Image Retargeting Quality Assessment: A Study of Subjective Scores and Objective Metrics

Lin Ma, *Student Member, IEEE*, Weisi Lin, *Senior Member, IEEE*, Chenwei Deng, *Member, IEEE*, and King Ngi Ngan, *Fellow, IEEE*

***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 have been developed, there is a need for a diverse independent public database of the retargeted images and the corresponding subjective scores to be freely available. We build an image retargeting quality database, in which 171 retargeted images (obtained from 57 natural source images of different contents) were created by several representative image retargeting methods. And the perceptual quality of each image is subjectively rated by at least 30 viewers, meanwhile the mean opinion scores (MOS) were obtained. It is revealed that the subject viewers have arrived at a reasonable agreement on the perceptual quality of the retargeted image. Therefore, the MOS values obtained can be regarded as the ground truth for evaluating the quality metric performances. The database is made publicly available (Image Retargeting Subjective Database, [Online]. Available: http://ivp.ee.cuhk.edu.hk/projects/demo/retargeting/index.html) to the research community in order to further research on the perceptual quality assessment of the retargeted images. Moreover, the built image retargeting database is analyzed from the perspectives of the retargeting scale, the retargeting method, and the source image content. We discuss how to retarget the images according to the scale requirement and the source image attribute information. Furthermore, several publicly available quality metrics for the retargeted images are evaluated on the built database. How to develop an effective quality metric for retargeted images is discussed through a specifically designed subjective testing process. It is demonstrated that the metric performance can be further improved, by fusing the descriptors of shape distortion and content information loss.**

***Index Terms*—Image quality assessment, image retargeting, objective metric, subjective evaluation.**

I. Introduction

THE diversity and versatility of the display devices have imposed new demands on digital image processing. The same image needs to be displayed with different resolutions on variant devices. The image retargeting methods [7]–[17] have been proposed to adjust the source images into arbitrary sizes and simultaneously keep the salient content of the source images. These developed methods, such as seam carving [10]–[12], warp [8], and multi-operator [13], try to preserve the salient shape and content information of the source image, and shrink (or expend) the unimportant regions of the image into the given resolution. For most of these methods, a simple visual comparison was conducted for the results (comparing the results of different retargeting methods based on a small set of images) to demonstrate the efficiencies of the retargeting methods. Such a method cannot be used for on-line manipulation. In order to obtain 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, there is a new challenge of objectively evaluating the retargeted image perceptual quality, 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 pre-defined 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 evaluate the image quality automatically [1]–[5]. Moreover, subjective studies can also enable the improvement in the performance of the quality metric towards attaining the ultimate goal of matching human perception. Then the developed quality metric can be utilized for guiding the corresponding application. Furthermore, the subjective studies can also benefit the image applications for better perceptual quality experience, specifically improving the perceptual quality of the retargeted image. Therefore, there is a 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.* [18]. The main purpose of building the database concentrates on a comparative study of existing retargeting methods. The authors compared

Manuscript received November 02, 2011; revised April 29, 2012 and July 29, 2012; accepted July 31, 2012. Date of publication August 07, 2012; date of current version September 12, 2012. This work was done while L. Ma was with Nanyang Technological University as a visiting student. This work was supported in part by the Singapore Ministry of Education Academic Research Fund (AcRF) Tier 2 under Grant T208B1218. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Amy Reibman.

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

W. Lin and C. Deng are with the School of Computer Engineering, Nanyang Technological University, Singapore 639798 (e-mail: wslin@ntu.edu.sg; cwdeng@ntu.edu.sg).

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/JSTSP.2012.2211996

which retargeting method generates the retargeted image with the highest perceptual quality. The subjective test 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. The resulting database comprises the retargeted image and the corresponding number of times that the retargeted image is favored over another one. This is distinct from the traditional subjective testing [1]–[4], 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 by the correlation between subjective MOS/DMOS values and the metric outputs [19], [20], where the statistical measurements are used to depict the matching score between metric values and MOS/DMOS values.

Moreover, as only the number of times that the retargeted image is favored over another image is recorded, the actual perceptual quality of the image is not clearly indicated. For the traditional subjective testing methods [1]–[4], the perceptual quality of the retargeted image can be directly indicated by the MOS or DMOS value. It is also the main reason why the quality metric cannot be evaluated as introduced in [19], [20]. Furthermore, the total number of possible paired comparisons is too large. It is unaffordable and unrealistic for one human subject to complete all the comparisons. Therefore, the authors in [18] sample the space of possible comparisons for each individual subject, which ensures that a satisfactory subset of the comparisons is built for each individual subject and several complete comparisons are obtained by accounting all the subjects' results. According to this sample strategy, no subjects perform a complete comparison between the retargeted images of each source image. Different subjects may have different interpretations of the retargeted image quality, which may affect the robustness of the subjective ratings. However, the most serious shortcoming of the database is that subjects have difficulties to arrive at an agreement on the perceptual quality of the retargeted image. While statistical tests demonstrate that viewers agreed more often than chance, the Kendall μ-coefficient [24] obtained of all the images is only 0.095. It is a relatively low value suggesting that the subjects in general had difficulty judging.

In this paper, a subjective study is conducted to assess the perceptual quality of the retargeted image to build a publicly available database. Totally 171 retargeted images (in two different scales) are generated by different retargeting methods from 57 source images. With the source image as the reference, the perceptual quality of each retargeted image has been subjectively rated by at least 30 human viewers on a pre-defined scale. After processing the subjective ratings, the MOS value and the corresponding standard deviation are obtained for each image. Based on the MOS values, the built image retargeting database is analyzed from the perspectives of the retargeting scale, the retargeting method, and the source image content. Moreover, some publicly available quality metrics for retargeted images are evaluated on the built database. Furthermore, a specifically designed subjective testing process is carried out to provide further information for developing an effective quality metric for retargeted images.

Our constructed database mainly focuses on evaluating perceptual quality of the retargeted images other than pair-wise comparing the retargeting methods [18]. Therefore, based on the built database, the objective quality metrics can be evaluated by the correlation value between the subjective MOS values and the objective metric outputs. While for the database [18], only the Kendall τ distance [44] is employed to measure the degree of correlation between two rankings. Same as traditional image/video quality assessments where multiple image/video databases [1]–[4], [45] were created, the image retargeting quality assessment also requires multiple image databases. When constructing different databases, different subjects participated in the subjective testing with different rating scales. Meanwhile, the source image content and image distortions introduced by retargeting are quite different. In these respects, the subjective quality databases can be ensured to be of great diversity, which can be employed to evaluate the effectiveness and robustness of the developed objective quality metric. Therefore, our built database and the one in [18] can be further viewed as complementary to each other.

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, the obtained subjective ratings will be processed and analyzed. In Section IV, some objective quality metrics are introduced and evaluated on the built database. Finally, Section V will conclude the paper.

II. Preparation of Database Building

A. Source Images

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. However, for some images with geometric structures and faces, the perceptual quality of the retargeted image cannot be ensured. 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). The detailed information of the attributes can be referred to the **supplementary** [6].

In order to build the 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. Fig. 1 illustrates some samples of the source images for generating the retargeted images. 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 attributes. For example, the image 'umdan' contains the attributes of people and geometric structure. The image 'bicycle1' contains the attributes of clear foreground object, people, and natural scenery. And the image 'fishing' contains the attributes of people and natural scenery. The attribute information of the source image can be found in the **supplementary** [6]. As the image retargeting methods

Fig. 1. Samples of the source images utilized in the subjective testing. The images in the top row mostly contain the attribute of face and people; the images in the second row mostly contain the attribute of clear foreground object; the images in the third row mostly contain the attribute of natural scenery; the images in the bottom row mostly contain the attribute of geometric structure.

are content-aware, the perceptual qualities of retargeted results from different source images will be different. The attributes of the images are critical to the perceptual quality of the final retargeted images. The human subjects are very sensitive to the distortion of the faces and geometric structures, while they can tolerate more distortions on the natural scenery, especially for the texture regions. By including the images with different attributes, the subjective database can reflect how the retargeted images are favored by the human subjects.

