No-Reference Retargeted Image Quality Assessment Based on Pairwise Rank Learning

Lin Ma, *Member, IEEE*, Long Xu, *Member, IEEE*, Yichi Zhang, Yihua Yan, and King Ngi Ngan, *Fellow, IEEE*

***Abstract*—In this paper, we propose a novel no-reference image quality assessment method for the retargeted image based on the pairwise rank learning approach. Each retargeted image needs to be first represented as a feature vector, which not only captures the image characteristics but also is sensitive to distortions during the retargeting process. As such, we investigate and examine different image representations for their abilities depicting the perceptual quality of retargeted image. Based on the image representations, we resort to the pairwise rank learning approach to discriminate the perceptual quality between the retargeted image pairs. Experimental results demonstrate that the proposed method can effectively depict the perceptual quality of the retargeted image, which can even perform comparably with the full-reference quality assessment methods.**

***Index Terms*—Image quality assessment, no-reference, rank learning, retargeted image.**

I. Introduction

NOWADAYS, with the rapid development of mobile devices, new applications of image/video have appeared in different terminal devices to improve the viewers' visual quality of experience (QoE) [1]–[5]. It is greatly demanded to display the same image/video content on all kinds of terminals, such as the mobile phone, tablets, and so on, to provide more conveniences and better QoEs for the viewers. However, different terminals are of different resolutions, which require a better displaying technique to better meet the viewers' perception of the visual contents. Traditionally, simple scaling and cropping methods are adopted to alter the image resolution arbitrarily. However, the salient content cannot be well preserved, which further introduces some unpleasant viewing experiences. In order to handle the drawbacks of scaling and cropping processes, several content-aware retargeting methods [1], [3], [6]–[12] are developed to adapt the image to different resolutions, which discard the unimportant information and meanwhile preserve the salient and important content information of the image. In order to demonstrate the superiorities of the retargeting methods, the authors usually performs the sample visual comparisons. However, such comparison cannot be performed at large scale due to a huge labor. More importantly, it cannot be employed for the online optimization and guidance of the retargeting process. Therefore, developing one metric to automatically evaluate the retargeted image quality is demanded, not only for evaluation and comparison but also for guiding the optimization process of retargeting.

Mean squared error (MSE) and the related peak signal-to-noise ratio (PSNR) are widely adopted to evaluate the fidelity between one signal and its corrupted version, because of its simplicity, easy optimization, and clear physical meaning. However, MSE and PSNR have been criticized that they are not able to simulate the perception of human visual system (HVS) [13], [14]. As such, many quality metrics have been developed to handle the drawbacks of MSE and PSNR and overcome their shortcomings. The most famous and representative one is the structural similarity (SSIM) [15], which is proposed to evaluate the image perceptual quality from three perspectives, specifically, the luminance, contrast, and structure, other than the image pixel difference. SSIM has been demonstrated to be effective to evaluate both natural image and video signals. Recently, more effective quality metrics have been proposed, such as feature similarity (FSIM) [16] for natural image, and motion-based video integrity evaluation (MOVIE) [17] for natural video. Theses quality metrics have achieved great success on characterizing the traditional distortions, such as JPEG compression, Gaussian noise, Gaussian blur, and so on. However, for the quality assessment of retargeted images, the most straightforward observation is that the resolutions between the original and retargeted images have been altered. The developed quality metrics, such as PSNR, SSIM, FSIM, and so on, cannot evaluate two images with different resolutions.

More specifically, the distortions in retargeted images are introduced from the following twofold perspectives [18]. Firstly, the retargeting process will inevitably discard partial information of the image content, even though they are not as important as other salient image information. Such discarding process will introduce the information loss to the retargeted image. Therefore, how to measure the effect of the discarded content on the perceptual quality of the retargeted image is very important and different from the traditional distorted image. Secondly, rather than information loss, one new distortion, namely the shape distortion, is introduced during the retargeting process. The

Manuscript received April 29, 2016; revised August 30, 2016; accepted September 21, 2016. Date of publication September 27, 2016; date of current version October 19, 2016. This work was supported in part by a grant from the Research Grants Council of the Hong Kong SAR, China (Project CUHK 415913), and by the National Natural Science Foundation of China under Grant 61572461 and Grant 11433006. The work of L. Xu was supported by CAS 100-Talents. The guest editor coordinating the review of this manuscript and approving it for publication was Prof. Daniel Keim. *(Corresponding author: Long Xu.)*

L. Ma is with Tencent AI Lab, Shenzhen 518057, China (e-mail: forest.linma@gmail.com).

L. Xu and Y. Yan are with the Key Laboratory of Solar Activity, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100049, China (e-mail: lxu@nao.cas.cn; yyh@nao.cas.cn).

Y. Zhang and K. N. Ngan are with the Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong, China (e-mail: yczhang@ee.cuhk.edu.hk; knngan@ee.cuhk.edu.hk).

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

Digital Object Identifier 10.1109/TMM.2016.2614187

retargeting methods [1], [3], [6]–[12] remove the image pixels, seams, and regions to meet the requirements of the final resolution. The shapes, such as the object boundaries and contours, will be inevitably distorted, which will severely degrade the perceptual quality of the image and significantly affect the viewing experiences of subjects. Overall, the retargeted image differs with the original image in resolution and distortion types, making its perceptual quality much more challenging for evaluation.

In this paper, we propose to evaluate the retargeted image quality in a no-reference (NR) manner. We firstly formalize the NR quality assessment as a pairwise ranking problem. Afterwards, a rank learning method is proposed to discriminate the perceptual quality of the retargeted image. Based on the learned ranking model, one can measure retargeted images without references. Our main contributions are listed in the following.

1) We first propose an NR quality assessment method for retargeted images, which achieves comparable performances with the full reference (FR) quality metrics on the public retargeted image quality databases.
2) We formalize the NR quality assessment as one pairwise ranking problem, where the image perceptual quality can be effective depicted by our proposed rank learning approach.
3) We examine and investigate the effects of different image representations on the retargeted image quality assessment (IQA).

The paper is organized as follows. Section II discusses the related work on retargeting and NR IQAs. We formulate the problem of NR IQA and introduce the pairwise rank learning framework in Section III. Detailed information about our proposed pairwise rank learning metric for retargeted images is introduced in Section IV. And the experimental results are illustrated in Section V. Finally, conclusions are provided in Section VI.

II. Related Work

A. Retargeted Image Quality Assessment

Nowadays, there are several research works discussing the retargeted IQA. Based on the approaches, the research work can be roughly categorized into the subjective [18], [19] and objective [4], [20]–[22] approaches. The subjective approach is the most reliable way for assessing the perceptual quality of the retargeted image. However, it requires many subjects to participate in the subjective testing process, which is very time consuming and of a huge labor work. Thus, it cannot be used for online manipulation and optimization. The subjective testing process is usually employed to construct the database, which provides the ground truth value of the image perceptual quality. In [19], the authors constructed the RetargetMe database, which employed one pairwise comparison method to indicate which retargeted image possesses a better perceptual quality, compared with the other retargeted images. The RetargeteMe database consists of the retargeted image and the number of times that the image is favored over other images. In [18], [23], the authors employ the simultaneous double stimulus for continuous evaluation (SDSCE) [24] to perform the subjective quality evaluation. The subjects provide their personal opinions on the perceptual quality of the retargeted image with the original image as reference. The database is composed by the retargeted image as well as its corresponding mean opinion score (MOS). Moreover, the impact of the retargeting process on human fixations is examined, by gathering eye-tracking data for a representative benchmark of retargeted images in [25].

