Saliency Detection in the Compressed Domain for Adaptive Image Retargeting

Yuming Fang, Zhenzhong Chen, *Member, IEEE*, Weisi Lin, *Senior Member, IEEE*, and
Chia-Wen Lin, *Senior Member, IEEE*

Abstract—Saliency detection plays important roles in many image processing applications, such as regions of interest extraction and image resizing. Existing saliency detection models are built in the uncompressed domain. Since most images over Internet are typically stored in the compressed domain such as joint photographic experts group (JPEG), we propose a novel saliency detection model in the compressed domain in this paper. The intensity, color, and texture features of the image are extracted from discrete cosine transform (DCT) coefficients in the JPEG bit-stream. Saliency value of each DCT block is obtained based on the Hausdorff distance calculation and feature map fusion. Based on the proposed saliency detection model, we further design an adaptive image retargeting algorithm in the compressed domain. The proposed image retargeting algorithm utilizes multioperator operation comprised of the block-based seam carving and the image scaling to resize images. A new definition of texture homogeneity is given to determine the amount of removal block-based seams. Thanks to the directly derived accurate saliency information from the compressed domain, the proposed image retargeting algorithm effectively preserves the visually important regions for images, efficiently removes the less crucial regions, and therefore significantly outperforms the relevant state-of-the-art algorithms, as demonstrated with the in-depth analysis in the extensive experiments.

Index Terms—Compressed domain, image retargeting, joint photographic experts group (JPEG), saliency detection, texture homogeneity.

I. INTRODUCTION

$\mathbf{V}$ISUAL attention is an important mechanism to process visual information in the human visual system (HVS). It is a cognitive process for selecting the significantly visual information and filtering out other redundant visual information for natural scenes. When observers look at a natural scene, they will focus on the salient regions while ignoring other non-salient regions. There are two approaches in visual attention mechanism: bottom–up and top–down approaches. Bottom–up approach, which is stimuli-driven and task-independent, is a perception process for automatic saliency detection. On the contrary, top–down approach is a recognition process influenced by the prior knowledge.

As to the bottom–up approach, studies have shown that the salient regions in the visual field would first pop out for their different low features from their surrounding [1], [2]. Based on the feature-integration theory [1], many saliency detection models have been proposed to extract the salient regions for various image processing applications [3]–[11]. One popular application is the image retargeting [12]–[16]. In some image retargeting algorithms [12]–[16], the saliency map is used to measure the visual importance of image pixels.

Existing saliency detection algorithms [3]–[10] and image retargeting algorithms [12]–[15], [17]–[22] are all implemented in the uncompressed domain. However, most images over Internet are typically stored in the compressed domain of joint photographic experts group (JPEG) [23]. The compressed JPEG images are widely used in various Internet-based applications, since they reduce the storage space and increase the downloading speed. In order to extract features from the compressed JPEG images, the existing saliency detection or image retargeting algorithms have to decompress these JPEG images from the compressed domain into the spatial domain. The full decompression from these saliency detection or image retargeting algorithms is not only computation-consuming but time-consuming as well. Compared with the existing saliency detection and image retargeting algorithms which operate in the uncompressed domain, it is crucial to design the efficient saliency detection and image retargeting algorithms in the compressed domain.

In this paper, we first propose a saliency detection model in the compressed domain. Furthermore, based on the proposed saliency detection model, we design an adaptive image retargeting algorithm in the compressed domain. As discrete cosine transform (DCT) is used in JPEG compression at 8×8 block level, the DCT coefficients are used to extract intensity, color, and texture features for each 8×8 block for saliency detection. Previous studies successfully extract various features by using DCT coefficients for various image processing applications [24]–[26]. In [24], the color and texture features are directly extracted from DCT coefficients for object localization. In [25], DCT coefficients in each macroblock are used for object indexing. In [26], the color and texture features are extracted from the DCT coefficients to measure the image similarity by using the statistical graph matching method. These studies have shown that the direct feature extraction from the compressed domain can obtain useful information
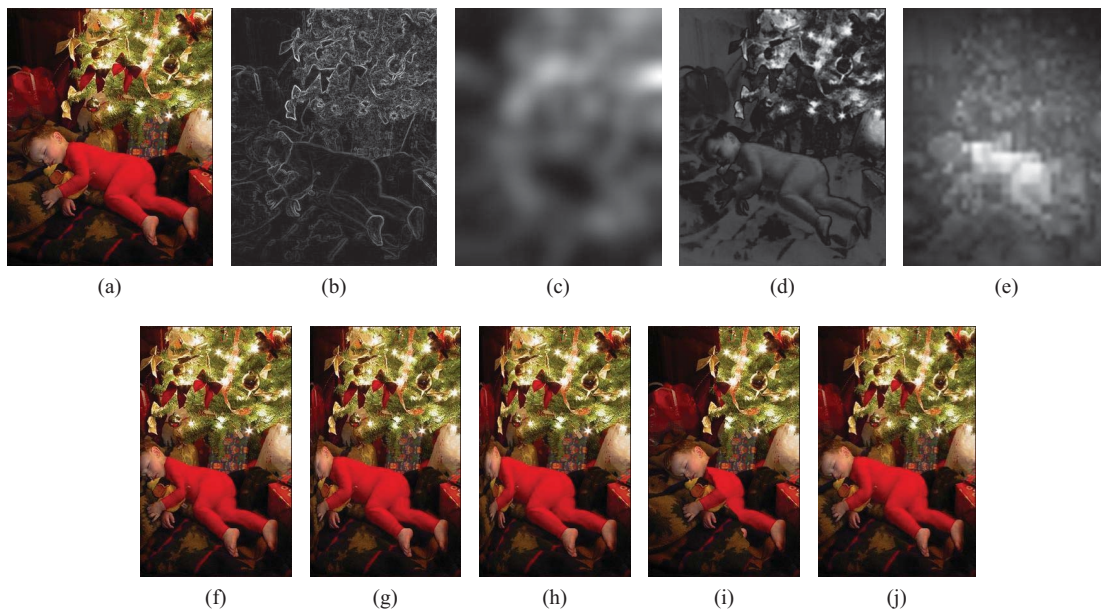
Fig. 1. Comparison of the different image retargeting algorithms (a) original image, (b) gradient map (used in [13] and [18]), (c) saliency map from Itti's model [3] (used in [14]), (d) saliency map from the model in [6] (used in [12]), (e) saliency map from our proposed model, and (f)–(j) retargeted images from [12]–[14], [18] and our proposed algorithm, respectively. The width of the retargeted images is 75% of that from the original image.

in many applications. Here, we directly extract the intensity, color, and texture features from the DCT coefficients for saliency detection for JPEG images. Although the minimum coded unit (MCU) blocks can be as large as 16×16 (for 4:2:0 component subsampling format) [23], [27], we perform our saliency detection and retargeting at the 8×8 block level for each DCT block. After obtaining the intensity, color, and texture features from the compressed domain, the saliency map for images is calculated based on weighted feature differences between DCT blocks. In addition, we design an adaptive image retargeting algorithm in the compressed domain based on the proposed saliency detection model. The saliency map in the compressed domain is used to determine the visual importance of each 8×8 block in JPEG images. The multioperators including the block-based seam carving and the image scaling are utilized to perform image resizing. The number of the removal block-based seams is determined by the texture homogeneity of images. Experimental results show the superior performance of the proposed saliency detection model as well as the proposed image retargeting algorithm.

The rest of this paper is organized as follows. In Section II, we first introduce the relevant existing saliency detection and image retargeting algorithms, respectively. Then we demonstrate the contributions of our work. In Section III, we describe the details of the proposed saliency detection and image retargeting algorithms. Section IV shows the experimental results by comparing the proposed algorithms with other existing ones. The final section concludes this paper.

II. RELATED WORK AND CONTRIBUTIONS OF OUR WORK

A. Saliency Detection

The classical saliency detection model proposed by Itti *et al.* is designed based on the neuronal architecture of the primates' early visual system [3]. In this paper, the saliency map is calculated according to the multiscale center-surrounding differences for three features: intensity, color, and orientation. The linear combination between these three feature maps is utilized to obtain the final saliency map. Based on Itti's model, Harel *et al.* proposed the graph-based visual saliency model by using a graph-based dissimilarity measure [9]. Ma *et al.* devised a saliency detection model based on the local contrast analysis [10]. In this paper, a fuzzy growing algorithm is adapted to extract the salient regions for images. In [16], Goferman *et al.* built a content-aware saliency detection with the consideration of the contrast from both local and global perspectives. Liu *et al.* utilized the machine learning technique to obtain the saliency map for images [4]. In this paper, they used a set of features, including multiscale contrast, center–surrounding histogram, and color spatial distribution to obtain the salient objects locally, regionally, and globally [4].

Hou *et al.* used a concept called spectral residual to build a saliency detection model in [5]. The authors claimed that the saliency map was obtained based on the log spectra representation for images. Later, Guo *et al.* found that Hou's model was actually caused by the phase spectrum and they devised a phase-based saliency detection algorithm [7]. This algorithm obtains the final saliency map by using inverse Fourier transform on a constant amplitude spectrum and the original phase spectrum of the input image. In [28], Bruce *et al.* decided saliency information based on the concept of information maximization. The saliency map is obtained according to the Shannon's self-information measure. In addition, this saliency detection algorithm is implemented in a neural circuit, which is explained as similar with the circuitry existent in the primate visual cortex [28]. Le Meur *et al.* proposed a visual attention model based on the understanding of the HVS behavior [29]. The contrast sensitivity functions, perceptual decomposition, visual masking, and center–surround interactions are utilized for detecting salient regions in this paper [29].