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. As shown in [7]–[17], most of the retargeting methods generate the retargeted images in two ratios, shrinking the image to 75% and 50%. Therefore, only these two retargeting ratios are employed to generate the retargeted image for constructing our database. In the built 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 image, no matter how it is generated and what the resolution is. For some source images, the retargeted results in 50% scale appear very high perceptual quality, which perfectly preserve the salient information of the source image. For some source images, even the retargeted images in the 75% scale are of low perceptual quality. For the subjective testing of different scales separately, how scale influences the perceptual quality may not be clearly revealed. Therefore, it is more reasonable by mixing retargeted images with different scales together to examine its perceptual quality through subjective testing. Ten recently developed retargeting methods are employed to generate the retargeted images, which are detailed in the following.

- Cropping (CROP): manually choosing a window of the target size from the source image to maximize the salient information.
- Scaling (SCAL): simple scaling the source image into the target size.
- Seam carving (SEAM) [10]–[12]: removing the contiguous chains of pixels that lie in the regions of the smallest gradient magnitude values in the source image. The dynamic programming is employed to find the seams for removing.
- Optimized seam carving and scale (SCSC) [17]: a measurement named as "seam carving distance" is proposed to measure the similarity of retargeted image and the source one. A combination of linear scaling and seam carving is considered to optimize the measurement.
- Non-homogeneous retargeting (WARP) [8]: a warping function is optimized to find the optimal squeezed image by reducing the image width. The gradient magnitude together with the face detection is employed to indicate the saliency region of the source image, which needs to be preserved with high priority during the retargeting process.
- Scale and stretch (SCST) [14]: an objective function is optimized by uniformly scaling the salient regions to preserve the shape information. The saliency map is detected by combining the gradient magnitude and the saliency map detected by Itti *et al.* [21].
- Shift-map editing (SHIF) [15]: graph cut is used to remove an entire object at a time rather than a seam. The smoothness is depicted by the color differences and the gradient information.

- Multi-operator process (MULT) [13]: seam carving, scaling, and cropping are combined together to generate the retargeted image. And a bi-directional warping measurement determines how to choose these operators.
- Energy-based deformation (ENER) [16]: similar as the SCST method, warping is also used to generate the retargeting image.
- Streaming video (STVI) [9]: the warping method is also used. The saliency map is obtained by combining the visual attention map, the line detection, and important objects.

Referring to these retargeting methods, it can be observed that the cropping, scaling, seam carving, and warping are the basic tools for image retargeting. Many research works are proposed to combine these tools together by optimizing a defined objective measurement. As the foreground objects, including the faces and people, represent the most salient information to the human viewers, the saliency map is incorporated into retargeting. It can be utilized to guide the image retargeting by preserving the shape information in the salient regions.

With these 10 retargeting methods, if each source image is to be retargeted into two different aspect ratios (75% and 50%), there should be 20 retargeted results for each source image. However, some retargeting methods, such as SCSC [17], MULT [12], ENER [16], and STVI [9], do not provide the source code or executable file. Therefore, we can only include the retargeted results provided by the developers of the corresponding retargeting methods. For some source images, the retargeted results cannot be generated. Including all of 20 retargeted images seems thus impossible. Secondly, we do not aim at comparing the performances between different image retargeting methods as the paper [18] did. Therefore, we need not include all the retargeted images at each retargeting ratio into our database. The database we built mainly focus on evaluating the perceptual quality of the retargeted image. It needs only to be ensured that the perceptual qualities of the selected images are sampled in an approximately uniform fashion as shown in [1], [2]. In this respect, 3 retargeted images for each source image are manually selected according to the coarse judgment of the authors. Although 3 out of 20 seems a bit sparse sampling, different retargeted images obtained by different methods are selected, whose perceptual qualities are expected to be distributed uniformly from low to high qualities. The built database demonstrates a uniform distribution and good separation of the perceptual quality, as will be illustrated in the following section.

C. Subjective Testing

ITU-R BT.500-11 [22] has specified several methodologies for the subjective assessment of the quality of television pictures. These methods can be roughly categorized into two types: the double stimulus and single stimulus approaches. The double stimulus approach asks the subjective viewers to rate the quality or change in quality between two videos/images (reference and impaired). For the single stimulus approach, the subjective viewers only rate the quality of just one impaired video/image. As discussed in [38], each subjective test methodology has its own advantages. The double stimulus approach is claimed to be less sensitive to the context, where the subjective ratings are less influenced by the severity and ordering of the impairments within the test session. The single stimulus approach yields more representative quality estimates for quality monitoring. Also the single stimulus approach can ensure a faster and more efficient subjective testing process [40], compared with the double stimulus one.

However, for our subjective testing process of retargeted images, we not only care about the distortions perceived in the retargeted image, but also how much information of the source image has been conveyed. Therefore, in order to provide more convincing results, the source image needs to be presented to the subjective viewers as the reference simultaneously. Otherwise, if we employ the single stimulus approach, the CROP method will always yield the best quality, as no distortions are introduced. Without the source image as the reference, the subjective viewers are not able to detect the discarded information, which may be the most important part of the source image. Therefore, in this paper, the simultaneous double stimulus for continuous evaluation (SDSCE) as specified in [22] is employed. 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. The subjects are requested to check the difference between the two images and judge the perceptual quality of the retargeted one. After that, they provide their own opinions on the retargeted image quality.

The only difference of the subjective testing in this paper was the use of the ITU-R absolute category rating (ACR) scale rather than a continuous scale. The ACR scale employs a 5-category discrete quality judgment, which has also been used in recent VQEG studies [23]. As discussed in [39], the subjective rating scales can be increased to more than 5 categories, such as 9 or 11 categories, which are particularly designed for the assessment of special applications, such as low bit-rate video codecs. Also an additional possibility is to use continuous scale rating, which can provide more precise subjective values. In [40], the experimental data has demonstrated that there are no overall statistical differences between different rating scales, which include (a) 5-category discrete scale, which is the one we employed in our subjective testing process; (b) 11-category continuous scale; (c) 5-category continuous scale; (d) 9-category discrete scale. Moreover, for the subjective testing of retargeted image, the resolutions of the images and the introduced shape distortions are very different. The subjective viewers may have difficulties to judge the perceptual quality of the image and provide a precise subjective value. Therefore, in order to make the scoring process simpler to the subjective viewers and the subjective values more distinguishable, the 5-category discrete scale is employed to obtain the subjective opinions to build the image retargeting subjective quality database.

The user interface for the subjective testing is developed by using MATLAB, as shown in Fig. 2. The two images, including the source and the retargeted one, are loaded into the memory before displaying. In order to avoid strong visual contrast, the remaining regions of the display area are gray (the pixel values are set equal to 128). The quality scales are labeled to help the human subjects to do the quality evaluation. The quality scales are labeled as “Bad”, “Poor”, “Fair”, “Good”, and “Excellent”,

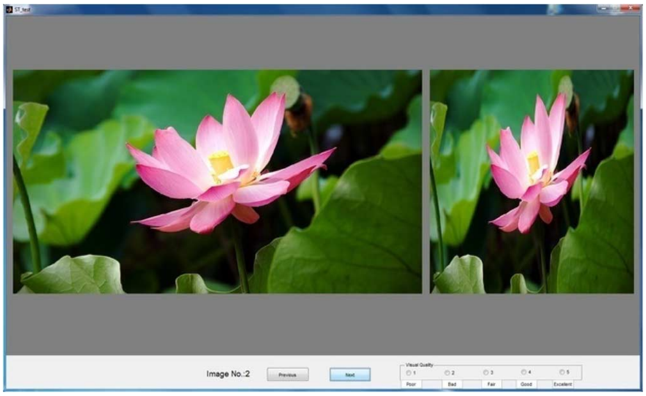


Fig. 2. Screenshot of the subjective study interface displaying the images to the human subject.