For the objective quality metrics for retargeted images, a metric named as bidirectional similarity (BDS) is developed in [20], [22]. Two visual signals, namely the original and retargeted images, are considered to be visually similar where as many as possible patches of one visual signal are shared by the other visual signal in a bidirectional manner. BDS tries to depict the similarity between two visual signals from the completeness and coherence perspectives, which can be further employed to generate a retargeted image by minimizing the similarity measurement. The scale-invariant feature transform (SIFT) [26] is proposed to characterize the view-invariant and brightness-independent image structures. Therefore, matching SIFT descriptors can help to establish meaningful correspondences across images with significantly different image contents. Moreover, the pixel displacements obtained from the SIFT flow [27] are spatially coherent. The matching cost can thus indicate the difference and pixel displacement between the original and retargeted images. In [4], Fang *et al.* proposed one metric, namely IR-SSIM, to create an SSIM quality map that indicates at each spatial location of the reference image how the structural information is preserved in the retargeted image. For each pixel in the original image, the best matching pixel in retargeted image is firstly located. The SSIM measurement is calculated between the local regions of the original and retargeted image around specific matching pixels. After obtaining the SSIM quality map for the reference image, a saliency map is developed to pool the SSIM quality map into a final quality score. Zhang and Ngan [28] proposed one metric named as RB-RIQA by employing a region-based framework to evaluate shape distortion and information loss of the retargeted images. Liang *et al.* [29] proposed a simple yet efficient objective assessment method based on similarity, satisfying aesthetic requirements, and maintaining symmetry features.

B. No-Reference Image Quality Assessment

The aforementioned metrics for retargeted images can be regarded as the FR metrics, where the original reference image needs to be present for quality evaluation. However, in practical scenarios, the reference image is always unavailable. Therefore, NR quality metrics are highly demanded for quality assessment. There are also several research works discussing NR IQA, which can be roughly classified into three categories. The first category approach takes the behavior of specific distortions into consideration. For example, in [30], Sheikh *et al.* employed the wavelet statistical model to capture the distortion introduced by JPEG 2000. Brandão and Queluz [31] proposed an NR quality metric based on the discrete cosine transform (DCT) domain statistics to evaluate the quality of JPEG coded image. The second category approach relies on the quality aware clustering. The image patches of training set are grouped into the given number of

classes based on local image features [32], [33]. Each cluster center has a quality score which is derived from the qualities of image patches falling into this cluster. Associating cluster centers with their qualities, a codebook is established. For the patches of a test image, the codebook is looked up to search the most similar codeword and retrieve the associated quality values. In [32], a visual codebook associated with the Gabor filter based local appearance descriptors as well as the MOS is proposed. The authors of [33] used FSIM [16] instead of MOS as image patch quality to establish the codebook. The third category approach is to utilize machine learning method [34] to map image features into image qualities. In [35], Moorthy and Bovik proposed to use support vector machine (SVM) [36], [37] and support vector regression (SVR) [38], [39] to learn a classifier and an ensemble of regressors for a distortion-aware quality metric. It deploys summary statistics called natural scene statistics (NSS) which is derived from wavelet decomposition of an image. In [40], Tang *et al.* proposed an approach similar to [35] but with more elaborate features, including texture statistics, blur/noise statistics, and histogram of each subband of image decomposition.

Current NR quality metrics are developed for natural images with traditional distortions, such as JPEG, JPEG 2000, Gaussian blur, and so on. However, for retargeted images, to our best knowledge, there is no literature discussing the NR quality assessment methods. In this paper, we make the first attempt to address the NR quality metrics for retargeted images based on the pairwise rank learning approach.

III. No-Reference Retargeted Images Quality Assessment

In this section, we first reformulate the NR IQAs as one pairwise comparison problem, based on which our proposed pairwise rank learning metric for retargeted images is thereby introduced.

A. Problem Formulation

Nowadays, there is a thread of work on the learning based NR IQAs [16], [32], [33], [35], [40]–[42]. The authors represent the image or image patch as one feature vector. Afterwards, the clustering or regression learning methods are employed to generate the perceptual quality index for the provided image or image patch. Such NR metrics are developed to measure the consistency between the generated quality index and the ground truth value obtained during the subjective testing process [24], [43]. However, as discussed in [2], such subjective testing process suffers from some problems, such as dissimilar interpretations of the quality scales, and so on. Therefore, regressing the image or image patch into a quality index without reference will be also very difficult. Recent work turns to an alternative approach with pairwise comparison for subjective evaluation [19], [44]. For the pairwise subjective comparison, a participant is simply asked to compare two visual signals simultaneously, and determine which one is of a better perceptual quality. As such, the decision in pairwise comparison is simpler than the traditional scale rating [24], [43]. Same as the pairwise subjective comparison, we also formulate the NR quality assessment as a pairwise comparison problem, with a fundamental departure from the family of existing machine learning method for quality assessment [16], [32], [33], [35], [40].

Inspired by the development of rank learning [45] in information retrieval, we do not regress the image to a specific quality index but learn to rank a pair of images. As such, we would like to develop a quality assessment model targeting at ranking images instead of assigning a quality score to each image. For general settings in information retrieval, it usually ranks the retrieved items by their relevance with the query. However, to our end concerning the quality evaluation of retargeted image, we measure the retargeted image qualities by their orders instead of quality score

$$f_r(x_i, x_j) = \begin{cases} 1, & \text{if} \quad y_i < y_j \\ 0, & \text{if} \quad y_i \geq y_j \end{cases} \tag{1}$$

where x_i and x_j are the image representations of I_i and I_j, respectively. y_i and y_j are the corresponding ground truth quality scores annotated by the human subjects. f_r is the ranking function that has the binary outputs of 0 and 1, representing the inconsistency and consistency between predicted order of image quality and ground truth subjective values respectively. Our target is to learn the ranking function f_r based on the training data. Afterwards, the learned ranking function f_r can help evaluating all the image pairs. We can further generate the image quality scores by referring to a exponential curve fitting function based on the rankings. As such, the image quality ranking information can be converted to its quality index. The traditional statistical measurements [46] can thus be employed for quality metric performance evaluation.

B. Pairwise Rank Learning Framework

The framework of the proposed rank learning based NR quality assessment of retargeted image is illustrated in Fig. 1, consisting of two phases, specifically the training and testing phases. During the training process, we firstly prepare the image pairs for the pairwise rank learning approaches. The retargeted image pairs are prepared and sampled by referring to their subjective quality scores. Specifically, the image with better subjective quality is admitted to rank higher than the one with poorer subjective quality. Afterwards, the generated retargeted image pairs are further represented as feature vectors, which are expected to be not only able to capture the characteristics of the images, but also sensitive to the distortions introduced during the retargeting process. Based on the extracted image feature pairs, we resort to the pairwise rank learning to learn a ranking model, which can discriminate and distinguish the perceptual quality of the two provided retargeted images. Based on the ranking results from the learned model as well as their subjective quality scores, an exponential function can be further fitted to map the rankings to the subjective scores.

