

# Saliency-Based Image Editing for Guiding Visual Attention

**Aiko Hagiwara**  
Chiba University  
Chiba, Japan  
hagi@chiba-u.jp

**Akihiro Sugimoto**  
National Institute of  
Informatics  
Tokyo, Japan  
sugimoto@nii.ac.jp

**Kazuhiko Kawamoto**  
Chiba University  
Chiba, Japan  
kawa@faculty.chiba-u.jp

## ABSTRACT

The most important part of an information system that assists human activities is a natural interface with human beings. Gaze information strongly reflects the human interest or their attention, and thus, a gaze-based interface is promising for future usage. In particular, if we can smoothly guide the user's visual attention toward a target without interrupting their current visual attention, the usefulness of the gaze-based interface will be highly enhanced. To realize such an interface, this paper proposes a method for editing an image, when given a region in the image, to synthesize the image in which the region is most salient. Our method first computes a saliency map of a given image and then iteratively adjusts the intensity and color until the saliency inside the region becomes the highest for the entire image. Experimental results confirm that our image editing method naturally draws the human visual attention toward our specified region.

## Author Keywords

visual attention, saliency map, image editing, gaze-based interface.

## ACM Classification Keywords

I.4.9 Image processing and computer vision: Applications.

## General Terms

Human Factors.

## INTRODUCTION

With the rapid progress of computer facilities, computer usage in every aspect of our daily life has become more and more popular. In fact, we have a drastically increased number of information systems such as the smartphone or the tablet PC. Moreover, electronic bulletin boards can be found everywhere and the electrical audio assistant is available in many cities. These information systems should be in good harmony with human beings and, thus, developing a natural interface is a key issue.

Human gestures, faces and voices are popularly used for an interface between information systems and human beings. Among them, a gaze-based interface is being recognized as

the most important. This is because our visual focus of attention is an important clue for inferring our internal state. A number of studies on a gaze-based interface have been reported. McNamara *et al.*[11] proposed to use an image space modulation method called Subtle Gaze Direction (SGD) [2] for a gaze-guided interaction with images where visual stimulus generated by SGD is presented in the user's peripheral vision area. A gaze-based interaction with a robot was studied in [14] where a robot estimates the human intention from his gaze point to assist him to build up lego-blocks. A gaze-based interface was also studied for wheelchair navigation[13], image editing/rendering[3, 4, 15] and for visual search[17].

It is generally understood that gaze-point selection is controlled by the user's interest and/or intention. Accordingly, a strong external visual stimulus should be presented to draw the user's attention[12]. In order to response the stimulus, we interrupt the current gaze and move the gaze point to the place where the stimulus is presented. This is useful for emergency alert. This, however, interrupts the user's natural action in the sense that it forces the movement of the gaze points using external visual stimulus. Guiding a user's visual attention using eye contact or pointing fingers is also useful[7]. But we cannot use either of these unless people are looking in an indicated direction itself.

It is true that gaze-point selection is controlled by human's interest and/or intention, but gaze-point selection is also affected by the image that a human is looking at. Namely, bottom-up information obtained from the image can be an effective cue for finding out where the gaze point of a human is. Such a cue is called visual saliency[10] and a computational model of it was introduced by Itti *et al.*[8]. The core of visual saliency is a bottom-up and stimulus-driven signal that reveals a point or an area sufficiently different from the surroundings to be worthy of attention. When an image is given, simple visual features are detected first over the image at multiple scales and then they are all combined into a single saliency map which represents a pure salient signal that is independent of the visual features. In the saliency map, the location that has high saliency is regarded to attract visual attention. If we do not care about the user's interest or intention, the highest saliency point in the image is assumed to be selected as the gaze point. This is supported by the insight that the visual attention persuasion filter through geometry modification increased saliency to attract the user's visual attention[9] and by the insight that visual saliency modulation affects the user's attention shift and memory[16].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

PETME'11, September 18, 2011, Beijing, China.

Copyright 2011 ACM 978-1-4503-0930-1/11/09...\$10.00.