All these saliency detection models mentioned above are implemented in the uncompressed domain. In this paper, we propose a novel saliency detection model in the compressed domain. We extract the intensity, color, and texture features from DCT coefficients in the JPEG bit-stream to calculate the block differences for saliency detection. The saliency map obtained in the compressed domain is used to determine the visual importance of each DCT block in the proposed image retargeting algorithm. Thanks to the directly derived saliency map from the compressed domain, the proposed image retargeting algorithm effectively preserves the objects of attention and remove the less crucial regions, as shown in Fig. 1. From Fig. 1, we can see that our saliency map identifies the salient regions more accurately than the gradient map and the saliency maps from [3] and [6]. Therefore, the retargeted image from our proposed algorithm is better than these from other algorithms, as shown in Fig. 1(f)–(j). More details and comparisons will be provided in the following sections.

B. Image Retargeting

One traditional image resizing method is to scale images by downsampling. The problem with the image scaling is that it will result in worse viewing experience and loss of some detailed information as the salient objects turn to be smaller. Image cropping is an alternative solution, which preserves regions of interest in images by discarding other noninterest regions. The defect of this technique is that the context information in images will be lost [15], [17], [30], [31]. To overcome the limitations of the image scaling and cropping, many effective image retargeting algorithms [12]–[15], [17]–[22] have been proposed. In these algorithms, the content awareness is taken into consideration and the visual significance map is designed for measuring the visual importance of each pixel for image resizing operation. The visual significance maps used in these algorithms are generally comprised of the gradient map, the saliency map, and some high-level feature maps, such as facial map, motion map, and so on [12]–[15], [17]–[22].

Avidan *et al.* proposed the popular image retargeting algorithm named seam carving [17]. A seam is defined as eight-connected path of low-energy pixels (from top to bottom or left to right) in images. These pixels include only one pixel in each row or column. The seam carving aims to reduce the width (or height) by removing those unimportant seams. A gradient map is used to determine the importance of each pixel in images. Later, Rubinstein *et al.* extended this algorithm to video retargeting by introducing the forward energy method [18]. Some similar algorithms were also designed based on seam carving [12], [20].

Other advanced image retargeting algorithms have also been proposed. Wolf *et al.* introduced a video retargeting algorithm through introducing a linear system to determine the new pixel position [13]. In this paper, the visual importance of each image pixel is measured by the visual importance map comprised of local saliency detection, face detection, and motion detection. Ren *et al.* proposed an image retargeting algorithm based on global energy optimization, in which the saliency map and face detection are combined to determine

the visual importance of each image pixel [14]. Jin *et al.* presented a content-aware image resizing algorithm through warping a triangular mesh over images by regarding salient line features and curved features as important regions [19]. Guo *et al.* advanced an image retargeting algorithm through utilizing saliency-based mesh parametrization [15].

Recently, Rubinstein *et al.* conducted a user study and found that applying multioperators (such as seam carving, cropping, and so on) can obtain better results than those from only single operator in image retargeting [21]. In this paper, the authors proposed a multioperator media retargeting algorithm that combines the seam carving, scaling, and cropping operators to resize images. The size amount for each operation is determined by optimal result for maximizing the similarity between the input image and the retargeted image. In [22], Dong *et al.* introduced a image retargeting algorithm by combining seam carving and scaling. The authors utilized a bidirectional similarity function of image Euclidean distance, a dominant color descriptor similarity, and seam energy variation to determine the best number of seam carving operation.

All these image retargeting algorithms introduced above are implemented in the spatial domain. As discussed in Section I, it is crucial to design an image retargeting algorithm in the compressed domain. In this paper, we design such an image retargeting algorithm in the compressed domain based on the proposed saliency detection model. We use multioperators including the block-based seam carving and the image scaling to perform image resizing. Different from the existing algorithms using the image similarity to determine the number of removal seams [21], [22], we define the texture homogeneity to determine the number of the removal block-based seams. Experimental results show our proposed image retargeting algorithm in the compressed domain outperforms the existing ones.

C. Contributions of Our Work

In this paper, we propose a saliency detection model and an adaptive image retargeting algorithm in the compressed domain. The saliency map is calculated based on the directly extracted features from JPEG bit-stream. To demonstrate the advantages and applications of the proposed saliency detection model, we design a novel adaptive image retargeting algorithm based on multioperators in the compressed domain. Our initial work was published in [32]. The main contribution of this paper include the following aspects.

1) We propose a novel saliency detection model in the compressed domain. We present how to extract intensity, color, and texture features directly from the JPEG bit-steam. In addition, we design a new algorithm for saliency detection based on DCT blocks in the compressed domain.

2) We design a novel adaptive image retargeting algorithm in the compressed domain. The block-based seam carving and image scaling are used for image resizing in the proposed algorithm. Different from the existing image retargeting algorithms, the texture homogeneity is defined to determine the number of the removal block-based seams.
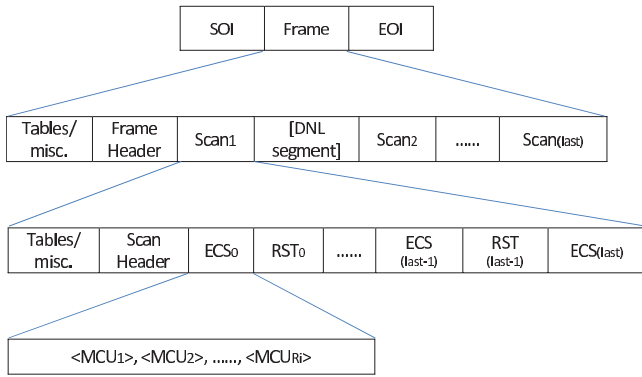
Fig. 2. Syntax for DCT-based modes of operation in JPEG standard [27].

Fig. 3. DCT coefficients and the zig–zig scanning in one 8×8 block.

III. FRAMEWORK

In this section, we first describe how to extract the intensity, color, and texture features for images in the compressed domain. Then we introduce the proposed saliency detection model based on these features. The adaptive image retargeting algorithm based on the block-based seam carving and image scaling is given in the final section.

A. Feature Extraction from the JPEG Bit-Stream

In this paper, we extract the saliency information for the image from the JPEG bit-stream. The saliency information is derived directly from the DCT coefficients rather than from the fully decoded JPEG images. Therefore, we should obtain the DCT coefficients from the JPEG bit-stream. We first introduce how to obtain DCT coefficients from the JPEG bit-stream. Then we give the description of the feature extraction based on DCT coefficients.

1) Obtaining DCT Coefficients from the JPEG Bit-Stream: The *Baseline method* of JPEG, which is implemented based on DCT, is the most widely used compression method [23]. In this paper, we mainly focus on the *Baseline method* of JPEG. The entropy decoding is used to decode the JPEG bit-stream to obtain the quantized DCT coefficients. As Huffman coding is utilized to encode the quantized DCT coefficients in the *Baseline method* of JPEG [23], [27], we should decode the JPEG bit-stream into quantized DCT coefficients according to the two sets of Huffman tables (one AC table and DC table per set). Then the dequantization operation is applied on these quantized DCT coefficients to obtain the DCT coefficients.

The syntax for DCT-based modes of operation in JPEG standard is shown in Fig. 2. In the JPEG standard, markers are used to identify various structural parts of the compressed data formats. We obtain the marker "SOI" in the JPEG bit-stream to identify the start of a compressed image. The frame header presented at the start of a frame (JPEG image) specifies the source image characteristics, the components in the frame, and the sampling factors for each component, and specifies the destinations from which the quantized tables to be used with each component are retrieved. The parameter of Tq in the frame header specifies the quantization table destination from which the quantization table to use for dequantization of DCT coefficients.
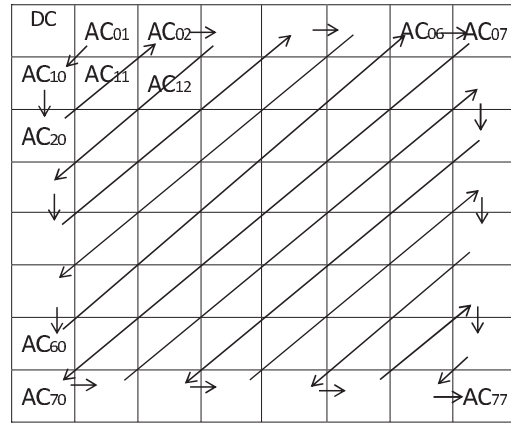
Following the frame header, the scan header specifies which components and which DCT quantized coefficients are contained in the scan. The parameters Td_j and Ta_j in the scan header specify the DC and AC entropy coding table destinations, respectively. The data following the scan header includes the entropy-coded segment (ECS) and restart marker (RST) data. Each ECS is comprised of a sequence of entropy-coded MCUs. The RST is a conditional marker placed between two ECSs only if restart is enabled. Detailed information of the JPEG bit-stream can be found in [27].