For the testing phase, all the possible retargeted image pairs are sampled to measure their perceptual quality rankings. Same as the training process, the image pairs are first represented

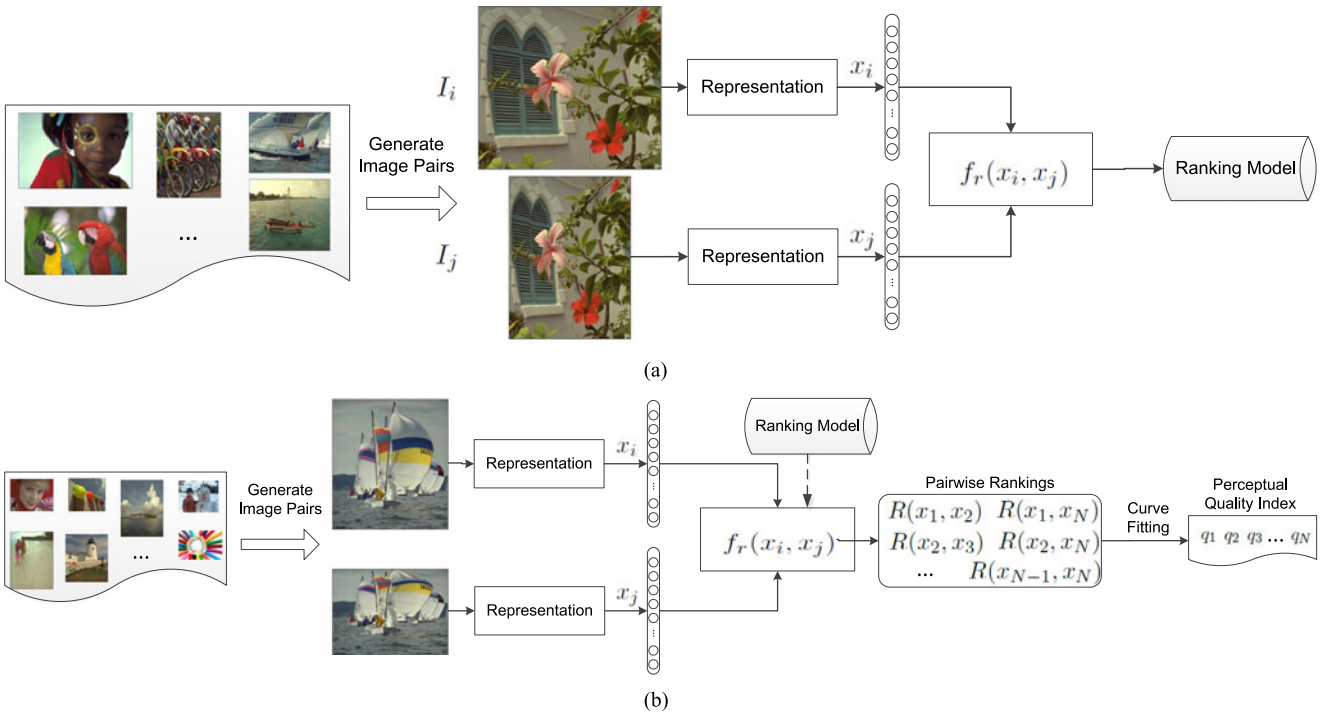

Fig. 1. Pairwise rank learning framework for NR quality assessment of retargeted images, consisting of the (a) training and (b) testing phases. During the training phase, the image representations of the sampled image pairs as well as their subjective quality scores are employed to train a ranking model. During the testing phase, the obtained training model is employed to discriminate perceptual qualities of the test image pairs. Based on the obtained pairwise rankings, the perceptual quality index of each test image is generated via a curve fitting function.

as feature vectors. The ranking model then discriminates their relevant rankings based on the extracted features. With all the possible ranking pairs on the testing images, the fitted exponential function further maps the rankings to the image perceptual quality indexes.

IV. Pairwise Rank Learning Based Quality Assessment for Retargeted Images

In this section, each process during the training and testing phases is introduced in details.

A. Image Representation

For retargeted image quality assessment, the image representation is very important for quality analysis. As discussed in [18], the features should be sensitive to the distortions introduced during the image retargeting process. More specifically, the features should not only capture the shape information, but also depict the image content information loss. The shape information is more important to the quality assessment of the retargeted image. In this paper, we examine different image representations for the purpose of retargeted IQA. We compare different image representations for retargeted images, specifically the BDS [20], [22], SIFT flow [27], earth mover's distance (EMD) [47], MPEG-7 [48], pyramid histogram of visual words (PHOW) [49], GIST [50], and the image representation from the deep neural networks, specifically the VGG [51]. As demonstrated in Section V, GIST outperforms other image representations on retargeted image quality measurement in terms of correlation with HVS perception. The reason is mainly attributed to that GIST is demonstrated to capture the global information. Therefore, it can effectively capture the shape distortions which are introduced during the retargeting process [52] and severely degrade the retargeted image perceptual quality. Detailed information about the different image representations can be found in Appendix.

B. Training Data Preparation

As illustrated in Fig. 1, we need to generate image pairs for both the training and testing processes. We carry out our work on the existing subjective retargeting image database, specifically the CUHK [18] and RetargetMe [19] databases. The two databases provide the subjective score for each retargeted image. Detailed information about this database can be found in Section V-A.

We denote the GIST feature vectors as $\{x_i, i = 1, 2, \ldots, n\}$ and the corresponding subjective scores as $\{y_i, i = 1, 2, \ldots, n\}$ for the retargeted images $\{I_i, i = 1, 2, \ldots, n\}$. y may be the MOS value in CUHK database or the favored times in RetargetMe database. In both cases, the larger value y_i means the better perceptual quality of the retargeted image I_i. To establish the pairwise rank learning task for quality assessment of retargeted images, two images (I_i, I_j) are simultaneously sampled as a pair to form the training set. The GIST feature vectors (x_i, x_j) are used to represent the retargeted images. The subjective scores y_i and y_j are compared to generate their ranking label $\{+1, -1\}$ for the feature vector pair (x_i, x_j). If $y_i < y_j$,

the ranking label is set as $+1$ for the training pair (x_i, x_j). Otherwise, we will assign the ranking label -1 to (x_i, x_j) if $y_i > y_j$.

C. Pairwise Rank Learning

Based on the prepared training data (retargeted image pairs as well as their ranking labels), we introduce the pairwise rank learning approach to learn the ranking model for NR retargeted IQA in this subsection.

