Focusness Extraction from Images for Automatic Camera Control by Measuring Objectness

Satomi Tanaka[1,2], Kaori Kiyomitsu[1], Masanori Maki[1], Akihiro Kakinuma[2], Naohiro Kamijo[2], Hiroshi Takahashi[2] and Norimichi Tsumura[1]

1) Graduate School of Advanced Integration Science
Chiba University
1-33 Yayoi-cho, Inage-ku, Chiba-shi, Chiba 263-8522, Japan
tsumura@faculty.chiba-u.ac.jp

2) Ricoh Company, Ltd.
16-1 Shinei-cho, Tsuzuki-ku, Yokohama-shi, Kanagawa 224-0035, Japan
satomi.st.tanaka@nts.ricoh.co.jp

Abstract

Current digital cameras have various automatic control systems. In automatic camera systems, extracting focusness from an image is a very important problem. Automatic extraction of the main subject makes taking photos very easy, even for an amateur photographer. Methods have been proposed to evaluate focusness by visual saliency, which assume that an area with high saliency also has high focusness. However, various differences exist between focusness and saliency. In this study, we compare the values between focusness and saliency maps. We evaluate the focusness of 80 images in an image evaluation experiment with 20 observers. Saliency maps are calculated using six conventional algorithms. We show that the individual variations of focusness are very few in images that include only one major object. Furthermore, we apply a GIST feature to the saliency method by using a center-surround histogram and extract focusness from images with high accuracy.

Keywords: Visual saliency, Salient object detection, Objectness, Focusness, GIST feature

1. Introduction

Current digital cameras have various automatic control systems, such as auto-focus, auto-white balance, and auto-exposure, which enable an amateur photographer to take excellent photos. In particular, an important function for an automatic camera is extracting the main subject from the photographic space. If the main subject(s) is (are) known, the various camera parameters, e.g., the point of focus, f-number, and depth of field (DOF), can be set automatically.

Many studies modify these parameters after taking a single photograph. Ng *et al.* [1] proposed a method to refocus images from the light field information from a plenoptic camera. Nagahara *et al.* [2] expanded the degrees-of-freedom (DOFs) by using a sweeping camera. Ito *et al.* [3] simulated a photographic image using compressive epsilon photography. Although these methods enable us to later change the point of focus and DOFs, they require unique and expensive equipment or long processing times. In contrast, the methods for extracting the main subject while taking photos are expected to provide the appropriate parameters to users.

Therefore, we define "focusness" as the metric indicating the degree to which the main subject is chosen in a photography space; that is, focusness indicates the point that a photographer would focus on when taking a picture. Typically, an object is selected as the main subject of the image. Several methods have been proposed to extract focusness from an image. Because the human face is typically chosen as the main subject, current digital cameras detect a human face in the photographic space and focus on it. However, recent cameras are also able to detect a moving object as the main subject. A new method has also been proposed to detect the main subject by using visual saliency [4].

Visual saliency is the perceptual degree of an object, person, or standout pixel relative to its neighbors and thus captures our attention. A number of researchers have proposed various types of algorithms to calculate visual saliency on an image [5]–[11]. The main subject detection method [4] assumes that an area with high saliency is the same as an area with high focusness. These methods are used not only for automatic camera control systems but also for cropping main subjects from images [12], [13]. It is also used in the navigation system [14] that performs visual feedback control using the salient region. In general, we pay attention to the area that has high saliency in the image; however, that area is not necessarily selected as the main subject. In other words, focusness and bottom-up visual

saliency are somewhat different from each other.

In this study, we evaluate the focusness of 80 images in an image evaluation experiment with 20 observers. We compare the values between focusness and saliency maps calculated by six bottom-up saliency algorithms [1], [6]–[11]. We find the properties of an image when conventional saliency algorithms can exactly extract focusness from the image as well as the properties when conventional saliency algorithms cannot extract focusness. Based on the results of the conventional saliency methods, we apply the GIST feature [15][16] to a saliency algorithm [11] which uses a center-surround histogram to calculate saliency from images. Our method is similar to the target detection method in satellite images [17]. This method detects boats, buildings and airplanes by evaluating the GIST features of patch regions divided from large-scale satellite images. In contrast, we evaluate the differences in GIST features between center of the image and the surrounding areas and we extract distinct objects with different textures (eg rough or fine, natural or artificial) with the surroundings. As a result, our saliency method is able to more accurately extract focusness from an image compared with the other six saliency methods, as shown in Figure 1. It should be noted that, in order to avoid copyright issues, some of the original images used in the experiments were replaced by alternatives. Similarity between an original and alternative image was determined by calculating the three-dimensional Euclidean distances between the RGB values (0 to 255) of pixels in the former and the corresponding pixels in the latter. Image pairs were considered to be similar if the average three-dimensional Euclidian distance was less than 10.

In Section 2, we describe the image evaluation experiment for focusness. In Section 3, we describe the extraction of focusness by using conventional saliency algorithms. In Section 4, we describe our approach to extract focusness by improving the saliency algorithm (improved GIST feature). In Section 5, we summarize our research.

2. Focusness evaluation

2.1 Evaluation experiment

In this section, we obtain focusness in an image-by-image evaluation experiment. Figure 2 shows the experimental environment. Twenty observers look at an image and search for the point that is the most preferred to be finely focused, and then select it as the focus point. The focus point selected by the

observers is measured by an eye-tracking system (Tobii X1 Light [18]). The system measurement error is 2 degrees of the viewing angle. To select the focus point, observers press the key while gazing at the focus point.

We use 80 pan-focus images taken with a low f-number. All images are the same size, 1920×1200 pixels, and the monitor resolution is the same as the image resolution. If an image is smaller than 1920×1200, the pixel value 0 is padded into the vacant pixels to keep the same image size. These images are classified as natural or man-made, as shown in Fig. 3(a). To determine the classification, we use the term "mean depth", which roughly shows the distance from the camera to the background. In a close-up image, the mean depth is near 0 meters. In a panoramic image, the mean depth is over 1000 meters. A portrait image and a near-landscape image are 10 meters and 100 meters, respectively. We classify images according to the type of four mean depths, as shown in Fig. 3(b). Images are displayed on the 24-inch monitor and the order of display is fixed for all observers.

2.2 Focus map

Figure 4 shows the steps taken to create a focus map that depicts the focus points selected by 20 observers for each image. First, the focus point is measured from the observer's gazing position as 1 pixel on the image by the eye-tracking system. Second, the focus point is blurred to 90×90 pixels with a filter to account for system measurement error. This process is implemented for the focus points of the 20 observers. Finally, we integrate these focus point images into an image as a focus map. By overlaying the focus map on the original image, we can present "focusness" visually on the original image. In Fig. 4, the individual variation of focusness is two objects.

We define the variance V of the observers' focus point on the image. The variance of the 20 observers is calculated, and the results are as shown in Fig. 4. According to the results, when the variance is less than or equal to 10, the focus map is not different between the observers.

2.3 Individual variation of focusness

We calculate the concordance rate of focusness on each image. If the same object in an image is selected by all 20 observers, this rate is 100%. Focusness has some individual variations, because the focus point

depends on individual preference. Figure 5 shows example images and their concordance rate. The rate is small in Fig. 5(a) and large in Fig. 5(b). We find that the individual variations of focusness vary due to the properties of the images. On an image including only one major object or an image that attracts the observers' attention by image composition, almost all observers select the same focus point, so that the individual variation of focusness is very small. In contrast, on an image including many objects or having no attention-getting area, the individual variation is larger. For images such as those shown in Fig. 5(a), we expect that focusness can be extracted by simply using bottom-up saliency algorithms.

Therefore, we select 16 images from all 80 images whose concordance is generally 100%. Then, we attempt to extract focusness on these selected images. In the next section, we calculate saliency maps by using the six conventional algorithms and compare their focusness and saliency maps.

3. Focusness evaluation using visual saliency

In this section, we compare focusness and saliency maps by using the six conventional methods. Here, a saliency map is a map indicating the image area where we tend to pay attention. The methods to create saliency maps calculate saliency or objectness, or both, from the images. First, we introduce these methods. Second, we calculate the accuracy of these algorithms for extracting focusness.

3.1 Saliency maps

Itti *et al*. [6] proposed the most general visual saliency method based on a biological model. Using the bottom-up visual attention theory, they calculated the saliency of pixels by the contrast against its neighbors in color, intensity and edge orientation. They assumed that the standout pixels attract human attention. Their method is here referred to as "IT". Harel *et al*. [7] proposed the "GB" method to improve IT as a coherent saliency map. Hou *et al*. [8] estimated saliency in the frequency domain, which is referred to as the "SR" method. The "IS" method [9] is a hybrid approach of IT and SR. The IT, GB, SR, and IS methods are based on edge detection, and the pixels whose intensity or color are very different from neighboring pixels are assumed to have high saliency.