Based on the above description, the JPEG bit-stream can be decoded into quantized DCT coefficients based on the DC and AC entropy coding tables (Td_j and Ta_j) from the scan header. According to the quantization table from Tq, the quantized DCT coefficients are further decoded through dequantization operation to get the DCT coefficients.

2) Feature Extraction Based on DCT Coefficients: Three features including the intensity, color, and texture are extracted based on the DCT coefficients to build the saliency detection model. The DCT coefficients in one 8×8 block are shown as Fig. 3. DCT coefficients in one block are comprised of the DC coefficient and AC coefficients. In each block, the DC coefficient is a measure of the average energy over all the 8×8 pixels, while the remaining 63 AC coefficients represent the detailed frequency properties of this block. The JPEG compression standard takes advantage of the fact that most of the energy is included in the first several low-frequency (LF) coefficients, which are in the left-upper corner of the block in Fig. 3. The high-frequency (HF) coefficients from the right-bottom of the block are almost close to zero and thus they are neglected during the quantization of DCT coefficients. The AC coefficients are ordered by zig–zag scanning from LF to HF, as shown in Fig. 3.

The YCrCb color space is used to encode color images in the JPEG standard. The Y channel represents the Luminance information while the Cr and Cb channels include the Chrominance information for JPEG images. As discussed above, the DC coefficients represent the average energy of each block including 8×8 pixels. Here, we first transfer the DC coefficients from YCrCb color space to the RGB color space to extract the intensity and color features for JPEG images. We calculate the color and intensity features
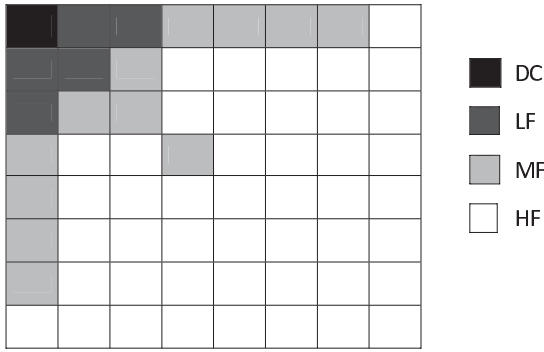
Fig. 4.   Different types of DCT coefficients in one 8×8 block.

by the following steps: let r, g, and b denote the red, green, and blue color components from DC coefficients, and four broadly-tuned color channels are generated as $R = r - (g + b)$ for new red component, $G = g - (r + b)/2$ for new green component, $B = b - (r + g)/2$ for new blue component, and $Y = (r + g)/2 - |r - b|/2 - b$ for new yellow component. The intensity feature can be calculated as: $I = (r + g + b)/3$. Each color channel is then decomposed into red/green and blue/yellow double opponency according to the related properties of the human primary visual cortex [33]: $C_{rg} = R - G$ and $C_{by} = B - Y$.

I, C_{rg}, and C_{by} are the three extracted intensity and color features for an 8×8 block in the JPEG image. It is noted that a 16×16 MCU consists of four 8×8 Luminance blocks and two 8×8 Chrominance blocks (one for Cb and the other for Cr). Thus, four Luminance blocks share the same Chrominance blocks in a typical 4:2:0 component subsampling JPEG encoding system.

The AC coefficients include the detailed frequency information for each image block, and previous studies have shown that the AC coefficients can be used to represent the texture information for images blocks [24], [26], [34]–[36]. Here, we use the AC coefficients in YCrCb color space to extract the texture feature for each 8×8 block. In YCrCb color space, Cr and Cb components represent the color information and their AC coefficients provide little information for texture. In addition, a 16×16 MCU consists of more Luminance blocks compared with those from Chrominance blocks in a typical 4:2:0 scheme. Thus, we use the AC coefficients from the Y component only to extract the texture feature T. As to the AC coefficients in one 8×8 DCT block, the LF components capture much of the detailed information, while the HF components include less information. Following the studies in [37] and [38], we classify the AC coefficients into three parts: LF, medium-frequency (MF), and HF parts, as shown in Fig. 4. The coefficients in each part are summed as one value to obtain three corresponding elements (t_{LF}, t_{MF}, and t_{HF}) to represent the texture feature for each DCT block. Therefore, the texture feature T for each DCT block can be expressed as follows:

$$T = \{t_{LF}, t_{MF}, t_{HF}\} \tag{1}$$

where t_{LF}, t_{MF}, and t_{HF} is the sum of all the coefficients in the LF, MF, and HF parts, respectively, in Fig. 4.

B. Saliency Detection in the Compressed Domain

From the above section, the intensity, color, and texture features (I, C_{rg}, C_{by}, and T) can be extracted from the DCT coefficients of each 8×8 block. In this paper, we use these four features (one intensity, two color features, and one texture feature) to obtain four feature maps, respectively. Then we use the coherent normalization-based fusion method to combine these four feature maps to get the saliency map for JPEG images. The details of the saliency detection model are shown in the following.

1) Feature Differences Between DCT Blocks: As to the intensity and color features (I, C_{rg}, and C_{by}), the feature differences between blocks i and j can be computed as

$$D_{ij}^k = \mathfrak{C}_i^k - \mathfrak{C}_j^k \tag{2}$$

where $k = 1, 2, 3$ represent the intensity and color features, respectively (one intensity feature and two color features); $\mathfrak{C}^k \in \{I, C_{rg}, C_{by}\}$.

We use the vector T from (1) including three elements to represent the texture feature for each DCT block in the JPEG image. The Hausdorff distance [39] is used here to calculate the difference between two vectors of texture feature from two different blocks. Hausdorff distance is widely used to calculate the dissimilarity between two point sets through examining the fraction of points in one set that lie near points in the other set (and perhaps vice versa). The texture difference D_{ij}^4 between two blocks i and j can be computed as follows:

$$D_{ij}^4 = \max(h(T_i, T_j), h(T_j, T_i)) \tag{3}$$

where 4 means the texture feature is the fourth feature (the first three features include one intensity and two color features described above); T_i and T_j represent the vectors of texture feature for blocks i and j, respectively. $h(T_i, T_j)$ is calculated as

$$h(T_i, T_j) = \max_{t_i \epsilon T_i} \min_{t_j \epsilon T_j} \| t_i - t_j \| \tag{4}$$

where $\| . \|$ is the L2 norm.

2) Feature Maps in the Compressed Domain: In this paper, the saliency value for each DCT block in each feature map is determined by two factors: one is the block differences between this DCT block and all other DCT blocks of the input image and the other is the weighting for these block differences. If these differences between this DCT block and all other DCT blocks are larger, then the saliency value for this DCT block is larger. In addition, a Gaussian model of the Euclidean distances between DCT blocks is used to determine the weighting for these DCT block differences. We use Gaussian model of the Euclidean distance for its generality. Here, we use S_i^k to represent the saliency value calculated from the kth feature for the DCT block i. The feature map for the kth feature can be obtained as follows:

$$S_i^k = \sum_{j \neq i} \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{d_{ij}^2}{2\sigma^2}} D_{ij}^k \tag{5}$$

where σ is a parameter for the Gaussian model; d_{ij} is the Euclidean distance between DCT blocks i and j; D_{ij}^k is calculated as (2) and (3). In this paper, we set $\sigma = 5$.

From (5), the saliency value of DCT block i considers all the block differences between this DCT block and the other DCT blocks in the image. The saliency value of the block i is larger with greater block differences from all other blocks in the image. Here, we use a Gaussian model of the Euclidean distances between DCT blocks to weight the block differences. From (5), the weights of the block differences from nearer neighbor blocks are larger compared with these from farther neighbor blocks. Therefore, the contributions of the block differences to the saliency value of the DCT block i will decrease with larger-distance DCT blocks from the DCT block i. On the contrary, the contributions of the block differences to the saliency value of the DCT block i will increase with smaller-distance DCT blocks from the DCT block i.

According to (5), we can obtain four feature maps (one intensity feature map, two color feature maps, and one texture feature map) based on the intensity, color, and texture features. The final saliency map is a combination of these four feature maps. We will describe how to get the final saliency map for images in the next section.

3) Final Saliency Map in the Compressed Domain: After obtaining the four feature maps $S^k (k \in \{1, 2, 3, 4\})$, the saliency map for the JPEG image can be obtained by integrating these four feature maps. In this paper, we use the coherent normalization-based fusion method to combine these four feature maps into the saliency map $\mathbb{S}$ as follows:

$$\mathbb{S} = \sum \gamma_\theta N_{(\theta)} + \beta \prod N_{(\theta)} \qquad (6)$$

where N is the normalization operation; $\theta \in \{S^k\}$; γ_θ and β are parameters determining the weights for each components in (6). In this paper, we set $\gamma_\theta = \beta = 1/5$. The second term in (6) represents these areas which all the four feature maps S^k detect as salient areas.

C. Image Resizing Operation

As described above, we build a saliency detection model in the compressed domain by extracting intensity, color, and texture features from the DCT coefficients. To demonstrate its advantages and applications, we design an novel adaptive image retargeting algorithm in the compressed domain. The proposed saliency detection model is used to obtain the saliency map to measure the visual importance of each DCT block for the image retargeting algorithm. Multioperators including the block-based seam carving and the image scaling are utilized to resize the JPEG images in the proposed image retargeting algorithm. The number of block-based seam carving is determined by the defined image homogeneity.

