Objective Quality Assessment for Image Retargeting Based on Structural Similarity

Yuming Fang, Kai Zeng, Zhou Wang, *Fellow, IEEE*, Weisi Lin, *Senior Member, IEEE*, Zhijun Fang, and Chia-Wen Lin, *Senior Member, IEEE*

Abstract—We propose an objective quality assessment method for image retargeting. The key step in our approach is to generate a structural similarity (SSIM) map that indicates at each spatial location in the source image how the structural information is preserved in the retargeted image. A spatial pooling method employing both bottom-up and top-down visual saliency estimations is then applied to provide an overall evaluation of the retargeted image. To evaluate the performance of the proposed IR-SSIM algorithm, we created an image database that contains images produced by different retargeting algorithms and carried out subjective tests to assess the quality of the retargeted images. Our experimental results show that IR-SSIM is better correlated with subjective evaluations than existing methods in the literature. To further demonstrate the advantages and potential applications of IR-SSIM, we embed it into a multi-operator image retargeting process, which generates visually appealing retargeting results.

Index Terms—Image quality assessment, image retargeting, image saliency, structural similarity.

I. INTRODUCTION

WITH the fast spread of various mobile devices, maximizing the visual experiences of end users on small display screens has become ever important. Image and video retargeting techniques are essential to adapt the original visual content to different sizes and aspect ratios of the display screens. Traditional retargeting methods such as linear scaling and cropping often suffer from serious distortions or loss of significant information [1]–[3]. To overcome this problem, many content-aware image retargeting algorithms have been proposed [3]–[10], so as to preserve important image content with minimal distortions during the image resizing process. Nevertheless, how to objectively evaluate the quality of retargeted images remains an unresolved problem.

One popular image retargeting algorithm is named seam carving [1], where a seam is defined as eight-connected path of pixels (from top to bottom or left to right) that includes only one pixel in each row or column. Seam carving reduces the width (height) of an image by removing the unimportant seams. The seam carving approach has been extended for video retargeting [4], [5]. In [6], a saliency map, a facial map and a motion map are integrated to calculate a significance map indicating the importance of image pixels [6]. In mesh parametrization based retargeting algorithm [3], the significance map is composed of a saliency map, a facial map and a body estimation map, and an optimal user input map. An adaptive image retargeting algorithm based on Fourier analysis was proposed in [7]. In [8], a scale-and-stretch warping algorithm is developed by iteratively computing the optimal local scaling factors. Another important class of image retargeting algorithms are patch-based. In [9], a shift-map is employed to describe the relative geometric relationship between the reference and retargeted images. In [10], a randomized algorithm is proposed to find approximate nearest neighbor matches between image patches. In [11], patch-based method is extended to the compressed domain based on saliency maps computed from DCT coefficients. A recent user study suggested that multi-operator operation can obtain better retargeting results than those using only a single operator [12]. Promising results were produced by dynamically applying a sequence of operators including seam carving, scaling and cropping [12]. In [13], a multi-operator algorithm based on seam carving and image scaling was proposed that uses a bidirectional similarity function of image Euclidean distance, a dominate color descriptor similarity, and seam energy variation, to determine the number of seam carving operations [13].

Much less has been done to evaluate the quality of retargeted images quantitatively. Most existing studies of image retargeting were validated using small-scale subjective tests. The subjective evaluation method is expensive, time-consuming, and cannot be embedded into retargeting algorithms for optimization purposes. As such, objective quality assessment is highly desirable. In the past decades, there has been significant progress in the field of objective image quality assessment (IQA) [14]–[16]. Unfortunately, popular IQA algorithms, such as peak signal-to-noise-ratio, the structural similarity index (SSIM) [17] and the visual information fidelity index (VIF) [18], do not apply in retargeting applications because they require the sizes of the reference and distorted images to be the same. To the best of our knowledge, only a few IQA studies has been conducted on image retargeting [19]–[23]. A large
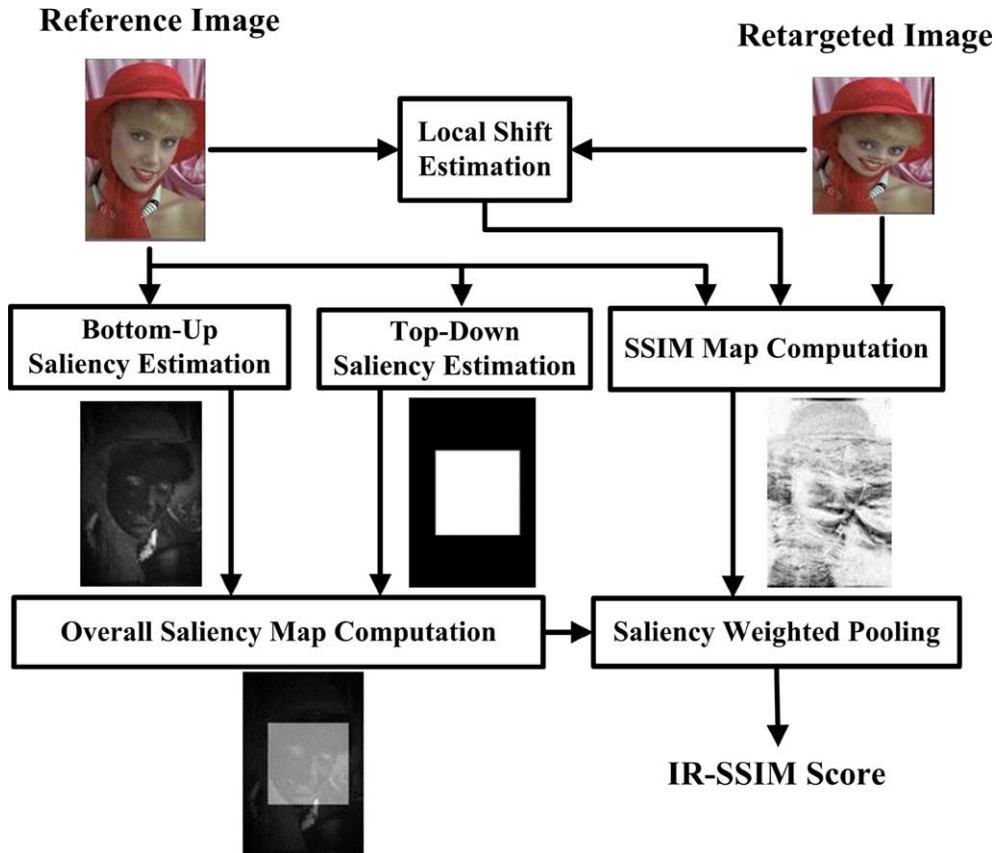
Fig. 1. Framework of the proposed approach.

scale user study was conducted using a paired comparison approach [19]. The study found that humans generally agree with each other upon the quality of retargeted images, and some retargeting algorithms are consistently more favorable than others. According to [19], some objective distance metrics, such as color layout descriptor (CL) [24], bidirectional warping (BDW) [12], and edge histogram (EH) [25], are useful in assessing the visual quality of retargeted images, though their correlations with subjective evaluations is not always consistent. In [20], it was shown that visual fixation maps obtained from eye tracking experiments can improve the visual quality assessment performance of objective metrics. Based on the eye tracking data, Chamaret *et al.* proposed an objective metric to assess the visual quality of video retargeting [21]. In [22], both global geometric structures and local pixel correspondence are established to evaluate the visual quality of the retargeted image. In [23], another user study was carried out, where subjects evaluated the quality of retargeted images on linear scales, rather than the paired comparison method used in [19].

