2015 IEEE International Conference on Systems, Man, and Cybernetics

Rank Learning Based No-Reference Quality Assessment of Retargeted Images

Lin Ma
Huawei Noah's Ark Lab,
Hong Kong
forest.linma@gmail.com

Long Xu
Key Laboratory of Solar Activity,
National Astronomical Observatories,
Chinese Academy of Sciences
lxu@nao.cas.cn

Yichi Zhang
Department of Electronic Engineering,
The Chinese University of Hong Kong, Hong Kong
yczhang@ee.cuhk.edu.hk

King Ngi Ngan
Department of Electronic Engineering,
The Chinese University of Hong Kong, Hong Kong
knngan@ee.cuhk.edu.hk

Yihua Yan
Key Laboratory of Solar Activity,
National Astronomical Observatories,
Chinese Academy of Sciences
yyh@nao.cas.cn

***Abstract*—In this paper, we first propose a novel no-reference (NR) image quality assessment (IQA) method for retargeted image based on the rank learning approach. Firstly, image features for each retargeted image are extracted, which should not only represent the image characteristics but also be sensitive to the retargeted distortions. Specifically, the image feature should be able to capture the shape distortions, which are the commonly encountered distortions of the retargeted image. Based on the extracted image features, the rank learning method is employed to train a model to discriminate the perceptual quality of the retargeted image. 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 (FR) quality assessment methods.**

Keywords-Retargeted image; image quality assessment (IQA); no-refernce (NR); rank learning

I. INTRODUCTION

Nowadays, with the rapid development of mobile devices, new applications of image/video have appeared in different terminal devices for better visual quality experiences. It is a great challenge to display the same image/video content in all kinds of terminals, such as the mobile phone, tablets, and so on, to provide good visual quality experiences for the viewers. Different terminals have different resolutions, which require a better displaying technique to better meet the viewers' perception. Traditionally, simple scaling and cropping methods are employed to change the image resolution arbitrarily. However, the salient content cannot be well preserved. Thus, several content-aware retargeting methods [1]-[5] are developed to adapt the image to different resolutions while preserve the salient content information of the image. In order to demonstrate the superiority of the retargeting method, simple visual comparisons were conducted for the result comparisons (comparing different retargeting methods based on a small set of images). However, such a method cannot be employed for online optimization and the guidance of the retargeting process. Therefore, developing one automatic quality metric for the retargeted images is very useful, not only for evaluation 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 for image processing, 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). Many quality metrics have been developed to overcome the drawbacks of MSE and PSNR. Structural similarity (SSIM) [6] 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 image and video signals. Recently, more effective quality metrics have been proposed, such as feature similarity (FSIM) [7] for natural image, and motion-based video integrity evaluation (MOVIE) [8] for natural video. Theses quality metrics have achieved great success on quality assessment of traditional distortions, such as JPEG compression, Gaussian noise, and Gaussian blur. However, for newly emerged retargeted images, the developed quality metrics cannot work well.

For retargeted images, the distortions are introduced from the following twofold perspectives [9]. Firstly, the retargeting process will inevitably discard partial information of the image content. How to evaluate the effect of the discarded content on the perceptual quality of the retargeted image is very different from the traditional distorted image. Secondly, other than the traditional noise, such as compression artifact, blurring, one new distortion, namely the shape distortion, is introduced, which will severely degrade the perceptual quality of the image, compared with other traditional distortions. Overall, the retargeted image differs with the original image in resolution and distortion types, making its perceptual quality much more challenging for evaluation.