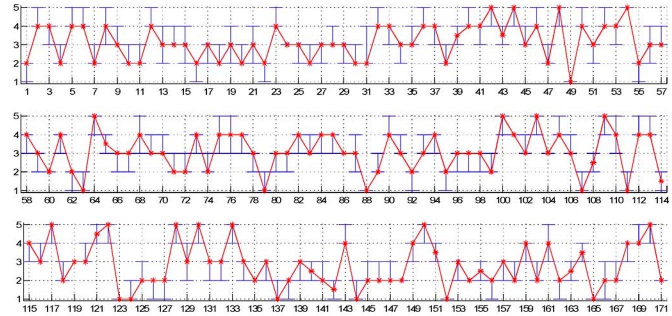


Fig. 3. The subjective scores for each image (the horizontal axes corresponds to the image number, and the vertical axes corresponds to the subjective scores of the viewers. The red star indicates the median value among all the viewers. And the blue error bar indicates the corresponding 25th and 75th percentiles of the subjective scores).

which range from the lowest to the highest perceptual quality index. During the subjective testing, the subjective values are recorded in numerical values. As shown in Fig. 2, the "Bad" corresponds to 1 and the "Excellent" corresponds to 5. Therefore, for the obtained subjective ratings, the larger the value, the better the image perceptual quality is. The human subjects select the appropriate quality index according to their own opinions. After choosing the quality of one image, the subjects can go on evaluating the next image. The subject was allowed to take as much time as needed to evaluate the image quality.

In order to reduce the effect of the viewer fatigue, the 171 retargeted images are divided into 2 sessions. In the first session, the subjective testing is performed in two steps. In the first step, the subjective viewers are asked to provide their personal opinions on the perceptual quality of the retargeted image. After that, in the second step they are further asked to provide their personal opinions on the two distortion levels: (1) the level of shape distortion; (2) the level of content information loss. The detailed process of the second step will be described in following Section IV-B. However, for the second session, the subjective viewers are only asked to take the first step of the subjective testing. Therefore, compared with the first session, the second one will take shorter time for each image. In order to reduce the effect of the viewer fatigue, the number of the images in the first session should be smaller than that of the second one. But the image numbers of the two sessions can be different. Considering this, the authors simply separate the images into two parts. The first session contains 69 images, while the second one contains 102 images. For each session, it will take the viewer about 10–20 minutes to accomplish the subjective testing. The order of the image pairs (the source image and the retargeted image) is randomly arranged, which is distinct for different viewers. 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, which is sufficient for displaying the images in their original resolution.

During the subjective testing, each viewer is briefed by the objective of this subjective study and told how to do the quality evaluation. Before starting the testing, a training session will be presented to all the human subjects. There are in total 7 retargeted images in the training session. They are generated from different source images by different methods in different scales. Also their corresponding perceptual qualities span from "Bad" to "Excellent". The authors of this paper explained to the subjects which quality scale is suggested for each training image. After the training process, each subject is clear on what they should do and how to provide their opinions on the retargeted image quality.

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. For the first session, 30 subjects provide their personal ratings on the perceptual quality of each image, where 15 viewers are experts in image processing and the others are not. And each image in the second session is rated by 34 subjects, where 18 viewers are experts in image processing and the others are not.

III. Data Processing and Analysis for the Database

A. Processing of Subjective Ratings

1) Subjective Agreement: Before we process the subjective ratings to build the database, we need to firstly examine the similarity of choices between participants. Each subject has its own opinion to interpret the image quality. However, for a large proportion of the images in the database, most of the participants should have similar agreements on the perceptual quality. If the subjective results demonstrate diversely among the human subjects, the corresponding image is not suitable for including into the database.

In this paper, we employ the quartiles of the subjective scores for each image to analyze the subject agreement, which is illustrated in Fig. 3. The lower and higher bound of the blue error bar denote the 25th and 75th percentiles of subjective ratings obtained for each image. After sorting the subjective scores, the central 50% of subject ratings lie within the range. The red star indicates the median value of the subjective scores. The detailed information of the image number and the corresponding retargeted image name can be found in the **supplementary** [6]. An outlier coefficient (OC) is introduced to quantify the subjective agreement of the database:

$$OC = \frac{N_{outlier}}{N_{total}}, \tag{1}$$

where N_{total} denotes the total number of the retargeted images in the database, and $N_{outlier}$ denotes the number of the images, which are regarded as the outlier. If the interval between the higher bound and lower bound error bar in Fig. 3 is larger than 1, the image is recognized as the outlier image. The reason is that viewers may have different opinions on the image quality, but they should at least have the similar judgment. For one image, different viewers may interpret the same image as "Good" or "Excellent", which are neighboring values. In most cases, the same image will not be scored with greatly differences, such as "Poor" or "Good". Therefore, if the central 50% subjective ratings are constrained within the interval of 1, we believe that the participants have arrived at an agreement of the retargeted image quality. For the built database, 15 out of 171 are recognized as the outlier images, which implies $OC = 8.77\%$. Therefore, 91.2% of the images in the database have shown the agreement among participants. It is believed that the images in the database will be rated as the similar quality if subjectively tested by the others. Consequently, these images can be included for building the database and further employed for evaluation of the quality metrics.

Furthermore, the subject agreement is checked between every two subjective ratings. Suppose that the subject rating values given by each viewer compose a vector, the normalized cross correlation (*NCC*) and the Euclidean distance (*EUD*) between every two vectors is examined:

$$\begin{aligned} NCC &= \frac{a^t \cdot b}{\|a\|\|b\|}, \\ EUD &= \frac{\|a-b\|_2^2}{k}, \end{aligned} \tag{2}$$

where a and b are the two vectors indicating the subjective ratings of the retargeted images, and $\|\cdot\|_2^2$ defines the norm of the vector, k denotes the dimension of the subjective rating vector. As there are 30 and 34 subjective ratings of each session, $C\binom{30}{2} = 435$ and $C\binom{34}{2} = 561$ NCC and EUD values are obtained for session 1 and session 2, respectively. The average value of each session is employed to examine the subjective agreement of the retargeted image qualities. The average NCC values are 0.9552 and 0.9493 for session 1 and session 2, respectively. The average EUD values are 0.1336 and 0.1046 for session 1 and session 2, respectively. The higher the NCC value and the lower the EUD values, the higher is the correlation between two subjective rating vectors. The NCC values are close to 1, indicating that angle difference between every two subjective rating vectors is very small. Also the EUD values are close to 0.1, demonstrating that the magnitude difference between every two subjective rating vectors is relatively small. Therefore, these two NCC and EUD values demonstrate that the subjects have achieved a great agreement on the perceptual qualities of the retargeted images. These images are thus reasonable to be included to build the database.

2) Screening of the Observers: In the previous section, we have examined the subject agreement on the retargeted image quality. The central 50% subjective ratings of the images have shown high agreement. However, in order to obtain the final MOS and standard deviation value for each image, the subject rejection process is suggested by [22]. 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 [25]:

$$\begin{aligned} \mu_{ik} &= \frac{1}{N_{ik}} \sum_{j=1}^{N_{ik}} S_{ijk}, \\ \sigma_{ik} &= \sqrt{\frac{1}{N_{ik}-1} \sum_{j=1}^{N_{ik}} (S_{ijk} - \mu_{ik})^2}, \\ Z_{ijk} &= \frac{S_{ijk} - \mu_{ik}}{\sigma_{ik}}, \end{aligned} \tag{3}$$

where N_{ik} is the number of the test images seen by the subject i in session k. It is noted that Z-scores are obtained per session, which accounts for any differences in subject preferences for reference images, and different human subjects between sessions.