Conventionally, the intrinsic principle of machine learning for IQA is explored to optimize the numerical distance between the subjective value y_i (e.g. MOS value) and the predicted image quality index $\psi_\omega(x_i)$

$$\hat{\omega} = \arg\min_{\omega} \left\{ \sum_{i=1}^{n} \parallel \psi_\omega(x_i) - y_i \parallel_p \right\} \tag{2}$$

where ψ_ω denotes the learned function based on the training data and used to compute the image quality index for an input image. ω denotes the parameters of the learned function ψ_ω. x_i represents the feature vector of the i-th image I_i. y_i is the subjective quality score of I_i. $\| \cdot \|_p$ represents p-norm distance. For simplicity, we usually assume that the learned function ψ_ω to be a linear formation

$$\psi_\omega(x) = \mathbf{w}x \tag{3}$$

where $\mathbf{w}$ is one matrix to be learned for mapping the image feature vector x to the predicted perceptual quality $\psi_\omega(x)$. Also nonlinear mapping functions can be employed, which explore the nonlinear and complicated relationships between image features and subjective values. By using kernel functions, nonlinear problems can be converted to linear problems. Observing the optimization objective of (2), the p-norm distance is optimized. However, for our pairwise image ranking function, a new optimization objective based on the binary comparison of image quality is established as

$$\arg\min \left\{ \sum_{i \neq j} \big[y_i < y_j\big]_{\mathbf{I}} \big[\psi_\omega(x_i) \geq \psi_\omega(x_j)\big]_{\mathbf{I}} \right\} \tag{4}$$

where $[\ell]_{\mathbf{I}} = 1$ if the logic decision ℓ holds; otherwise $[\ell]_{\mathbf{I}} = 0$. Instead of the p-norm distance, (4) is established on the ranks of image qualities other than the numerical values. From (4), a false rank prediction, i.e., the order of two images in violation of the ground-truth, would result in the increase of its cost. In order to make a comprehensive utilization of the training data, (4) concerns all possible pairwise comparisons of image qualities among all training images. Obviously, an image which has the distinct quality different from others would contribute more to the defined objective function. Intuitively, if the rank is wrongly predicted, i.e., contradictory to the ground-truth given by the subjective preference, the penalty should be emphasized to refrain from such an occurrence. On the contrary, the images with similar image qualities tend to contribute less to the optimization objective. In practice, we set a threshold $\mathbf{T}$ to realize the task of training data selection by excluding the cases of $|y_i - y_j| \leq T$ in (4) for compressing noise and reducing computational complexity.

For simplicity, we also employ the linear function as in (3) to optimize the objective function defined in (4). As such, the optimization objective is to seek a matrix $\mathbf{w}$, which results in the minimum value of (4) on the generated training set. With a linear function $\psi_\omega(x)$, (4) is reformulated as

$$\arg\min_{\mathbf{w}} \left\{ \sum_{i \neq j} \big[y_i < y_j\big]_{\mathbf{I}} \big[\mathbf{w}x_i \geq \mathbf{w}x_j\big]_{\mathbf{I}} \right\}. \tag{5}$$

We further define an empirical loss $L(\mathbf{w})$ as

$$L(\mathbf{w}) = \sum_{i \neq j} \big[y_i < y_j\big]_{\mathbf{I}} \big[\mathbf{w}x_i \geq \mathbf{w}x_j\big]_{\mathbf{I}}. \tag{6}$$

We aim to minimize $L(\mathbf{w})$ based on the training data. Since $[\ell]_{\mathbf{I}}$ is non-convex, we encounter a non-convex optimization problem. In order to handle such problem, the Boolean term in (5) is replaced by its upper bound to relax the non-convex optimization to the convex one

$$\big[\mathbf{w}x_i \leq \mathbf{w}x_j\big]_{\mathbf{I}} \leq e^{\mathbf{w}(x_i - x_j)}. \tag{7}$$

The exponential upper bound $e^{\mathbf{w}(x_i - x_j)}$ is employed as it is convex and can efficiently facilitate the optimization process. The derivative of the exponential upper bound $e^{\mathbf{w}(x_i - x_j)}$ is obtained by

$$\frac{\partial}{\partial \mathbf{w}} e^{\mathbf{w}(x_i - x_j)} = (x_i - x_j) e^{\mathbf{w}(x_i - x_j)}. \tag{8}$$

After replacing the term containing the variable $\mathbf{w}$ in (5), the empirical loss function $L(\mathbf{w})$ would turn out to be convex. Afterwards, the gradient decedent method can be employed to solve (5). The gradient decent direction of $L(\mathbf{w})$ can be written as

$$\Delta\mathbf{w} = \sum_{i \neq j} \big[y_i < y_j\big]_{\mathbf{I}} (x_i - x_j) e^{\mathbf{w}(x_i - x_j)}. \tag{9}$$

λ acting as an iteration step is further employed to control the convergence speed during the training process, with the multiplication of $\Delta\mathbf{w}$.

For (5), given x_i, y_i, and an initial $\mathbf{w}$, the empirical loss $L(\mathbf{w})$ can be thus initialized. Replacing $\mathbf{w}$ by $\mathbf{w} + \Delta\mathbf{w}$, $L(\mathbf{w})$ can be updated. By iteratively updating $\mathbf{w}$ and $L(\mathbf{w})$, the global minimum objective can be reached, which results in the final learned ranking function for NR retargeted IQA.

D. Image Quality Generation From Ranking Results

It should be pointed out that the optimization objective in (5) is established intrinsically on image quality ranking instead of image quality score. However, it cannot directly output image qualities. For CUHK database, all the MOS values are available during training. The relation between MOS values and their rankings can be fitted by a polynomial function, which helps to output image qualities in the form of quality scores.

In [53], a term "gain" is defined as the number of times of that an image is preferable against the others. For IQA, the "gain" of a retargeted image is proportional to its perceived quality because

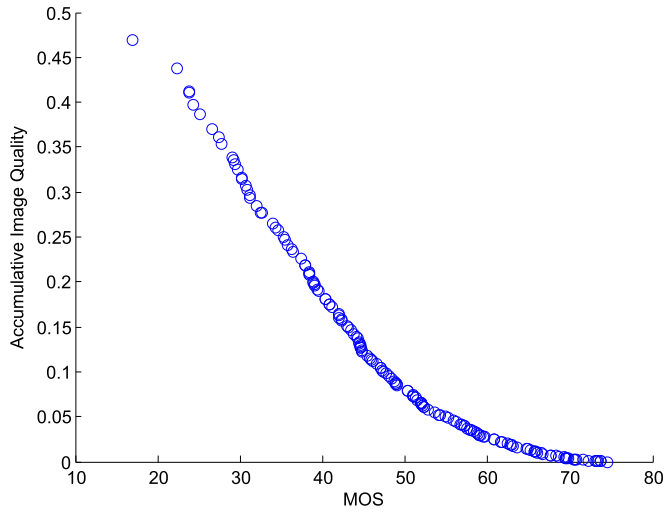

Fig. 2. Scatter plot of the MOS value and accumulative image quality difference.

the preference (better or worse image quality) essentially reflects the relative quality (ranking instead of numerical value) of this image relative to others. Moreover, a linear mapping between the "gain" and the quality score is assumed and can be fitted by training data. However, the differences between the test image and the training image needs to be computed [53], which make it impractical in real applications as the training images are not available. In this paper, our proposed rank function generates a rank list of all the images instead of only binary preference of each image pair. Afterwards, a mapping function between the rank list and image quality scores can thus be fitted. During the training process, a nonlinear fitting function is deduced from mapping the predicted rank list to the subjective score rank list. In the test stage, the nonlinear fitting function can thus tell quality score of each test image without the need of any information from the training set.

