EUROGRAPHICS 2011 / M. Chen and O. Deussen
(Guest Editors)

Volume 30 (2011), Number 2

Image Retargeting Quality Assessment

Yong-Jin Liu[1], Xi Luo[1], Yu-Ming Xuan[2], Wen-Feng Chen[2], Xiao-Lan Fu[2]

[1]TNList, Department of Computer Science and Technology, Tsinghua University, P. R. China
[2]State Key Lab of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, P. R. China

Abstract

Content-aware image retargeting is a technique that can flexibly display images with different aspect ratios and simultaneously preserve salient regions in images. Recently many image retargeting techniques have been proposed. To compare image quality by different retargeting methods fast and reliably, an objective metric simulating the human vision system (HVS) is presented in this paper. Different from traditional objective assessment methods that work in bottom-up manner (i.e., assembling pixel-level features in a local-to-global way), in this paper we propose to use a reverse order (top-down manner) that organizes image features from global to local viewpoints, leading to a new objective assessment metric for retargeted images. A scale-space matching method is designed to facilitate extraction of global geometric structures from retargeted images. By traversing the scale space from coarse to fine levels, local pixel correspondence is also established. The objective assessment metric is then based on both global geometric structures and local pixel correspondence. To evaluate color images, CIE $L^*a^*b^*$ color space is utilized. Experimental results are obtained to measure the performance of objective assessments with the proposed metric. The results show good consistency between the proposed objective metric and subjective assessment by human observers.

Categories and Subject Descriptors (according to ACM CCS): I.3.0 [Computing Methodologies]: Computer Graphics—General I.4.10 [Computing Methodologies]: Image Processing And Computer Vision—Image Representation

1. Introduction

Image retargeting is a technique that adjusts input images into arbitrary sizes and simultaneously preserves the salient regions of the input images. The basic idea of image retargeting is to find an importance map of an input image, and expand (or shrink) the image using less important regions in the image, so that observers perceive few changes in the retargeted image.

Recently many retargeting techniques were proposed, including 1D [AS07] and 2D [ZCHM09] distortion diffusion methods, homogeneous [WTSL08] and non-homogeneous grid transformation methods [WGCO07], graph labeling method [PKVP09], patch match method [BSFG09] and optimal multi-operator combination method [RSA09], etc. Given these retargeting techniques, an evaluation metric to judge their qualities is useful for a wide range of retargeting applications. Ideally assessment by human beings with normal color vision is suitable for this task. However, subjective assessments such as mean opinion scores (MOSs) metric is time-consuming and expensive. An objective assessment providing computational models to measure the perceptual quality of images is therefore much desired.

Objective image assessment has been extensively studied [PS00, WB06]. Notably three categories exist: full reference (FR), reduced reference (RR) and no reference (NR). Assuming that original images have perfect quality, FR methods require full access to original images as references. RR methods only require partial information of original images for quality assessment. NR methods evaluate distorted images in a blind manner which is an extremely difficult task. Usually application-domain-knowledge (e.g., JPEG compression) must be provided for NR assessments. For both FR and RR methods, original images and distorted images are usually required to have the same size, and thus are not suitable for image retargeting assessment. If an NR method is applied to image retargeting, the information of original images is completely discarded and then the assessment may not be accurate as it could be. So a new objective assessment method for image retargeting should be developed.

In this paper, an objective quality assessment method for

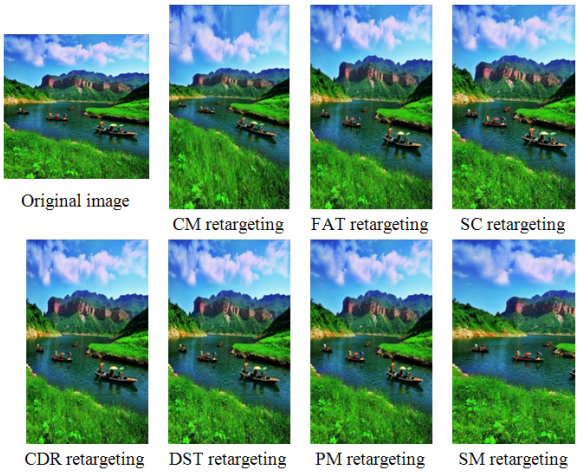

Figure 1: *An original image of size* 512×512 *is retargeted to the size of* 420×720 *using the methods: CM [ZCHM09], FAT [GSCO06], SC [AS07], CDR [WGCO07], DST [ZHM08], PM [BSFG09] and SM [PKVP09].*

image retargeting is proposed. The method is based on a global topological property and image scale space is used to extract this global structure in an efficient way. The extension to color images is also discussed. Experimental results show that the objective quality values closely match the subjective scores evaluated by observers, indicating that our proposed objective metrics are congruent with human perception mechanism.

2. Related work

2.1. Image retargeting

To fit arbitrary image sizes of different aspect ratios, the image retargeting techniques reduce/expand images by automatically removing/adding less-important image portions while keeping important features intact. This property is useful to create images of different sizes, adapted for state-of-the-art display devices (mobile, PDA and TV, etc.) that usually have widely differing resolutions.

Many image retargeting techniques have been proposed [AS07, BSFG09, KLHG09, PKVP09, RSA09, WGCO07, WTSL08, ZCHM09]. These works are different in two aspects: (1) how to define an image importance map; (2) how to propagate the distortion from less-important regions of retargeted images with low perceptual error. The seam carving algorithm [AS07] performed horizontal and vertical retargeting separately and thus the distortion propagation is non-homogeneous. The scale-and-stretch warping algorithm [WTSL08] used a quad-mesh to characterize the overall image structure and for retargeting, important quads are constrained to scale homogeneously. By minimizing the bending energy of grid lines, this warping algorithm distributed distortion in all directions. A similar algorithm using sparse linear system solver was proposed in [WGCO07]. An elegant conformal distortion energy was introduced in [ZCHM09] to diffuse distortion in all directions and preserve well important edges in the image. By observing that no single retargeting operator outperforms others in all the cases, a novel algorithm finding an optimal combination of different operators is proposed in [RSA09]. Figure 1 illustrates seven different retargeting methods.