The image resizing operation steps are given as follows: 1) determine the number of block-based seam carving operation based on the defined image homogeneity; 2) use the block-based seam carving operation to resize the original image; and 3) use the image scaling operation to resize the retargeted image from the block-based seam carving to obtain the final retargeted image. The details will be described in the following sections.
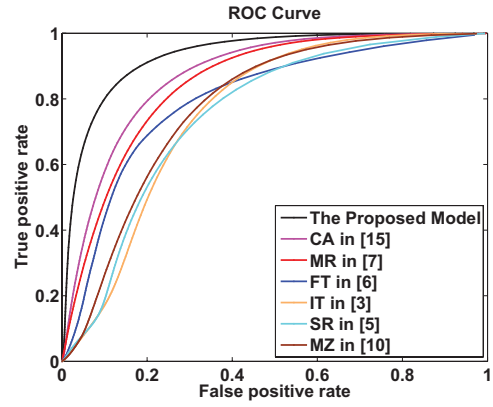

Fig. 5. Comparison result of the ROC curves between different saliency detection models.

1) Block-Based Seam Carving Operation: It is noted that since our final saliency map is at block-level, each seam indicates connected blocks instead of connected pixels in the original image. We use DCT blocks with the size of 8×8 to calculate the saliency map, thus the final saliency map is only 1/64 times of the original image and each value in the final saliency map represent the saliency value for one 8×8 DCT block. We design a block-based seam carving method based on the forward energy [18] to determine the optimal block seams. Based on the saliency map $\mathbb{S}$ in (6), the block-based seam carving uses the following dynamic programming technique to determine the optimal block-based seams:

$$M(i, j) = \mathbb{S}(i, j) + \min \begin{cases} M(i-1, j-1) + C_L(i, j) \\ M(i-1, j) + C_U(i, j) \\ M(i-1, j+1) + C_R(i, j) \end{cases} \qquad (7)$$

where $M(i, j)$ determines the position (i, j) of the saliency map for the optimal block-based seams; $C_L(i, j)$, $C_U(i, j)$, and $C_R(i, j)$ are the costs due to the generation of new neighbor blocks separated by the removal seam previously. These costs are calculated as follows:

$$\begin{cases} C_U(i, j) = \left\| \mathbb{S}(i, j+1) - \mathbb{S}(i, j-1) \right\| \\ C_L(i, j) = C_U(i, j) + \left\| \mathbb{S}(i-1, j) - \mathbb{S}(i, j-1) \right\| \\ C_R(i, j) = C_U(i, j) + \left\| \mathbb{S}(i-1, j) - \mathbb{S}(i, j+1) \right\|. \end{cases} \qquad (8)$$

2) Adaptive Image Retargeting: The optimal block-based seams can be determined by (7) and (8). As introduced previously, the proposed image retargeting algorithm first utilizes the block-based seam carving operation to resize the image. Then the image scaling operation is used to obtain the final retargeted images. In this paper, we propose to use the image homogeneity to decide the number of removal block-based seams. The number of removal block-based seams in dimension p (horizontal or vertical) can be calculated as follows:

$$n_p^s = \frac{\lambda_p \left(n_p - n_p^r \right)}{8} \qquad (9)$$

where n_p^s is the number of removal block seams in dimension p; λ_p represents the texture homogeneity of the image in dimension p, which is used to determine the number of the
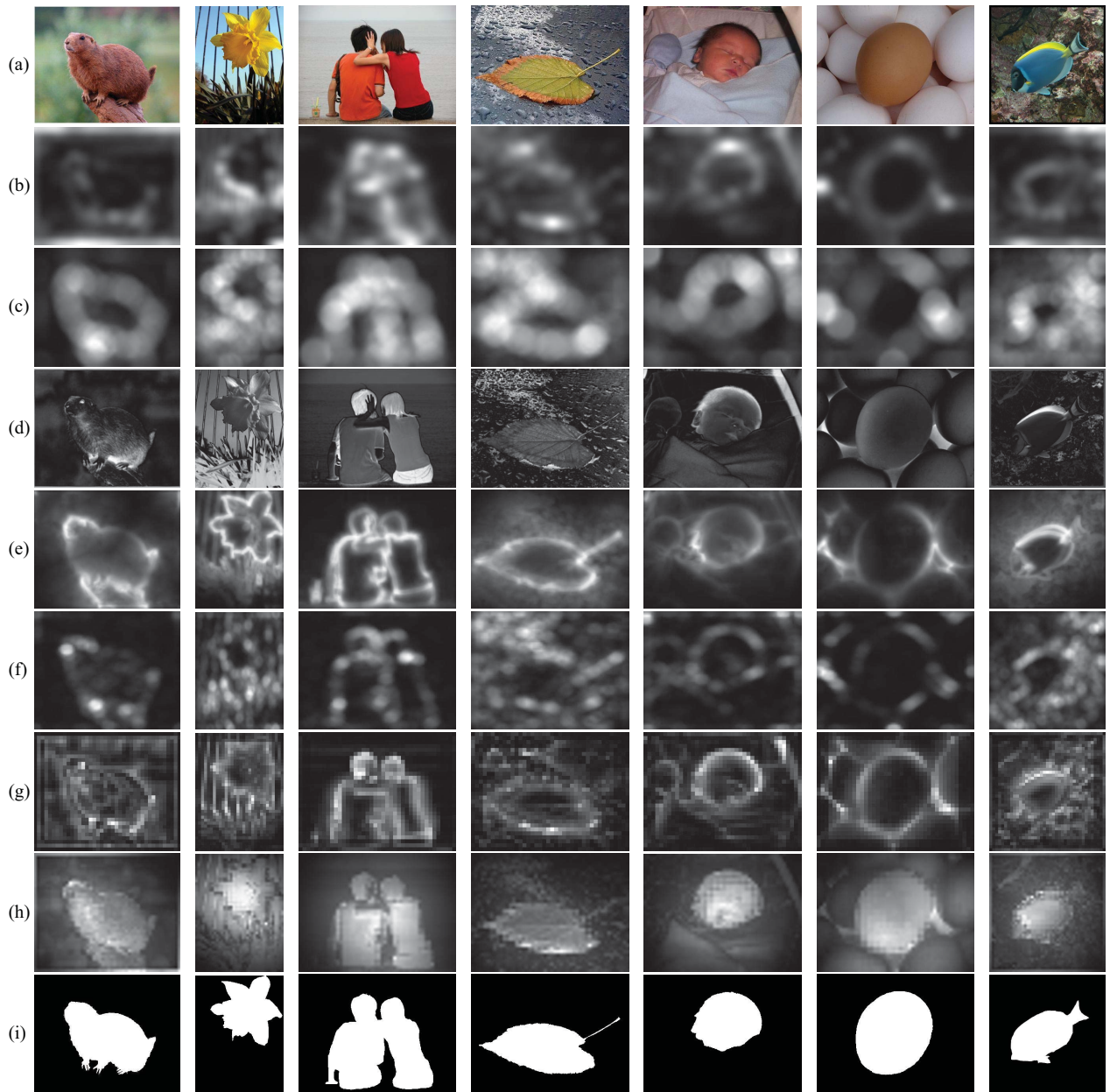
Fig. 6.   Comparison of the different saliency detection algorithms. (a) Original images. (b)–(i) Saliency maps from [3], [5]–[7], [10], and [16], our proposed algorithm and the ground truth, respectively.

removal block seams; n_p is the length of the original image in dimension p (width or height); n_p^r is the length of the retargeted image in dimension p (width or height). The value of n_p^r is decided by the size of the display screen of the client based on the initial communication between the server and client in real applications. As the proposed algorithm is based on DCT blocks and the size of DCT blocks is 8×8, we use the number eight to calculate the number of the removal block-based seams in (9).

Texture homogeneity are widely used in various applications [4], [40]–[42]. In this paper, we define a measurement for texture homogeneity λ to determine the number of removal block seams. The texture homogeneity defined here is dependent on the energy spatial-distribution and the energy connectedness (here, the saliency map from the proposed saliency detection model is regarded as the energy map). If the image

energy is more centralized and connected, there may be only one or several small salient objects in the image with simple background. In this case, we will use more seam carving operation to remove the block-based seams. On the contrary, with more disconnected and decentralized energy distribution, the image may include one or several big salient objects, or the context of the image is complex. In this case, we should use more image scaling operation to resize the image for preserving these salient objects or the context information.

The texture homogeneity of the image in dimension p (horizontal or vertical) can be computed as follows:

$$\lambda_p = (1 - \tau_p) * v_p \tag{10}$$

where τ_p represents the spatial variance of the energy pixels in dimension p; v_p represents the connectedness of the energy pixels in the dimension p. In this paper, the Otsu's

Fig. 7. Overall statistical result for retargeted images from four different algorithms.
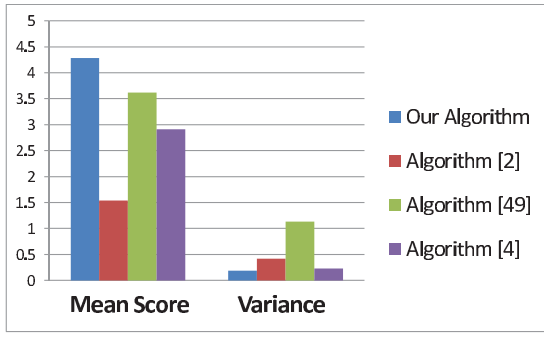

Fig. 8. Detailed statistical result for retargeted images from four different algorithms.

thresholding algorithm [43] is used to binarize the energy map into energy pixels (the energy value is 1) and non-energy pixels (the energy value is 0).

