# Study of Subjective and Objective Quality Assessment of Retargeted Images

Lin Ma[†], Weisi Lin[‡], Chenwei Deng[‡], and King N. Ngan[†]
[†]Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, NT., Hong Kong
[‡]School of Computer Engineering, Nanyang Technological University, Singapore
Email: {lma, knngan}@ee.cuhk.edu.hk; {cwdeng, wslin}@ntu.edu.sg

***Abstract*— This paper presents the result of a recent large-scale subjective study of image retargeting quality on a collection of images generated by several representative image retargeting methods. Owning to many approaches to image retargeting that are developed, there is a need for a diverse independent public database of the retargeted images and the corresponding subjective scores that is freely available. We build an image retargeting quality database, in which 171 retargeted images (obtained from 57 natural source images of different contents) were generated by several representative image retargeting methods. The perceptual quality of each image is evaluated by at least 30 human subjects and the mean opinion scores (MOS) were recorded. Furthermore, several publicly available quality metrics for the retargeted images are evaluated on the built database. The database is made available [1] to the research community in order to further research on the perceptual quality assessment of the retargeted images.**

## I. Introduction

The diversity and versatility of the display devices have imposed new demands on digital images. The same images need to be displayed with different resolutions on variant devices. The image retargeting methods [2]-[9] are proposed to adjust the source images into arbitrary sizes and simultaneously keep the salient content of the source images. In order to retarget an image with good quality, quality assessment of retargeted images should be performed and used to maximize the perceptual quality during the retargeting process. Therefore, a new challenge of evaluating the retargeted image perceptual quality is issued, where the resolution has been changed, the objective shape may be distorted, and some content information may be discarded.

Given that the ultimate receivers of images are human eyes, the human subjective opinion is the most reliable value for indicating the image perceptual quality. The subjective opinions are obtained through the subjective testing, where a large number of viewers participate in the testing and provide their personal opinions of the image quality on some predefined scale. After processing these subjective scores across the human subjects, a score is finally generated to indicate the perceptual quality of the image. The subjective testing method is time-consuming and expensive, which makes it impractical for most image applications. However, the obtained subjective rating value can be recognized as the ground truth of the image perceptual quality. Therefore, they can be employed to evaluate the performances of the objective quality metrics, which can evaluate the image/video quality automatically [11]-[13]. Moreover, subjective studies can also enable the improvement in the performance of the quality metric towards attaining the ultimate goal of matching human perception. Therefore, there is a clear need to build an image retargeting database with subjective testing results, based on which we can evaluate the current developed quality metrics for retargeted images.

Till now, the only publicly available subjective image retargeting database is built by M. Rubinstein et al. [10]. The main purpose of building the database concentrates on a comparative study of existing retargeting methods. The subjective testing is performed in a pair comparison way, where the participants are shown two retargeted images at a time, side by side, and are asked to simply choose the one they like better. It is distinct from the traditional subjective testing [11]-[13], where the mean opinion score (MOS) or difference mean opinion score (DMOS) of each image/video is obtained. Therefore, the perceptual quality metric for retargeted images cannot be evaluated in the standardized way [14] [15]. Moreover, as only the number of the times that the retargeted image is favored over another image is recorded, the actual perceptual quality of the image is not clearly indicated. For one image with a larger number of favored times, it may possess a low perceptual quality if it is compared with the images of even lower perceptual qualities. The image may be favored the most by comparing with other images, whereas its perceptual quality may still not be acceptable. It is also the main reason why the quality metric cannot be evaluated in the standardized way. The most serious shortcoming of the database is that participants have difficulties to arrive at an agreement on the perceptual quality of the retargeted image. The Kendall $\mu$-coefficient [16] obtained of all the images is only 0.095. It is a relatively low value, suggesting that the subjects in general had difficulty judging. Therefore, the images in the database are not suitable for the subjective testing and also not appropriate to evaluate the performances of the quality metrics.

In this paper, we first present a study that we conducted to assess the subjective quality of retargeted images. 171

This work was done while L. Ma was with Nanyang Technological University as a visiting student.