The importance map plays a crucial role in image retargeting. Low level vision techniques have been used to define the importance map. Five energy functions, including L_1 and L_2-norm of the gradient, combined corner and edge detectors, neuron-response-like saliency measure [IKN04], and eye gaze measurement, are used in [AS07] for assigning different weights to image pixels. Face and motion detectors are also used to auxiliarily define the importance map in [WGCO07]. In PatchMatch system [BSFG09], users can interactively specify constraints in the retargeting process. With these different importance maps, the above retargeting methods give rise to a wide variety of retargeting distortions. An objective quality assessment for image retargeting is therefore useful to predict perceived image quality.

2.2. Image quality assessment

Image quality assessment is a fundamental issue in both computer and human vision [Gre98]. An obvious and accurate way is the subjective assessment based on the human perception. A widely used subjective assessment computes mean opinion scores (MOSs) from the human ratings [Esk01,ITU02]. But this method is time-consuming and not suitable for practical use. Objective assessments by computer programs whose evaluations are in close agreement with human judgement have been extensively studied in the past.

Early work about objective assessments characterized the similarity of two images of same size using peak signal-to-noise ratio (PSNR) and mean squared errors (MSE) [Mar86]. Although PSNR and MSE are simple to calculate, it is well known that they are not well matched to perceived visual quality [Gir93, EF95]. Later the error-sensitivity-based metrics were comprehensively extended by considering more characteristics of human vision system (HVS), such as decomposing signals in subband channels [Bra99, SFAH97], contrast sensitivity function (CSF) masking [CR90], just noticable difference (JND) threshold and normalization [WS97] and choosing an appropriate color space for HVS [PW93]. A fundamental different framework [WBSS04], called SSIM, was proposed based on the assumption that HVS is highly adapted for extracting statistic structural information. Experimental results show that SSIM is one of most successful metrics for image quality assessment (IQA).

If full information in original images is needed for IQA, it is referred to as FR-IQA. RR-IQA and NR-IQA require partial and none information in original images, respectively. Currently most FR-IQA methods studies distorted images that have the same size of original images. So they are not suitable for retargeted image assessment. In our study, the retargeted images to be evaluated range from natural scenes to artificial objects such as buildings and cartoons. Since state-of-the-art RR-IQA and NR-IQA mainly utilizes strong hypotheses such as JPEG compression degradation [MDWE04, WSB02], natural image statistics [LW09, SO01] or image multiscale geometric information [GLTL09], a simple extension of RR-IQA is not feasible to meet our task. In this paper we propose a new IQA that utilizes the full information in original images to evaluate the perceived quality of retargeted images.

3. Basic framework

The original and distorted images studied in previous IQA methods usually have the same size, e.g., in the applications of image compression, network communication, printing, displaying and restoration, etc. Given two images of the same size, there is a natural one-to-one correspondence that map one pixel in an image to the pixel at the same position in another image. In this case, most existing IQA methods start from low level vision that works with extraction of certain physical properties at the pixel level [Mar82]. Then the perceptual process is mimicked by assembling pixel-level features in a local-to-global manner [Bie87]. We regard this class of methods as working in a bottom-up manner, i.e., the image features are detected and organized from local to global viewpoints.

We argue that the bottom-up manner is not suitable for image retargeting quality assessment, since retargeted images have quite different aspect ratios and humans usually observe global structure changes before comparing subtle changes pixel by pixel. In this paper we propose to use a reverse order (top-down manner) that organizes image features from global to local viewpoints, leading to a new image retargeting IQA method. Experimental results show that our IQA method performs well for image retargeting quality assessment. The success of our top-down-manner IQA method is consistent with the postulate [Che82] in cognitive science that *the human visual system is sensitive to global topological properties and extraction of global topological properties is a basic factor in perceptual organization*.

3.1. A practical algorithm

At an abstract level, the proposed method can be outlined in two steps. First, a correspondence between the global geometric structures of two images is established, which characterizes the global topological properties in original and retargeted images. Let the global geometric structure of an image be depicted by an adjacent graph $G(V,E)$: each node in V corresponds to a salient region and there is an edge in E if its two nodes are sufficiently close. Given two images of different sizes or even different aspect ratios, the global topological property of two adjacent graphs G_1, G_2 is given by finding maximal subgraphs in G_1, G_2 that is isomorphic. Secondly, given the pixel correspondence in two images (or node mapping between G_1 and G_2), we compute the similarity between local windows of corresponding pixels in two images. Let the global topological similarity of two graphs G_1, G_2 be S_{global} and the sum of the similarity of local window correspondence be S_{local}. The perceived quality of retargeted images is measured by a weighted combination of S_{global} and S_{local}.

To establish graphs G_1, G_2 and their correspondences for the two images, we utilize scale-space theory [Lin94] that is a basic tool to analyze the global geometric structure in an image. Given a hierarchical view of an image scene, to find the topological properties in G_1, G_2 by establishing the graph correspondence, we need a robust and stable operator that extracts distinct invariant features from images and performs reliable mapping between feature points. There is a considerable body of previous work on local invariant feature detection in scale space [TM08], in which we use the scale invariant feature transform (SIFT) [Low04], since it achieves the state-of-the-art performance. For similarity measure between local windows of pixels in two images, we implement one perceptual error measure [WLB95] and two matrix-based error measures [WBSS04, SGE06]. In our experiments all the three measures have the similar performance and we choose the structure similarity (SSIM) [WBSS04] in our approach, since it performs consistently well with the human perception of quality in a variety of measures and is simple for programming.

3.2. Algorithmic details

Our proposed method is based on the scale invariant feature transform (SIFT) [Low04]. Given an original image I_{ori} and a retargeted image I_{ret}, two scale spaces $SP(I_{ori}) = \{I^0_{ori}, I^1_{ori}, \cdots, I^n_{ori}\}$, $SP(I_{ret}) = \{I^0_{ret}, I^1_{ret}, \cdots, I^n_{ret}\}$ of I_{ori} and I_{ret} are constructed, respectively, with the same Gaussian convolution kernel. The scale space $SP(I)$ is then converted to a difference-of-Gaussian space $DoG(I) = \{D^0 = I, D^i(x,y,\sigma) = (G(x,y,k\sigma) - G(x,y,\sigma)) * I^{i-1}(x,y), i = 1, \cdots, n\}$. The parameter k will be specified in Section 3.4. Distinctive invariant feature points (DIFPs) are detected in both $SP(I_{ori})$ and $SP(I_{ret})$ using the local extrema detection method in [Low04]. The attributes of each DIFP include location, scale and orientation. If a DIFP is detected at scale i, its counterparts are recorded in images at all scales $0, 1, \cdots, i$.

