HOW DOES THE SHAPE DESCRIPTOR MEASURE THE PERCEPTUAL QUALITY OF THE RETARGETING IMAGE?

Lin Ma[†], *Long Xu*[‡], *Huanqiang Zeng*[♭], *King N. Ngan*[♮], *Chenwei Deng*[♯]

[†]Huawei Noah's Ark Lab, Hong Kong, China,

[‡]National Astronomical Observatories, Chinese Academy of Sciences, Beijing, China

[♭]School of Information Science and Engineering, Huaqiao University, Xiamen, China

[♮]School of Electronic Engineering, the Chinese University of Hong Kong, Hong Kong, China

[♯]School of Information and Electronics, Beijing Institute of Technology, Beijing, China

Email: forest.linma@gmail.com

ABSTRACT

Perceptual quality evaluation of the retargeting image plays an important role in benchmarking different retargeting methods, as well as guiding or optimizing the retargeting process. The distortions introduced during the retargeting process are mainly categorized into shape distortion and content information loss [1]. The shape distortion measurement is critical to the evaluation of retargeting image perceptual quality. In this paper, the performances of different shape descriptors, such as PHOW [2], GIST [3], MPEG-7 descriptors [4], EMD [5], for evaluating the perceptual quality of the retargeting image are examined based on the public image retargeting subjective quality database [6]. Experimental results demonstrated that most of the shape descriptors can hardly capture the characteristics representing the quality of the retargeting image, but the global shape descriptor GIST [3] presents significant performance gains. Moreover, by incorporating with the measurements from the perspective of content information loss, a better performance is further obtained.

Index Terms— Perceptual quality assessment, retargeting image, shape descriptor, shape distortion, content information loss

1. INTRODUCTION

The rapid development of mobile devices has issued great challenges for the visual experiences of end users in different terminals, such as the mobile phone, tablets, and so on. In order for the visual experience optimization, the same source image needs to be displayed on different terminals of different resolutions. Simple scaling and cropping methods can change the image resolution arbitrarily. However, the content information are not considered during the process, where the desired information cannot be well preserved. Recently, content-aware retargeting methods [7] [8] [9] [10] [11] are developed to adapt the image to different resolutions while preserve the content information of the image. In these papers, simple visual comparisons were conducted for the results (comparing different retargeting methods based on a small set of images) to illustrate the effectiveness of the retargeting method. Such a method cannot be employed for online manipulation and extended for large scale comparisons. Therefore, automatic evaluation of the retargeting image quality is desired.

In the past decades, there is a great progress in the research area of objective quality assessment [12] [13] [14]. The well-adopted one is peak signal-to-noise ratio (PSNR) for its simple formulation, clear physical meaning, and easy optimization. However, PSNR is not able to reflect the perception property of human visual system (HVS). Among the developed quality metrics, the most important and well-recognized quality metric is structural similarity (SSIM) [15]. Three perspectives of comparisons are considered for signal comparison, specifically the luminance, contrast, and structure comparison. SSIM is proved to be efficient and effective to evaluate the perceptual quality of the visual signal, including image and video. However, most of the developed quality metrics cannot be simply employed for the retargeting image quality assessment, as the resolutions of the original and retargeting images are not identical. Nowadays, many researchers are working on the retargeting image quality metric for not only benchmarking the retargeting method performances but also guiding the retargeting process.

The quality evaluation of retargeting images can be categorized into subjective [1] [16] and objective approaches [17] [18] [19] [20]. In [16], M. Rubinstein *et al.* built a subjective database concentrating on a comparative study of existing retargeting methods. The subjetive test is performed in a pair-wise comparison manner. Two retargeted images are shown to the viewers (with or without the original image) side by side. The viewers simply indicate which one is of a better perceptual quality. The database comprises the retargeted image and the number of times that the image is fa-

vored over another retargeting image. In [1] [6], L. Ma *et al.* built the subjective quality database by referring the simultaneous double stimulus for continuous evaluation (SDSCE) as specified in [21]. There are 171 retargeting images in total in the database. And the perceptual quality of the each retargeting image is subjectively rated by at least 30 viewers, meanwhile the mean opinion scores (MOS) were obtained, which can be regarded as the ground truth for further evaluating the performances of the quality metrics. For the objective quality metrics, a metric named as bidirectional similarity (BDS) is developed in [17] [20]. Two visual signals, specifically the original and retargeting 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 'complete' and 'cohere' perspectives. 'Complete' measures whether all patches of one visual signal are preserved in the other visual signal. 'Cohere' measures whether there are any 'newborn' patches in one visual signal which do not appear in the other visual signal. Also the BDS metric can be employed to generate a retargeting image by minimization of the similarity measurement. SIFT flow [22] characterizes the view-invariant and brightness-independent image structures. Matching SIFT descriptors [23] allows establishing meaningful correspondences across image with significantly different image content. Also the pixel displacement (indicating by the SIFT correspondence matching) should be spatial coherent, which means that close-by pixels should have similar displacement. In [18], 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. Each pixel in the original image, the best matching pixel in retargeting image is firstly located. The SSIM measurement is calculated between the local regions of the original and retargeting 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 [19], 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 image. 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 structure and local pixel correspondence, the objective quality metric for retargeting image is developed.