Achanta *et al*. [10] proposed the "IG" method, which takes the difference between the mean pixel value among all pixels in the image and each pixel value. Liu *et al*. [11] proposed the method "LI" to

extract “objectness” by using a center-surround histogram. Objectness is the degree to which a given image area belongs to an object. People, vehicles, animals typically have high objectness, while areas such as background, sky and general landscape have low objectness. Therefore, not only the visual saliency but also the objectness influence human attention. LI assumes that the difference between color histograms of the object and its surroundings is very large. As shown in Fig. 6, the LI method extracts the objectness of the center area by sliding the center-surround rectangle. The IG and LI methods are based on color.

We calculate the saliency maps on 16 images by using the six methods. The examples of the results are shown in Fig. 7. Here, the padding area in an image is removed from the original image before calculating saliency.

3.2 Comparison between focusness and saliency

We show examples of focus maps and the saliency maps calculated by the six methods. We compare a focus map and the saliency maps by using receiver operating characteristic (ROC) curves [19] through the steps shown in Fig. 8. Generally, ROC curves indicate the performance of detection methods. In our research, a ground truth is a focus map and the detected result is a saliency map. We describe the method to quantify the similarity of a saliency map and a focus map by using a ROC curve, as follows. First, we binarize a focus map with an intensity image value k. Then, we obtain a ground truth referred to as the goal map. We use one fixed value, k = 46, for 16 focus maps for the binarization. Second, a corresponding saliency map is assigned a threshold with cut-off points. We use 32 cut-off points from 0 to 256 in intensity value. Therefore, 32 binary maps are created. Equation (1) shows the metric used to evaluate the similarity between a focus map and a saliency map; these are general terms in ROC curve analyses.

$$TPF = \frac{\left| P_{saliency} \cap P_{goal} \right|}{\left| P_{goal} \right|}, \; FPF = \frac{\left| P_{saliency} \cap \left(P_{all} \quad P_{goal} \right) \right|}{\left| P_{all} \quad P_{goal} \right|} \tag{1}$$

where P_{all} is the total number of pixels composing an original image, and P_{goal} and $P_{saliency}$ are the positive area in each map. The ROC curve is composed of dots (FPF, TPF) with the vertical axis as the true positive fraction (TPF) and the horizontal axis as the false positive fraction (FPF). By carrying out the

above steps, we obtain the ROC curves. We show some examples of ROC curves for comparing a focus map and saliency map in Fig. 9. Figure 10 shows the ROC curves of the 6 methods' detected results.

We fit the ROC curves to Eq. (2) and calculate the coefficient α. As the focusness and the saliency maps become increasingly similar, the coefficient α becomes larger. Figure 11 shows the coefficient α for all 16 images.

$$\mathrm{TPF} = -\exp(\alpha \cdot \mathrm{FPF}) + 1 \ . \tag{2}$$

As shown on the "flower" image in Fig. 7, the saliency map using LI is similar to the focus map as the ground truth of focusness. In contrast, on the "tree" image, the edge-based saliency maps are similar to the focus map. However, on the "stone sculpture" image, all methods extract focusness inaccurately, because the background of the stone sculpture image has high intensity on all six saliency maps. As stated previously, focusness indicates the degree that the main subject is chosen in the photo. Then, a method such as LI extracts focusness from an image by regarding a standout object as having high objectness. In the LI method, the center-surround histogram compares only the color value without edge intensity. Therefore, in this study, we proposed to apply the GIST feature [15], [16] to the LI method by using the center-surround histogram.

4. Improved method

We consider objectness to be a major factor of focusness, because we assume that high objectness strongly influences focusness. Specifically, a single object stays in the center of an image. On the other hand, high visual saliency doesn't necessarily influence on focusness. For example, in the case of two objects—one with high saliency and one with low saliency—in an image, the former will have higher 8 focusness than the latter. Therefore, in this paper, we improve the conventional LI method by using GIST. GIST [15], [16] is the global feature of an image for obtaining the abstract representation of a scene, an impression of the image, and the included main object. This feature can classify images into scene categories, such as a city or mountain. Recently, this feature is used to the recognition systems of scene [15] or layout [14]. It is also used to describe the structure to for matching between stereo images [20] and high speed object detection from satellite image [17].

We next describe our proposed approach applying GIST to the center-surround histogram for extracting a main subject that does not stand out in color. First, we introduce the GIST feature. Second,

we describe our approach. Finally, we confirm our improvement of the conventional method LI by comparison to the other conventional methods.

4.1 Improved saliency method using GIST

Figure 12 shows the steps of calculating GIST. First, the input intensity image is divided into blocks. Each blocked image is a convoluted steerable pyramid [21] with four scales and four orientations, and then we obtain 16 thresholded gradient images. We express these gradient images as $I_{k,l}$ with orientation k and scale l, where $m_{k,l}$ is the mean value of $I_{k,l}$. GIST is a vector composed of a 4×4 coefficient $m_{k,l}$ for each scale and each orientation of one block image. Thus, the GIST abstract image represents the scene with scale, orientation, and location of the block. In a previous study [22], GIST was extracted from a color image and these images are decomposed into four scales and eight orientations.

The polar graphs in Fig. 13 show that the GIST feature roughly represents the type of object. The four color lines of blue, green, red, and black indicate the coefficients of scale 1–4, respectively. In man-made objects such as buildings or signatures, GIST is larger at 0 and 90 degrees. GIST is largest at 0 degrees in the shrub image that has hard vertical lines. In contrast, GIST is largest at 90 degrees in the sky image. A natural image such as the flower is orientation invariant. By calculating GIST, therefore, we can recognize the type of object included in the image. To obtain the difference between two GIST feature vectors, we integrate the absolute values of the difference between coefficients with the same scale, orientation, and block.

In our approach, we average the intensities of the edge detection images among the center and surround rectangles, and then compare the two GIST feature vectors of the center and surround to calculate objectness.

4.2 Our approach

Figure 14 shows the concept of our approach. If different objects are in the center and the surround rectangles, these GISTs are not similar and the differences are large. If the same objects are in two rectangles, the difference in GIST is small. That is, our approach calculates objectness by using the GIST feature against the LI method that uses the color histogram. In our research, we assume that the center-

surround difference in GIST becomes larger when the standout object exists in the center rectangle, and we assume that focusness is the same as objectness when high objectness is concentrated at a local area, because no other object for focus is present.

In the case that objectness by using GIST is higher than that by using color, we apply our improved method to calculate focusness. In contrast, in the case that objectness by using color is higher than that by using GIST, we use the LI method to calculate focusness. In this way, we can extract focusness from an image whenever the color or edge of the main subject does not stand out.

4.3 Results of our approach

Figure 15 shows the results of our method. A red square indicates an area with the highest objectness. Figure 15(a) shows the results of measuring objectness by using only color information, that is, by the conventional LI method. In this case, we find that the red squares and the focus maps are similar. Figure 15(b) shows the results of measuring objectness by using GIST, which is our proposed approach. Although the red squares in the results of LI and the focus maps are different, the red squares in our approach and the focus maps are similar. Therefore, we can extract focusness more exactly than we can with LI. We calculate the accuracy as log (TPF/FPF) against the conventional LI and SR methods. TPF and FPF are the values when log(TPF/FPF) is the highest value among the 32 cut-off points. It is better to not use alpha in Eq. (3) because alpha is not always available; the ROC curve is not necessarily in a shape to calculate alpha, as in the LI curve shown in Fig. 10(b). Therefore, alpha is not suitable to compare a focus map and saliency map in our research. Therefore, we regard log(TPF/FPF) as being more suitable than alpha.

By using our proposed method, a total of 12 images were found to have a higher accuracy of focusness when the accuracy of log(TPF/FPF) is higher than a threshold value (here, the value is set as 1). On the other hand, among the 16 images, seven images had higher accuracy when using the conventional LI or SR methods.

Figure 16 shows a graph indicating the results of the improvement achieved with our approach for calculating objectness by using either color or GIST. The accuracy is calculated as log (TPF/FPF) for our approach and the conventional LI and SR methods. Our approach more accurately extracted a total of 12

images for focusness with the threshold of the accuracy value as (1) among the 16 images.