The principle of the practical top-down matching method is as follows. Assume that a correspondence from pixels of D^{i+1}_{ori} to pixels of D^{i+1}_{ret} has been established. The correspondence at scale i is established in both intra- and inter-scale

manners. First, if DIFPs exist in both D^i_{ori} and D^i_{ret}, each DIFP pair (p_{DIFP}, q_{DIFP}), $p_{DIFP} \in D^i_{ori}$ and $q_{DIFP} \in D^i_{ret}$, is matched and evaluated using the local image descriptor (LID) in [Low04]. This offers the intra-scale constraints. The inter-scale constraints are achieved by propagating pixel pair matching from coarse scale $(D^{i+1}_{ori}, D^{i+1}_{ret})$ to fine scale (D^i_{ori}, D^i_{ret}) in the following way.

Given a matched pair (p^{i+1}, q^{i+1}) at level $i+1$, it offers a constraint that defines two 5×5 local windows w^i_p and w^i_q in D^i_{ori} and D^i_{ret}, respectively. All pixels in w^i_p are matched to all pixels in w^i_q using the SSIM metric and form a small local bipartite graph G^i_{pq}. The restriction of pixel correspondence in a local window inherited from a coarser scale correspondence implicitly imposes an overall geometric structure in a hierarchical manner. All the intra- and inter-scale constraints contribute to a large, sparse bipartite graph G^i between D^i_{ori} and D^i_{ret}. The edges in G^i are further pruned using a scale-dependent threshold T^i of matching cost. The value of T^i is specified in Section 3.4. Note that in this process the DIFPs and ordinary pixels have different characteristics and thus two different metrics, LID [Low04] and SSIM [WBSS04], are used to match them respectively. To equalize the contributions of LID and SSIM measures, both LID and SSIM matching costs are normalized to $[0,1]$.

To start up the above process, at the coarsest scale, the intra-scale constraints are first established. In more detail, two images D^n_{ori} and D^n_{ret} are matched using the SSIM metric for each pixel pair (p,q), $p \in D^n_{ori}$ and $q \in D^n_{ret}$. Given the dense bipartite graph G^n, a correspondence between pixels of D^n_{ori} and D^n_{ret} is established by finding a maximum cardinality matching in G^n with the Hungarian method that maximizes the total value of matching cost.

At the end of the hierarchical constraint-matching propagation process, a many-to-many mapping between pixels in I^0_{ori} and I^0_{ret} is established at the finest scale 0. This mapping can again be interpreted as a bipartite graph G_{geo_struct} that serves as the correspondence of two geometric structures in I^0_{ori} and I^0_{ret}. The similarity of two images I^0_{ori} and I^0_{ret} is defined as the similarity of two geometric structures measured as a weighted summation of edge-matching costs in G_{geo_struct}. A simplified, non-weighted similarity metric is given by:

$$Sim(I^0_{ori}, I^0_{ret}) = \frac{\#_{ver}}{pn(I^0_{ori})+pn(I^0_{ret})} \cdot \frac{1}{\#_{edge}} \cdot \sum_{i=1}^{\#_{edge}} SSIM(v_0(e_i), v_1(e_i)) \quad (1)$$

where $pn(I)$ is the number of pixels in image I, $e_i \in G_{geo\_struct}$, $\#_{ver}$ and $\#_{edge}$ is the number of vertices and edges in G_{geo_struct} respectively, $v_0(e_i), v_1(e_i)$ are two vertices of e_i, and $SSIM(\cdot)$ is the SSIM metric in [WBSS04] using a local 8×8 square window. The more similar I^0_{ori} and I^0_{ret} are, the more correspondences between pixels of I^0_{ori} and I^0_{ret} and the weight $\frac{\#_{ver}}{pn(I^0_{ori})+pn(I^0_{ret})}$ is closer to 1. The value of $Sim(\cdot)$ ranges between $[0,1]$. Given two identical images, their similarity is maximized to be 1.

One hypothesis on the human vision system is that its intermediate or high level process seems to selectively focus on salient regions [KU95, NK98]. Not all pixels in images have the same saliency. We use the saliency-based visual attention model in [IKN04] to compute the saliency of two images I^0_{ori} and I^0_{ret}. For each pixel in salient regions, if there is no corresponding pixels in the other image, we link it to a dummy vertex dv of that image in G_{geo_struct} and set $SSIM(\cdot, dv) = 0$. This gives rise to a modified, saliency-based graph SG_{geo_struct}. For each edge in SG_{geo_struct}, if one of its vertices is in a salient map, its weight is set to be

$$w_s = \frac{pn(I^0_{ori}) + pn(I^0_{ret}) + C}{pn(I^{salience}_{ori}) + pn(I^{salience}_{ret}) + C}$$

where $I^{salience}_{ori}$ and $I^{salience}_{ret}$ are the salient regions in I^0_{ori} and I^0_{ret}, respectively, and C is a small constant that prevents denominator very close to zero. The image size we handled in experiments is in the magnitude of 10^5, and we use the scale 10^{-4} of the image size, i.e., $C = 10$, in the experiments in this paper. If the area of salience regions is small, the weight w_s is large. If all pixels in images are salient, the weight is minimized to be one. For the remainding edges in SG_{geo_struct}, the weight is set to be one. The saliency-based similarity metric is given by:

$$SalSim(I^0_{ori}, I^0_{ret}) = \frac{\#_{ver}}{pn(I^0_{ori})+pn(I^0_{ret})} \cdot \frac{1}{\sum_{i=1}^{\#_{edges}} w_i} \cdot \sum_{i=1}^{\#_{edge}} w_i \cdot SSIM(v_0(e_i), v_1(e_i)) \quad (2)$$

where w_i is the weight of edges in SG_{geo_struct}.