In this work, we aim to develop an objective quality assessment algorithm for image retargeting (Section II). Our work is inspired by the success of the SSIM approach [17] in general-purpose IQA [15]. A critical step in our approach is to create an SSIM quality map that indicates at each spatial location of the reference image how the structural information is preserved in the retargeted image. This is appealing because an SSIM quality map not only can be pooled to produce an overall quality evaluation of the retargeted image, but also provides additional infor-

mation about how the structural details are preserved over the image space. Such information is useful because future image retargeting algorithms may be designed accordingly to better preserve local structural information. In the pooling stage of our IQA algorithm, we employ an automatic saliency prediction approach that combines both bottom-up and top-down mechanisms. A saliency map is thus generated and used as a spatially varying weighting factor of the SSIM map. The resulting image retargeting SSIM (IR-SSIM) algorithm demonstrates promising quality prediction performance when tested using subject-rated image retargeting databases (Section III). Its potentials are also demonstrated when it is embedded into a multi-operator image retargeting process, which produces visually appealing retargeting results (Section IV).

II. PROPOSED METHOD

The general framework of the proposed method is depicted in Fig. 1. For each image pixel p in the reference image, we first find its best matching pixel p' in the retargeted image. The SSIM measure is then calculated between the local regions surrounding p and p' in the reference and retargeted images, respectively. An SSIM map is generated by applying this process to all pixels in the reference image. Based on the reference image, a bottom-up saliency map is computed using data-driven, task-independent visual attention models, and a top-down saliency map is created based on task-dependent, high-level prior knowledge about the scene. The bottom-up and top-down saliency maps are combined to form an overall saliency map of the scene.

The final IR-SSIM score of the retargeted image is computed by weighting the SSIM map using the overall saliency map. In the current study, we only use a face detector in the calculation of the top-down saliency map. In the future, saliency detection models based on other high level features may also be integrated into the proposed framework. The general framework may be implemented in different ways, depending on the selection of the shift estimation, saliency estimation, and SSIM computation methods. Details about our current implementation are given in the following subsections.

A. SSIM Map Estimation

The reference and retargeted images have different sizes and pixel correspondences are missing, thus the original SSIM algorithm [17] cannot be applied directly. To overcome this problem, we would need to employ shift estimation methods to build dense pixel correspondence between the images. The scale-invariant feature transform (SIFT) descriptors [26] allow robust matching across different scenes. Analogous to the computational framework of optical flow, the SIFT flow algorithm matches SIFT descriptors instead of raw pixels between two images [27]. SIFT flow has been proven to be effective in finding dense correspondence between two images [27]. Here we use SIFT flow to find the matching pixel p' in the retargeted image for each pixel p in the original image. Given the SIFT images S_o and S_r computed from the original and retargeted images, respectively, the SIFT flow algorithm [27] attempts to find the pixel-match dense correspondence by minimizing the energy function.

Ideally, one would desire the local structural details in the reference image to be well preserved in the corresponding regions in the retargeted image. Once the pixel correspondence is established, the local structures in the corresponding regions can be compared. The SSIM index has been shown to be a simple but effective measure in predicting the perceptual structural distortions between image patches [15], [17]. Given two image patches $\mathbf{x}$ and $\mathbf{x}'$ centered at p and p' in the reference and retargeted images, respectively, the SSIM index between them is given by

$$\text{SSIM}(\mathbf{x}, \mathbf{x}') = \frac{(2\mu_{\mathbf{x}}\mu_{\mathbf{x}'} + C_1)(2\sigma_{\mathbf{x}\mathbf{x}'} + C_2)}{(\mu_{\mathbf{x}}^2 + \mu_{\mathbf{x}'}^2 + C_1)(\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{x}'}^2 + C_2)} \quad (1)$$

where $\mu_{\mathbf{x}}$, $\mu_{\mathbf{x}'}$, $\sigma_{\mathbf{x}}^2$, $\sigma_{\mathbf{x}'}^2$, and $\sigma_{\mathbf{x}\mathbf{x}'}$ denote the local mean of $\mathbf{x}$ and $\mathbf{x}'$, the local variance of $\mathbf{x}$ and $\mathbf{x}'$, and the local covariance between $\mathbf{x}$ and $\mathbf{x}'$, respectively. C_1 and C_2 are small positive stability constants that account for the saturation effects of the visual system at low luminance and contrast [17].

Applying the local SSIM computation using a sliding window approach at every pixel across the reference image space, we obtain an SSIM map that indicates how the structural details are preserved at each spatial location. Fig. 2 provides visual samples of two retargeted images created from the same reference image together with their SSIM maps. It appears that the SSIM maps reasonably estimate the perceived local structure preservations of the retargeted images. In particular, Fig. 2(d) is generated by directly cropping the center part of the reference image. Therefore, the structures at the center region are perfectly preserved (corresponding to the white center region in the SSIM
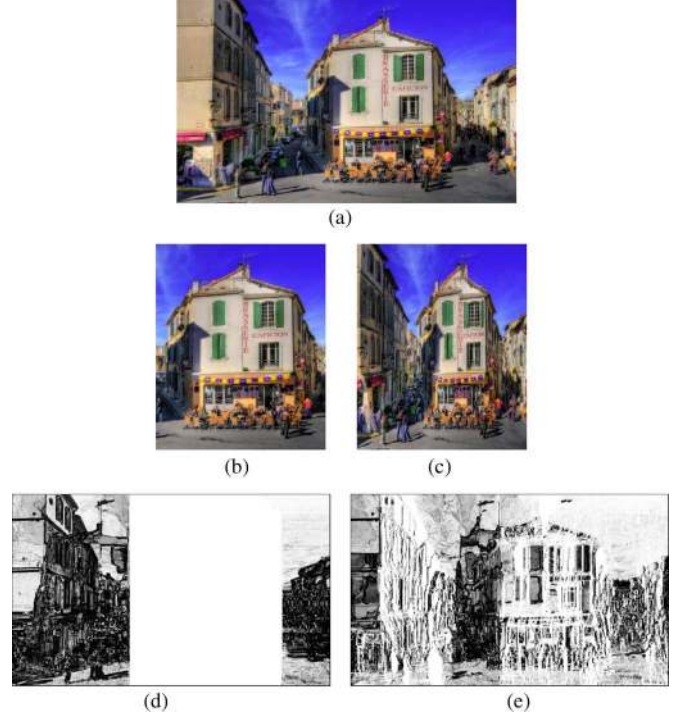

Fig. 2. Examples of SSIM maps. (a) Reference image. (b) Retargeted image by cropping (by 50% of the reference image). (c) Retargeted image by seam carving [4] (by 50% of the reference image). (d) SSIM map of the retargeted image in (b) (the structures of the center region are perfectly preserved, since the SSIM map is generated by directly cropping the center part of the reference image.). (e) SSIM map of the retargeted image in (c). In (d) and (e), brighter pixels indicate higher SSIM values.

map), and most distortions occur in the left and right parts of the image (corresponding to the dark regions in the SSIM map). Fig. 2(e) is obtained by seam carving, and the center regions are distorted more than Fig. 2(d), while more structural information at the two sides of the image is preserved. All the observations are well captured by the SSIM maps.