To simplify the description, we just demonstrate how to calculate the horizontal variance of the energy pixels here. The calculation process of vertical variance of the energy pixels is similar. The horizontal variance of the energy pixels τ_1 in the image can be calculated as follows:

$$\tau_1 = \frac{1}{P} \sum_{(i,j)} |i - H|^2 * \mathbb{E}_{(i,j)} \qquad (11)$$

$$H = \frac{1}{P} \sum_{(i,j)} i * \mathbb{E}_{(i,j)} \qquad (12)$$

where $\mathbb{E}_{(i,j)}$ is the energy value for the position (i, j); P represents all the energy pixels in the image, $P = \sum_{(i,j)} \mathbb{E}_{(i,j)}$.

From (12), we can see that H is the expected value of the spatial location for the energy value in the image. Thus, we can obtain the horizontal variance of the energy pixels for the image based on (11) and (12). We set $\tau_1 = 1$ when all the energy pixels are centralized into one square in the image. In this case, the image texture is most homogeneous. On the contrary, we set $\tau_1 = 0$ when all the energy pixels are distributed uniformly over the image. In this case, the image texture is considered the most inhomogeneous. We normalize the τ_1 from (11) based on these two cases and then use the normalized τ_1 to calculate λ_1 in (10).

The connectedness of the energy pixels in the image is measured by the number of the energy pixels in the neighborhood of all energy pixels in the image. For each dimension (horizontal or vertical) of the image, there are at most six neighbor pixels for each energy pixel. The other two neighbor pixels are from the other dimension and thus not considered. We use v_p^i to represent the connectedness of the energy pixel i for the dimension p. It can be computed as follows [40]:

$$v_p^i = \frac{1}{6} \sum_{z \in M_i} f(z) \qquad (13)$$

where M_i includes all six neighbor pixels around i; $f(z)$ is the function to denote whether the neighbor pixel z is a energy pixel or not.

The connectedness of the image energy in dimension p is obtained as the sum of the connectedness of all energy pixels
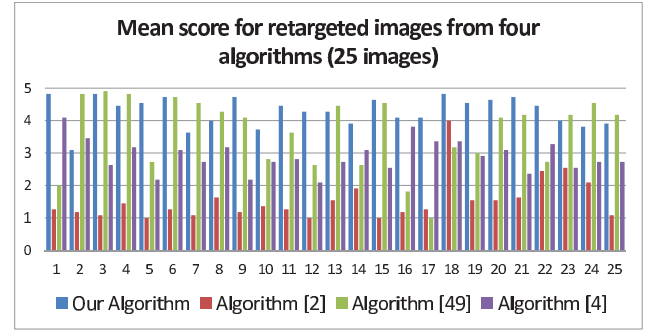
in the image as follows:

$$v_p = \frac{1}{K} \sum_i v_p^i \qquad (14)$$

where K is the number of the energy pixels in the image.

We can obtain the connectedness of the image energy according to (13) and (14). We set $v_p = 1$ when the energy pixels in the image are centralized as a connected square. In this case, the image texture owns the greatest connectedness with this amount of energy pixels. On the contrary, we set $v_p = 0$ when the energy pixels in the image are distributed uniformly over the image. Based on these above two cases, the v_p from (13) is normalized between 0 and 1 as the relative connectedness to be used in (10).

Therefore, the amount of removal block-based seams for images can be obtained according to (9)–(14). After we utilize the block-based seam carving to remove the optimal block-based seams, the image scaling is used to scale the retargeted image from the block-based seam carving to obtain the final retargeted image.

IV. EXPERIMENTS

In this section, we evaluate the overall performance of the proposed algorithm from two aspects: one is the performance of the proposed saliency detection algorithm and the other is the performance of the proposed image retargeting algorithm in resizing images. The following two sections give the performance comparisons between the proposed algorithms and the existing ones from these two aspects, respectively.

A. Saliency Detection Evaluation

Saliency detection models are widely used to detect the salient objects [4], [7], [8], for it can provide the positions of the salient objects in images. Therefore, a efficient qualitative evaluation method for the performance of saliency detection algorithms is to extract the salient objects in images. Actually, salient object detection is widely utilized in the performance evaluation of saliency detection algorithms [3]–[7]. In this paper, we use the public database provided by Achanta *et al.* [6] to evaluate the proposed saliency detection algorithm. This database includes 1000 images with accurately human-labeled ground truth for the salient objects.

TABLE I

COMPARISON RESULT OF ROC AREAS BETWEEN DIFFERENT SALIENCY DETECTION MODELS

	IT in [3]	FT in [6]	CA in [16]	MR in [7]	SR in [5]	MZ in [10]	Our model
ROC area	0.8028	0.8229	0.88834	0.8630	0.8025	0.7951	0.9397

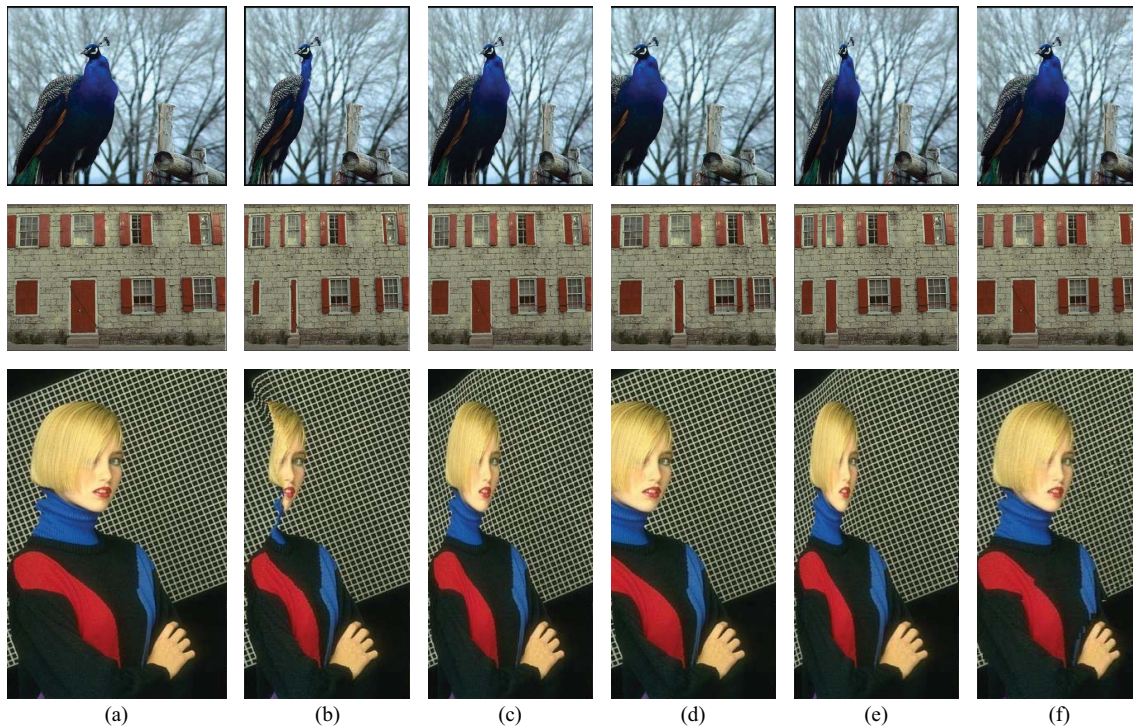(a)　　　　(b)　　　　(c)　　　　(d)　　　　(e)　　　　(f)

Fig. 9.　Comparison of the different image retargeting algorithms. (a) Original images. (b)–(f) Retargeted images from [13], [14], [18], and [45], and our proposed algorithm, respectively. The width of the retargeted images is 75% of the width of the original images.

We compare the proposed saliency detection algorithm with other six existing ones: IT [3], FT [6], SR [5], CA [16], MR [7], and MZ [10]. Three (FT, SR, and MR) of these algorithms are implemented in the frequency domain while the other three (IT, CA, and MZ) are built in the spatial domain. In the public database [6], the sizes of the human-labeled ground truth are the same as those of the original images. Therefore, the saliency maps from all the saliency detection algorithms are resized to the sizes of the original images for the fair comparison.

Receiver operating characteristic (ROC) is widely used to evaluate the performance of the saliency map [16], [7]. The saliency map obtained by a computational saliency detection model can be divided into the salient points and the non-salient points through defining a threshold. The ground truth map marked by subjects includes the target points and the background points. The percentage of target points falling into the salient points from a computational saliency detection model is true positive rate (TPR), while the percentage of background points falling into the salient points is false positive rate (FPR). The ROC curve for a specified saliency detection algorithm can be achieved as the curve of TPR versus FPR through choosing different thresholds. The overall quantitative performance for this specified saliency detection algorithm can be given by the area under the ROC curve, which is ROC area. Generally, the larger the ROC area is, the better the

predication performance of the saliency detection algorithm is for the salient object detection.