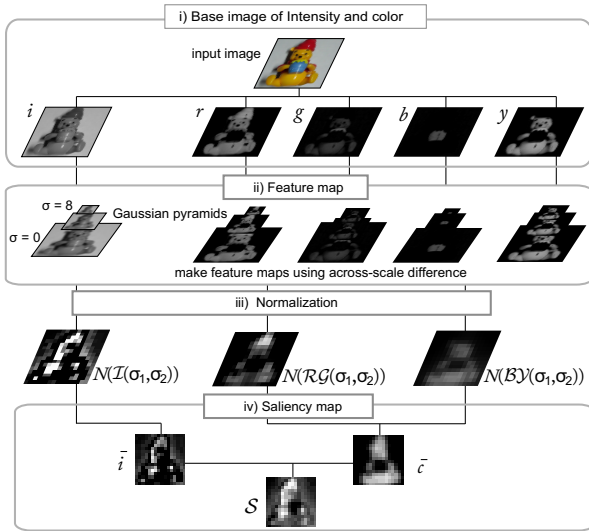


Figure 1. Saliency map

In order to develop an acceptable gaze-based interface, we put ourselves in the position of using visual saliency and directly modulate saliency for guiding the user’s visual attention. Namely, we propose a method for editing an image, for a given region in the image, to synthesize the image in which the region shows the highest saliency. Our method iteratively adjusts the intensity and color so that the saliency inside the region increases while that outside the region decreases. This iteration is carried out until we obtain the image in which a given region is most salient over the entire image. With the image obtained in this way, we can smoothly attract the user’s visual attention to the specified region, without any interruption to the user’s gaze. Experimental results show the effectiveness of our method for drawing visual attention.

### SALIENCY MAP MODEL

Koch and Ullman[10] proposed the visual saliency map concept, which extracts and fuses simple visual features from an input image. After that, Itti *et al.*[8] proposed and developed a fully bottom-up computational saliency map model. Studies on actual gaze measurement have demonstrated that such a saliency map matches the distributions of actual human attention quite well. This model computes saliency using the intensity, color, and orientation contrast as visual features.

Orientation contrast is influenced by the shape of an object in the image. If we adjust this feature, we then need to move the object, and it leads to a great change in image, resulting in an almost equivalent effect that strong external visual stimulus is presented to attract visual attention. This means that adjusting orientation contrast does not match our purpose of naturally guiding visual attention. Therefore, our image editing is restricted to adjusting only the intensity and color. We thus compute the saliency map using only the intensity and color. The outline for computing our saliency map is given in Figure 1. The procedures to compute the saliency map follow [8] except for excluding orientation contrast.

### Base Images of Intensity and Color

Let  $R_{pq}$ ,  $G_{pq}$  and  $B_{pq}$  be a red, green and blue channel for point  $(p, q)$  in an input image. We create five base images. The intensity base image  $i_{pq}$  is given by  $i_{pq} = (R_{pq} + G_{pq} + B_{pq})/3$ . To decouple the hue from the intensity, we normalize  $R$ ,  $G$  and  $B$  channels to obtain the four color base images  $r_{pq}$ ,  $g_{pq}$ ,  $b_{pq}$  and  $y_{pq}$ . Namely,  $r_{pq} = R_{pq} - (G_{pq} + B_{pq})/2$ ,  $g_{pq} = G_{pq} - (R_{pq} + B_{pq})/2$ ,  $b_{pq} = B_{pq} - (R_{pq} + G_{pq})/2$ ,  $y_{pq} = (R_{pq} + G_{pq})/2 - |R_{pq} - G_{pq}|/2 - B_{pq}$ . We assume that hue variations are not perceivable at a low luminance. Thus, if  $r, g, b$  and  $y$  at a given point is smaller than  $1/10$  of the maximum value of  $i$  in the image, we set  $r, g, b$  and  $y$  at that point to zero. In addition, negative values are set to zero.

### Feature Maps

We smooth each base image using a Gaussian filter, and then down-sample the obtained images by skipping every other row and column of the images. This smoothing and down-sampling works to generate Gaussian pyramids[6]. We set scale  $\sigma = 0$  for an input image, and whenever we apply a pair of smoothing and down-sampling, we increase the scale level  $\sigma$  by one. We iterate this procedure to have  $\sigma = 0, 1, 2, \dots, 8$ ; we denote the obtained images by  $i(\sigma), r(\sigma), g(\sigma), b(\sigma)$  and  $y(\sigma)$ , where, for example,  $i(\sigma)$  is a Gaussian pyramid of image  $i$  with scale  $\sigma$ . We note that the size of image  $i(8)$  is  $1/256$  times that of  $i(0)$ .