As discussed in [1] [16], the distortions introduced in retargeting image can be roughly categorized into shape distortion and content information loss. According to the subjective evaluation result, the shape distortion impacts more heavily on the final perceptual quality of the retargeting image. Therefore, a good shape distortion measurement will significantly benefit the quality metric for retargeting image. In this paper, the authors will examine recently developed and well-recognized shape descriptors. These shape descriptors are compared between the original and retargeting images to depict the shape distortions for the quality evaluation.

The rest of the paper is organized as follows. Section 2 will review the shape descriptors for quality evaluation of retargeting image. The implementation details and experimental results are introduced in Section 3. Finally, discussion and conclusion are presented in Section 4 .

2. SHAPE DESCRIPTORS FOR IMAGE REPRESENTATION

Recently, in order to develop an effective and efficient image retrieval, recognition, detection, such computer vision tasks, the image is always firstly represented as a vector consisting many different features. And the geometric and shape information is very important to depict the property of the object in the image. Therefore, many research works discussed how to represent the object shape or geometric information, which will be greatly helpful for image representation. As discussed before, shape information is critical to the perceptual quality of the retargeting image. In this section, we will review recent works on shape descriptors for image representation. Based on the representation (a feature vector representing shape information), quality analysis can be performed to generate the final quality score of the retargeting image.

2.1. MPEG-7 descriptors

MPEG-7 [4] considered many descriptors from the color and texture perspectives, such as scalable color descriptor (SCD), color layout descriptor (CLD), color structure descriptor (CSD), homogeneous texture descriptor (HTD), and edge histogram descriptor (EHD). Detailed information is introduced in the following.

- SCD is defined in the hue-saturation-value (HSV) color space with fixed color space quantization, and uses a novel Haar transform encoding. The Haar transform based encoding facilitates a scalable representation of the description, as well as complexity scalability for feature extraction and matching procedures.

- CSD expresses local color structure in an image using an 8×8-structuring element. It counts the number of times a particular color is contained within the structuring element as the structuring element scans the image. Suppose $c_0, c_1, c_2, ..., c_{M-1}$ denote the M quantized colors. A color structure histogram can then be denoted by $h(m), m = 0, 1, ..., M-1$, where the value in each bin represents the number of structuring elements in the image containing one or more pixels with color c_m. The hue-min-max-difference (HMMD) color is used for CSD extraction.

- CLD specifies the spatial distribution of colors. The extraction for the descriptor consists of four stages; image partitioning, dominant color selection, DCT transform, and non-linear quantization of the zigzag-scanned DCT coefficients. In the first stage, an input picture is partitioned into 64 blocks. The size of the each block is $W/8 \times H/8$, where W and H denote the width and height of an input picture, respectively. In the second stage, a single dominant color is selected in each block to build a tiny image whose size is 8×8. Any method for dominant color selection can be applied. Simple average colors is calculated as the dominant colors. In the third stage, each of the three components (Y, Cb, Cr) is transformed by 8×8 DCT, and we obtain three sets of DCT coefficients. A few low frequency coefficients are extracted using zigzag scanning and quantized to form the CLD for a still picture. Image-to-image or sketch-to-image search can be implemented by calculating a distance of the descriptors.

- HTD is computed by first filtering the image with a bank of orientation and scale sensitive filters, and computing the mean and standard deviation of the filtered outputs in the frequency domain. Specifically, the frequency space is partitioned into 30 channels with equal divisions in the angular direction and octave division in the radial direction. The individual feature channel is modeled using 2-D Gabor function. Then the image texture in each of the filtered channels is computed. The HTD is extracted by concatenating mean intensity, the standard deviation of the image texture, the energy, and energy deviation.