978-1-4673-0219-7/12/$31.00 ©2012 IEEE 

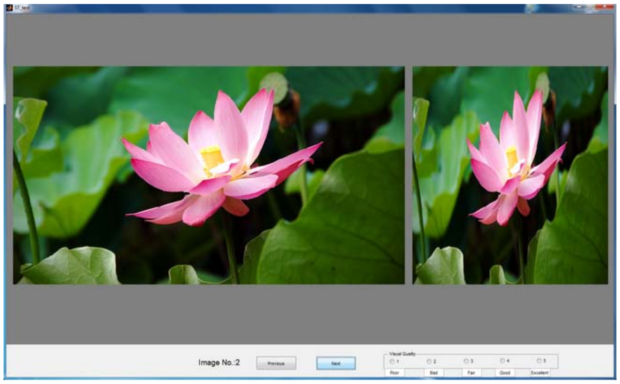


Figure 1. Screenshot of the subjective study interface displaying the images to the human subject.

retargeted images are generated by different retargeting methods from 57 source images. With the source image as the reference, the perceptual quality of each retargeted image is subjectively rated by at least 30 human viewers on a predefined scale. After processing the subjective ratings, the MOS value and the corresponding standard deviation are computed for each image. Based on the MOS values, some publicly available quality metrics for retargeted images are evaluated on the built database in the standardized way.

The rest of the paper is organized as follows. In Section II, we will introduce the subjective testing process for building the image retargeting database. In Section III, some objective quality metrics are introduced and evaluated over the built database. Finally, the conclusion will be given in Section IV.

## II. Detailes of Subjective Testing

### A. Source Image

Content-aware retargeting methods generate images with high perceptual quality where some background content can be removed or efficiently compacted, and the clear foreground object will be preserved. In order to build a reasonable image retargeting database, we need to consider the source images containing the frequently encountered attributes, such as the face and people, clear foreground object, natural scenery (containing smooth or texture region), and geometric structure (evident lines or edges). For building our database, we select 57 source images in which the frequently encountered attributes have been included. The corresponding resolutions of source images are diverse, in order to alleviate the influence of the image resolution on the subjective testing. The source images are roughly categorized into four classes according the aforementioned attributes. It should be pointed out that one image may contain more than one attribute. Detailed information about the source image can be found in [1].

### B. Retargeting Methods

In order to efficiently demonstrate the perceptual quality of the retargeted images, the resolution changes are restricted in only one dimension. The retargeting methods change the resolution of the source images in either the width or height dimension. Two resizing scales (the source image is shrunk to 75% and 50%) are considered to generate the corresponding retargeted images. In our database, three retargeted results of each source image are included. They may be in different retargeting scales. The reason why the database is built in this way is that we only care about the perceptual quality of the retargeted images, no matter how it is generated and what the resolution is. We have considered ten retargeting methods which are recently developed. These retargeting methods are: cropping (CROP), scaling (SCAL), seam carving (SEAM) [4], optimized seam carving and scale (SCSC) [9], Non-homogeneous retargeting (WARP) [2], scale and stretch (SCST) [6], shift-map editing (SHIF) [7], multi-operator (MULT) [5], energy-based deformation (ENER) [8], and steaming video (STVI) [3]. As we do not focus on the comparisons of different retargeting methods as [10] did, only three retargeting results of each source image are selected.

### C. Subjective Testing

During the subjective testing, the source image should be presented to the subjective viewers as the reference. In this paper, the subjective study conducted is simultaneous double stimulus for continuous evaluation (SDSCE), as specified in [17]. Two images are juxtaposed on the screen for the human subject. One is the source image for reference and the other is the retargeted image to be evaluated. The human subjects are aware of which one is the reference image and which one is the retargeted image. The subjective testing uses the ITU-R absolute category rating (ACR) scale [18]. The user interface for the subjective testing is developed by using MATLAB, as shown in Figure 1. In order to avoid strong visual contrast, the remaining regions of the display area are gray (pixel values are set equal to 128). The quality scales have been labeled to help the human subjects to do the quality evaluation. The quality scales are labeled as “Bad”, “Poor”, “Fair”, “Good”, and “Excellent”, which ranges from 1 to 5.

In order to reduce the effect of the viewer fatigue, the 171 retargeted images are divided into 2 sessions. The first session contains 69 retargeted images, while the second contains 102 retargeted images. The order of the image pairs (the source image and the retargeted image) is randomly arranged. Furthermore, in order to avoid the contextual and memory effects on the subjects’ judgment of the quality, the retargeted images which are generated from the same source image will not be presented consecutively. In order to prevent the scaling effect, which is critical to the image retargeting results, the source image and the retargeted image must be displayed in their native resolution. In our experiment, the resolution of the screen for subjective testing is 1920×1280.