When given two images at different scales, we first apply the bilinear interpolation[5] to the smaller image so that the two images have the same size. We then take the corresponding pixel-wise subtraction between the two images. This series of procedures is denoted by  $\ominus$ . The intensity feature maps are computed through the across-scale subtraction  $\ominus$  between images at different scales. On the other hand, color feature maps  $\mathcal{RG}(\sigma_1, \sigma_2)$  and  $\mathcal{BY}(\sigma_1, \sigma_2)$  follow the opponent color theory. Thus, we first compute the subtractions in each pair of  $r$  and  $g$ , and  $b$  and  $y$ . Then, we apply the across-scale subtraction to the obtained images. Now, the intensity feature map  $\mathcal{I}(\sigma_1, \sigma_2)$  and color feature maps  $\mathcal{RG}(\sigma_1, \sigma_2), \mathcal{BY}(\sigma_1, \sigma_2)$  are obtained<sup>†</sup>:

$$\mathcal{I}(\sigma_1, \sigma_2) = |i(\sigma_1) \ominus i(\sigma_2)|, \quad (1)$$

$$\mathcal{RG}(\sigma_1, \sigma_2) = |(r(\sigma_1) - g(\sigma_1)) \ominus (g(\sigma_2) - r(\sigma_2))|, \quad (2)$$

$$\mathcal{BY}(\sigma_1, \sigma_2) = |(b(\sigma_1) - y(\sigma_1)) \ominus (y(\sigma_2) - b(\sigma_2))|. \quad (3)$$

We remark that we set  $\sigma = \sigma_1, \sigma_2$  ( $\sigma_1 \in \{2, 3, 4\}, \sigma_2 = \sigma_1 + \delta, \delta \in \{3, 4\}$ ). We have six different combinations of  $\sigma_1$  and  $\sigma_2$  in total. We therefore have 18 feature maps: six for intensity and 12 for color.

From the across-scale subtractions in (1), (2) and (3), we see that the point that is sufficiently different from its surround points has a larger value.

### Normalization

Stimulus that is discriminatively different from that in the surrounding area is important for attracting visual attention. Thus, we normalize the feature maps to prevent important

<sup>†</sup>In this paper, calculation  $+$ ,  $-$  between two images indicates the corresponding pixel-wise summation or subtraction.

stimulus from being restrained after independently merging with other feature maps. This normalization will repress the hoarding of noise. Operator  $\mathcal{N}$  below that acts on a feature map is applied to all feature maps. We note that  $M$  is given beforehand.

---

**Algorithm 1** Normalization operator  $\mathcal{N}$

---

- 1: Normalize the values in the feature map to a fixed range  $[0, M]$ .
  - 2: Set  $\bar{m}$  as the average of the local maximal values over the feature map, each of which is strictly less than  $M$ .
  - 3: Multiply  $(M - \bar{m})^2$  to all the values in the map.
- 

If the value of a point is sufficiently different from that of the surroundings, the concerned feature is effective enough to discriminate the point from the others. In this case,  $(M - \bar{m})^2$  becomes larger and multiplying  $(M - \bar{m})^2$  results in increasing the distinction of the point. On the other hand, when the values are almost the same over a given area, a point is not discriminative even if it has a large value. In that case,  $(M - \bar{m})^2$  becomes smaller and multiplying  $(M - \bar{m})^2$  does not affect the distinction of the point. We denote by  $N(n)$  the normalization of feature map  $n$  with operator  $\mathcal{N}$ .

**Saliency Map**

Superposing all the normalized feature maps allows us to obtain the saliency map. Given two normalized feature maps at different scales, we first apply the bilinear interpolation to the smaller feature map so that the two maps have the same size. We then take the corresponding pixel-wise summation between the two feature maps. This series of procedures is denoted by  $\oplus$ . We apply this operation to all the intensity-based feature maps and the color-based feature maps separately. Combined intensity feature map  $\bar{i}$  and color feature map  $\bar{c}$  are obtained as follows.

$$\bar{i} = \bigoplus_{\sigma_1=2}^4 \bigoplus_{\sigma_2=\sigma_1+3}^{\sigma_1+4} N(\mathcal{I}(\sigma_1, \sigma_2)),$$

$$\bar{c} = \bigoplus_{\sigma_1=2}^4 \bigoplus_{\sigma_2=\sigma_1+3}^{\sigma_1+4} [N(\mathcal{RG}(\sigma_1, \sigma_2)) + N(\mathcal{BY}(\sigma_1, \sigma_2))].$$

We finally combine  $\bar{i}$  and  $\bar{c}$  after their normalization to obtain the saliency map:

$$\mathcal{S} = \frac{1}{2}(N(\bar{i}) + N(\bar{c})). \tag{4}$$

Each point in the saliency map has the  $\mathcal{S}$  value. If a point has a large  $\mathcal{S}$  value, the point highly attracts the visual focus of attention.

