Bottom-Up Saliency Detection Model Based on Human Visual Sensitivity and Amplitude Spectrum

Yuming Fang, Weisi Lin, *Senior Member, IEEE*, Bu-Sung Lee, *Member, IEEE*, Chiew-Tong Lau, *Member, IEEE*, Zhenzhong Chen, *Member, IEEE*, and Chia-Wen Lin, *Senior Member, IEEE*

***Abstract*—With the wide applications of saliency information in visual signal processing, many saliency detection methods have been proposed. However, some key characteristics of the human visual system (HVS) are still neglected in building these saliency detection models. In this paper, we propose a new saliency detection model based on the human visual sensitivity and the amplitude spectrum of quaternion Fourier transform (QFT). We use the amplitude spectrum of QFT to represent the color, intensity, and orientation distributions for image patches. The saliency value for each image patch is calculated by not only the differences between the QFT amplitude spectrum of this patch and other patches in the whole image, but also the visual impacts for these differences determined by the human visual sensitivity. The experiment results show that the proposed saliency detection model outperforms the state-of-the-art detection models. In addition, we apply our proposed model in the application of image retargeting and achieve better performance over the conventional algorithms.**

***Index Terms*—Amplitude spectrum, Fourier transform, human visual sensitivity, saliency detection, visual attention.**

I. Introduction

With the rapid increase in multimedia services, the efficient perceptual-aware image or video processing technology becomes more important for delivering high-quality images or videos. The saliency detection technologies, which exploit the most important areas for natural scenes, are very useful in practice, since they make the perceptual-friendly image or video processing possible by understanding the functionalities of the human visual system (HVS). The saliency detection technologies have already been used widely in many multimedia applications such as coding, retrieval, adaptation, and streaming [12], [14], [36], [38].

Manuscript received May 17, 2011; revised July 12, 2011; accepted September 12, 2011. Date of publication September 26, 2011; date of current version January 18, 2012. This work was supported in part by MoE AcRF Tire 2, Singapore, Grant Number: T208B1218. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Maja Pantic.

Y. Fang, W. Lin, B.-S. Lee, and C.-T. Lau are with the School of Computer Engineering, Nanyang Technological University, Singapore 639798 (e-mail: fa0001ng@ntu.edu.sg; wslin@ntu.edu.sg; ebslee@ntu.edu.sg; asctlau@ntu.edu.sg).

Z. Chen is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: zzchen@ntu.edu.sg).

C.-W. Lin is with the Department of Electrical Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan (e-mail: cwlin@ee.nthu.edu.tw).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2011.2169775

Visual attention is an important characteristic in the HVS and the research on visual attention has been reported in 1890 [1]. It is a cognitive process of selecting the relevant areas while acquiring the most significant information from the visual scene. Generally, the information captured by the human eyes is much more than that the central nervous system can process. When observers look at a scene, it is impossible for them to recognize all the objects and their relationships in the scene immediately. Thus, the selective attention will allocate processing resources to these salient areas rather than the entire scene [40], [41]. There are two different approaches in visual attention mechanism: bottom-up approach and top-down approach [2], [19], [42], [43]. Bottom-up approach, which is data-driven and task-independent, is a perception processing for automatic salient region selection for images. On the contrary, top-down approach is related to the recognition processing influenced by the prior knowledge such as tasks to be performed, the feature distribution of the target, the context of the visual scene, and so on [20]–[22].

In this paper, we focus on the bottom-up approach. During the past several decades, researchers have tried to understand the visual attention mechanism and developed computational models for predictions [2], [45], [46]. In the 1980s, Treisman *et al.* developed the well-known Feature-Integration Theory (FIT) [2]. According to this theory, some salient locations in a natural scene automatically stand out due to specific low-level features (such as color, intensity, orientation, and so on) when observers look at the visual scene. Many computational models of visual attention based on FIT have been proposed [3]–[11], [13]. Itti *et al.* devised a visual attention model based on the behavior and the neuronal architecture of the primates' early visual system [3]. This model obtains the feature maps through calculating the multi-scale center-surround differences for color, orientation, and intensity channels [3]. The final saliency map is achieved by the linear combination for these feature maps. Later, Harel *et al.* proposed a graph-based visual saliency (GBVS) model by using a better dissimilarity measure for saliency based on Itti's model [4]. In [5], Hou *et al.* devised a saliency detection model based on a concept defined as spectral residual (SR). Guo *et al.* later found that Hou's model was caused by phase spectrum and they designed a phase-based saliency detection model [6]. The model in [6] achieves the final saliency map by inverse Fourier transform (IFT) on a constant amplitude spectrum and the original phase spectrum of the image. Bruce *et al.* described visual attention based on the principle of maximizing information [7]. Liu *et al.* used the technology of machine learning to achieve the saliency map for images [9]. Gao

et al. calculated the center-surround discriminant for saliency detection [10]. The saliency value for a location is obtained by the power of a Gabor-like feature set to discriminate the center-surround visual appearance [10]. Gopalakrishnan *et al.* built a saliency detection model based on the color and orientation distributions in images [11]. Recently, a saliency detection model by Valenti *et al.* is advanced through calculating the center-surround differences of edges, color, and shape for images [13].

Most of these saliency detection models mentioned above obtain the saliency map for images by calculating the center-surround differences. They neglected some key characteristics of the HVS, such as the human visual sensitivity change due to foveation (i.e., the influence of an image patch decreases with the increase of the spatial distance). In this paper, we propose a novel saliency detection model based on the FIT and the human visual sensitivity variations. According to the FIT, the salient area in an image can be distinguished according to the differences of low-level features between this area and its neighbors. In our proposed model, we first divide the input image into small image patches and measure the saliency value for each image patch through calculating the differences of color, intensity, and orientation distributions between this image patch and all other patches (all the neighbor patches) in the image. Unlike existing methods which only consider local contrast or global contrast [3], [5], we exploit both local and global contrast by considering the differences between this patch and all the other image patches in the image. In addition, the contributions of these differences to the saliency value of image patches are different with the consideration of foveation behavior. We use the foveation-tuned human visual sensitivity to determine the weightage for these patch differences.

In essence, our proposed model first divides images into small image patches. Then it uses quaternion Fourier transform (QFT) instead of Fourier transform (FT) to obtain the amplitude spectrum of each image patch. Compared with FT which processes each feature channel of the color images separately, QFT allows color images to be transformed as a whole [24]. The saliency value of each patch is obtained by two factors: the differences of QFT amplitude spectrum between this image patch and other image patches in the whole image, and the weights for these patch differences determined by the human visual sensitivity. The novel saliency detection utilizes the characteristics of the HVS and is proven promising, as shown in Sections II–VI.

In this study, we also explore the application of the proposed saliency detection model in image retargeting. Recently, many efficient algorithms for image retargeting have been devised. Liu *et al.* introduced an image retargeting algorithm for mobile devices through generating the optimal browsing path [27]. Ren *et al.* designed an image retargeting algorithm based on global energy optimization [30]. Guo *et al.* proposed an image retargeting algorithm combining saliency map and object information to resize images [28]. Wang *et al.* presented an image retargeting algorithm based on saliency map and gradient map [37]. Recently, a popular image resizing technique called seam carving has been studied in [15] and [29]. In this study, we use the proposed saliency detection model in the framework of seam carving in [29]. Experimental results demonstrate the effectiveness of the proposed model in the application of image retargeting.

The rest of this paper is organized as follows. In Section II, we discuss the limitations of the relevant existing saliency detection models, and the contributions of the proposed model. In Section III, we describe the details of the proposed model. Section IV shows the experiment results by comparing the proposed model with other existing methods. Section V explores the application of the proposed model in image retargeting. The final section concludes the paper by summarizing our findings.

