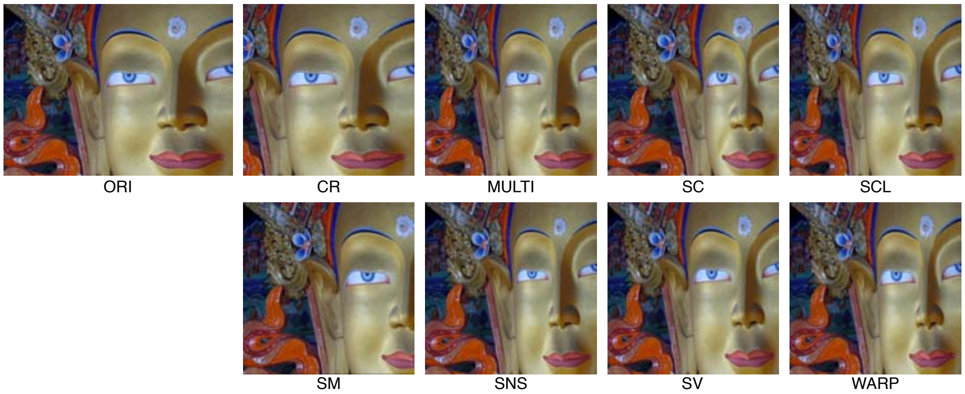
[4] W. Dong, N. Zhou, J.-C. Paul, and X. Zhang. Optimized image resizing using seam carving and scaling. *ACM Transactions on Graphics (TOG)*, 28(5):125, 2009.

[5] J. Harel, C. Koch, P. Perona, et al. Graph-based visual saliency. *Advances in neural information processing systems*, 19:545, 2007.

[6] Z. Karni, D. Freedman, and C. Gotsman. Energy-based image deformation. In *Computer Graphics Forum*, volume 28, pages 1257–1268. Wiley Online Library, 2009.

[7] P. Krähenbühl, M. Lang, A. Hornung, and M. Gross. A system for retargeting of streaming video. In *ACM Transactions on Graphics (TOG)*, volume 28, page 126. ACM, 2009.

[8] W. Lin and C.-C. Jay Kuo. Perceptual visual quality metrics: A survey. *Journal of Visual Communication and Image Representation*, 22(4):297–312, 2011.



τ = -0.357

Rank	1	2	3	4	5	6	7	8
Subjective	MULTI	SCL	SV	SM	CR	SNS	WARP	SC
Objective	CR	SC	WARP	SV	MULTI	SM	SNS	SCL

Figure 12: An example where the proposed GLS index is poorly correlated with the subjective rank (with Kendall rank coefficient $\tau = -0.357$). The original and eight retargeted images of the *Buddha* image and the corresponding subjective rank [17] and the objective rank computed using the GLS index are shown.

[9] C. Liu, J. Yuen, A. Torralba, J. Sivic, and W. T. Freeman. Sift flow: Dense correspondence across different scenes. In *Computer Vision–ECCV 2008*, pages 28–42. Springer, 2008.

[10] Y.-J. Liu, X. Luo, Y.-M. Xuan, W.-F. Chen, and X.-L. Fu. Image retargeting quality assessment. In *Computer Graphics Forum*, volume 30, pages 583–592. Wiley Online Library, 2011.

[11] D. G. Lowe. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. IEEE, 1999.

[12] L. Ma, W. Lin, C. Deng, and K. N. Ngan. Image retargeting quality assessment: a study of subjective scores and objective metrics. *Selected Topics in Signal Processing, IEEE Journal of*, 6(6):626–639, 2012.

[13] B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada. Color and texture descriptors. *Circuits and Systems for Video Technology, IEEE Transactions on*, 11(6):703–715, 2001.

[14] O. Pele and M. Werman. Fast and robust earth mover's distances. In *Computer vision, 2009 IEEE 12th international conference on*, pages 460–467. IEEE, 2009.

[15] Y. Pritch, E. Kav-Venaki, and S. Peleg. Shift-map image editing. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 151–158. IEEE, 2009.

[16] S. Purushotham, Q. Tian, and C.-C. J. Kuo. Picture-in-picture copy detection using spatial coding techniques. In *Proceedings of the 2011 ACM international workshop on Automated media analysis and production for novel TV services*, pages 25–30. ACM, 2011.

[17] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir. A comparative study of image retargeting. In *ACM transactions on graphics (TOG)*, volume 29, page 160. ACM, 2010.

[18] M. Rubinstein, A. Shamir, and S. Avidan. Multi-operator media retargeting. In *ACM Transactions on Graphics (TOG)*, volume 28, page 23. ACM, 2009.

[19] A. Shamir and O. Sorkine. Visual media retargeting. In *ACM SIGGRAPH ASIA 2009 Courses*, page 11. ACM, 2009.

[20] D. Simakov, Y. Caspi, E. Shechtman, and M. Irani. Summarizing visual data using bidirectional similarity. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[21] Y.-S. Wang, C.-L. Tai, O. Sorkine, and T.-Y. Lee. Optimized scale-and-stretch for image resizing. In *ACM Transactions on Graphics (TOG)*, volume 27, page 118. ACM, 2008.

[22] Z. Wang and A. C. Bovik. Modern image quality assessment. *Synthesis Lectures on Image, Video, and Multimedia Processing*, 2(1):1–156, 2006.

[23] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612, 2004.

[24] A. B. Watson and J. A. Solomon. Model of visual contrast gain control and pattern masking. *JOSA A*, 14(9):2379–2391, 1997.

[25] L. Wolf, M. Guttmann, and D. Cohen-Or. Non-homogeneous content-driven video-retargeting. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–6. IEEE, 2007.

[26] J. Zhang and C.-C. Kuo. Region-adaptive texture-aware image resizing. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pages 837–840. IEEE, 2012.

[27] L. Zhang, D. Zhang, and X. Mou. Fsim: a feature similarity index for image quality assessment. *Image Processing, IEEE Transactions on*, 20(8):2378–2386, 2011.