3.3. Color space transformation

For color images, most traditional IQA methods use a common approach that separates the luminance component from the color information and uses the luminance channel only [WB06]. Despite its simplicity, this approach only takes partial information of a color image into account. Color has been shown to play an important role in the visual process [TWW01]. Neurophysiological experiments have revealed that color is a useful cue for scene organization and understanding [GR00]. In our application scenario, a color space whose metric has uniform visual perception is desired. We use the CIE $L^*a^*b^*$ space based on opponent-color coordinates, which is computed via simple formulas from CIE XYZ but is more perceptually uniform than several well-known color spaces such as RGB, HSV and CIE XYZ.

Given the saliency-based similarity metric (2) for gray-level images, we use the following color metric for retargeting evaluation:

$$ColSim(C^0_{ori}, C^0_{ret}) = w_L SalSim(L^{*0}_{ori}, L^{*0}_{ret}) + w_a SalSim(a^{*0}_{ori}, a^{*0}_{ret}) + w_b SalSim(b^{*0}_{ori}, b^{*0}_{ret}) \quad (3)$$

In all experiments in this paper, we use $w_L = w_a = w_b = 1/3$.

Figure 2: *Image retargeting and pixel correspondence. Top row: image retargeting using seam carving method [AS07]. Middle row: exact pixel correspondence between original* (392×320) *and retargeted* (196×320) *images. Bottom row: pixel correspondence output from the proposed scale space method. Only* 0.1% *correspondence is presented.*

3.4. Parameter settings

To establish pixel correspondence, two scale spaces are built up in the algorithm proposed in Section 3.2, for original and retargeted images, respectively. Lowe [Low04] shows that the value of multiplicative factor k in kernel $G(x,y,k\sigma)$ is insensitive on the stability of local extrema detection. In our algorithm, we use $k = \sqrt{2}$, since by the property $G(x,y,\sqrt{m^2+n^2}) = G(x,y,m) \otimes G(x,y,n)$, we can use a fixed kernel $m = n = 1$ to iteratively smooth the blurred images.

We denote the size of the original image by $a \times b$ and the size of retargeted image by $c \times d$. We denote the minimum number among a,b,c,d being m. We set the octave number s in both scale spaces by satisfying $\lfloor \frac{m}{2^s} \rfloor \leq w$, where w is the minimum width or height at the top level of scale spaces. The larger w is, the more pixel correspondence relies on the maximum cardinality matching. The smaller w is, the more pixel correspondence relies on the scale space transformation.

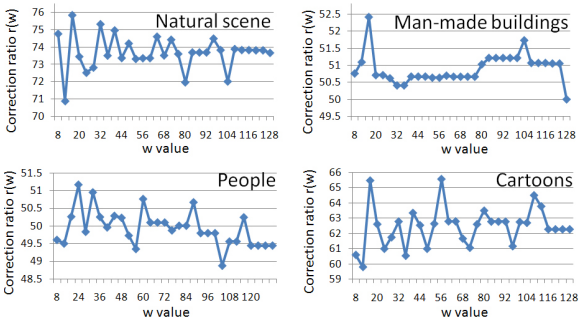

Figure 3: *The correction ratio* $r(w)$ *using the seam carving method. The results are averaged on 100 retargeting tests selected from the classes of natural scenes, man-made buildings, people and cartoons, respectively.*

The experimental determination of an optimal value w is presented below. If we know the exact algorithm for retargeting, we can determine pixel correspondence exactly. Let EPC be the exact pixel correspondence between original image and retargeting image, we can compute the correction ratio of pixel correspondence $PC(w)$ output from our scale-space method using value w. The correction ratio $r(w)$ is defined as:

$$r(w) = \frac{\#correct\_PC(w)}{\#EPC} \cdot \frac{\#correct\_PC(w)}{\#PC(w)} \cdot 100\%$$

where $\#correct\_PC(w)$ is the number of correct correspondence in $PC(w)$, $\#(EPC)$ is the total number of the exact pixel correspondence in EPC, and $\#PC(w)$ is the total number of correspondence in $PC(w)$. Since $\#correct\_PC \leq \min\{\#EPC, \#PC\}$, $r(w)$ ranged between $[0,1]$. In the ideal case $\#correct\_PC = \#EPC = \#PC$, $r(w) = 100\%$.

The results of correction ratio $r(w)$ for different w using the seam carving method are shown in Figure 3 and other retargeting methods [AS07, WGCO07, WTSL08, ZCHM09] exhibit similar pattern as in Figure 3. The results are obtained by average on a collection of 40 real images which are selected from classes of natural scenes, man-made buildings, people and cartoons. All the testing original images have size of 512×512, and retargeting images have sizes that are uniformly randomly sampled (URS) in $[256,1024] \times [256,1024]$. Each class has 10 original images and each image is retargeted 10 times by URS. Totally 100 results are averaged for each class. From the results, we conclude that the correction ratio of pixel correspondence is not seriously dependent on the value w. For the value of w ranging from 8 to 128, the correction ratios $r(w)$ in the classes of natural scenes, buildings, people and cartoons, ranged in $[70.8,75.8]$, $[50.0,52.4]$, $[48.8,51.2]$ and $[59.8,65.6]$, respectively. Based on the results, we choose the optimal value

of w being 16, since it achieves a good balance between time and space (in terms of octave number) complexities. Although detailed retargeting algorithms can provide exact pixel correspondence, our scale-space matching method is developed for blind assessment of retargeted images without knowing how to retarget.

Another parameter needed to be specified in the algorithm is the scale-dependent threshold T^i that is used to prune the edges in G^i. Generally the larger the scale is, a smaller threshold should be used, so that more geometric structures can be preserved. At scale i, T^i is defined as $\alpha^{n-i} \cdot Median(G^i)$, where $Median(G^i)$ is the median of matching costs in all edges in G^i. In all our experiments, we use $\alpha = 1.15$.