In the experiment, the accuracies of our proposed method are equal to or better than the those of LI. However, the accuracy becomes worse than LI when we apply our method to some images of flowers. The reason for this is that LI uses only RGB color features, while our method uses only the edge features as the GIST of intensity. For an image that has only color disparity compared with the surroundings and does not have the disparity of texture, as in flower images, the LI method may provide a better result than our method. In our latest study, the method that we approached in Section 4.2 was improved by replacing the grayscale input images with the RGB color input image, as shown in Fig.17. In this method, objectness is calculated from each image of RGB using GIST features, and linearly combine with multiplying arbitral weight parameters. Fig.18 shows the results of focusness using the improved method shown in Fig.17. Even if an image have only color disparity, this method's result (Fig.17 f) matches to ground-truth (Fig.17 b) as is the case with the LI's result (Fig.17 d), whereas the method introduced in Section 4.2 doesn't match (Fig.17 e).

5. Summary and Conclusion

We presented an improved model of extracting focusness by applying GIST to the visual saliency algorithm. We calculated objectness by comparing the GIST of the center and the surround. If the objectness is high and concentrated at a point, we assumed that objectness is equal to focusness. Objectness obtained using our proposed approach was similar to the experimentally obtained focus map.

On the other hand, if the objectness is low or scattered around an image, objectness is insufficient to predict focusness. Our research goal is to support people taking photos by automatically focusing on the main subject, on the region where most people usually want to focus. The 80 images used in Section 2 are various types of photographs, such as close-up, portrait, near-landscape and panoramic images. In particular, a panoramic photograph has either no object or a too small object; in this case, focusness is scattered and has a large individual variance. In the case that an image has no object, a small object, or many objects, we should not apply our method, because it may discourage photographers. Therefore, in this research, we selected 16 images including a major object with low individual variety for our experiment.

In future work, we intend to develop a method to judge whether the automatic focus function should be applied. We suggest that this task will be solved by using the variance of objectness in images. Furthermore, it is valuable to apply our method to images with focus points that vary for different users. We suggest that learning the image features of an object and a scene will enable individuals to select the main object in that scene. Then, we should calculate the priority of the detected objects in an image based on individual preference. Herrera *et al.* [23] stated that the objects that are semantically inconsistent with the context of the scene are gazed more quickly and prolonged times than objects that are semantically consistent. We suggest that the image context consisting of objects and its surroundings is an important point to select the most important object from other objects.

In addition, we will consider the effect of image composition on focusness for attracting attention. For example, the vanishing point, the vanishing line, and the horizontal line in images not only guide the eye's attention but also influence the impression of an image. Thus, image composition is assumed to have an effect on focusness as well as objectness. GIST is also useful in scene recognition [15] for distinguishing the composition of images. Therefore, we will attempt to find the relevance between focusness and composition.

Our method does not require prior conditions. However, to consider the individual variations of focusness, we should learn an observer's personal preferences, e.g., a favorite color, an object, and his/her photo-taking habits, to extract individual focusness more accurately. At this time, the OS we use is Windows 7 Professional (64 bit), the CPU is Intel Core i7-4790K (3.60 GHz), and our source code is written in MATLAB. Our approach requires a 0.5 second calculation per image, so we need to achieve a shorter processing time for applying this method in a practical camera system. The processing time will be shorter than 0.01 seconds per image by rewriting the source code with C/C++.

References

[1] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," *Comput. Sci. Tech. Rep. CSTR*, vol. 2, no. 11, 2005.

[2] H. Nagahara, S. Kuthirummal, C. Zhou, and S. K. Nayar, "Flexible depth of field photography," in *Computer Vision–ECCV 2008*, Springer, 2008, pp. 60–73.

[3] A. Ito, S. Tambe, K. Mitra, A. C. Sankaranarayanan, and A. Veeraraghavan, "Compressive Epsilon Photography for Post-Capture Control in Digital Imaging," *ACM Trans. Graph.*, vol. 33, no. 4, 2014.

[4] C. LTD, "IMAGE PROCESSOR AND METHOD," *JP002011035633A*.

[5] T. Judd, F. Durand, and A. Torralba, "A Benchmark of Computational Models of Saliency to Predict Human Fixations," *Mit-Csail-Tr-2012*, vol. 1, pp. 1–7, 2012.

[6] L. Itti, C. Koch, and E. Niebur, "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis," vol. 20, no. 11, pp. 1254–1259, 1998.

[7] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Advances in neural information processing systems*, 2006, pp. 545–552.

[8] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, 2007, pp. 1–8.

[9] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *Pattern Anal. Mach. Intell. IEEE Trans.*, vol. 34, no. 10, pp. 1915–1926, 2012.

[10] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Computer vision and pattern recognition, 2009. cvpr 2009. ieee conference on*, 2009, pp. 1597–1604.

[11] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H.-Y. Shum, "Learning to detect a salient object," *Pattern Anal. Mach. Intell. IEEE Trans.*, vol. 33, no. 2, pp. 353–367, 2011.

[12] M. Nishiyama, T. Okabe, Y. Sato, and I. Sato, "Sensation-based photo cropping," in *Proceedings of the 17th ACM international conference on Multimedia*, 2009, pp. 669–672.

[13] J. Luo, "Subject content-based intelligent cropping of digital photos," in *Multimedia and Expo,*

2007 IEEE International Conference on, 2007, pp. 2218–2221.

[14] C. K. Chang, C. Siagian, and L. Itti, "Mobile robot vision navigation & localization using gist and saliency," in *IEEE/RSJ 2010 International Conference on Intelligent Robots and Systems, IROS 2010 - Conference Proceedings*, 2010, pp. 4147–4154.

[15] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vis.*, vol. 42, no. 3, pp. 145–175, 2001.

[16] A. Torralba and A. Oliva, "Depth estimation from image structure," *Pattern Anal. Mach. Intell. IEEE Trans.*, vol. 24, no. 9, pp. 1226–1238, 2002.

[17] Z. Li and L. Itti, "Saliency and gist features for target detection in satellite images," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 2017–2029, 2011.

[18] T. T. AB, "Tobii X1 Light Eye Tracker." .

[19] A. K. Akobeng, "Understanding diagnostic tests 3: receiver operating characteristic curves," *Acta Paediatr.*, vol. 96, no. 5, pp. 644–647, 2007.

[20] J. L. Herrera, J. Konrad, C. R. Del-Bianco, and N. Garcia, "Learning-based depth estimation from 2D images using GIST and saliency," in *Proceedings - International Conference on Image Processing, ICIP*, 2015, vol. 2015-December, pp. 4753–4757.

[21] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *Image Processing, International Conference on*, 1995, vol. 3, p. 3444.

[22] A. Oliva, A. Torralba, M. S. Castelhano, and J. M. Henderson, "Top-down control of visual attention in object detection," in *Image processing, 2003. icip 2003. proceedings. 2003 international conference on*, 2003, vol. 1, pp. I–253–6 vol. 1.

[23] G. Underwood, L. Humphreys, and E. Cross, "Congruency, saliency and gist in the inspection of objects in natural scenes," in *Eye Movements*, 2007, pp. 563–579.

Figure captions

Figure 1 Our results extracts focusness from images. First row is original image, second row is saliency map using Itti *et al.* [8] method, and third row is our method's result. The red rectangles indicate the extracted focusness.

Figure 2 Experimental environment for evaluating focusness on images. Focusness is measured by the eye tracking system Tobii.

Figure 3 The classification of images. These images are classified into (a) natural or man-made, (b) mean depth of image.

Figure 4 Focus map are measured from the observer's gazing position.

Figure 5 Examples of focus map and their concordance rate. The rate is small in (a) and large in (b).

Figure 6 LI method. To extract objectness of the center area by sliding the center-surround rectangle.

Figure 7 Saliency map on three images by six conventional methods. The bottom row is the focus map obtained in section 2.

Figure 8 The calculating steps for ROC curves and comparing steps between the focus map and the saliency maps.

Figure 9 The ROC curve with 5 cut-off points with the vertical axis as TPF and the horizontal axis as FPF.

Figure 10 The ROC curves corresponding (a) image # 79 and (b) image #12. On the left, top image is

the focus map, middle image is the goal map thresholded from the focus map as is top image, and the bottom image is the original image.

Figure 11 The graph of comparison results. This Accuracy is sorting order of accuracy of LI method.

Figure 12 The steps to calculate GIST feature.

Figure 13 GIST represent the type of object.

Figure 14 The concept of our approached method.

Figure 15 Results of our method. We extract focusness by calculating objectness using (a) color, or (b) GIST. To decide which method is used, we compare objectness using color and GIST, and big one method is used.

Figure 16 The result of improved accuracy by using our method.

Figure 17 The overview of our approached method using not only texture disparity but also color disparity to calculate objectness.

Figure 18 The result of the method shown in Figure 17.