978-1-4799-8697-2/15 $31.00 © 2015 IEEE
DOI 10.1109/SMC.2015.185

Nowadays, there are several research works discussing the quality assessment of retargeted images. Based on the approaches, the works can be roughly categorized into the subjective [9]-[11] and objective [12]-[15] 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 cannot be applied for online manipulation. Therefore, as the ground truth value of the image perceptual quality can be provided, the subjective testing process is always employed to construct the database, based on which the developed objective quality assessment methods are validated and evaluated. In [11], the authors employed one pair-wise comparison method to indicate which retargeted image possesses a better perceptual quality. The database consists of the retargeted image and the number of times that the image is favored over other images. In [9] [10], the authors employ the simultaneous double stimulus for continuous evaluation (SDSCE) [16] to perform the subjective quality evaluation. Each image with its mean opinion score (MOS) composes the whole database.

For the objective quality metrics, a metric named as bidirectional similarity (BDS) is developed in [19] [20]. Two visual signals, namely the original and retargeted image, 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 can be accurately depicted from the 'completeness' and 'coherance' perspectives. 'Completeness' measures whether all the patches of one visual signal are preserved in the other visual signal. 'Coherence' measures whether there are any 'newborn' patches in one visual signal which do not appear in the other visual signal. BDS can be employed to generate a retargeted image by minimizing the similarity measurement. SIFT flow [21] characterizes the view-invariant and brightness-independent image structures. Matching SIFT descriptors [22] allows establishing meaningful correspondences across images with significantly different image contents. The pixel displacements obtained from the SIFT flow are spatially coherent and therefore the matching cost can indicate the difference between the original and retargeted images. In [13], a critical step is to create an SSIM quality map that indicates at each spatial location of the reference image how the structural information is preserved in the retargeting 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. 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. In [14], the authors employed a top-down manner to organize the image features from global to local viewpoints, leading to a new quality metric for retargeting. A scale-space matching method is designed to facilitate extraction of global geometric structures. And by traversing the scale space from coarse to fine levels, local pixel correspondence is established. By considering both the global geometric structures and local pixel correspondences, the objective quality metric for retargeting image is formulated.

The aforementioned metrics can be regarded as the full-reference (FR) quality assessment (QA) for retargeted images, where the original reference image needs to be present for quality evaluation. However, in practical scenarios, the reference image is always unavailable. Therefore, no-reference (NR) QA methods are demanded. Many research works discuss NR QA methods, which can be roughly classified into three categories. The first category of approaches takes the behavior of specific distortions into consideration. For example, in [23], Sheikh *et al.* employed wavelet statistical model to capture the distortion introduced by JPEG 2000. Brandao *et al.* [24] proposed an NR-IQA approach based on the DCT domain statistics to evaluate the quality of JPEG coded image. The second category of approaches uses quality aware clustering. They group the image patches of training set into the given number of classes based on local image features. 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, the researchers established a codebook. 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 [25], a visual codebook associated Gabor filter based local appearance descriptors with MOS is proposed. The authors of [26] used FSIM [7] instead of MOS as image patch quality to establish the codebook. The third category is to utilize machine learning method to map image features into image qualities. In [28], Moorthy *et al.* proposed to use support vector machine (SVM) and support vector regression (SVR) to learn a classifier and an ensemble of regressors for a distortion-aware IQA metric. It deploys summary statistics called natural scene statistics (NSS) which is derived from wavelet decomposition of an image. In [27], Tang *et al.* proposed an approach similar to [28] but with more elaborate features, including distortion, texture statistics, blur/noise statistics, and histogram of each subband of image decomposition.

However, for retargeted images, to our best knowledge, there is no literature discussing the NR IQA method for retargeted images. In this paper, we first address the NR quality metrics for retargeted images, where a new rank learning approach is proposed. First, the image is represented as a feature vector, which is sensitive to the distortions introduced during the retargeting process. Afterwards, a rank learning method is proposed to discriminate the perceptual quality based on the extracted image features as well as its accompanied quality values (MOS values). Based on the learning results, we can further evaluate the perceptual quality of the retargeted image.

The paper is organized as follows. Section II introduces our proposed NR IQA method for retargeted images. And the experimental results are illustrated in Section III. Finally, conclusions are provided in Section IV.