II. Limitations of the Most Relevant Models and Contributions of the Proposed Model

In the area of visual attention modeling, some studies [5], [6], [11] use the FT to get the final saliency map. As a basic transform, FT has been widely used in image processing with many applications such as convolutions, filtering, compression, and reconstruction. Because of the importance of the FT, a number of studies have been carried out to find what the Fourier phase and amplitude components represent in images. It is commonly accepted that the phase spectrum carries location information, while the amplitude spectrum includes the appearance and orientation information for visual scenes [17], [25], [32]. Based on this understanding, FT has been used in various studies of the human visual perception and understanding [5], [6], [11], [34], [35]. In [5] and [6], the saliency map is obtained based on the phase spectrum, whereas in [11], the amplitude spectrum of image patches is applied to obtain the orientation saliency submap. We analyze these two types of saliency detection models in details below.

In the phase-based saliency detection models [5], [6], the amplitude spectrum and phase spectrum are first obtained by FT. The saliency map is then calculated using inverse Fourier transform (IFT) on a user-defined constant amplitude spectrum and the original phase spectrum. This will amplify the intensity of the areas with less periodicity or less homogeneity and suppress the intensity of the areas with more periodicity or more homogeneity in the original images [5], [6]. In these approaches, the FT is operated on the whole images, and thus they mainly consider the contrast for images from the global perspective. These approaches suffer some defects in saliency detection. One problem is that these models cannot detect smooth-texture salient objects in the complex-texture background, as shown in the first row of Fig. 1(c), since the complex-texture areas are less homogeneous in those images. On the other hand, the phase-based saliency detection models have to resize the images into appropriately smaller sizes to allow the *major part* of a salient object to be less homogeneous, or they will just get the contour of the salient object, as shown in the second row of Fig. 1(b). Even when the original image is resized into a smaller one, the saliency map from phase-based models will still ignore some information of salient objects, as shown in the second row of Fig. 1(c). On the contrast, the proposed model can obtain the salient smooth-texture objects with the complex-texture background in images and will not

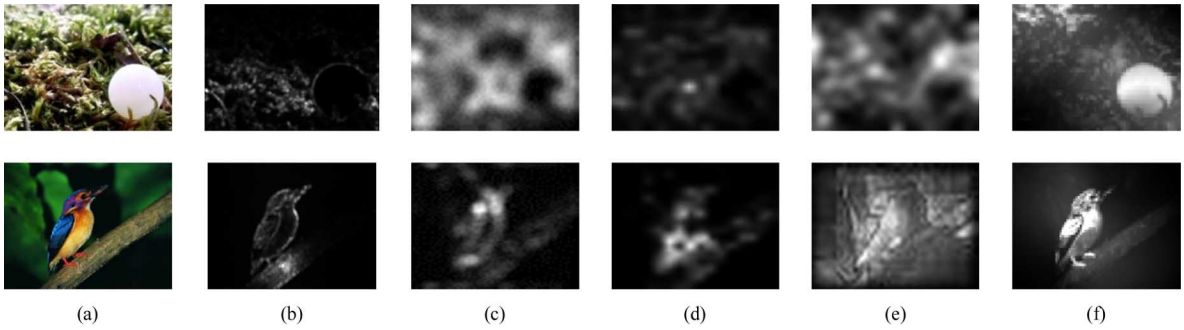

Fig. 1. Original images and the saliency maps. (a) Original images. (b) Saliency maps obtained from phase-based model without resizing the original images before FT. (c) Saliency maps obtained from phase-based model with resizing the original images into smaller ones (64 pixels for the input image width) [5]. (d) Saliency maps obtained from the saliency detection model in [11]. (e) Saliency maps obtained from the saliency detection model in [3]. (f) Saliency maps obtained from our proposed saliency detection model.

ignore much information of salient objects, as shown in the first and second rows of Fig. 1(f), respectively.

The study in [11] gets the final saliency map based on color and orientation distributions for images. The color and orientation saliency sub-maps are achieved separately by using two difference algorithms. The final saliency map will be selected as either the color saliency sub-map or the orientation saliency sub-map by identifying which of the sub-maps leads to the identification of the salient region [11]. This model uses amplitude spectrum to calculate the orientation distribution for images by computing the global orientation and orientation entropy contrast. One problem with this approach is that the orientation distribution in [11] is calculated by the histograms of 18 special orientations for image patches, which causes the loss of other orientation information. In addition, the final saliency map chosen as either color saliency sub-map or orientation saliency sub-map means that the final saliency map is determined by only color distribution or orientation distribution. Therefore, the saliency map from the model in [11] will lose much information of salient objects, as shown in Fig. 1(d). On the contrary, our proposed model uses the color, intensity, and orientation features together to get the final saliency map and it uses all the orientation information in the calculation of the final saliency map. Thus the obtained saliency map can better preserve the information of the salient objects [Fig. 1(f)].

Some classic visual attention models such as the model in [3] also use the low-level features including color, intensity, and orientation to calculate the saliency map for images. As these models calculate the multi-scale center-surround differences of low-level features from images to get the final saliency map, they mainly focus on the local contrast of low-level features for saliency detection [3]. One problem with the model in [3] is that it might regard the non-salient areas (such as areas in the background) as salient, as shown in Fig. 1(e), due to the lack of consideration for global characteristics in the image. In the saliency map from the first row of Fig. 1(e), some non-salient areas from the background are considered as salient areas. On the contrary, our proposed model considers both local and global contrast for images, and thus it can obtain much better saliency maps [Fig. 1(f)].

As can be seen from the analysis above, a key factor of successful saliency detection is the proper treatment of local and global information. Another important and related issue is how to combine (or pool) different features at a location and a feature from different locations. In [3] and [11], linear combinations are used; in [11], the Euclidian distances are used for the weighting of patch differences. However, there is lack of perceptual ground for all these approaches.

Compared with these saliency detection models discussed above, our proposed saliency detection model based on the human visual sensitivity and the amplitude spectrum achieves a higher accuracy in the detection of salient areas. The main contributions of our proposed model include the following: 1) we propose to divide an image into small image patches for local information extraction and combine information from different image patches in a global perspective; 2) we investigate into the visual impacts of the image patch differences based on the human visual sensitivity, which is a key characteristics of the HVS; 3) we utilize the amplitude spectrum of QFT to represent the color, intensity, and orientation distributions for image patches, to overcome the difficulty with linear or ad hoc feature combination/pooling in the existing models; 4) we exploit the characteristics of the HVS to determine the patch size and perform the multi-scale operations. Different from our initial work [18], we use the human visual sensitivity to determine the weights of image patch differences and justify the effectiveness of the proposed scheme with more aspects and applications. In addition, the choice of patch size and multi-scale operation are justified in this paper. We will explain the proposed model in details in Section III.

III. Saliency Detection Model Based on Human Visual Sensitivity and Amplitude Spectrum

In this section, we describe the proposed model in details. As mentioned above, the model first divides each original input image into small image patches for gathering local information. For simplicity, we take image patches from the original images and do not do any preprocessing operation for extracting image patches. In our work, the patch size is chosen as 8×8 and partially overlapping. Here we select the size of image patches based on the fovea size. Details of defining image patches will be given in Section III-D. The saliency value for each image patch is obtained through calculating the QFT amplitude spectrum differences between a patch and its neighbor patches, and

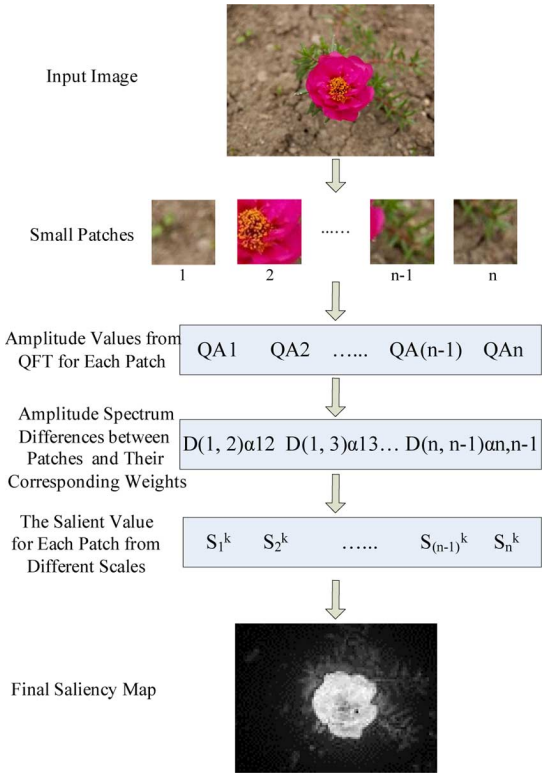

Fig. 2. Framework of the proposed saliency detection model.

the weights for these differences determined by the human visual sensitivity. The proposed model is illustrated as Fig. 2. We will describe the details step by step in Sections III-A–III-D.

