Signal Processing: Image Communication 56 (2017) 12–19

Contents lists available at ScienceDirect

Signal Processing: Image Communication

journal homepage: www.elsevier.com/locate/image

# Learning quality assessment of retargeted images

Bo Yan[a,*,1], Bahetiyaer Bare[a], Ke Li[a], Jun Li[a], Alan C. Bovik[b,1]

[a] School of Computer Science, Shanghai Key Laboratory of Intelligent Information Processing, Fudan University, Shanghai 200433, China
[b] Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712-1084, USA

ARTICLE INFO

*Keywords:*
Image retargeting
Machine learning
RBF neural network
CW-SSIM
SIFT
Image aesthetics

ABSTRACT

Content-aware image resizing (or image retargeting) enables images to be fit to different display devices having different aspect ratios while preserving salient image content. There are many approaches to retargeting, although no "best" method has been agreed upon. Therefore, finding ways to assess the quality of image retargeting has become a prominent challenge. Traditional image quality assessment methods are not directly applicable to image retargeting because the retargeted image size is not same as the original one. In this paper, we propose an open framework for image retargeting quality assessment, where the quality prediction engine is a trained Radial Basis Function (RBF) neural network. Broadly, our approach is motivated by the observation that no single method can be expected to perform well on all types of content. We train the network on ten perceptually relevant features, including a saliency-weighted, SIFT-directed complex wavelet structural similarity (CW-SSIM) index, and a new image aesthetics evaluation method. These two features and eight other features are used by the neural network to learn to assess the quality of retargeted images. The accuracy of the new model is extensively verified by simulations.

## 1. Introduction

As the dimensions and sizes of display devices on mobile phones and tablets continues to diversify, image retargeting has become an important way to adjust original images to fit user-defined resolutions while simultaneously preserving important relationships between objects and picture content. The general idea is to shrink image regions of less importance while causing as little salient change as possible [1].

Recent image retargeting methods can be classified into two main types: discrete and continuous methods. Discrete methods including seam carving [2,3] and shift map [4], which remove and shift pixels of an image. Continuous methods including scale-and-stretch [5] and warping [6], which are processed on a quad mesh and merge pixels of an image. Other types of approaches, such as [7], apply a sequence of operators such as seam carving, scaling and cropping, objectively optimizing each step of the process. Given the rapid growth and use of automatic image retargeting techniques, it is becoming more necessary to also develop methods to automatically ensure the high quality of the resulting retargeted images.

If the perceptual quality of an image is high, then the retargeted image should also be obtained with good quality. Thus a perceptual evaluation process is needed to aid image retargeting. While human subjective judgments of images, including retargeted images, are the most reliable assessment, these are time consuming and difficult to obtain. Although highly effective image quality models exist [8–11], they require considerable modification to be able to be applied to assess the quality of retargeted images. This motivates the development of special-purpose objective retargeted image assessment models.

In [12], Rubinstein *et al*. conducted a large scale subjective study to compare eight state-of-the-art retargeting algorithms. They made algorithm comparisons against a variety of objective distance metrics, including a color layout descriptor, bidirectional warping and an edge histogram, to assess the quality of retargeted images. The authors of [13] also built a subject image retargeting database that includes 171 images produced by several representative retargeting methods on 57 natural source image contents. Each image has a mean opinion score (MOS) drawn from the subjective ratings of at least 30 viewers.

The authors of [14] measured global geometric structures and local pixel correspondences to evaluate the visual quality of retargeted images, using an objective metric. The method is top-down, organizing the features from global to local scales. The method proposed in [15] creates a SSIM [8] quality map that indicates, at each spatial location of the reference image, how well the structural information is preserved in a corresponding retargeted image. A saliency map is generated to spatially weight a computed SSIM map to estimate the visual quality of a retargeted image. In [16], three factors predictive of human judge-

* Corresponding author.
*E-mail address:* byan@fudan.edu.cn (B. Yan).
[1] This work is supported by NSFC (61522202 and 61370158).

http://dx.doi.org/10.1016/j.image.2017.04.005
Received 28 May 2016; Received in revised form 6 April 2017; Accepted 6 April 2017
Available online 09 April 2017

ments of the visual quality experienced when viewing retargeted images were analyzed. In [17], Hsu et al. proposed a full-reference objective metric for assessing the visual quality of a retargeted image, based on measurements of perceptual geometric distortion and information loss. This method is highly predictive of the subject quality of retargeted images, but when it is applied on images containing redundant, repeated texture patterns or very smooth areas, the SIFT flow [18] algorithm used to establish correspondences between the original and retargeted images may fail. In [19], Zhang et al. developed an aspect ratio similarity (ARS) metric to evaluate the visual quality of retargeted images by measuring local block changes. This method achieved state-of-the-art performance on two image retargeting quality assessment databases. Although many approaches have been devised for assessing the quality of retargeted images, no single method has achieved good results on the database in [12]. Given the generality and complexity of the problem, modern machine learning based methods offer a possible way to achieve acceptable performance [20].

The aesthetic value of a retargeted image is also relevant to image retargeting quality assessment. In [21], Tong et al. developed a method of image classification to identify whether photos were professional or amateur snapshots. They deployed a large set of heuristic low-level features. In [22], Murray et al. built a large scale database to assist with the development of aesthetic visual analysis models, called AVA. It contains more than 250,000 images, each supplied with a subjective aesthetic score and a semantic label from over 60 categories. The authors of [23–25] designed perceptual features which they used to create predictive algorithms for image aesthetics assessment. Ke et al. [23] proposed a principled method based on high level semantic features to determine perceptual aesthetic differences. Datta et al. [24] extracted 56 attributes from photos which they used to train an SVM classifier to classify the photos into two categories. They obtained an accuracy of approximately 70%. Cerosaletti and Loui [25] designed a few image features for image aesthetics prediction which they demonstrate on 450 photos. Jiang et al. [26] used the same data to train and verify their model. Many existing techniques for image aesthetics evaluation seek to classify photos into only two classes. Certainly, since retargeting affects image aesthetics as well as (distortion related) quality, it is of great interest to find ways to automatically assign aesthetics scores to retargeted images. The relationships between quality assessment and aesthetics evaluation on retargeted images is also potentially quite interesting. In [27], Liang et al. used aesthetics evaluation as part of a method of quality assessment on retargeted images. Only two aesthetic features were used: one descriptive of the rule of thirds, and the other of visual balance. However, many other aesthetics features have been developed that could be used in this application.