II. Rank Learning Based No-Referece Quality Assessment for Retargeted Images

A. Framework

The framework of the proposed rank learning based NR IQA method for retargeted images is illustrated in Figure 1.

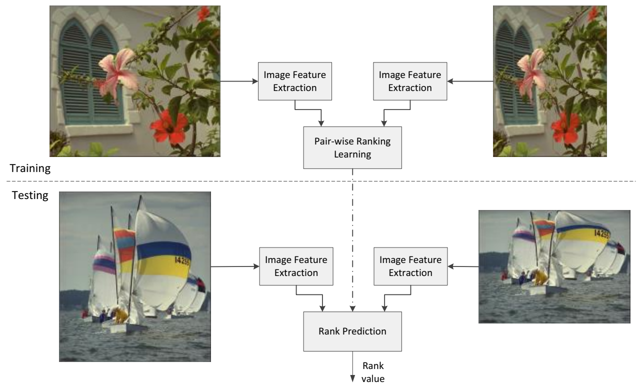

Figure 1. The framework of the proposed rank learning based NR IQA for retargeted images.

There are two phases, specifically the training and testing phases. For the training phase, the image feature needs to be extracted for each image. Afterwards, for each pair of images for training, we have the image features as well as their subjective quality values, based on which the rank learning is applied. The supervision is obtained from their subjective scores. Specifically, the image with better subjective quality should rank higher than the one with poorer subjective quality. With the training process, the parameters can be learned. For the testing phase, the two images should also undergo the image feature extraction process same as the training process. Based on the learned parameters and the two image features, the rank prediction process will output the rank value of these two images.

B. Feature Extraction

For retargeted image quality assessment, the extracted features are very important for quality analysis. As discussed in [9], 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. Moreover, the shape information is more important to the quality assessment of the retargeted image. Therefore, in this paper, we only employ the shape descriptors to capture the shape information of the retargeted image. In order to measure the whole content information of the image, the GIST [32] feature is extracted to depict the global image shape.

The GIST descriptor is extracted based on a very low dimensional representation of the scene, termed as the spatial envelope in [32]. A set of perceptual dimensions, such as naturalness, openness, roughness, expansion, ruggedness, is employed to represent the dominant spatial structure of a scene. For naturalness, the structure of a scene strongly differs between man-made and natural environments. Straight horizontal and vertical lines dominate man-made structures whereas most natural landscapes have textured zones and undulating contours. Therefore, scenes with edges biased toward vertical and horizontal orientation would have a low degree of naturalness. For openness, a scene can have a closed spatial envelop full of visual references or it can be vast and open to infinity. The existence of a horizon line and the lack of visual reference indicate a higher degree of openness in the scene. For roughness, it depends on the size of elements at each partial scale. Roughness is correlated with the fractal dimension of the scene and thus, its complexity. For expansion, the convergence of parallel lines gives the perception of depth gradient of the space. A flat view of a building would have a low degree of expansion. On the contrary, a street with long vanishing lines would have a high degree of expansion. For ruggedness, it refers to the deviation of the ground with respect to the horizon. A rugged environment produces oblique contours in the picture and hides the horizon line. Therefore, rugged environments are mostly natural.

C. Rank Learning

Based on the extracted GIST features, the rank learning method is employed to rank the objective qualities of the retargeted images. We firstly formulate the NR IQA problem as a rank problem. Afterwards, the training data is introduced. The pair-wise rank learning is introduced for the NR IQA of the retargeted image.

1) Problem Formulation

Inspired by the development of rank learning [17] [18] in information retrieval, we make a fundamental departure from the family of existing machine learning method for quality assessment. Regarding learning to rank image perceptual quality, the deduced computer model targets at ranking images instead of assigning a quality score (like PSNR) to each image. Usually, in information retrieval, it ranks the retrieved items by their relevance with the query. To our concerned quality metric for retargeted images, we measure image qualities by their orders instead of quality scores. Thus, the computational model from rank learning is firstly used to rank the retargeted images. Then, a relation between the relative order and MOS can be established by the quadratic polynomial curve fitting. Additionally, the pairwise approach is employed to establish the optimization objective function, where the binary comparisons of MOS values are the ground-truth for training the rank model. The pairwise comparison has the binary outputs of 0 and 1, representing inconsistence and consistence between predicted order of image quality and ground-truth respectively.

