Multiple-Level Feature-Based Measure for Retargeted Image Quality

Yabin Zhang, Weisi Lin, *Fellow, IEEE*, Qiaohong Li, Wentao Cheng, and Xinfeng Zhang, *Member, IEEE*

***Abstract*— Objective image retargeting quality assessment aims to use computational models to predict the retargeted image quality consistent with subjective perception. In this paper, we propose a multiple-level feature (MLF)-based quality measure to predict the perceptual quality of retargeted images. We first provide an in-depth analysis on the low-level aspect ratio similarity feature, and then propose a mid-level edge group similarity feature, to better address the shape/structure related distortion. Furthermore, a high-level face block similarity feature is designed to deal with sensitive region deformation. The multiple-level features are complementary as they quantify different aspects of quality degradation in the retargeted image, and the MLF measure learned by regression is used to predict the perceptual quality of retargeted images. Extensive experimental results performed on two public benchmark databases demonstrate that the proposed MLF measure achieves higher quality prediction accuracy than the existing relevant state-of-the-art quality measures.**

***Index Terms*— Retargeted image quality, edge group similarity, multiple-level feature.**

I. Introduction

WITH the increasing diversity and versatility of display devices, image retargeting which adapts the image to different resolutions and aspect ratios becomes ever more important to improve the viewing experience of users. The conventional methods like manual cropping (CR) and uniform scaling (SCL) are not satisfactory since they do not consider image content and suffer from the information loss and visual distortion. In the last decade many content-aware image retargeting operators [1]- [7] have been proposed and show promising performance by preserving the important content with minimal distortion. However, there is still no single retargeting operator that can work well for every image, thus it is meaningful to develop an effective measure for the image retargeting quality assessment (IRQA) to further advance retargeting techniques.

The most reliable way to estimate the visual quality of the retargeted image is the subjective test since the human visual system (HVS) is usually the ultimate receiver in retargeting applications. While subjective evaluation is generally cumbersome, non-automatic and expensive, objective IRQA methods [8]- [11] are more favorable to predict the retargeted image quality. The IRQA is actually a semantic high-level task, where the visual distortions due to artificial modifications are arbitrary such as the twisted line/structure, deformed face and broken symmetry pattern. However, existing objective methods still lack effective IRQA features to depict these characteristics and measure the corresponding quality degradation evoked by image retargeting.

In general, information loss and visual distortion are two major factors in quality degradation of the retargeted image. Previous works [10]- [14] measure the retargeted image quality in terms of low-level features and have achieved promising prediction accuracy. However, they still suffer from several problems. Firstly, the low-level features usually pool the local quality scores with visual importance map, thus the visual distortions in unimportant regions are likely to be overlooked. Secondly, image retargeting quality evaluation is a high-level task where the semantic high-level understanding of the image is important. The low-level features are not sufficient to cover all the quality degradation factors.

In this paper, we propose a multiple-level feature (MLF) based measure to predict the perceptual quality of the retargeted image and the major contributions are summarized as follows. Firstly, to address the limitations of the existing low-level features and visual importance dependence, we propose a MLF based framework to model various kinds of quality degradation factors in the retargeted image. Secondly, we give an in-depth analysis on the low-level aspect ratio similarity feature, and propose one mid-level edge group similarity feature and one high-level face block similarity feature to measure the shape/structure related distortion and sensitive facial region deformation, respectively. Thirdly, we conduct a comprehensive feature analysis, and extensive experimental results demonstrate that the proposed MLF measure outperforms other existing methods and shows promising generalization ability as well.

The rest of this paper is organized as follows. Section II introduces the related work. We discuss the motivation and show the framework of the proposed method in Section III. We introduce the feature design and fusion in Section IV and present the comprehensive experimental results in Section V. Finally, conclusions are provided in Section VI.

Manuscript received November 21, 2016; revised May 11, 2017 and August 28, 2017; accepted September 27, 2017. Date of publication October 9, 2017; date of current version November 3, 2017. This work was supported in part by the National Research Foundation, Singapore, within the Rapid-Rich Object Search Laboratory, in part by Nanyang Technological University through the Interactive Digital Media Strategic Research Programme, and in part by the Tier 2 Fund of Ministry of Education, Singapore: MOE2016-T2-2-057 (S). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Kalpana Seshadrinathan. *(Corresponding author: Weisi Lin.)*

The authors are with the School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798 (e-mail: zhan0398@ntu.edu.sg; wslin@ntu.edu.sg; qli013@ntu.edu.sg; wcheng005@ntu.edu.sg; xfzhang@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/TIP.2017.2761556

II. Related Work

A. Image Retargeting

Over the past decade many different content-aware image retargeting operators have been proposed, which can be mainly classified as discrete or continuous approaches [15]. The discrete approaches like seam-carving (SC) [1], and shift-map (SM) [6] remove pixels or patches judiciously, while the continuous approaches like non-homogeneous warping (WARP) [4], streaming video (SV) [2], and scale-and-stretch (SNS) [3] seek an optimal mapping (warping) to adapt an image to the target size with certain content protection constraints.

As discussed in [16] the main objectives shared by most image retargeting operators are to preserve important content and internal structure, as well as prevent visual artifacts. Most retargeting operators work well on images with removable regions like sky and water, where the information loss can be easily controlled. When it comes to images with dense information or global patterns, there will be inevitable visual distortion when we minimize the information loss and vice versa. It becomes a challenging problem to achieve an appropriate tradeoff between information loss and visual distortion in the retargeted image.

B. Image Retargeting Quality Assessment

There has been a significant progress in the image quality assessment (IQA) studies [17]- [22], [23] in the past few decades. However, the IQA metrics like structural similarity index (SSIM) [23] cannot be applied to IRQA because they require the reference and distorted images to be of the same resolution.

During recent years there have been a number of studies developing quality measures for image retargeting. In the comparative study, Rubinstein *et al.* [16] evaluated the existing image similarity measures like bidirectional similarity (BDS) and bidirectional warping (BDW) on their benchmark database. This study showed that SIFT flow [24] and earth-mover's distance (EMD) [25] achieved better quality prediction accuracy. Ren *et al.* [26] developed an automatic image retargeting quality method based on the evaluation criteria derived from real user requirements in image retargeting. Liu *et al.* [8] proposed a quality measure based on global geometric structures and local pixel correspondence established by one scale-space matching method. Ma *et al.* [27] examined the quality prediction performance of different existing shape descriptors like MPEG-7 descriptors.

Fang *et al.* [10] proposed an IRQA method called IR-SSIM by evaluating the retained structural information in the retargeted image based on a structural similarity (SSIM) map. Zhang *et al.* [28] extracted the global/local distortions and salient information loss features, and proposed a GLS method using the logistic regression fusion. Liu *et al.* [29] fused spatial and frequency domain features using the machine learning method and achieved promising prediction accuracy. Hsu *et al.* [9] predicted the visual quality based on the measurement of perceptual geometric distortions and information loss. Liang *et al.* [30] considered five factors like the inspiring aesthetics rules and symmetry preservation, and the combination of them correlates well with the human preference. Zhang *et al.* [11] identified the pixel-level correspondence between the original and retargeted images and proposed an effective aspect ratio similarity (ARS) by measuring the aspect ratio changes of local blocks. Recently, Chen *et al.* [31] leveraged the relative quality difference and proposed a general regression neural network (GRNN) model based two-step learning method to rank the retargeted images.

C. Visual Importance Map

Visual importance map is necessary for content-aware image retargeting to identify the important regions, and even for the identical retargeting operator, different visual importance maps may lead to quite different retargeted results. The common visual importance maps such as edge maps (L1 or L2 norm of the gradient), Itti's saliency map [32] and eye gaze measurement [33] are adopted by different retargeting operators. To achieve better performance, some works have further refined the importance map such as the multiplication of the edge map and saliency map [3], and the importance filtering [34], where the saliency map is filtered with the guidance of the image itself to preserve structure consistency. There are also some saliency detection methods like [35], [36] developed specifically for image retargeting, and these visual importance maps play an important role in most existing IRQA works.

III. Overview

A. Motivation for High-Level Features

The visual importance map is the key to achieve accurate quality prediction performance in existing IRQA studies. It ranks image pixels according to their visual importance, so that the information loss and visual distortion in the important regions are prevented as much as possible. The low-level feature based quality measures [10]- [11] highly depend on the visual importance map to achieve the content-aware purpose. Here we show the limitation of the visual importance dependence and the necessity to develop effective features to model the high-level quality degradation aspects.