Here we describe an open framework for image retargeting quality assessment that builds on several existing quality related features, a novel aesthetics evaluation method, and a new saliency weighted CW-SSIM feature. We use these features to train a radial basis function (RBF) neural network to predict retargeted image quality. Ten of the features used as input to the neural network are the scalar outputs of independent image retargeting quality assessment methods. The experimental results show that the integrated method outperforms all of these ten methods.

The main contributions of our work are: i) We propose an open framework for image retargeting quality assessment; ii) that embodies a novel image retargeting quality assessment model which combines CW-SSIM [28], SIFT [29] and image saliency [30]; iii) and that also embodies a novel no-reference image aesthetics quality assessment method designed for retargeted images.

The rest of the paper is organized as follows. In Section 2, ten features are introduced including two important new ones, a saliency-weighted, SIFT-directed CW-SSIM feature and an integrated aesthetics feature. The radial basis function (RBF) network and the training process are described in Section 3. In Section 4, we present experimental results. The training and testing data are presented and the results of our integrated method are demonstrated and compared with existing retargeting assessment approaches. Finally, we draw conclusions in Section 5.

## 2. Quality-aware retargeting features

The success of any quality assessment model depends heavily on the features being used. Here we describe two new and effective features: a method of applying CW-SSIM to retargeted images by a SIFT-directed mapping process, which weights the CW-SSIM values using a saliency model, and a new method of evaluating aesthetics of retargeted images. We also briefly summarize the other eight existing retargeting assessment features.

### 2.1. Saliency-weighted, SIFT-directed CW-SSIM feature

SSIM is a classic method of measuring the perceptual similarity between two images which greatly improves upon traditional methods such as PSNR and MSE. The CW-SSIM index [28] is an application of the SSIM method in the complex wavelet domain. Thus, it is a version of SSIM that accords with cortical models of processing [31].

An advantage of the CW-SSIM index is that it possesses a useful degree of shift-invariance. However, CW-SSIM requires the ostensibly distorted image to have the same resolution as the original image, hence CW-SSIM cannot be implemented directly for image retargeting assessment. To handle this problem, we deploy the scale invariant feature transform (SIFT) to aid the CW-SSIM algorithm to find matching elements between images of possibly different resolutions. This saliency-weighted, SIFT-directed CW-SSIM feature builds upon the ideas in [15], by using sparse SIFT matching and to drive a shift-robust CW-SSIM index.

Given an original image and a distorted image, two scale spaces $SP(I_{original})$ and $SP(I_{retarget})$ are constructed. Using the local extrema detection method in [29], key points are detected on each image: the images are convolved with Gaussian filters at different scales, and the differences of successive Gaussian-blurred images are taken. Key points are defined as the *maxima/minima* of these bandpass differences of Gaussians over multiple scales. Key points which have low contrast or that are poorly localized on an edge [29] are rejected. Each key point is assigned one or more orientations based on the local image gradient histogram.

The key points in the original image and the retargeted image are subjected to a matching process. An example of the matching process is illustrated in Fig. 1. Assuming there are *nPair* matched pairs of key points [29], then define a simplified, non-weighted similarity metric:

$$Sim_{non-w}(I_{ori}, I_{ret}) = \frac{1}{nPair} \sum_{i=1}^{nPair} CWSSIM(p_i, p'_i), \tag{1}$$

where $p_i$ is a key point in the original image and $p'_i$ is a corresponding key point in the retargeted image. The $CWSSIM()$ implementation used here uses a $9 \times 9$ square window.

Given that the human vision system is attracted to certain "salient" regions and salient objects, not all of the matched pixels may have the same perceptual importance. Thus we utilize a saliency weighting scheme to improve the SIFT directed CW-SSIM similarity algorithm just described. We calculate the average saliency using the same $9 \times 9$ window and weight the similarity by the average saliency. We use Itti's well-known algorithm [30] (a MATLAB implementation from http://www.saliencytoolbox.net) to compute saliency. The saliency weight of each window is

$$\omega_i = \frac{1}{n} \sum_{p_j \in window_i} saliency(p_j), \tag{2}$$

where $n = 81$ is the number of pixels in the window. Thus define the

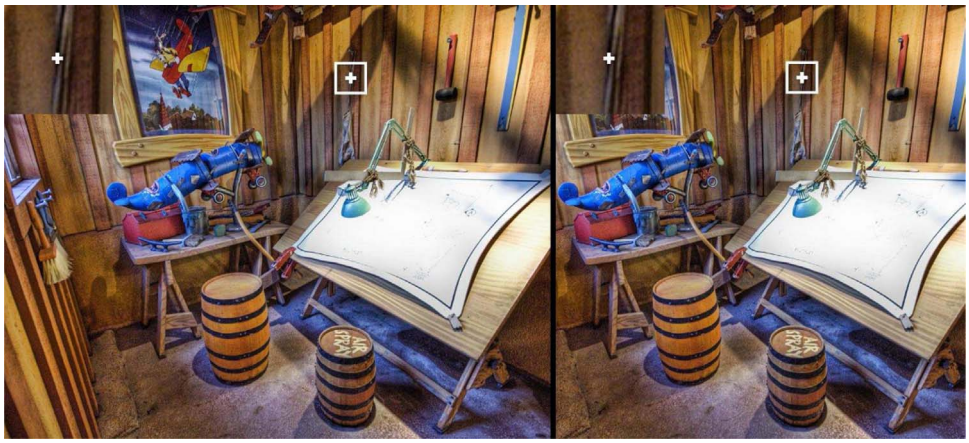

**Fig. 1.** An example of key point matching.

saliency-weighted, SIFT-directed CW-SSIM index

$$Sim_w(I_{ori}, I_{ret}) = \frac{1}{nPair} \sum_{i=1}^{nPair} \omega_i \times CWSSIM(p_i, p'_i). \quad (3)$$

### 2.2. Integrated aesthetics evaluation feature

We have also developed a novel aesthetics model for evaluating this aspect of retargeted images. This allows us to define a unique combination of quality assessment and aesthetics evaluation. Current aesthetics models typically deliver only a binary judgment to identify whether a photo is a professional or an amateur snapshot. Rather than predicting a range of aesthetic values, our integrated method computes and assigns an aesthetic score to a photo, thereby yielding graded predicted judgments.