2) Traing Data for Rank Learning

We carry out our work on the existing subjective retargeting image database with MOS rating value, specifically the retargeting image database [9], which provides the MOS value for each retargeted image. Detailed information about this database can be found in Section III. For the training process, we denote the GIST feature vectors as $\{x_i\}$ (i=1,2,…,n), and labels $\{y_l\}$ (k=1,2,…,k) given by MOS values. The GIST feature vectors concern high level information of a visual scene, depicting the image global shape information.

To establish the pairwise rank learning task for quality assessment of retargeted images, the training set is derived from existing MOS rating systems. The feature vectors are still the same as the conventional ones extracted from each retargeted image. Labels are given by MOSs, which have a rank order $y_1>y_2>\ldots>y_k$ and each instance x_l is associated with a label y_l. For each pair of images, a binary label $\{+1,-1\}$ represents the case for $y_i \geq y_j$ and the case for the inverse order respectively.

3) Pairs-wise Rank Learning for Retargeted Image Quality Assessment

Support vector machine (SVM) is a good representative of machine learning approaches. It performs with a sophisticated optimization objective, and specifically it optimizes the maximum margin between two classes of samples. With different loss functions, there are lots of variants of SVM, such as L^1-SVM, L^2-SVM and least squares (LS) SVM. We explore the intrinsic principle of machine learning for IQA, by optimizing the numerical distance between predicted image quality $\varphi_\omega(x_i)$ and its MOS value y_i as:

$$\omega^* = \underset{\omega}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} \left\| \varphi_\omega(x_i) - y_i \right\|_p \right\}, \tag{1}$$

where φ_ω is learned from the provided training data, and used to compute image quality for unknown input image. x_i represents the feature vector of the i-th image, y_i is the label of x_i given by MOS, and $\|\bullet\|_p$ represents p-norm operation. The linear form

$$\varphi_\omega(x) = \omega^T x, \tag{2}$$

is widely used in the literature. For fitting in more situations, nonlinear functions are employed, which explore the nonlinear relationships between image features and MOS values. By using kernel functions, nonlinear problems can be converted to linear problems. Observing the optimization objective of (1), the p-norm is optimized, while a new optimization objective is based on binary comparison of image quality is established in this work as

$$\min_{\varphi} \left\{ \sum_{i \neq j} \left[y_i < y_j \right]_{\mathrm{I}} \left[\varphi(x_i) \geq \varphi(x_j) \right]_{\mathrm{I}} \right\}, \tag{3}$$

where $[x]_{\mathrm{I}}=1$ if the logic decision x holds; otherwise $[x]_{\mathrm{I}}=0$. Eq. (3) is established on the ranks of image qualities instead of the numerical values. From (3), a false rank prediction, *i.e.*, the order of two images in violation of the ground-truth, would result in the increase of the cost of Eq. (3). Eq. (3) concerns all pairwise comparisons of image qualities among all images. Obviously, an image which has the distinct quality difference from others would contribute more to objective function of Eq. (3). Intuitively, if the rank is wrongly predicted, *i.e.*, contradictory to the ground-truth given by the MOS preference, the penalty should be large to refrain from such an occurrence. The images with similar image qualities tend to have low weights to the optimization objective. In practice, by assigning a threshold T, we can realize the task of training data selection by excluding the cases of $|y_i\text{-}y_j|>T$ in Eq. (3) for compressing noise and reducing computational complexity.