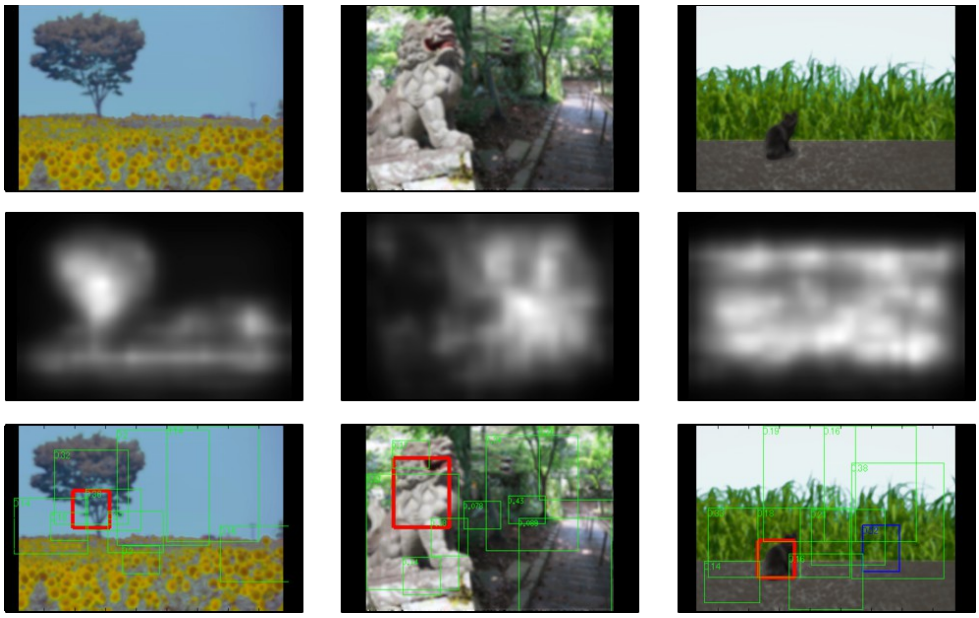


Figure 1

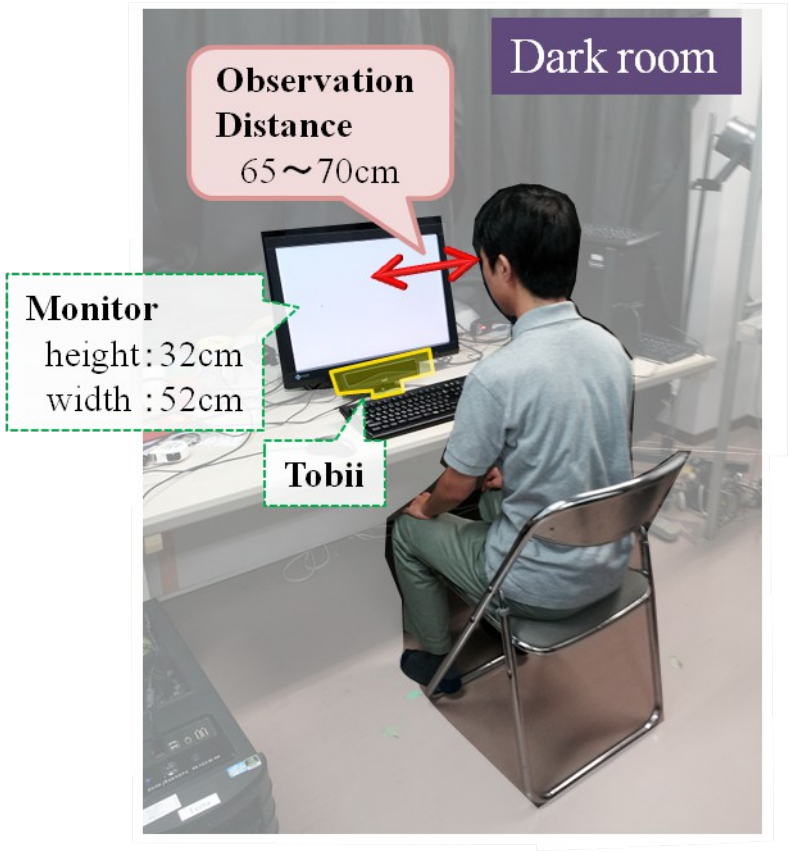
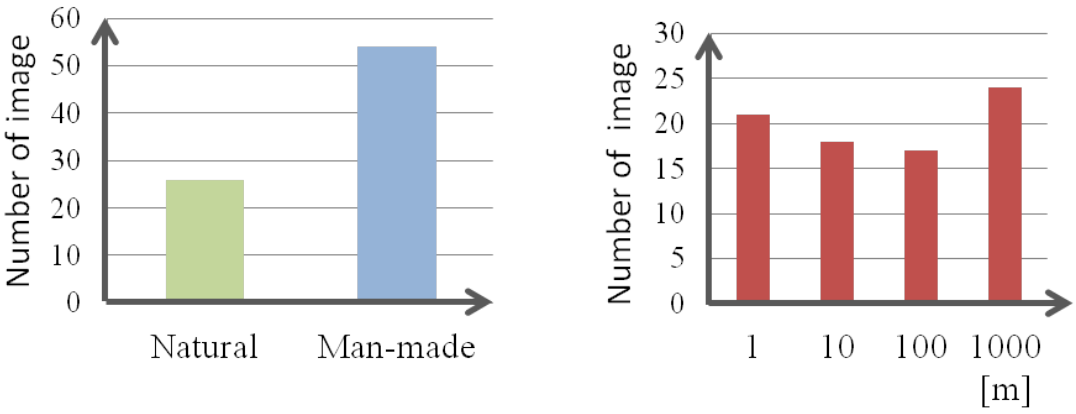

Dark room
Observation
Distance
65～70cm
Monitor
height : 32cm
width : 52cm
Tobii

Figure 2



(a) Natural or Man-made

(b) Mean depth

Figure 3

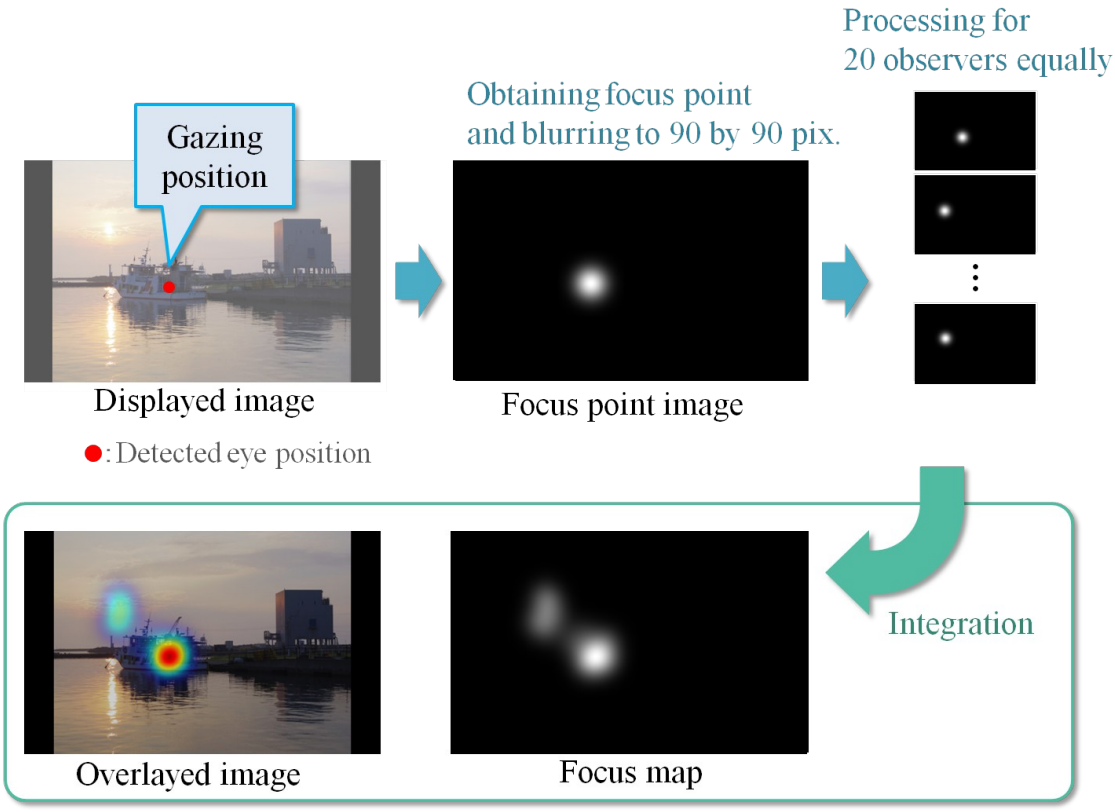
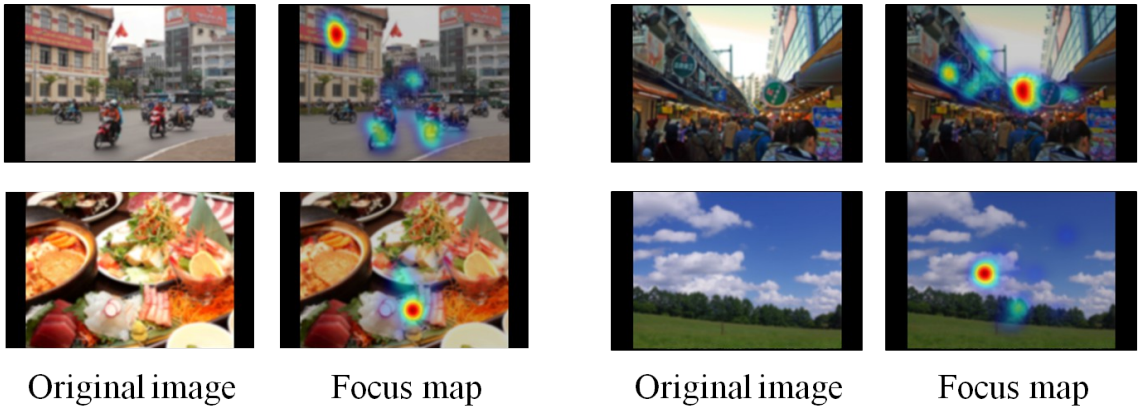