The i-th image is compared with other images in the training set. We compute the accumulative image quality difference as

$$d(i) = \frac{\sum_{j\neq i}^{n} \left[\mathbf{w}x_i - \mathbf{w}x_j\right]_+}{\sum_{j\neq i}^{n} \left[\mathbf{w}x_i > \mathbf{w}x_j\right]_{\mathbf{I}}}. \tag{10}$$

The numerator $\sum_{j\neq i}^{n}[\mathbf{w}x_i - \mathbf{w}x_j]_+$ summarizes the perceptual quality differences between x_i and all the other images, which are of lower perceptual quality than x_i. The denominator $\sum_{j\neq i}^{n}[\mathbf{w}x_i > \mathbf{w}x_j]_{\mathbf{I}}$ captures the number of images, which are of lower perceptual quality than x_i. Therefore, (10) generates the accumulative image quality difference $d(i)$ for x_i by representing the relative quality of the i-th image against others. We further illustrate the scatter plot of the MOS value and $d(i)$ of the training images in the CUHK dataset, which is illustrated in Fig. 2.

We can further fit a mapping function between $d(i)$ and image subjective score $y(i)$ by an exponential function as

$$y(i) = \alpha_1 + \alpha_2 \times e^{\alpha_3 \times d(i)}. \tag{11}$$

The parameters α_1, α_2, and α_3 can be easily obtained by the nonlinear least square regression. As such, the image quality score can be easily deduced based on the given image quality rankings. As a accumulative summation is proposed in (10) for generating $d(i)$, the fitting process of (11) is expected to be robust to the noise interference.

V. Experimental Results

In this section, we present the performances of our proposed method and other quality metrics for retargeted images. Firstly, we briefly introduce the image retargeting quality databases. Afterwards, different image representations as well as their effectiveness on retargeted IQA are presented. Finally, the performances of our proposed NR metric in terms of different statistical measurements are illustrated.

A. Image Retargeting Quality Database

As mentioned before, there are two public subjective quality databases for retargeted images, specifically the RetargetMe and CUHK databases.

1) CUHK database [18] is composed by 171 retargeted images in total generated from 57 original images. Each retargeted image is accompanied with its MOS value provided by the subjective viewers, which is denoted as its ground truth perceptual quality.
2) RetargetMe database [19] consists of 37 images. Each image is used to generate the retargeted results using 8 different retargeting algorithms. The ground truth rank of each image is obtained by sorting the number of times that it is preferred over other retargeted images by the subjective viewers.

B. Experimental Results

1) Performance Comparisons Between Different Image Representations: For the CUHK database, each image is accompanied with the MOS value to indicate its perceptual quality. The predicted image quality scores are obtained from the pairwise rankings as introduced in Section IV-D. Afterwards, three measurements, specifically the Spearman's rank correlation coefficient (SROCC), Pearson linear correlation coefficient (PLCC), and outlier ratio (OR) are employed to evaluate the relationships between the ground truth subjective score and the predicted quality score from each quality metric [46]. Table I illustrates the performance comparisons between different image representations. For EMD, the histogram is constructed to represent the feature distribution of the image, which is not able to capture enough information for the retargeted IQA. PHOW can somewhat extract the shape information. However, we need to maintain a visual vocabulary to compose the corresponding histogram at each pyramid scale. Consequently, the shape information is mostly extracted from the local perspective, although a pyramid structure is employed. As illustrated in [18], the global shape information is very important for retargeted image. The local features to evaluate the quality of retargeted image cannot yield a good performance. For the descriptors of MPEG-7, the EHD performs the best. The reason is that the local shape information is depicted by the edge histogram in local regions. The

TABLE I
PERFORMANCE COMPARISONS OF DIFFERENT IMAGE REPRESENTATIONS ON THE CUHK DATABASE

	SROCC	LCC	OR
SCD	0.1792	0.1508	0.2164
CSD	0.1688	0.1520	0.5322
CLD	0.0850	0.1033	0.2398
HTD	0.0890	0.0829	0.5673
EHD	0.2729	0.3031	0.2047
BDS	0.2887	0.2896	0.2161
EMD	0.2904	0.2760	0.1696
PHOW	0.2308	0.3706	0.1579
SIFT flow	0.2899	0.3141	0.1462
VGG	0.3784	0.1728	0.2339
GIST	**0.5114**	**0.5443**	**0.1576**

TABLE II
MEAN KENDALL CORRELATION COEFFICIENTS OF THE OBJECTIVE QUALITY METRICS FOR DIFFERENT IMAGE TYPES

	Lines/ Edges	Faces/ People	Texture	Foreground Objects	Geometric Structures	Symmetry
BDS	0.040	0.190	0.060	0.167	− 0.004	− 0.012
EH	0.043	− 0.076	− 0.060	− 0.079	0.103	0.298
SIFT flow	0.097	0.252	0.119	0.218	0.085	0.071
EMD	0.220	0.262	0.107	0.226	0.237	0.500
CLD	− 0.023	− 0.181	− 0.071	− 0.183	− 0.009	0.214
CSim	0.005	− 0.143	− 0.125	− 0.134	0.000	0.071
Q/Q'	0.159	0.057	0.054	**0.250**	0.250	**0.571**
GIST	**0.229**	**0.273**	**0.218**	0.182	**0.252**	0.484

TABLE III
PERFORMANCE COMPARISON OF DIFFERENT QUALITY METRICS ON THE CUHK DATABASE

	SROCC	PLCC	OR
GIST	0.5114	0.5443	0.1576
GLS (linear regression)	0.4939	0.4402	0.2046
GLS (logistic regression with L1 penalty)	0.4002	0.3961	0.2163
GLS (support vector regression with linear kernel)	0.4038	0.3656	0.2339
GLS (support vector regression with polynomial kernel)	0.3821	0.3711	0.2022
GLS (support vector regression with RBF kernel)	0.3961	0.3658	0.2267
GLS (logistic regression)	0.4760	0.4622	0.1345
Proposed NR metric	0.4926	0.5371	0.1928

global shape information is somewhat captured by concatenating the local edge histograms. But the other descriptors, such as CSD, SCD, and CLD mostly focus on the color component. And HTD concatenates the energy of each frequency channel, which does not pay much attention on the shape information of the image. These are the main reasons why the MPEG-7 descriptors cannot well evaluate the perceptual quality of the retargeted image. SIFT flow depicts the matching relations between the original and retargeted images, which can help depicting the scene difference. VGG uses the convolutional neural network to represent the whole image as one feature vector for classification. As such, VGG and SIFT flow can somewhat depict the image global information, which present good correlations with subjective rating values than MPEG-7 and EMD.