4. Experiments

We conducted four experiments to examine the validity of the proposed metric (3) for objective assessment, using the MOS subjective assessments as the baseline. In Experiment 1, ten retargeted images were obtained from each of twelve original images, using the seam carving method [AS07]. Each retargeted image was evaluated against its original image. The purpose of Experiment 1 was to test whether the proposed metric was consistent with the subjective assessment.

Experiment 2 has the same procedure as Experiment 1, but with different performance of the proposed objective assessment algorithm. The purpose of Experiment 2 was to examine the necessity and sensibility of different components in the proposed algorithm.

In Experiment 3, twenty original images were selected from four types: natural scenes, buildings, people and cartoons. Each original image was retargeted to five images of the same size using seven different retargeting methods [AS07, BSFG09, GSCO06, PKVP09, WGCO07, ZCHM09, ZHM08]. Observers were required to select the best retargeted image for each original image and their voting results were compared to the ranking computed by the proposed objective metric. Experiment 3 was designed to test whether the proposed objective metric was sensitive to different retargeting methods.

Experiment 4 has a similar procedure as in Experiment 3, but using a larger set of benchmark images [RGSS10] in which the subjective data is collected from 210 participants. Six objective metrics [KY01, LYT*08, MOVY01, PW09, RSA09, SCSI08] are also compared to the objective metric (3) in Experiment 4. The purpose of Experiment 4 is to validate the objective metric (3) over a large sample size, such that chance variation will be ruled out and more confidence can be achieved in the statistical results.

4.1. Experiment 1

Participant. In this experiment, twelve paid university students (5 females and 7 males) with normal or corrected-to-normal visual acuity and normal color vision participated in the study. They were all novel to the test.

Apparatus and stimuli. The experiments were run on a Pentium-IV PC with a 17-inch monitor at a 1280×1024 resolution. The experiment generator E-Prime 1.2 [SEZ02] was used to control the stimuli presentation. The luminance was constant and moderate in the testing laboratory. Participants sat approximately 60cm from the screen. The identical apparatus were used in all experimental tests.

In this experiment, participants provided their rating scores according to the interval scales of "excellent", "good", "fair", "poor" and "bad", which is a variant of the ITU-R five-point quality scale [ITU02]. The rating scores for the five intervals of the scale were 1-5 (bad), 6-10 (poor), 11-15 (fair), 16-20 (good), and 21-25 (excellent).

From each of the 12 original images, 10 different retargeted images were obtained and totally 120 retargeted images were used. To prevent the impact of complicated retargeting methods, we used a simple implementation of the seam carving algorithm [AS07] to obtain the 120 retargeted images. In fact, in Experiment 3 presented below, it is shown that the objective quality metric itself is insensitive to different retargeting methods.

Procedure. This experiment consisted of a learning stage and a test stage. In the learning stage, a sequence of 11 images were presented on the center of the screen. Participants were told that the sequence began with the original image, followed by its retargeted images, and the loss of quality and information of the image increased gradually one by one. Each image stayed on the screen for 3 seconds.

In the test stage, a pair of images at full size is simultaneously displayed side by side against a completely black background on the screen, subtending $33.8^o \times 13.5^o$ visual angle at the view distance of 60cm. The left image of the pair was always the original image, while the right one was always a retargeted image of the left. The image pairs were displayed in random order. For each pair of images, participants were instructed to rate how well the right image had kept the fidelity compared to the left original image, and then write down their ratings on score sheets. Participants rated each retargeted image only once. All participants completed 120 trials in this experiment.

Results. The raw scores were first normalized by the mean and variance of scores for each participant. Then the score was converted into Z-scores [SEZ02] and the entire data set was rescaled to fill the range from 1 to 100. Mean opinion scores (MOSs) were then computed for each image, after removing outliers by the interval method in [ITU02]. Higher MOSs corresponded to higher image quality.

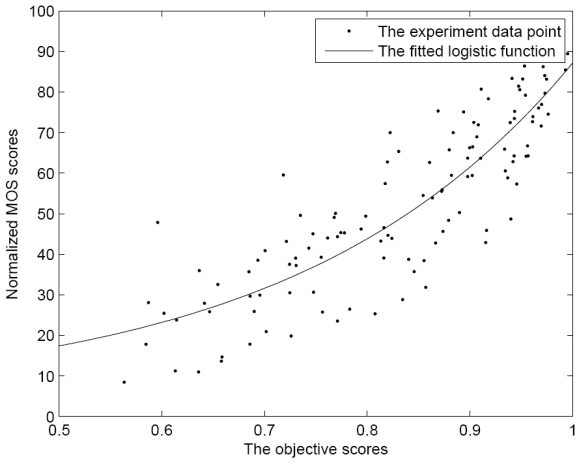

Figure 4: *The fitted curve using logistic function.*

To measure the performance of objective quality assessment models, the evaluation method proposed by video quality experts group [VQE00] is applied. A nonlinear mapping between the objective(X)/subjective(Y) scores is used with a logistic function:

$$Y = \frac{\beta_1 - \beta_2}{1 + e^{-\left(\frac{X-\beta_3}{|\beta_4|}\right)}} + \beta_2$$

where the initial estimates of parameters are:

$$\beta_1 = \max(Y),\ \beta_2 = \min(Y), \beta_3 = \text{mean}(X), \beta_4 = 1$$

An iterative process is evoked to find optimal $\beta_1, \beta_2, \beta_3, \beta_4$ using the SPSS software. With the nonlinear mapping, the fitting curve is shown in Figure 4 and four evaluation metrics were used. Metric A is the correlation coefficient between objective/subjective scores after variance-weighted regression analysis. Metric B is the correlation coefficient between objective/subjective scores after nonlinear regression analysis. Metric C is the Spearman rank-order correlation coefficient between the objective/subjective scores. Metric D is the outlier ratio of the predictions after the nonlinear mapping. The evaluation results are given in Table 1 (under Condition I), which demonstrates the consistency between subjective and objective measurements.