After converting the obtained subjective ratings into Z-scores, the subject rejection procedure specified in the ITU-R BT 500-11 [22] is then used to discard scores from unreliable subjects. The converting process and subject rejection procedure used should be superior to the VQEG studies [26]–[28]. The ITU-R BT 500-11 first determines whether the scores assigned by a subject are normally distributed by computing the kurtosis β_j of the scores:

$$\beta_j = \frac{m_4}{(m_2)^2} \quad with \quad m_\Delta = \frac{\sum_{j=1}^{N_{ik}} (S_{ijk} - u_{ik})^\Delta}{N_{ik}}, \tag{4}$$

If the kurtosis value β_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 procedure, 1 out of 30 subjects and 3 out of 34 subjects are rejected in session 1 and session 2, respectively.

Algorithm 1. Detailed information of the 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 $(P_{ik} + Q_{ik})/N_{ik} > 0.05$ and $|(P_{ik} - Q_{ik})/(P_{ik} + Q_{ik})| < 0.3$, then **REJECT** the subject i.

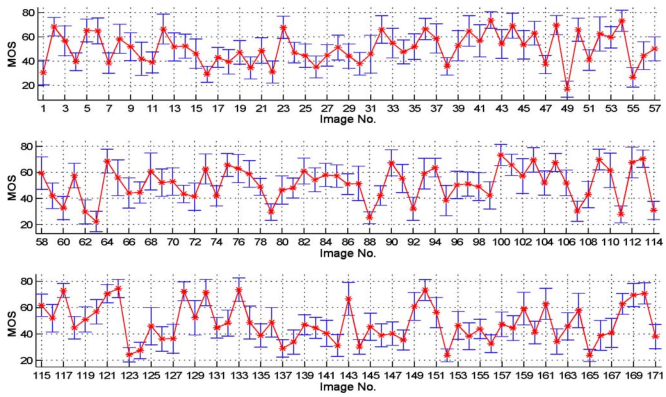


Fig. 4. The obtained MOS value of each retargeted image after processing (the horizontal axes corresponds to the image number, and the vertical axes corresponds to the MOS value. The red star indicates the obtained MOS value. And the blue error bar indicates the standard deviation of the subjective scores).

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 [2], 99% of the scores will lie in the range $[-3, +3]$. Re-scaling is accomplished by linearly mapping the range $[-3, +3]$ to $[0, 100]$ by:

$$\widetilde{Z}_{ijk} = \frac{100(Z_{ijk} + 3)}{6}. \tag{5}$$

Finally, the MOS value of each retargeted image is computed as the mean of the rescaled Z-scores, together with the standard deviation:

$$MOS_{jk} = \frac{1}{M_k} \sum_{i=1}^{M_k} \tilde{Z}_{ijk},$$
$$std_{jk} = \sqrt{\frac{1}{M_k - 1} \sum_{i=1}^{M_k} (\tilde{Z}_{ijk} - MOS_{jk})^2}, \tag{6}$$

where M_k is the number of remaining subjects of session k after the subject rejection. The MOS value together with the standard deviation is recorded for each retargeted image, which is recognized as the ground truth representing the retargeted image perceptual quality. They can be further analyzed and used for evaluating the performances of the quality metrics. The final subjective scores after conversion, with the standard deviation indicating the error bar, are illustrated in Fig. 4.

As we mentioned above, the perceptual qualities of the retargeted images in the database should span the entire range of visual quality and exhibit good perceptual quality separation [1], [2]. The histogram of the MOS values is shown in Fig. 5. It can be observed that the perceptual qualities of the images range from low to high values. Also it demonstrates that the subjective study samples a range of perceptual quality in an approximately uniform fashion. The image perceptual qualities exhibit a good separation.

B. Analysis and Discussion of the Subjective Ratings

After the processing of the subjective ratings, the image retargeting database is built, which comprises the retargeted images and their corresponding MOS values. The database is analyzed

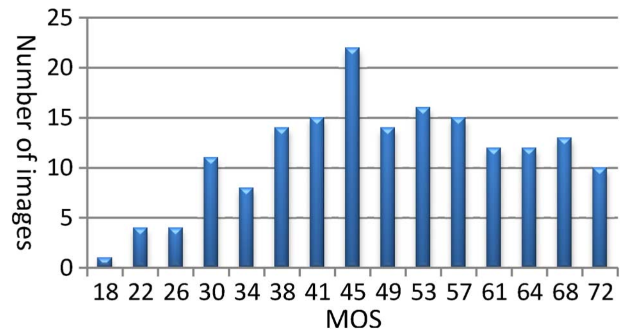


Fig. 5. Histogram of the MOS values in 15 equally spaced bins between the minimum and maximum MOS values of the image retargeting database.

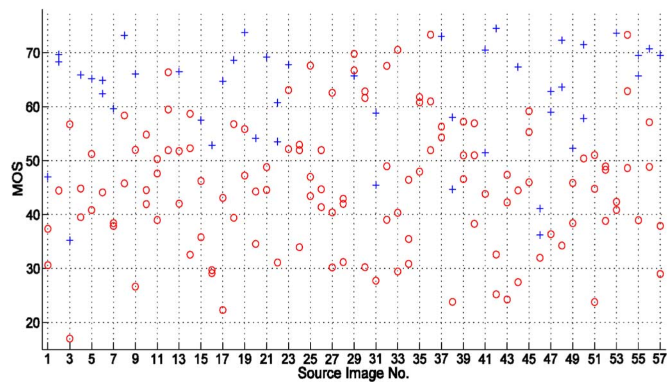


Fig. 6. The obtained MOS value versus the source image from the scale perspective. (The blue cross indicates the retargeted image in 75% scale; the red circle indicates the retargeted image in 50% scale.)

from three aspects, specifically the retargeting scale, the retargeting method, and the source image content.

1) Retargeting Scale: The MOS values of the retargeted images in two different scales are illustrated in Fig. 6. The detailed information of the image number and the corresponding source image can be found in the **supplementary** [6]. Generally, it can be observed that the retargeted images in 75% scale (with average MOS value as 61.79) exhibit higher perceptual quality than the retargeted images in 50% scale (with average MOS value as 45.66). There are two exceptions, which are generated from the source images ‘kodim04’ and ‘bicycle1’. For the ‘kodim04’ containing the human face, the CROP method in 50% scale can preserve the shape information but sacrifice some content information, while the SCSC method in 75% will distort the human face. For ‘bicycle1’ with clear foreground object, the SEAM and SHIF methods in 50% scale will accurately preserve the shape and the content information, while the SCAL method in 75% scale will introduce some shape distortion. Therefore, the two images in 50% scale present better quality than the images in 75% scale. The reason is that the subjects prefer information loss rather than shape deformation.

Furthermore, it can be observed that the MOS values of the retargeted images in 75% scale are mostly larger than 50, except ‘kodim01’, ‘kodim04’, ‘buddha’, ‘face’, and ‘kodim15’. Referring to the attribute information of the source images, these images only contain either ‘face and people’ or ‘geometric structure’ attributes. It is known that human eyes are very sensitive to these attributes, which will greatly influence the perceptual quality of the retargeted image. For ‘buddha’ and ‘face’

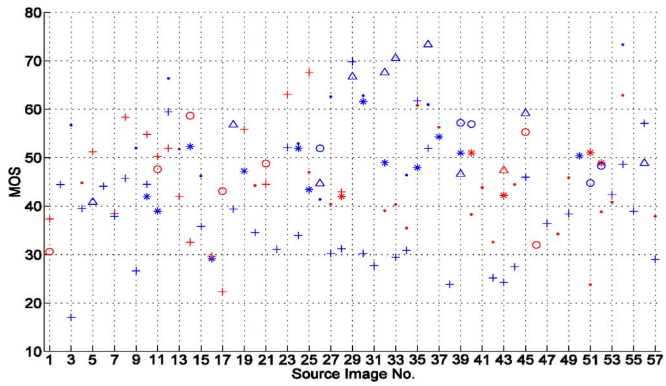


Fig. 7. The obtained MOS value versus the source image from the retargeting method perspective (in 50% scale). The blue dot is the CROP method; the blue star is the SCAL method, the blue cross is the SEAM method [10]–[12]; the blue triangle is the SHIF method [15]; the blue circle denotes the MULT algorithm [13]; the red dot denotes the WARP algorithm [8]; the red star denotes the ENER algorithm [16]; the red cross denotes the SCST [14]; the red triangle denotes the STVI method [9]; the red circle denotes the SCSC method [17].

images, other retargeting methods can generate higher quality images. Therefore, retargeting methods should be carefully selected for these images, which should not distort the shape information. For the retargeted images in 50% scale, the MOS values vary greatly. Some source images, such as "bicycle1" and 'eagle', generate retargeted images with very good quality. Also some source images, such as 'volleyball', generate retargeted images with very poor quality. Therefore the source image content will influence the perceptual quality of the retargeted images. Moreover, the retargeted images from the same source image also possess perceptual qualities with great differences, such as 'blueman'. It means that the retargeting method will also affect the image perceptual qualities. In the following sub-sections, the perceptual qualities of the retargeted images in 50% scale are analyzed from the two aspects: retargeting method, and source image content.