The basic SSIM computation is a single-scale approach, while the perceived image quality varies with the sampling density, the viewing distance, and the perceptual ability of the observer's visual system [28]. The multi-scale SSIM approach [28] provides a flexible framework to account for such variations. Following this framework, we iteratively apply low-pass filtering and downsample the reference and retargeted images by a factor of 2, and then compute the SSIM map at each scale. We index the original image scale as Scale 1, and the highest scale as Scale M. The overall SSIM value is computed by fusing the SSIM measures across all scales

$$\text{SSIM}(\mathbf{x}, \mathbf{x}') = \frac{1}{M} \sum_{j=1}^{M} w_j \text{SSIM}_j(\mathbf{x}, \mathbf{x}') \quad (2)$$

where SSIM_j is the SSIM measure at the jth scale, and w_j for $j \in \{1, 2, \ldots, M\}$ are weighting parameters determining the relative importance of different scales. Here, we up-sample the SSIM maps from different scales based on bilinear interpolation to match the size of the final SSIM map. As in [28], we set $M = 5$ and adopt the weighting factors w_j given in [28] that are obtained through psychovisual experiments.

Fig. 3. First row: reference image and its saliency map. Second to last rows: retargeted images (left), SSIM maps (middle), and product of SSIM and saliency maps (right). Retargeted images from the second to last rows are generated by MO, SCL, SC, SM, and WARP algorithms, respectively. IR-SSIM scores of the retargeted images from MO, SCL, SC, SM, and WARP are 0.920, 0.893, 0.859, 0.888, and 0.901, respectively.
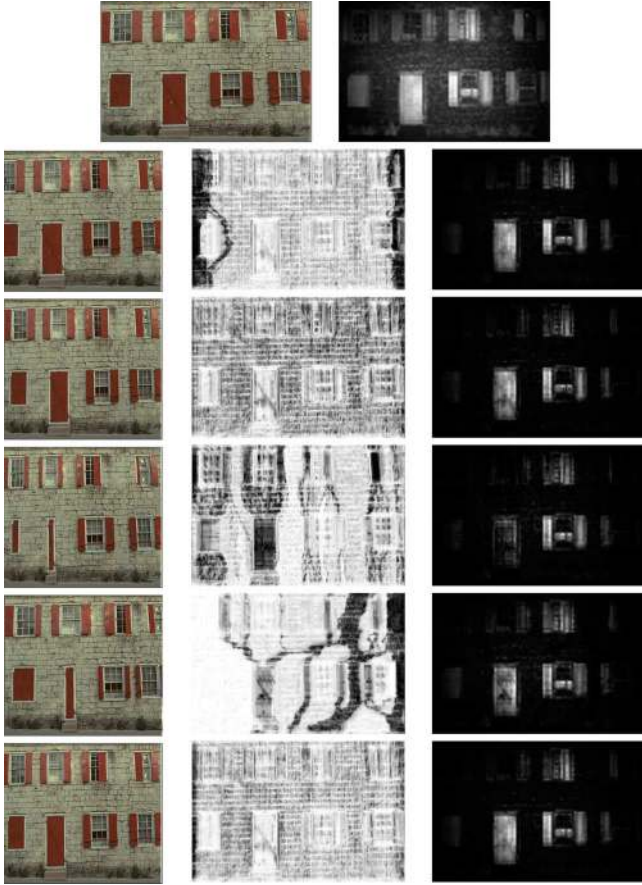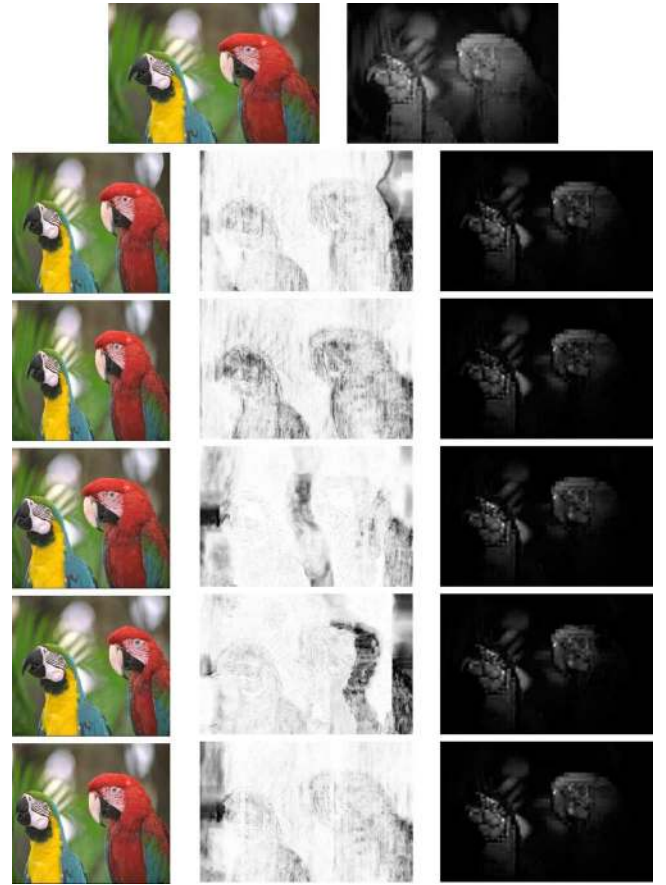

Fig. 4. First row: reference image and its saliency map. Second to last rows: retargeted images (left), SSIM maps (middle), and product of SSIM and saliency maps (right). Retargeted images from the second to last rows are generated by MO, SCL, SC, SM, and WARP algorithms, respectively. IR-SSIM scores of the retargeted images from MO, SCL, SC, SM and WARP are 0.957, 0.937, 0.964, 0.943, and 0.956, respectively.