The ground truth map from the database [6] includes the target points (the accurately labeled salient objects in images) and the background points (the accurately labeled non-salient areas), which can be shown in the Fig. 6(i). After obtaining the saliency maps from different saliency detection models, we choose different thresholds from 0 to 255 for these saliency maps to get the ROC curves for different saliency detection models. The ROC curves for different saliency detection models are shown in Fig. 5. Table I shows the ROC areas for all these saliency detection algorithms. From Fig. 5 and Table I, our proposed algorithm obviously outperforms other existing ones. According to Table I, the ROC areas from [7] and [16] are larger than those from [3], [5], and [10]. This is coincident with the experimental results in the studies [7] and [16].

We give some comparison results in Fig. 6. From Fig. 6(b) and (g), we can see that the saliency maps from [3] and [10] are mainly detect the contour of the salient objects. The reason for this is that the saliency detection models in [3] and [10] mainly consider the local contrast for calculating the saliency map for the image. Furthermore, the saliency detection models from [3] and [10] also detect some non-salient areas as salient, as shown in saliency maps from Fig. 6(b) and (g). The saliency maps in Fig. 6(c) and (f) are obtained from
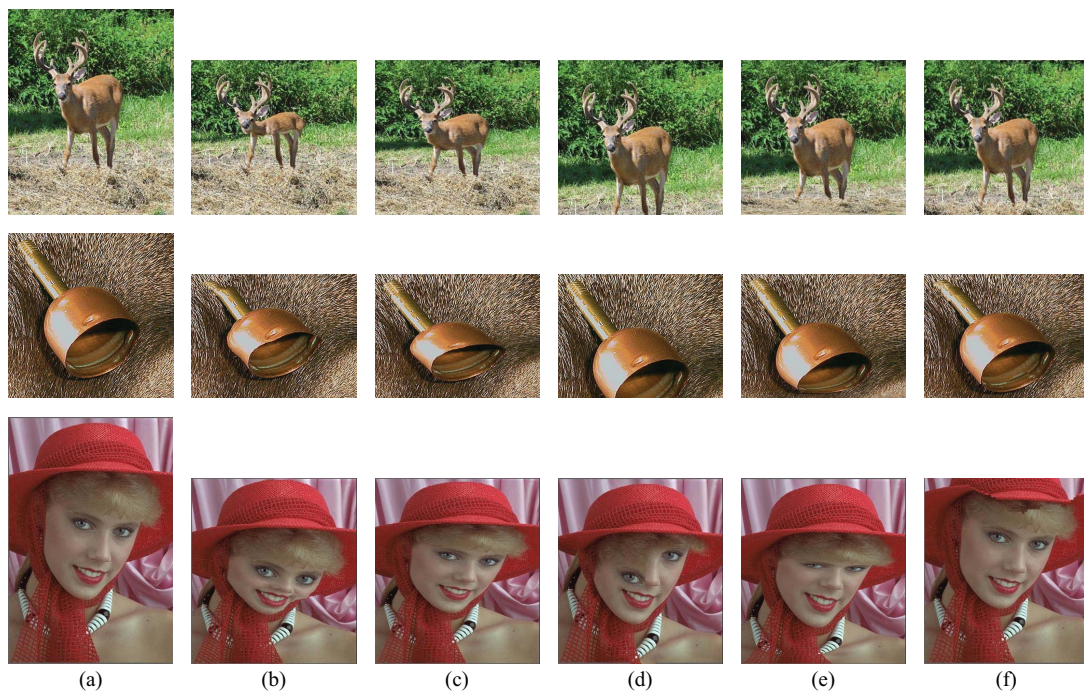
Fig. 10. Comparison of the different image retargeting algorithms. (a) Original images. (b)–(f) Retargeted images from [13], [14], [18], and [45], and our proposed algorithm, respectively. The height of the retargeted images is 75% of the height of the original images.

the phase-based saliency detection models [5], [7]. In these models, the Fourier transform is used for images to obtain the amplitude and phase first, then the saliency map is calculated as the reconstructed image through inverse Fourier transform on a constant amplitude (always set as 1) and the original phase. These models obtain the saliency maps for images mainly by the global contrast. Thus, the saliency maps from Fig. 6(c) and (f) mainly include the HF areas (such as edges) as salient. To improve the saliency detection model [3], the authors in [6] built the saliency detection model by retaining more frequency content for images. However, this model still loses some visually important information in the saliency map [saliency maps from the first, fifth, sixth, and seventh columns in Fig. 6(d)] or detect some background areas as salient [saliency maps from the second and fourth columns in Fig. 6(d)]. Although the saliency maps in the fifth row seem better than those from the models in [3], [5], [7], and [10], they also suffer the defect that the contour of the salient objects is much more salient than other parts of the salient objects in images, as shown in Fig. 6(e). On the contrary, our proposed saliency detection can obtain more accurate salient areas for images, as shown from the saliency maps from the Fig. 6(h).

It is commonly accepted that the texture features for DCT blocks will change in different compression ratios. However, the change will not significantly affect the detected saliency map of the image in the proposed algorithm. We have conducted the experiments to investigate the influence of the compression ratio. Experimental results demonstrate that the image compression has little influence on the saliency map. On the other hand, one recent study has shown that video coding impairments would not disturb the visual attention regions from observers [44]. Therefore, the normal image/video compression will not change the saliency results from the

proposed algorithm and the observers. In addition, we have also conducted the experiments for the images with different sizes. Experimental results have shown that the saliency map will not change greatly with the changed image sizes. In this paper, the saliency value for each DCT block is calculated through computing the weighted differences between this DCT block and all other DCT blocks in the image. In the JPEG standard, the size of each DCT block is 8×8 and the DCT blocks are less in the images with smaller sizes. Therefore, the computational cost will decrease with smaller image sizes. In this paper, we do not provide the experimental results for the saliency map of the images with different sizes and compression ratios for the consideration of saving space.

B. Image Retargeting Evaluation

The performance of image retargeting algorithms greatly depends on two factors: the visual significance map and the image resizing operation. In this paper, the saliency map in the compressed domain is used as the visual significance map to measure the visual importance for each pixel in images. The performance of the proposed saliency detection in the compressed domain is demonstrated in the above experiment. In this experiment, we compare the performance of the proposed image resizing operation with that from other existing ones.

To demonstrate the performance of our proposed image retargeting algorithm in the compressed domain, we use an image dataset including 25 images to conduct a user study. The compression ratio is about 1.34 b/p for these images. Three state-of-the-art image retargeting algorithms [13], [18], [45] are adopted for comparisons. Eleven participants (four are female while others are male) involve in this experiment.
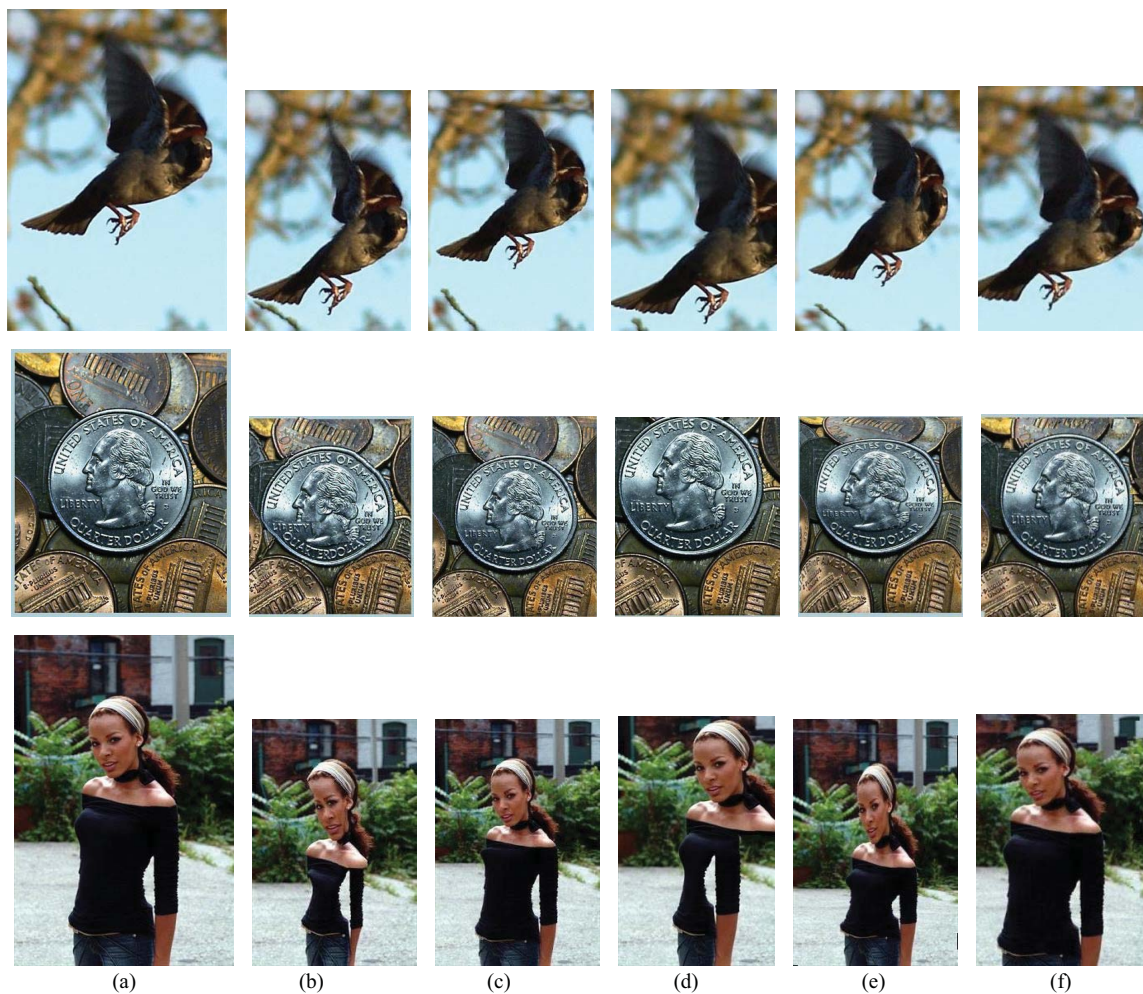
Fig. 11.   Comparison of the different image retargeting algorithms. (a) Original images. (b)–(f) Retargeted images from [13], [14], [18], and [45], and our proposed algorithm, respectively. The width and height are 75% of the width and height of the original images, respectively.