- EHD captures the spatial distribution of edges in the image. In order to depict the local edge distribution, the image is divided into 4×4 sub-images, each of which is examined by 5 different orientations: vertical, horizontal, two diagonals, and isotropic (non-directional). For each sub-image, a normalized 5-bin histogram is obtained by classifying apparent edges to these five categories. The feature is defined to be the combination of these histograms, which results in $4 \times 4 \times 5 = 80$ length description. Only the intensity component is employed for edge detection. And the L_1-norm distance is employed to measure the feature distance between two images, which is defined as $EH(S,T) = \| EHF(S) - EHF(T) \|_1$, where EHF is the edge histogram feature.

2.2. Earth mover's distance (EMD)

EMD is based on the minimal cost that must be paid to transform one distribution into the other. The signature $\{S_j = (m_j, w_j)\}$, which represents a set of feature clusters, is viewed as the histogram distribution. The point m_j is the central value in bin j of the histogram, and w_j is to indicate the corresponding proportion. The definition of cluster is open. The color, position, and texture information can be employed to obtain the feature clusters. Only the size of the clusters in the feature space needs to be limited. Let $P = \{(p_1, w_{p_1}), \cdots, (p_m, w_{p_m})\}$ be the first signature with m clusters; $Q = \{(q_1, w_{q_1}), \cdots, (q_n, w_{q_n})\}$ is the second signature with n clusters. And $D = [d_{ij}]$ is the ground distance matrix, where d_{ij} is the ground distance between clusters p_i and q_j. d_{ij} can be any distance and will be chosen according to the problem at hand. The purpose is to find a flow $F = [f_{ij}]$, with f_{ij} as the flow between p_i and q_j, that minimizes the overall cost:

$$WORK(P, Q, F) = \sum_i^m \sum_j^n d_{ij} f_{ij} \quad (1)$$

After obtaining the optimal flow F, EMD is defined as the work normalized by the total flow:

$$EMD(P, Q) = \frac{\sum_i^m \sum_j^n d_{ij} f_{ij}}{\sum_i^m \sum_j^n f_{ij}} \quad (2)$$

2.3. Pyramid Histogram of Visual Words (PHOW)

First, SIFT descriptors [23] are computed at points on a regular grid with spacing M pixels. At each grid point the descriptors are computed over four circular support patches with different radii. Consequently, each point is represented by four SIFT descriptors. Multiple descriptors are computed to allow for scale variation between images. The dense features are vector quantized into N visual words using K-means clustering. Based on the SIFT descriptor and image spatial layout, we can obtain the PHOW representation. In forming the pyramid the grid at level l has 2^l cells along each dimension. Consequently, level 0 is represented by N-vector corresponding to the N bins of the histogram, level 1 by a $4N$-vector *etc*. The PHOW is a vector with dimensionality of $N \sum_{l=0}^{L} 4^l$. Detailed information about the implementation can be found in Section 3.

2.4. GIST Descriptor

GIST descriptor is extracted based on a very low dimensional representation of the scene, which is termed as the *Spatial Envelope* in [3]. 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

Fig. 1. Examples of the original (left most) and retargeting images (the rest).

reference confer to the scene a high degree of openness. For roughness, it depends on the size of elements at each spatial 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. Detailed information of the implementation can also be found in Section 3.

3. EXPERIMENTAL RESULTS

In this section, the performances of the shape descriptors introduced in the previous section are compared and examined. Before the performance evaluation, the subjective database will be briefly introduced. Consequently, the implementation detail of each shape descriptor will be introduced. Finally, the performance comparisons in terms of statistical measurements are illustrated.

3.1. Image retargeting subjective quality database

As mentioned before, there are only two public subjective quality databases for retargeting images. For the database [16], the authors mainly focused on the comparisons of retargeting methods, which generate images together with its favored times over the other retargeting images. For the database [6], each retargeting 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. Therefore, for simplicity and easy comparison, the subjective database [6] is employed. 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%. As demonstrated in [1], the MOS values present a unique distribution, which cover both the low and high perceptual quality, which is reasonable for quality metric evaluation. Some original image and its retargeting images are illustrated in Fig. 1.