Although there are many successful visual attention models over the past two decades, the visual importance map for IRQA is not always reliable, or even misleading sometimes. As we can see in Fig. 5(b), the importance map identifies the vase as important region but ignores the woman in the bottom left corner, which leads to inconsistent prediction results with human perception. The low-level feature based measures become vulnerable due to its high visual importance dependence. Although one possible way is to develop better importance map, it will become another very challenging problem. Furthermore, even the ideal importance map has quite limited ability to avoid the visual distortion like the deformation of lines and structures. The importance map only identifies the important region that should be intact during image retargeting process, but it cannot guarantee their intactness after retargeting, especially when the percentage of important region is large or the important structure distributes

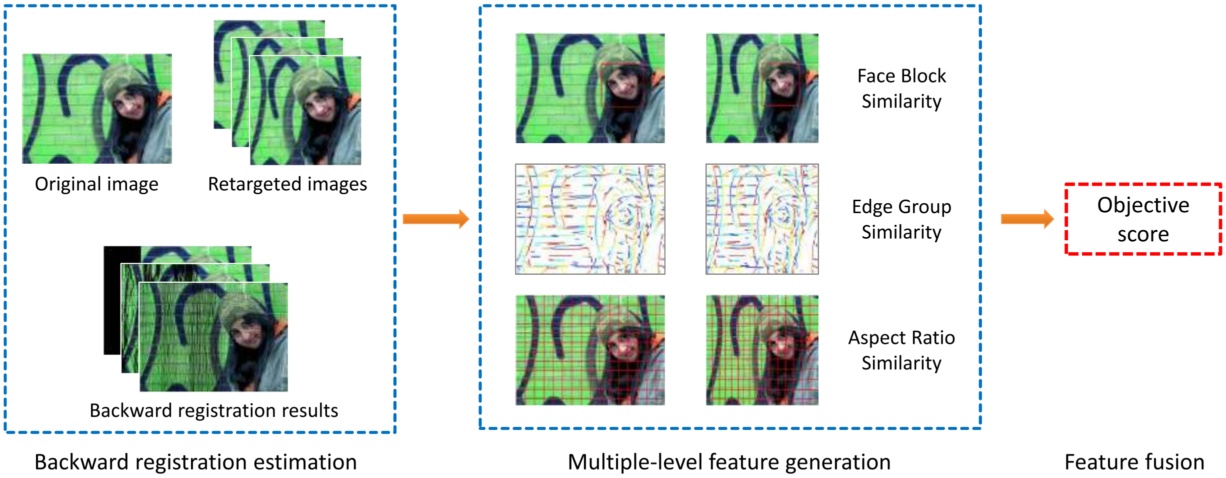


Fig. 1. The block diagram of the proposed multiple-level feature based method. In the first stage, the backward registration [11] is used to estimate the pixel-level correspondence between the original and retargeted images; in the second stage, we generate the multiple-level features to model different aspects of the quality degradation; in the last stage, a learned feature fusion model is used to predict the perceptual quality of the retargeted image.

over the entire image. This is also the reason that additional constraints, e.g. straight line preservation, are adopted in works [2] and [37] to suppress the structure distortions.

Generally we think that visual importance map is effective to measure information loss by estimating the preservation of the important region. On the other hand, it may be inappropriate to only focusing on the distortions in the important region, because the deformations in the unimportant regions with dense information like shape/structure patterns can also degrade the image quality in a significant way. Currently the existing low-level features are ineffective to capture the annoying artifacts like face and structure deformations, which are mostly the semantic high-level distortions. As we are seeking an optimal tradeoff between information loss and visual distortion, it is necessary to accurately measure both of them to develop an effective IRQA measure. Therefore, in this paper we aim to develop multiple-level features to model the visual quality degradation in the retargeted image better.

B. Framework of the Proposed MLF Based Measure

The framework of the proposed measure is shown in Fig. 1. Due to the complicated relationship between the original and retargeted images, we adopt the backward registration method [11] to estimate the pixel-level correspondence in the first stage. With the estimated correspondence, in the second stage we generate the multiple-level features to model different kinds of property changes during image retargeting. ARS captures the information loss and visual distortion of local blocks at the same time. EGS is designed to include the shape/structure related visual distortion by estimating the edge spatial arrangement change. FBS penalizes the sensitive facial region deformation in a top-down manner. In the last stage, a fusion model of multiple-level features is learned to predict the retargeted image quality.

IV. Feature Design and Fusion

In general, each of the features is designed to measure the quality change of certain properties such as shape/structure and human face. The property is first detected or labelled in either the original or retargeted image, and the corresponding property is matched in the other image based on the pixel-level correspondence revealed in the first stage. After the property matching procedure, it is feasible to measure the corresponding quality degradation by estimating to what extent the concerned properties change during the retargeting process. While the pixel-level correspondence matching error in the first stage will possibly disturb the matching and the property changes are complicated in different retargeted images, the designed features are in the form of relative large blocks or edge group representations, to achieve the robust property matching and reliable feature generation.

A. Aspect Ratio Similarity (ARS)

We adopt the low-level feature ARS [11] due to its simplicity and effectiveness. After a brief introduction, we present an in-depth analysis on ARS feature.

1) Brief Introduction: As shown in Fig. 1, the original image is partitioned into regular blocks (e.g. 16 × 16) indicated by the red lines, then the corresponding modified blocks are matched in the retargeted image based on the estimated pixel-level correspondence. The retargeted blocks usually have irregular shapes and we use their bounding boxes as an approximation to capture the local block deformation. For each block pair, the similarity is calculated by

$$s_{ar} = \left[\frac{2 \cdot r_w \cdot r_h + C}{r_w^2 + r_h^2 + C} \right] \cdot \left[e^{-\alpha (r_m - 1)^2} \right], \tag{1}$$

where the width and height change ratios r_w and r_h of their bounding boxes are used for the aspect ratio change, and $r_m = (r_w + r_h)/2$ takes account of the absolute size change. C is a small constant to avoid the division by zero and we choose $C = 10^{-6}$ in our experiments.

The similarity score for the entire image is given by Eq. (2), where N_1 is the number of regular blocks in the original image