**MODULATING SALIENCY**

A given region is not always salient in an input image. We thus iteratively edit the image so that the given region becomes the most salient in the entire image. Our editing method is restricted to adjusting only the intensity and color. Abdollahian *et al.*[1] proposed a method for controlling saliency by adding the motion to the video. However, we are dealing with still images here.

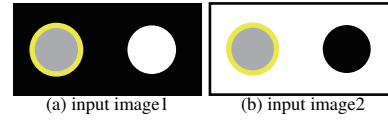


Figure 2. Example of different input images

	Inside region (=raise saliency)	Outside region (=reduce saliency)
Surrounded by darker area	raise intensity	reduce intensity
Surrounded by brighter area	reduce intensity	raise intensity

Table 1. Required change in intensity

**Control Saliency**

In order to raise the saliency of a given region, we have to strengthen visual features inside the region. Furthermore, weakening the visual features outside the region also contributes to reducing the saliency outside the region; as a result, the saliency inside the region increases. The visual features mentioned here refer to the intensity and color.

Controlling the visual features depending on the location in the image is more effective than uniformly controlling them over the entire image. The reason for this is explained using the example in Figure 2. There are two different input images in (a) and (b). The given region are shown in the yellow circles. In order to raise the saliency of the given region in (a), we have to introduce a change in intensity of the gray ball to make it brighter because the surrounding area is darker. Making the intensity of the white ball darker also contributes to raising the saliency of the given region because the surrounding area of the white ball is darker. In contrast, in the case of (b), the gray ball should be changed to darker and the black ball should be changed to brighter for our purpose. As we see, the intensity relationship between a point/region of interest and its surroundings is a key in controlling saliency. Table 1 illustrates how to control the intensity depending on the relationship between a point/region and its surroundings.

Taking into account the relationship given in Table 1, we introduce two parameters: the point variation and the intensity coefficient. The point variation determines the degree of change of the visual features when taking into account the difference from the surrounding area. The intensity coefficient adjusts the direction of change when taking into account whether the point of interest is inside or outside the region.

Let  $R_{pq}, G_{pq}$  and  $B_{pq}$  be a red, green and blue channels of point  $(p, q)$  in an image before updating, respectively.  $R'_{pq}, G'_{pq}$  and  $B'_{pq}$  denote the corresponding channels in the image after updating. The updating procedure in one iteration is illustrated in **Algorithm 2**, where  $V_{R_{pq}}, V_{G_{pq}}$  and  $V_{B_{pq}}$  are the point variations, each of which corresponds to the color channels, and  $w_{pq}$  is the intensity coefficient. Our method it-

erates **Algorithm 2** until the region becomes the most salient in the entire image.

---

**Algorithm 2** Updating each channel with  $V_{R_{pq}}, V_{G_{pq}}, V_{B_{pq}}$  and  $w_{pq}$

---

- 1: Compute point variation  $V_{R_{pq}}, V_{G_{pq}}$  and  $V_{B_{pq}}$ .
  - 2: Compute intensity coefficient  $w_{pq}$ .
  - 3:  $\alpha'_{pq} = \alpha_{pq} + w_{pq}V_{\alpha_{pq}}(\alpha = R, G, B)$ .
- 

### Point Variation $V_{R_{pq}}, V_{G_{pq}}$ and $V_{B_{pq}}$

The point variation reflects how much a feature influences the saliency of a concerned point. For example, if the red color heavily influences the saliency, we enlarge  $V_{R_{pq}}$  of the point. We thus evaluate which color or intensity provides great influence on the saliency of a point. To compute the influence, we return to the procedures for computing the saliency map. We can identify how each visual feature influences on the saliency at a given point by backtracking one by one through the saliency-computation procedures.