2) Retargeting Methods: As we discussed in the previous subsection, most of the algorithms produce the retargeted images in 75% scale with acceptable perceptual quality. In order to analyze the influence of the retargeting method, only the retargeted images in 50% scale are considered. The MOS values of the images by different retargeting methods are illustrated in Fig. 7. As we mentioned before, the basic tools for retargeting are CROP, SCAL, WARP and SEAM. We firstly analyze these basic tools and then discuss the performances of the other methods.

The images generated by SEAM method [10]–[12](denoted by the blue cross in Fig. 7) are always of the worst perceptual quality. The reason is that the SEAM method tries to remove the seams in the regions with low gradient magnitudes. For some images, such as "kodim04" and "kodim15", some regions of the salient object appear to be very smooth, which will be discarded during the retargeting process. Therefore, some annoying shape distortion will be introduced. And as revealed by [18], the human subjects prefer sacrificing some image information rather than having deformation. The SEAM method does not consider any approaches to preserve the object shape. Therefore, it exhibits the worst perceptual quality, especially for images containing salient objects.

The CROP method (denoted by the blue dot) can only retarget some images with good perceptual quality. As it only keeps part information of the source image, its performance depends on the source image content. For some images with a small region containing the salient content, the CROP method can retarget a good quality image, such as "surfer". For some images, such as "perissa_santorini", where all regions contain meaningful information, the CROP method retargets images with bad quality. In [18], the CROP method is suggested as the most reliable and simplest method to retarget images.

The WARP algorithm [8] (denoted by the red dot) tries to squeeze the source image to a target size by optimizing a warping function. The shape of the object cannot be preserved. Therefore, the retargeted images are of bad perceptual quality. In most cases, it only outperforms SEAM method, while is inferior to other methods. The SCAL method (denoted by the blue star) retargets images with medium perceptual quality. It will introduce some shape deformation into the retargeted image, but not as severe as the SEAM and WARP method. Therefore, the SCAL method always outperforms SEAM and WARP, but worse than the other methods under study.

The other methods try to combine these basic tools together to produce an optimal retargeted image. Some methods, such as SCST [14] and SHIF [15], have considered using the saliency map to guide the retargeting. The shape information of the objects in the salient regions is preserved to avoid introducing unpleasant deformation. Therefore, these methods can obtain better performances. As shown in Fig. 7, in most cases the SHIF algorithm (denoted by the blue triangle) and SCST (denoted by the red cross) can retarget the test images with better perceptual quality, compared with the other methods.

3) Source Image Contents: As mentioned above, the source images can be categorized by the containing attributes, which are 'face and people', 'clear foreground object', 'natural scenery', and 'geometric structure'. The 'clear foreground object' attribute is defined as that the salient object occupies the image regions smaller than 50% of the source image. If the salient object is preserved, the perceptual quality of the retargeted image (in 50% and 75% ratios) will not be very bad, as the crop margin (how much can be cropped without losing the object/regions of interest) is larger than 50%. The 'natural scenery' attribute means that a large proportion of the image contains the texture or smooth information. These images contain information with symmetric or similar patterns. Therefore, cropping or scaling some part of the image will not introduce significant degradations in perceptual quality. The crop margin of these images is large. Therefore, the retargeted images in 50% and 75% ratios are of good perceptual quality. These two attributes are regarded as non-salient. The 'geometric structure' attribute denotes that there are evident edges or lines in the source image, and 'face and people' attribute means that the faces or persons occupy most regions of one source image. The subjective viewers will be very sensitive to the evident edges, shapes, and faces. The distortion introduced by the retargeting method will severely affect the judgment of the subjective viewer. For some images containing 'face and people' attribute, such as the image 'face', 'kodim15', 'kodim04', and 'buddha', the entire image is a human face, which is of great saliency. If we crop some part of the image, some important content is discarded, which will result in very bad perceptual quality. In this respect, the crop margin of these images is very small (nearly 0). Therefore, if we retarget these images with any kind of methods, the perceptual quality will not be good. These two attributes are regarded as the salient attributes.

Each source image may contain more than one attribute. However, one attribute dominates each source image, while other attributes are not so significant. The detailed attribute information of each source image is illustrated in the **supplementary** [6]. The attributes are sorted according to their significances. According to the attribute saliency, the source images are divided into two classes. Note that we only utilized the most significant attribute to classify the source images. After the separation, we got 30 images with salient attributes and the other 27 images with non-salient attributes. The MOS values versus the source images of different attributes are illustrated in Fig. 8. In this subsection, as we only care about the influence of the image content on the perceptual quality, the retargeting methods are not considered. The retargeted image with the worst perceptual quality is utilized for comparison. They are all in the 50% scale, which ensures a fair comparison. We calculate the mean MOS values of the retargeted images in the two classes. The mean MOS value of the images with non-salient attributes is 45.55, which is higher than the average MOS value of the database. The images with non-salient attributes contain some texture and smooth information, such as ‘kodim13’ and ‘fishing’, which ensures a large crop margin value. Therefore, the shape deformation will not be easily detected. And the region discarded during the retargeting process mostly contains information with symmetric or similar patterns, or unimportant background information. Therefore, the perceptual quality will not be significantly influenced. However, the mean MOS value of the images with salient attributes is 31.1292, which is lower than that of non-salient attribute image. As most regions of the source image contain salient or meaningful information, the crop margin of such image is very small (nearly 0). And the contents and shapes of the objects, faces, or humans are critical for judging the perceptual quality. Retargeting these images into 50% scale will significantly distort the shapes or discard important content information. Therefore, the perceptual quality will be very unpleasant.

For the source images with salient attributes, Fig. 8 shows that the CROP method always retargets image with the highest MOS values, such as ‘kodim04’ and ‘sanfrancisco’. Although only a few source images employ the CROP method to retarget image, it can be deduced that the CROP will retarget the other images with highest quality as claimed by [18]. The reason is that the shape deformation is much more annoying to the human viewers, compared with the information discarded. For the salient attributes, such as ‘face and human’ and ‘geometric structures’, the shape distortion can be easily detected and rated badly by the subjective viewers. Therefore, for the images containing salient attributes, the CROP process is recommended, not only because of its simplicity but also for its best performance. This is also applied to retarget image into 75% scale. For the images with salient attributes, such as ‘kodim01’, ‘kodim04’, ‘buddha’, ‘face’, and ‘kodim15’, the other methods other than CROP can introduce deformation to the object shape. That is the reason why the retargeted images in 75% are of low MOS values. For the images with non-salient attributes, most of the retargeted images are of good qualities, with MOS values larger than 45. However, there are several exceptions, such as ‘perissa_santorini’, ‘butterfly’, and ‘fishing’. It can be observed from Fig. 8 that SCAL and SEAM also generate images with bad quality. The other methods, such as SCST, will preserve much more information, while the introduced shape deformation can be hardly detected by the human viewers.