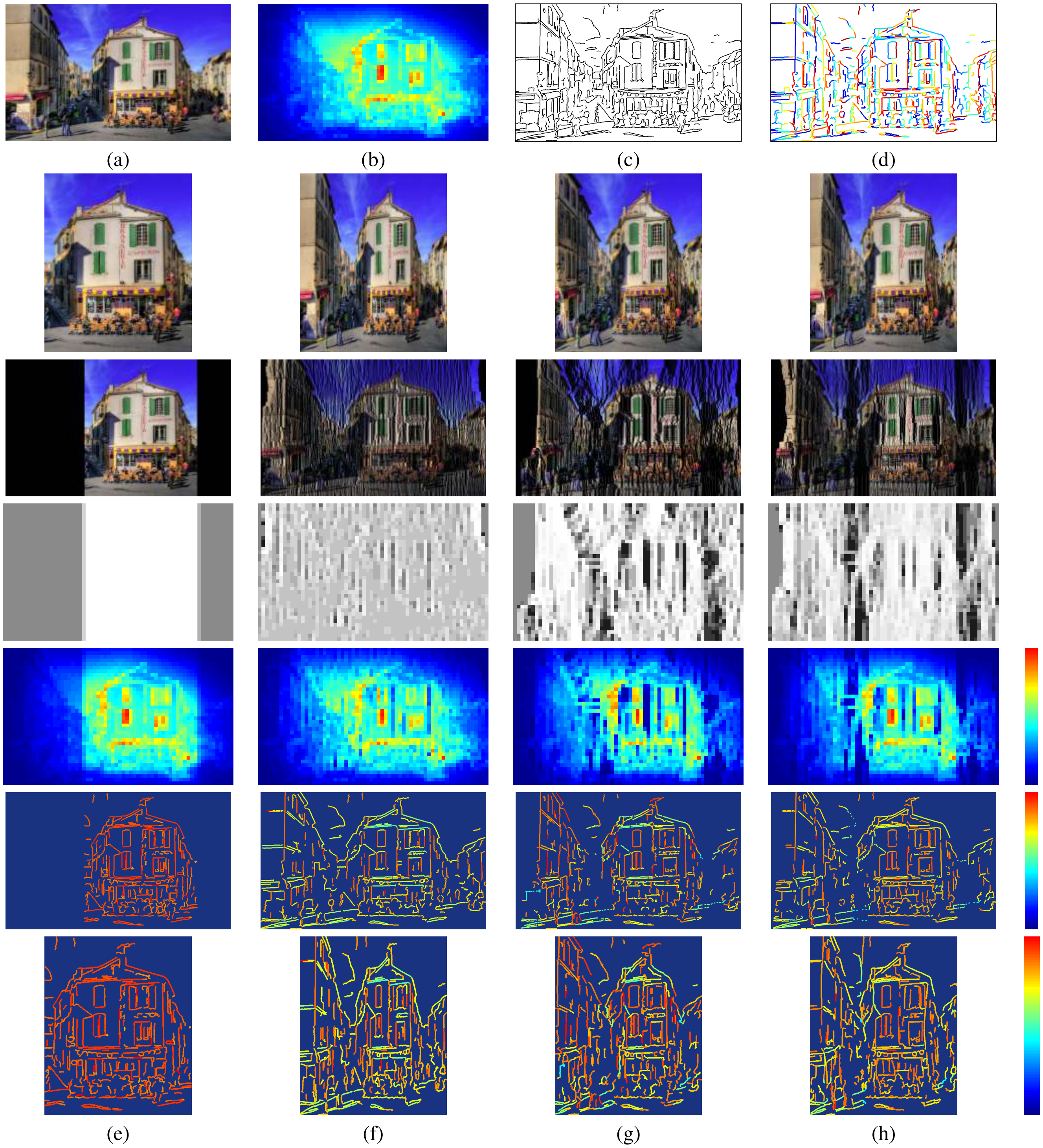


Fig. 2. ‘Brasserie_L_Aficion’ image set. (a)∼(d) are the original image, the corresponding visual importance map, initial edge map, and clustered edge group map. Note that (d) is clustered on the original image individually and not used in following EGS feature generation. For each column in (e) CR, (f) SCL, (g) SC, and (h) WARP, the first row is the image retargeted by the mentioned operators. The second row is the visualized backward registration result. The third row is the calculated ARS map (brightness indicates the quality score) and the fourth row is the importance weighted ARS map using (b). The fifth and sixth rows are EGS maps for the original and retargeted images, where the edge group’s quality is indicated by the colorbar on the right side.

and $w(i)$ is the visual importance [35] weight for the ith block.

$$Q_{AR} = \sum_{i=1}^{N_1} w(i) s_{ar}(i) \tag{2}$$

More ARS maps and importance weighted ARS maps are shown as the third and fourth rows in Fig. 2(e∼h). We can see that ARS relies on the visual importance map to achieve the content-aware quality measurement.

2) Tradeoff Between Information Loss/Visual Distortion: In Eq. (1), the first aspect ratio change term mainly corresponds to the visual distortion occurred in images retargeted by operators like SCL and SNS, while the other absolute size change term complements the information loss missed by

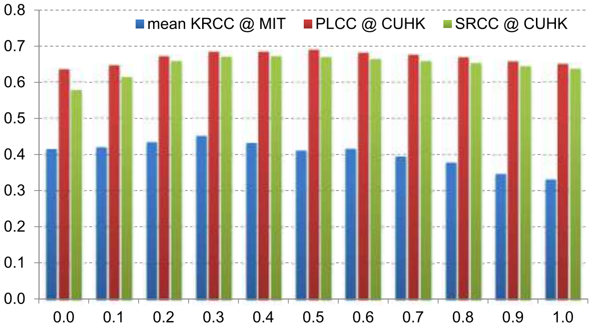


Fig. 3. Influence of α from ARS [11] on MIT and CUHK databases. The prediction accuracies are evaluated using KRCC on MIT database, and PLCC and SRCC on CUHK database with detailed explanations in Section V-A.

the aspect ratio term, which avoids that images retargeted by CR are always favored. Parameter α adjusts the relative importance between the aspect ratio change and the absolute size change terms. Larger α indicates more information loss penalty. It controls the tradeoff between the information loss and visual distortion. The performance of ARS with different α values on the MIT [38] and CUHK [39] databases is shown in Fig. 3. On both databases α has the similar influence on the overall performance. The prediction accuracy increases from 0 to 0.3 and decreases from 0.3 to 1. In the previous work [11], the parameter setting is fixed as $\alpha = 0.3$ for both databases. As we try to develop multiple-level features to measure the information loss and visual distortion in different aspects, the tradeoff between information loss and visual distortion will change obviously when ARS works as one of the features. Since we introduce more features in the following to address visual distortion, we adopt ARS feature Q_{AR} at $\alpha = 0.7$ with more information loss penalty to achieve a new appropriate balance between information loss and visual distortion.

B. Edge Group Similarity (EGS)

In the retargeted image, artifacts like the shape/structure deformation are common distortions that seriously degrade the viewing experience of users, even when they occur in less important regions. The low-level features have limited abilities to model these visual distortions. Due to their visual importance dependence, the distortions in the unimportant region are likely to be ignored as well. We are eager to develop higher level features to capture the distortions like shape/structure deformation directly. As humans can perceive the object using different kinds of cues including color, texture, shape information etc., the features like edge/contour provide effective information for some object-level tasks like object detection and recognition [40]- [44]. Hence, we are intuitively inspired to use the sparse edge groups based shape/structure representation to model the related mid-level information, and accordingly develop the EGS feature to measure the shape/structure related distortion.

1) Edge Group Representation and Matching: For the gradient edge information, we are more interested in the gradient spatial arrangement for the shape/structure information rather than the gradient width or magnitude. Given an image we utilize the structure edge detector [45] to generate the initial edge map. The non-maximum suppression (NMS) [46] and edge group clustering procedures [42] are performed to obtain sparse edge group representation as shown in Fig. 1 and Fig. 2(d), where neighboring edge groups are marked in different colors. The NMS is an edge thinning technique to help reduce the redundant gradient information and preserve the sparse spatial arrangement of the edge response, which is more suitable as the representation of shape/structure.

In the resultant edge map, each edge pixel has the orientation along the edge response. To facilitate the subsequent similarity measurement, it is preferred to cluster the sparse edge pixels into edge groups as the basic evaluation unit. The edge pixels are clustered greedily along the 8-connected neighboring edge pixels until the accumulative orientation difference summation is larger than the specified threshold $\pi/2$, and the small edge group will be merged with the neighboring edge groups further following the setting in [42].

The designed EGS is to measure the similarity between edge groups from the original and retargeted images, but the separate greedy edge group clustering cannot guarantee that the obtained two edge groups are corresponding to the exact same content. This will lead to the unreliable similarity estimation between the edge group pair. Therefore, it is necessary to utilize the pixel-level correspondence revealed in the first stage to match the edge groups.

As EGS feature measures visual distortion in the retargeted image, the main problem is to measure the shape deformation of the edge groups preserved in the retargeted image. Therefore, we first cluster the edge groups in the retargeted image and then match their corresponding edge groups in the original image based on the pixel-level correspondence. To be specific, each edge group $EG' = \{e'_j\}$ in the retargeted image is mapped to the original image to search the corresponding edge group $EG = \{e_i\}$ therein. Since the backward registration estimation is not exactly accurate, we allow the edge pixel searching in the local neighborhood to refine the matching results using the edge information itself. As shown in Fig. 1, the neighboring edge groups are marked in different colors and the edge group pairs are generally matched in high accuracy.

2) Edge Group Similarity: The Chamfer matching proposed by Barrow *et al.* [41] is a popular technique used for contour alignment, object detection and human pose estimation [47], [48]. For the purpose of shape based object searching, it cannot effectively handle the shape variations like the scale, rotation and aspect ratio changes as shown in Fig. 4, and is also unreliable when the background clutter exists.

Since our edge group representations are always in the clean sparse form as shown in Fig. 1, the background clutter influence is negligible. After the edge group representation and matching in the previous part, the Chamfer matching, however, can be used to estimate how the shape/structure represented by the edge group deviates from the original one. At the current stage, we regard all the shape/structure deformations as the perceivable visual distortions, so the Chamfer matching works as an appropriate distance metric to estimate the similarity between the edge group pair.