B. Saliency Estimation

When observers look at a natural scene, the information captured by human eyes is much more than the visual information processing system can handle. As a result, the observers focus on the most important regions in the scene, characterized by the visual attention mechanisms of the human visual system (HVS) [29]–[31]. There are two different strategies of selective attention: the bottom-up strategy is a data-driven, task-independent process for automatic salient region selection [11], [32], while the top-down strategy is related to the prior knowledge or specific tasks to be performed [33]–[35]. Generally, observers would pay more attention on salient regions in an image. Here we perform both bottom-up and top-down estimations to create saliency maps, which are subsequently used as a weighting factor in predicting the overall quality of the retargeted image.

Previous studies have demonstrated that salient regions often pop out in natural scenes due to their differences in low-level features from the surrounding regions [31], [32]. As a result, they may be detected based on center-surround differences in low-level features, such as color, intensity, orientation, and motion [11], [32]. Here we adopt a modified version of the saliency detection model in [11] to extract the bottom-up saliency map.

Given an image, we first convert it into YCbCr color space and divide the image into 8×8 patches. Four features are extracted from the DCT coefficients of each patch, including one luminance feature L, two color features H_1 and H_2, and one texture feature T. The luminance feature is represented as the dc coefficient of the Y component, and the color features are the dc coefficients of the Cb and Cr components, respectively. The texture feature is calculated as the total ac energy of the Y component. Four feature maps are created by computing these features for all patches. The saliency of each patch is estimated by a contrast-of-feature approach. The saliency value S_i^k for Patch i based on the contrast from feature k is given by

$$S_i^k = \sum_{j \neq i} \left[\frac{1}{\sqrt{2\pi}\sigma_s} e^{-d_{ij}^2/2\sigma_s^2} \right] D_{ij}^k \qquad (3)$$

where $k \in \{L, H_1, H_2, T\}$, D_{ij}^k is the absolute feature difference between Patches i and j, d_{ij} is the spatial distance between Patches i and j, and σ_s is a width parameter of the Gaussian weighting function, which imposes more importance to the patches closer to the current Patch i. Here, the texture difference between image patches is calculated by the Euclidean distance rather than Hausdorff distance in [11] to reduce the computational complexity. For each given feature

Fig. 5. Left to right: reference image; retargeted image; bottom-up saliency map; overall saliency map (combining bottom-up and top-down saliency maps); SSIM map; point multiplication of SSIM and bottom-up saliency map; point multiplication of SSIM and overall saliency map.

TABLE I
PERFORMANCE COMPARISONS OF IR-SSIM WITH DIFFERENT
SALIENCY WEIGHTING OPTIONS

Models	KRCC			SRCC		
	Mean	STD	p-value	Mean	STD	p-value
Average Subject	0.641	0.101	3e-25	0.661	0.108	9e-25
IR-SSIM w/o SW	0.360	0.569	2e-4	0.375	0.636	1e-3
IR-SSIM w/ TP-SW	0.392	0.538	2e-4	0.407	0.612	1e-3
IR-SSIM w/ BU-SW	0.464	0.372	3e-5	0.471	0.442	5e-4
IR-SSIM	0.480	0.352	2e-5	0.511	0.426	2e-4

k, we normalize the saliency map S_i^k for all i to the dynamic range of $[0,1]$. The normalized values are denoted by $\tilde{S}_i^k$. The final bottom-up saliency map S_{BU} is obtained by averaging the saliency maps calculated by all features

$$S_i^{BU} = \frac{1}{K} \sum_{k=1}^{K} \tilde{S}_i^k \qquad (4)$$

where K denotes the number of features ($K = 4$).

Various factors could contribute to the top-down mechanisms of visual attention. Without additional knowledge about the scene and the visual task, it is difficult to determine a full list of top-down factors to be included in saliency modeling. Here, we pay special attention to human faces, which have been repeatedly shown to be a major factor that attract visual attention. Therefore, face detection based top-down saliency would be useful in general visual attention prediction tasks. Here, we adopt the Viola-Jones face detection algorithm [36], which has shown consistent performance in a variety of tests. By applying the algorithm on the reference image, a face map is created, which is subsequently normalized and employed as the top-down saliency map S^{TD} in our algorithm.

Finally, the overall saliency map S_i is obtained by averaging bottom-up and top-down saliency maps

$$S_i = \frac{\left(S_i^{BU} + S_i^{TD}\right)}{2}. \qquad (5)$$

C. Saliency Weighted Pooling

The SSIM map provides a useful local indicator about how the image structures perceived in the reference image is preserved in the retargeted image. Meanwhile, recent studies show that eye tracking data may be used to improve the performance of objective quality metrics for retargeted images [20]. It was found that even strong artifacts may be unnoticed in the locations outside the regions of interest [20]. This suggests that a saliency map that predicts the locations in the reference image

that are more likely to attract visual attention would be useful in assessing the quality of the retargeted image. Given the SSIM value SSIM_i and the saliency value S_i at each spatial location i, we define the final image retargeting SSIM index (IR-SSIM) based on saliency weighted pooling given by

$$\text{IR} - \text{SSIM} = \frac{\sum_i S_i \text{SSIM}_i}{\sum_i S_i}. \qquad (6)$$

Figs. 3 and 4 show visual examples of retargeted images produced by different retargeting algorithms, together with the saliency map of the reference image, the SSIM maps of the retargeted image, and the product of the saliency and SSIM maps. By close inspection of these images, we observe that different retargeting algorithms lead to substantially different structure preserving behaviors across the image space while saliency weighting help to focus on the regions of higher interests.

III. VALIDATION

We use two subject-rated image retargeting databases to validate the proposed IR-SSIM method. The first is built by ourselves, and the second is a widely used public benchmark [19].

Before conducting our subjective user study, we built a database of 25 source images with diverse resolutions and image content, including scenes with outstanding objects (such as human faces, cars, etc.), scenes without any outstanding object, scenes containing smooth or texture regions, and scenes with sharp features and edges. Five image retargeting algorithms are employed to create retargeted images. These include seam caving (SC) [4], scaling (SCL), warping (WP) [6], shift-map (SM) [9], and multi-operator (MO) [11] retargeting. Among these algorithms, SC is a popular and discrete retargeting algorithm; SCL is a traditional and continuous resizing algorithm; WP is a popular, content-aware, and continuous retargeting algorithm; SM is a patch-based and discrete resizing algorithm; and MO is a content-aware retargeting algorithm by combining two resizing operators of SC and SCL. The source images are resized in one dimension (either horizontally or vertically). To generate the retargeting results, the image sizes are reduced by 25%, a resizing factor widely used in existing studies [2], [11], [19].

