PAPER Image Modification Based on a Visual Saliency Map for Guiding Visual Attention

Hironori TAKIMOTO^{†a)}, *Member*, Tatsuhiko KOKUI[†], *Nonmember*, Hitoshi YAMAUCHI[†], Mitsuyoshi KISHIHARA[†], and Kensuke OKUBO[†], *Members*

SUMMARY It is commonly believed that improved interaction between humans and electronic device, it is effective to draw the viewer's attention to a particular object. Augmented reality (AR) applications can call attention to real objects by overlaying highlight effects or visual stimuli (such as arrows) on a physical scene. Sometimes, more subtle effects would be desirable, in which case it would be necessary to smoothly and naturally guide the user's gaze without external stimuli. Here, a novel image modification method is proposed for directing a viewer's gaze to specific regions of interest. The proposed method uses saliency analysis and color modulation to create modified images in which the region of interest is the most salient region in the entire image. The proposed saliency map model that is used during saliency analysis reduces computational costs and improves the naturalness of the image using the LAB color space and simplified normalization. During color modulation, the modulation value of each LAB component is determined in order to consider the relationship between the LAB components and the saliency value. With the image obtained in this manner, the viewer's attention is smoothly attracted to a specific region very naturally. Gaze measurements as well as a subjective experiments were conducted to prove the effectiveness of the proposed method. These results show that a viewer's visual attention is indeed attracted toward the specified region without any sense of discomfort or disruption when the proposed method is used.

key words: visual attention, saliency map, image processing, guiding attention

1. Introduction

Estimates show that 80% of the information entering a human brain is visual. Furthermore, humans tend to use information obtained from their peripheral vision when deciding where to direct their attention [1], [2]. Recently, many studies have reported attentive user interfaces (AUIs) that use optical head-mounted displays such as Google Glass. Effective AUIs naturally attract a viewer's visual attention to the region specified by the gaze-based interface. Therefore, it is commonly believed that by drawing a viewer's attention to particular objects, many types of human activities (for example driving, manufacturing, and studying) and human-machine interaction can be effectively facilitated and directed.

Humans choose important information from an enormous volume of visual information; this is called "visual attention". Visual saliency may be defined as an estimation of

Manuscript received March 11, 2015.

Manuscript revised July 12, 2015.

Manuscript publicized August 13, 2015.

[†]The authors are with the Faculty of Computer Science and System Engineering, Okayama Prefectural University, Soja-shi, 719–1197 Japan.

a) E-mail: takimoto@c.oka-pu.ac.jp

DOI: 10.1587/transinf.2015EDP7087

how likely a given region is to attract human visual attention. Several image modification methods that use visual saliency maps to effectively and naturally guide a viewer's attention to a specified region have recently been proposed. These can be divided into orientation-based and color-based methods. Typical orientation-based methods include those proposed by Hata et al. [3] and Hitomi et al. [4], which are based on spatial filters. The basic concept of such orientation-based methods is to modify the visual saliency of an image by blurring the color or intensity outside the region of interest such that the viewer's attention is guided to the nonblurred region. Mateescu et al. analyzed how manipulating the orientation of a particular region of an image affects human visual attention [5]. In that method, the region selected by a user is rotated so that the selected region's saliency is maximized. However, it is difficult to apply these orientationbased methods to specified regions with low texture.

On the other hand, color-based methods modify lightness and color components so that the visual saliency inside a specified region increases whereas that outside the region decreases. Color-based methods can generally guide a viewer's attention to the target region while keeping the rest of the image at high resolution. Hagiwara et al. [6] proposed such a modification method that iteratively modulates lightness and color components. However, despite this method being highly effective, unnatural colors were observed in the modified image because very little attention was given to the naturalness of the modified image. In addition, iterative saliency map calculations and color modulations require large computational time. Therefore, a high-speed, color-based image modification technique for guiding visual attention without causing discomfort to the viewer is still required for practical AUIs.