Here is the Chamfer distance to calculate the similarity between the edge group pair. Let $EG_k = \{e_i\}$ and $EG'_k = \{e'_j\}$ be the kth pair of edge groups in original and retargeted

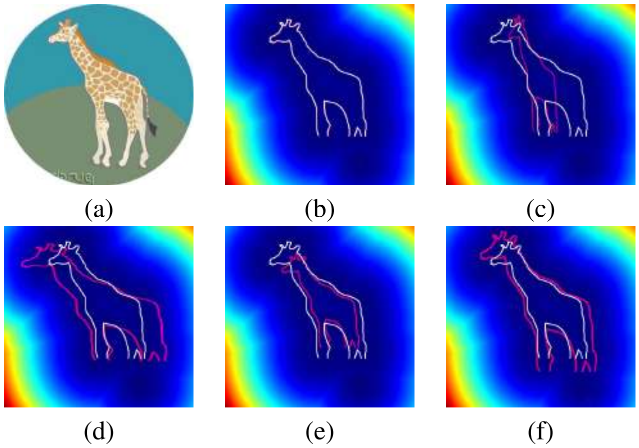


Fig. 4. The illustration of Chamfer matching for different transformed shapes. (a) original image. (b) the shape contour of the giraffe (white). (c∼f) the query shapes (red) with aspect ratio or scale transformation applied on the original shape. The matching distances for (c∼f) are 148.5, 759.4, 84.3, and 217.0, respectively. The distance variations show its limitation for the object search.

images respectively. In Eq. (3), Chamfer distance d_{CM} between EG_k and EG'_k is the average distance of each edge pixel in $e'_j \in EG'_k$ and its nearest match in EG_k, where $L = |EG'_k|$ is the length of edge group EG'_k.

$$d_{CM}(EG_k, EG'_k) = \frac{1}{L} \sum_{e'_j \in EG'_k} \min_{e_i \in EG_k} |e_i - e'_j| \qquad (3)$$

We treat each edge group equally, and EGS for the retargeted image is calculated by Eq. (4), where N_2 is the total number of the edge group pairs. The feature score is transformed into the range of 0 to 1 by an exponential operator,

$$Q_{EG} = e^{-\beta \cdot \sqrt{\frac{1}{N_2}\sum_{k=1}^{N_2} d_{CM}(EG_k, EG'_k)}}, \qquad (4)$$

where β controls the distribution of transformed scores. As the parameter study shown in Fig. 10, different β settings in the range of $[0.1, 1]$ achieve similar performance and in the experiments, we choose the setting $\beta = 0.2$.

More EGS maps for the original and retargeted images are shown in the fifth and sixth rows in Fig. 2(e∼h), and the edge group quality is indicated by the colorbar on the right side. EGS maps for the original and retargeted images are actually equivalent, but we present both of them to show their matching relationship as well. Since EGS is formulated to measure the shape/structure related distortion, the images retargeted by CR are always in high quality. The comparisons between EGS feature and ARS [11] are shown in Fig. 5. From the results, we can see that the low-level feature ARS, which relies on the visual importance, tends to ignore the distortion in the unimportant region, which leads to the prediction results inconsistent with the subjective votes in these cases. Even though EGS only measures the visual distortion, the predictions Q_{EG} by the EGS feature correlate better with the human preference than ARS, when there is no obvious content removal as shown in Fig. 5.

C. Face Block Similarity (FBS)

Face is an important element in a large number of images, and usually attracts more attentions. The study [16] shows that viewers show consistently high sensitivity to face deformation. Therefore, it will be beneficial to develop specific high-level features to capture the deformation in facial region.

As shown in Fig. 1, FBS feature is designed to explicitly address the annoying deformation in the facial region indicated by the red boxes. Beyond the local distortions measured by ARS, FBS feature calculates the facial region deformation at the face block scale and measures the visual distortions with explicit semantic meaning. To obtain a satisfied face detection rate, we adopt the Face++ toolkit [49] to detect human faces in the original image, and the retargeted faces are established using the bounding box based on the pixel-level correspondence. In Fig. 6, we show more examples of the detected faces in the original and their matched face blocks in retargeted images. The deformation between each face block pair is calculated with Eq. (1) denoted as $s_{ar}(n)$ and FBS score is given by

$$Q_{FB} = \begin{cases} \frac{1}{N_3} \sum_{n=1}^{N_3} s_{ar}(n) & N_3 > 0 \\ 1 & N_3 = 0, \end{cases} \qquad (5)$$

where N_3 is the total number of detected faces.

Since human faces appear in a portion of natural images, for the general purpose, only the face images will be penalized by FBS and other images will be assigned with the high quality score $Q_{FB} = 1$. The similarity scores for the multiple faces are averaged within one image. The novel FBS feature, capturing the sensitive facial region deformation, is nontrivial compared to previous works, where the face detection is only used to refine the visual importance map [10], [34] or classify the images into binary categories [29].

D. Multiple-Level Features Fusion

A feature fusion model is learned to map feature scores into quality indices, providing a final measure of the retargeted image quality. In our implementation, a support vector machine (SVM) regression (SVR) [50] is adopted for learning on CUHK [39] database. SVR has been widely applied to IQA problems [51]- [53]. We use the LIBSVM package [54] to implement SVR on CUHK database with radial basis function (RBF) kernel. Different from the mean opinion score (MOS) provided on CUHK database, pair-wise comparison votes are provided on MIT RetargetMe [16], where only rankings among retargeted images are reliable. The direct regression using SVR is not efficient. By this consideration, we utilize the SVM^{rank} implementation [55] with RBF kernel for ordinal regression on MIT database instead.

V. Experimental Results

In this section we introduce the MIT RetargetMe [16] and CUHK [39] databases and their performance evaluation criteria, and then present a series of performance evaluation of the proposed MLF on the two public databases.

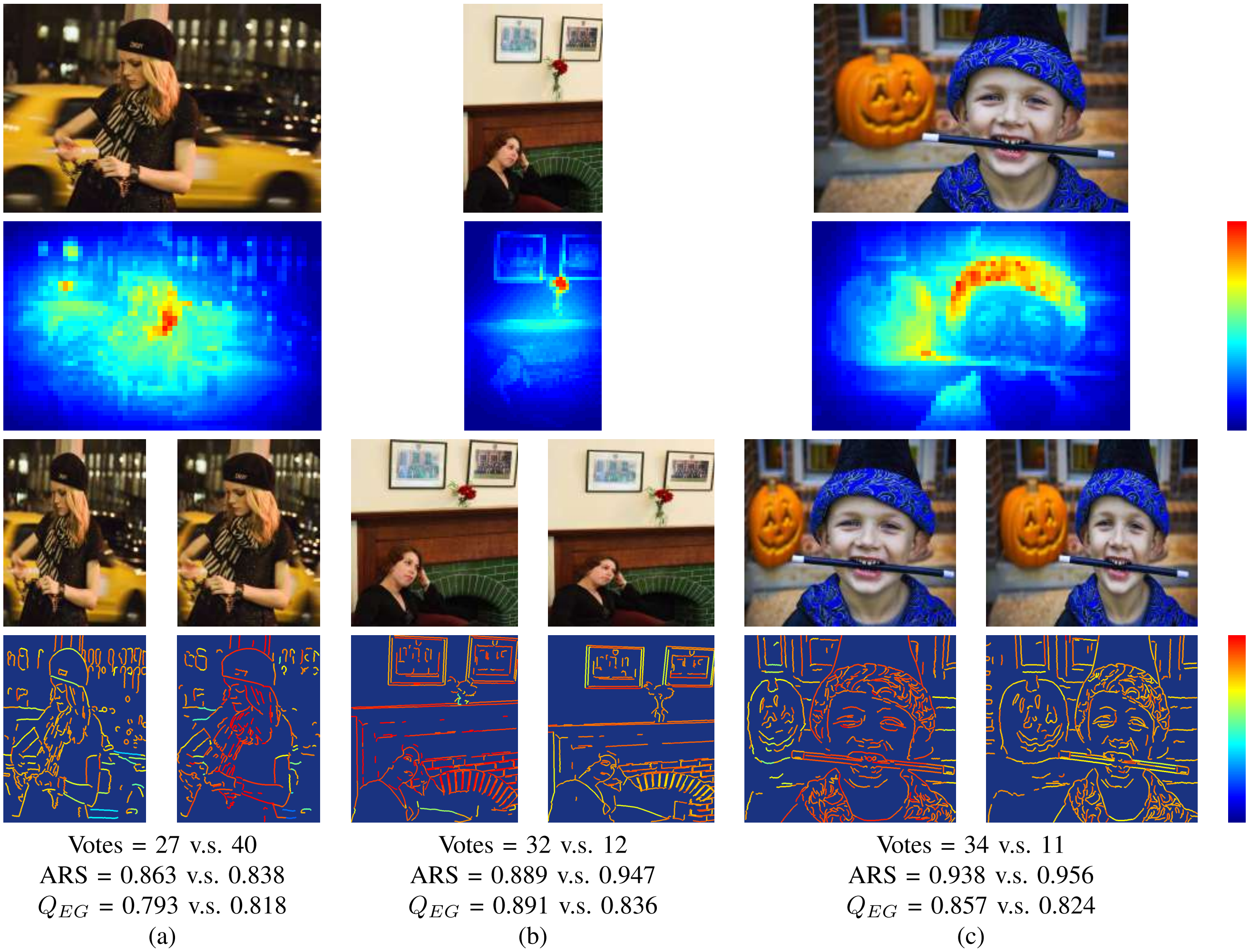


Fig. 5. Comparisons of the EGS feature Q_{EG} and ARS [11]. From the top to the bottom are the original image, visual importance map, retargeted images and the EGS maps. (a)'DKNYgirl' image with the MOP (left) and SC (right) retargeting operators. (b)'Woman' image with the SC (left) and SCL (right) retargeting operators. (b)'child' image with the SC (left) and SCL (right) retargeting operators.