A. Saliency Value for Each Patch

In the proposed model, the saliency value of each image patch is determined by two factors: one is the patch differences between this image patch and all other image patches in the input image; the other is the weighting for these patch differences. If these differences between an image patch and all other image patches are big, then the saliency value for this image patch is large. In addition, we take the influence of the foveation behavior into consideration in the proposed model. Here, we use $\mathcal{D}_{(i,j)}$ to represent the difference between image patch i and image patch j, the saliency value for image patch i can be expressed as follows:

$$S_i = \sum_{j \neq i} \alpha_{ij} \mathcal{D}_{(i,j)} \tag{1}$$

where α_{ij} is the weight for the patch difference between image patches i and j, which is determined by the human visual sensitivity.

It is generally believed that the HVS is highly space-variant because the retina in the human eye has different density of cone photoreceptor cells [26]. On the retina, the fovea owns the highest density of cone photoreceptor cells. Thus, the focused area has to be projected on the fovea to be perceived at the highest resolution. The density of the cone photoreceptor cells becomes lower with larger retinal eccentricity. Therefore,

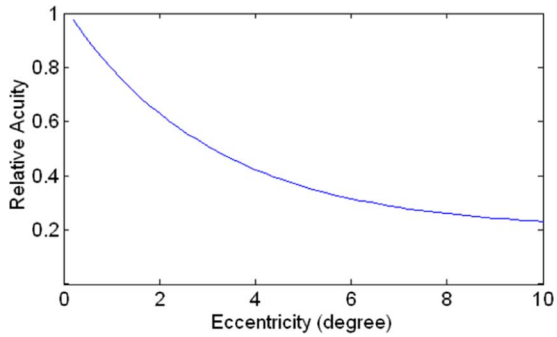

Fig. 3. Relationship between visual acuity and eccentricity [47].

the visual sensitivity decreases with the increased eccentricity from the fixation point, as shown in Fig. 3 [26], [47], [48].

As to the saliency value of image patch i in (1), all patch differences between the image patch i and the other image patches are considered and summed together. We use the human visual sensitivity to determine the weights for the patch differences. In this study, the eccentricity from the center of the fixation (the center of the image patch i) is not directly used as a weighting factor for calculating the saliency value of the image patch i but a weighting factor for calculating the importance of patch-difference pairs. Here, the weights for the patch differences are determined by the human visual sensitivity, and this means that the weights of the patch differences from its nearer neighbor patches (with smaller eccentricities) are larger compared with these from farther neighbor patches. With larger eccentricity of image patches from the image patch i (which means farther image patches from the image patch i), the visual sensitivity decreases and thus the weighting for the patch differences between these image patches and image patch i becomes smaller. Therefore, the contributions of the patch differences to the saliency value of image patch i will decrease with larger-eccentricity image patches from image patch i. On the contrary, the contributions of the patch differences to the saliency value of the image patch i will increase with smaller-eccentricity image patches from image patch i. This is reasonable, as the human eyes are more sensible to the patch differences from nearer image patches compared with those from farther image patches. Our proposed saliency detection model takes both local and global center-surround differences into account, for it uses the patch differences from all other image patches in the image to calculate the saliency value of image patch i. We will describe how to get the α_{ij} and $\mathcal{D}_{(i,j)}$ in details in Sections III-B–III-D.

B. Amplitude Spectrum for Each Image Patch

In the proposed model, we use the color and intensity channels for QFT to get the amplitude spectrum for each image patch, which is used to compute the differences between image patches. As we know, the amplitude spectrum indicates the presence of the respective spatial frequencies and their strengths can represent the orientation distribution in images [25]. Thus, the amplitude spectrum of QFT can represent the color, intensity, and orientation distributions for image patches when we use the color and intensity channels as the input into QFT. Then

the differences between amplitude spectrums of QFT for image patches can show the differences for color, intensity, and orientation distributions between image patches. Here we use opponent color space to represent the color information for image patches. If r, g, and b denote the red, green, and blue color components, four broadly-tuned color channels are generated as $R = r - (g + b)/2$ for red, $G = g - (r + b)/2$ for green, $B = b - (r + g)/2$ for blue, and $Y = (r + g)/2 - |r - g|/2 - b$ for yellow. Each color channel is then decomposed into red-green and blue-yellow double opponency according to the related property of the human primary visual cortex [23]:

$$\mathcal{C}_{\mathcal{RG}} = R - G \tag{2}$$

$$\mathcal{C}_{\mathcal{BY}} = B - Y. \tag{3}$$

The intensity channel can be computed as $\mathcal{J} = (r + g + b)/3$. We use one intensity channel $\mathcal{J}$, and two color channels $\mathcal{C}_{\mathcal{RG}}$ and $\mathcal{C}_{\mathcal{BY}}$, as three features for calculating the amplitude spectrum of QFT. Based on the three features, the quaternion representation for each image patch is as follows:

$$q(n,m) = \mathcal{J}(n,m)\mu_1 + \mathcal{C}_{\mathcal{RG}}(n,m)\mu_2 + \mathcal{C}_{\mathcal{BY}}(n,m)\mu_3 \tag{4}$$

where μ_1, μ_2, and μ_3 are unit pure quaternion; $\mu_1^2 = \mu_2^2 = \mu_3^2 = -1$; $\mu_1 \perp \mu_2$, $\mu_2 \perp \mu_3$, $\mu_1 \perp \mu_3$, and $\mu_3 = \mu_1\mu_2$.

The symplectic decomposition for the above quaternion image patch is given by

$$q(n,m) = f_1(n,m) + f_2(n,m)\mu_2 \tag{5}$$

$$f_1(n,m) = \mathcal{J}(n,m)\mu_1 \tag{6}$$

$$f_2(n,m) = \mathcal{C}_{\mathcal{RG}}(n,m) + \mathcal{C}_{\mathcal{BY}}(n,m)\mu_1. \tag{7}$$

The study of [24] indicates that the QFT can be calculated by using two standard complex fast Fourier transform. The QFT of $q(n,m)$ in (5) can be computed as follows:

$$\mathcal{Q}[u,v] = F_1[u,v] + F_2[u,v]\mu_2 \tag{8}$$

$$F_i[u,v] = \frac{1}{\sqrt{MN}} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \times e^{-\mu_1 2\pi((mv/M)+(nu/N))} f_i(n,m) \tag{9}$$

where $i \in \{1,2\}$; (n,m) and (u,v) are the locations for image patches in spatial and frequency domains, respectively; N and M are the height and width of image patches; $f_i(n,m)$ is obtained from (6) and (7).

According to (4)–(9), we can get the QFT result $\mathcal{Q}[u,v]$ for each image patch. Now we describe $\mathcal{Q}[u,v]$ in polar form as follows:

$$\mathcal{Q}[u,v] = \mathcal{A}e^{\mu\varphi} \tag{10}$$

where $\mathcal{A}$ is the QFT amplitude spectrum of the image patch; φ is the corresponding QFT phase spectrum; μ is a unit pure quaternion.