In this paper, we propose a high-speed, color-based modification method to obtain an image in which the specified region achieves its highest saliency. The proposed method consists of two phases – visual saliency analysis and color modulation. First, a novel visual saliency map based on the CIE LAB color space (which is a good match with the distribution of actual human attention) is proposed; by omitting some normalization processing, the saliency map's calculation costs are reduced. Next, the proposed method iteratively modulates lightness and color components in the LAB color space so that the visual saliency inside the specified region increases whereas that outside the region decreases. By optimizing the step size of the iteration, highspeed image modification is then achieved without decreasing the image quality. When this modification method is applied to images, the human gaze is guided to the specified region because the saliency inside that region increases while that outside the region decreases.

The aim of this paper is to modify the original image while keeping it looking "natural" (i.e. unmodified) to reduce discomfort for the viewer. In this paper, the naturalness of modified image is defined as the preservation of color tone in a scene. The proposed method improves the tradeoff between effectiveness of the attention guiding and the naturalness of the modified image. From gaze measurement results with test subjects, we demonstrate that the proposed saliency map matches well with the distribution of actual human attention. Moreover, we confirm that the viewer's visual attention is indeed attracted toward the specified regions without discomfort based on our gaze measurement results and their answers to a questionnaire.

2. Related Works

Here, we discuss other studies on attention-guiding methods. One conventional approach for guiding the human gaze is to present visual stimuli such as arrows [1] or lightemitting diodes (LEDs) [7] in the peripheral visual field. Bailey et al. [8] proposed a method, called "subtle gaze direction", which terminates modulation before the user has an opportunity to scrutinize the image by monitoring saccadic velocity. However, from the viewer's standpoint, this approach is more coercive than persuasive. A better approach would be to smoothly guide the viewer's attention toward a target without disturbing their current visual attention.

Here, we are more interested in attention-guiding methods based on visual saliency. The orientation-based image modification method proposed by Hata et al. [3] creates a modified image by using a Gaussian filter to blur areas outside the specified region. In this method, the high frequency component is suppressed by applying spatial filtering for controlling the resolution. Therefore, this method is classified as an orientation-based method. However, it is difficult to directly guide a viewer's gaze toward the specified region after the first saccade, because this method uniformly applies a Gaussian filter to various frequency components without using a visual saliency map. To address this issue, Hitomi et al. [4] proposed a saliency map based on the wavelet transform and an image modification method to direct a viewer's gaze to a given region in an image. This method modified the frequency components based on the obtained saliency map to suppress visual saliency outside the specified region. Mateescu et al. [5] rotated the selected region by predicting the angle of rotation at which the region becomes the most salient of the entire image. The method proposed by Su et al. [9], which enhances the saliency of image regions that are different from their surroundings in intensity or color, makes full use of the semantic depth of a field technique. However, all of these methods are difficult to apply to the specified regions with low texture.

On the other hand, color-based methods modify the lightness and color components so that the visual saliency inside a specified region increases whereas that outside the region decreases. Color-based methods can generally guide a viewer attention to the target region while non-target region maintains high resolution. Therefore, it is expected that the color-based method is applicable to a wide range of applications. Veas et al. [10] proposed a saliency modulation technique, which prompts attention shifts and influences the recall of selected regions without perceptible change to visual input. Mendez et al. [11] proposed a method for dynamically directing a viewer's gaze by analyzing and modulating the bottom-up salient features. In this method, the specified region is adaptively darkened, lightened, and manipulated in hue according to local contrast information rather than according to global parameters. Although these methods show good results compared with other approaches, a threshold map for each image needs to be manually preset. Recently, Shi et al. [12] proposed a video saliency modulation method based on the HSI color space. Hagiwara et al. [6] proposed a method for editing an image in order to obtain the image in which the given region has its highest saliency. Although this method can direct the gaze to specified regions, false border lines, which cause discomfort to users, are generated in the specified region. Kokui et al. [13] proposed a modification method of RGB color components based on multi-resolution using the conventional saliency map proposed by Itti et al. [14]. Nguyen et al. [15] proposed a new computational framework which actively recolors only the target region to make it stand out, in both a local and global sense. This method utilizes a salient patch dataset that includes the recorded fixation data. Then, the color transfer from the salient patches onto the target region that maximizes the region's saliency is performed using graph-based optimization. A disadvantage of this method is that is computationally expensive and requires access to a sufficiently comprehensive database of gaze data. None of these conventional color-based methods gave much attention to computational costs or the naturalness of the modified image.

Mateescu et al. [16] proposed a saliency manipulation method that modifies hue while keeping intensity and chromaticity constant. However, the original hue of the target object is not considered, although this method modifies only the hue. In addition, when other objects with high saliency exists in the original image, it is difficult to effectively guide a gaze to the modified target object. Figure 1 shows examples of attention re-targeting to the bread (right hand side) by conventional and proposed methods. Viewers may feel unnaturalness when Mateescu's methods are applied to a particular object where an original hue has the special meaning such as its category color. On the other hand, our method and Hagiwara's method modify a color by considering the original color of target region to improve the naturalness.



Fig.1 Examples of attention re-targeting to the bread (right hand side): (a) original image, (b) modified image (our method), (c) modified image (Hagiwara's method [6]), (d) modified image (Mateescu's method [16])



Fig. 2 Flowchart of the proposed method

3. Image Modification for Guiding Visual Attention

In this section we describe our method for indirectly adjusting the saliency of an entire image by changing each component in the CIE LAB color space in order to subtly guide a viewer's visual attention. Our proposed method has two phases, firstly we create a visual saliency map from the original image, and secondly we modulate the color components using the obtained saliency map. A flowchart of the proposed method is shown in Fig. 2; the processing details are explained in the following subsections.

3.1 Creating the Visual Saliency Map

Itti et al. [14] proposed a computational model of visual saliency on the basis of Koch and Ullman's early vision model [17] and demonstrated that a saliency map matches well with the distribution of actual human attention based on their human gaze measurements. This means that by adjusting the features of the whole image based on a saliency map, a specified region can attract the attention of a user. However, our modification method requires iterative cal-



Fig. 3 Saliency map model based on the CIE LAB color space

culations of the saliency map, which increases processing costs. The computational costs associated with the conventional saliency map [14] cause serious problems for real-time processing.

The model we propose here requires less computational time, it can compute visual saliency in a short time using the CIE LAB color space (which is designed to be perceptually uniform) and omitting normalization.

A flowchart of the novel saliency map is shown in Fig. 3. Bottom-up visual attention is related to the low-level features of the scene such as intensity, color, and orientation. Most conventional saliency maps based on visual attention consider these features where orientation is influenced by the shape and texture of an object. Modulation of the orientation means that the object in the image is moved, leading to a significant change in the image. We prefer not to implement this technique as the objective of this study is to naturally guide a user's gaze without causing discomfort.

First, we convert the image *I* from the RGB to the CIE LAB color space. This LAB frame is split into three channels – L*, a*, and b* – where a* and b* are normalized by L* and their small values are set to zero. Gaussian pyramids of each channel are created by smoothing and down-sampling the input image: $L*(\sigma)$, $a*(\sigma)$, $b*(\sigma)$, where $\sigma \in [0, 8]$ represents a Gaussian pyramid with nine scales. These pyramids are used for the center-surround difference operation to create feature maps.

Center-surround difference is implemented by the pixel difference between finer scale *c* and coarser scale *s*. We set c = 2, 3, 4 and s = c + d, with d = 3, 4. Each feature map FM_k ($k = L^*, a^*, b^*$) is defined as follows:

$$FM_k(c,s) = |k(c) \ominus k(s)| \quad (k = L^*, a^*, b^*)$$
(1)

where \ominus indicates the corresponding pixel-wise subtraction between the two scale images. Next, the feature maps are normalized and combined into three conspicuity maps CM_k at a scale of 4.

$$CM_k = \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} FM_k(c,s)$$
(2)

where \bigoplus indicates the corresponding pixel-wise summation between the two feature maps. Finally, the three conspicuity maps are normalized and combined into a master saliency map SM.

$$SM = \frac{1}{3}(\mathcal{N}(CM_{L^*}) + \mathcal{N}(CM_{a^*}) + \mathcal{N}(CM_{b^*}))$$
(3)

where N is a normalization operator. The simple max-local normalization [14] is applied to each conspicuity map to highlight the most discriminative feature within each map.

Unlike the model given by Itti et al. [14], in Eq. (2), the normalization for combining each feature map is removed for our method because of the high calculation costs of normalization. On the other hand, in [14], four Gaussian pyramids for each color channel (red, green, blue, and yellow) are created to represent a color double-opponent system. In the center of each receptive field of the eye, neurons are excited by one color and inhibited by another, while the converse is true in the surrounding region. Such spatial and chromatic opponency exists for the red/green, green/red, blue/yellow, and yellow/blue color pairs in the human primary visual cortex [18]. In our implementation, channels a* and b* for the color-opponent dimensions represent a color double-opponent system. Therefore, the calculation costs associated with creating Gaussian pyramids and feature maps are reduced.

Achanta et al. [19], [20] and Chuang et al. [21] have also proposed saliency map models using the LAB color space. Achanta's novelty-based models have been developed to try to find spatial irregularities and temporal nonstationarity as saliency in an image. Our saliency map model is a psychophysics-based model based on the feature integration theory [22]. We believe that a psychophysics-based model is efficient for applying a saliency map to image modification for guiding visual attention. Chuang's model also includes psychophysics-based saliency, however, here we improve on this by omitting several normalization steps to achieve real time processing. In addition, our saliency map does not include an orientation component since the image modification method modifies each pixel based on the color components. Therefore, the saliency map presented here is specialized for color modification (discussed in more detail later). In Sect. 4, the effectiveness of this saliency model is evaluated experimentally.

3.2 Color Modulation Based on the Visual Saliency Map

The basic concept of our color modulation is that, by iteratively modulating intensity and color components, saliency inside the region of interest increases whereas that outside this region decreases. The procedures of this method are as follows:

- 1. Selection of a region D.
- 2. Calculation of saliency map $S M^t$ of input image I^t .
- 3. Calculation of intensity coefficient *P^t* and modification value *Q^t*.
- 4. Color modulation using P^t and Q^t .

First, a user selects a region *D* to which he wants to guide a viewer's attention. If $k^t(i, j)$ is the intensity or color $k \in \{L^*, a^*, b^*\}$ of the input image I^t at pixel (i, j), then $k^{t+1}(i, j)$ is the corresponding value in the modified image I^{t+1} . Here, t = 0 is the original image. Each channel is then modified by the following:

$$k^{t+1}(i,j) = k^{t}(i,j) + W \cdot P^{t}(i,j) \cdot Q_{k}^{t}(i,j)$$
(4)

where W is the weight coefficient for color modulation. Although the computational time decreases as parameter W increases, the image quality decreases. Therefore, we use a subjective experiment to optimize W. The intensity coefficient $P^t(i, j)$ in Eq. (4), which is the weight for the modulation value of each pixel, is defined by:

$$P^{t}(i,j) = \begin{cases} S M_{ave}^{t} & (i,j) \in D \\ -S M^{t}(i,j) & otherwise \end{cases}$$
(5)

$$SM_{ave}^{t} = \frac{1}{n} \sum_{(i,j)\in D} SM^{t}(i,j)$$
(6)

where *n* is the number of pixels in the given region *D*. $P^{t}(i, j)$ determines whether the intensity or color $k^{t}(i, j)$ is emphasized or suppressed. Moreover, to prevent the generation of a false border, which would give the viewer a sense of incongruity, we set the intensity coefficient average inside *D*.

The modification value $Q_k^t(i, j)$ is defined by

$$Q_k^t(i,j) = \operatorname{Sgn}_k^t(i,j)\varphi_k^t \tag{7}$$

 $Q_k^t(i, j)$ reflects the amount of influence a given feature has

over saliency and allows us to obtain (by back calculation) the saliency map. Here, $Sgn_k^t(i, j)$ is defined as follows:

$$\operatorname{Sgn}_{k}^{t}(i,j) = \begin{cases} 1 & k^{t}(i,j) > k_{ave}^{t}(i,j) \\ -1 & otherwise \end{cases}$$
(8)

where $k_{ave}^t(i, j)$ is the average of $k^t(i, j)$ in a region of 256 × 256 pixels around a pixel (i, j). This area for averaging is defined by the minimum scale $\sigma = 8$ of the Gaussian pyramid. Next, the influence degree φ_k^t (meaning the rate of influence that each intensity or color has to saliency) is defined by

$$\varphi_{k}^{t} = \frac{\mathcal{N}(CM_{k}^{t})}{\mathcal{N}(CM_{L^{*}}^{t}) + \mathcal{N}(CM_{a^{*}}^{t}) + \mathcal{N}(CM_{b^{*}}^{t})}$$
(9)

After image modification of an input image I^{t} , the saliency map SM^{t+1} is calculated. If saliency SM_D^{t+1} inside D is the highest saliency in the whole modified image I^{t+1} , then the image modification is completed. Otherwise, each feature k^{t+1} is iteratively modulated on the basis of SM^{t+1} .

4. Experiments

4.1 Saliency Map Evaluation

The proposed saliency map matches well with the distribution of actual human attention, as demonstrated by our human gaze measurements.

4.1.1 Experimental Setup

We used a QG-PLUS tracking device (Ditect Inc.) to track the eye movements of our subjects. With this technology, each subject's head is fixed at a distance of 65 cm from the front of the display. We define a fixation as a set of consecutive data points that are within certain proximity of the visual angle for a minimum of 100 ms. The radius of the visual angle for fixation was 1.3 degrees.

Nineteen subjects (11 male, 8 female, 18–24 years old, $\bar{x} = 21.4$), each with normal color vision, participated in the experiment that included 30 images that were 640×480 pixels in size. The resolution of the 19 inches display was chosen so that each image could be viewed at its native resolution. Each image was displayed for 3 s in random order with a 1 s pause in between. During the pause, a small cross-hair was shown in the center of the display and the participants were asked to fixate on it.

4.1.2 Experimental Results and Discussion

For comparison, a typical location-based, bottom-up saliency map (based on Itti's saliency map [14]) used in a conventional method [6] was employed along with our proposed method. We chose this saliency map for comparison because, like ours, it is calculated only from intensity and color features.

Evaluational results of normalized scanpath saliency

Table 1NSS and Percentail

Ave. (SD)	NSS	Percentail
Proposed saliency map without norm.	0.472 (0.219)	0.649 (0.062)
Proposed saliency map with norm.	0.415 (0.250)	0.650 (0.060)
Conventional saliency map [6]	0.431 (0.238)	0.625 (0.075)

Table 2	Computational times of saliency map	calculatio	ns [ms]
-	Proposed saliency map without norm.	9.5	

rioposed surferey map without norm.	1.5
Proposed saliency map with norm.	11.4
Conventional saliency map [6]	24.6

(NSS) and percentail are shown in Table 1. The averages and standard deviations of NSS and percentail for each test image are included in that table. NSS measures the performance of saliency models using fixations. First, the saliency scores of all regions in an image are normalized to have a zero mean and unit standard deviation [23]; then, the saliency scores at fixed locations are used to measure the model performance. On the other hand, percentail (first proposed in [24]) measures the percentage of fixations whose predicted saliency values fall below the value of a fixed location. "Proposed saliency map without norm." means the saliency map we propose in Sect. 3.1. "Proposed saliency map with norm." means normalization is performed when each feature map FM is combined into one conspicuity map CM, which is defined by:

$$CM_k = \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} \mathcal{N}(FM_k(c,s))$$
(10)

It can be seen that all criteria of the proposed saliency map are equal to or greater than those of the other saliency maps within experimental uncertainty. In other words, the proposed saliency map matches well with the distribution of human attention. In addition, the computational time of each saliency map is shown in Table 2, where it can be seen that our proposed method is significantly faster than the other techniques and is hence suitable for color modulation.

4.2 Evaluation of Color Modulation

Using a gaze-measurement system, we confirm that the modified image obtained by the proposed method guides viewers' attention. We also evaluate the naturalness of the modified image using a subjective experiment and confirm the processing costs associated with our method.

4.2.1 Experimental Setup

First, we confirmed that the modified image successfully guides the viewer's attention compared with conventional methods. Twenty subjects (12 male, 8 female, 18–24 years old, $\bar{x} = 21.6$), each with normal color vision, participated in the experiment where the number and size of the images was the same as for the saliency map experiments. Our method was compared to the comparison method for all original images. We used the color-based modification method proposed by Hagiwara et al. [6], to evaluate how well each



Fig.4 Example of original and modified images: (a) original images, (b) saliency maps of (a), (c) images modified by the proposed method, (d) saliency maps of (c), (e) images modified by [6], and (f) saliency maps of (e)

method guides visual attention. We minimized the influence of participant observing an image of the same scene in the past. We created 90 test images from 30 original images (2 modified images per original image). These 90 images were divided into 3 groups, where every group had 30 different scenes. In each group, the images were shown in random order. In addition, we controlled a recess between each group to reduce the influence of the past experiment. For target regions, we selected the objects with relatively low saliency that were not located at the center of the image. The weight coefficient for color modulation was set to a value of 10 for these experiments.

We also performed a subjective experiment to evaluate whether the proposed method modified images without causing discomfort to the viewer. In this paper, the "naturalness" of modified image is defined with respect to the color tone in a scene; a deterioration of color tone is perceived by the viewer as a less natural image. We explained this definition of naturalness for a modified image to the test subjects before the experiments began. In addition, all subjects performed training in the evaluation technique using different datasets. The subjective experiment was performed on the basis of the double-stimulus impairment scale (DSIS) and pair comparison (PC) proposed in ITU-R [25]. With DSIS, an original image and its corresponding modified image were simultaneously shown to subjects. Subjects were asked "how natural is the modified image compared with the original image?". Subjects evaluated the naturalness of the modified image using the five-grade impairment scale (where 5 is excellent and 1 is bad). High DSIS values mean high-quality images. With PC, modified images created by the proposed and comparison methods were simultaneously shown to subjects. Subjects were asked "which of the two images looks more natural?". Subjects evaluated the images on a quality assessment scale based on five categories (where +2 means that the image created by proposed method is better and -2 means that the image created by the comparison method is better).

The hardware platform for the experiment was a personal computer equipped with an Intel Core i7-3770 3.4GHz



(a) original image



(d) original image



(b) modified image of (a) (coventional method[6])



(e) modified image (d) (coventional method[6])





(c) modified image of (a) (proposed method)



(f) modified image of (d) (proposed method)

CPU with 8GB RAM. To evaluate performance in terms of execution time, we implemented the saliency calculations and color modulation schemes using the Intel OpenCV library using a PC running Microsoft Windows 7.

4.2.2 Experimental Results and Discussion

Examples of original images, modified images obtained by the proposed method and the comparison method [6] and their respective saliency maps are shown in Fig. 4. Examples of scan-paths obtained from the gaze-tracking system are shown in Fig. 5, where the numbers in circles show the order of fixation movement, and the size of the circles shows the time of fixation. In the images in the top row, the watermelon at the top right of the image was selected as the target region, and the number of fixation movements that occur prior to reaching the specified region was 3, 2, and 2, respectively. In the original image (Fig. 5 (d)), the specified region was not observed by the subject. In Fig. 5 (d), the specified region was not observed by the subject. The number of fixation movements that occur prior to reaching the specified region in Figs. 5 (e) and (f) were 6 and 1, respectively.

The rates of fixation movements occurring prior to reaching the specified regions are shown in Table 3. The proposed method more effectively guides gazes to the specified regions than the comparison method. The reason our method is more effective than the comparison method is the deterioration of the tone of a color by color modulation. Furthermore, the comparison method fails to sufficiently suppress saliency outside the specified region. Using results from these experiments, we performed a pairwise t-test (two-tailed) to compare the methods statistically, as shown in Table 4, which shows that there are indeed statistically significant differences between the methods. However, despite these good results, our method was unable

 Table 3
 Fixation movement rates

%	Original image	Proposed method	Conventional method [6]
1st	6.0	41.0	25.8
2nd	9.0	21.5	29.5
3rd	10.7	9.5	10.8
4th	5.8	5.0	5.0
5th	2.5	1.8	3.8
other	66.0	21.2	25.0

 Table 4
 The results of pairwise t-test (two-tailed)

Condition	t-Value	p-Value
Original vs. Proposed method	23.63	< 0.001
Original vs. Conventional method	17.94	< 0.001
Proposed method vs. Conventional method	-4.11	< 0.001

to guide gazes in some modified images where the specified region was small; the ratio of the specified region to the whole image ranged in our experiments from 0.16% to 19.0% ($\bar{x} = 3.19\%$).

After determining the superiority of our method in directing gazes to the specified regions, we analyzed the subjective naturalness of each modified image. The average DSIS result for our method was 3.31 (SD = 0.93) and that for the comparison method was 2.61 (SD = 1.27). The results of a pairwise t-test (two-tailed) on these results were t(599) = 10.36, p < 0.001. Thus, there are statistically significant differences between our method and the comparison method. On the other hand, the average result of pair comparison was 0.48. Hence, our proposed method improved the image quality without lowering the effect of gaze direction.

Finally, we evaluated the computational costs of the proposed method and the conventional method. The number of iterations and computational times for color modulation are shown in Table 5. It can be seen that our method

 Table 5
 Comparison of the number of iterations and computational time for color modulation

	# of iterations	Computational time [ms]
Proposed method	6.6	277.2
Conventional method [6]	153.7	9280.3

achieves significantly fewer iterations with a computational time more than thirty times shorter than the conventional method, without lowering image quality and while maintaining influence over a viewer's gaze.

Gaze movement is affected by both bottom-up and topdown factors. There are several computational models based on top-down approaches that consider the prior knowledge, intentions and cognitive states of humans [26]. When a viewer looks at the specific scene continuously, efficient guiding of visual attention is realized by considering an influence of top-down attention. Therefore, a guiding method based on top-down attention is considered as future work.

5. Conclusions

We proposed a method for modifying the color map of an image to highlight a specific region of interest. This method uses saliency analysis and color modulation to iteratively adjust the intensity and color so that the saliency inside a user-specified region increases while that outside the region decreases. We performed gaze measurements and a subjective analysis of image naturalness to evaluate the effectiveness of the proposed method. This method showed significantly lower computational times than other methods and produced images that were considered much more natural. It was shown that this new method can be used successfully to subtly and naturally guide the human gaze to the target region.

Acknowledgments

This research was partially supported by the Wesco Scientific Promotion Foundation and a Grant-in-Aid for Scientific Research (C) from the Japan Society for the Promotion of Science (grant no. 24500653).

References

- M.I. Posner, "Orienting of Attention, Quarterly Journal of Experimental Psychology," vol.32, no.1, pp.3–25, 1980.
- [2] 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.
- [3] H. Hata, H. Koike, and S. Yoichi, "Visual attention guidance using image resolution control," Proc. of IPSJ Interaction 2014, pp.57–64, 2014 (in Japanese).
- [4] S. Hitomi, T. Kokui, H. Takimoto, H. Yamauchi, M. Kishihara, and K. Okubo, "Guiding Visual Attention Using Saliency Map Based on Wavelet Transform," Proc. of RISP International Workshop on NCSP 2015, pp.90–93, 2015.
- [5] V.A. Mateescu and I.V. Bajić, "Guiding visual attention by manipulating orientation in images," Proc. of 2013 IEEE International Conference on Multimedia and Expo, pp.1–6, 2013.

- [6] A. Hagiwara, A. Sugimoto, and K. Kawamoto, "Saliency-based image editing for guiding visual attention," Proc. of PETMEI 2011, pp.43–48, 2011.
- [7] H. Inome, E. Aiba, T. Shimotomai, T. Fujisawa, and N. Nagata, "Visibility Improvement in LED Warning Lights: An Evaluation Method of the Flash Patterns, Journal of the Institute of Image Information and Television Engineers," vol.34, no.10 pp.89–92, 2010 (in Japanese).
- [8] R. Bailey, A. McNamara, N. Sudarsanam, and C. Grimm, "Subtle gaze direction," ACM TOG, vol.28, no.4, pp.1–14, 2009.
- [9] Z. Su and S. Takahashi, "Real-time Enhancement of Image and Video Saliency using Semantic Depth of Field," VISAPP 2, pp.370–375, INSTICC Press, 2010.
- [10] E.E. Veas, E. Mendez, S.K. Feiner, and D. Schmalstieg, "Directing attention and influencing memory with visual saliency modulation," Proc. of the SIGCHI Conference on Human Factors in Computing Systems, pp.1471–1480, 2011.
- [11] E. Mendez, S. Feiner, and D. Schmalstieg, "Focus and Context in Mixed Reality by Modulating First Order Salient Features," Proc. of 10th International Symposium on Smart Graphics, vol.6133, pp.232–243, 2010.
- [12] T. Shi and A. Sugimoto, "Video Saliency Modulation in the HSI Color Space for Drawing Gaze," Proc. of Pacific-Rim Symposium on Image and Video Technology, vol.8333, pp.206–219, 2013.
- [13] T. Kokui, H. Takimoto, Y. Mitsukura, M. Kishihara, and K. Okubo, "Color Image Modification Based on Visual Saliency for Guiding Visual Attention," Proc. of 22th IEEE International Symposium in RO-MAN2013, pp.467–472, 2013.
- [14] L. Itti, C. Koch, and E. Niebur, "A model of saliency based visual attention for rapid scene analysis," IEEE Trans. on PAMI, vol.20, no.11 pp.1254–1259, 1998.
- [15] 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.
- [16] V.A. Mateescu and I.V. Bajić, "Attention Retargeting by Color Manipulation in Images," Proc. of ACM Multimedia PIVP, pp.15–20, 2014.
- [17] C. Koch and S. Ullman, "Shifts in selective visual attention: Towards the underlying neural circuitry," Human Neurobiology, vol.4, no.4, pp.219–227, 1985.
- [18] S. Engel, X. Zhang, and B. Wandell, "Colour Tuning in Human Visual Cortex Measured With Functional Magnetic Resonance Imaging," Nature, vol.388, no.6, pp.68–71, 1997.
- [19] R. Achanta, F. Estrada, P. Wils, and S. Susstrunk, "Salient region detection and segmentation," Proc. of International Conference on Computer Vision Systems, pp.66–75, 2008.
- [20] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequencytuned salient region detection," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp.1597–1604, 2009.
- [21] Y. Chuang, L. Chen, G. Chen, and J. Woodward, "Isophote Based Center-Surround Contrast Computation for Image Saliency Detection," IEICE Trans. Inf. & Syst., vol.E97-D, no.1, pp.160–163, 2014.
- [22] A.M. Treisman and G. Gelade, "A feature-integration theory of attention," Cognitive Psychology, vol.12, no.1, pp.97–136, 1980.
- [23] D. Pang, A. Kimura, T. Takeuchi, J. Yamato, and K. Kashino, "A stochastic model of selective visual attention with a dynamic Bayesian network," Proc. of IEEE international Conference of Multimedia and Expo 2008, pp.1073–1076, 2008.
- [24] R.J. Peters and L. Itti, "Applying computational tools to predict gaze direction in interactive visual environments," ACM Transactions on Applied Perception, vol.5, no.2(9), 2008.
- [25] ITU-R BT.500-11. Methodology for the subjective assessment of the quality of television pictures, 2002.
- [26] A. Kimura, R. Yonetani, and T. Hirayama, "Computational Models of Human Visual Attention and Their Implementations: A Survey," IEICE Trans. Inf. & Syst., vol.E96-D, no.3, pp.562–578, 2013.



Hironori Takimoto received his B.E., M.E., and Ph.D. degrees from the University of Tokushima, Japan, in 2002, 2004, and 2007, respectively. He joined Sasebo National College of Technology in 2005. Since 2009, he has worked as an Assistant Professor at Okayama Prefectural University. His research interests include pattern recognition, image processing and human sensing. He is a member of IEEE, IEICE, SICE and IEEJ.



Tatsuhiko Kokui received his B.E. and M.E. degrees from Okayama Prefectural University, Japan, in 2013 and 2015, respectively. His research interests include image processing and human sensing.



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 become 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.