GIST significantly outperforms the other image representations, which may be attributed to two reasons. Firstly, GIST tries to depict each image from different perspectives, such as the horizontal and vertical information, fractal dimensions of the scene, the convergence of parallel lines, oblique contours, and so on. As such, GIST image feature is able to capture the most information which is sensitive to the retargeted image perceptual quality assessment. That is also the main reason why we employ GIST as the image feature to develop our NR quality metric for retargeted image. Secondly, the metric performs in an FR manner. By computing the difference of GIST features from original and retargeted images, the quality can be more accurately captured.

Moreover, we further examine the generality of GIST over different groups of images, according to the image attributes as described in RetargetMe [19]. Besides BDS, EH, SIFT flow, EMD, CLD, two recently developed objective quality metrics CSim [21] and Q/Q' [29] are included for performance comparison. The experimental results are listed in Table II. The best objective metric for each group is highlighted in boldface. It can be observed that GIST can yield the best performances for most of the image groups. The reason is that GIST depicts the global shape information, which can easily capture the distortions introduced to some image attributes, such as the "lines/edges", "faces/people", "geometric structures", and so on. Q/Q' is designed by dealing with salient content preservation and maintaining the symmetry features. Thus it provides the best performances on the image groups with "foreground objects" and "symmetry". Such experimental results also demonstrate the generality of GIST, which is also the main reason that we employ GIST as the image representation to design our NR IQA for retargeted images.

2) Performances of Our Proposed NR Quality Metric: For our proposed NR quality metric, as we need to obtain the parameters during the training process, we employ the standard split for the evaluation. Specifically, the images in the subjective database are randomly divided into training and testing sets. A training set consists of 80% of the reference images and their associated distorted versions, and a testing set consists of the remaining 20% of the reference images and their associated distorted versions. In order to ensure that the proposed method is robust across content and is not biased by the specific train-test split, random 80% train 20% test split is repeated 1000 times.

Table III provides the performance comparisons between our proposed NR quality metric with GIST and GLS [54]. GLS extracts and fuses different features of the retargeted image from the perspectives of the global structural distortion, local region distortion, and loss of salient information. Several regression models are used to fuse these features together as shown in Table III. Comparing the performances in Tables I and III, it can be observed that the proposed NR quality metric outperforms all the other FR metrics except GIST. As we employ GIST as the image feature for quality assessment, the performance of our proposed NR IQA is slightly inferior to that of GIST. For GLS, although the image is depicted from global, local, and salient perspectives, the experimental results demonstrated that GLS

TABLE IV
PERFORMANCE COMPARISONS OF DIFFERENT QUALITY METRICS ON THE RETARGETME DATABASE

	BSD	EHD	EMD	SIFT flow	IR-SSIM	GLS	RB-RIQA	Proposed NR metric
Kendall τ distance	0.083	0.004	0.145	0.251	0.363	0.382	0.503	0.4770

TABLE V
CROSS DATASET PERFORMANCES OF THE PROPOSED PAIRWISE RANK LEARNING METRIC

	SROCC	LCC
Training on CUHK and Testing on RetargetMe	0.3121	0.3264
Training on RetargetMe and Testing on CUHK	0.2973	0.3107

cannot well capture the meaningful features for the retargeted IQA.

For RetargetMe database, the image perceptual quality is indicated by the favored times over the other images. Therefore, same as [19], the Kendall τ distance [55] is employed to evaluate the ground-truth and predicted rankings. The experimental results are illustrated in Table IV. In addition, we also compare the performances with the newly developed quality metrics IR-SSIM [4], GLS (logistic regression with L1 penalty), and RB-RIQA [28]. It can be observed that our proposed metric outperforms other competitor models, except the RB-RIQA metric. RB-RIQA and IR-SSIM try to capture the information loss and the shape distortion in local regions, compared with other metrics, such as EMD, EHD, and so on. GLS considers both the global and local information, yielding better performance than IR-SSIM. However, these quality metrics are all FR, which require the original images for quality analysis. On the contrary, our proposed metric works in a NR fashion, where the original image does not need to be present for quality analysis. As such, the proposed NR metric can be easily adopted for practical applications, such as monitoring the visual QoE during the retargeting process.

C. Cross Dataset Performance

In order to further demonstrate the effectiveness of our proposed NR metric, we perform the cross dataset performances by training on one retargeted image dataset and testing on the other. The results are illustrated in Table V. It can be observed that the performances are not as good as that tested solely on each database. However, compared with the performances in Table I, it can be observed that the proposed NR IQA based on pairwise rank learning outperforms some FR metrics developed by the image representations, such as SIFT-flow, PHOW, EMD, and so on, but performs inferiorly to VGG and GIST. The superior performance of the proposed metric to SIFT-flow, EMD, and so on, is due to the effectiveness of GIST for capturing the shape distortions of retargeted image. The inferior performance to VGG and GIST is because that the NR metric does not have the original image for comparison, which issues great challenges for perceptual quality assessment.

D. Limitations

In this paper, we consider pariwise rank learning for NR IQA of retargeted images. The proposed method has several limitations. Firstly, the training phase needs to know all the distortions to make a more robust training result, which can provide a more reliable result during testing phase. That is the main reason that we cannot provide a good performance when training on one retargeted image database and testing on the other. The retargeted images generated by the retargeted images have not been encountered during the training process. Therefore, the pairwise learning strategy cannot learn such behaviors over the other retargeted images. Secondly, for pairwise rank learning, the pairwise data needs to be performed during the training and testing phase. Suppose that there are m samples, the total number of the prepared pair data is C_m^2, which is a huge number, especially when m is extreme large. How to handle such large number of training samples and make a more reliable training will be of great challenge. In the future, we will consider how to select the most representative pairs from all the possible pairs, which can not only speed up the training process but also make a reliable training.

VI. CONCLUSION AND FUTURE WORK

In this paper, we made the first attempt on the NR quality assessment for retargeted images. The GIST feature extracted for each image as well as its accompanied subjective quality value are employed for the pairwise rank learning approach. The experimental results demonstrate the effectiveness of our proposed NR quality metrics for the retargeted image, which can achieve comparable performances with GIST and significantly outperform other FR metrics.

GIST is employed as the image feature for the pairwise rank learning of retargeted images. In the future, we will consider new image features to more accurately represent the retargeted image. Also we are considering the deep learning methods to learn the features from image pixels for retargeted IQA. With better image features, a better NR IQA for the retargeted image is expected. Moreover, besides the linear relationship in the pairwise rank learning, in the future we will consider to employ a more complicated model, such as deep neural network, to model and learn the nonlinear relationships between image representation and its subjective quality score. Also we will consider how to handle the large scale pairwise training by selecting the most representative pairs from all the possible pairs, in order to not only speed up the training process but also make a reliable training.

APPENDIX
IMAGE REPRESENTATIONS FOR RETARGETED IMAGE

1) MPEG-7 [48] considers different descriptors from the color and texture perspectives, specifically the scalable color descriptor (SCD), color layout descriptor (CLD),

color structure descriptor (CSD), homogeneous texture descriptor (HTD), and edge histogram descriptor (EHD).