Actually, the FT amplitude spectrum $\mathcal{A}$ can be calculated as

$$\mathcal{A} = |\mathcal{Q}[u,v]|. \tag{11}$$

From (11), we can get the amplitude spectrum of QFT for each image patch, to be used to represent each image patch. In this study, we use the color and intensity channels for QFT, so this amplitude spectrum from QFT includes color information as well as intensity information. As we know, the amplitude spectrum indicates the presence of the respective spatial frequencies in images. The value of the amplitude spectrum in a special direction of the center-shifted FT indicates the strength of the orientation in a perpendicular direction [11], [25]. We can get the strength of different orientations through using the amplitude spectrum of the center-shifted FT, as shown in [11]. Thus, the amplitude spectrum used here can represent the orientation distribution of image patches. As we use the color and intensity channels for QFT in this study, the QFT amplitude spectrum can represent color, intensity, and orientation distributions for image patches.

C. Differences Between Image Patches and Their Weighting to Saliency Value

As described previously, the saliency value for each image patch is determined by the weighted differences between this patch and its patch neighbors including all other image patches in the image. If an image patch is significantly different from its neighbors, it has a higher probability to be a salient area. The saliency value for an image patch should be larger with the larger differences between this patch and its neighbors. As the spatial distance (eccentricity) between the patch and its neighbor increases, the weight of this difference to the saliency value of the patch decreases. The saliency value for an image patch is calculated according to (1). Now we discuss how $\mathcal{D}_{(i,j)}$ in (1) is obtained.

We use the Euclidian distance of the amplitude spectrum of QFT to represent the differences between each patch and its neighbors. To reduce the dynamic range of the amplitude coefficients, we use logarithm operation and add the constant 1 to each original amplitude coefficient value to avoid the undefined case when $\mathcal{A}$ approaches zero. Using this algorithm, we can calculate the difference between image patches i and j as

$$\mathcal{D}_{(i,j)} = \sqrt{\Sigma_m \left(\log\left(\mathcal{A}_m^i + 1\right) - \log\left(\mathcal{A}_m^i + 1\right)\right)^2} \tag{12}$$

where m indexes all pixels in an image patch.

We use the visual sensitivity to determine the weights of the QFT amplitude spectrum differences between image patches. Here we use the model developed in [47] to measure the human contrast sensitivity as a function of eccentricity. The contrast sensitivity $C_s(f,e)$ is defined as the reciprocal of the contrast threshold $C_t(f,e)$ as follows:

$$\mathcal{C}_s(f,e) = \frac{1}{\mathcal{C}_t(f,e)}. \tag{13}$$

According to the model in [47], the contrast threshold is defined as

$$\mathcal{C}_t(f,e) = \mathcal{C}_0 \exp\left(\alpha f \frac{e + e_2}{e_2}\right) \tag{14}$$

where f is the spatial frequency (cycles/degree), e is the retinal eccentricity (degree); C_0 is the minimum contrast threshold; α is

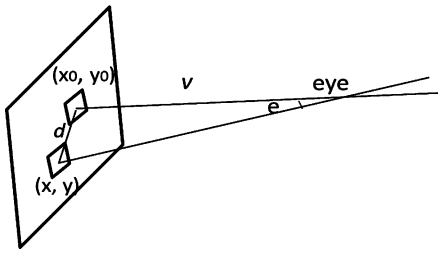

Fig. 4. Relationship between viewing distance and retina eccentricity.

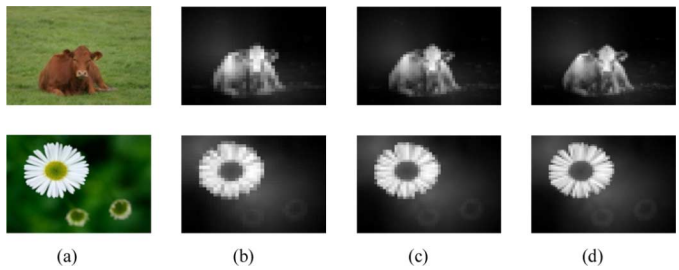

Fig. 5. Original images and its different saliency maps with different patch sizes. (a) Original images. (b) Saliency maps with the image patch, the eccentricity of whose width is 0.25 degree. (c) Saliency maps with the image patch, the eccentricity of whose width is 0.2 degree. (d) Saliency maps with the image patch, the eccentricity of whose width is 0.15 degree.

the spatial frequency decay constant; e_2 is the half-resolution eccentricity. According to the experiments reported in [47], these parameters are set to $C_0 = 1/64$, $\alpha = 0.106$, and $e_2 = 2.3$.

We can calculate the retina eccentricity e according to its relationship with viewing distance v as Fig. 4. Given the position of the fixation point (x_0, y_0) (the center of an image patch), the retinal eccentricity e for the position (x, y) (the center of another image patch) can be computed as follows:

$$e = \tan^{-1}\left(\frac{d}{v}\right) \tag{15}$$

where d is the Euclidian distance between (x, y) and (x_0, y_0).

The typical ratio of the viewing distance to the picture height is in the range of 3 to 6 [49]. Here we use a ratio of 4 to determine the viewing distance.

Thus, we can get the weight α_{ij} as the normalized $C_s(f, e)$ based on (13)–(15). The weighting parameters α_{ij} in (1) can be calculated as follows:

$$\alpha_{ij} = \frac{1}{\mathcal{C}_0 \exp\left(\alpha f \frac{e+e_2}{e_2}\right)}. \tag{16}$$

From the description above, the saliency value for the image patch i is represented as all the contributions from the patch differences between the image patch i and all other image patches in the image, as calculated in (1).

D. Patch Size and Scale for Final Saliency Value

In this study, the final saliency map is influenced by the image patch size. The conventional computational visual attention models [7], [11] choose a fixed patch size empirically. In this paper, we consider the characteristics of the HVS and the fovea size. Given an image patch with the size $p \times p$, the relationship between the eccentricity e and the viewing distance v can be computed as follows:

$$e = \tan^{-1}\left(\frac{p}{2v}\right). \tag{17}$$

Studies show that the 1 to 2 degree retinal area in the fovea is with the best visual acuity and the parafovea surrounding the fovea has lower visual acuity [39]. Here we use e_0 to represent the eccentricity for the best visual acuity, which is set as 1 degree. We set $e = \beta e_0$, where $\beta < 1$ to make sure that with e, good visual acuity is maintained. As mentioned previously, the typical viewing distance is 3 to 6 times of the image height. Here we set the view distance as 4 times of the image height, while setting β as 0.2. Setting $\beta = 0.2$ means that the maximum eccentricity for the width of the image patch is 0.2 degree, which can guarantee that the whole image patch is in area with the best visual acuity. In addition, for better effect, we divide the input images into partially overlapping image patches, which is determined by the overlap-eccentricity $\gamma\beta e_0$. We choose the parameter $\gamma = 0.5$.

As we can see, the saliency values for all pixels in an image patch obtained based on $\mathcal{D}_{(i,j)}$ and α_{ij} are similar. Thus, the patch size would influence the final saliency map. With a smaller patch size, the final saliency map will become more distinguishable, as shown in Fig. 5 where the saliency map with the smallest image patch size (the eccentricity of 0.15 degrees) is more distinguishable than the other two with larger patch sizes. Of course, to obtain more accurate saliency map, we can divide images into smaller image patches with larger overlapping; however, in this situation, the computational complexity will increase. Given an input image with size of $M \times N$ (where M is the width and N is the height): with the patch size of $k \times k$, the computational complexity of the proposed algorithm is $(M*N)^2/((1-\lambda)^2 k)^2$ with λ overlapping. Therefore, with the smaller patch size or more overlapping, the computational complexity will increase. Thus, we choose the suitable patch size to compute the saliency map based on the consideration of fovea characteristics, saliency detection performance, and the computational complexity.