From the AVA [22], we selected 1320 images from 66 different category (20 images for each), which provides aesthetic scores ranging from 1 to 10 evaluated by professional photographers, and extracted 11 aesthetics features from each. We then trained the neural network to predict aesthetic scores. The training details will be introduced in Section 3.

Aesthetics evaluation is a complex problem with limited available scientific evidence regarding the human sense of this broad attribute. Hence, we began our development process by developing a large number of potential aesthetic features. We conducted extensive testing to determine the relative efficacies of these features in regards to their ability to predict human subjective scores, as reported in [12]. In the end we settled on 11 features having positive correlations with human subjectivity. These 11 features can be divided into three types: global features, saliency features, and facial features.

#### 2.2.1. Global features

Luminance correlates with aesthetic value [24]. As illustrated in Fig. 2, photos of higher average luminance often have better aesthetic value. Hence we transform an image to be analyzed into HSV color space and use the mean value of $V$ to represent the global picture luminance as the first aesthetic feature:

$$f_1 = \frac{1}{WH} \sum_{x=1}^{W} \sum_{y=1}^{H} V(x, y), \quad (4)$$

where $W$ is the width of the photo, $H$ is the height of the photo, and $V(x, y)$ is the value of $V$ at pixel $(x,y)$.

The depth of field (DOF) is the range of distances within a scene that appear acceptably sharp in a photograph [24]. Often, if the DOF is narrow, objects of interest within the DOF lie in front of a blurred background outside the DOF. If the DOF is large, as is often the case when taking photos of landscapes, the background may be as sharp as the subjects. Professional photographers often reduce the depth of field (DOF) when shooting single objects by using larger aperture settings, macro lenses, or telephoto lenses. We adopted Datta et *al.*'s method [24] to detect a condition of low DOF. The low DOF aesthetic features that we use (labeled $f_2$ and $f_3$) are computed only on the saturation $I_s$ and intensity $I_v$ channels.

The rule of thirds is a guideline which has been widely applied in the composition of visual images [32]. For example, photographers often seek to approximately incise a scene along two vertical and two horizontal lines, the four intersections of which are potential locations on which to position objects of high interest (Fig. 3). This feature is computed on the $S$ and $V$ components of the image. Thus define the two aesthetic features


**Fig. 2.** Average luminance correlates with aesthetic value. (a) Image with low average luminance. (b) Image with high average luminance.

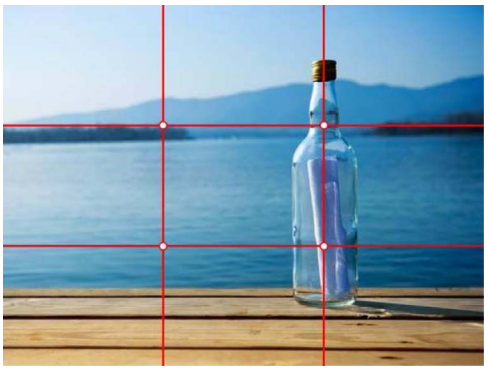

**Fig. 3.** Rule of thirds.

$$f_4 = \frac{9}{WH}\sum_{x=W/3}^{2W/3}\sum_{y=H/3}^{2H/3} S(x, y), \tag{5}$$

$$f_5 = \frac{9}{WH}\sum_{x=W/3}^{2W/3}\sum_{y=H/3}^{2H/3} V(x, y). \tag{6}$$

#### 2.2.2. Saliency features

We again use Itti's algorithm [30], this time to compute a saliency value at each pixel, designating those pixels whose saliency values are greater than $2\times Th$ as salient pixels. $Th$ is defined as the mean saliency value over the entire picture. That is,

$$Th = \frac{1}{WH}\sum_{x=1}^{W}\sum_{y=1}^{H} SM(x, y), \tag{7}$$

where $SM(x, y)$ is the saliency value at pixel $(x,y)$.

#### 2.2.3. Facial features

Since the human face is perhaps the most common and subjectively significant type of object in photographs, we use several features related to the face to complete the aesthetic image assessment process. Specifically, we apply a face detection technique [33] based on skin color. This method computes two $x$ coordinates $k_1$, $k_2$ and two $y$ coordinates $k_3$, $k_4$ of the vertices of a rectangular face region $A$, which has area

$$|A| = |k_2 - k_1| \times |k_4 - k_3|. \tag{8}$$

Faces having high saliency are more likely to be of high subjective value, hence define

$$f_6 = \frac{\sum_{(x,y)\in A} T(x, y)}{|A|}, \tag{9}$$

where $T(x, y)=1$ when $(x,y)$ is a salient pixel as defined earlier, else $T(x, y)=0$.

Also define $S$, $V$ color space features on the face region that capture the saturation and brightness of the face. These aesthetic features are given by:

$$f_7 = \frac{\sum_{(x,y)\in A} S(x, y)}{|A|}, \tag{10}$$

$$f_8 = \frac{\sum_{(x,y)\in A} V(x, y)}{|A|}. \tag{11}$$

Also define texture features using a three-level wavelet transform [24] on $H$, $S$, $V$ color space:

$$f_9 = \frac{\sum_{(x,y)\in A} w_h(x, y)}{|A|}, \tag{12}$$

$$f_{10} = \frac{\sum_{(x,y)\in A} w_s(x, y)}{|A|}, \tag{13}$$

$$f_{11} = \frac{\sum_{(x,y)\in A} w_v(x, y)}{|A|}, \tag{14}$$

where $w_h$, $w_s$, $w_v$ are $\{w^{lh}, w^{hl}, w^{hh}\}$ on $H$, $S$, $V$ color components, where LH, HL and HH are three bands of wavelet coefficients [24].

### 2.3. Other eight existing retargeting assessment features

In addition to the new CW-SSIM and aesthetics features, we also use 8 other features that have previously proven relevant for predicting retargeted image quality. These include Fang et *al.*'s [15] method which is taken as an input feature to the neural network. The other 7 features are drawn from experiments on the RetargetMe dataset [12] and include: Bidirectional Similarity (BDS) [34], Bidirectional Warping (BDW) [7], SIFT flow [18], Earth Movers Distance (EMD) [35], Color Layout Descriptor Patch Match (BDSPM) [36], Edge Histogram (EH) [37], and Color Layout Descriptor (CL) [38].