Processing for
20 observers equally
Obtaining focus point
and blurring to 90 by 90 pix.
Gazing
position
Displayed image
●: Detected eye position
Focus point image
Integration
Overlayed image
Focus map

Figure 4

(a) Small individual variation, 95% ~ 100%



(b) Large individual variation, 30% ~ 45%

Figure 5

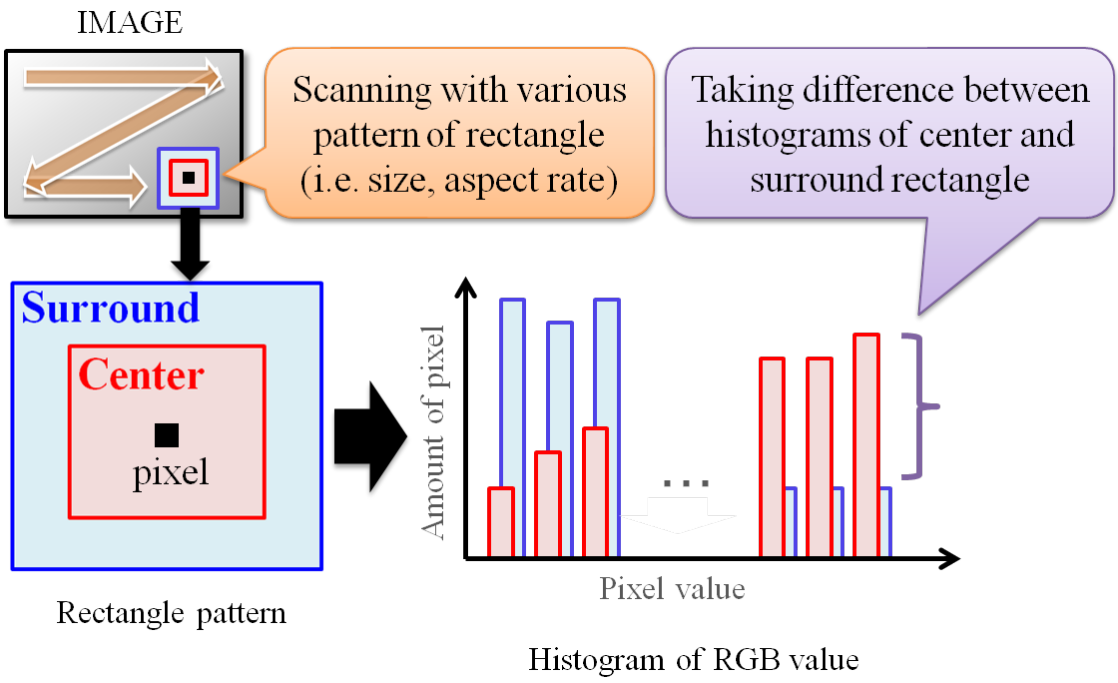


IMAGE
Scanning with various pattern of rectangle (i.e. size, aspect rate)
Taking difference between histograms of center and surround rectangle
Surround
Center
pixel
Amount of pixel
Pixel value
Rectangle pattern
Histogram of RGB value