All the subjects participating in the subjective testing are the students from the Chinese University of Hong Kong in Hong Kong, and Nanyang Technological University in Singapore. They have normal vision (with or without corrective glasses) and have passed the color blindness test. 30 subjects attend the first session, where 15 viewers are experts in image processing. And 34 subjects attend the second session, where 18 viewers are experts in image processing.

### D. Screening of the Observers

In order to obtain the final MOS and standard deviation value for each image, the subject rejection process is suggested by [17]. Let $S_{ijk}$ denotes the subjective rating by the subject $i$ to the retargeted image $j$ in session $k = \{1,2\}$. The $S_{ijk}$ values are firstly converted to $Z$-scores per session [19]:

$$\mu_{ik} = \frac{1}{N_{ik}} \sum_{j=1}^{N_{ik}} S_{ijk}$$
$$\sigma_{ik} = \sqrt{\frac{1}{N_{ik}-1} \sum_{j=1}^{N_{ik}} \left(S_{ijk} - \mu_{ik}\right)^2}, \quad (1)$$
$$z_{ijk} = \frac{S_{ijk} - \mu_{ik}}{\sigma_{ik}}$$

where $N_{ik}$ is the number of the test images seen by the subject $i$ in session $k$. After converting the obtained subjective ratings into $Z$-scores, the subject rejection procedure specified in [17] is then used to discard scores from unreliable subjects. The algorithm first determines whether the scores assigned by a subject are normally distributed by computing the kurtosis $\beta_j$ of the scores:

$$\beta_j = \frac{m_4}{(m_2)^2} \quad with \quad m_\Delta = \frac{\sum_{j=1}^{N_{ik}} \left(S_{ijk} - u_{ik}\right)^\Delta}{N_{ik}}, \quad (2)$$

If the kurtosis value $\beta_j$ falls between 2 and 4, the scores are regarded to be normally distributed. The subject rejection procedure is depicted in Algorithm 1. By performing the algorithm, 1 out of 30 subjects and 3 out of 34 subjects are rejected in session 1 and 2, respectively.

Algorithm 1. Subject rejection process

For each subject $i$, find the $P_{ik}$ and $Q_{ik}$

if $2 \le \beta_j \le 4$ (normally distributed)
    if $S_{ijk} \ge u_{ik} + 2\sigma_{ik}$, then $P_{ik} = P_{ik} + 1$;
    if $S_{ijk} \le u_{ik} - 2\sigma_{ik}$, then $Q_{ik} = Q_{ik} + 1$;
else
    if $S_{ijk} \ge u_{ik} + \sqrt{20}\sigma_{ik}$, then $P_{ik} = P_{ik} + 1$;
    if $S_{ijk} \le u_{ik} - \sqrt{20}\sigma_{ik}$, then $Q_{ik} = Q_{ik} + 1$;
end

if $\frac{P_{ik}+Q_{ik}}{N_{ik}} > 0.5$ and $\left|\frac{P_{ik}-Q_{ik}}{P_{ik}+Q_{ik}}\right| < 0.3$, **REJECT** the subject $i$.

After subject rejection, $Z$-scores are then linearly rescaled to lie in the range of [0,100]. Assuming that $Z$-scores assigned by a subject are distributed as a standard Gaussian [12], 99% of the scores lie in the range [-3,+3]. Re-scaling is accomplished by linearly mapping the range [-3,+3] to [0,100] by:

$$\tilde{z}_{ijk} = \frac{100(z_{ijk}+3)}{6}. \quad (3)$$

Finally, the MOS value of each retargeted image is computed as the mean of the rescaled $Z$-scores, the corresponding variance as the standard deviation, which is illustrated in Figure 2.

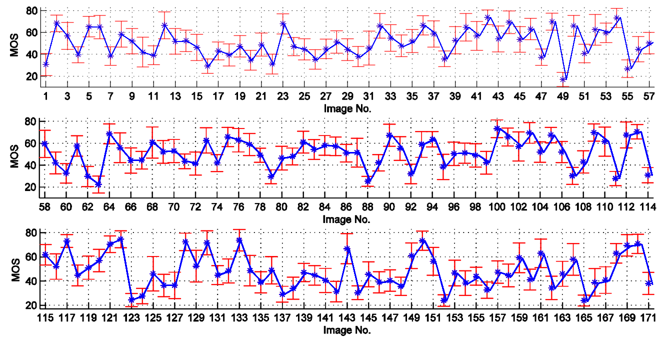


Figure 2. The obtained MOS value of each retargeted image after processing (the horizontal axes corresponds to the image number as shown in [1], and the vertical axes corresponds to the MOS value. The blue star indicates the obtained MOS value. And the red error bar indicates the standard deviation of the subjective scores).