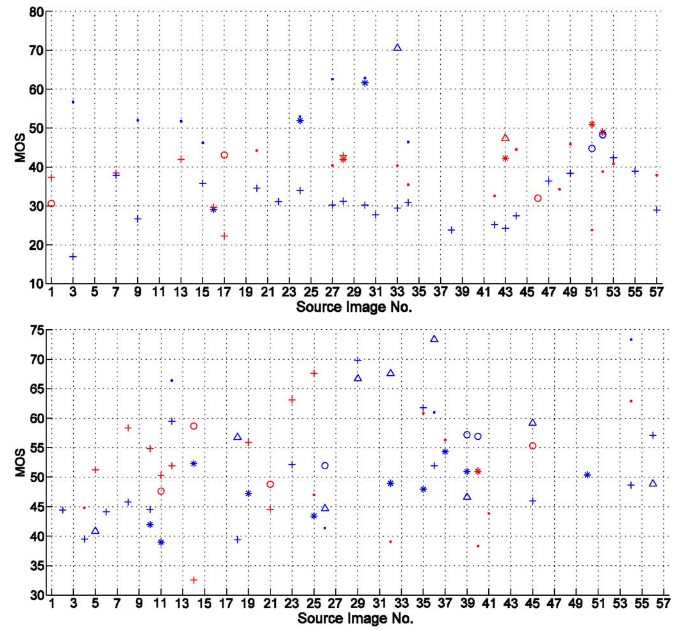


Fig. 8. The obtained MOS value versus the source image. Top: source images with salient attributes; bottom: source images with non-salient attributes.

Considering the above analysis, different retargeting methods are recommended for different images, as shown in Fig. 9. For images with salient attribute, the CROP methods are suggested for its effectiveness and low complexity. For images with non-salient attribute, if retargeting them into 75% scale, all the retargeting method can generate acceptable results, because the shape distortion can hardly perceived and the loss of the image content is negligible. To retarget the images with non-salient attribute into 50% scale, we recommend the SCSC and SHIF method. They have considered the saliency map, which can help to preserve the object in the image.

IV. Objective Quality Metric for Retargeted Images

A. *Quality Metric Performances on the Built Image Retargeting Database*

Image retargeting quality metric has been recently researched [29]–[35], in order to not only evaluate the retargeted image quality automatically and reliably instead of the subjective testing, but also 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 [13], and the quality metric in [34], which are not made publicly. In this section, we only tested the metrics which are publicly

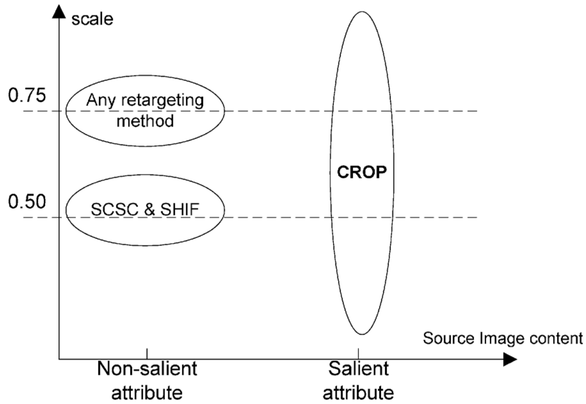


Fig. 9. Recommended retargeting methods by considering both the retargeting scale and source image content.

available and suggested in [18], specifically the earth mover's distance (EMD) [29], [30], the bidirectional similarity (BDS) [31], [32], edge histogram (EH) [35], and SIFT-flow [33]. The information about the metrics is detailed in the following.

- **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 $P = \{(p_1, w_{p_1}), \ldots, (p_m, w_{p_m})\}$ be the first signature with m clusters; $Q = \{(q_1, w_{q_1}), \ldots, (q_n, w_{q_n})\}$ 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:

$$WORK(P,Q,F) = \Sigma_i^m \Sigma_j^n d_{ij} f_{ij}. \quad (7)$$

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

$$EMD(P,Q) = \frac{\Sigma_i^m \Sigma_j^n d_{ij} f_{ij}}{\Sigma_i^m \Sigma_j^n f_{ij}}. \quad (8)$$

- **BDS**: Two signals S (original image) and T (retargeted image) are considered to be 'visually similar' is as many as possible patches of S (at multiple scales) are contained in T, and vice versa. The dissimilarity can be formulated as:

$$(S,T) = \overbrace{\frac{1}{N_S} \Sigma_{P \subset S} \min_{Q \subset T} D(P,Q)}^{d_{complete}(S,T)} + \overbrace{\frac{1}{N_T} \Sigma_{Q \subset T} \min_{P \subset S} D(P,Q)}^{d_{cohere}(S,T)}. \quad (9)$$

where P and Q denote patches in S and T, respectively. And let N_S and N_T denote the number of patches in S and T. For each patch $Q \subset T$ we search for the most similar patch $P \subset S$, and measure their distance $D(P,Q)$, and vice-versa. The patches are taken around every pixel at multiple scales, resulting in significant patch overlap. $D(P,Q)$ can be any distance measurements between two patches, such as sum squared distances (SSD) or SSIM [36]. The two terms have important commentary roles. The first term, $d_{complete}(S,T)$ measures the deviation of the target T from 'completeness' w.r.t. S. Namely, it measures if all patches of S have been preserved in T. The second term $d_{cohere}(S,T)$ measures if there are any 'newborn' patches in T which have not originated from S. Therefore, the $d_{complete}(S,T)$ tries to represent the input image well (be complete), and the $d_{cohere}(S,T)$ makes sure the retargeted image to be visually pleasing (coherent). The dissimilarity measurement is minimized in order to generate a retargeted image [31], [32].

- **EH** 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.
- SIFT-flow descriptors characterize view-invariant and brightness-independent image structures. Matching SIFT descriptors allows establishing meaningful correspondences across image with significantly different image content. Furthermore, the pixel displacement (indicating by the SIFT correspondence matching) should be spatial coherent, which means that close-by pixels should have similar displacement. The cost function is defined as:

$$\begin{aligned} E(w) = {} & \Sigma_p \left\| s_1(p) - s_2(p+w) \right\|_1 + \frac{1}{\sigma^2} \Sigma_p \left(u^2(p) + v^2(p) \right) \\ & + \Sigma_{(p,q)\in\epsilon} \left(min\left(\alpha \left| u(p) - u(q) \right|, d \right) \right. \\ & \left. + \min\left(\alpha \left| v(p) - v(q) \right|, d \right) \right). \end{aligned} \quad (10)$$

where $w(p) = (u(p), v(p))$ is the displacement vector at pixel location $p = (x, y)$, $s_i(p)$ is the SIFT descriptor extracted at location p in image i and ϵ is the spatial neighborhood of a pixel. SIFT flow employs the SIFT for feature matching. And the local smoothness is preserved by the vector difference constraint.

The algorithms are provided by the respective 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 [28] and that in [37], we follow the procedure to evaluate

TABLE I
PERFORMANCES OF DIFFERENT METRICS ON THE IMAGE RETARGETING DATABASE

	EH	EMD	BSD	SIFT-flow	Fusion (EH, EMD and SIFT-flow)	Fusion (EH, EMD, BSD, SIFT-flow)
LCC	0.3422	0.2760	0.2896	0.3141	0.4361	0.5217
SROCC	0.3288	0.2904	0.2887	0.2899	0.4203	0.4514
RMSE	12.686	12.977	12.922	12.817	12.149	11.484
OR	0.2047	0.1696	0.2164	0.1462	0.1462	0.1287

the performance of each metric. 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:

$$V_j = \beta_1 \times \left(0.5 - \frac{1}{1 + e^{\beta_2 \times (x_j - \beta_3)}}\right) + \beta_4 \times x_j + \beta_5. \quad (11)$$

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 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 metrics are illustrated in Table I. 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 by 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 correspondence 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 in 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.