Figure 6

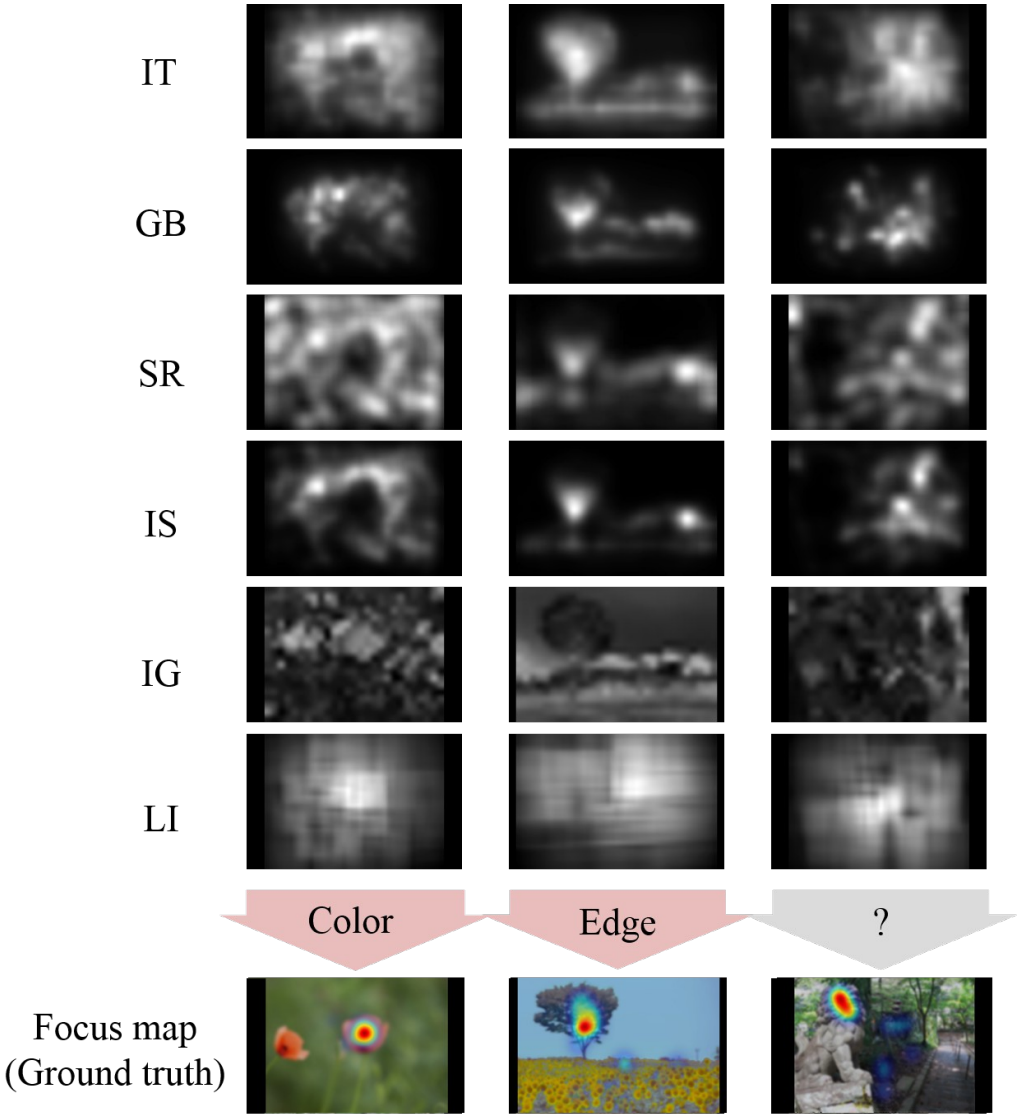


Figure 7

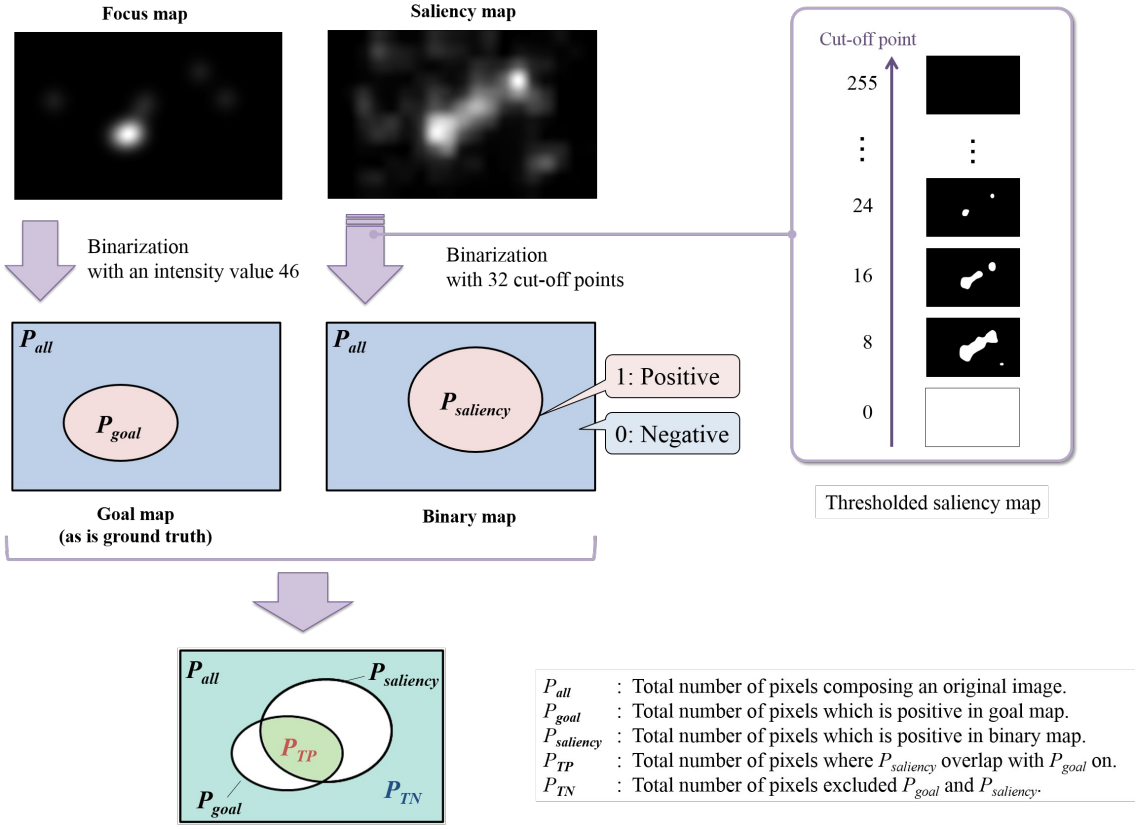

Focus map
Saliency map
Cut-off point
255
24
16
8
0
Binarization
with an intensity value 46
Binarization
with 32 cut-off points
P_{all}
P_{goal}
P_{all}
$P_{saliency}$
1: Positive
0: Negative
Thresholded saliency map
Goal map
(as is ground truth)
Binary map
P_{all}
$P_{saliency}$
P_{TP}
P_{goal}
P_{TN}
P_{all} : Total number of pixels composing an original image.
P_{goal} : Total number of pixels which is positive in goal map.
$P_{saliency}$: Total number of pixels which is positive in binary map.
P_{TP} : Total number of pixels where $P_{saliency}$ overlap with P_{goal} on.
P_{TN} : Total number of pixels excluded P_{goal} and $P_{saliency}$.

Figure 8

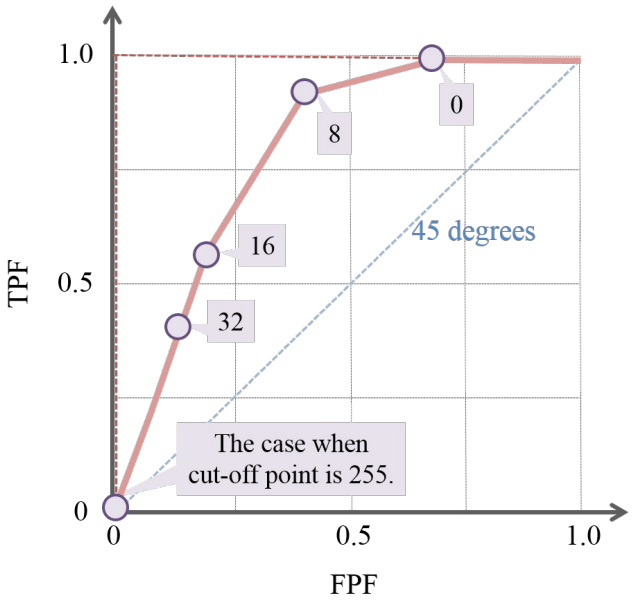
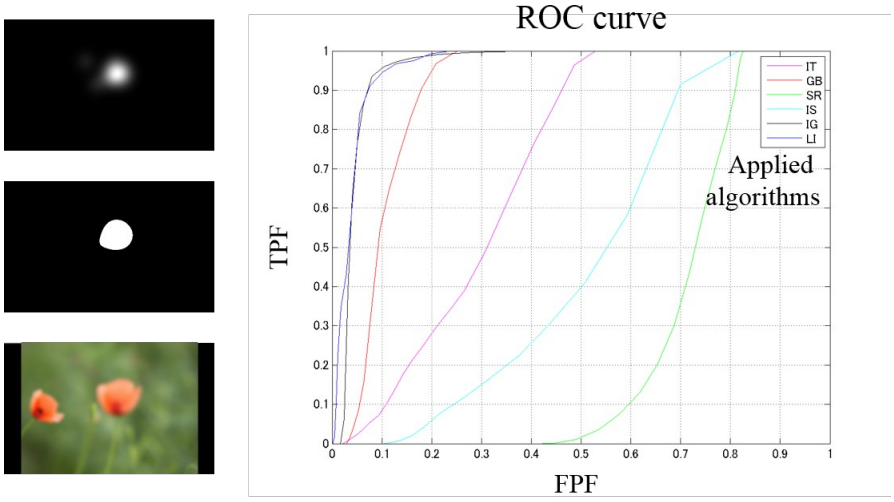
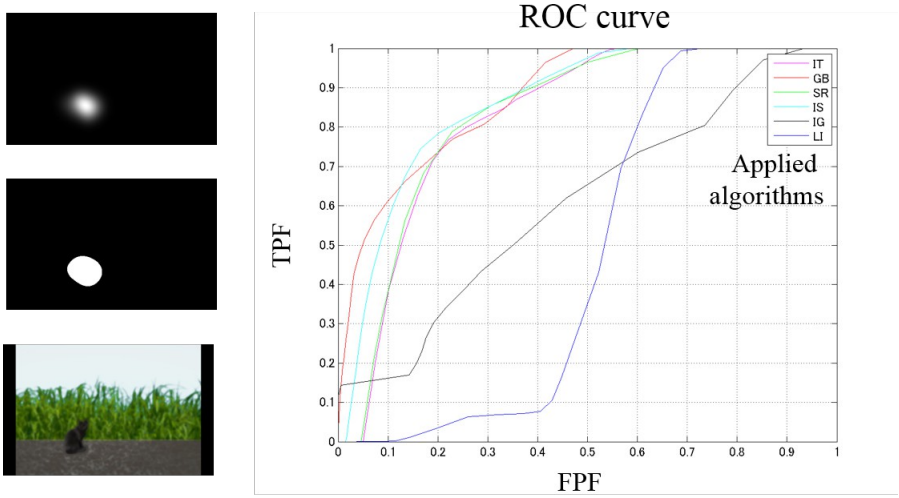

1.0
0.5
0
TPF
0
0.5
1.0
FPF
0
8
16
32
45 degrees
The case when
cut-off point is 255.