Backtracking through the procedures for computing  $S$  in (4) allows us to obtain the following point variations.

$$\begin{aligned} V_{R_{pq}} &= \text{sgn}(i, p, q)\varphi_i + \text{sgn}(r, p, q)\varphi_r - \text{sgn}(g, p, q)\varphi_g \\ &\quad - \text{sgn}(b, p, q)\varphi_b + \text{sgn}(y, p, q)\varphi_y, \\ V_{G_{pq}} &= \text{sgn}(i, p, q)\varphi_i - \text{sgn}(r, p, q)\varphi_r + \text{sgn}(g, p, q)\varphi_g \\ &\quad - \text{sgn}(b, p, q)\varphi_b + \text{sgn}(y, p, q)\varphi_y, \\ V_{B_{pq}} &= \text{sgn}(i, p, q)\varphi_i - \text{sgn}(r, p, q)\varphi_r - \text{sgn}(g, p, q)\varphi_g \\ &\quad + \text{sgn}(b, p, q)\varphi_b, \end{aligned}$$

where  $\varphi_i, \varphi_r, \varphi_g, \varphi_b$  and  $\varphi_y$  are the influence rates of each intensity/color channel at a given point.  $\text{sgn}(\beta, p, q)$  ( $\beta = i, r, g, b, y$ ) is defined by

$$\text{sgn}(\beta, p, q) = \begin{cases} 1 & (\beta_{pq} > \beta_{\text{ave}}) \\ -1 & (\beta_{pq} \leq \beta_{\text{ave}}), \end{cases}$$

where  $\beta_{pq}$  is the value of the base image  $\beta$  at point  $(p, q)$  and  $\beta_{\text{ave}}$  is the average of  $\beta_{pq}$  over the entire image. We remark that  $\varphi_y$  does not contribute to  $V_{B_{pq}}$  because of the shape of  $y$ , which is different from that of the other color base images.

If a point is salient because of an intensity difference from its surrounding points,  $\varphi_i$  becomes larger. If the point is salient because of a difference in color, then  $\varphi_r, \varphi_g, \varphi_b$  and  $\varphi_y$ , become larger. Taking into account the ratio between the intensity and color, we define  $\varphi_i$  as follows (cf. (4)).

$$\varphi_i = \frac{N(\bar{i})}{N(\bar{i}) + N(\bar{c})}.$$

On the other hand, we obtain  $\varphi_r, \varphi_g, \varphi_b$  and  $\varphi_y$  by computing each color's influence on the saliency. In the saliency computation, we first compared opponent color pairs,  $r$  and  $g$ , and  $b$  and  $y$ , and then applied the across-scale subtraction. We thus follow these procedures in the reverse way to identify the influence rate on the saliency of each color channel.

$$\varphi_r = \frac{N(\bar{c})}{N(\bar{i}) + N(\bar{c})} \frac{N(rg)}{N(rg) + N(by)} \frac{N(r^*)}{N(r^*) + N(g^*)},$$

$$\begin{aligned} \varphi_g &= \frac{N(\bar{c})}{N(\bar{i}) + N(\bar{c})} \frac{N(rg)}{N(rg) + N(by)} \frac{N(g^*)}{N(r^*) + N(g^*)}, \\ \varphi_b &= \frac{N(\bar{c})}{N(\bar{i}) + N(\bar{c})} \frac{N(by)}{N(rg) + N(by)} \frac{N(b^*)}{N(b^*) + N(y^*)}, \\ \varphi_y &= \frac{N(\bar{c})}{N(\bar{i}) + N(\bar{c})} \frac{N(by)}{N(rg) + N(by)} \frac{N(y^*)}{N(b^*) + N(y^*)}, \end{aligned}$$

where

$$\begin{aligned} rg &= \mathcal{RG}(2, 5), \\ by &= \mathcal{BY}(2, 5), \\ r^* &= |r(2) \ominus r(5)|, \\ g^* &= |g(2) \ominus g(5)|, \\ b^* &= |b(2) \ominus b(5)|, \\ y^* &= |y(2) \ominus y(5)|. \end{aligned}$$

We remark that we employed the case of  $\sigma_1 = 2, \sigma_2 = 5$  for defining  $rg, by, r^*, g^*, b^*$  and  $y^*$ . We may also consider all the possible cases for  $\sigma_1$  and  $\sigma_2$ , however, it becomes too much complex while not so effective. The finest scale-level is sufficient to investigate an input image in most detail.