## 3. RBF neural network

We learn the image retargeting assessment model using an RBF neural network [39,40]. Artificial neural networks have a long history of applications in image analysis [41]. Among the various flavors of neural networks, the RBF neural network model is particularly well suited for learning to approximate continuous or piecewise continuous real-valued mappings, when the input dimension is sufficiently small.

The radial basis function network is an artificial neural network that uses radial basis functions as activation functions. By training the neural network, the output is expressed as a linear combination of radial basis functions of the inputs. A function is a radial basis (RBF) if its output is a non-increasing function of the distance of the input from a given stored vector.

There are three layers in an RBF network: an input layer, a hidden layer with a non-linear RBF activation function, and a linear output layer. The input is modeled as a vector of real numbers so the output is then a scalar function of the input vector and is given by:

$$\varphi(x) = \sum_{i=1}^{N} a_i \rho(\|x - c_i\|), \tag{15}$$

where $N$ is the number of neurons in the hidden layer; $c_i$ is the center vector for neuron $i$ and $a_i$ is the weight of neuron $i$ in the linear output neuron. The norm is typically taken to be the Euclidean distance and the radial basis function is commonly taken to be Gaussian:

$$\rho(\|x - c_i\|) = \exp[-\beta \|x - c_i\|^2]. \tag{16}$$

For a Gaussian RBF this sensitivity may be tuned by adjusting the spread $\sigma$, where a larger spread implies less sensitivity. An RBF network with enough hidden neurons can approximate any continuous function with arbitrary precision.

In the training process, the following parameters are determined:

- The number of neurons in the hidden layer.
- The center vectors in the hidden layer.
- The radius of each RBF function in each dimension.
- The weights applied to the RBF function.

In order to train the RBF network, we use a two-step algorithm. In the first step, we choose the center vectors in the hidden layer. The centers can be randomly sampled from some sets of examples, and a third back-propagation step is applied to tune all the parameters [39,40].

The second step simply fits a linear model with coefficients $\omega_i$ to the hidden layer's outputs with respect to some objective function such as

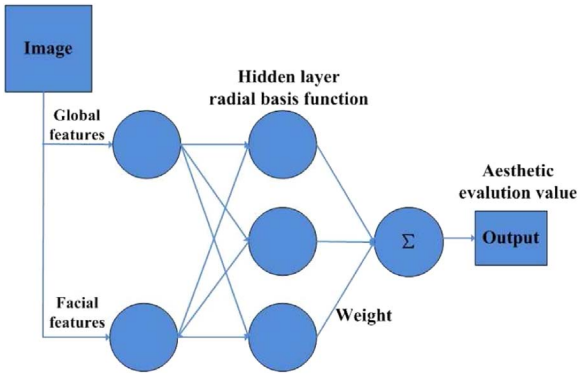


**Fig. 4.** Flow diagram of our aesthetics evaluation method.

the least squares function. By choosing optimal weights, minimization of the least squares objective function optimizes the accuracy of fit. The object is to obtain a linear combination of the above aesthetic features. The goal of learning is to estimate the linear weights of these features.

Since existing image retargeting databases, like [12] are insufficiently large to train deep learners on, we utilized the efficient RBF neural network as our learning engine. We deployed RBF neural networks to train both the aesthetics evaluation engine and the quality assessment engine.

### 3.1. Training of aesthetics evaluation engine

Fig. 4 shows the flow of our aesthetics evaluation method. We selected 1320 images from the AVA [22] as a training set, where each of the images have associated subjective aesthetics scores ranging from 1 to 10. We trained a model to perform image aesthetics evaluation by computing the 11 features on the 1320 images, yielding $1320 \times 11$ features fed as input to the neural network, which outputs predicted subjective image aesthetic scores. That is

$$Input_i = f_i \tag{17}$$

$$Output \in [1, 10] \tag{18}$$

where $Input_i$ is the value of the $i^{th}$ feature and *Output* is the predicted retargeted image aesthetic score.

### 3.2. Training of quality assessment engine

Fig. 5 shows the framework of our proposed learning based image retargeting quality assessment method. In this process, the input is the above-mentioned ten features of the retargeted images processed by different retargeting methods and the output is the quality score, which ranges from 0 to 1. Following the training process, the trained RBF network is able to assess the quality of retargeted images with high accuracy. The details of training the proposed overall quality assessment engine are explained in the next section.

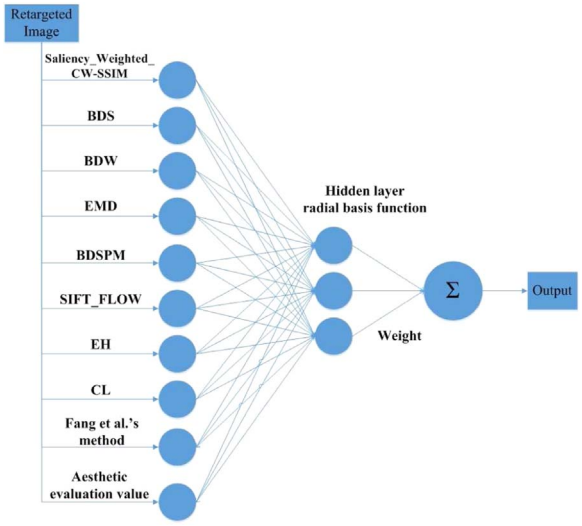


**Fig. 5.** Framework of the integrated retargeted image quality evaluation engine.

## 4. Evaluation

### 4.1. Dataset

We conducted the following experiment to examine the validity of our integrated model. We tested our integrated model on two benchmark datasets: the RetargetMe dataset [12] and the CUHK dataset [13]. In this experiment, we used the images and content in the RetargetMe and CUHK datasets.