Figure 9



(a) Image #79



(b) Image #12

Figure 10

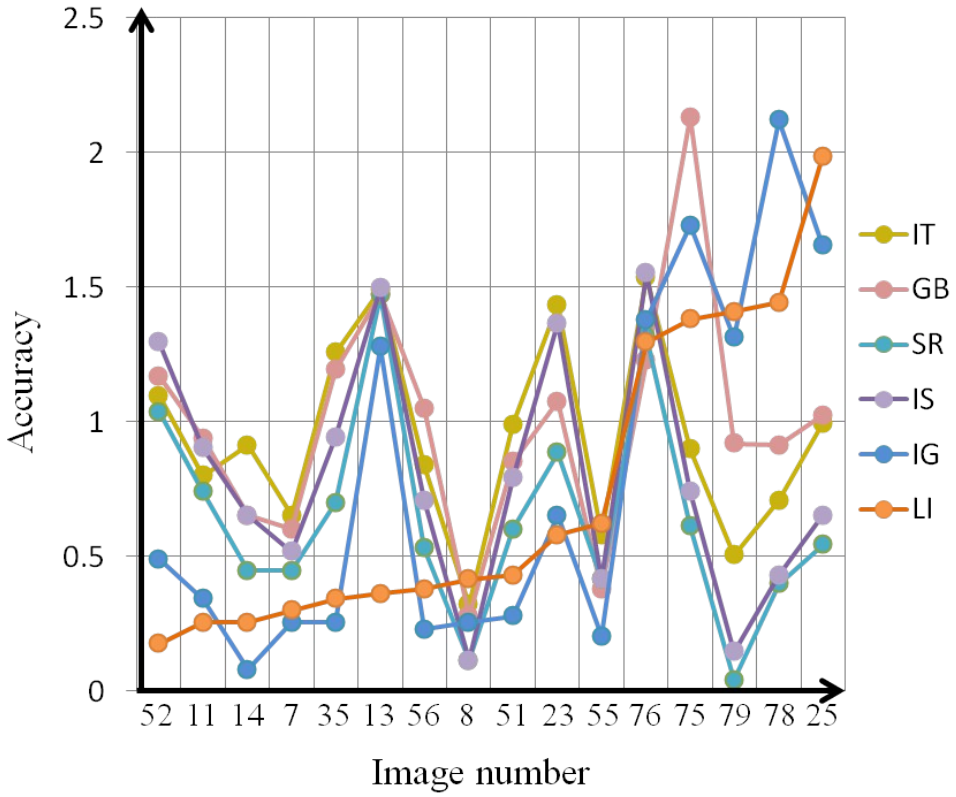

2.5
2
1.5
1
0.5
0
Accuracy
52 11 14 7 35 13 56 8 51 23 55 76 75 79 78 25
Image number
IT
GB
SR
IS
IG
LI

Figure 11

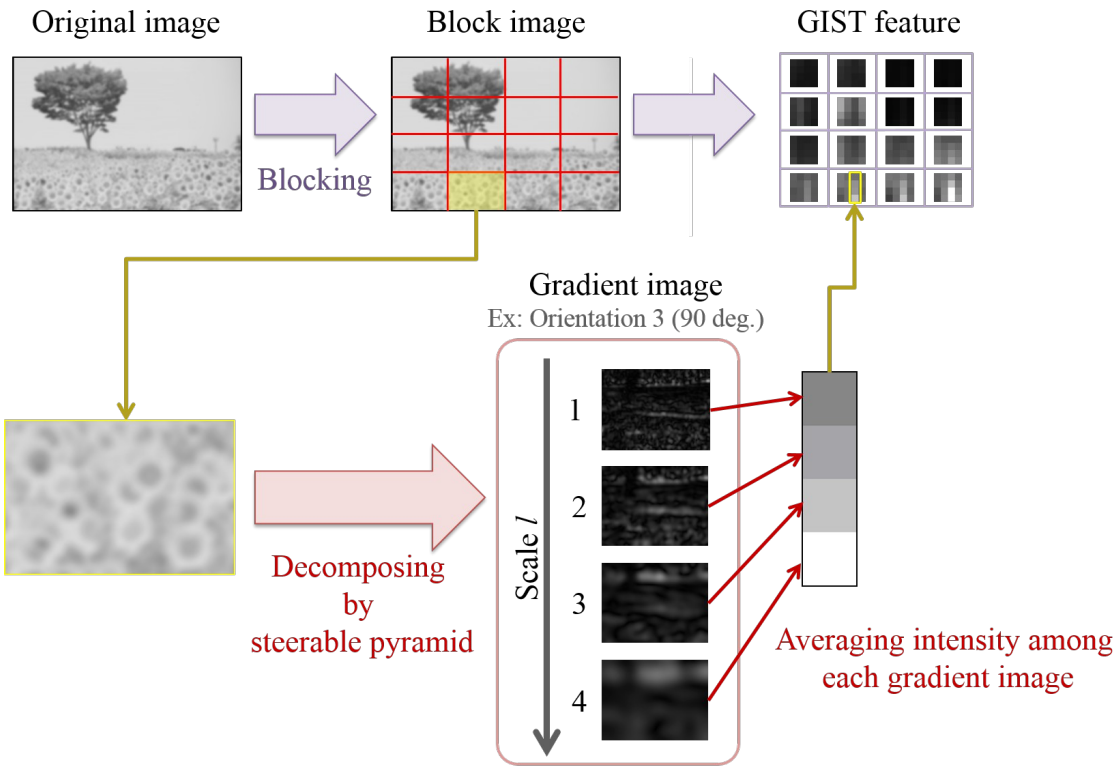

Original image
Block image
GIST feature
Blocking
Gradient image
Ex: Orientation 3 (90 deg.)
Scale l
1
2
3
4
Decomposing
by
steerable pyramid
Averaging intensity among
each gradient image