For simplicity, the linear function as in Eq. (2) is assumed in Eq. (3). Thus, the optimization objective is to seek a vector ω which results in the minimum of Eq. (3) on the training set. With a linear function $\varphi(x)$, Eq. (3) is formulated as:

$$\min_{\omega} \left\{ \sum_{i \neq j} \left[y_i < y_j \right]_{\mathrm{I}} \left[\omega^T x_i \geq \omega^T x_j \right]_{\mathrm{I}} \right\}. \tag{4}$$

Let $L(\omega) = \sum_{i \neq j} \left[y_i < y_j \right]_{\mathrm{I}} \left[\omega^T x_i \geq \omega^T x_j \right]_{\mathrm{I}}$, we call $L(\omega)$ the empirical loss. Since $[x]_{\mathrm{I}}$ is non-convex, we encounter a non-convex optimization problem. As in [33], the Boolean terms related to ω in Eq. (4) is replaced by their upper bounds to facilitate the optimization as:

$$[\omega^T x_i \leq \omega^T x_j]_I \leq \mathrm{e}^{(\omega^T x_i - \omega^T x_j)}, \tag{5}$$

where the exponential upper bound is used since it is convex and can facilitate the optimization. After the only one term containing the variable ω in Eq. (4) being replaced, the empirical loss function would turn out to be convex. Then, the gradient decedent method can be employed to solve Eq. (4). Note that we have

$$\frac{\partial}{\partial \omega} e^{(\omega^T x_i - \omega^T x_j)} = (x_i - x_j)\, \mathrm{e}^{(\omega^T x_i - \omega^T x_j)}. \tag{6}$$

So the gradient decent direction can be written as

$$\Delta\omega = \lambda \times \sum_{i \neq j} \left[y_i < y_j \right]_{\mathrm{I}} (x_i - x_j)\, \mathrm{e}^{(\omega^T x_i - \omega^T x_j)}, \tag{7}$$

where λ acts as an iteration step controlling the convergence speed.

From Eq. (4), given $\{y_i\}$, $\{x_i\}$ and an initial ω, the empirical loss $L(\omega)$ can be initialized. Replacing ω by $\omega+\Delta\omega$, $L(\omega)$ can be updated. By iteratively updating ω and $L(\omega)$, the global minimum objective can be reached. It should be pointed out that the optimization objective Eq. (4) is established intrinsically on image quality ranking instead of image quality score, so it is used for ranking images in terms of their qualities; however, it cannot directly output image qualities. Since all MOS values are available during training, the relation between MOSs and their ranks can be fitted by a polynomial function. This polynomial function can output image qualities in the form of quality scores, which can be further used for the retargeted image quality assessment task.

III. EXPERIMENTAL RESULTS

In this section, we present the performance of our proposed method and other quality metrics for retargeted images. Firstly, we will briefly introduce the subjective retargeting database. Afterwards, the metrics for comparison are introduced. Finally, the performance comparisons in terms of statistical measurements are illustrated.

A. Image Retargeting Subjective Quality Database

As mentioned before, there are two public subjective quality databases for retargeted images. For the subjective database [9], each retargeted image is associated with its MOS value, which can be employed to evaluate the quality metric in a traditional way by matching the correlation between the outputs of the quality metric and the MOS values. There are 171 retargeting images in total generated from 57 original images. The retargeting images are obtained by employing 10 retargeting methods. And the retargeting scales are 50% and 75%.

Table 1. Performance comparison of different quality metrics on the image retargeting database

	LCC	SROCC	OR
SCD	0.1508	0.1792	0.2164
CSD	0.1520	0.1688	0.5322
CLD	0.1033	0.0850	0.2398
HTD	0.0829	0.0890	0.5673
EHD	0.3031	0.2729	0.2047
EMD	0.2760	0.2904	0.1696
PHOW	0.3706	0.2308	0.1579
GIST	0.5443	0.5114	0.1576
Proposed	**0.5371**	**0.4926**	**0.1928**

B. Quality Metrics for Retargeted Images

There are many quality metrics, which can be employed for retargeted image quality assessment. We detailed them as follows.