3.2. Implementation Details of the Shape Descriptors

For the MPEG-7 descriptors, we employ the public MPEG-7 low level feature extraction tools ([Online]. Available: `http://www.cs.bilkent.edu.tr/~bilmdg/bilvideo-7/Software.html`) to extract the corresponding descriptors, such as SCD, CSD, CLD, HTD, and EHD. According to the default settings, the lengths of the MPEG-7 descriptors are different. The lengthes of SCD, CSD, CLD, HTD, and EHD are 128, 64, 120, 62, and 80, respectively. For the EMD, the code ([Online]. Available: `http://www.seas.upenn.edu/~ofirpele/FastEMD/`) is employed to depict the perceptual quality of the retargeting image. For PHOW, SIFT descriptors are first computed at points on a regular grid with spacing M pixels, here $M = 10$. At each grid point the descriptors are computed over circular support patches with radii $r = 4, 8, 12$ and 16 pixels. The patches with radii 4 do not overlap and the other radii do. As all the images are color, this process will provide a 128×3 D-SIFT descriptor for each point. These descriptors are rotation invariant. The K-means clustering is performed over 2000 training images. Finally the vocabulary consisting of 2000 visual words is used here. Then the vocabulary is used to extract PHOW features, which generates a vector with the dimensionality as 2000. The authors employed VLFeat [24] to extract the PHOW descriptors. For GIST, the code provided by the authors ([Online] Available: `http://people.csail.mit.edu/torralba/code/spatialenvelope/`) is employed. As a result, the dimension of the GIST feature is 960.

3.3. Experimental Results

The performances of the shape descriptors are evaluated based on the subjective quality database by depicting the relationship of the obtained metric outputs and the provided MOS values. As suggested by video quality experts group (VQEG) HDTV test [25], we follow the procedure to evaluate the performance of each metric. Let x_j represent the visual quality index of the j-th retargeting image obtained from the corre-

Table 1. Performances of different shape descriptors on the image retargeting database

	LCC	SROCC	RMSE	OR
SCD	0.1508	0.1792	13.347	0.2164
CSD	0.1520	0.1688	32.731	0.5322
CLD	0.1033	0.0850	13.429	0.2398
HTD	0.0829	0.0890	35.151	0.5673
EHD	0.3031	0.2729	12.866	0.2047
EMD	0.2760	0.2904	12.977	0.1696
PHOW	0.3706	0.2308	12.540	0.2222
GIST	**0.5443**	**0.5114**	**11.326**	**0.1579**
PHOW+GIST	0.5440	0.5090	11.329	0.1579
MPEG-7 features	0.1164	0.1502	24.357	0.4094
ALL	0.1168	0.1504	24.257	0.4094
Combination	**0.5999**	**0.5609**	**10.801**	**0.1228**

sponding metric. Five parameters $\{\beta_1, \beta_2, \beta_3, \beta_4, \beta_5\}$ monotonic logistic function is employed to map x_j into V_j:

$$V_j = \beta_1 \times \left(0.5 - \frac{1}{1+e^{\beta_2 \times (x_j - \beta_3)}}\right) + \beta_4 \times x_j + \beta_5 \quad (3)$$

The corresponding five parameters are determined by minimizing the sum of squared differences between the mapped objective scores V_j and the MOS values. In order to evaluate the performances, four statistical measurements are employed. The linear correlation coefficient (LCC) measures the prediction accuracy. The Spearman rank-order correlation coefficient (SROCC) provides an evaluation of the prediction monotonicity. The root mean square prediction error (RMSE) is introduced for evaluating the error during the fitting process. The outlier ratio (OR) evaluates the consistency attributes of the objective metric, which represents the ratio of 'outlier-points' to the total points. According to the definitions, larger values of LCC and SROCC mean that the objective and subjective scores correlate better, that is to say, a better performance of the metric. And the smaller RMSE and OR values indicate smaller errors between the two scores, therefore a better performance.

The performances of different shape descriptors are illustrated in Table 1. Firstly, we compared the performances of each shape descriptors. It can be observed the GIST can achieve the best performances, which significantly outperforms the other shape descriptors. The reason is that GIST consider the image shape from several perspectives, such as naturalness, openness, roughness, expansion, and ruggedness. By considering the shape information from these perceptual dimensions, the object shape can be accurately depicted. Therefore, as some retargeting methods significantly degrade the shape information, such as seam-carving [7], the distortions introduced can be more precisely captured. Moreover, GIST is regarded as a global descriptor, which is believed to be able to capture more shape information from the global viewpoint compared with other shape descriptors. For EMD, the composed histogram only represents the feature distribution of the image, which cannot accurately depict the object shape information. PHOW can somewhat extract some shape information. However, a visual vocabulary is introduced to compose the corresponding histogram at each pyramid scale. Therefore, the shape information is mostly extracted from the local perspective, although a pyramid structure is employed for PHOW. The global shape information cannot be accurately described. 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 region. The global shape information is somewhat captured by concatenating the local edge histogram. For the other shape descriptors, CSD, SCD, and CLD mostly focus color part. Although color can somewhat represent the shape information, the accuracy cannot be ensured by the color features. HTD concatenates the energy of each frequency channel, which does not pay much attention on the shape description. These are the main reasons why the other shape descriptors cannot depict the perceptual quality of the retargeting image, compared with GIST.