Discussion. In Experiment 1, Spearman rank-order correlation coefficient ($r = 0.868$) indicates that there is a high similarity between MOSs and ranking scores of the objective assessment, which means objective computed qualities of images, assessed by the metric (3) proposed in Section 3, are consistent with human subjective visual perception. The correlation coefficient of variance-weighted regression analysis ($r = 0.898$) and the correlation coefficient of nonlinear regression ($r = 0.868$) together provide the evidence that objective assessment scores can be used to predict the quality of retargeted images perceived by the human visual system

Condition	Correlation coefficients			(Outlier)
	Metric A	Metric B	Metric C	Metric D
I	0.898	0.868	0.868	0.008
II	0.635	0.518	0.515	0.108
III	0.795	0.715	0.713	0.042
IV	0.401	0.367	0.293	0.116
V	0.829	0.752	0.748	0.058

Table 1: *Four evaluation metrics reveal the consistency between subjective and objective measurements, with different conditions in the proposed algorithm.*

with high accuracy. Likewise, the outlier ratio ($p = 0.008$) indicates accurate prediction of objective assessment proved by regression analysis is consistent for any given retargeted image processed by the seam carving algorithm.

4.2. Experiment 2

Experiment 2 has the same apparatus and procedure as Experiment 1, but with new sets of participants. In the objective assessment algorithm proposed in Section 3, we refer to the full set of components below

I. SIFT-based matching, saliency-based filtering, SSIM-based local assessment and utilization of CIE $L^*a^*b^*$ color space

as the Condition I, which is measured in Experiment 1. For each new set of participants, the stimuli are obtained from an identical set of original images as in Experiment 1, by applying the proposed algorithm with one of the following conditions:

II. Match two images by maximizing the sum of SSIM local measures, the other conditions being the same as in I.
III. Ignore the saliency-based filtering, the other conditions being the same as in I.
IV. Use MSE (instead of SSIM) local assessment, the other conditions being the same as in I.
V. Convert color images into gray images, the other conditions being the same as in I.

The evaluation results are summarized in Table 1.

Discussion. The proposed assessment algorithm makes use of all the components in the full condition I. As demonstrated by evaluation results, the algorithm relies heavily on the correct matching and saliency-based filtering. If we use the strategy of maximizing the sum of SSIM local measures to replace the SIFT-based matching, all the correlation coefficients are low (r between $[0.50, 0.65]$) and the outlier ratio is increased ($p = 0.108$). The same conclusion is held if we simply ignore the saliency-based filtering: the correlation coefficients are also lower (in $[0.71, 0.80]$) and the outlier ratio is larger ($p = 0.042$) than those with full condition I. The other components (SSIM-based local assessment and CIE

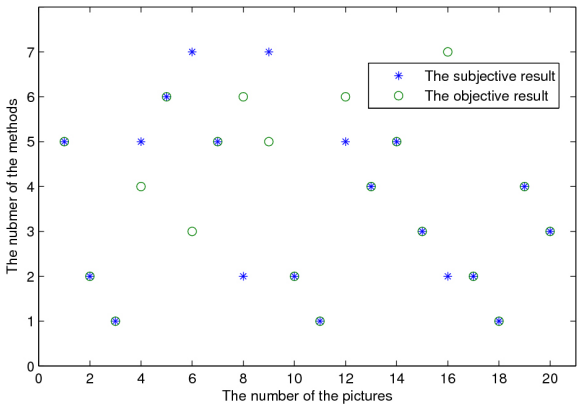

Figure 5: *Match of the best retargeted images chosen by the participants and ranked by the objective quality metric (3).*

$L^*a^*b^*$ color space) also have significant influence on objective assessment. The results show that using the full condition I achieves the best performance of retargeted image assessment.

4.3. Experiment 3

Participant. In this experiment, sixteen paid university students (7 females and 9 males) with normal or corrected-to-normal visual acuity and normal color vision participated in the study. They were all novel to the test.

Apparatus and stimuli. In this experiment, the experimental condition was the same as with Experiment 1, but the test images were different. From each of 20 original images, seven retargeted images were obtained using the seven different retargeting methods [AS07, BSFG09, GSCO06, PKVP09,WGCO07,ZCHM09,ZHM08]. The 20 original images were selected from four types according to their contents: natural scenes, buildings, people and cartoons:

- Natural scenes. They are well known to be statically redundant [SO01]: among all the visual cues, human subjects can only see a small fraction.
- Man-made buildings and roads. The contents of these images consist of different line segments, either intersected or parallel in a perspective view. The image contents also frequently show repeated patterns such as windows and balconies. Since human vision is sensitive to the abrupt changes in line segments and perspective relations between line segments [Bie87], building and road images can be good test stimuli.
- People. Human vision system has been proved to be sensitive to human faces, human bodies and human related characteristics [DBRC04]: For example, when human attention resources have been already occupied by a stimulus, compared to other task-irrelevant stimuli, the stimuli relating to human character will automatically snatch the attention resource from previous stimuli, and then jump into consciousness.
- Cartoons. They are quite different from the above three types of images. The cartoon contents show a strong non-photorealistic (NPR) fashion (see Figure 2 for an example). We use the cartoon images as the last class to test human vision sensitivity on man-made, NPR artistic figures.

Procedure. In this experiment, for each trial a sequence of 8 images was presented at the center of the screen. The first image was an original image. Participants were asked to choose the image of the best quality from the seven retargeted images following the original image. The image selected most frequently was considered as the best retargeted image from the original one. The participants could go back and forth to see any of the images in the sequence until he/she made a choice.

Results. The experimental results are shown in Figure 5, the best retargeted images chosen by the participants were consistent with the results ranked by the proposed objective quality metric. To be specific, in 20 sets of images, 14 best retargeted images selected from subjective and objective assessments were the same. In the six inconsistent cases, the retargeted images were so similar to each other that the participants had difficulty to decide which was the best. Therefore, their selections might be at random. In fact, in subjective assessment these images win as the best only had one or two more votes than the other images. Similarly in the objective assessment, the scores also showed minor differences between images, and the order exhibits certain randomness.

Discussion. The high match rate 70% (14 over 20) reveals that the objective assessment with the proposed objective quality metric (3) captures human perception well and is insensitive to different retargeting methods. Also observed from Figure 5, there does not exist a single retargeting method which achieves the best quality among all five methods. Given this diversity, the proposed objective quality metric (3) still matches human perception well.