When using the RetargetMe dataset, we chose the 37 analysis images for which users' votes are supplied. These images have one or more attributes from six major attributes: Line/Edge, Face/People, Foreground Objects, Texture, Geometric Structures and Symmetry. For each image, eight different retargeting methods were applied, thus every image has eight retargeted images generated by distinct methods. These eight retargeting methods include Scaling (SCL), Cropping (CR), Seam Carving (SC) [3], Shift-maps (SM) [4], Nonhomogeneous warping (WARP) [6], and Streaming Video (SV) [42]. These Eight retargeting methods were ranked by the number of votes each received. We regarded the users' votes as the ground truth of image perceptual quality and used normalized votes to train our neural network as well as to validate our integrated method. The correlations between the objective and subjective scores are measured by the Kendall's rank-order correlation coefficient (KROCC) [43]:

$$KROCC = \frac{n_c - n_d}{0.5n\,(n-1)} \tag{19}$$

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

We also tested our method on the CUHK dataset, which includes 57 natural images, and three retargeted versions of each image. The retargeted images were created using ten state-of-the-art retargeting methods, including optimized seam carving and scaling (SCSL) [44], and energy-based deformation (ENER) [45], as well as eight retargeting methods associated with the RetargetMe dataset. Each of the 171 retargeted image in the CUHK dataset has an associated MOS score, which was obtained by averaging the opinion scores of at least 30 human viewers.

In the CUHK dataset, four evaluation metrics have been used to evaluate the correlations between objective quality prediction scores and the MOS. These metrics are: Pearson's Linear Correlation Coefficient (LCC), Spearman's rank-order correlation coefficient (SROCC) [46], Root Mean Squared Error (RMSE) and Outlier Ratio (OR) [47,48]. LCC is computed between MOS and the objective scores after nonlinear regression. The nonlinear regression is the mapping function [49]:

$$f(x) = \beta_1 \left( \frac{1}{2} - \frac{1}{1 + e^{\beta_2 (x - \beta_3)}} \right) + \beta_4 x + \beta_5. \tag{20}$$

SROCC evaluates the prediction monotonicity of the objective scores. RMSE evaluates the difference between MOS and the objective scores after nonlinear regression. OR evaluates the false objective score ratio. A false objective score is one that lies outside the interval $[MOS - 2\theta, MOS + 2\theta]$ following nonlinear regression, where $\theta$ is the standard deviation of the corresponding MOS values.

**Table 1**
Performance of different methods on RetargetMe dataset.

| Metric | Attribute | | | | | | Total | | |
|---|---|---|---|---|---|---|---|---|---|
| | Lines/Edges | Faces/ People | Texture | Foreground objects | Geometric structures | Symmetry | mean | std | p-value |
| BDS | 0.040 | 0.190 | 0.060 | 0.167 | −0.004 | −0.012 | 0.083 | 0.268 | 0.017 |
| BDS-PM | 0.054 | 0.162 | 0.083 | 0.167 | 0.062 | −0.024 | 0.097 | 0.232 | 0.016 |
| BDW | 0.031 | 0.048 | −0.048 | 0.060 | 0.004 | 0.119 | 0.046 | 0.181 | 0.869 |
| EH | 0.043 | −0.076 | −0.060 | −0.079 | 0.103 | 0.298 | 0.004 | 0.334 | 0.641 |
| CL | −0.023 | −0.181 | −0.071 | −0.183 | −0.009 | 0.214 | −0.068 | 0.301 | 0.384 |
| SIFT_flow | 0.097 | 0.252 | 0.119 | 0.218 | 0.085 | 0.071 | 0.145 | 0.262 | 0.031 |
| EMD | 0.220 | 0.262 | 0.107 | 0.226 | 0.237 | 0.500 | 0.251 | 0.272 | 1.00E−05 |
| Fang et al.'s method | 0.211 | 0.319 | 0.143 | 0.290 | 0.192 | 0.321 | 0.249 | 0.239 | 2.36E−07 |
| Aesthetic evaluation value | 0.117 | 0.190 | 0.095 | 0.242 | 0.174 | −0.048 | 0.143 | 0.323 | 0.011 |
| Saliency Weighted CW-SSIM | 0.154 | 0.343 | 0.214 | 0.222 | 0.134 | 0.155 | 0.183 | 0.332 | 0.002 |
| Liu et al.'s method | 0.154 | 0.143 | 0.202 | 0.127 | 0.156 | 0.238 | 0.166 | 0.251 | 2.88E−04 |
| Hsu et al.'s method | 0.431 | 0.390 | 0.389 | 0.286 | 0.438 | 0.523 | 0.415 | 0.296 | 6.0E−10 |
| Liang et al.'s mehod | 0.351 | 0.271 | 0.304 | 0.381 | 0.415 | **0.548** | 0.399 | / | / |
| ARS | 0.463 | **0.519** | 0.444 | 0.330 | 0.505 | 0.464 | 0.452 | 0.283 | 1.00E−11 |
| Existing eight features | 0.429 | 0.443 | 0.452 | 0.413 | 0.424 | 0.524 | 0.427 | 0.154 | 1.13E−18 |
| Existing eight features + Saliency Weighted CW-SSIM | 0.497 | 0.467 | 0.488 | 0.444 | 0.478 | 0.512 | 0.471 | 0.157 | 9.73E−20 |
| Existing eight features + Aesthetic evaluation value | 0.497 | 0.476 | 0.488 | 0.464 | 0.491 | 0.512 | 0.481 | 0.151 | 3.27E−15 |
| Our integrated mehod | **0.520** | 0.505 | **0.548** | **0.496** | **0.509** | 0.524 | **0.510** | **0.145** | **4.74E−22** |

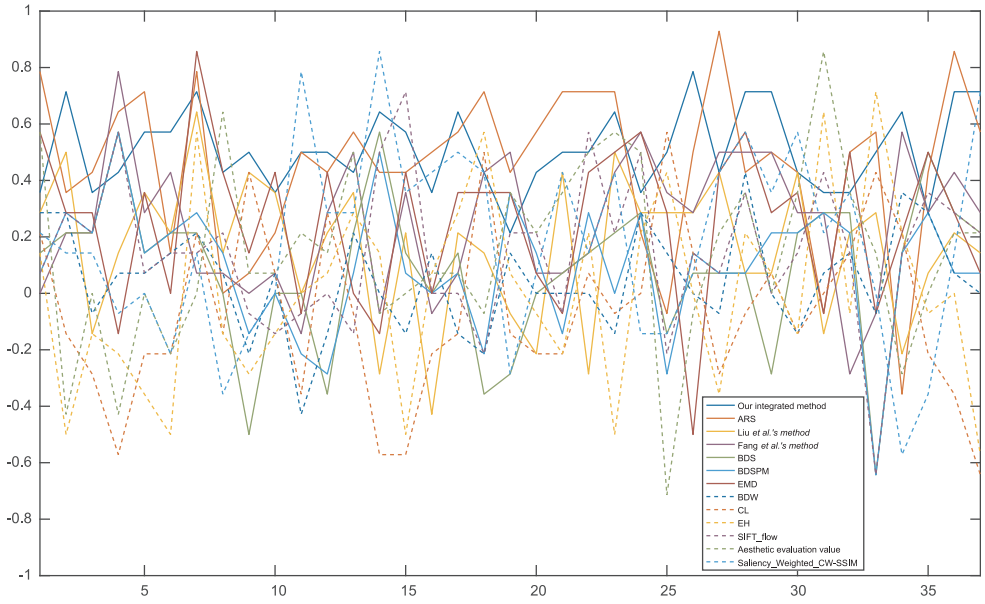


**Fig. 6.** The KROCC score at each iteration time. We compared our integrated method with Liu et al.'s method [14], ARS [19], and other ten methods that we used as features. The *X* label is the iteration time and the *Y* label is the KROCC score.