### Intensity Coefficient $w_{pq}$

The intensity coefficient depends on not only the  $S$  value of a concerned point but also whether or not the point is inside the region. The intensity coefficient of a point inside the region should be positive while that outside the region should be negative. This is because we increase the saliency inside the region and decrease the saliency outside the region. The magnitude of the intensity coefficient of each point may be proportional to the  $S$  value of the point. But we smooth it over the entire image in order to avoid drastic effect caused by change in sign. Therefore, after attaching an appropriate sign to the  $S$  value of each point, we apply the Gaussian filter to have  $w_{pq}$ .

## EXPERIMENT

### Experimental Procedures

#### Image Editing

1. For a given input image, compute the saliency map to have the  $S$  value of each point in the image.
2. Set a given region  $D$  in the image wherever you want to raise  $S$  values.
3. Set  $k$  to count the number of updating times. Apply **Algorithm 2**. Compute the saliency map after each updating. Iterate the updating (**Algorithm 2**) until the given region  $D$  has the largest  $S$  values over the entire image.

#### Check Visual Attention

Present a set of images to ten subjects. The image set is composed of four different input images and 12 edited images. The 12 edited images were made from the four input images with three different regions. Present each image for five seconds in a random order. Check whether or not subjects gaze at the given region. Here, if the gaze point of a subject does not move more than 0.9 degrees for 0.1 seconds, we regarded that the subject was fixating that point.

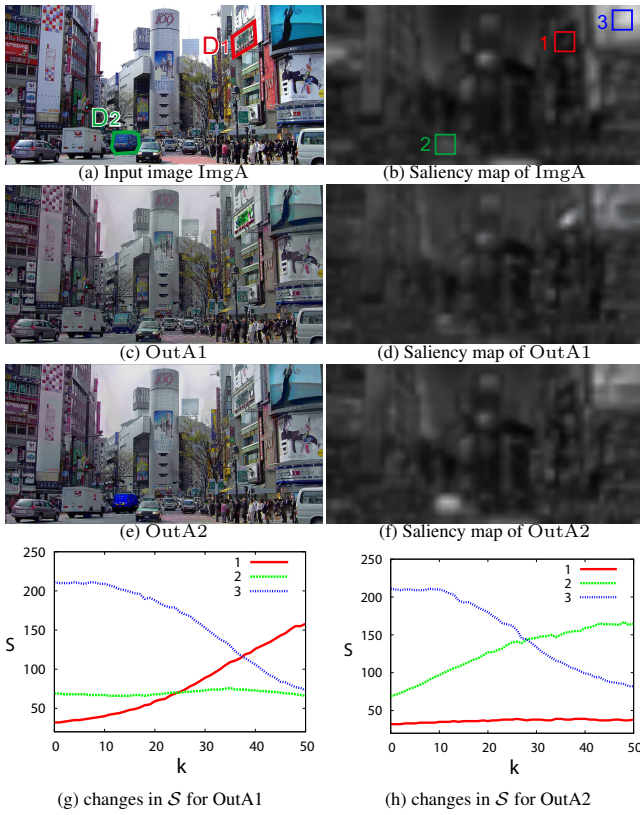


Figure 3. Results from  $\text{ImgA}^\ddagger$  ( $S$  is the value of saliency, and  $k$  is the number of updating times).

We used a Tobii TX300 to track the eye-movements of our subjects.

### Results and Discussion

We show our results of image editing in Figures 3 and 4. The input image and given regions  $D_1, D_2$  are shown in (a). The saliency map of (a) is illustrated in (b) where the locations of sample points 1, 2 and 3 for comparing  $S$  values are specified. Here, sample points 1 and 2 are inside given regions  $D_1, D_2$  while sample point 3 is outside the regions. Sample point 3 is the point with the largest  $S$  value in the input image. (c) and (e) are the output images, which indicate that the given region mostly stands out. The saliency maps of the output images are shown in (d) and (f). (g) and (h) show the changes in  $S$  value depending on  $k$  (the number of updating times) for (c) and (e).

By comparing (a), (c) and (e) we can see that the appearance of output images are different even for the same input image. We may have the impression that color of the output images tends to become gray. The reason for this can be understood as follows. The basic idea of our editing is to strengthen/weaken the visual features that influence the saliency of the point of interest. Accordingly, the white or black colors in the area outside a given region will get close to a medium color, resulting in this change of appearance.