Except the patch size, the scale will also influence the final saliency map. In the saliency map, the saliency value for the image patches from salient areas are much higher than that for these patches belonging to background. For the images with different scales, the saliency values of background are similarly low, while the saliency values of salient areas are high. Thus, using multi-scale can strengthen the saliency for these salient areas. We adopt the steerable pyramid algorithm [33] to get multi-scale images. This algorithm obtains multi-scale images through low-pass filtering and subsampling the input image. For simplicity, here we use the linear combination to obtain the final saliency map. Thus, the saliency value for pixel i is expressed as follows:

$$\mathcal{S}_i = \frac{1}{N}\Sigma_k \mathcal{S}_i^k \tag{18}$$

where N is the scale number; S_i^k is the saliency value for image patch i in the kth scale. The image with the lowest scale level should not be too small for a good performance of the final saliency map. Our experiments show that the lowest scale level should not be smaller than one fourth of the original scale. Therefore, we use 3 different scales to get the final saliency map: the original scale, a half of the original scale, and one fourth of the original scale.

IV. Evaluations

From the discussion in Section II, we know that there are some defects for the saliency maps from the existing saliency detection models. We have also compared the performance of the proposed method with others in Fig. 1. As a further experiment, we give a quantitative evaluation of the saliency map obtained from our proposed model and other relevant saliency detection models on a public database.

Saliency map can give the salient areas for images, which can provide the locations for salient object candidates. One efficient quantitative evaluation method for saliency detection algorithms is to detect salient objects for natural images. Many studies have used saliency map to detect objects for natural images [5], [6], [9], [44]. The quantitative evaluation of this experiment is based on a database of 5000 images from Microsoft database [9]. This image database includes the original images and their corresponding ground-truth indicated with bounding boxes by 9 subjects. We calculate the ground-truth saliency map for images by averaging the 9 users' labeled-data (similar with [9]). Thus, the quantitative evaluation for a saliency detection algorithm is to see how much the saliency map from the algorithm overlaps with the ground-truth saliency map. Here, we use the precision, recall, and F-measure to evaluate the performance of our proposed model. Precision is computed as the ratio of correctly detected saliency region to the detected salient region from the saliency detection algorithm. Recall is calculated as the ratio of correctly detected salient region to the ground-truth salient region. Given a ground-truth saliency map G and the detected saliency map S for an image, we have

$$\text{precision} = \frac{\Sigma_x g_x s_x}{\Sigma_x s_x} \tag{19}$$

$$\text{recall} = \frac{\Sigma_x g_x s_x}{\Sigma_x g_x}. \tag{20}$$

F-measure, a harmonic mean of precision and recall, is a measure that combines precision and recall. It is calculated as follows:

$$F_\alpha = \frac{(1+\alpha) * precision * recall}{\alpha * precision + recall} \tag{21}$$

where α is a positive parameter to decide the importance of precision over recall in computing the F-measure.

Generally, the precision indicates the performance of the saliency detection algorithms compared with ground-truth saliency map. To compare the proposed model with others, we always see the precision value for different algorithms, for the precision value is the ratio of the correctly detected region over the whole detected region. We set $\alpha = 0.3$ in this experiment as in [11] for fair comparison. The comparison results are shown in Fig. 6. Here we use the original experiment results of other models including the models [5], [11], [31], [44] from [11]. The model in [44] is an improved saliency detection model based on [3]. In Fig. 6, we can see that the overall performance of our proposed model for the 5000 images is better than the others under comparison in terms of all three measures.

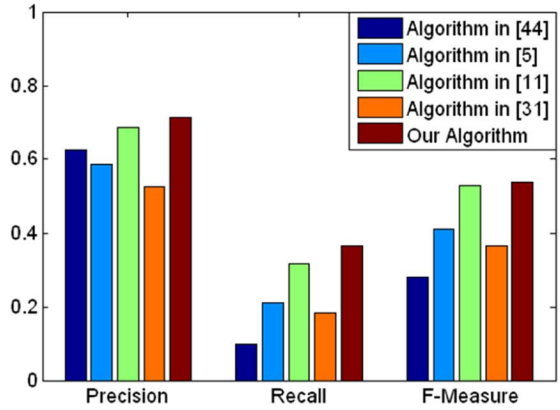

Fig. 6. Experiment results for the comparison between our proposed model and others.

In Fig. 7, we show some comparison samples of saliency maps from our proposed model and the others. From this figure, we can see that the saliency maps from the proposed model are better than those from other existing ones. From fifth and sixth rows in Fig. 7, all other models just detect the contour of the salient objects, while the proposed model can detect the whole salient objects exactly.

V. Application: Image Retargeting

Saliency map plays an important role in many image processing applications such as image retargeting [14], [28], [38]. With different screen resolutions in devices such as smart phones and PDAs, the displayed image has to be resized to fit for different display sizes and aspect ratios. An effective image retargeting algorithm should preserve the visually important content without much distortion to the image context. The saliency map, which can detect the salient areas for images, measures the visual importance for image pixels in image retargeting for better representations [28], [30], [37]. In this work, we show how the better detected saliency map can improve the image retargeting performance and therefore demonstrate the effectiveness of the proposed saliency detection algorithm.

The performance of image retargeting algorithms greatly depends on the visual significance map, which is used to measure the visual importance for each pixel in images. The visual significance map used in the existing image retargeting algorithms includes the gradient map, the saliency map, and some high-level feature maps such as the facial map [15], [16], [27]–[30]. Here, we apply the saliency map from the proposed model in the framework of [29] to demonstrate the effectiveness of the proposed model in the application of image retargeting.

In this experiment, we compare the performance of the proposed image retargeting algorithm and three existing image retargeting techniques [16], [29], [30] based on the Microsoft

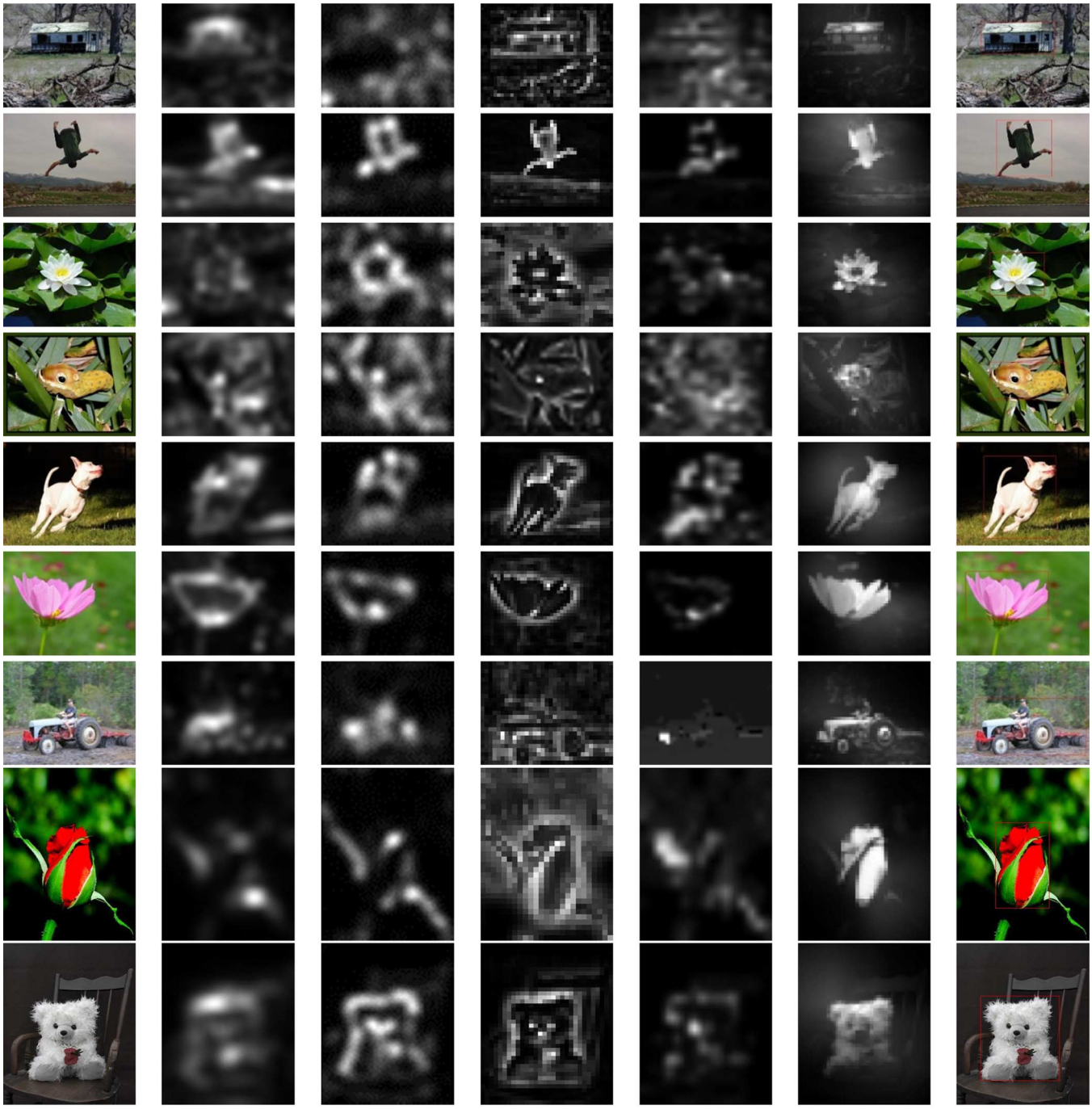

Fig. 7. Saliency maps from different saliency detection models. (a) Original images. (b) Saliency maps from Itti's model [3]. (c) Saliency maps from Hou's model [5]. (d) Saliency maps from Ma's model [31]. (e) Saliency maps from Gopalakrishnan's model [11]. (f) Saliency maps from the proposed model. (g) Human-labeled maps from 9 subjects.

database [9]. We found that our image retargeting algorithm using the saliency map from the proposed model outperforms the others greatly for the images with complex background; while the performance of our proposed image retargeting algorithm is similar with the others for other images with simple background. The reason is that the saliency map from the proposed saliency detection model measures the importance of each pixel in images with complex background more accurately compared with the visual significance map from other algorithms, as shown in the first row of Fig. 8. For the images with simple background, the visual significance maps used in other image retargeting algorithms can get similar result with the saliency map from our proposed model, as shown in the second row of Fig. 8.

To better compare the performance of different visual significance maps and saliency maps in image retargeting, we have conducted a user study based on two image datasets: one includes 23 images with complex background, while the other

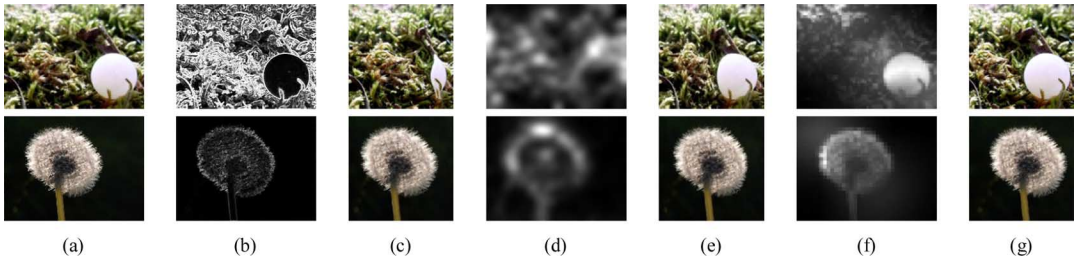

Fig. 8. Original images, visual significance maps, and retargeted images from different algorithms. (a) Original images. (b) Gradient maps from algorithm [29]. (c) Retargeted images from algorithm [29]. (d) Saliency maps from Itt's model [3]. (e) Retargeted images from [30] (Itti's model is integrated into this algorithm to measure the importance of each pixel for images). (f) Saliency maps from the proposed model. (g) Retargeted images from our image retargeting algorithm based on the proposed saliency detection model.

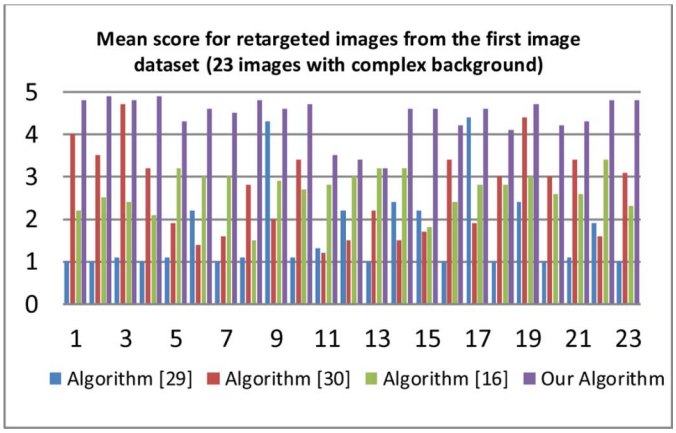

Fig. 9. Detailed statistical results of the mean score for each retargeted image from ten participants for four different algorithms for the first image dataset.

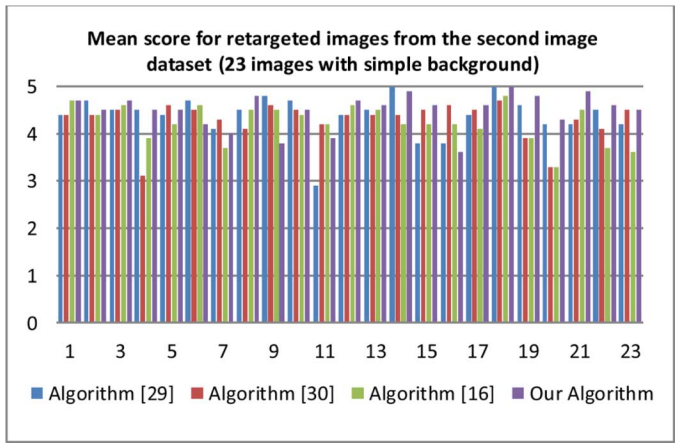

Fig. 10. Detailed statistical results of the mean score for each retargeted image from ten participants for four different algorithm for the second image dataset.

includes 23 images with simple background. All these 46 images are selected from Microsoft database [9]. Three existing image retargeting algorithms [16], [29], [30] are utilized for performance comparison. Ten participants (three female and seven male) were involved in this experiment. The experiment was conducted in the typical laboratory environment. The original image was displayed in the middle of the display screen as the reference image, while four retargeted images from four different algorithms were displayed in random orders surrounding the reference image. Mean opinion scores (1–5) were recorded by participants where 1 means bad viewing experience and 5 means excellent viewing experience. Each participant voted for these 46 images. The mean score for each retargeted image from ten participants for four different algorithms are shown in Figs. 9 and 10, respectively, for these two image datasets.

Fig. 9 shows that the mean scores of the retargeted images from the proposed algorithm are much higher than the mean scores from other algorithms for the first dataset. Fig. 11 shows the overall mean scores of the 23 retargeted images in the first dataset for four different algorithms. These two figures demonstrate that the performance of the proposed algorithm is the best and the performance of algorithms [16] and [30] are better than that from algorithm [29] for the images with complex background in the first image dataset. In Fig. 10, it might be noted that the results of the four algorithms are close for the images with simple background in the second dataset. Fig. 12 shows the overall mean scores of 23 images in the second dataset for four different image retargeting algorithms. In this figure, it can be

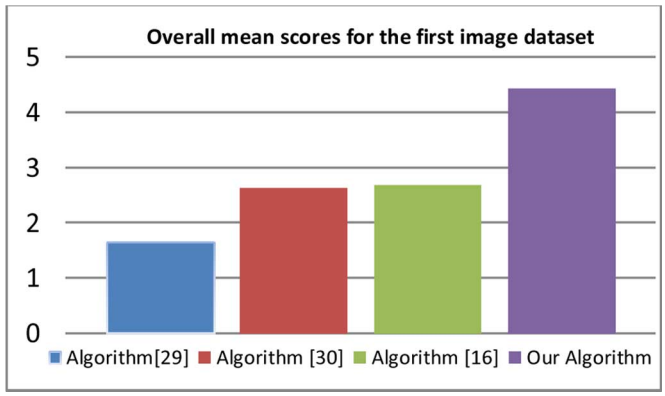

Fig. 11. Overall mean scores of 23 retargeted images from four different algorithms for the first image dataset.

seen that the overall mean score of the retargeted images from our algorithm is slightly higher than those from the other algorithms. As the overall mean scores from four algorithms are close to each other, the overall performances of four algorithms are comparable for the images with simple background.