4.4. Experiment 4

In this experiment, a public avaliable[†] benchmark of retargeting images [RGSS10] was used to validate the objective metric (3) with the comparision to seven other objective metrics. The *RetargetMe* benchmark collected subject evaluation from 210 participants of 37 images. Each image was retargeted by eight methods (simple scaling, cropping and six in [AS07,KLHG09,PKVP09,RSA09,WTSL08,WGCO07]). The scores of six objective metrics (BDS [SCSI08], BDW [RSA09], EH [MOVY01], CL [KY01], SIFTf [LYT*08] and EMD [PW09]) were also presented in *RetargetMe* as well as their correlations with the subjective data.

[†] http://people.csail.mit.edu/mrub/retargetme/

Procedure and results. For each image and its eight retargeted images, objective similarity scores were computed using metric (3). Given these objective scores, firstly a similar procedure as in Experiment 3 was performed by matching the best retargeted images chosen by the participants with the results ranked by objective metrics. In 37 sets of images, the match rate of metric (3) is 40.54%. As a comparison, the match rates of metrics BDS, BDW, EH, CL, SIFTf and EMD are 18.92%, 24.32%, 21.62%, 0.0%, 24.32%, 27.03%, respectively. Secondly, following the definition in [RGSS10], the rank correlation vector (Mean, std, p-value) of metric (3) was computed with the three highest rank results. The rank correlation vectors of metric (3), BDS, BDW, EH, CL, SIFTf and EMD are (0.400,0.3752,1e-4), (0.108, 0.532, 0.005), (0.200,0.395,0.002), (-0.071,0.593,0.013), (-0.320,0.543,1e-6), (0.298,0.483,1e-6) and (0.326,0.496,1e-6), respectively.

Discussion. The highest match rate 40.54% and the best rank correlation vector of metric (3) shows that metric (3) outperforms the metrics BDS, BDW, EH, CL, SIFTf and EMD. These results may be explained by that BDS, BDW, EH, CL, SIFTf and EMD only use low- and mid-level image characteristics, while metric (3) takes high-level perceptual organization into account. In particular, both BDS and BDW use the sum-of-square-differences of pixel values, which are well known that they do not match well to perceived visual quality [Gir93, EF95].

5. Conclusions

In this paper, a quality metric (3) for objectively assessing the quality of image retargeting is proposed. Different from traditional full-reference, reduced reference and no reference assessment methods, our proposed assessment is based on a top-down manner that first extracts the global geometric structures of two images and then establishes the detailed pixel correspondence for assessment. Elaborated experiments are developed, demonstrating that (1) the objective assessment made by the proposed metrics is consistent with human perception, and (2) the proposed objective metrics are insensitive to different retargeting methods. So the objective metrics proposed in this paper can be used to faithfully assess the performance of retargeting operations. For example, in some applications, many retargeting methods can be used and no single method is absolutely superior to others. When retargeting an image, we can use the proposed metric to assess the quality of different retargeted images, in order to choose the best method for that image.

Acknowledgements

The authors thank the reviewers for their comments that help improve this paper. The program of *PatchMatch*, *ShiftMap* and the *RetargetMe* benchmark are taken from original publication websites. The authors thank Mr. G.X. Zhang for implementing other retargeting methods in Experiment 3. This work was supported by the National Science Foundation of China (Project 60970099), the National Basic Research Program of China 2011CB302201 and Tsinghua University Initiative Scientific Research Program 20101081863.

References

[AS07] AVIDAN S., SHAMIR A.: Seam carving for content-aware image resizing. In *Proc. SIGGRAPH '07* (2007), Article No. 10.

[Bie87] BIEDERMAN I.: Recognition-by-components: a theory of human image understanding. *Psychological Review 94*, 2 (1987), 115–147.

[Bra99] BRADLEY A.: A wavelet visible difference predictor. *IEEE Trans. on Image Processing 8*, 5 (1999), 717–730.

[BSFG09] BARNES C., SHECHTMAN E., FINKELSTEIN A., GOLDMAN D.: Patchmatch: a randomized correspondence algorithm for structural image editing. In *Proc. SIGGRAPH '09* (2009), Article No. 24.

[Che82] CHEN L.: Topological structure in visual perception. *Science 218* (1982), 699–700.

[CR90] CHITPRASERT B., RAO K.: Human visual weighted progressive image transmission. *IEEE Trans. on Communication 38*, 7 (1990), 1040–1044.

[DBRC04] DOWNING P., BRAY D., ROGERS J., CHILDS C.: Bodies capture attention when nothing is expected. *Cognition 93*, 1 (2004), 27–38.

[EF95] ESKICIOGLU A., FISHRE P.: Image quality measures and their performance. *IEEE Trans. on Communications 43*, 12 (1995), 2959–2965.

[Esk01] ESKICIOGLU A.: Quality measurement for monochrome compressed images in the past 25 years. *J. Electron. Imaging 10*, 1 (2001), 20–29.

[Gir93] GIROD B.: What's wrong with mean-squared error. In *Digital Images and Human Vision* (1993), Cambridge, MA: MIT Press, pp. 207–220.

[GLTL09] GAO X., LU W., TAO D., LI X.: Image quality assessment based on multiscale geometric analysis. *IEEE Trans. on Image Processing 18*, 7 (2009), 1409–1423.

[GR00] GEGENFURTNER K., RIEGER J.: Sensory and cognitive contributions of color to the recognition of natural scenes. *Current Biology 10*, 13 (2000), 805–808.

[Gre98] GREGORY R.: *Eye and Brain: the psychology of seeing*. Oxford University Press, 1998.

[GSCO06] GAL R., SORKINE O., COHEN-OR D.: Feature-aware texturing. In *Proc. Eurographics Symposium on Rendering* (2006), p. 297IC303.

[IKN04] ITTI L., KOCH C., NIEBUR E.: A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence 20*, 11 (2004), 1254–1259.

[ITU02] ITU: *ITU-R Recommendation BT.500-11. Methodology for the subjective assessment of the quality of television images*. International Telecommunication Union: Geneva, 2002.

[KLHG09] KRAHENBUHL P., LANG M., HORNUNG A., GROSS M.: A system for retargeting of streaming video. In *Proc. SIGGRAPH Asia '09* (2009), Article No. 126.