$\ddagger$ (C)2011 BIZUTART PHOTOGRAPH  
<http://www.bizutart.com/photograph/>

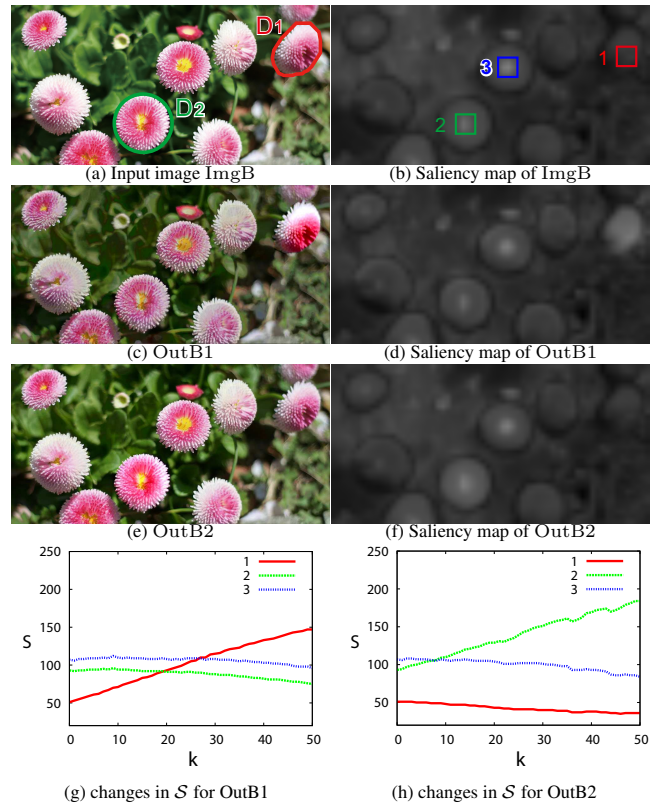


Figure 4. Results from  $\text{ImgB}^\ddagger$

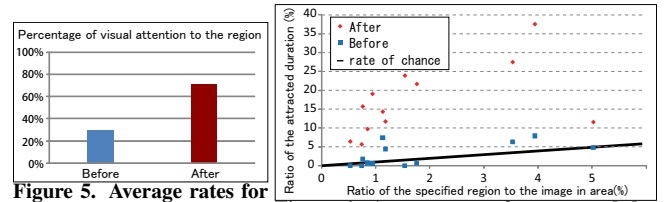


Figure 5. Average rates for drawing visual attentions

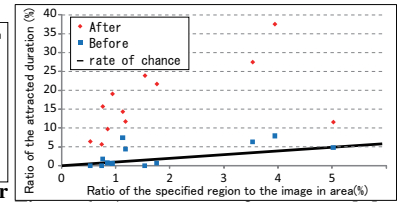


Figure 6. Average rates for attracted duration and region area

From (g) and (h), we see that the  $S$  values inside the given region became largest after editing, so that it was possible to control saliency by our presented editing method. The number of updating times required to make the given region most salient depends on the input image and it is small if the  $S$  value of a given region is large before editing. The amount of increases of saliency also depends on the input image. We see that it is small during the first few times in Figure 3(g). This is caused by the intensity coefficient  $w_{pq}$ . The  $S$  value of sample point 1 in  $\text{ImgA}$  is far smaller than that of the other sample points. Since  $w_{pq}$  depends on the  $S$  value of point  $(p, q)$ ,  $w_{pq}$  becomes small, and thus the amount of changes during one updating is small. This caused the slow increase in the  $S$  value during the first few times in Figure 3(g).

Figure 5 shows the average of rates for drawing visual attention to a given region. If the point fixated by a subject comes inside the given region of an image while presenting the image, we regarded that the region draws the subject's visual attention. "Before" indicates the results when using the input images in the image set, and "After" does the results

using our edited images. We can see that the average rates of drawing visual attention to a given region significantly increase by our editing. We should remark that there was a case where the given region in an edited image never drew any subject's visual attention. This may be because human eye-movements depends on the level of interest as well. We should accept the fact that guiding the user's visual attention with 100% is impossible.