### 4.2. Performance on the retargetme dataset

We applied ten assessment methods to each of the retargeted images in the RetargetMe dataset to obtain $37 \times 8 \times 10$ assessment results. To verify the performance of our metric and compare it to other objective methods, we applied leave-one-out cross validation (LOOCV). Specifically, at each iteration of LOOCV, one original image and its eight retargeted results were used as the test set while the remaining images were used as the training set.

In the training process, the RBF network takes the ten assessment results on a retargeted image as input and the normalized votes of the image as label. In this case:

$$IN_i = result_i \tag{21}$$

$$OUT \in [0, 1] \tag{22}$$

where $result_i$ is the evaluation of the $i$th assessment method and $OUT$ is the predicted quality score on the image. By trained on the $36 \times 8$ sets of data, we obtained our assessment model.

In the testing process, we used the remaining one set of data as the test set. We evaluated each of the eight retargeted images using several different approaches. These evaluations form the $IN$ vector, which indicate the assessment result of each approach to be analyzed. The RBF neural network took these evaluations as inputs and predicted the quality score of the remaining eight images using the model learned from the 36 sets of data. In this process the values were weighted and transformed in a non-linear manner and the output $OUT$ was determined from the processed values. By computing the output of the eight images, a ranking of the eight methods was obtained.

For each of the tested original images, both subjective and objective rankings were compared. The KROCC was computed to measure the level of agreement between the subjective rankings and the objective

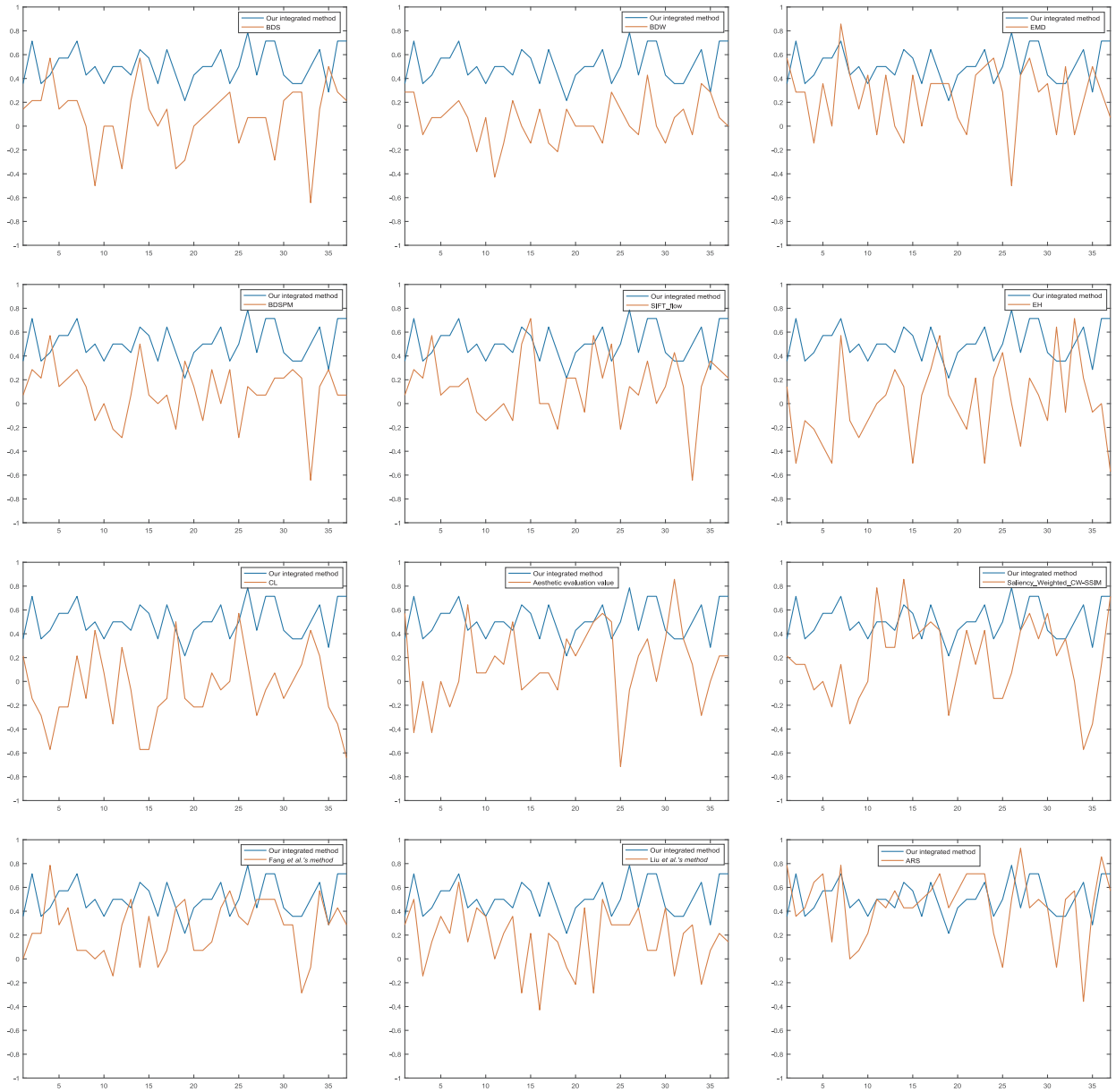


**Fig. 7.** KROCC score comparisons at each iteration time. Each sub-figure compares the KROCC score between our integrated method and a competing retargeting quality assessment measure at each iteration time. The *X* label is the iteration time and the *Y* label is the KROCC score.

rankings obtained using the different assessment methods, including the proposed one. In both cases, higher correlation coefficient indicates higher agreement. If the rank order is the same, the correlation coefficient will be 1.