Table I. Performances of different metrics on the built image retargeting database

| | EH | EMD | BSD | SIFT-flow | Fusion |
|---|---|---|---|---|---|
| LCC | 0.3422 | 0.2760 | 0.2896 | 0.3141 | 0.5013 |
| SROCC | 0.3288 | 0.2904 | 0.2887 | 0.2899 | 0.4578 |
| RMSE | 12.686 | 12.977 | 12.922 | 12.817 | 11.682 |
| OR | 0.2047 | 0.1696 | 0.2164 | 0.1462 | 0.1345 |

The algorithms are provided by the authors, which are tested on our built image retargeting quality database. The performance can be evaluated by depicting the relationship of the obtained metric values and the provided MOS values. As suggested by video quality experts group (VQEG) HDTV test [27] and that in [28], we follow their evaluation procedure for evaluating the performances of the metrics. Let $x_j$ represent the visual quality index of the $j$-th retargeted image obtained from the corresponding metric. The five parameter $\{\beta_1, \beta_2, \beta_3, \beta_4, \beta_5\}$ monotonic logistic function is employed to map $x_j$ into $V_j$:

$$V_j = \beta_1 \times \left(0.5 - \frac{1}{1+e^{\beta_2 \times (x_j - \beta_3)}}\right) + \beta_4 \times x_j + \beta_5. \quad (4)$$

The corresponding five parameters are determined by minimizing the sum of squared differences between the mapped objective score $V_j$ and the MOS value. In order to evaluate the performance, 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 utilized 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 smaller RMSE and OR values indicate smaller errors between the two scores, therefore a better performance. The performances of different metrics are illustrated in Table I.

## III. Objective Quality Metric for Retargeted Image

Image retargeting quality metric has been recently researched [20]-[26], in order to not only evaluate the retargeted image quality automatically and reliably instead of the subjective testing, but also to help improving the performances of the retargeting methods. One problem is that several quality metrics are licensed or patented, such as the bidirectional warping in [5] and the quality metric in [25], which are not made publicly. In this paper, we only tested the metrics which are publicly available and suggested in [10], specifically the earth mover's distance (EMD) [20] [21], the bidirectional similarity (BDS) [22][23], edge histogram (EH) [26], and SIFT-flow [24].

It can be observed that all of the metrics perform poorly on the built database. For the EMD, the composed histogram only represents the feature distribution of the image, which cannot accurately depict the object shape and the content information of the image. Therefore, the shape distortions and content information loss, introduced during the retargeting process, are not effectively described. BDS tries to capture how much information one image conveys of the other image in a bidirectional way. However, although it is claimed that the spatial geometric relationship is considered through a multiple scale approach, the order-relationship can still not be preserved, such as the local-order of each pixel or patch. Therefore, the dissimilarity metric of BDS does not accurately depict the object shape distortion either. SIFT-flow employs the SIFT descriptor to detect the correspondences between two images. It is claimed that the order-relationship of the pixels or patches is captured. However, the content information loss during the retargeting process is not considered. EH employs the edge histograms to describe the image, which are organized in order for comparison. EH can somehow represent the object shape information of the image. Same as the SIFT-flow, the content information loss is not accounted. These are the reasons why the metrics cannot perform effectively on the built image retargeting database.

The perceived quality of the retargeted image depends on both the distortions perceived in the retargeted image and how much information conveyed of the source image. Therefore, the descriptors of shape distortion and content information loss should be fused together for evaluating the retargeted image quality. For the current available metrics, EH, EMD and SIFT-flow try to capture the object shape of the image. BSD tries to depict the content information loss in a bidirectional way. If they are combined together, these two distortions are fused to build a quality metric, the performance of which is illustrated in Table I. It can be observed that by fusing these metrics together, a better performance can be obtained. However, the performance is still not good enough, which attributes to two reasons. First, the employed metrics, such as SIFT-flow, and BSD, cannot accurately depict levels of the shape distortion and content information loss. Second, for different images shape distortion and content information loss may play different roles on influencing the perceptual quality. Therefore, in the future, two following aspects will be researched in order to develop an effective quality metric for retargeted images. One is to find the accurate descriptors for depicting the shape distortion and content information loss. The other is to develop an adaptive fusion method to combine the aforementioned descriptors together.