As in [19], we carried out our subjective test using a paired comparison approach, where the reference image and two retargeted images created by two different retargeting algorithms were shown to the subject each time. The subject was asked to vote for the retareted image with relatively better visual quality by referring to the reference image. For each source image,

TABLE II
PERFORMANCE COMPARISON BETWEEN IR-SSIM USING SINGLE
SCALE SSIM AND MULTI-SCALE SSIM

Models	KRCC			SRCC		
	Mean	STD	p-value	Mean	STD	p-value
Single-scale IR-SSIM	0.368	0.684	4e-2	0.399	0.669	5e-2
IR-SSIM	0.480	0.352	2e-5	0.511	0.426	2e-4

TABLE III
PERFORMANCE COMPARISONS OF IMAGE RETARGETING QUALITY
ASSESSMENT MODELS ON OUR DATABASE

Models	KRCC			SRCC		
	Mean	STD	p-value	Mean	STD	p-value
EH	0.088	0.520	0.406	0.095	0.608	0.444
BDW	0.072	0.476	0.457	0.063	0.574	0.590
SIFT Flow	0.176	0.504	9e-2	0.195	0.577	0.104
CL	0.216	0.374	8e-3	0.284	0.485	7e-3
GLM	0.160	0.668	0.243	0.182	0.756	0.241
IR-SSIM w/o SW	0.360	0.569	2e-4	0.375	0.636	1e-3
IR-SSIM	0.480	0.352	2e-5	0.511	0.426	2e-4

TABLE IV
KRCC PERFORMANCE COMPARISONS OF IMAGE RETARGETING
QUALITY ASSESSMENT MODELS ON DATABASE [19]

Models	Mean	STD	p-value
BDS	0.083	0.268	0.017
BDW	0.046	0.181	0.867
EH	0.004	0.334	0.641
CL	-0.068	0.301	0.384
SIFT Flow	0.145	0.262	0.031
EMD	0.251	0.272	1e-5
GLM	0.106	0.553	0.447
IR-SSIM	0.363	0.271	1e-3

Fig. 6. First row: reference image and its saliency map. Second to last rows: retargeted images (left), SSIM maps (middle), and product of SSIM and saliency maps (right). Retargeted images from the second to last rows are generated by CR, MO, SC, SCL, SM, SNS, SV, and WARP algorithms, respectively.

given five retargeting algorithm, there are 10 possible combinations of paired comparisons, and thus the total number of paired comparisons for all 25 source images is 250. A total of 30 subjects, including 20 males and 10 females, were involved in the test. They all have normal vision and have no prior knowledge about the experiment. They voted for the retargeted image with relatively better visual quality by clicking a button on the screen.

For each reference image, the subjective rankings of the retargeted images are averaged and used to compare with the rankings given by objective measures. Kendall's rank-order correlation coefficient (KRCC) (or Kendall τ Distance [37]) and Spearman's rank-order correlation (SRCC) [38] are computed to measure the level of agreement between subjective and objective rankings. In both cases, higher correlation coefficient represents higher agreement, and both measures are upper-bounded by 1, which corresponds to perfect agreement. Before applying the test to evaluate objective quality assessment methods, we first compute how an average subject would perform in such as test. This is done by computing the KRCC and SRCC values between the average rankings of all subjects and the rankings given by any particular individual subject (out of the 30 subjects in total). When this is done for all 30 subjects, we compute the mean and standard deviation of the KRCC and SRCC values across all subjects. Such a quantitative measure of average subject performance gives a useful reference point on how an objective method behaves relative to a typical human subject. The average subject performance is given in Table I, where the average KRCC and SRCC values of 0.641 and 0.661 suggests that

although subjects generally agree with each other on the quality of retargeted images, unsurprisingly, there also exist significant variations between subject opinions.

Based on our subjective test results, we first test the significance of saliency weighting in the IR-SSIM measure. Specifically, we compare the performance between IR-SSIM without any saliency weighting (w/o SW), IR-SSIM with top-down saliency weighting only (w/TD-SW), IR-SSIM with bottom-up saliency weighting only (w/BU-SW), and full IR-SSIM. The results are given in Table I. It can be seen that both bottom-up and top-down saliency weighting are helpful
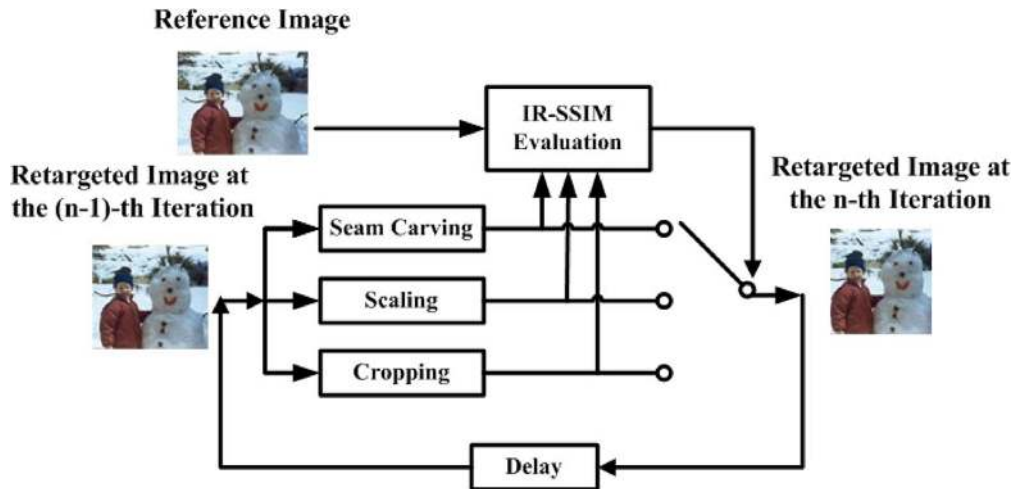
Fig. 7. Diagram of IR-SSIM directed iterative multi-operator image retargeting.

in improving the performance and the impact of bottom-up saliency weighting is more significant. This is not surprising because the top-down saliency maps are based on face detection and only a small proportion of the test images contain faces. The usefulness of top-down saliency estimation is demonstrated in Fig. 5, where the bottom-up saliency detection algorithm fails to emphasize on the face region. By contrast, the face detection based top-down saliency map allows us to focus on the face region that are of high interests to human observers. Overall, the best performance is obtained when both saliency weighting factors are included, as shown in Table I.

To demonstrate the advantage of the multi-scale SSIM approach in the proposed method. We compare single-scale IR-SSIM and the proposed multi-scle IR-SSIM. The experimental results are shown in Table II, where the KRCC and SRCC values show that the IR-SSIM using multi-scale SSIM can obtain better performance than that using single scale SSIM.

We compare IR-SSIM with five existing image retargeting quality assessment methods, including EH [25], BDW [12], SIFT flow [27], CL [24], and GLM [22].