A. Image Retargeting Databases

1) MIT RetargetMe Database: In the MIT RetargetMe database, there are 37 original images for our evaluation. Each image has been retargeted by eight typical retargeting operators including CR, SV [2], MOP [5], SC [1], SCL, SM [6], SNS [3], WARP [4] and there are 296 retargeted images in total. There are 23 original images with 25% height or width reduction and 14 images with 50% height or width reduction. There are 210 people participated in the subjective tests. The tests are the pairwise comparisons, where the subjects vote for their favored one from two retargeted images shown side-by-side. The database provides the subjective votes as the subjective scores for the objective quality measure evaluation. We adopt the Kendall rank correlation coefficient (KRCC) [56] to measure the correlation between the objective scores and subjective rankings:

$$\mathrm{KRCC} = 1 - \frac{4N_d}{N(N-1)} \tag{6}$$

where N is the ranking length (here $N = 8$) and N_d is the number of discordant pairs from all the pairs (the maximal $N_d = 28$). To investigate results for specific characteristics, the images are also labelled with one or more attributes including *Line/Edge, Faces/People, Foreground Objects, Texture, Geometric Structures* and *Symmetry*.

2) CUHK Database: In the CUHK database there are 57 original images and 171 retargeted images. The optimized seam-carving and scale [57] and energy-based deformation [58] retargeting operators, are also included along with the eight operators from the MIT database. Each image is retargeted with 25% or 50% of width or height reduction using the selected operators. Different from the pairwise way in MIT database, the subjective experiments utilize 5-category scales to generate MOSs, which is used for the correlation evaluation similar to that in IQA [17], [23].

Four commonly used performance metrics are used to evaluate the objective quality measures. The first one is the Pearson linear correlation coefficient (PLCC). To compute the PLCC we need to apply a nonlinear mapping between MOSs and the objective scores. The second one is the Spearman rank-order correlation coefficient (SRCC), which measures the prediction monotonicity. The third one is the root mean squared error (RMSE) between MOSs and the objective scores. The last one, the outlier ratio (OR) [59] is the percentage of false

TABLE I
PERFORMANCE COMPARISON OF DIFFERENT IRQA MEASURES ON MIT DATABASE BY KRCC

Methods	mean KRCC for each subset						Total		
	Line Edges	*Faces People*	*Foreground Objects*	*Texture*	*Geometric Structure*	*Symmetry*	mean KRCC	std KRCC	*p*-val
BDS [61]	0.040	0.190	0.167	0.060	-0.004	-0.012	0.083	0.268	0.017
EH [62]	0.043	-0.076	-0.079	-0.060	0.103	0.298	0.004	0.334	0.641
SIFT flow [24]	0.097	0.252	0.218	0.161	0.085	0.071	0.145	0.262	0.031
EMD [25]	0.220	0.262	0.226	0.205	0.237	0.500	0.251	0.272	1e-5
CSim [8]	0.097	0.290	0.293	0.161	0.053	0.150	0.164	0.263	0.028
PGDIL [9]	0.431	0.390	0.389	0.286	0.438	0.523	0.415	0.296	6e-10
ARS [11]	0.463	0.519	0.444	0.330	0.505	0.464	0.452	0.283	1e-11
MLF	**0.486**	**0.605**	**0.544**	**0.384**	**0.536**	**0.536**	**0.512**	**0.251**	**1e-14**

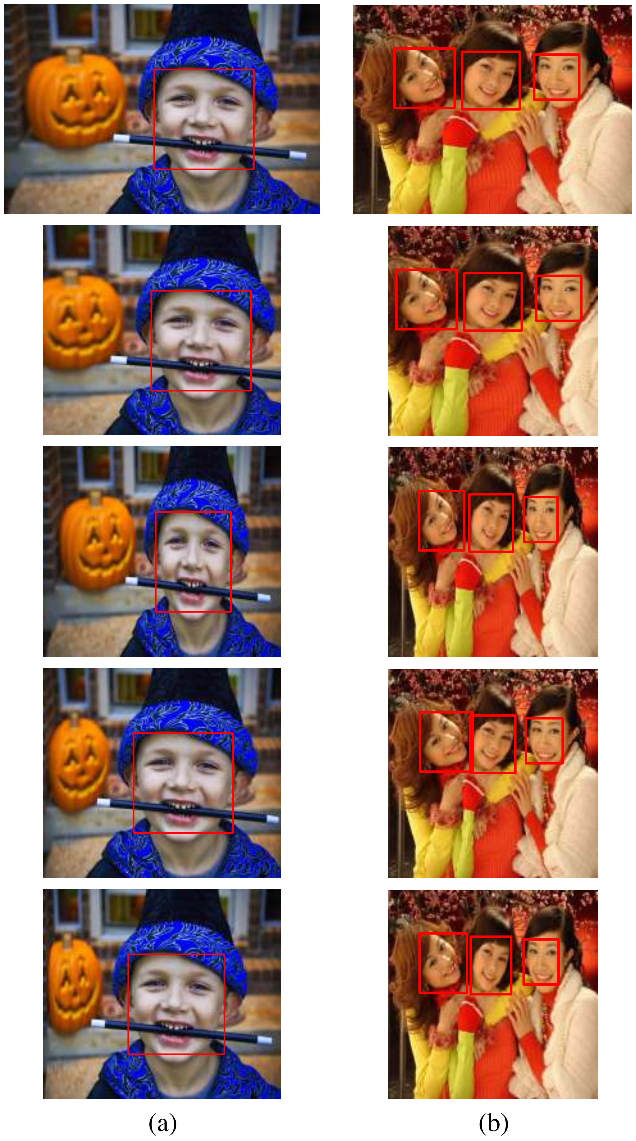


Fig. 6. Examples of face images. The 'child' (a) and 'girls' (b) images are retargeted by CR, SCL, SC and WARP, respectively. As indicated by the red boxes, the facial regions are first detected in the original image and the corresponding facial regions are identified in retargeted images using the pixel-level correspondence estimated in the first stage.

prediction scores, which lie outside the interval [MOS-2σ, MOS+2σ] (where σ is the standard deviation of MOS) after the nonlinear regression. For the nonlinear regression, we use the mapping function as suggested by Sheikh *et al* [60]:

$$f(x) = \beta_1\left(\frac{1}{2} - \frac{1}{1 + e^{\beta_2(x-\beta_3)}}\right) + \beta_4 x + \beta_5 \qquad (7)$$

where $\beta_1, \beta_2, \ldots, \beta_5$ are parameters to be fitted. In general a good IRQA measure should have high PLCC, SRCC scores and low RMSE, OR scores.

B. Performance Evaluation on MIT Database

Following the previous work [30] we apply leave-one-out cross-validation (LOOCV) on MIT database with SVM^{rank} [55] using RBF kernel for the feature fusion. In each image set, we treat the original image as a query and the subjective votes of the 8 retargeted images as their quality ranking. There are 37-folds in total and in each fold we use one set for testing and the rest 36 sets for training. After LOOCV, each set is tested once exactly, and then their results are evaluated accordingly.

The rank correlation results on MIT database [38] are shown in Table I. We compare the proposed MLF measure with BDS [61], EH [62], SIFT flow [24], EMD [25], CSim [8], PGDIL [9] and ARS [11]. We present the mean and standard deviation of KRCC for each IRQA measure, as well as the mean KRCC for each image subset with certain attributes. From the results, we can see that our proposed measure outperforms other measures in each image subset as well as on the overall database. In Fig. 7, we compare the individual KRCCs of ARS and MLF for the 37 image sets from MIT database. Compared with ARS, the proposed MLF shows more stable and reliable quality prediction accuracy.