For each assessment method, we computed the mean and standard deviation of the KROCC values across all test sets. Comparisons were made between our integrated method and the other assessment approaches. The compared methods include the ten features used in our model (applied individually) and four state-of-the-art image retargeting quality assessment methods: Liu et al.'s method [14], Liang et al.'s method [27], Hsu et al.'s method [17] and ARS [19]. For better comparison, we also computed the mean KROCC values on each subset. The result is shown in Table 1. From this table, it may be observed that our integrated method significantly outperformed all other objective assessment measures. Fig. 6 plots the KROCC scores of our integrated method and the twelve other methods at each iteration. Beyond comparison with the ten individual features composing our model, we also compared performance against Liu et al.'s method [14] (implemented using Itti's saliency maps [30]) and ARS [19]. For clear explanation, Fig. 7 plots comparisons of the KROCC scores between our integrated method and each competing retargeting quality assessment measure. As may be observed from these two figures, our integrated method was able to provide more stable performance than the other methods.

In order to further demonstrate the relevance of the new features we have introduced, we also tested the performance of our framework using only the eight previously existing features, against our integrated method. The comparison is summarized in Table 1, which clearly indicates that when the two new features are removed, the performance drops significantly.

It should be noted that for our integrated method the average KROCC value was around 0.5, indicating that there remains room for significant improvement.

### 4.3. Performance on the CUHK dataset

We also tested our method by using LOOCV on the CUHK dataset. At each iteration of LOOCV, one original image and its three retargeted

**Table 2**
Performance of different retargeted image quality assessment models on the CUHK dataset.

| Metric | SROCC | LCC | RMSE | OR |
|---|---|---|---|---|
| Liu et al.'s method | 0.4662 | 0.4374 | 12.141 | 0.1520 |
| Hsu et al.'s method | 0.5409 | 0.5403 | 11.361 | 0.1520 |
| ARS | 0.6693 | 0.6835 | 9.855 | 0.0702 |
| Our integrated method | **0.9123** | **0.9211** | **3.201** | **0.0117** |

results were used as the test set, while the remaining images are used as the training set. We then used our trained model to predict the quality scores on the three retargeted versions of each of the 57 source images in this dataset; after 57 iterations, 57 quality score triples were obtained. To evaluate performance, we calculated the mean of the SROCC, LCC, RMSE and OR on the 171 retargeted images. We compared our method with three state-of-the-art methods: Liu et al.'s method [14], Hsu et al.'s method [17] and ARS [19]. The results are shown in Table 2. From this table, it may be observed that our integrated method significantly outperformed the three state-of-the-art objective quality assessment methods.

## 5. Conclusion

We have described an open framework for image retargeting quality assessment that uses a saliency-weighted, SIFT-directed CW-SSIM model and an aesthetics evaluation model, that are combined with eight existing retargeting assessment methods. These new and old models supply features that we used to train a radial basis function network to learn to predict the quality of retargeted images. We used the images and subjective votes in the RetargetMe dataset to train and test our integrated prediction model. In order to verify the cross-database stability of our method, we also tested it on the CUHK dataset. Promising quality prediction results are achieved using the trained neural network. The experimental results confirm a performance leap relative to compared state-of-the-art methods.