B. Subjective Analysis of the Shape Distortion and Content Information Loss

As shown in the previous subsection, accounting for the object shape or content information loss alone cannot effectively evaluate the retargeted image quality. In order to investigate how the object shape and content information loss influence the perceptual quality, a subjective testing is designed.

During the first session of our subjective testing, after the human subjects providing their personal opinions on the retargeted image quality, they are also asked to provide their personal opinions on the two distortion levels: (1) the level of shape distortion; (2) the level of content information loss. The shape distortion depicts the distortion, such as face deforming, object squeezing, object boundary discontinuity, and so on. The content information loss depicts that part information of the object or content is missing in the retargeted image, compared with the source image. Both of the two distortion levels are recorded in 5-scale, same as introduced in Section III. After the subjective testing, not only the visual quality of the retargeted image is evaluated, but also the distortion levels of the two factors (shape distortion and content information loss) that may affect the visual quality are recorded.

Same as in Section III, the level scores of the shape distortion and content information loss are processed independently by following the Z-score conversion, the subject rejection, and Z-score inverse conversion. After these procedures, the level values are re-scaled in the range [0, 100], same as the MOS values. LCC and SROCC between the level scores and the MOS values are utilized to evaluate their correlation, which is shown in Table II. It can be observed that the level of shape distortion correlates much more closely with MOS values than the content information loss. It means that the subject viewers are more sensitive to the shape distortions introduced in the retargeted images. In most cases, the human subjects tend to sacrifice the information loss rather than the shape distortion for recognizing a good quality image. For the information loss, although it correlates badly with the MOS values, it still affects the visual quality of the retargeted image.

From Table II, it can be observed that the shape distortion correlates closely with the final perceptual quality of the retargeted image. However, the employed three metrics, EH, EMD, and SIFT-flow describing the shape distortion do not prove to be efficiency, as shown in Table I. The reason may attribute to that none of them are able to accurately capture the shape distortion. Therefore, a fusion strategy is tested by combining the three metrics together through average process. The performance is also illustrated in Table I. Compared with the three metrics, the fusion one performs better. It means that the current descriptor

TABLE II
RELATIONSHIP BETWEEN MOS VALUES AND THE LEVELS OF SHAPE DISTORTION AND INFORMATION LOSS

	LCC	SROCC
MOS vs. Shape Distortion	0.8243	0.8371
MOS vs. Content Information Loss	0.3264	0.4680
MOS vs. Fusion of Shape Distortion and Content Information Loss	0.9218	0.9267

for capturing the shape distortion is not accurate enough. Furthermore, a fusion strategy by summing the shape distortion and content information loss together is tested. As shown in Table I, the fusion result correlates more closely with the MOS value. The observation provides us some hints for designing the quality metric from the perspective of shape distortion and content information loss. The descriptors of shape distortion and content information loss should be combined together for evaluating retargeted image quality. For the current available metrics, EH, EMD and SIFT-flow tries 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 considered to build a quality metric, the performance of which is illustrated in Table I. It can be observed that a better performance is obtained, which means that considering the shape distortion and content information loss together can help to improve the performances.

C. Discussion

As demonstrated in previous subsections, the performances of the objective quality metrics for retargeted images are still not good enough. The statistical correlations between the subjective MOS values and the metric outputs are not close. Even fusing EH, EMD, BSD, and SIFT-flow together, the LCC and SROCC values are smaller than 0.6, which indicates a bad performance of the objective metric. In this subsection, we will discuss and try to figure out how to design an effective objective quality metric for evaluating the perceptual quality of the retargeted image. The source image content, retargeting scale, the shape distortion and content information loss measurement, and the HVS properties are the candidate factors, which are believed to benefit the objective metric performance.

- Shape distortion description. As illustrated Table II, the shape distortion is closely related to the perceptual quality of the retargeted image. Therefore, the recently developed metrics, such as EH, EMD, and SIFT-flow, try to capture the object shape of the image and measure the corresponding differences between the source and retargeted image. However, the performances are not good enough, where the LCC and SROCC values are only about 0.35 as shown in Table I. Even combining these metrics together, a better performance can be ensured. But the result is still unpleasant. Therefore, in order to accurately depict the perceptual quality of the retargeted image, the shape distortions that introduced by retargeting process need to be captured more precisely. Recently, A. D'Angelo [41], [42] proposed a full-reference quality metric to evaluate the geometrical distortions of the images. The approaches are based on that the HVS is sensitive to the image structures, such as edges and bars, which are indentified by employing the Gabor filter. By considering this descriptor for evaluating the geometrical distortion, the shape distortion introduced during the retargeting process is believed to be more accurately described. Therefore, it can help to improve the performance of the objective quality metric.
- Fusion of the shape distortion and content information loss. As illustrated in Table II, the content information loss alone is not closely related to the final perceptual quality of the retargeted image. But fusion the shape distortion and content information loss together can improve the performance, which has also been illustrated in Table I. The combinations of the four objective quality metrics can beat the other metrics. Therefore, if we develop accurate metrics to capture the shape distortion and content information loss, how to fuse them together needs to be further considered. The fusion strategy of the two factors should consider their corresponding contributions to the final retargeted image quality.
- Source image quality and retargeting scale. The source images that we employed to build our database are of different resolutions and different qualities, which may affect the subjective viewers' judgment of the retargeted image perceptual quality. Moreover, the retargeting scale will also affect the retargeted image quality. Given one source image, the larger the retargeting ratio, the better the perceptual quality of the retargeted image is. Therefore, the final perceptual quality index of the retargeted image needs to account for the quality of the source image as well as the retargeting scale.
- Image content. As discussed in previous sections, the image content correlates closely to the crop margin of the source image (how much can be cropped without losing the object/regions of interest). If the source image contains the 'clear foreground object' or 'natural scenery' attribute, the crop margin will be very large. Therefore, retargeting the source image into 75% and 50% ratios will not significantly affect the perceptual quality. Otherwise, if the source image contains the 'face and people' or 'geometric structure' attribute, the crop margin will be very small. Then any retargeting methods will severely degrade the perceptual quality. In this respect, the image content and the crop margin of each source image need to be included to depict the perceptual quality of the retargeted image.
- HVS saliency. Additionally, the HVS demonstrates different conspicuities over different regions of the image. The shape distortions and content information loss in the saliency regions are more sensitively perceived by the subjective viewers than those in the non-saliency regions. That is also the reason why several retargeting methods consider the saliency or visual attention map during the retargeting process, such as WARP [8], SCST [14], and STVI [9]. The

subjective viewers' assessment on the quality of the retargeted image is prejudiced during the subjective testing process. Therefore, the effect of the HVS saliency needs to be considered to model the subjective viewer's behavior, which will lead to a more effective quality metric for retargeted images. The simplest way for incorporating the HVS saliency is to weight the corresponding shape distortion and content information loss by the saliency map detected from the source image, which has been demonstrated to be effective to evaluate the perceptual quality of the traditional distorted image [43].

V. Conclusion

An image retargeting database is built through the subjective study in this paper. Based on the subjective ratings of the human viewers, the database is analyzed from the perspectives of retargeting scale, retargeting method, and source image content. Also the publicly available quality metrics for the retargeted images are evaluated on the built database. By fusing the metrics together, which independently depict shape distortion and content information loss, the performance can be improved.

Acknowledgment

The authors would like to thank the researchers who provide the source images, the retargeting methods, and the quality metrics to build and analyze the image retargeting database. In particular, we thank M. Rubinstein *et al.* for providing the Retargetme benchmark, Y. Wang *et al.* for providing the SCST retargeting method, Y. Pritch *et al.* for providing the SHIF retargeting method, W. Dong *et al.* for providing the SCSC retargeting results, and S. Avidan for providing the SEAM retargeting method. The authors are grateful to O. Pele *et al.* for providing the EMD metric, C. Barnes *et al.* for providing the BSD metric, and C. Liu *et al.* for providing the SIFT-flow metric.