The original images are used as reference images. The experiments are conducted in the typical laboratory environment. All retargeted images from four different algorithms are displayed in the random order on the screen. Mean opinion scores (1–5) are recorded by participants where 1 means bad viewing experience and 5 means excellent viewing experience. Each participant votes for this image dataset. The statistical results for the retargeted images are shown in Figs. 7 and 8.

Fig. 7 shows the overall mean scores and variances from four different image retargeted algorithms based on the 25 images. From this figure, we can see that the mean score from our proposed algorithm is higher than those from other three ones. This means that the overall viewing experience from the retargeted images from our proposed algorithm is better than those from others. Meanwhile, the score variance from our proposed algorithm is lower than those from others. This means that the retargeted results from our proposed algorithm are more stable than those from others. From this figure, we can see that the overall viewing experience from the algorithm [45] is better than other two algorithms [13], [18]. However, the stability from the algorithm [45] is not as good as the other two [13], [18]. Fig. 8 presents the number of the retargeted images under each score. From this figure, most of the retargeted images from

our proposed algorithm provide better viewing experience for users.

Here, we give some visual comparison results in Figs. 9–12. In these figures, we add the results of one saliency-based image retargeting algorithm [14] for the comparison. From Figs. 9–11, we can see that the retargeted images from the algorithm [18] suffer serious distortion. The viewing experience of the retargeted images from the algorithms [13], [14] is better than that from algorithm [18]. However, there is still some distortion in these retargeted images from [13] and [14], as shown in the first and third rows in Fig. 9, and the second row in Fig. 11. The retargeted coin from [13] and [14] in the second row of Fig. 11 obviously suffer serious distortion. In addition, compared with the original images, the salient objects in the retargeted images from [13] and [14] become much smaller, as shown in the first row in Fig. 9 and the third row in Fig. 11. From Figs. 9(d)–11(d), we can see that the retargeted images from [45] obviously lose some important visually information. In addition, there is some serious distortion in some retargeted images from [45], as shown in the third row in Figs. 10 and 11. On the contrary, the retargeted images from our proposed image retargeting algorithm are obviously better than those from other algorithms. In Fig. 12, the retargeted images with

images over Internet were typically stored in the compressed format of JPEG. In this paper, we proposed a novel saliency detection model in the compressed domain. In addition, we also design a novel adaptive image retargeting algorithm in the compressed domain based on the proposed saliency detection model.

First, we extracted the intensity, color, and texture features from DCT coefficients in the JPEG bit-stream to calculate the DCT block differences based on Hausdorff distance. Combining the Gaussian model for the Euclidean distances between the DCT blocks, we utilized the DCT block differences to obtain the saliency map for JPEG images. Experimental results showed that the proposed saliency detection model in the compressed domain outperforms the existing ones.

Furthermore, based on the proposed saliency detection model, we design a novel adaptive image retargeting algorithm in the compressed domain. The saliency map from our proposed saliency detection model was used as the visual significance map for our image retargeting algorithm. The multioperator operation including the block-based seam carving and the image scaling is utilized for image resizing. Different from the existing studies which use the image similarity to determine the number of seam carving operation, the texture homogeneity was defined to determine the number of the removal block-based seams in this paper. Experimental results showed that the performance of the proposed image retargeting algorithm is better than those from existing ones.
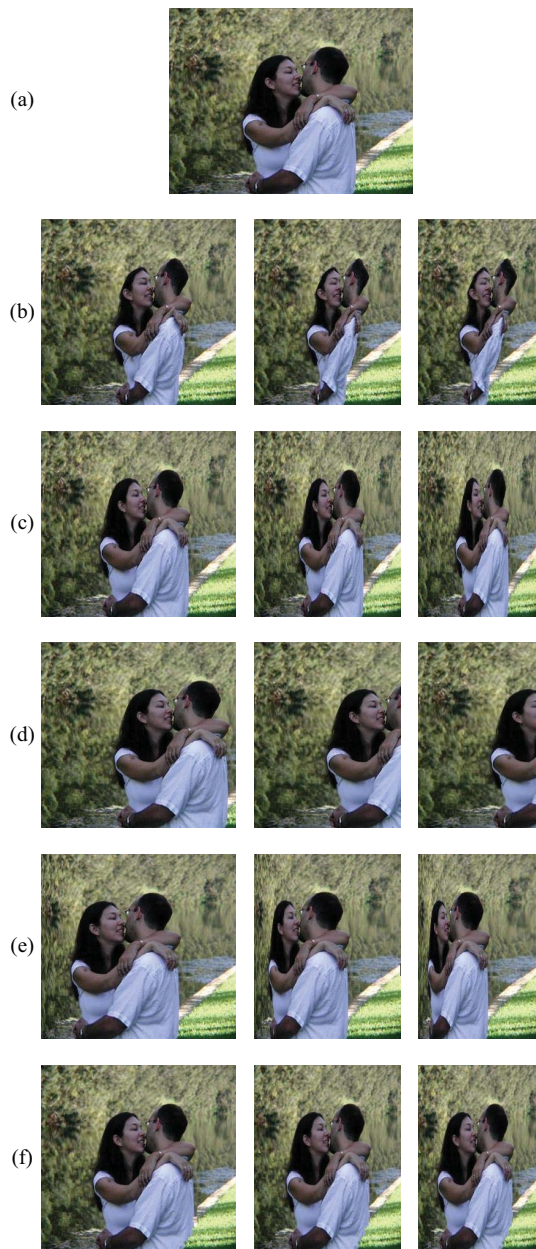

Fig. 12. Comparison of the different image retargeting algorithms. (a) Original image. (b)–(f) Retargeted images from [13], [14], [18], and [45], and our proposed algorithm, respectively. The width of the three retargeted images is 50%, 60%, and 80% of the width of the original image, respectively.

different sizes from one sample image are given for different image retargeting algorithms. From this figure, we can see that there is some distortion with the retargeted images from [13], [14], and [18]. From the Fig. 12(d), the retargeted image with 80% width from [45] can preserve the salient objects in the image. However, the retargeted images with smaller width sizes lose much important visually information, as shown in the second and third retargeted images from [45] in the Fig. 12(d).

V. CONCLUSION

Saliency detection was widely used in various image processing applications. Existing saliency detection algorithms were implemented in the uncompressed domain. However,

REFERENCES

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

[2] J. Wolfe, K. R. Cave, and S. L. Franzel, "Guided search: An alternative to the feature integration model for visual search," *J. Experim. Psychol.: Human Percept. Perform.*, vol. 15, no. 3, pp. 419–433, 1989.

[3] 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.

[4] T. Liu, J. Sun, N. Zheng, X. Tang, and H. Y. Shum, "Learning to detect a salient object," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jul. 2007, pp. 1–8.

[5] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[6] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1597–1604.

[7] C. Guo and L. Zhang, "A novel multi-resolution spatiotemporal saliency detection model and its applications in image and video compression," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 185–198, Jan. 2010.

[8] D. Walther and C. Koch, "Modeling attention to salient proto-objects," *Neural Netw.*, vol. 19, no. 9, pp. 1395–1407, 2006.

[9] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 19. 2007, pp. 545–552.

[10] Y. Ma and H. Zhang, "Contrast-based image attention analysis by using fuzzy growing," in *Proc. ACM Int. Conf. Multimedia*, 2003, pp. 228–241.

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

[12] R. Achanta and S. Susstrunk, "Saliency detection for content-aware image resizing," in *Proc. IEEE Int. Conf. Image Process.*, Nov. 2009, pp. 1005–1008.

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

[14] T. Ren, Y. Liu, and G. Wu, "Image retargeting based on global energy optimization," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jun.–Jul. 2009, pp. 406–409.

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

[16] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2376–2383.

[17] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," *ACM Trans. Graph.*, vol. 26, no. 3, pp. 1–9, 2007.

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

[19] Y. Jin, L. Liu, and Q. Wu, "Nonhomogeneous scaling optimization for realtime image resizing," *Visual Comput.*, vol. 26, nos. 6–8, pp. 769–778, 2010.

[20] 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.

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

[22] 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.