Secondly, we test the performance by combining these shape descriptors together. The MPEG-7 feature combines CSD, SCD, CLD, EHD, and HTD together. PHOW+GIST concatenates PHOW and GIST together to generate a vector feature with the dimensionality as 2960. ALL is evaluated by adding all the shape features (MPEG-7 features, GIST, PHOW) together. From the experimental results in Table 1, the performances of these combinations cannot outperform its best component. Therefore, we cannot expect better performance by simply combining as many as shape descriptors. The shape descriptors may conflict with each other for evaluating the retargeting image perceptual quality.

Finally, as discussed in [1], the distortion introduced in retargeting process can be categorized into shape distortion and content information loss. And the measurements for these two distortions are complementary for each other. By combining these measurements together, a better performance is expected. In this paper, we combine shape descriptors (GIST, and EMD) together with content information loss measurements (BDS and SIFT-flow). The combination process is a simple summation process. The quality score is obtained by $Q = \alpha \times log_2(GIST) + log_2(EMD) + log_2(BDS) + log_2(SIFT - flow)$, where α is simply set as 10. With the evaluation on the database, the best performance is observed, which greatly outperforms GIST.

4. DISCUSSION

Based on the experimental results presented in Section 3, firstly it should be noted that more shape descriptors are not able to ensure a better performance. Some shape descriptors, such as HTD and CLD, cannot effectively evaluate the retargeting

image quality. Therefore, shape descriptors should be carefully selected for metric design which needs to represent the shape distortions introduced by retargeting. Secondly, some shape descriptors, such CLD, HTD, CSD, and SCD, focuses on color/energy distribution or layout. Although they can somewhat capture the shape information, the structure of the retargeting image is preferred to be beneficial for retargeting image quality measurement. Moreover, GIST is a global shape descriptor, which significantly outperforms the other descriptors. The other descriptors, such PHOW and EHD, tried to depict the global information in a bottom-up manner, where the local information are grouped or concatenated together to represent global information. The experimental results demonstrate the the quality evaluation of retargeting image should concentrate on the global information. It truly matches the HVS property. During the subjective test, the viewer's preference is highly affected by the global shape information. If the shape information appears to be very annoying globally, the subjective score will be absolutely very low, no matter how well the local shape information is preserved. While the global shape information is well preserved, the viewer will then clearly check the local shape information. Therefore, during the quality metric design, global shape information should be of the highest priority. In the following, the local shape information is considered to be complementary. Furthermore, it can be observed that the content loss information descriptors are very helpful, which should be included for quality metric design.

5. CONCLUSION

In this paper, the authors examined recent shape descriptors for quality evaluation of retargeting images. With extensive experiments on the image retargeting quality database, we concludes that the global shape descriptor, specifically the GIST, outperforms the other shape descriptors for depicting shape distortion of retargeting image. By combining with content loss information descriptors, the performance of the quality metric can be further improved.

6. ACKNOWLEDGMENT

The work was partially supported by a grant from the Research Grants Council of the Hong Kong SAR, China (Project CUHK 415913); the National Nature Science Foundation of China under Grant No. 61301090; the Supporting Program for Beijing Excellent Talents under Grant No. 2013D0090110000001, and the National Natural Science Foundation of China under Grant No. 61202242.

7. REFERENCES

[1] 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.

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

[3] 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.

[4] 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.

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

[6] L. Ma W. Lin C. Deng and K. N. Ngan, "Image retargeting subjetive quality database," `http://ivp.ee.cuhk.edu.hk/projects/demo/retargeting/index.html`, 2012.

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

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

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

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

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

[12] W. Lin and C.-C. Jay Kuo, "Perceptual visual quality metrics: a survey," *J. Vis. Commun. Image Represent.*, vol. 22, no. 2, pp. 297–312, 2011.

[13] Z. Wang and A. C. Bovik, "Mean squared error: love it or leave it? - a new look at fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, Jan. 2009.

[14] Z. Wang and A. C. Bovik, "Modern image quality assessment," Synthesis Lectures on Image, Video & Multimedia. Morgan & Claypool Publishers, 2006.

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

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

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

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

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

[21] 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.

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

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

[24] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," `http://www.vlfeat.org/`, 2008.

[25] VQEG HDTV, "Final report from the video quality experts group on the validation of objetive models of multimedia quality assesssment, phase 1," `ftp://vqeg.its.bldrdoc.gov/Documents/Projects/multimedia/MM_Final_Report/`.