References

[1] A. K. Moorthy, K. Seshadrinathan, R. Soundararajan, and A. C. Bovik, "Wireless video quality assessment: A study of subjective scores and objective algorithms," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 4, pp. 587–598, Apr. 2010.

[2] K. Soundararajan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1427–1441, Jun. 2010.

[3] K. Soundararajan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "A subjective study to evaluate video quality assessment algorithms," in *SPIE Proc. Human Vision and Electronic Imaging*, Jan. 2010.

[4] P. L. Callet and F. Autrusseau, Subjective Quality Assessment IRCCyN/IVC Database, [Online]. Available: http://www2.irccyn.ec-nantes.fr/ivcdb/

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

[6] L. Ma, W. Lin, C. Deng, and K. N. Ngan, Image Retargeting Subjective Database, [Online]. Available: http://ivp.ee.cuhk.edu.hk/projects/demo/retargeting/index.html

[7] A. Shamir and O. Sorkine, "Visual media retargeting," in *ACM SIGGRAPH Asia Courses*, 2009.

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

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

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

[11] M. Rubinstein, A. Shamir, and A. Avidan, "Improved seam carving for video retargeting," in *Proc. SIGGRAPH*, 2008.

[12] A. Shamir and S. Avidan, "Seam-carving for media retargeting," *Commun. ACM*, vol. 52, no. 1, pp. 77–85, Jan. 2009.

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

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

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

[16] Z. Karni, D. Freedman, and C. Gotsman, "Energy-based image deformation," in *Proc. Symp. Geometry Process.*, 2009.

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

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

[19] 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.

[20] VQEG, Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, 2000 [Online]. Available: http://www.vqeg.org

[21] L. Itti, C. Koch, and E. Neibur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

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

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

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

[25] A. M. van Dijk, J. B. Martens, and A. B. Watson, "Quality assessment of coded images using numerical category scaling," in *Proc. SPIE Adv. Image Video Commun. Storage Technol.*, 1995.

[26] Final report from the Video Quality Experts Group on the validation of objective quality metrics for video quality assessment 2000, [Online]. Available: http://www.its.bldrdoc.gov/vqeg/projects/frtv_phaseI/

[27] The Video Quality Experts Group Final,, Final VQEG Report on the validation of objective models of video quality assessment, [Online]. Available: http://www.its.bldrdoc.gov/vqeg/projects/frtv_phaseII/

[28] Final report from the Video Quality Experts Group on the validation of objective models of multimedia quality assessment, Phase 1 [Online]. Available: ftp://vqeg.its.bldrdoc.gov/Documents/Projects/multimedia/MM_Final_Report/

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

[30] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *Int. J. Comput. Vis.*, vol. 40, no. 2, pp. 99–121, 2000.

[31] D. Simakov Yaron Caspi, E. Shechtman, and M. Irani, "Summarizing visual data using bidirectional similarity," in *Porc. CVPR*, 2008.

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

[33] C. Liu, J. Yuen, A. Torralba, J. Sivic, and W. T. Freeman, "SIFT flow: Dense correspondence across different scenes," in *Proc. ECCV*, 2008.

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

[35] 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.

[36] 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.

[37] 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.

[38] M. H. Pinson and S. Wolf, "Comparing subjective video quality testing methodologies," *Proc. SPIE*, vol. 5150, no. 3, pp. 573–582, 2003.

[39] ITU-T Recommendation P. 910, "Subjective video quality assessment methods for multimedia applications," [Online]. Available: http://www.videoclarity.com/PDF/

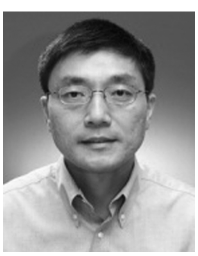
[40] Q. Huynh-Thu, N. N. Garcia, F. Speranza, P. Corriveau, and A. Raake, "Study of rating scales for subjective quality assessment of high-definition video," *IEEE Trans. Broadcasting*, vol. 57, no. 1, pp. 1–14, Mar. 2011.

[41] A. D'Angelo, G. Menegaz, and M. Barni, "Perceptual quality evaluation of geometric distortions in images," *Proc. SPEI, Human Vis. Electron. Imaging*, vol. 6492, 2007.
[42] A. D'Angelo, Z. Zhao, and M. Barni, "A full-reference quality metric for geometrically distorted images," *IEEE Trans. Image Process.*, vol. 19, no. 4, pp. 867–881, Apr. 2010.
[43] 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.
[44] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, vol. 30, pp. 81–93, 1938.
[45] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, LIVE Image Quality Assessment Database Release 2, [Online]. Available: http://live.ece.utexas.edu/research/quality/

Lin Ma (S'09) received the B.E. and M.E. degrees from Harbin Institute of Technology, Harbin, China, in 2006 and 2008, respectively, both in computer science. He is now pursuing the Ph.D. degree in the Department of Electronic Engineering at the Chinese University of Hong Kong (CUHK).

He was a Research Intern with Microsoft Research Asia 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, from July 2011 to September 2011. His research interests include image/video quality assessment, super-resolution, restoration, and compression.

Mr. Ma received the Best Paper Award in the Pacific-Rim Conference on Multimedia (PCM) 2008. He was awarded the Microsoft Research Asia fellowship in 2011.



Weisi Lin (M'92–SM'98) received his B.Sc. and M.Sc. from Zhongshan University, Guangzhou, China, and the Ph.D. from King's College, London University, London, U.K. He was the Lab Head, Visual Processing, and the Acting Department Manager, Media Processing, for the Institute for Infocomm Research, Singapore. Currently, he is an Associate Professor in the School of Computer Engineering, Nanyang Technological University, Singapore. His areas of expertise include image processing, perceptual modeling, video compression, multimedia communication and computer vision. He has published 200+ refereed papers in international journals and conferences.

He is on the editorial boards of IEEE TRANSACTIONS ON MULTIMEDIA, IEEE SIGNAL PROCESSING LETTERS and *Journal of Visual Communication and Image Representation*. He served as the Lead Guest Editor for a special issue on perceptual signal processing, IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, 2012. He is a Chartered Engineer (U.K.), a fellow of Institution of Engineering Technology, and an Honorary Fellow, Singapore Institute of Engineering Technologists. He chairs the IEEE MMTC Special Interest Group on Quality of Experience. He has been elected as a Distinguished Lecturer of APSIPA (2012/3). He is the Lead Technical Program Chair for Pacific-Rim Conf. On Multimedia (PCM) 2012, and a Technical Program Chair *IEEE International Conference on Multimedia and Expo* (ICME) 2013.

Chenwei Deng (M'09) received the Ph.D. degree in signal and information processing from Beijing Institute of Technology, Beijing, China, in 2009. Since May 2009, he has been a Postdoctoral Research Fellow at Nanyang Technological University, Singapore. His research interests include image/video coding, processing, and computer vision.

King Ngi Ngan (M'79–SM'91–F'00) received the Ph.D. degree in Electrical Engineering from the Loughborough University in U.K. He is currently a chair professor at the Department of Electronic Engineering, Chinese University of Hong Kong. He was previously a full professor at the Nanyang Technological University, Singapore, and the University of Western Australia, Australia. He holds honorary and visiting professorships of numerous universities in China, Australia and South East Asia.

Prof. Ngan served as associate editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS for Video Technology, Journal on Visual Communications and Image Representation, EURASIP Journal of Signal Processing: Image Communication, and Journal of Applied Signal Processing. He chaired a number of prestigious international conferences on video signal processing and communications, and served on the advisory and technical committees of numerous professional organizations. Prof. Ngan co-chaired the IEEE International Conference on Image Processing (ICIP) held in Hong Kong in September 2010. He has published extensively including 3 authored books, 6 edited volumes, over 300 refereed technical papers, and edited 9 special issues in journals. In addition, he holds 10 patents in the areas of image/video coding and communications.

Prof. Ngan is a Fellow of IEEE (U.S.A.), IET (U.K.), and IEAust (Australia), and an IEEE Distinguished Lecturer in 2006–2007.