Generation of an Importance Map for Visualized Images

Akira Egawa[1] and Susumu Shirayama[2]

[1] