Image Modification Based on Spatial Frequency Components for Visual Attention Retargeting

Hironori TAKIMOTO†(a), Member, Syuhei HITOMI†, Nonmember, Hitoshi YAMAUCHI†, Mitsuyoshi KISHIHARA†, and Kensuke OKUBO†, Members

SUMMARY  It is estimated that 80% of the information entering the human brain is obtained through the eyes. Therefore, it is commonly believed that drawing human attention to particular objects is effective in assisting human activities. In this paper, we propose a novel image modification method for guiding user attention to specific regions of interest by using a novel saliency map model based on spatial frequency components. We modify the frequency components on the basis of the obtained saliency map to decrease the visual saliency outside the specified region. By applying our modification method to an image, human attention can be guided to the specified region because the saliency inside the region is higher than that outside the region. Using gaze measurements, we show that the proposed saliency map matches well with the distribution of actual human attention. Moreover, we evaluate the effectiveness of the proposed modification method by using an eye tracking system.

key words: visual attention, visual saliency map, attention retargeting, wavelet analysis

1. Introduction

It is well known that 80% of the information entering the human brain is visual information [1], [2]. A human shifts his/her gaze by employing visual information obtained through his/her peripheral vision when deciding where to direct his/her attention [3], [4]. A system for supporting human activity requires a visual attention retargeting method that naturally guides human gaze to a region of interest (ROI) in order to realize natural interaction between the system and the human. The usability of such a system can be improved if the system enables a user to easily find and access the required information by guiding his/her gaze. Attentive user interfaces (AUIs) that use optical head-mounted displays, such as Google Glass, are attracting increasing interest. Effective AUIs naturally attract human visual gaze to the region specified by the gaze-based interface. Therefore, it is widely believed that drawing human gaze to particular objects can effectively facilitate and direct many types of human activities (e.g., driving, manufacturing, and studying) as well as human-machine interfaces and other applications, including education, gaming, design, and media arts [5].

Humans choose important information from an enormous volume of visual information; this is called “visual attention.” Visual saliency is defined as an estimation of how likely a given region is to attract human visual attention, and there is much evidence indicating a correlation between visual attention and saliency maps [6]. Recently, several image modification methods that use visual saliency maps to guide human attention to a ROI have been proposed. These methods are categorized into two types: color-based methods and orientation-based methods.

Color-based methods modify color components such that the visual saliency inside an ROI increases whereas that outside the ROI decreases. In general, color-based methods can guide a viewer’s attention to the ROI while keeping the rest of the image at a high resolution. Hagiwara et al. [7] and Kokui et al. [8] proposed image editing methods in order to obtain an image in which the given region has the highest saliency. These methods adaptively modify the RGB color components by reverse engineering a typical visual saliency map model. Takimoto et al. [9] proposed a modification method that iteratively modulates the lightness and color components on the basis of a novel saliency map using an L*a*b* color space. This method requires significantly less computational time than other methods and produces images that are considered much more natural. The above-mentioned methods [7]–[9] are able to guide human attention effectively because they are strictly based on a visual saliency computation model for bottom-up attention, i.e., they are categorized under “feedback from saliency computation.”

On the other hand, Kim et al. [10] proposed a manipulation method as a typical orientation-based method by using bilateral displacements for sharpening and smoothing operations to alter the frequencies in the geometry. Mateescu et al. [11] analyzed how manipulating the orientation of an ROI of an image affects human visual attention. Specifically, they rotated the ROI in order to maximize its saliency. However, it is difficult to apply these orientation-based methods to ROIs with low texture. The most significant drawback of these methods is that it is difficult to effectively guide attention to a modified ROI if other regions with high saliency exist, because these methods manipulate only the features of the ROI. Hata et al. [12] obtained a modified image by using a Gaussian filter to blur each pixel outside the ROI. They analyzed how blurring an image affects human visual attention. In other words, they suppressed the high-frequency components by adopting spatial filtering to control the resolution. However, it is difficult to di-
rectly guide a viewer’s gaze toward the ROI after the first saccade, because this method uniformly applies a Gaussian filter to various frequency components without using a visual saliency model. In addition, although many color-based methods guide visual attention by controlling the saliency in an image on the basis of existing saliency map models, there is no orientation-based method that is strictly based on a saliency model for bottom-up attention.

In this paper, we propose an orientation-based attention retargeting method to obtain an image in which the ROI achieves the highest saliency in order to guide human attention. The proposed method is based on reverse engineering of a visual saliency model for bottom-up attention. It consists of two phases: visual saliency analysis and frequency component modulation. First, a novel visual saliency map based on frequency components is obtained by octave division using frequency analysis. The obtained visual saliency map matches well with the distribution of actual human attention. Next, the proposed method iteratively modulates each frequency component such that the visual saliency outside the ROI decreases. When this modification method is applied to an image, human gaze is guided to the ROI because the saliency inside the ROI increases while that outside the ROI decreases.

2. Related Work

Various studies have investigated visual attention retargeting. A traditional attention retargeting approach that is used with TV is to present visual stimuli, such as arrows [3] or bounding boxes, in the peripheral visual field. Bailey et al. [13] proposed “subtle gaze direction” by performing brief image-space modulations on the ROI of human peripheral vision in order to direct human gaze. However, these traditional approaches are more coercive than effective from the viewer’s standpoint. A better approach would be to smoothly and effectively direct human attention toward an ROI without impeding the current visual attention.

A human detects the next target to be closely observed by using the information in his/her peripheral vision and then confirms the details of the target in the central visual field. The region having the most remarkable features is observed closely in the field of vision, if factors such as the viewer’s interest and the task are not considered. Bottom-up attention induced by visual features obtained from a visual stimulus dominantly influences visual attention in the early stages, i.e., immediately after the visual stimulus is presented [14], [15]. Therefore, attention retargeting methods that modify visual features on the basis of bottom-up visual saliency have been investigated in several studies.

In a typical bottom-up saliency map model, the saliency in a given region is determined primarily by how different the region is from its surroundings in terms of features such as color, orientation, depth, and motion. Hence, conventional attention retargeting methods based on a bottom-up saliency map model for a static image are categorized into two types: color-based methods and orientation-based methods.

Color-based methods modify each color component such that the visual saliency inside an ROI increases whereas that outside the ROI decreases. The main advantage of this approach is that it generally guides attention to the ROI while keeping the rest of the image at a high resolution. Veas et al. [16] proposed a saliency modulation technique that prompts attention shifts and influences the recall of an ROI without perceptible changes in the visual input. Mendez et al. [17] proposed a method for dynamically directing human gaze by analyzing and modulating the bottom-up salient features. The ROI is adaptively darkened, lightened, and manipulated in hue according to local contrast information rather than global parameters. Although these methods show better results than the traditional approaches, they require a threshold map to be manually set for each image. Hagiwara et al. [7] and Kokui et al. [8] proposed image editing methods to obtain an image in which the ROI has the highest saliency. These methods adaptively modulate the RGB color components by reverse engineering a visual saliency map model proposed by Itti et al. [18]. Although such methods can guide human gaze to the ROI, false borders that cause discomfort to users are generated in the modified image. Nguyen et al. [19] proposed a novel computational framework that actively recolors only the ROI in order to make it stand out, in both the local and the global sense. This method uses a salient patch dataset that includes the recorded fixation data. Then, through graph-based optimization, the color is transferred from the salient patches to the ROI such that the saliency of the ROI maximized. The disadvantages of this method are that it is computationally expensive and requires access to a sufficiently comprehensive gaze dataset. Mateescu et al. [20] proposed a saliency manipulation method that modifies the hue while keeping the intensity and chromaticity constant. However, although this method modifies only the hue, the original hue of the ROI is not considered. In addition, when other objects with high saliency exist in an original image, it is difficult to effectively guide human gaze to the modified ROI. Takimoto et al. [9] used a novel saliency analysis method and color modulation to create modified images in which the ROI is the most salient region in the entire image. The saliency map model used in their saliency analysis reduces the computational cost and improves the naturalness of the image by using an L*A*B* color space and simplified normalization.

The above-mentioned methods [7]–[9] are strictly based on a visual saliency computation model for bottom-up attention; thus, they are able to guide human attention effectively. Although state-of-the-art color-based method [9] achieves attention retargeting without degradation of visibility, it is not necessarily effective for all images. In addition, both color- and orientation-based methods can be applied to an image simultaneously because the characteristics used by each method are different. A more robust attention retargeting is realized by applying both methods in combination. Therefore, it is important to realize an efficient orientation-based method.
The aim of this paper is to develop an effective orientation-based attention retargeting method that is strictly based on a bottom-up computational model of visual attention. To guide human gaze to an ROI, the saliency inside the ROI should increase while that outside the ROI should decrease, as shown in Fig. 1. Figure 1 (c) shows the result of applying the proposed method to the image shown in Fig. 1 (a). The lower right bowl, which has relatively low saliency, is chosen as an ROI. Figures 1 (b) and (d) show the saliency maps of the original and modified images, respectively, in which brighter regions have higher saliency values. Therefore, our objective is to propose an image modification method that is strictly based on an orientation-based saliency map model in order to obtain a modified image that matches the desired saliency as the inverse of the ideal saliency map in which the ROI has the highest saliency.

Our method consists of two phases: visual saliency analysis and image modulation. For visual saliency analysis, several orientation-based saliency map models have been proposed by extracting other features, such as spatial and transform domains [18], [23]–[26]. However, solving the above-mentioned saliency inversion problem for effective attention retargeting requires a simple computational model of a saliency map that estimates human visual attention with high accuracy. Therefore, we propose a novel visual saliency model using the wavelet transform (WT) for bottom-up attention. For image modulation, we propose an image modulation algorithm that maximizes the saliency inside an ROI.

The procedure of the proposed method is summarized in Algorithm 1. First, a user selects an ROI from an original image to which a viewer’s attention should be guided. Next, the WT is applied to the target image in order to create multi-scale feature maps that can represent different features (e.g., edges and texture). The WT can achieve multi-scale spatial and frequency analysis at the same time [27]. The orientation-based saliency map is computed by applying the center-surround difference operation to the obtained multi-scale feature maps. Then, the wavelet coefficients of the target image are modified such that the saliency inside the ROI is maximized. After the inverse WT (IWT) operation, the

---

**Algorithm 1**: Attention retargeting based on visual saliency map

1. Select an ROI from an original image.
2. Compute the saliency based on frequency analysis.
3. Modify the target image such that the saliency in the ROI becomes higher.
4. Check whether the ROI has the highest saliency.
   - If the ROI does not have the highest saliency, go to Step 2.
   - Apply a Gaussian filter to the modified image to remove high-frequency noise. Output the modified image.

---

(a) Original image  
(b) Saliency map of (a)  
(c) Our result  
(d) Saliency map of (c)

**Fig. 1**: Example of image and saliency map in proposed method
ROI of the original image is overwritten with the modified image. If the ROI of the overwritten image has the highest saliency in the entire modified image, the algorithm is completed by applying a Gaussian filter to the modified image in order to remove high-frequency noise. Otherwise, the target image is modulated by iteratively applying saliency analysis and modifying the wavelet coefficients.

The novel orientation-based saliency model and the saliency-map-based image modulation algorithm for frequency components are described in Sects. 4 and 5, respectively.

4. Saliency Map Based on Frequency Components

Itti et al. [18] proposed a visual saliency computation model based on the early vision model proposed by Koch and Ullman [1]. Using human gaze measurements, they demonstrated that their saliency map matches well with the distribution of actual human attention. Their algorithm obtains a saliency map based on the intensity, color, and orientation conspicuity maps obtained by cross-scale addition of the feature maps. This means that adjusting the features of the entire image on the basis of a saliency map can attract the attention of a user to a specified region. Therefore, a modified image matching the desired target saliency is created from a given original image and the desired target saliency map. Although this can be regarded as a saliency inversion problem, it is an ill-posed problem because there is no one-to-one mapping between saliency and images. In this study, a modified image that matches the desired target saliency map is created by an iterative optimization procedure based on saliency map calculation.

Itti et al. [18] obtained a saliency map based on orientation features via spatial filtering using a Gabor filter. Imamoglu et al. [23] proposed a novel bottom-up saliency model that consists of local and global saliency maps created from multi-scale features maps based on wavelet coefficients. However, it is difficult to apply such saliency maps, which are based on a complex calculation model or spatial filtering, to the above-mentioned saliency inversion problem for attention retargeting. Hence, we propose a novel visual saliency model using the WT for bottom-up attention.

Wavelet decomposition has the advantage of extracting the orientation details from a multi-scale perspective. It offers high spatial resolution with higher frequency components and low spatial resolution with lower frequency components without information loss (i.e., loss of details) during the decomposition process [27].

The framework of the proposed saliency model based on the WT is shown in Fig. 2. First, the WT is applied to each RGB color channel to obtain the sub-bands of the input image I. The obtained sub-band for each color channel \( I^\alpha \) (\( \alpha \in \{R, G, B\} \)) is defined by

\[
[A^\alpha_N, H^\alpha(s), V^\alpha(s), D^\alpha(s)] = WT_N(I^\alpha)
\]

where \( N \) is the maximum number of scaling iterations for the WT; \( s \ (s \epsilon \{1, 2, \ldots, N\}) \) is a resolution index; \( A \) is the approximation output at the coarsest resolution; and \( H, V, \) and \( D \) are the wavelet coefficients of the horizontal, vertical, and diagonal directions, respectively. We use the Daubechies wavelet (Daub.3) as the mother wavelet for the WT, and we set \( N = 6 \) as the maximum number of scaling iterations for the WT. Wavelet coefficient pyramids \((H^\alpha(s), V^\alpha(s), D^\alpha(s))\) for each color channel are used for saliency detection.

The center-surround difference operation, which is implemented by the pixel difference between the finer scale \( c \) and the coarser scale \( s \), is applied to these pyramids for the creation of feature maps. We set \( c = 1, 2, 3 \) and \( s = c + d \), with \( d = 2, 3 \). Each feature map \( FM_k \ (k \epsilon \{H, V, D\}) \) is defined as follows:

\[
FM_k(c, s) = ||k^c(c)|| \oplus ||k^s(s)||
\]

where \( \oplus \) denotes the corresponding pixel-wise subtraction between the two scaled images. Next, the feature maps are normalized and combined into three conspicuity maps \( CM_k^\alpha \).

\[
CM_k^\alpha = \bigoplus_{c=1}^{3} \bigoplus_{s=c+2}^{3} N(FM_k^\alpha(c, s))
\]

where \( \bigoplus \) denotes the corresponding pixel-wise summation between the two feature maps. In addition, \( N \) is a normalization operator. Simple max-local normalization [18] is applied to each conspicuity map to highlight the most discriminative feature within the map.

Then, the three conspicuity maps are combined into a saliency map \( SM^\alpha \) for each color channel.

\[
SM^\alpha = \frac{1}{3} (N(CM^H_k) + N(CM^V_k) + N(CM^D_k))
\]
Finally, the three saliency maps $SM^a$ for each color channel are combined into a master saliency map $SM$.

$$SM = \frac{1}{3}(SM^R + SM^G + SM^B)$$  \hspace{1cm} (5)

5. Image Modification for Guiding Visual Attention

The basic concept of our image modulation method is that the saliency inside an ROI increases while that outside the ROI decreases by iteratively modulating the wavelet coefficients. First, a user selects an ROI to which a viewer’s attention should be guided. If $k^a_{(i,j)}(s)$ holds for all $k \in \{H, V, D\}$, and $a \in \{R, G, B\}$ is the wavelet coefficient pyramid for each color channel in the target image $I'$ at pixel $(i, j)$, then $k^a_{(i,j)}(s)$ is the corresponding coefficient in the modified image $I'^+$. Here, $t = 0$ denotes the original image.

Each wavelet coefficient is then modified as follows:

$$k^a_{(i,j)}(s) = k^a_{(i,j)}(s) + SM^a \cdot P^a_{(i,j)} \cdot (k^a_{(i,j)}(s) - \bar{k}^a_{(i,j)}) \hspace{1cm} (6)$$

where $\bar{k}^a_{(i,j)}$ is the average of the wavelet coefficients at pixel $(i, j)$ in color channel $a$, defined by

$$\bar{k}^a_{(i,j)} = \frac{1}{N} \sum_{s=1}^{N} |k^a_{(i,j)}(s)| \hspace{1cm} (7)$$

where $N$ is the maximum number of scaling iterations for the WT. If the absolute value of wavelet coefficient $|k^a_{(i,j)}(s)|$ is higher than the average of wavelet coefficients $\bar{k}^a_{(i,j)}$ at pixel $(i, j)$, the wavelet coefficient should be suppressed because it is the primary factor in increasing the saliency at pixel $(i, j)$.

Note that $SM^a$ in Eq. (6) is the saliency value of color channel $a$ in the input image $I'$ at pixel $(i, j)$. The saliency value is used as a weight coefficient to efficiently suppress the wavelet coefficients at a pixel with high saliency.

The intensity coefficient $P^a_{(i,j)}$, which is the weight of the modulation value of each wavelet coefficient, is defined by

$$P^a_{(i,j)} = \frac{N(CM^a_{H})}{(N(CM^a_{H}) + N(CM^a_{V}) + N(CM^a_{D}))} \hspace{1cm} (8)$$

Note that $P^a_{(i,j)}$ reflects the influence of a given feature on the saliency, and it enables us to obtain (by back calculation) the saliency map.

Next, a reconstructed image $\hat{I}$ is created by applying the inverse WT (IWT) to each modified color sub-band. Then, the modified image is created by applying the ROI of the original image $I^0$ to the reconstructed image $\hat{I}$.

After image composition of the reconstructed image $\hat{I}$ and the ROI of the original image $I^0$, the saliency map $SM'$ is calculated. If the maximum saliency $SM'$ outside the ROI is lower than an arbitrary threshold $Th$, then the iterative modification of the wavelet coefficients is completed.

Finally, we can obtain the modified image $I'$ by applying a 2D Gaussian low-pass filter to the reconstructed image $\hat{I}$ for the removal of false edges generated by modifying the wavelet coefficients.

$$I' = \hat{I} * g_{\text{norm}} \hspace{1cm} (9)$$

where $g_{\text{norm}}$ is the $m \times m$ 2D Gaussian filter and $*$ denotes the convolution operation.

6. Experiments and Results

We conducted experiments to evaluate visual saliency analysis and image modulation for attention retargeting.

6.1 Evaluation of Visual Saliency Analysis

6.1.1 Experimental Setup

To demonstrate the effectiveness of the proposed saliency map model for estimation of visual attention, we evaluate the proposed method using public datasets CAT2000 [28]. Although this dataset contains two sets of images, the training set and the test set, we only use the training set. The training set has 20 different categories (100 images for each category) and fixations for 18 observers.

6.1.2 Experimental Results and Discussion

For comparison, a typical saliency map used in a conventional method [18] was employed. It should be noted that Itti’s saliency map consists of three conspicuity maps based on intensity, color, and orientation. In addition, we also used a conspicuity map of orientation as a reference saliency map for comparison because our saliency map focuses only on the orientation component.

The evaluation results in terms of the normalized scanpath saliency (NSS) and percentile are listed in Table 1. In this table, “Conv. method (Case 1)” refers to Itti’s conspicuity map of only orientation, and “Conv. method (Case 2)” refers to Itti’s saliency map comprising three conspicuity maps. Specifically, the average NSS and percentile for all test images and participants are listed. NSS measures the performance of saliency models using fixations. First, the saliency scores of all the regions in an image are normalized to have zero mean and unit standard deviation [29]; then, the saliency scores at fixed locations are used to measure the model performance. On the other hand, the percentile (first proposed in [30]) measures the percentage of fixations whose predicted saliency values fall below the value of a fixed location [31].

From Table 1, it can be seen that the NSS and percentile values of the proposed saliency map are greater than those of the conventional saliency map using only orientation features. Using the results of this experiment, we performed a pairwise t-test (two-tailed) to compare the saliency map models statistically. Thus, we confirmed that the differences
Table 1  NSS, percentile, and pairwise t-test (two-tailed) values for visual saliency

|                  | NSS | Percentile |
|------------------|-----|------------|
| Prob. method     | 0.781 | 0.708     |
| Conv. method (Case 1) | 0.751 | 0.699 |
| t-value (p-value) | 3.36 (<0.001) | 3.87 (<0.001) |
| Conv. method (Case 2) | 1.014 | 0.743 |
| t-value (p-value) | 23.75 (<0.001) | 19.62 (<0.001) |

between the methods based on orientation features are indeed statistically significant.

The accuracy of the proposed saliency map model is higher than that of the traditional orientation-based saliency map model based on spatial filtering on the results of saliency map evaluation. However, state-of-the-art saliency map models based on various features have been proposed recently [32]–[34]. It should be noted that we focus not only on the accuracy of the saliency map model but also on the ease of solution of the inverse problem. Therefore, it is possible to apply the proposed saliency map to our image modification method for attention retargeting.

6.2 Evaluation of Image Modulation

6.2.1 Experimental Setup

Using a gaze measurement system, we confirmed that the modified image obtained by the proposed method guides attention effectively. We used a QG-PLUS tracking device (Ditect Inc.) to track the eye movements of our participants. Each participant's head was fixed at a distance of 65 cm from the front of the display. We defined a fixation as a set of consecutive data points within a certain range of the visual angle for a minimum of 100 ms. The radius of the visual angle for fixation was 1.3°.

Fifteen subjects (all male, 21–36 years old, average age = 23.7 years), each with normal color vision, participated in this experiment. Further, fifteen images, each having a size of 640 × 480 pixels, were used. Fifteen images were obtained from the photo database [35] and the internet. Examples of the obtained images are shown in Fig. 3.

The proposed method and two reference methods were applied to all the original images. For comparison, we used the orientation-based method proposed by Hata et al. [12] and the color-based modification method proposed by Takimoto et al. [9] in order to evaluate how well each method guides visual attention.

We minimized the influence of previous observation of an image of the same scene by a participant. Further, we created 60 test images from the 15 original images (3 modified images per original image). These 60 images were divided into 4 groups, each having 15 different scenes. The images in each group were shown in random order. In addition, we controlled the intervals between the groups to reduce the influence of the previous group. For the ROI, we selected objects with relatively low saliency located away from the center of the image.

Using the concept of SDOF, Hata et al. [12] achieved attention retargeting by blurring the region outside an ROI using a Gaussian filter. Therefore, to fairly compare the effectiveness of attention retargeting, we determined the optimum Gaussian filter parameter $\sigma$ for each original image. The procedure for determining the optimum parameter $\sigma$ is summarized in Algorithm 2. In Step 4, if the following condition is satisfied, the saliency suppression is judged to have converged.

$$SM_{ROI}(I'_{o,r}) - SM_{ROI}(I'_{o,r}) \leq SM_{MAX}/100 \quad (10)$$

where $SM_{MAX}$ is the maximum saliency in the original image, $I'_o$ is the modified image obtained using a Gaussian filter.
filter with parameter $\sigma_t$, and $SM_{\text{ROI}}(I_{\sigma_t})$ is the maximum saliency outside the ROI in image $I$ modified with parameter $\sigma_t$.

On the other hand, in our method, the saliency outside the ROI is gradually suppressed by iterative calculation of the saliency map and image modification. Hence, when the maximum saliency outside the ROI in the image modified by our method is less than the maximum saliency outside the ROI in the image modified by the method of Hata et al. [12] with optimum parameter $\sigma_t$, our iterative calculation is com-
6.2.2 Experimental Results and Discussion

Examples of the original images and modified images obtained using the proposed method and reference methods [9], [12] as well as their respective saliency maps are shown in Fig. 4. These saliency maps were computed by the proposed visual saliency model using the WT.

The circular object in the upper part of the image was chosen as the ROI of the image in the first row. Further, the pear in the upper right part of the image and the picture on the wall were chosen as the ROIs of the images in the third and fifth rows, respectively. In the saliency maps of the modified images obtained using the method of Hata et al. [12], the saliency inside the ROI was slightly greater than that of the original images. However, it is difficult to guide attention to the ROI because there are other regions having high saliency. In the method of Takimoto et al. [9], the orientation-based saliency is not suppressed effectively because this method controls the intensity and color on the basis of a color-based saliency map. The detailed results for the third row in Fig. 4 are listed in Table 2. Specifically, the average and maximum saliency inside and outside the ROI of the original and modified images are listed. From these results, we confirmed that saliency outside the ROI can be suppressed most effectively by the proposed method. However, result of the proposed method is less natural in terms of its visibility when compared with other methods.

Examples of scanpaths obtained from the same participant using the gaze tracking system are shown in Fig. 5, where the numbers in the circles denote the order of fixation movement and the size of the circles denote the duration of fixation. In Fig. 5, the original image hardly attracts any attention to the ROI. From the first to third rows, according to the results of the conventional methods, the participants were slightly attracted to the ROIs. On the other hand, the image modified by our method effectively attracted the participants’ attention at an early stage. In addition, we confirmed that the duration of fixation at the ROI was longer than that at other regions. In the fourth and fifth rows, the contrast of color outside of ROI is salient. Therefore, it is difficult for orientation-based methods to guide a first-saccade to the ROI. However, the duration of fixation inside the ROI is longer than that outside the ROI. In the sixth row, the image modulated by Hata’s method blurred the original image significantly. Therefore, there is almost no difference between our proposed method and Hata’s method. Hence, our method is not necessarily effective for all images as compared with color-based method, and vice versa. Nevertheless, we confirmed that results obtained by our method are sufficient when compared with the orientation-based method with spatial filtering.

The rates of fixation movements occurring prior to reaching the ROI for all the datasets are listed in Table 3. The proposed method guides gaze to the specified regions more effectively than the reference methods. Using the results of these experiments, we performed a pairwise t-test (two-tailed) to compare the methods statistically, as shown in Table 4. Thus, we confirmed that the differences between the methods are indeed statistically significant.

Our method is more effective than the reference methods because of its optimum suppression of wavelet coefficients on the basis of the visual saliency map. Thus, the saliency outside the ROI is efficiently suppressed; consequently, the ROI attracts greater attention. The conventional orientation-based method [12] based on spatial filtering fails

| Table 2 | Example of saliency analysis inside/outside the ROI |
| Ave. inside ROI | Ave. outside ROI | Max. inside ROI | Max. outside ROI |
| Original | 60 | 52 | 221 | 255 |
| Prop. | 64 | 36 | 255 | 140 |
| Conv. [12] | 70 | 48 | 255 | 180 |
| Conv. [9] | 57 | 47 | 189 | 255 |

| Table 3 | Fixation movement rates |
| % | Original image | Prop. method | Conv. method [12] | Conv. method [9] |
| 1st | 3.1 | 53.1 | 30.4 | 16.9 |
| 2nd | 4.4 | 25.4 | 25.0 | 32.0 |
| 3rd | 8.9 | 8.5 | 18.8 | 15.1 |
| 4th | 9.8 | 3.1 | 10.7 | 7.1 |
| 5th | 6.2 | 3.1 | 4.9 | 6.7 |
| other | 67.6 | 6.7 | 10.3 | 22.2 |

| Table 4 | Results of pairwise t-test (two-tailed) |
| (p-value) | Prop. method | Conv. method [12] | Conv. method [9] |
| Orig. | 24.28 (<0.001) | 17.61 (<0.001) | 13.91 (<0.001) |
| Prop. | - | 4.90 (<0.001) | 8.27 (<0.001) |
| Conv. [12] | - | - | 4.20 (<0.001) |

Algorithm 2: Determination of optimum parameter $\sigma$ in [12]

Step 1: Select an ROI for an arbitrary original image.
Step 2: Compute the saliency based on frequency analysis.
Step 3: Set parameter $\sigma_1 = 1$.
Step 4: Apply the Gaussian filter to the original image outside the ROI.
Step 5: Check whether saliency suppression outside the ROI has converged.

if Saliency suppression outside the ROI has converged then $\sigma_t$ is output as the optimum parameter for the image.
else $\sigma_{t+1} = \sigma_t + 1$. Go to Step 4.
end if
to sufficiently suppress the saliency outside the ROI. The proposed method strongly blurs the region outside the ROI because the effectiveness of attention retargeting is related to the naturalness of the modified image. Therefore, it is necessary to determine the termination condition of iterative image modification for different purposes in accordance with the goal. Most importantly, the proposed method, which is based on a visual saliency map, achieves efficient and ef-
fective attention retargeting compared with the conventional
orientation-based method, which uses spatial filtering without
depending on a visual saliency map.

Although state-of-the-art color-based methods achieve attention retargeting without degradation in visibility, it is not necessarily effective for all images, and vice versa. The color-based method and the proposed orientation-based method modify different image features on the basis of different visual saliency maps. Therefore, more efficient and effective attention retargeting can be realized by applying both methods to an image.

7. Conclusion

We proposed an orientation-based attention retargeting method for a given region in an image in order obtain the image in which the region shows the highest saliency. A novel visual saliency map for bottom-up attention, which matches well with the distribution of actual human attention, was proposed on the basis of the WT. The proposed image modification method iteratively adjusts wavelet coefficients such that the saliency inside the ROI increases while that outside the ROI decreases. We performed gaze measurements to evaluate the effectiveness of the proposed method. The results of our experiments clearly showed that our method can successfully guide human gaze to an ROI.

The main contribution of this work is an image modification method using wavelet transform and saliency map based on frequency components. It should be noted that the objective of our proposed method is not natural attention retargeting such as color-based retargeting [9] or other orientation-based method [12]. It may be difficult to understand what the object in the region outside an ROI is because the region outside the ROI is strongly blurred. Therefore, it may cause a decline in the quality of viewers’ activity. However, this issue is resolved by estimating a viewer’s gaze. We would suggest that the quality of viewers’ activity is preserved by changing region outside the ROI to original image after the viewers’ gaze is shifted to ROI.

Gaze movement is affected by both bottom-up and top-down factors. Several computational models based on top-down approaches consider the intentions, cognitive states, and prior knowledge of humans [6]. When a viewer looks at a specific scene continuously, efficient attention retargeting can be realized by considering the influence of both bottom-up and top-down attention. Therefore, an attention retargeting method based on top-down attention should be considered in the future.

Acknowledgments

This research was partially supported by a Grant-in-Aid for Scientific Research (C) from the Japan Society for the Promotion of Science (grant no. 15K00282).

References

[1] C. Koch and S. Ullman, “Shifts in selective visual attention: Towards the underlying neural circuitry,” Human Neurobiolgy, vol.4, no.4, pp.219–227, 1985.
[2] A.M. Treisman and G. Gelade, “A feature-integration theory of attention,” Cognitive Psychology, vol.12, no.1, pp.97–136, 1980.
[3] M.I. Posner, “Orienting of Attention,” Quarterly Journal of Experimental Psychology, vol.32, no.1, pp.3–25, 1980.
[4] J. Shen, E.M. Reingold, and M. Pomplun, “Distractor ratio influences patterns of eye movements during visual search,” Perception, vol.29, no.2, pp.241–250, 2000.
[5] V.A. Mateescu and I.V. Bajić, “Visual Attention Retargeting,” IEEE Multimedia Mag., vol.23, no.1, pp.82–91, 2016.
[6] A. Kimura, R. Yonetani, and T. Hirayama, “Computational models of human visual attention and their implementations: A survey,” IEICE Trans. on Inf. and Syst., vol.96, no.3, pp.562–578, 2013.
[7] A. Hagiwara, A. Sugimoto, and K. Kawamoto, “Saliency-based image editing for guiding visual attention,” in Proc. of 1st Int. Workshop on Pervasive Eye Tracking and Mobile Eye-Based Interaction, pp.43–48, 2011.
[8] T. Kokui, H. Takimoto, Y. Mitsuoka, M. Kishihara, and K. Okubo, “Color Image Modification Based on Visual Saliency for Guiding Visual Attention,” in Proc. of 22th IEEE Int. Symp. on Robot and Human Interactive Comm., pp.467–472, 2013.
[9] H. Takimoto, T. Kokui, H. Yamauchi, M. Kishihara, and K. Okubo, “Image Modification Based on a Visual Saliency Map for Guiding Visual Attention,” IEICE Trans. on Inf. and Syst., vol.E98-D, no.11, pp.1967–1975, 2015.
[10] Y. Kim and A. Varshney, “Persuading Visual Attention through Geometry,” IEEE Trans. Visualization and Comp. Graphics, vol.14, no.4, pp.772–782, 2008.
[11] V.A. Mateescu and I.V. Bajić, “Guiding visual attention by manipulating orientation in images,” in Proc. of IEEE Int. Conf. on Multimedia and Expo, pp.1–6, 2013.
[12] H. Hata, H. Koike, and S. Yoichi, “Visual Attention Guidance Using Image Resolution Control,” Inf. Process., vol.56, no.4, pp.1152–1161, 2015. (in Japanese)
[13] R. Bailey, A. McNamara, N. Sudarsanam, and C. Grimm, “Subtle gaze direction,” ACM Trans. on Graphics, vol.28, no.4, pp.1–14, 2009.
[14] A. Borji and L. Itti, “State-of-the-Art in Visual Attention Modeling,” IEEE Trans. on Pattern Anal. and Mach. Intell., vol.35, no.1, pp.185–207, 2013.
[15] A. Borji, D.N. Sihite, and L. Itti, “Quantitative Analysis of Human-Model Agreement in Visual Saliency Modeling: A Comparative Study,” IEEE Trans. on Image Proc., vol.22, no.1, pp.55–69, 2013.
[16] E.E. Veas, E. Mendez, S. Feiner, and D. Schmalstieg, “Directing attention and influencing memory with visual saliency modulation,” in Proc. of the SIGCHI Conf. on Human Factors in Computing Syst., pp.1471–1480, 2011.
[17] E. Mendez, S. Feiner, and D. Schmalstieg, “Focus and Context in Mixed Reality by Modulating First Order Salient Features,” in Proc. of Int. Sympo. on Smart Graphics, pp.232–243, 2010.
[18] L. Itti, C. Koch, and E. Niebur, “A model of saliency based visual attention for rapid scene analysis,” IEEE Trans. on Pattern Anal. and Mach. Intell., vol.20, no.11, pp.1254–1259, 1998.
[19] T.V. Nguyen, B. Ni, H. Liu, W. Xia, J. Luo, M. Kankanhalli, and S. Yan, “Image Re-Attentionizing,” IEEE Trans. Multimedia, vol.15, no.8, pp.1910–1919, 2013.
[20] V.A. Mateescu and I.V. Bajić, “Attention Retargeting by Color Manipulation in Images,” in Proc. of ACM Multimedia PIVP, pp.15–20, 2014.
[21] Z. Su and S. Takahashi, “Real-time Enhancement of Image and Video Saliency using Semantic Depth of Field,” in Proc. of Int. Conf. on Comp. Vision Theory and App., pp.370–375, 2010.
[22] N.R. Kadaba, X.-D. Yang, and P.P. Irani, “Facilitating Multiple Target Tracking using Semantic Depth of Field,” in Proc. of 27th Int. Conf. Extended Abstracts on Human Factors in Computing Syst.,
[23] N. İmamoğlu, W. Lin, and Y. Fang, “A Saliency Detection Model Using Low-Level Features Based on Wavelet Transform,” IEEE Trans. Multimedia, vol.15, no.1, pp.96–105, 2013.

[24] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, “Frequency tuned salient region detection,” in Proc. of IEEE Conf. on Comp. Vision and Pattern Recog., pp.1597–1604, 2009.

[25] C. Guo, Q. Ma, and L. Zhang, “Spatio-temporal Saliency detection using phase spectrum of quaternion fourier transform,” in Proc. of IEEE Conf. on Comp. Vision and Pattern Recog., pp.1–8, 2008.

[26] X. Hou and L. Zhang, “Saliency Detection: A Spectral Residual Approach,” in Proc. of IEEE Conf. on Comp. Vision and Pattern Recog., pp.1–8, 2007.

[27] R.J.E. Merry, Wavelet Theory and Applications: A Literature Study, DCT 2005.53, Eindhoven, The Netherlands: Eindhoven Univ. Technol., 2005.

[28] A. Borji and L. Itti, “CAT2000: A Large Scale Fixation Dataset for Boosting Saliency Research,” arXiv:1505.03581, pp.1–4, May 2015.

[29] D. Pang, A. Kimura, T. Takeuchi, J. Yamato, and K. Kashino, “A stochastic model of selective visual attention with a dynamic Bayesian network,” in Proc. of IEEE Int. Conf. of Multimedia and Expo, pp.1073–1076, 2008.

[30] R.J. Peters and L. Itti, “Applying computational tools to predict gaze direction in interactive visual environments,” ACM Trans. on Applied Perception, vol.5, no.2, pp.1–19, 2008.

[31] N. Riche, M. Duvinage, M. Mancas, B. Gosselin, and T. Dutoit, “Saliency and Human Fixations: State-of-the-art and Study of Comparison Metrics,” in Proc. of IEEE Int. Conf. on Computer Vision, pp.1151–1160, 2013.

[32] E. Vig, M. Dorr, and D. Cox, “Large-Scale Optimization of Hierarchical Features for Saliency Prediction in Natural Images,” in Proc. of Computer Vision and Pattern Recognition, pp.2798–2805, 2014.

[33] J. Zhang and S. Sclaroff, “Saliency Detection: A Boolean Map Approach,” in Proc. of IEEE Internatinoal Conference on Computer Vision, pp.153–160, 2013.

[34] M.-M. Cheng, N.J. Mitra, X. Huang, P.H.S. Torr, and S.-M. Hu, “Global contrast based salient region detection,” in IEEE Trans. Pattern Anal. Mach. Intell., vol.37, no.3, pp.569–582, 2015.

[35] http://www.sozaijiten.com/

Syuhei Hitomi received his B.E. and M.E. degrees from Okayama Prefectural University, Japan, in 2014 and 2016, respectively. His research interests include image processing and perceptual information processing.

Hitoshi Yamauchi received his B.E., M.E. and Ph.D. degrees from University of Osaka Prefecture in 1993, 1995 and 2008, respectively. In 1995, he joined the Faculty of Computer Science and Systems Engineering, Okayama Prefectural University. In 2007, he became an Assistant Professor. He is currently an Associate Professor since 2012. His research interests include image processing. He is a member of IEICE, IPSJ, ISCIE, SICE and IEEE.

Mitsuyoshi Kishihara received his B.S. degree in Electrical Engineering from Setsunan University, Osaka, Japan, in 1996, and M.S. and Ph.D. degrees in Electronics Engineering from Himeji Institute of Technology, Hyogo, Japan, in 1998 and 2002, respectively. In October 2000, he joined Okayama Prefectural University, as a Research Associate. He is currently an Associate Professor. He is a member of IEICE and IEEE.

Kensuke Okubo received his B.E., M.E. and D.Eng. degrees in electrical engineering from the Kyoto Institute of Technology, Kyoto, Japan, in 1987, 1989, and 1992, respectively. In April 1993, he joined the Faculty of Computer Science and System Engineering, Okayama Prefectural University, Okayama, Japan, where he is now a Professor. He is a member of IEICE and IEEE.