## References


[1] Y. Xia, L. Zhang, R. Hong, L. Nie, Y. Yan, L. Shao, Perceptually guided photo retargeting, IEEE Trans. Cybern. PP 99 (2016) 1–13.
[2] S. Avidan, A. Shamir, Seam carving for content-aware image resizing, in: ACM Transactions on Graphics (TOG), vol. 26, 2007, p. 10.
[3] M. Rubinstein, A. Shamir, S. Avidan, Improved seam carving for video retargeting, in: ACM Transactions on Graphics (TOG), vol. 27, 2008, p. 16.
[4] Y. Pritch, E. Kav-Venaki, S. Peleg, Shift-map image editing, in: Proceedings of the 2009 IEEE 12th International Conference on Computer Vision, 2009, p. 151–158.
[5] Y.-S. Wang, C.-L. Tai, O. Sorkine, T.-Y. Lee, Optimized scale-and-stretch for image resizing, in: ACM Transactions on Graphics (TOG), vol. 27, ACM, 2008, p. 118.
[6] L. Wolf, M. Guttmann, D. Cohen-Or, Non-homogeneous content-driven video-retargeting, in: Proceedings of the IEEE 11th International Conference on Computer Vision, 2007. ICCV 2007, 2007, pp. 1–6.
[7] M. Rubinstein, A. Shamir, S. Avidan, Multi-operator media retargeting, ACM Trans. Graph. (TOG) 28 (3) (2009) 1–11.
[8] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[9] T.-J. Liu, W. Lin, C.-C.J. Kuo, Image quality assessment using multi-method fusion, IEEE Trans. Image Process. 22 (5) (2013) 1793–1807.
[10] A.C. Bovik, Automatic prediction of perceptual image and video quality, Proc. IEEE 101 (9) (2013) 2008–2024.
[11] A.K. Moorthy, A.C. Bovik, Visual quality assessment algorithms: what does the future hold? Multimed. Tools Appl. 51 (2) (2011) 675–696.
[12] M. Rubinstein, D. Gutierrez, O. Sorkine, A. Shamir, A comparative study of image retargeting, in: ACM Transactions on Graphics (Proceedings SIGGRAPH Asia), 29(5), 2010, pp. 160:1–160:10.
[13] L. Ma, W. Lin, C. Deng, K.N. Ngan, Image retargeting quality assessment: a study of subjective scores and objective metrics, IEEE J. Select. Top. Signal Process., 6(6).
[14] Y. Liu, X. Luo, Y. Xuan, W. Chen, X. Fu, Image retargeting quality assessment, in: Computer Graphics Forum, vol. 30, 2011, pp. 583–592.
[15] Y. Fang, K. Zeng, Z. Wang, W. Lin, Z. Fang, C.-W. Lin, Objective quality assessment for image retargeting based on structural similarity, IEEE J. Emerg. Sel. Top. Circuits Syst. 4 (1) (2014) 95–105.
[16] J. Zhang, C.-C. J. Kuo, An objective quality of experience (qoe) assessment index for retargeted images, in: Proceedings of the ACM International Conference on Multimedia, ACM, 2014, pp. 257–266.
[17] C. Hsu, C. Lin, Y. Fang, W. Lin, Objective quality assessment for image retargeting based on perceptual geometric distortion and information loss, IEEE J. Sel. Top. Signal Process. 8 (3) (2014) 377–389.
[18] C. Liu, J. Yuen, A. Torralba, J. Sivic, W.T. Freeman, Sift flow: dense correspondence across different scenes, in: European Conference on Computer Vision (ECCV 2008), 2008, pp. 28–42.
[19] Y. Zhang, Y. Fang, W. Lin, X. Zhang, Backward registration based aspect ratio similarity (ars) for image retargeting quality assessment, IEEE Trans. Image Process. 25 (9) (2016) 1.
[20] L. Xu, W. Lin, C.-C.J. Kuo, Visual Quality Assessment by Machine Learning, Springer, 2015.
[21] H. Tong, M. Li, H. jiang Zhang, J. He, C. Zhang, Classification of digital photos taken by photographers or home users, in: Proceedings of Pacific Rim Conference on Multimedia, 2004, pp. 198–205.
[22] N. Murray, L. Marchesotti, F. Perronnin, Ava: A large-scale database for aesthetic visual analysis, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2408–2415.
[23] Y. Ke, X. Tang, F. Jing, The design of high-level features for photo quality assessment, in: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, 2006, pp. 419–426.
[24] R. Datta, D. Joshi, J. Li, J.Z. Wang, Studying aesthetics in photographic images using a computational approach, in: Computer Vision–ECCV 2006, 2006, pp. 288–301.
[25] C. Cerosaletti, A. Loui, Measuring the perceived aesthetic quality of photographic images, in: 2009. QoMEx 2009. International Workshop on Quality of Multimedia Experience, 2009, pp. 47–52.
[26] W. Jiang, A. Loui, C. Cerosaletti, Automatic aesthetic value assessment in photographic images, in: 2010 IEEE International Conference on Multimedia and Expo (ICME), 2010, pp. 920–925.
[27] Y. Liang, Y.J. Liu, D. Gutierrez, Objective quality prediction of image retargeting algorithms, IEEE Trans. Vis. Comput. Graph. PP 99 (2016) 1.
[28] M.P. Sampat, Z. Wang, S. Gupta, A.C. Bovik, M.K. Markey, Complex wavelet structural similarity: a new image similarity index, IEEE Trans. Image Process. 18 (11) (2009) 2385–2401.
[29] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vision. 60 (2) (2004) 91–110.
[30] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, IEEE Trans. Pattern Anal. Mach. Intell. 20 (11) (1998) 1254–1259.
[31] M. Clark, A.C. Bovik, Experiments in segmenting texton patterns using localized spatial filters, Pattern Recognit. 22 (6) (1989) 707–717.
[32] S. Meech, Contemporary Quilts: Design, Surface and Stitch, Harper Collins, 2003.
[33] X. Ma, H. Zhang, X. Zhang, A face detection algorithm based on modified skin-color model, in: Control Conference (CCC), 2013 32nd Chinese, 2013, pp. 3896–3900.
[34] D. Simakov, Y. Caspi, E. Shechtman, M. Irani, Summarizing visual data using bidirectional similarity, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008), 2008, pp. 1–8.
[35] O. Pele, M. Werman, Fast and robust earth mover's distances, in: Proceedings of IEEE 12th international conference on Computer vision (ICCV 2009), 2009, pp. 460–467.
[36] C. Barnes, E. Shechtman, A. Finkelstein, D. Goldman, Patchmatch: a randomized correspondence algorithm for structural image editing, ACM Trans. Graph. TOG 28 (3) (2009) 24.
[37] B.S. Manjunath, J.-R. Ohm, V.V. Vasudevan, A. Yamada, Color and texture descriptors, IEEE Trans. Circuits Syst. Video Technol. 11 (6) (2001) 703–715.
[38] E. Kasutani, A. Yamada, The MPEG-7 color layout descriptor: a compact image feature description for high-speed image/video segment retrieval, in: International Conference on Image Processing (ICIP'2001), vol. 1, IEEE, 2001, pp. 674–677.
[39] D.S. Broomhead, D. Lowe, Radial basis functions, multi-variable functional interpolation and adaptive networks, Tech. Rep., DTIC Document, 1988.
[40] F. Schwenker, H.A. Kestler, G. Palm, Three learning phases for radial-basis-function networks, Neural Netw. 14 (4) (2001) 439–458.
[41] O.K. Sethi, A.K. Jain, Artificial Neural Networks and Statistical Pattern Recognition: Old and New Connections, Machine Intelligence and Pattern Recognition, North-Holland, Amsterdam, 1991.
[42] P. Krähenbühl, M. Lang, A. Hornung, M. Gross, A system for retargeting of streaming video, in: ACM Transactions on Graphics (TOG), vol. 28, ACM, 2009, p. 126.
[43] M.G. Kendall, A new measure of rank correlation, Biometrika 30 (1/2) (1938) 81–93.
[44] W. Dong, N. Zhou, J.-C. Paul, X. Zhang, Optimized image resizing using seam carving and scaling, in: ACM Transactions on Graphics (TOG), vol. 28, ACM, 2009, p. 125.
[45] Z. Karni, D. Freedman, C. Gotsman, Energy-based image deformation, in: Computer Graphics Forum, vol. 28, Wiley Online Library, 2009, pp. 1257–1268.
[46] C. Spearman, The proof and measurement of association between two things, Am. J. Psychol. 15 (1) (1904) 72–101.
[47] Z. Wang, L. Lu, A.C. Bovik, Video quality assessment based on structural distortion measurement, Signal Process. Image Commun. 19 (2) (2004) 121–132.
[48] D. Jayaraman, A. Mittal, A.K. Moorthy, A.C. Bovik, Objective quality assessment of multiply distorted images, in: Asilomar Conference on Signals, Systems and Computers, 2012, pp. 1693–1697.
[49] H.R. Sheikh, M.F. Sabir, A.C. Bovik, A statistical evaluation of recent full reference image quality assessment algorithms, IEEE Trans. Image Process. 15 (11) (2006) 3440–3451.