[23] G. K. Wallace, "The JPEG still picture compression standard," *Commun. ACM*, vol. 34, no. 4, pp. 30–44, Apr. 1991.

[24] Y. Zhong and A. K. Jain, "Object localization using color, texture and shape," *Pattern Recognit.*, vol. 33, no. 4, pp. 671–684, 2000.

[25] J. Wei, "Color object indexing and retrieval in digital libraries," *IEEE Trans. Image Process.*, vol. 11, no. 8, pp. 912–922, Aug. 2002.

[26] C. Theoharatos, V. K. Pothos, N. A. Laskaris, G. Economou, and S. Fotopoulos, "Multivariate image similarity in the compressed domain using statistical graph matching," *Pattern Recognit.*, vol. 39, no. 10, pp. 1892–1904, 2006.

[27] *10918-1: Information Technology–Digital Compression and Coding of Continuous-Tone Still Images–Requirements and Guidlines*, ISO/IEC Standard JTC1/SC29/WG1, 1994.

[28] N. D. Bruce and J. K. Tsotsos, "Saliency based on information maximization," *Adv. Neural Inf. Process. Syst.*, vol. 18, pp. 155–162, Jun. 2006.

[29] O. Le Meur, P. Le Callet, D. Barba, and D. Thoreau, "A coherent computational approach to model the bottom-up visual attention," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 5, pp. 802–817, May 2006.

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

[31] D. Vaquero, M. Turk, K. Pulli, M. Tico, and N. Gelfand, "A survey of image retargeting techniques," *Proc. SPIE*, vol. 7798, pp. 779814-1–779814-15, Sep. 2010.

[32] Y. Fang, Z. Chen, W. Lin, and C. Lin, "Saliency-based image retargeting in the compressed domain," in *Proc. 19th ACM Int. Conf. Multimedia*, 2011, pp. 1049–1052.

[33] S. Engel, X. Zhang, and B. Wandell, "Colour tuning in human visual cortex measured with functional magnetic resonance imaging," *Nature*, vol. 388, no. 6, pp. 68–71, 1997.

[34] Y. L. Huang and R. F. Chang, "Texture features for DCT-coded image retrieval and classification," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 6. Mar. 1999, pp. 3013–3016.

[35] Y. Li, X. Chen, X. Fu, and S. Belkasim, "Multi-level discrete cosine transform for content-based image retrieval by support vector machines," in *Proc. IEEE Int. Conf. Image Process.*, Sep.–Oct. 2007, pp. 21–24.

[36] H. Li, G. Liu, and Z. Zhang, "A new texture generation method based on pseudo-DCT coefficients," *IEEE Trans. Image Process.*, vol. 15, no. 5, pp. 1300–1312, May 2006.

[37] H. Y. Tong and A. N. Venetsanopoulos, "A perceptual model for JPEG applications based on block classification, texture masking, and luminance masking," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 1998, pp. 428–432.

[38] Y. Jia, W. Lin, and A. A. Kassim, "Estimating just-noticeable distortion for video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 7, pp. 820–829, Jul. 2006.

[39] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*. New York: Springer-Verlag, 2005.

[40] L. Macaire, N. Vandenbroucke, and J. G. Postaire, "Color image segmentation by analysis of subset connectedness and color homogeneity properties," *Comput. Vis. Image Understand.*, vol. 102, no. 1, pp. 105–116, 2006.

[41] K. Hirakawa and T. W. Parks, "Adaptive homogeneity-directed demosaicing algorithm," *IEEE Trans. Image Process.*, vol. 14, no. 3, pp. 360–369, Mar. 2005.

[42] H.-D. Cheng and Y. Sun, "A hierarchical approach to color image segmentation using homogeneity," *IEEE Trans. Image Process.*, vol. 9, no. 12, pp. 2071–2082, Dec. 2000.

[43] N. Otsu, "A threshold selection method from gray level histograms," *IEEE Trans. Syst., Man Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.

[44] O. Le Meur, A. Ninassi, P. Le Callet, and D. Barba, "Do video coding impairments disturb the visual attention deployment?" *Signal Process.: Image Commun.*, vol. 25, no. 8, pp. 597–609, Sep. 2010.

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

Yuming Fang received the B.E. degree in software engineering from Sichuan University, Chengdu, China, in 2006, and the M.S. degree in communication and information system from the Beijing University of Technology, Beijing, China, in 2009. He is currently pursuing the Ph.D. degree with the School of Computer Engineering, Nanyang Technological University, Singapore.

His current research interests include image and video processing and computer vision.

Zhenzhong Chen (M'07) received the B.Eng. degree from the Huazhong University of Science and Technology (HUST), Wuhan, China, and the Ph.D. degree from the Chinese University of Hong Kong (CUHK), Hong Kong, both in electrical engineering.

He is currently a Lee Kuan Yew Research Fellow and Principal Investigator with Nanyang Technological University (NTU), Singapore. Before joining NTU, he was an ERCIM fellow with the National Institute for Research in Computer Science and Control (INRIA), France. He held visiting positions at Polytech'Nantes, France, and Universite Catholique de Louvain (UCL), Belgium, Belgium. His current research interests include visual perception, visual signal processing, and multimedia communications.

Dr. Chen is a Voting Member of the IEEE Multimedia Communications Technical Committee (MMTC), an Invited Member of the IEEE MMTC Interest Group of Quality of Experience for Multimedia Communications from 2010 to 2012. He has served as a Guest Editor of special issues for the IEEE MMTC E-letter and the *Journal of Visual Communication and Image Representation*. He has co-organized several special sessions at international conferences, including IEEE ICIP 2010, IEEE ICME 2010, and Packet Video 2010, and has served as a Technical Program Committee Member of IEEE ICC, GLOBECOM, CCNC, and ICME. He was a recipient of the CUHK Faculty Outstanding Ph.D. Thesis Award, the Microsoft Fellowship, and the ERCIM Alain Bensoussan Fellowship. He is a member of SPIE.

Weisi Lin (M'92–SM'98) received the B.Sc. degree in electronics and the M.Sc. degree in digital signal processing from Zhongshan University, Guangzhou, China, and the Ph.D. degree in computer vision from King's College, London University, London, U.K.

He has been the Project Leader of over ten major successfully-delivered projects in digital multimedia technology development. He also served as the Lab Head, Visual Processing, and the Acting Department Manager, Media Processing, for the Institute for Infocomm Research. Currently, he is an Associate Professor with the School of Computer Engineering, Nanyang Technological University, Singapore. He has published over 190 refereed papers in international journals and conferences. His current research interests include image processing, perceptual modeling, video compression, multimedia communication, and computer vision.

Dr. Lin is a Chartered Engineer, U.K., a fellow of the Institution of Engineering Technology, and an Honorary Fellow, Singapore Institute of Engineering Technologists. He organized special sessions in the IEEE International Conference on Multimedia and Expo (ICME) in 2006 and 2012, the IEEE International Workshop on Multimedia Analysis and Processing in 2007, the IEEE International Symposium on Circuits and Systems (ISCAS) in 2010, the Pacific-Rim Conference on Multimedia (PCM) in 2009, the SPIE Visual Communications and Image Processing (VCIP) in 2010, the Asia Pacific Signal and Information Processing Association (APSIPA) in 2011, and the MobiMedia 2011. He gave invited, keynote, and panelist talks in International Workshop on Video Processing and Quality Metrics in 2006, the IEEE International Conference on Computer Communications and Networks in 2007, SPIE VCIP in 2010, and the IEEE Multimedia Communication Technical Committee (MMTC) Interest Group of Quality of Experience for Multimedia Communications in 2011, and tutorials in PCM 2007, PCM in 2009, IEEE ISCAS in 2008, IEEE ICME in 2009, APSIPA in 2010, and the IEEE International Conference on Image Processing in 2010. He is currently on the editorial boards of the IEEE Transactions on Multimedia, the IEEE Signal Processing LettersSignal Processing Letters, and the *Journal of Visual Communication and Image Representation*, and four IEEE Technical Committees. He co-chairs the IEEE MMTC Special Interest Group on Quality of Experience. He has been elected as a Distinguished Lecturer of APSIPA in 2012.

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

He is currently an Associate Professor with the Department of Electrical Engineering and the Institute of Communications Engineering, NTHU. He was with the Department of Computer Science and Information Engineering, National Chung Cheng University, Chiayi, Taiwan, from 2000 to 2007. Prior to joining academia, he worked for the Information and Communications Research Laboratories, Industrial Technology Research Institute, Hsinchu, from 1992 to 2000. His current research interests include video content analysis and video networking.

Dr. Lin is an Associate Editor of the IEEE Transactions on Circuits and Systems for Video Technology, the IEEE Transactions on Multimedia, the IEEE Multimedia, and the *Journal of Visual Communication and Image Representation*. He is an Area Editor of the *EURASIP Signal Processing: Image Communication*. He served as the Technical Program Co-Chair of the IEEE ICME in 2010, and Special Session Co-Chair of the IEEE ICME in 2009. He was a recipient of the Ph.D. Thesis Awards in 2001 presented by the Ministry of Education, Taiwan. His paper won the Young Investigator Award from SPIE VCIP in 2005. He received the Young Faculty Awards from CCU in 2005 and the Young Investigator Awards from National Science Council, Taiwan, in 2006.