2) EMD [47] is based on the minimal cost that must be paid to transform one distribution into the other. The original and retargeted images are represented as histograms. The EMD between these two histograms indicates the retargeted image quality.
3) PHOW [49] is obtained based on the SIFT descriptor and image spatial layout. Multiple descriptors are computed to allow for scale variations between images.
4) BDS [20], [22] captures how much information of the original image is covered and preserved by the retargeted image in a bidirectional manner and measures the newly introduced artifacts.
5) SIFT flow [27] characterizes view-invariant and brightness-independent image structures, which establishes the meaningful correspondences across the original and retargeted images and evaluates the the spatial coherence of the pixel displacement (indicated by the SIFT correspondence matching).
6) GIST [50] is extracted based on a very low dimensional representation of the scene. A set of perceptual dimensions, such as naturalness, openness, roughness, expansion, ruggedness, is employed to represent the dominant spatial structure of a scene, which can accurately depict the global information of the image.
7) VGG [51] employs a deep convolutional neural network to represent each image as a fixed length image feature vector, where convolution and max-pooling processes are employed to compose and summarize the image pixels values to semantic representation for further classification, detection, localization, and so on.

REFERENCES

[1] Y. Fang, Z. Chen, W. Lin, and C. Lin, "Saliency detection in the compressed domain for adaptive image retargeting," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 3888–3901, Sep. 2012.

[2] Q. Xu *et al.*, "Hodgerank on random graphs for subjective video quality assessment," *IEEE Trans. Multimedia*, vol. 14, no. 3, pp. 844–857, Jun. 2012.

[3] B. Li *et al.*, "Spatiotemporal grid flow for video retargeting," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1615–1628, Apr. 2014.

[4] Y. Fang *et al.*, "Objective quality assessment for image retargeting based on structural similarity," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 4, no. 1, pp. 95–105, Mar. 2014.

[5] W. Lin and C. J. Kuo, "Perceptual visual quality metrics: A survey," *J. Vis. Commun. Image Representation*, vol. 22, no. 4, pp. 297–312, 2011.

[6] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," in *Proc. ACM SIGGRAPH 2007 Papers*, 2007, Art. no. 10.

[7] W. Dong, N. Zhou, J.-C. Paul, and X. Zhang, "Optimized image resizing using seam carving and scaling," *ACM Trans. Graph.*, vol. 28, no. 5, pp. 125:1–125:10, Dec. 2009.

[8] M. Rubinstein, A. Shamir, and S. Avidan, "Multi-operator media retargeting," *ACM Trans. Graph.*, vol. 28, no. 3, pp. 23:1–23:11, Jul. 2009.

[9] A. Shamir and O. Sorkine, "Visual media retargeting," in *Proc. ACM SIGGRAPH ASIA 2009 Courses*, 2009, Art. no. 11.

[10] F. Banterle *et al.*, "Multidimensional image retargeting," in *Proc. SIGGRAPH Asia 2011 Courses*, 2011, Art. no. 15.

[11] Y. Pritch, E. Kav-Venaki, and S. Peleg, "Shift-map image editing," in *Proc. Int. Conf. Comput. Vis.*, 2009, pp. 151–158.

[12] Y.-S. Wang, C.-L. Tai, O. Sorkine, and T.-Y. Lee, "Optimized scale-and-stretch for image resizing," *ACM Trans. Graph.*, vol. 27, no. 5, Dec. 2008, Art. no. 118.

[13] L. Ma, S. Li, and K. N. Ngan, "Visual horizontal effect for image quality assessment," *IEEE Signal Process. Lett.*, vol. 17, no. 7, pp. 627–630, Jul. 2010.

[14] S. Li, F. Zhang, L. Ma, and K. N. Ngan, "Image quality assessment by separately evaluating detail losses and additive impairments," *IEEE Trans. Multimedia*, vol. 13, no. 5, pp. 935–949, Oct. 2011.

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

[16] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "Fsim: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

[17] K. Seshadrinathan and A. C. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 335–350, Feb. 2010.

[18] L. Ma, W. Lin, C. Deng, and K. N. Ngan, "Image retargeting quality assessment: A study of subjective scores and objective metrics," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 6, pp. 626–639, Oct. 2012.

[19] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir, "A comparative study of image retargeting," *ACM Trans. Graph.*, vol. 29, no. 6, Dec. 2010, Art. no. 160.

[20] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, "Patchmatch: A randomized correspondence algorithm for structural image editing," *ACM Trans. Graph.*, vol. 28, no. 3, Jul. 2009, Art. no. 24.

[21] Y. Liu, X. Luo, Y. Xuan, W. Chen, and X. Fu, "Image retargeting quality assessment," *Comput. Graph. Forum*, vol. 30, no. 2, pp. 583–592, 2011.

[22] D. Simakov, Y. Caspi, E. Shechtman, and M. Irani, "Summarizing visual data using bidirectional similarity," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2008, pp. 1–8.

[23] L. Ma, W. Lin, C. Deng, and K. N. Ngan, "Study of subjective and objective quality assessment of retargeted images," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2012, pp. 2677–2680.

[24] "BT.500.11: Methodology for the subjective assessment of the quality of television pictures," 2012. [Online]. Available: https://www.itu.int/rec/R-REC-BT.500/en

[25] S. Castillo, T. Judd, and D. Gutierrez, "Using eye-tracking to assess different image retargeting methods," in *Proc. ACM SIGGRAPH Symp. Appl. Perception Graph. Vis.*, 2011, pp. 7–14.

[26] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, Sep. 1999, vol. 2, pp. 1150–1157.

[27] C. Liu, J. Yuen, A. Torralba, J. Sivic, and W. T. Freeman, "Sift flow: Dense correspondence across different scenes," in *Proc.10th Eur. Conf. Comput. Vis.*, 2008, pp. 28–42.

[28] Y. Zhang and K. N. Ngan, "Region-based image retargeting quality assessment," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2015, pp. 1757–1761.

[29] Y. Liang, Y.-J. Liu, and D. Gutierrez, "Objective quality prediction of image retargeting algorithms," *IEEE Trans. Vis. Comput. Graph.*, 2016, to be published. [Online]. Available: doi:10.1109/TVCG.2016.2517641

[30] H. R. Sheikh, A. C. Bovik, and L. K. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1918–1927, Nov. 2005.

[31] T. Brandão and M. P. Queluz, "No-reference image quality assessment based on DCT domain statistics," *Signal Process.*, vol. 88, no. 4, pp. 822–833, 2008.

[32] P. Ye and D. S. Doermann, "No-reference image quality assessment using visual codebooks," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3129–3138, Jul. 2012.

[33] W. Xue, L. Zhang, and X. Mou, "Learning without human scores for blind image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2013, pp. 995–1002.

[34] X. Wen, L. Shao, Y. Xue, and W. Fang, "A rapid learning algorithm for vehicle classification," *Inf. Sci.*, vol. 295, pp. 395–406, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.ins.2014.10.040

[35] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.