Figure 12

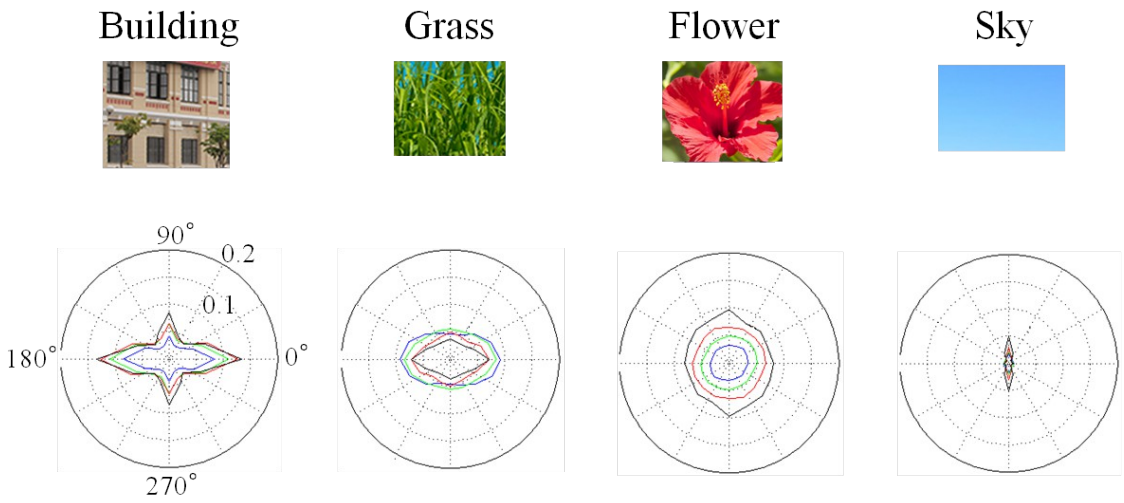


Figure 13

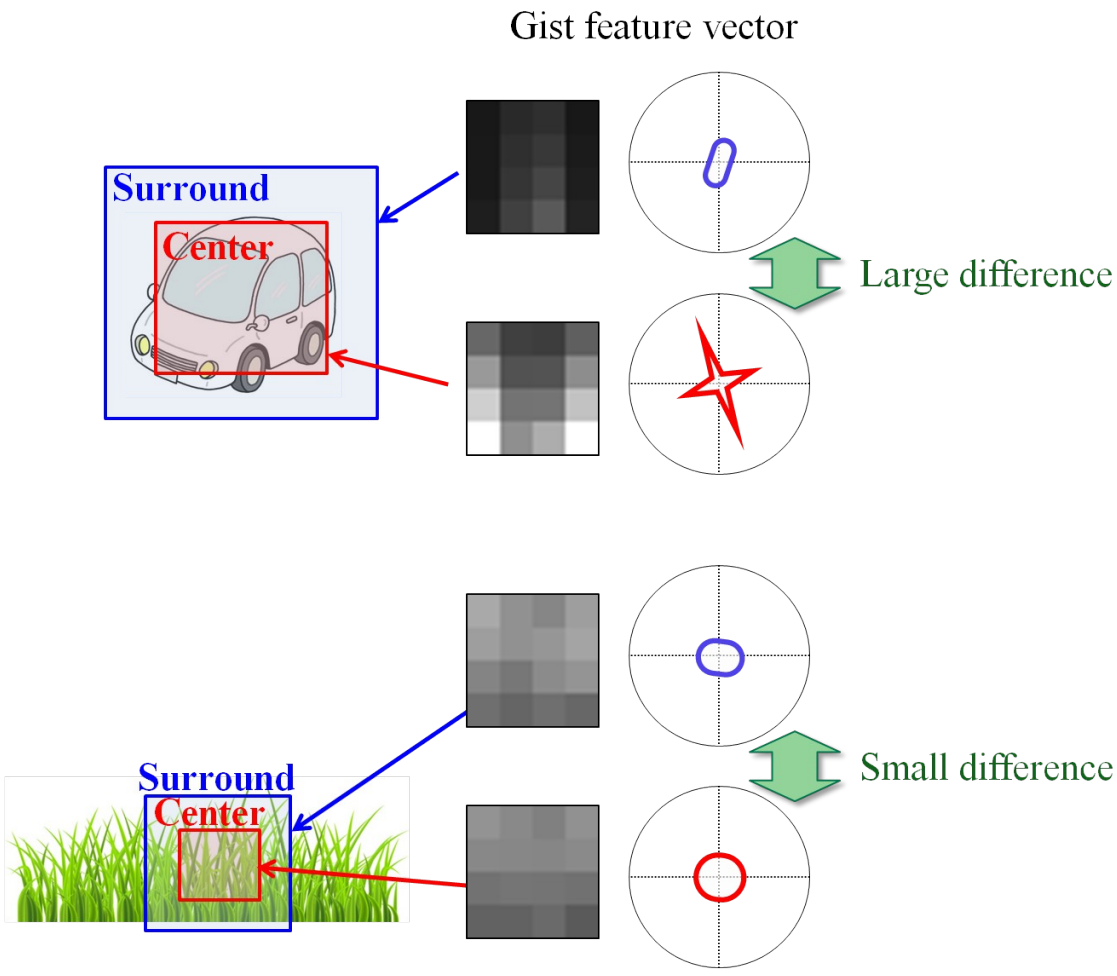
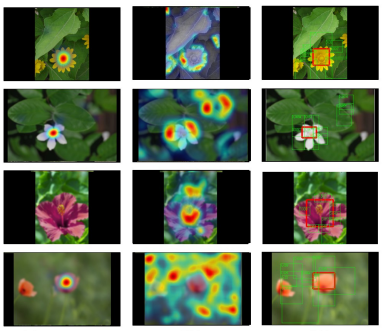
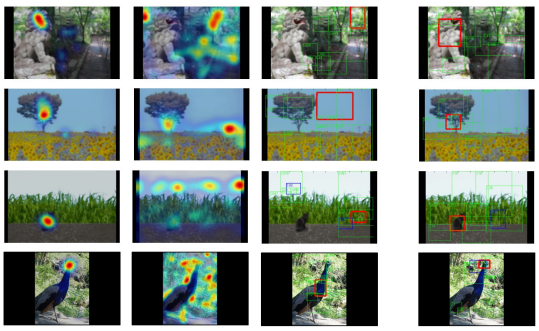

Gist feature vector
Surround
Center
Large difference
Surround
Center
Small difference

Figure 14



(a) Objectness using color



(b) Objectness using GIST

Figure 15

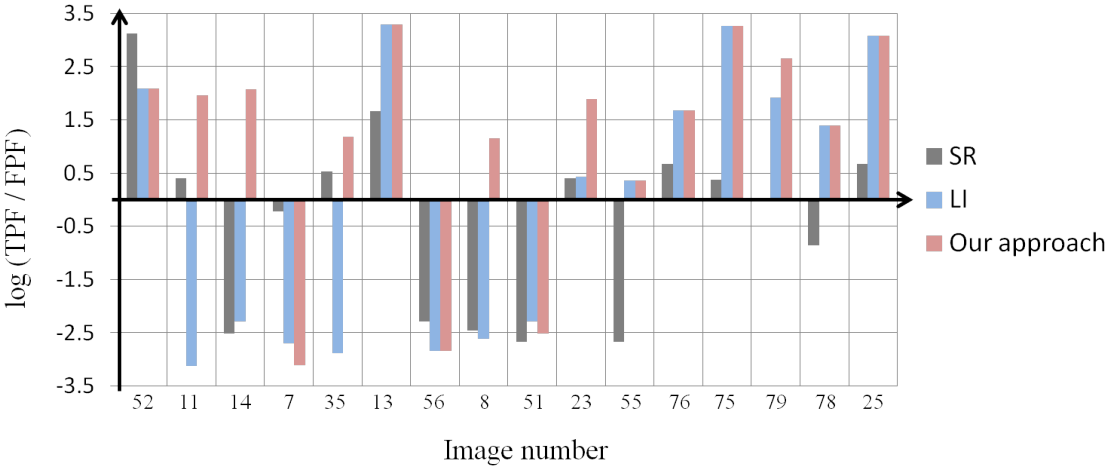

3.5
2.5
1.5
0.5
-0.5
-1.5
-2.5
-3.5
log (TPF / FPF)
52 11 14 7 35 13 56 8 51 23 55 76 75 79 78 25
Image number
SR
LI
Our approach

Figure 16

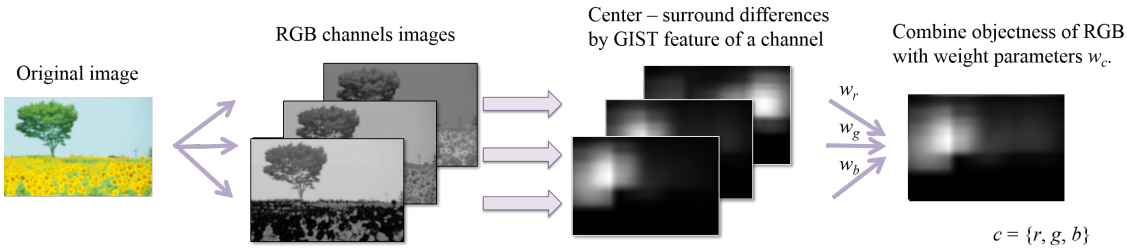

Original image
RGB channels images
Center – surround differences
by GIST feature of a channel
Combine objectness of RGB
with weight parameters w_c.
w_r
w_g
w_b
$c = \{r, g, b\}$

Figure 17

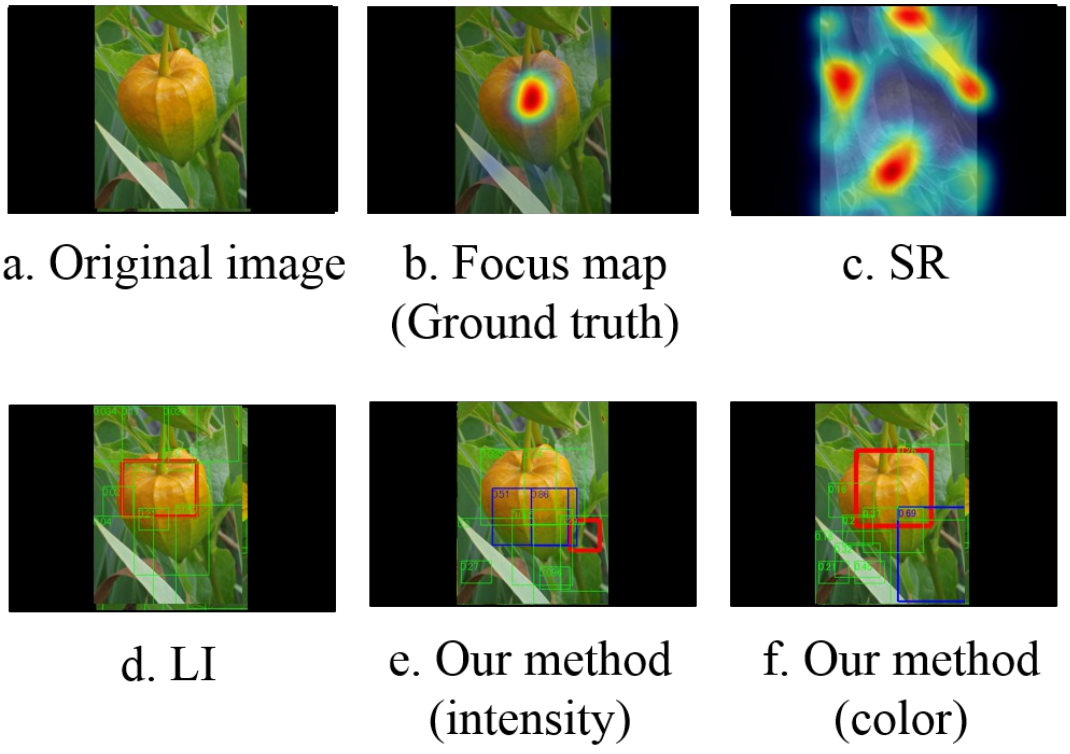

a. Original image
b. Focus map (Ground truth)
c. SR
d. LI
e. Our method (intensity)
f. Our method (color)

Figure 18