We also follow the *top 3 ranking* KRCC evaluation strategy adopted in [16] and [9], which we denote it as KRCC3 here. For the $N = 8$ length ranking, all 28 ranking pairs are compared for the complete rank correlation calculation, while the KRCC3 evaluation *s.t.* $(rank_s(i) \leqslant 3 \vee rank_s(j) \leqslant 3) \wedge (rank_o(i) \leqslant 3 \vee rank_o(j) \leqslant 3)$, focuses more on the prediction accuracy for the top ranking images. As the constraint shows, there are the *top 3 ranking* selections for both subjective ranking and objective ranking. With the *top 3 ranking* selection for the subjective ranking, we compare the constant 18 high ranking pairs. The additional *top 3 ranking* selection for the objective ranking will further reduce it to a uncertain smaller number. For example, the average number of the compared pairs for EMD is 13.30 and its standard deviation

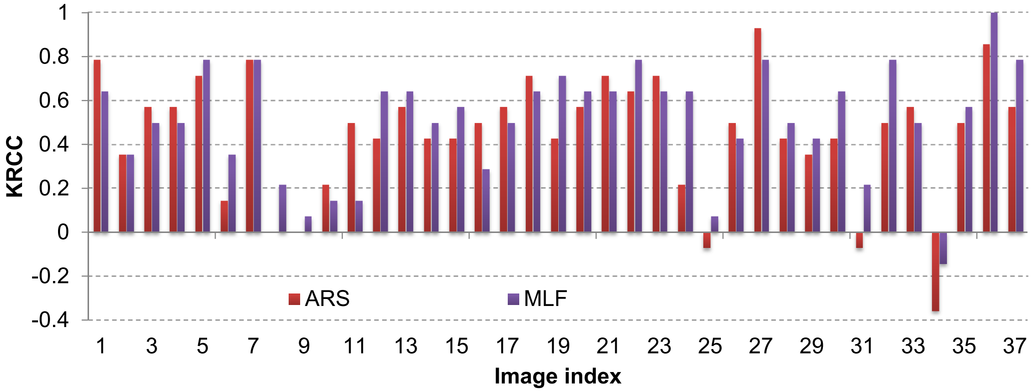


Fig. 7. The individual KRCC results of ARS [11] and MLF measure for the 37 image sets on MIT database. For each image set (including eight retargeted images), the KRCC is calculated as the rank correlation between the ranking of subjective and objective scores.

TABLE II

PERFORMANCE COMPARISON OF DIFFERENT MEASURE ON MIT DATABASE BY TWO KINDS OF KRCC3. THE FIRST ONE (LEFT) CONSIDERS THE PAIRS *s.t.* $(rank_s(i) \leqslant 3 \vee rank_s(j) \leqslant 3) \wedge (rank_o(i) \leqslant 3 \vee rank_o(j) \leqslant 3)$, WHILE THE SECOND ONE (RIGHT) CONSIDERS THE PAIRS *s.t.* $rank_s(i) \leqslant 3 \vee rank_s(j) \leqslant 3$

Methods	*top 3* subjective ranking *top 3* objective ranking		*top 3* subjective ranking	
	mean KRCC3	std KRCC3	mean KRCC3	std KRCC3
BDS [61]	0.108	0.532	0.132	0.372
EH [62]	-0.071	0.593	0.009	0.413
SIFT flow [24]	0.298	0.483	0.255	0.348
EMD [25]	0.326	0.496	0.339	0.442
CSim [8]	0.277	0.467	0.240	0.305
PGDIL [9]	**0.533**	**0.383**	0.532	**0.280**
ARS [11]	0.450	0.469	0.520	0.347
MLF	0.518	0.459	**0.589**	0.310

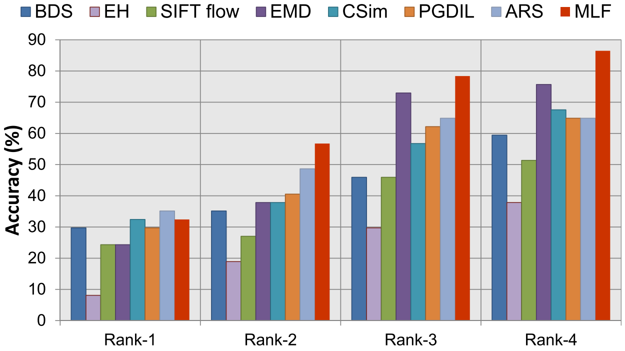


Fig. 8. *Rank-n* accuracy of different IRQA measures. The rank prediction by MLF shows better overall prediction accuracy.

is 2.47. It indicates that we are comparing a varying number of pairs for the objective measures and it may be biased towards certain measures. To exclude this possible influence factor, we also compare the KRCC3 performance only with the *top 3 ranking* selection for the subjective ranking. As shown in Table II, we can observe that our proposed MLF measure shows the comparable or better prediction performance.

TABLE III

PERFORMANCE OF DIFFERENT QUALITY MEASURES ON CUHK DATABASE

Methods	PLCC	SRCC	RMSE	OR
BDS [61]	0.2896	0.2887	12.922	0.2164
EH [62]	0.3422	0.3288	12.686	0.2047
SIFT flow [24]	0.3141	0.2899	12.817	0.1462
EMD [25]	0.2760	0.2904	12.977	0.1696
CSim [8]	0.4374	0.4662	12.141	0.1520
GLS [28]	0.4622	0.4760	10.932	0.1345
PGDIL [9]	0.5403	0.5409	11.361	0.1520
ARS [11]	0.6835	0.6693	9.855	0.0702
MLF	**0.7577**	**0.7383**	**8.525**	**0.0294**

TABLE IV

INVESTIGATION OF INDIVIDUAL FEATURES ON CUHK DATABASE

Features	PLCC	SRCC	RMSE	OR
SCD	0.1508	0.1792	13.347	0.2164
CSD	0.1520	0.1688	32.731	0.5322
CLD	0.1033	0.0850	13.429	0.2398
HTD	0.0829	0.0890	35.151	0.5673
EHD	0.3031	0.2729	12.866	0.2047
EMD	0.2760	0.2904	12.977	0.1696
PHOW	0.3706	0.2308	12.540	0.2222
GIST	0.5443	0.5114	11.326	0.1579
Q_{AR}	0.6752	0.6587	9.960	0.0760
Q_{EG}	0.5706	0.5617	11.087	0.1287
Q_{FB}	0.8052	0.6984	8.717	0.0909

Furthermore, to examine the ability to select the best subject-rated retargeted image, we measure the *rank-n* accuracy for different IRQA measures. Given the eight retargeted images for each original image, the *rank-n* accuracy is the percentage of the objective scores where the subjective-rated best one is within their *top n* positions. In Fig. 8, we present the *rank-1*, *rank-2*, *rank-3*, and *rank-4* accuracies for MLF and other measures on MIT database. We can see that the proposed MLF measure shows better overall prediction accuracy than other measures.

C. Performance Evaluation on CUHK Database

On CUHK database MLF measure is evaluated using SVR with 5-fold cross-validation following IQA method

TABLE V
FEATURE ANALYSIS ON MIT DATABASE

No.	Feature combination			mean KRCC for each Attribute						Total	
	Q_{AR}	Q_{EG}	Q_{FB}	*Line Edges*	*Faces People*	*Foreground Objects*	*Texture*	*Geometric Structure*	*Symmetry*	mean KRCC	std KRCC
1	✓			0.369	0.424	0.401	0.313	0.424	0.536	0.396	0.244
2		✓		0.217	0.257	0.270	0.268	0.250	0.024	0.222	0.251
3*			✓	-	0.464	-	-	-	-	0.464	-
4	✓	✓		0.463	0.576	0.524	0.357	0.522	0.512	0.492	0.249
5	✓		✓	0.383	0.462	0.425	0.339	0.424	0.536	0.411	0.249
6		✓	✓	0.229	0.310	0.306	0.304	0.250	0.024	0.243	0.243
7	✓	✓	✓	0.486	0.605	0.544	0.384	0.536	0.536	0.512	0.251

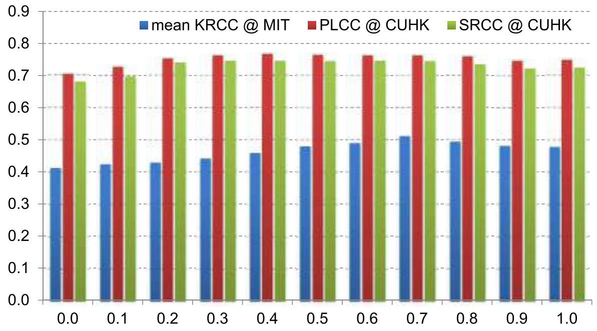


Fig. 9. Influence of α from MLF measure on MIT and CUHK databases.