## IV. Conclusion

An image retargeting database is built through the subjective study in this paper. Based on the subjective ratings of the human viewers, the publicly available quality metrics for the retargeted images are evaluated. By fusing the metrics together, which independently depict shape distortion and content information loss, the performance can be improved. The image retargeting subjective database has been made publicly for the research community.

## References


[1] Image Retargeting Subjective Database: [Online] Available: http://ivp.ee.cuhk.edu.hk/projects/demo/retargeting/index.html

[2] L. Wolf, M. Guttmann, D. Cohen-Or, "Non-homogeneous content-driven video-retargeting", in *Proc. ICCV* 2007.

[3] P. Krahenbuhl, M. Lang, A. Hornung, and M. Gross, "A system for retargeting of streaming video", in *Proc. SIGGRAPH Asia* 2009.

[4] A. Shamir, and S. Avidan, "Seam-carving for media retargeting", *Communication of the ACM*, vol. 52, no. 1, pp. 77-85, Jan. 2009.

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

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

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

[8] Z. Karni, D. Freedman, and C. Gotsman, "Energy-based image deformation", in *Proc. Symposium on Geometry Processing* 2009.

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

[10] M. Rubinstein, et al., "A comparative study of image retargeting", in *Proc. SIGGRAPH Asia* 2010.

[11] A. K. Moorthy, et al., "Wireless video quality assessment: a study of subjective scores and objective algorithms", *IEEE Trans. Circuits Syst. Video Techno.*, vol. 20, no. 4, pp. 587-598, Apr. 2010.

[12] K. Soundararajan, et al., "Study of subjective and objective quality assessment of video", *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1427-1441, Jun. 2010.

[13] P. L. Callet, et al., Subjective quality assessment IRCCyN/IVC database. [Online] Available: http://www2.irccyn.ec-nantes.fr/ivcdb/.

[14] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full ref-erence image quality assessment algorithms", *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440-3451, Nov. 2006.

[15] VQEG. Final report from the video quality experts group on the validation of objective models of video quality assessment, 2000. Available [Online]: http://www.vqeg.org.

[16] M. G. Kendall, and B. Babington Smith, "On the method of paired comparisons", *Biometrika*, vol. 31, pp.324-345, 1940.

[17] ITU-R Recommendation BT.500-11, "Methodology for the subjective assessment of the quality of television pictures", International Telecommunications Union, Tech. Rep., 2000.

[18] Video Quality Experts Group. [Online]. Available: http://www.its.bldrdoc.gov/vqeg/.

[19] A. M. van Dijk, et al., "Quality assessment of coded images using numerical category scaling", in *Proc. SPIE Advanced Image and Video Communications and Storage Technologies*, 1995.

[20] O. Pele, and M. Werman, "Fast and Robust Earth Mover's Distances", in *Proc. ICCV*, 2009.

[21] Y. Rubner, et al. "The earth mover's distance as a metric for image retrieval". *IJCV*, vol. 40, no. 2, pp. 99-121, 2000.

[22] D. Simakov Yaron Caspi et al., "Summarizing visual data using bidirectional similarity", in *Porc. CVPR*, 2008.

[23] C. Barnes, et al., "Patchmatch: a randomized correspondence algorithm for structural image editing", in *Proc. SIGGRAPH*, 2009.

[24] C. Liu, et al., "SIFT flow: dense correspon-dence across different scenes", in *Proc. ECCV*, 2008.

[25] Y. Liu, X. Luo, Y. Xuan, W. Chen, and X. Fu, "Image Retargeting Quality Assessment", in *Proc. EUROGRAPHICS*, 2011.

[26] B. S. Manjunath, J. R. Ohm, V. V. Vasudevan, and A. Yamada, "Color and texture descriptors", *IEEE Trans. Circuits Syst. Video Techno.*, vol. 11, no. 6, pp. 703-715, Jun. 2001.

[27] Final report from the video quality experts from group on the validation of objective models of multimedia quality assessment, Phase 1. ftp://vqeg.its.bldrdoc.gov/Documents/Projects/multimedia/MM_Final_Report/.

[28] H. R. Sheikh, et al., "A statistical evaluation of recent full reference image quality assessment algorithms", *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440-3451, Nov. 2006.