In EH algorithm [25], the distortion of the retargeted image is computed based on the difference between edge distributions of the reference and retargeted images. BDW calculates the distortion of the retargeted image based on nonuniform alignment between the reference and retargeted images subject to order constraints [12]. SIFT Flow measures the distortion of the retargeted image by the distance between the reference and retargeted images, which is represented as the energy cost from pixel-matching correspondence between the reference and retargeted images [27]. In CL algorithm [24], the distortion of the retargeted image is calculated by the difference between the color distributions of the reference and retargeted images. GLM calculates the distortion of the retargeted image based on the difference of global geometric structures (from local pixel correspondence) between the reference and retargeted images [22]. Among these methods, BDW and GLM are the quality assessment methods designed specifically for retargeted images; SIFT flow is the pixel-matching algorithm used in this paper;

BDW is a bidirectional similarity measure from mid-level semantic correspondence between images; while EH and CL are low-level similarity measures which treat the image as a whole [19]. For fair comparison, we use the same saliency map created in Section II-C in the implementation of the GLM model [22]. The results are shown in Table III, where IR-SSIM obtains better performance than existing metrics, even without saliency weighting. The reason may be that IR-SSIM directly evaluates local quality within each image patch based on SSIM, while existing methods often extract local features but leave quality analysis at global level. The local SSIM measure not only supplies a useful spatial quality map, but also makes it naturally integrated with saliency weighting or other spatial pooling approaches.

From the Table III, the proposed IR-SSIM with saliency consideration outperforms all other models by a sizable margin in terms of mean correlations with subjective rankings. Its lower standard deviations of KRCC and SRCC values suggest that IR-SSIM performs more consistently than other models across different reference images. Visual examples are given in Figs. 3 and 4. Despite the promising results, significant gap still exists between the performance of IR-SSIM and that of an average subject given in Table I. Such a comparison is helpful for us to better understand the current status of state-of-the-art image retargeting quality assessment techniques, and gives a sensible measure on the space for future improvement.

To further validate the proposed method, we test and compare image retargeting algorithms using a public benchmark image retargeting database [19]. The benchmark was collected by a user study from 210 participants using a database containing 37 source reference images. Each reference image was resized by eight image retargeting algorithms, which include scaling, cropping, SV [41], MO [12], SC [4], SM [9], SNS [8], and WARP [6]. In addition, the subjective evaluation data, together with the correlations of six objective quality assessment models (BDS [39], BDW [12], EH [25], CL [24], SIFT Flow [27], and EMD [40]) are also provided [19]. The performance comparison results are shown in Table IV, where we computed the results for IR-SSIM, and the results for all other models are extracted from [19] directly. Again, IR-SSIM outperforms all other methods in
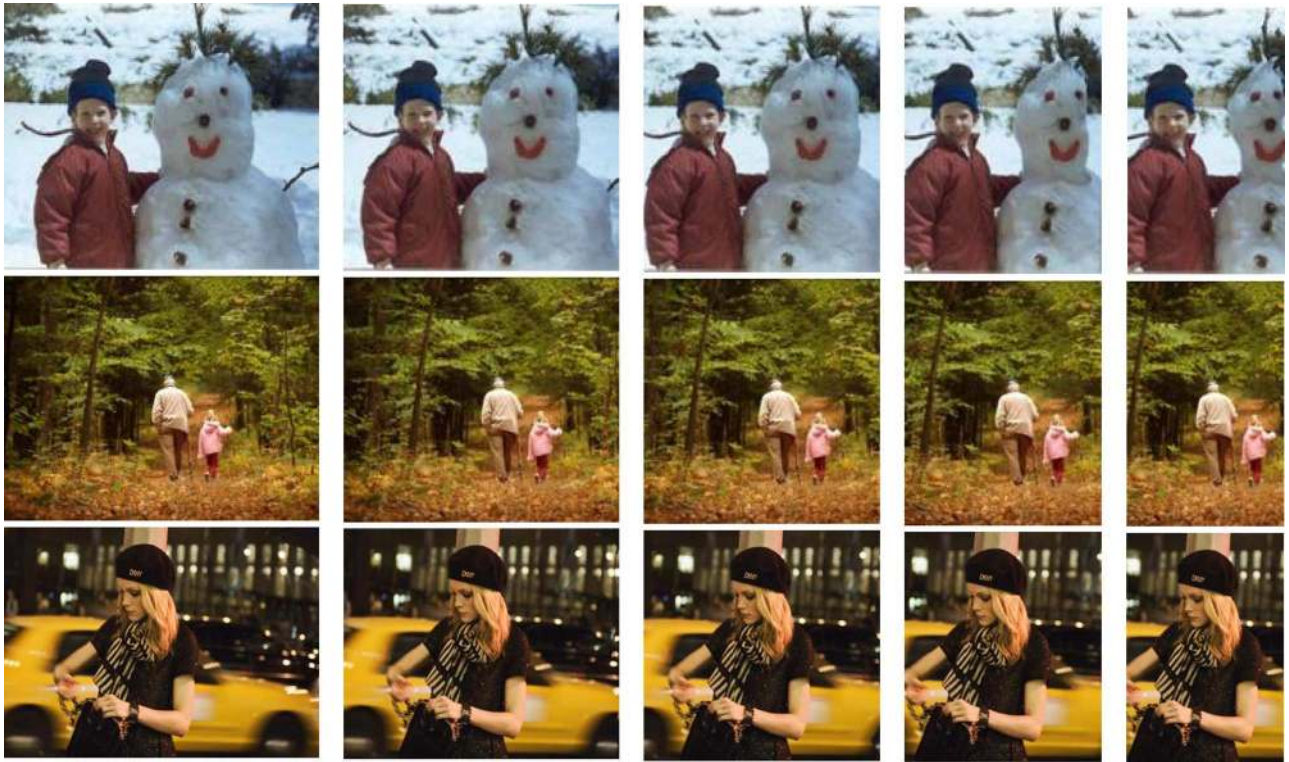
Fig. 8. Sample images created by IR-SSIM directed multi-operator retargeting. Column 1: reference images. Columns 2–5: retargeted images for resizing factors of 12.5%, 25%, 37.5%, and 50%, respectively.

this benchmark in terms of higher mean KRCC values and lower standard deviation across different reference images. More visual examples are given in Fig. 6, where it can be observed that different image retargeting algorithms preserve the structural information in different ways, leading to substantially different SSIM maps.

IV. APPLICATION: IR-SSIM DIRECTED MULTI-OPERATOR RETARGETING

The value of IQA models is far beyond assessing and comparing images or image processing algorithms. A broader potential field of applications is to apply them in the design and optimization of image processing algorithms and systems [15]. Here we demonstrate the value of IR-SSIM by embedding it into a multi-operator image retargeting algorithm.

Previous studies have demonstrated that multi-operator image retargeting algorithms often achieve better performance in image resizing than single-operator methods [11], [12], [19]. Multi-operator algorithms iteratively apply a sequence of resizing operations dynamically selected from several candidate resizing operators, which often include scaling, cropping, and seam carving. The most critical step is to determine the optimal image resizing operator sequence for image resizing. For this purpose, a quality/distortion measure is desired to assess the similarity/dissimilarity between the reference and retargeted images (generated by different image resizing operator sequences), and to decide on which image resizing operator sequence creates the best result. In [12], the BDW metric is playing such a role, but according to our results reported in Section III, BDW is not among the best options in predicting

perceived quality of retargeted images. Naturally, one would be curious to see the retargeted image obtained by replacing the role of BDW with IR-SSIM.