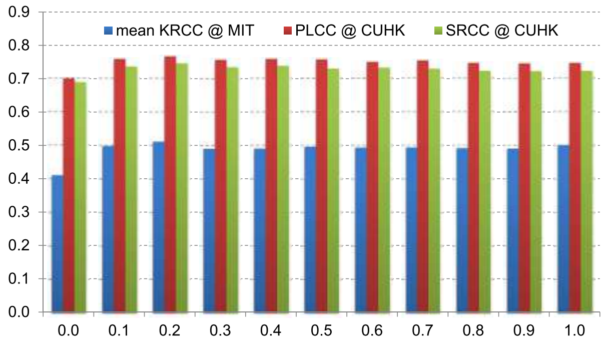


Fig. 10. Influence of β from MLF measure on MIT and CUHK databases.

BRISQUE [52]. We divide the database into two randomly chosen subsets, 80% training and 20% testing. There will be no overlap between train and test subsets. We repeat the random train-test procedure 1000 times and report the median performance of these 1000 iterations.

In Table III, we compare the performance of the proposed MLF measure with BDS [61], EH [62], SIFT flow [24], EMD [25], CSim [8], GLS [28], PGDIL [9] and ARS [11] on CUHK database using PLCC, SRCC, RMSE and OR. As the results show, MLF achieves significantly better prediction performance compared with other existing measures.

In Table IV we also compare the MLF features individually with other shape descriptor features evaluated in [27] like Pyramid Histogram of Visual Words (PHOW) [63] and GIST [64]. We can see that all our features show competitive performance compared with other existing shape features. When the Q_{FB} works as a individual feature, it is evaluated among the 33 detected face images from the 171 retargeted images. It should be clear that FBS feature can predict the quality of face images rather effectively and it is worthy to develop the specific high-level features for these sensitive attributes.

D. Parameter Study

We provide the sensitivity studies for parameters α and β from the proposed MLF measure on MIT and CUHK databases in Figs. 9 and 10. Since the major part of information loss is captured by ARS feature in MLF measure, α still plays an important role in the tradeoff between information loss and visual distortion similar to that in ARS [11]. The influence of α in Fig. 9 is similar to that in Fig. 3, but the peak shifts from 0.3 to 0.7. Visual distortions measured by EGS and FBS features have broken the balance in [11], and a larger α with more information loss penalty helps to achieve the new optimal balance. We have also observed that the variation tendency of the overall performance on CUHK database is relatively less obvious compared with that on MIT database. In Fig. 10 we show the influence of β from EGS feature on MLF measure. The overall performance on both databases is relatively stable with small variations when β is larger than 0.1. In general, the performance of MLF measure is insensitive to the change of β in the range of [0.1, 1], but still influenced by α with regard to the tradeoff between information loss and visual distortion.

E. Feature Analysis

To further investigate the significance of different features, in Tables V and VI we conduct the performance evaluation using different feature combinations on MIT and CUHK databases, respectively. In combination #3*, only face images are focused, since other images are assigned with the same scores. From the tables, we can see that the combination #7 with all features outperforms other feature combinations with noticeable margins on both databases. In combination #6, EGS and FBS features mainly measure visual distortion and without effective information loss measurement, their performance has decreased obviously compared with #7. ARS feature measures information loss and visual distortion at the same time, and is able to work as a baseline part individually. Based on the comparisons of #4, #7 and #5, #7, we can observe

TABLE VI
FEATURE ANALYSIS ON CUHK DATABASE

No.	Feature combination			PLCC	SRCC	RMSE	OR
	Q_{AR}	Q_{EG}	Q_{FB}				
1	✓			0.6752	0.6587	9.960	0.0760
2		✓		0.6179	0.5725	10.616	0.1228
3*			✓	0.8052	0.6984	8.717	0.0909
4	✓	✓		0.7314	0.7089	9.029	0.0588
5	✓		✓	0.7015	0.6900	9.393	0.0588
6		✓	✓	0.5898	0.5527	10.642	0.1176
7	✓	✓	✓	0.7577	0.7383	8.525	0.0294

that EGS and FBS features both are able to further improve the overall performance effectively.

F. Cross Database Performance Evaluation

In order to check the generalization ability of MLF, we conduct the cross database performance evaluation by training on one entire database and testing on the other one. First, we use SVR to learn the feature fusion model on the entire CUHK database, and the mean KRCC and its standard deviation tested on MIT database are 0.469 and 0.256, respectively. Next, we perform the learning on MIT database using SVM^{rank}. The PLCC, SRCC, RMSE and OR tested on CUHK database are 0.730, 0.713, 9.230, and 0.047, respectively. Compared with Tables I and III, we can observe that MLF's cross database performance is inferior to the performance when training and testing data come from the same database but still better than that of other existing quality measures. It should be clear that MLF performs better in terms of correlation with human perception and has good generalization ability across databases as well.

VI. CONCLUSION

For the high-level IRQA task, we have proposed a multiple-level feature based framework to predict the perceptual retargeted image quality and developed three effective general-purpose features to model various kinds of quality degradation factors in the retargeted image. On the public MIT and CUHK databases, the proposed MLF measure shows better and more reliable quality prediction performance compared with other existing methods. We pointed out that the existing low-level feature based methods are limited due to the visual importance dependency and the lack of high-level visual distortion modelling. While the proposed method addresses the related problems and shows promising performance, it is still necessary to develop more effective features as well as the correspondence estimation and visual importance map algorithms for better IRQA measures in the future work.

REFERENCES

[1] M. Rubinstein, A. Shamir, and S. Avidan, "Improved seam carving for video retargeting," *ACM Trans. Graph.*, vol. 27, no. 3, p. 16, 2008.

[2] P. Krähenbühl, M. Lang, A. Hornung, and M. H. Gross, "A system for retargeting of streaming video," *ACM Trans. Graph.*, vol. 28, no. 5, p. 126, 2009.

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

[4] L. Wolf, M. Guttmann, and D. Cohen-Or, "Non-homogeneous content-driven video-retargeting," in *Proc. IEEE 11th Int. Conf. Comput. Vis. (ICCV)*, Oct. 2007, pp. 1–6.

[5] M. Rubinstein, A. Shamir, and S. Avidan, "Multi-operator media retargeting," *ACM Trans. Graph.*, vol. 28, no. 3, p. 23, 2009.

[6] Y. Pritch, E. Kav-Venaki, and S. Peleg, "Shift-map image editing," in *Proc. IEEE 12th Int. Conf. Comput. Vis. (ICCV)*, Sep./Oct. 2009, pp. 151–158.

[7] F. Shao, W. Lin, W. Lin, Q. Jiang, and G. Jiang, "QoE-guided warping for stereoscopic image retargeting," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4790–4805, Oct. 2017.

[8] Y.-J. Liu, X. Luo, Y.-M. Xuan, W.-F. Chen, and X.-L. Fu, "Image retargeting quality assessment," *Comput. Graph. Forum*, vol. 30, no. 2, pp. 583–592, Apr. 2011.

[9] C.-C. Hsu, C.-W. Lin, Y. Fang, and W. Lin, "Objective quality assessment for image retargeting based on perceptual geometric distortion and information loss," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 3, pp. 377–389, Jun. 2014.

[10] Y. Fang, K. Zeng, Z. Wang, W. Lin, Z. Fang, and C.-W. Lin, "Objective quality assessment for image retargeting based on structural similarity," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 4, no. 1, pp. 95–105, Mar. 2014.

[11] Y. Zhang, Y. Fang, W. Lin, X. Zhang, and L. Li, "Backward registration-based aspect ratio similarity for image retargeting quality assessment," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4286–4297, Sep. 2016.

[12] Y. Zhang, W. Lin, X. Zhang, Y. Fang, and L. Li, "Aspect ratio similarity (ARS) for image retargeting quality assessment," in *Proc. IEEE ICASSP*, Mar. 2016, pp. 1080–1084.

[13] Q. Jiang, F. Shao, W. Lin, and G. Jiang, "Learning sparse representation for objective image retargeting quality assessment," *IEEE Trans. Cybern.*, to be published. [Online]. Available: http://ieeexplore.ieee.org/document/7898810/

[14] M. Karimi, S. Samavi, N. Karimi, S. M. R. Soroushmehr, W. Lin, and K. Najarian, "Quality assessment of retargeted images by salient region deformity analysis," *J. Vis. Commun. Image Represent.*, vol. 43, pp. 108–118, Feb. 2017.

[15] A. Shamir and O. Sorkine, "Visual media retargeting," in *Proc. ACM SIGGRAPH ASIA Courses*, 2009, p. 11.

[16] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir, "A comparative study of image retargeting," *ACM Trans. Graph.*, vol. 29, no. 5, pp. 160:1–160:10, 2010.

[17] 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, 2011.

[18] L. Li, W. Lin, X. Wang, G. Yang, K. Bahrami, and A. C. Kot, "No-reference image blur assessment based on discrete orthogonal moments," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 39–50, Jan. 2016.