Overall, Figs. 9–12 show that the image retargeting algorithm based on our proposed saliency detection algorithm outperforms the others. Some retargeted images from the two datasets are depicted in Figs. 13 and 14. As can be seen in Fig. 13, the retargeted images from the other three existing image retargeting algorithms suffer some distortion in salient objects, whereas the retargeted images from our algorithm preserve

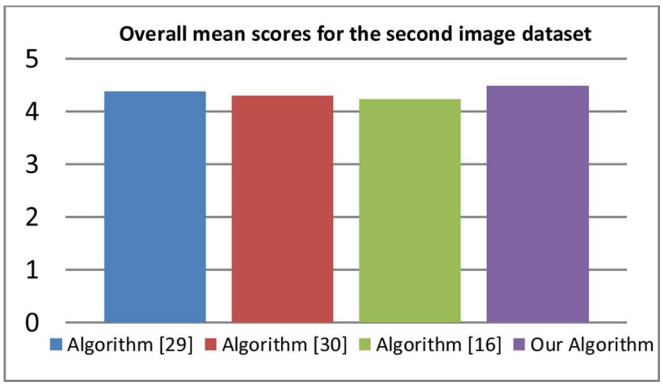

Fig. 12. Overall mean scores of 23 retargeted images from four different algorithms for the second image dataset.

Fig. 13. Retargeted image samples from the first image dataset (images with complex background) for four algorithms. (a) Original images. (b) Retargeted images from [29]. (c) Retargeted images from [30]. (d) Retargeted images from [16]. (e) Retargeted images from our algorithm. The sizes of the original images and the retargeted images are 300×400 and 300×300, respectively.

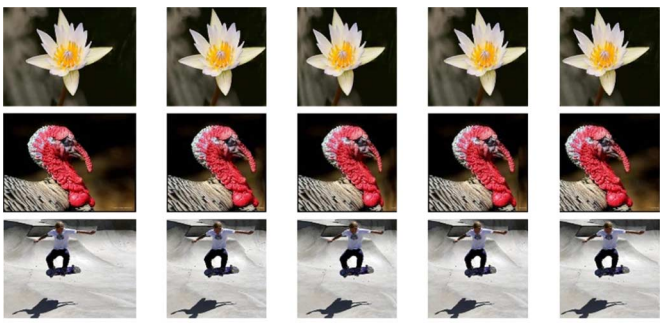

Fig. 14. Retargeted image samples from the second image dataset (images with simple background) for four algorithms. (a) Original images. (b) Retargeted images from [29]. (c) Retargeted images from [30]. (d) Retargeted images from [16]. (e) Retargeted images from our algorithm. The sizes of the original images and the retargeted images are 300×400 and 300×300, respectively.

the salient objects accurately. In Fig. 14, it can be seen that the retargeted images from all four algorithms are somewhat similar and without much distortion. More comparison results can be found at: https://sites.google.com/site/leofangyuming/Home/saliencydetection.

VI. Conclusions

In this paper, we proposed a novel bottom-up saliency detection model based on both local and global feature contrast, the human visual sensitivity and QFT amplitude spectrum. The proposed model first divides the input images into small image patches. It then uses the QFT amplitude spectrum to represent the color, intensity, and orientation distributions for image patches. The saliency value for each image patch is obtained by computing the differences between the QFT amplitude spectrum of this patch and all other patches in the image, and the weights for these differences determined by the visual impacts of the human visual sensitivity. The proposed saliency detection model also utilizes the characteristics of the HVS for the selection of patch size and multi-scale operations. Compared with the existing models, the proposed model has better performance with regard to the ground truth of human-labeled salient objects. In addition, we demonstrate the advances of our proposed saliency detection model in image retargeting. Experiment results from a user study show that the results of the image retargeting algorithm based on our proposed model are better than those of other existing ones.

References

[1] W. James, *The Principles of Psychology*. Cambridge, MA: Harvard Univ. Press, 1890.
[2] A. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognit. Psychol.*, vol. 12, no. 1, pp. 97–136, 1980.
[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] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," *Adv. Neural Inf. Process. Syst.*, vol. 19, pp. 545–552, 2006.
[5] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, 2007.
[6] C. Guo, Q. Ma, and L. Zhang, "Spatio-temporal saliency detection using phase spectrum of quaternion Fourier transform," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, 2008.
[7] N. D. Bruce and J. K. Tsotsos, "Saliency based on information maximization," *Adv. Neural Inf. Process. Syst.*, vol. 18, pp. 5–162, 2006.
[8] L. Itti and C. Koch, "Computational modeling of visual attention," *Nature Rev. Neurosci.*, vol. 2, no. 3, pp. 194–203, Mar. 2001.
[9] T. Liu, J. Sun, N. Zheng, X. Tang, and H. Y. Shum, "Learning to detect a salient object," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, 2007.
[10] D. Gao and N. Vasconcelos, "Bottom-up saliency is a discriminant process," in *Proc. IEEE Int. Conf. Computer Vision*, 2007.
[11] V. Gopalakrishnan, Y. Hu, and D. Rajan, "Salient region detection by modeling distributions of color and orientation," *IEEE Trans. Multimedia*, vol. 11, no. 5, pp. 892–905, Aug. 2009.
[12] L. Itti, "Automatic foveation for video compression using a neurobiological model of visual attention," *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1304–1318, Oct. 2004.
[13] R. Valenti, N. Sebe, and T. Gevers, "Image saliency by isocentric curvedness and color," in *Proc. IEEE Int. Conf. Computer Vision*, 2009.
[14] L. Q. Chen, X. Xie, X. Fan, W. Y. Ma, H. J. Zhang, and H. Q. Zhou, "A visual attention model for adapting images on small displays," *Multimedia Syst.*, vol. 9, no. 4, pp. 353–364, 2003.
[15] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," *ACM Trans. Graph.*, vol. 26, no. 3, pp. 267–276, 2007.
[16] L. Wolf, M. Guttmann, and D. Cohen-Or, "Nonhomogeneous content-driven video-retargeting," in *Proc. IEEE Int. Conf. Computer Vision*, 2007, pp. 1–6.
[17] A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," *Proc. IEEE*, vol. 69, pp. 529–541, 1981.
[18] Y. Fang, W. Lin, B. Lee, C. Lau, and C. Lin, "Bottom-up saliency detection model based on amplitude spectrum," in *Proc. Int. Conf. Multi-Media Modeling*, 2011.
[19] J. M. Wolfe, S. J. Butcher, and M. Hyle, "Changing your mind: On the contributions of top-down and bottom-up guidance in visual search for feature singletons," *J. Experiment. Psychol.: Human Percept. Perform.*, vol. 29, pp. 483–502, 2003.