We develop a modified MO method [12] that applies an iterative method to decide on the sequence of retargeting operators based on IR-SSIM. The diagram of the algorithm is shown in Fig. 7, which searches for the best operator at each iteration rather than in the entire multi-operator space as in [12], and thus the computational cost is significantly lower. The whole image retargeting task is divided into subtasks, each associated with one iteration with only small changes in image size. In the nth iteration, IR-SSIM between the reference and retargeted images is employed to pick the best quality image among those created by seam carving, scaling and cropping operators applied to the input image generated form the last iteration (in the first iteration, the original reference image is used as the input image). The iteration continues until the target resizing factor is achieved.

Fig. 8 provides examples that demonstrate the iterative process, where the reference images are gradually resized until the target size is reached, generating a series of retargeted images as mid-results. In Fig. 9, IR-SSIM directed multi-operator retargeting is compared with the original BDW based multi-operator method [12]. Both subjective and objective IR-SSIM measures prefer the retargeted images produced by the proposed approach. In the experiment, we also find that IR-SSIM directed multi-operator retargeting algorithm leads to cropping being chosen more frequently during the image resizing process compared with BDW based method. The results from 37 images in [19] show that the cropping operator

Fig. 9. Left: reference images; middle: retargeted images by multi-operator retargeting [12]; right: retargeted images by multi-operator retargeting directed by IR-SSIM. The number below each retargeted image is the IR-SSIM value computed from the reference and retargeted images.

occupies about 73% of the total resizing operation in IR-SSIM, as opposed to 40% in BDW, based multi-operator algorithms. This might be due to the fact that cropping preserves the structural information better than scaling or seam carving, and often leads to better perceptual preference, especially when the cropped area coincides with visual saliency. This conjecture is supported by the subjective experimental results in [19], where cropping operator typically achieves better performance than scaling and seam carving operators.

V. CONCLUSION

We propose an IR-SSIM method for image retargeting quality assessment by combining a series of ideas including SIFT flow based local shift estimation, SSIM based local quality assessment, and saliency based perceptual pooling. Promising quality prediction performance is achieved when testing the proposed IR-SSIM algorithm using two subject-rated image retargeting databases. To further demonstrate the advantages and potentials of IR-SSIM, we embed it into a multi-operator image retargeting process, resulting in visually appealing retargeted images.

Besides the good correlation with subjective evaluations of image quality, a perhaps more interesting and more useful output of the IR-SSIM approach is the SSIM map that indicates the strength of structural information preservation at each spatial location in the reference image. This is useful not only in understanding how the image retargeting algorithms work, but also in the design and optimization of novel image regargeting algorithms. This would be an interesting and promising future research direction to be explored.

REFERENCES

[1] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," *ACM Trans. Graph.*, vol. 26, no. 3, 2007, 10.

[2] D. Vaquero, M. Turk, K. Pulli, M. Tico, and N. Gelfand, "A survey of image retargeting techniques," in *SPIE Appl. Digital Image Process. XXXIII*, 2010, vol. 779814.

[3] Y. Guo, F. Liu, J. Shi, Z.-H. Zhou, and M. Gleicher, "Image retargeting using mesh parameterization," *IEEE Trans. Multimedia*, vol. 11, no. 5, pp. 856–867, Aug. 2009.

[4] M. Rubinstein, A. Shamir, and S. Avidan, "Improved seam carving for video retargeting," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 1–9, 2008.

[5] M. Grundmann, V. Kwatra, M. Han, and I. Essa, "Discontinuous seam-carving for video retargeting," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 569–576.

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

[7] J.-S. Kim, J.-H. Kim, and C.-S. Kim, "Adaptive image and video retargeting technique based on Fourier analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1730–1737.

[8] Y.-S. Wang, C.-L. Tai, O. Sorkine, and T.-Y. Lee, "Optimized scale-and-stretch for image resizing," *ACM Trans. Graph. (Proc. SIGGRAPH Asia)*, vol. 27, no. 5, 2008, 118.

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

[10] C. Barnes, E. Shechtman, A. Finkelstein, and D. Goldman, "Patch-Match: A randomized correspondence algorithm for structural image editing," *ACM Trans. Graph. (Proc. SIGGRAPH)*, vol. 28, no. 3, 2009, 24.

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

[12] M. Rubinstein, A. Shamir, and S. Avidan, "Multi-operator media retargeting," *ACM Trans. Graph.*, vol. 28, no. 3, pp. 1–11, 2009.

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

[14] Z. Wang and A. C. Bovik, *Model Image Quality Assessment*, ser. Syntheses Lectures Image, Video Multimedia Processing. San Rafael, CA: Morgan Claypool, 2006.

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

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

[17] 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.

[18] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

[19] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir, "A comparative study of image retargeting," *Proc. ACM SIGGRAPH Asia*, vol. 29, no. 6, pp. 1–9, 2010.

[20] S. Castillo, T. Judd, and D. Gutierrez, "Using eye-tracking to assess different image retarget-ing methods," in *ACM SIGGRAPH Symp. Appl. Percept. Graph. Visualizat.*, Toulouse, France, 2011, pp. 7–14.

[21] C. Chamaret, O. L. Meur, P. Guillotel, and J.-C. Chevet, "How to measure a relevance of a retargeting approach," in *Eur. Conf. Comput. Vis. Workshop Media Retarget.*, Crete, Greece, 2011, pp. 156–168.

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

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

[24] E. Kasutani and A. Yamada, "The mpeg-7 color layout descriptor," in *Proc. IEEE Int. Conf. Image Process.*, 2001, pp. 674–677.

[25] 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.

[26] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE Int. Conf. Comput. Vis.*, 1999, vol. 2, pp. 1150–1157.

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

[28] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *IEEE Asilomar Conf. Signals, Syst. Comput.*, 2003, vol. 2, pp. 1398–1402.

[29] K. Koch, J. McLean, R. Segev, M. A. Freed, M. J. Berry, V. Balasubramanian, and P. Sterling, "How much the eye tells the brain," *Current Biol.*, vol. 16, no. 14, pp. 1428–1434, 2006.

[30] H. Pashler, *The Psychology of Attention*. Cambridge, MA: MIT Press, 1997.

[31] A. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognit. Psychol.*, vol. 12, no. 1, pp. 97–136, 1980.

[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] Z. Lu, W. Lin, X. Yang, E. Ong, and S. Yao, "Modeling visual attentions modulatory aftereffects on visual sensitivity and quality evaluation," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1928–1942, Nov. 2005.

[34] Y. Fang, W. Lin, C. T. Lau, and B.-S. Lee, "A visual attention model combining top-down and bottom-up mechanisms for salient object detection," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2011, pp. 1293–1296.

[35] A. Torralba, A. Oliva, M. S. Castelhano, and J. M. Henderson, "Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search," *Psychol. Rev.*, vol. 113, no. 4, pp. 766–786, 2006.