- MPEG-7 [29] considers many 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).
- Earth mover's distance (EMD) [30] 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.
- Pyramid histogram of visual words (PHOW) [31] is obtained based on the SIFT descriptor and image spatial layout. Multiple descriptors are computed to allow for scale variation between images.

C. Experimental results

We compare the performances of our proposed NR IQA for the retargeted image with other quality metrics, such as EMD, PHOW, and MPEG-7 descriptors. The performance comparison is listed in Table 1. As there is no NR IQA method for retargeted images, we compare our method with the FR quality metrics. For all the other quality metrics except the proposed method, we extract the image features from the reference and retargeted image, respectively. Their difference is computed as the quality index of the retargeted image.

For our proposed NRI QA method, as we need to obtain the parameters during the training process, we employ the standard split for the evaluation. Specifically, the images in the database are randomly divided into training sets 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 on subjective retargeting image database [9].

From the results listed in Table 1, it can be observed GIST can achieve the best performance, which significantly outperforms the other metrics. The reasons may be attributed to two reasons. Firstly, GIST image feature is able to capture the most information which is sensitive to the retargeting image perceptual quality assessment. That is also the main reason that we employ GIST as the image feature for developing our NR IQA 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. 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 image quality assessment. 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 compared with the GIST feature, although a pyramid structure is employed for PHOW. As illustrated in [9], 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 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.

For our proposed method, we only extract GIST from the retargeted image to evaluate its perceptual quality. It can be observed that we can outperform 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 worse than that of GIST. However, our proposed method is an NR approach, which does not require the original image as the reference, making it suitable for online manipulation. For example, we can use the proposed metric to monitor the perceptual quality of the retargeted image during the retargeting process. As such, the image with best perceptual quality can serve as the final retargeting result.

In this paper, we only employ GIST as the image feature. In the future, we will consider new image features to more accurately represent the retargeted image. Also we can turn to deep learning method to scratch the features from image pixels for retargeted image quality assessment. With better image features, a better NR IQA for the retargeted image is expected. Moreover, we only test our proposed method on a small subjective database [9]. In future, we will consider performing the evaluation of our proposed method on [11]. We need to further extend our work on the database, which consists of the retargeted image as well as its preference time over other retargeted images.

IV. CONCOLUTIONS

In this paper, we proposed a rank learning method for NR IQA for retargeted images. GIST feature is firstly extracted for each image and the rank learning approach is performed on the GIST features as well as its accompanied subjective quality values. Experimental results demonstrate that our proposed NR IQA for the retargeted image can achieve comparable performances with GIST and significantly outperform the other FR metrics.

ACKNOWLEDGMENT

This work was partially supported by a grant from the Research Grants Council of the Hong Kong SAR, China (Project CUHK 415913), the National Natural Science Foundation of China under Grant 61202242, 100-Talents Program of Chinese Academy of Sciences (No. Y434061V01).

REFERENCES

[1] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," in *Proc. SIGGRAPH*, 2007.

[2] W. Dong N. Zhou J. C. Paul and X. Zhang, "Optimized image resizing using seam carving and scaling," in *Proc. SIGGRAPH*, 2009.

[3] Y. Pritch E. Kav-Venaki and S. Peleg, "Shift-map image editing," in *Proc. ICCV*, 2009.

[4] M. Rubinstein A. Shamir and A. Avidan, "Multi-operator media retargeting," in *Proc. SIGGRAPH*, 2009.

[5] Y. Wang C. Tai O. Sorkin and T. Lee, "Optimized scale-and-stretch for image resizing," in *Proc. SIGGRAPH Asia*, 2008.

[6] 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.

[7] 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, 2011.

[8] 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.