[19] L. Li, W. Lin, and H. Zhu, "Learning structural regularity for evaluating blocking artifacts in JPEG images," *IEEE Signal Process. Lett.*, vol. 21, no. 8, pp. 918–922, Aug. 2014.

[20] L. Li, Y. Zhou, W. Lin, J. Wu, X. Zhang, and B. Chen, "No-reference quality assessment of deblocked images," *Neurocomputing*, vol. 177, pp. 572–584, Feb. 2016.

[21] X. Zhang, R. Xiong, X. Fan, S. Ma, and W. Gao, "Compression artifact reduction by overlapped-block transform coefficient estimation with block similarity," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 4613–4626, Dec. 2013.

[22] S. Wang, K. Gu, X. Zhang, W. Lin, S. Ma, and W. Gao, "Reduced-reference quality assessment of screen content images," *IEEE Trans. Circuits Syst. Video Technol.*, to be published. [Online]. Available: http://ieeexplore.ieee.org/document/7552436/

[23] 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.

[24] C. Liu, J. Yuen, and A. Torralba, "SIFT flow: Dense correspondence across scenes and its applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 978–994, May 2011.

[25] O. Pele and M. Werman, "Fast and robust earth mover's distances," in *Proc. IEEE 12th Int. Conf. Comput. Vis. (ICCV)*, Sep./Oct. 2009, pp. 460–467.

[26] T. Ren and G. Wu, "Automatic image retargeting evaluation based on user perception," in *Proc. IEEE ICIP*, Sep. 2010, pp. 1569–1572.

[27] L. Ma, L. Xu, H. Zeng, K. N. Ngan, and C. Deng, "How does the shape descriptor measure the perceptual quality of the retargeting image?" in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICME)*, Jul. 2014, pp. 1–6.

[28] J. Zhang and C. C. J. Kuo, "An objective quality of experience (QoE) assessment index for retargeted images," in *Proc. ACM Multimedia*, 2014, pp. 257–266.

[29] A. Liu, W. Lin, H. Chen, and P. Zhang, "Image retargeting quality assessment based on support vector regression," *Signal Process. Image Commun.*, vol. 39, pp. 444–456, Nov. 2015.

[30] Y. Liang, Y.-J. Liu, and D. Gutierrez, "Objective quality prediction of image retargeting algorithms," *IEEE Trans. Vis. Comput. Graph.*, vol. 23, no. 2, pp. 1099–1110, Feb. 2017.

[31] Y. Chen, Y.-J. Liu, and Y. Lai, "Learning to rank retargeted images," in *Proc. IEEE CVPR*, Jun. 2017, pp. 3994–4002.

[32] L. Itti, C. Koch, and E. Niebur, "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.

[33] D. DeCarlo and A. Santella, "Stylization and abstraction of photographs," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 769–776, 2002.

[34] Y. Ding, J. Xiao, and J. Yu, "Importance filtering for image retargeting," in *Proc. IEEE CVPR*, Jun. 2011, pp. 89–96.

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

[36] R. Achanta and S. Süsstrunk, "Saliency detection for content-aware image resizing," in *Proc. IEEE ICIP*, Nov. 2009, pp. 1005–1008.

[37] C.-H. Chang and Y.-Y. Chuang, "A line-structure-preserving approach to image resizing," in *Proc. IEEE CVPR*, Jun. 2012, pp. 1075–1082.

[38] *RetargetMe Benchmark*. [Online]. Available: http://people.csail.mit.edu/mrub/retargetme

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

[40] M. Eitz, J. Hays, and M. Alexa, "How do humans sketch objects?" *ACM Trans. Graph.*, vol. 31, no. 4, p. 44, Jul. 2012.

[41] H. G. Barrow, J. M. Tenenbaum, R. C. Bolles, and H. C. Wolf, "Parametric correspondence and chamfer matching: Two new techniques for image matching," in *Proc. IJCAI*, 1977, pp. 659–663.

[42] C. L. Zitnick and P. Dollár, "Edge boxes: Locating object proposals from edges," in *Computer Vision—ECCV* (Lecture Notes in Computer Science), vol. 8693. Zürich, Switzerland: Springer, 2014, pp. 391–405.

[43] Y.-J. Liu, C.-C. Yu, M.-J. Yu, and Y. He, "Manifold SLIC: A fast method to compute content-sensitive superpixels," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 651–659.

[44] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, Apr. 2002.

[45] P. Dollár and C. L. Zitnick, "Structured forests for fast edge detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 1841–1848.

[46] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[47] M.-Y. Liu, O. Tuzel, A. Veeraraghavan, and R. Chellappa, "Fast directional chamfer matching," in *Proc. IEEE CVPR*, Jun. 2010, pp. 1696–1703.

[48] A. Thayananthan, B. Stenger, P. H. S. Torr, and R. Cipolla, "Shape context and chamfer matching in cluttered scenes," in *Proc. IEEE CVPR*, Jun. 2003, pp. 127–133.

[49] M. Inc. (Dec. 2013). *Face++ Research Toolkit*. [Online]. Available: www.faceplusplus.com

[50] B. Schölkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, "New support vector algorithms," *Neural Comput.*, vol. 12, no. 5, pp. 1207–1245, 2000.

[51] M. Narwaria and W. Lin, "SVD-based quality metric for image and video using machine learning," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 2, pp. 347–364, Apr. 2012.

[52] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

[53] M. Narwaria and W. Lin, "Objective image quality assessment based on support vector regression," *IEEE Trans. Neural Netw.*, vol. 21, no. 3, pp. 515–519, Mar. 2010.

[54] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, 2011. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm

[55] T. Joachims, "Training linear SVMs in linear time," in *Proc. ACM KDD*, 2006, pp. 217–226.

[56] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, vol. 30, pp. 81–93, Jun. 1938.

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

[58] Z. Karni, D. Freedman, and C. Gotsman, "Energy-based image deformation," *Comput. Graph. Forum*, vol. 28, no. 5, pp. 1257–1268, 2009.

[59] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Process., Image Commun.*, vol. 19, no. 2, pp. 121–132, 2004.

[60] 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.

[61] D. Simakov, Y. Caspi, E. Shechtman, and M. Irani, "Summarizing visual data using bidirectional similarity," in *Proc. IEEE CVPR*, Jun. 2008, pp. 1–8.

[62] 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.

[63] A. Bosch, A. Zisserman, and X. Muñoz, "Image classification using random forests and ferns," in *Proc. IEEE ICCV*, Oct. 2007, pp. 1–8.

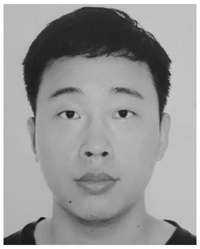
[64] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vis.*, vol. 42, no. 3, pp. 145–175, 2001.

Yabin Zhang received the B.E. degree in electronic information engineering from the Honors School, Harbin Institute of Technology, in 2013. He is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His research interests include video coding, image/video processing, image quality assessment, and computer vision.

Weisi Lin (M'92–SM'98–F'16) received the B.Sc. degree in electronics and the M.Sc. degree in digital signal processing from Sun Yat-Sen University, Guangzhou, China, in 1982 and 1985, respectively, and the Ph.D. degree in computer vision from King's College, London University, London, U.K., in 1993.

He was with Sun Yat-Sen University, Shantou University, Shantou, China, Bath University, Bath, U.K., the National University of Singapore, the Institute of Microelectronics, Singapore, and the Institute for Infocomm Research, Singapore. He has been the Project Leader of more than ten major successful projects in digital multimedia technology development. He was the Laboratory Head of Visual Processing and the Acting Department Manager of Media Processing, Institute for Infocomm Research. He is currently a Professor with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His research interests include image processing, perceptual signal modeling, video compression, multimedia communication, and computer vision.

Qiaohong Li received the B.E. and M.E. degrees from the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China, in 2009 and 2012, respectively, and the Ph.D. degree from the School of Computer Science and Engineering, Nanyang Technological University, Singapore, in 2017. She is currently a Research Fellow with the School of Electrical and Electronic Engineering, Nanyang Technological University. Her research interests include image quality assessment, speech quality assessment, computer vision, and visual perceptual modelling.



Wentao Cheng received the B.E. degree in computer science and engineering from the Harbin Institute of Technology in 2012. He is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His research interests include image-based localization and 3-D SfM point cloud simplification.

Xinfeng Zhang (M'16) received the B.S. degree in computer science from the Hebei University of Technology, Tianjin, China, in 2007, and the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2014.

He is currently a Research Fellow with Nanyang Technological University, Singapore. His research interests include image and video processing and image and video compression.