[KU95] KOCH C., ULLMAN S.: Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology 4* (1995), 219–227.

[KY01] KASUTANI E., YAMADA A.: The mpeg-7 color layout descriptor. In *Proc. IEEE Int. Conf. Image Processing* (2001), pp. 674–677.

[Lin94] LINDEBERG T.: Scale-space theory. *Journal of Applied Statistics 21*, 2 (1994), 225–270.

[Low04] LOWE D.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision 60*, 2 (2004), 91–110.

[LW09] LI Q., WANG Z.: Reduced-reference image quality assessment using divisive normalization-based image representation. *IEEE Journal of Selected Topics in Signal Processing 3*, 2 (2009), 202–211.

[LYT*08] LIU C., YUEN J., TORRALBA A., SIVIC J., FREEMAN W.: Sift flow: Dense correspondence across different scenes. In *Proc. ECCV '08* (2008), pp. 28–42.

[Mar82] MARR D.: *Vision: a computational investigation into the human representation and processing of visual information*. San Francisco, W.H. Freeman, 1982.

[Mar86] MARMOLIN H.: Subjective mse measures. *IEEE Trans. on Systems, Man, and Cybernetics 16*, 3 (1986), 486–489.

[MDWE04] MARZILIANO P., DUFAUX F., WINKLER S., EBRAHIMI T.: Perceptual blur and ringing metrics: application to jpeg2000. *Signal Process. Image Commun. 19* (2004), 163–172.

[MOVY01] MANJUNATH B., OHM J., VASUDEVAN V., YAMADA A.: Color and texture descriptors. *IEEE Trans. on Circuits and Systems for Video Technology 11*, 6 (2001), 703–715.

[NK98] NIEBUR E., KOCH C.: Computational architectures for attention. In *The Attentive Brain* (1998), Cambridge, Mass., pp. 163–186.

[PKVP09] PRITCH Y., KAV-VENAKI E., PELEG S.: Shift-map image editing. In *Proc. ICCV '09* (2009), pp. 151–158.

[PS00] PAPPAS T., SAFRANEK R.: Safranek. perceptual criteria for image quality evaluation. In *Handbook of Image and Video processing* (2000), New York: Academic.

[PW93] POIRSON A., WANDELL B.: Appearance of colored patterns: pattern-color separability. *Journal of the Optical Society of America A 10*, 12 (1993), 2458–2470.

[PW09] PELE O., WERMAN M.: Fast and robust earth mover's distances. In *Proc. ICCV '09* (2009).

[RGSS10] RUBINSTEIN M., GUTIERREZ D., SORKINE O., SHAMIR A.: A comparative study of image retargeting. In *Proc. SIGGRAPH Asia '10* (2010), Article No. 160.

[RSA09] RUBINSTEIN M., SHAMIR A., AVIDAN S.: Multi-operator media retargeting. In *Proc. SIGGRAPH '09* (2009), Article No. 23.

[SCSI08] SIMAKOV D., CASPI Y., SHECHTMAN E., IRANI M.: Summarizing visual data using bidirectional similarity. In *Proc. CVPR '08* (2008), pp. 1–8.

[SEZ02] SCHNEIDER W., ESCHMANN A., ZUCCOLOTTO A.: *E-Prime User's Guide: Psychology Software Tools*. Pittsburgh, PA, 2002.

[SFAH97] SIMONCELLI E., FREEMAN W., ADELSON E., HEEGER D.: Shiftable multi-scale transforms. *IEEE Trans. on Information Theory 38*, 2 (1997), 587–607.

[SGE06] SHNAYDERMAN A., GUSEV A., ESKICIOGLU A.: An svd-based grayscale image quality measure for local and global assessment. *IEEE Trans. on Image Processing 15*, 2 (2006), 422–429.

[SO01] SIMONCELLI E., OLSHAUSEN B.: Natural image statistics and neural representation. *Annual Review of Neurosicnece 24* (2001), 1193–1216.

[TM08] TUYTELAARS T., MIKOLAJCZYK K.: Local invariant feature detectors: a survey. *Foundations and Trends in Computer Graphics and Vision 3*, 3 (2008), 177–280.

[TWW01] TANAKA J., WEISKOPH D., WILLIAMS P.: The role of color in high-level vision. *Trends in Cognitive Science 5*, 5 (2001), 211–215.

[VQE00] VQEG: *Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment*. 2000. http://www.vqeq.org/.

[WB06] WANG Z., BOVIK A.: *Modern Image Quality Assessment*. New York: Morgan & Claypool, 2006.

[WBSS04] WANG Z., BOVIK A., SHEIKH H., SIMONCELLI E.: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. on Image Processing 13*, 4 (2004), 600–612.

[WGCO07] WOLF L., GUTTMANN M., COHEN-OR D.: Non-homogeneous content-driven video-retargeting. In *Proc. ICCV '07* (2007), pp. 1–6.

[WLB95] WESTEN S., LAGENDIJK R., BIEMOND J.: Peceptural image quality based on a multiple channel hvs model. In *Proc. Int. Conf. Acoustics, Speech, Signal Processing* (1995), vol. 4, pp. 2351–2354.

[WS97] WATSON A., SOLOMON J.: Model of visual contrast gain control and pattern masking. *Journal of the Optical Society of America A 14*, 9 (1997), 2379–2391.

[WSB02] WANG Z., SHEIKH H., BOVIK A.: No-reference perceptual quality assessment of jpeg compressed images. In *Proc. IEEE Int. Conf. Image Processing* (2002), pp. 477–480.

[WTSL08] WANG Y., TAI C., SORKINE O., LEE T.: Optimised scale-and-stretch for image resizing. In *Proc. SIGGRAPH Asia '08* (2008), Article No. 118.

[ZCHM09] ZHANG G., CHENG M., HU S., MARTIN R.: A shape-preserving approach to image resizing. In *Proc. Pacific Graphics '09* (2009), pp. 1897–1906.

[ZHM08] ZHANG Y., HU S., MARTIN R.: Shrinkability maps for content-aware video resizing. In *Proc. Pacific Graphics '08* (2008), pp. 1797–1804.