[9] L. Ma W. Lin C. Deng and K. N. Ngan, "Image retargeting quality assessment: a study of subjecitive scores and objective metrics," *IEEE J. Select. Top. Signal Process.*, vol. 6, no. 6, pp. 626–639, Oct. 2012.

[10] L. Ma, W. Lin, C. Deng, and K. N. Ngan, "Study of Subjective and Objective Quality Assessment of Retargeted Images", in *Proc. ISCAS*, May 20-23, 2012.

[11] M. Rubinstein D. Gutierrez O. Sorkine and A. Shamir, "A comparative study of image retargeting," in *Proc. SIGGRAPH Asia*, 2010.

[12] C. Barnes E. Shechtman A. Finkelstein and D. B. Goldman, "Patchmatch: a randomized correspondence algorithm for structrual image editing," in *Proc. SIGGRAPH*, 2009.

[13] Y. Fang K. Zeng Z.WangW. Lin Z. Fang and C. Lin, "Objective quality assessment for image retargeting based on structrual similarity," *IEEE J. Emerging Select. Top. Circuits Syst.*, 2014.

[14] Y. Liu X. Luo Y. Xuan W. Chen and X. Fu, "Image retargeting quality assessment," in *Proc. EUROGRAPHICS*, 2011.

[15] D. Simakov Y. Caspi E. Shechtman and M. Irani, "Summarizing visual data using bidirectional similarity," *in Proc. CVPR*, 2008.

[16] ITU-R REC. BT.500.11, "Methodoloty for the subjective assessment of the quality of television pictures," http://www.itu.int/rec/ R-REC-BT.500-11-200206-S/en, 2000.

[17] H. Li, Rank learning for information retrieval and natural language processing. *Synthesis Lectures on Human Language Technologies*, 4(1):1-113, 2011.

[18] T. Liu. Rank learning for information retrieval. Foundations and Trends *in Information Retrieval*. 3(3): 225-331, 2009.

[19] C. Barnes E. Shechtman A. Finkelstein and D. B. Goldman, "Patchmatch: a randomized correspondence algorithm for structrual image editing," in *Proc. SIGGRAPH*, 2009.

[20] D. Simakov Y. Caspi E. Shechtman and M. Irani, "Summarizing visual data using bidirectional similarity," in *Proc. CVPR*, 2008.

[21] C. Liu J. Yuen A. Torralba J. Sivic andW. T. Freeman, "Sift flow: dense correspondence across differet scenes," in *Proc. ECCV*, 2008.

[22] D. Lowe, "Object recognition from local scale-invariant features," in *Proc. ICCV*, 1999.

[23] H. R. Sheikh, A. C. Bovik, and L. Cormack, "No-reference quality assessment using nature scene statistics: JPEG 2000", *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1918-1927, Nov. 2005.

[24] T. Brandao, and M. P. Queluz, "No-reference image quality assessment based on DCT domain statistics", *Signal Processing*, vol. 88, no. 4, pp. 822-833, Apr. 2008.

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

[26] W. Xue, L. Zhang, X. Mou, "Learning without Human Scores for Blind Image Quality Assessment," In *Proc.CVPR*, pp. 995- 1002, 2013.

[27] H. Tang, N. Joshi and A. Kapoor, "Learning a Blind Measure of Perceptual Image Quality," In *Proc. CVPR*, Colorado Springs, 2011.

[28] A.K. Moorthy, 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, 2011.

[29] B. S. Manjunath J. R. 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.

[30] O. Pele and M. Werman, "Fast and robust earth mover's distance," in *Proc. ICCV*, 2009.

[31] A. Bosch A. Zisseman and X. Munoz, "Image calssification using random forests and ferns," in *Proc. ICCV*, 2007.

[32] 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, 2001.

[33] J. Li, Y. H. Tian, T. J. Huang and W. Gao, "Cost-Sensitive Rank Learning from Positive and Unlabeled Data for Visual Saliency Estimation," *IEEE Signal Process. Lett.*, vol. 17, no. 6, pp. 591-594, 2010.