[36] B. Gu and V. S. Sheng, "A robust regularization path algorithm for ν-support vector classification," *IEEE Trans. Neural Netw. Learn. Syst.*, 2016, to be published. [Online]. Available: 10.1109/TNNLS.2016.2527796

[37] B. Gu and V. S. Sheng, "Structural minimax probability machine," *IEEE Trans. Neural Netw. Learn. Syst.*, 2016, to be published. [Online]. Available: 10.1109/TNNLS.2016.2544779

[38] B. Gu, V. S. Sheng, K. Tay, W. Romano, and S. Li, "Incremental support vector learning for ordinal regression," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 7, pp. 1403–1416, Jul. 2015. [Online]. Available: http://dx.doi.org/10.1109/TNNLS.2014.2342533
[39] B. Gu, V. S. Sheng, Z. Wang, D. Ho, S. Osman, and S. Li, "Incremental learning for ν-support vector regression," *Neural Netw.*, vol. 67, pp. 140–150, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.neunet.2015.03.013
[40] H. Tang, N. Joshi, and A. Kapoor, "Blind image quality assessment using semi-supervised rectifier networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2014, pp. 2877–2884.
[41] L. Xu *et al.*, "Multi-task rank learning for image quality assessment," *IEEE Trans. Circuits Syst. Video Technol.*, 2016, to be published. [Online]. Available: 10.1109/TCSVT.2016.2543099
[42] L. Xu *et al.*, "Pairwise comparison and rank learning for image quality assessment," *Displays*, vol. 44, pp. 21–26, 2006.
[43] *Methods for Subjective Determination of Transmission Quality*, ITU-T P.800, 1996.
[44] K. Chen, C. Wu, Y. Chang, and C. Lei, "A crowdsourceable QoE evaluation framework for multimedia content," in *Proc. 17th Int. Conf. Multimedia*, 2009, pp. 491–500.
[45] T.-Y. Liu, "Learning to rank for information retrieval," *Found. Trends Inf. Retrieval*, vol. 3, no. 3, pp. 225–331, Mar. 2009.
[46] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.
[47] O. Pele and M. Werman, "Fast and robust earth mover's distances," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep.-Oct. 2009, pp. 460–467.
[48] B. S. Manjunath, J. Ohm, V. V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 703–715, Jun. 2001.
[49] A. Bosch, A. Zisserman, and X. Muñoz, "Image classification using random forests and ferns," in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–8.
[50] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vis.*, vol. 42, no. 3, pp. 145–175, May 2001.
[51] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, 2014. [Online]. Available: http://arxiv.org/abs/1409.1556
[52] L. Ma, L. Xu, H. Zeng, K. N. Ngan, and C. Deng, "How does the shape descriptor measure the perceptual quality of the retargeting image?" in *Proc. IEEE Int. Conf. Multimedia Expo. Workshops*, Jul. 2014, pp. 1–6.
[53] F. Xia, T. Liu, J. Wang, W. Zhang, and H. Li, "Listwise approach to learning to rank: Theory and algorithm," in *Proc. Int. Conf. Mach. Learn.*, 2008, pp. 1192–1199.
[54] J. Zhang and C.-C. J. Kuo, "An objective quality of experience (QoE) assessment index for retargeted images," in *Proc. ACM Int. Conf. Multimedia*, 2014, pp. 257–266.
[55] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, vol. 30, pp. 81–93, 1938.

Lin Ma (M'13) received the B.E. and M.E. degrees from Harbin Institute of Technology, Harbin, China, in 2006 and 2008, respectively, both in computer science, and the Ph.D. degree from the Department of Electronic Engineering, the Chinese University of Hong Kong (CUHK), Hong Kong, in 2013.

He is currently a Senior Researcher with Tencent AI Lab, Shenzhen, China. He was a Research Intern with Microsoft Research Asia, Beijing, China, from October 2007 to March 2008. He was a Research Assistant with the Department of Electronic Engineering, CUHK, from November 2008 to July 2009. He was a Visiting Student with the School of Computer Engineering, Nanyang Technological University, Singapore, from July 2011 to September 2011. He was a Researcher with Huawei Noah's Ark Laboratory, Hong Kong, China. His research interests include deep learning and multimodal learning, specifically image and language, image/video processing, and quality assessment.

Dr. Ma was the recipient of the Best Paper Award at the Pacific Rim Conference on Multimedia 2008. He was awarded the Microsoft Research Asia fellowship in 2011. He was a finalist to HKIS Young Scientist Award in engineering science in 2012.

Long Xu (M'12) received the M.S. degree in applied mathematics from Xidian University, Xi'an, China, in 2002, and the Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China.

He was a Postdoctoral Researcher with the Department of Computer Science, City University of Hong Kong, Hong Kong, China, and the Department of Electronic Engineering, Chinese University of Hong Kong, Hong Kong, China, from July 2009 to December 2012. From January 2013 to March 2014, he was a Postdoctoral Researcher with the School of Computer Engineering, Nanyang Technological University, Singapore. He is currently with the Key Laboratory of Solar Activity, National Astronomical Observatories, Chinese Academy of Sciences. His research interests include image/video processing, solar radio astronomy, wavelet, machine learning, and computer vision.

Dr. Xu was selected in the 100-Talents Plan, Chinese Academy of Sciences, in 2014.

Yichi Zhang received the B.S. degree in information engineering from Shanghai Jiao Tong University, Shanghai, China, in 2013, and is currently working toward the Ph.D. degree at the Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong, China.

His research interests include image/video processing, image quality assessment, and visual saliency.

Yihua Yan received the B.E. and M.E. degrees from Northwestern Polytechnical University, Xi'an, China, in 1982 and 1985, respectively, and the Ph.D. degree from Dalian University of Technology, Dalian, China, in 1990.

He is currently a Professor and Chief Scientist of the Solar Radio Research and the Director of the Key Laboratory of Solar Activity and the Solar Physics Division, National Astronomical Observatories, Chinese Academy of Sciences, Beijing, China. He was a Foreign Research Fellow with the National Astronomical Observatory of Japan, Tokyo, Japan, from 1995 to 1996 and an Alexander von Humboldt Fellow at Astronomical Institute, Wurzburg University, Germany, from 1996 to 1997. He is currently the President of the International Astronomical Union Division E: Sun & Heliosphere for 2015–2018. His research interests include solar magnetic fields, solar radio astronomy, space solar physics, and radio astronomical methods.

King Ngi Ngan (M'79–SM'91–F'00) received the Ph.D. degree in electrical engineering from Loughborough University, Loughborough, U.K.

He was a Full Professor with Nanyang Technological University, Singapore, and the University of Western Australia, Perth, WA, Australia. He is currently a Chair Professor with the Department of Electronic Engineering, Chinese University of Hong Kong, Hong Kong, China. He holds honorary and visiting professorships with numerous universities in China, Australia, and South East Asia. He has published extensively, including three authored books, six edited volumes, and more than 300 refereed technical papers and has edited nine special issues in journals. He holds ten patents in image or video coding and communications.

Dr. Ngan is a Fellow of the IET, U.K., and IEAust, Australia. He was an IEEE Distinguished Lecturer from 2006 to 2007. He has served as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the *Journal on Visual Communications and Image Representation*, the *EURASIP Journal of Signal Processing: Image Communication*, and the *Journal of Applied Signal Processing*. He has chaired a number of prestigious international conferences on video signal processing and communications, and has served on the advisory and technical committees of numerous professional organizations. He co-chaired the IEEE International Conference on Image Processing, Hong Kong, China, in 2010.