To closely look into subject's visual attention, we calculated the total duration in which a subject kept fixating points inside the region of an image, and we call it the attracted duration. The average ratio of the attracted duration to five seconds, i.e., the duration of presenting the image, was 2.9% for "Before" while it was 17.1% for "After". Moreover, the duration in which a subject kept fixating points inside the region for the first time was 0.09 seconds on the average for "Before" whereas it was 0.39 seconds for "After". These indicate that our editing method indeed attracted visual attention. We note that our specified regions dominated 1.83% of the image on the average.

Figure 6 shows the relationship between the average ratio of the attracted duration to five seconds and the ratio of the specified region to the entire image in area. The solid line indicates the rate of chance, namely, the locus of the relationship if we randomly and uniformly gaze points over the image. "Before" closely correlates with the rate of chance while "After" is far above the rate of chance, resulting in the conclusion that the specified region significantly attracted visual attention from the statistical point of view.

## CONCLUSION

We presented image editing method for naturally guiding the user's visual focus of attention. Our method is different from conventional methods in that our method is based on a fully bottom-up information gathering obtained from an image. Namely, we stood on the position of using visual saliency and presented iteratively editing the intensity and color of an image until the saliency of a given region becomes the highest over the entire image. We confirmed that the user's visual attention is indeed attracted toward our specified region when using our method. The advantage of our method is that we do not need to present any visual stimulus to a user in order to attract his visual attention.

## ACKNOWLEDGEMENTS

This work was in part supported by JST, CREST.

## REFERENCES

1. G. Abdollahian, C. Taskiran, Z. Pizlo, and E. Delp. Camera motion-based analysis of user generated video. *IEEE Transactions on Multimedia*, 12(1):28–41, 2009.
2. R. Bailey, A. McNamara, N. Sudarsanam, and C. Grimm. Subtle gaze direction. *ACM Trans. Graph.*, 28(100):1–14, 2009.
3. J. Collomosse and P. Hall. Painterly rendering using image salience. In *Proc. of Eurographics UK Conference (EGUK'02)*, pages 122–128, 2002.
4. D. DeCarlo and A. Santella. Stylization and abstraction of photographs. *ACM Trans. Graph.*, 21:769–776, 2002.
5. M. Gangnet, D. Perny, and P. Coueignoux. Perspective mapping of planar textures. *Computers & Graphics*, 8(2):115–123, 1984.
6. H. Greenspan, S. Belongie, R. Goodman, P. Perona, S. Rakshit, and C. Anderson. Overcomplete steerable pyramid filters and rotation invariance. In *Proc. of Computer Vision and Pattern Recognition*, pages 222–228, 1994.
7. M. Imai, T. Ono, and H. Ishiguro. Physical relation and expression: Joint attention for human-robot interaction. *IEEE Transactions on Industrial Electronics*, 50(4):636–643, 2003.
8. L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, 1998.
9. Y. Kim and A. Varshney. Persuading visual attention through geometry. *IEEE Transactions on Visualization and Computer Graphics*, 14:772–782, 2008.
10. C. Koch and S. Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. *Human neurobiology*, 4(4):219–227, 1985.
11. A. McNamara, R. Bailey, and C. Grimm. Improving search task performance using subtle gaze direction. In *Proc. of the 5th symposium on Applied perception in graphics and visualization (APGV'08)*, pages 51–56, 2008.
12. Y. Ogata and K. Uchikawa. Does a stimulus not detected induce saccade? *Journal of Vision*, 9(14):47, 2009.
13. T. Razali, R. Zhu, K. Yoshinori, and K. Yoshinori. Smart wheelchair navigation based on user's gaze on destination. In *Proc. of Intelligent Computing in Computer Vision*, 3, pages 387–394, 2010.
14. K. Sakita, K. Ogawara, S. Murakami, K. Kawamura, and K. Ikeuchi. Flexible cooperation between human and robot by interpreting human intention from gaze information. *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pages 845–851, 2004.
15. A. Santella, M. Agrawala, D. DeCarlo, D. Salesin, and M. Cohen. Gaze-based interaction for semi-automatic photo cropping. In *Proc. of CHI '06*, pages 771–780, 2006.
16. E. Veas, E. Mendez, S. Feiner, and D. Schmalstieg. Directing attention and influencing memory with visual saliency modulation. In *Proc. of CHI'11*, 2011.
17. M. Verma and P. McOwan. Generating customised experimental stimuli for visual search using genetic algorithms shows evidence for a continuum of search efficiency. *Vision research*, 49(3):374–382, 2009.