[20] 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.
[21] 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.
[22] C. Kanan, M. Tong, L. Zhang, and G. Cottrell, "SUN: Top-down saliency using natural statistics," *Visual Cognit.*, vol. 17, no. 6, pp. 979–1003, 2009.
[23] S. Engel, X. Zhang, and B. Wandell, "Colour tuning in human visual cortex measured with functional magnetic resonance imaging," *Nature*, vol. 388, no. 6,637, pp. 68–71, Jul. 1997.
[24] T. Ell and S. Sangwin, "Hypercomplex Fourier transforms of color images," *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 22–35, Jan. 2007.
[25] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Englewood Cliffs, NJ: Prentice-Hall.
[26] B. A. Wandell, *Foundations of Vision*. Sunderland, MA: Sinauer Associates, 1995.
[27] H. Liu, X. Xie, W.-Y. Ma, and H.-J. Zhang, "Automatic browsing of large pictures on mobile devices," in *Proc. ACM Int. Conf. Multimedia*, 2003, pp. 148–155.
[28] Y. Guo, F. Liu, J. Shi, Z.-H. Zhou, and M. Gleicher, "Image retargeting using mesh parametrization," *IEEE Trans. Multimedia*, vol. 11, no. 5, pp. 856–867, Aug. 2009.
[29] M. Rubinstein, A. Shamir, and S. Avidan, "Improved seam carving for video retargeting," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 1–9, 2008.
[30] T. Ren, Y. Liu, and G. Wu, "Image retargeting based on global energy optimization," in *Proc. IEEE Int. Conf. Multimedia and Expo*, 2009, pp. 406–409.
[31] Y. F. Ma and H. J. Zhang, "Contrast-based image attention analysis by using fuzzy growing," in *Proc. ACM Int. Conf. Multimedia*, 2003, pp. 374–381.
[32] L. N. Piotrowski and F. W. Campbell, "A demonstration of the visual importance and flexibility of spatial-frequency amplitude and phase," *Perception*, vol. 11, pp. 337–346, 1982.
[33] H. Greenspan, S. Belongie, R. Goodman, P. Perona, S. Rakshit, and C. H. Anderson, "Overcomplete steerable pyramid filters and rotation invariance," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, Seattle, WA, Jun. 1994, pp. 222–228.
[34] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelop," *Int. J. Comput. Vision*, vol. 42, pp. 145–175, 2001.
[35] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Optic. Soc. Amer.*, vol. 4, no. 12, pp. 2379–2394, 1987.
[36] 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.
[37] Y.-S. Wang, C.-L. Tai, O. Sorkin, and T.-Y. Lee, "Optimized scale-and-stretch for image resizing," *ACM Trans. Graph.*, vol. 27, no. 5, pp. 1–8, 2008.
[38] H. Liu, S. Jiang, Q. Huang, C. Xu, and W. Gao, "Region-based visual attention analysis with its application in image browsing on small displays," in *Proc. ACM Int. Conf. Multimedia*, 2007.
[39] M. A. Just and P. A. Carpenter, *The Psychology of Reading and Language Comprehension*. Newton, MA: Allyn & Bacon, 1987.
[40] H. Pashler, *The Psychology of Attention*. Cambridge, MA: MIT Press, 1997.
[41] , H. Pashler, Ed., *Attention*. Hove, U.K.: Psychology Press, 1988.
[42] L. Itti, "Models of bottom-up and top-down visual attention," Ph.D. dissertation, Dept. Comput. Neural Syst., California Inst. Technol., Pasadena, 2000.
[43] J. Braun and D. Sagi, "Vision outside the focus of attention," *Percept. Psychophys.*, vol. 48, no. 1, pp. 45–48, 1990.
[44] D. Walther and C. Koch, "Modeling attention to salient proto-objects," *Neural Netw.*, vol. 19, pp. 1395–1407, 2006.
[45] J. Wolfe, K. R. Cave, and S. L. Franzel, "Guided search: An alternative to the feature integration model for visual search," *J. Experiment. Psychol.: Human Percept. Perform.*, vol. 15, no. 3, pp. 419–433, 1989.
[46] J. Wolfe, "Guided search 2.0: A revised model of visual search," *Psychonom. Bull. Rev.*, vol. 1, no. 2, pp. 202–238, 1994.
[47] W. S. Geisler and J. S. Perry, "A real-time foveated multi-solution system for low-bandwidth video communication," in *Proc. SPIE*, Jul. 1998, vol. 3299, pp. 294–305.
[48] Z. Chen and C. Guillemot, "Perceptually-friendly H.264/AVC video coding based on foveated just-noticeable-distortion model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 6, pp. 806–819, Jun. 2010.
[49] ITU, Methodology for the Subjective Assessment of the Quality of Television Pictures, Geneva, Switzerland, 2002, ITU-R BT.500-11.

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 Beijing University of Technology, Beijing, China, in 2009. He is currently pursuing the Ph.D. degree from the School of Computer Engineering, Nanyang Technological University, Singapore.

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

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, in 1982 and 1985, respectively, and the Ph.D. degree in computer vision from King's College, London University, London, U.K., in 1992.

He taught and conducted research at Zhongshan University, Shantou University (China), Bath University (U.K.), the National University of Singapore, the Institute of Microelectronics (Singapore), and the Institute for Infocomm Research (Singapore). He has been the Project Leader of 13 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 in the School of Computer Engineering, Nanyang Technological University, Singapore. His areas of expertise include image processing, perceptual modeling, video compression, multimedia communication, and computer vision. He holds ten patents, edited one book, authored one book and five book chapters, and has published over 160 refereed papers in international journals and conferences. He believes that good theory is practical, so has kept a balance of academic research and industrial deployment throughout his working life.

Dr. Lin is a fellow of Institution of Engineering Technology (IET) and a Chartered Engineer (U.K.). He organized special sessions in IEEE ICME06, IEEE IMAP07, IEEE ISCAS10, PCM09, SPIE VCIP10, APSIPA11, and MobiMedia11. He gave invited/keynote/panelist talks in VPQM06, IEEE ICCCN07, SPIE VCIP10, and IEEE MMTC QoEIG (2011), and tutorials in PCM07, PCM09, IEEE ISCAS08, IEEE ICME09, APSIPA10, and IEEE ICIP10. He currently serves on the editorial boards of the IEEE TRANSACTIONS ON MULTIMEDIA, IEEE SIGNAL PROCESSING LETTERS, and *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 on Technical Program Committees and/or Organizing Committees of a number of international conferences.

Bu-Sung Lee (M'08) received the B.Sc. (Hons) and Ph.D. degrees from the Electrical and Electronics Department, Loughborough University of Technology, Loughborough, U.K., in 1982 and 1987, respectively.

He is an Associate Professor with the Nanyang Technological University, Singapore. He is active in the development of the Research and Education Network in Singapore and the Asia Pacific Region, e.g., Trans-EuroAsia Network. His research interest covers grid/cloud computing, wireless communication, and multimedia communication.

Chiew-Tong Lau (M'84) received the B.Eng. degree from Lakehead University, Thunder Bay, ON, Canada, in 1983, and the M.A.Sc. and Ph.D. degrees in electrical engineering from the University of British Columbia, Vancouver, BC, Canada, in 1985 and 1990, respectively.

He is currently an Associate Professor in the School of Computer Engineering, Nanyang Technological University, Singapore. His research interests include wireless communications systems and signal processing.

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

He is currently a Lee Kuan Yew Research Fellow and Principal Investigator at Nanyang Technological University (NTU), Singapore. Before joining NTU, he was an ERCIM fellow at 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. 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) and an invited member of the IEEE MMTC Interest Group of Quality of Experience for Multimedia Communications (QoEIG) (2010–2012). He has served as a guest editor of special issues for *IEEE MMTC E-letter* and *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, ICME, etc. He received the CUHK Faculty Outstanding Ph.D. Thesis Award, Microsoft Fellowship, and the ERCIM Alain Bensoussan Fellowship. He is a member of SPIE.

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

He is currently an Associate Professor with the Department of Electrical Engineering, NTHU. He was with the Department of Computer Science and Information Engineering, National Chung Cheng University (CCU), Taiwan, during 2000–2007. Prior to joining academia, he worked for the Information and Communications Research Laboratories, Industrial Technology Research Institute (ICL/ITRI), Hsinchu, Taiwan, during 1992–2000, where his final post was Section Manager. He has authored or coauthored over 100 technical papers. He holds more than 20 patents. His 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 and the *Journal of Visual Communication and Image Representation*, and an Area Editor of *EURASIP Signal Processing: Image Communication*. He has served as a Guest Co-Editor of four special issues for the IEEE TRANSACTIONS ON MULTIMEDIA, the *EURASIP Journal on Advances in Signal Processing*, and the *Journal of Visual Communication and Image Representation*. 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. He was a recipient of the 2001 Ph.D. Thesis Awards presented by the Ministry of Education, Taiwan. His paper won the Young Investigator Award presented by SPIE 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.