[36] P. A. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.

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

[38] L. J. Myers and A. D. Well, *Research Design and Statistical Analysis*. Mahwah, NJ: Lawrence Erlbaum, 2003.

[39] D. Simakov, Y. Caspi, E. Shechtman, and M. Irani, "Summarizing visual data using bidirectional similarity," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[40] O. Pele and M. Werman, "Fast and robust earth movers distances," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2009, pp. 460–467.

[41] P. Krahenbuhl, M. Lang, A. Hornung, and M. Gross, "A system for retargeting of streaming video," *ACM Trans. Graph.*, vol. 28, no. 5, 2009, 126.
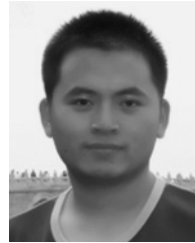
Kai Zeng received the M.A.Sc. degree in electrical engineering from Xidian University, Xi'an, China, in 2009, and the Ph.D. degree in electrical and computer engineering, in 2013, from the University of Waterloo, Waterloo, ON, Canada, where he is currently a Post-Doctoral Fellow in the Department of Electrical and Computer Engineering.

His research interests include computational video and pattern analysis, multimedia communications, and image and video processing (coding, denoising, analysis, and representation), with emphasis on image and video quality assessment and their applications.

Dr. Zeng was a recipient of IEEE Signal Processing Society student travel grant at the 2010 and 2012 IEEE International Conference on Image Processing.
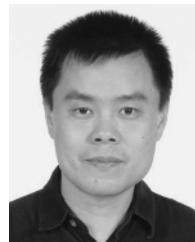
Zhou Wang (S'99–M'02–SM'12–F'14) received the Ph.D. degree in electrical and computer engineering from The University of Texas at Austin, Austin, TX, USA, in 2001.

He is currently an Associate Professor in the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. His research interests include image processing, coding, and quality assessment; computational vision and pattern analysis; multimedia communications; and biomedical signal processing. He has more than 100 publications in these fields with over 17 000 citations (Google Scholar). He has been serving as an Associate Editor of *Pattern Recognition* (2006–present). He has served as a Guest Editor of *EURASIP Journal of Image and Video Processing* (2009–2010) and *Signal, Image and Video Processing* (2011–2013).

Dr. Wang is a member of IEEE Multimedia Signal Processing Technical Committee (2013–2015). He had served or has been serving as an Associate Editor of IEEE Transactions on Image Processing (2009–2013) and IEEE Signal Processing Letters (2006–2010), and a Guest Editor of *IEEE Journal of Selected Topics in Signal Processing* (2013–present and 2007–2009). He is a recipient of 2014 NSERC E.W.R. Steacie Memorial Fellowship Award, 2013 IEEE Signal Processing Best Magazine Paper Award, 2009 IEEE Signal Processing Society Best Paper Award, 2009 Ontario Early Researcher Award, and ICIP 2008 IBM Best Student Paper Award (as senior author). He is a registered Professional Engineer in Ontario, Canada.

Yuming Fang received the B.E. degree from Sichuan University, Sichuan, China, the M.S. degree from Beijing University of Technology, Beijing, China, and the Ph.D. degree in computer engineering from Nanyang Technological University, Singapore, in February 2013.

He is currently a lecturer in the School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, China. From October 2011 to January 2012, he was a visiting Ph.D. student in National Tsinghua University, Taiwan. From September 2012 to December 2012, he was a visiting scholar in University of Waterloo, Canada. He was also a (visiting) Postdoc Research Fellow in IRCCyN lab, Poly-Tech' Nantes and University Nantes, Nantes, France, University of Waterloo, Waterloo, Canada, and Nanyang Technological University, Singapore. His research interests include visual attention modeling, visual quality assessment, image retargeting, computer vision, 3-D image/video processing, etc.

Dr. Fang was a Secretary for HHME 2013 and a Special Session Organizer for the IEEE Conference on Visual Communications and Image Processing (VCIP 2013).

Weisi Lin (M'92–SM'98) received the Ph.D. degree from King's College London, London, U.K.

He was the Lab Head and Acting Department Manager for Media Processing, in Institute for Infocomm Research, Singapore. Currently, he is the Associate Chair (Graduate Studies) in Computer Engineering, Nanyang Technological University, Singapore. His research areas include image processing, perceptual multimedia modeling and evaluation, and video compression. He published over 250 refereed papers in international journals and conferences. He is on the editorial board of the *Journal of Visual Communication and Image Representation*.

He is on the editorial boards of IEEE Transactions on Multimedia and IEEE Signal Processing Letters. He chairs the IEEE MMTC IG on Quality-of-Experience. He has been elected as an APSIPA Distinguished Lecturer (2012/3). He is the Lead Technical-Program Chair for Pacific-Rim Conference on Multimedia (PCM) 2012 and International Workshop on Quality of Multimedia Experience (QoMEX) 2014, and a Technical-Program Chair for IEEE International Conference on Multimedia and Expo (ICME) 2013. He is a Fellow of Institution of Engineering Technology, and an Honorary Fellow, Singapore Institute of Engineering Technologists.

Zhijun Fang received the Ph.D. degree from Shanghai Jiaotong University, Shanghai, China.

He is currently a Professor and Dean in the School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, Jiangxi, China. His current research interests include image processing, video coding, and pattern recognition.

Dr. Fang was a General Chair of HHME 2013. He received the awards of Jiangxi Provincial "the Gan Po Elite 555 Plan" and "One-hundred, One-thousand, Ten-thousand Talent Project."

Chia-Wen Lin (S'94–M'00–SM'04) received the Ph.D. degree in electrical engineering from National Tsing Hua University, Hsinchu, Taiwan, in 2000.

He is currently an Associate Professor with the Department of Electrical Engineering and the Institute of Communications Engineering, National Tsing Hua University, Hsinchu, Taiwan. He was with the Department of Computer Science and Information Engineering, National Chung Cheng University, Taiwan, during 2000–2007. Prior to joining academia, he worked for the Information and Communications Research Laboratories, Industrial Technology Research Institute, Hsinchu, Taiwan, during 1992–2000. His research interests include image and video processing and video networking. He is an Associate Editor of the *Journal of Visual Communication and Image Representation*. He is an Area Editor of *EURASIP Signal Processing: Image Communication*.

Dr. Lin is an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON MULTIMEDIA, and the IEEE MULTIMEDIA MAGAZINE. He is currently Chair of the Multimedia Systems and Applications Technical Committee of the IEEE Circuits and Systems Society. He served as Technical Program Co-Chair of the IEEE International Conference on Multimedia & Expo (ICME) in 2010, and Special Session Co-Chair of the IEEE ICME in 2009. His paper won the Young Investigator Award presented by VCIP 2005. He received the Young Faculty Awards presented by CCU in 2005 and the Young Investigator Awards presented by National Science Council, Taiwan, in 2006.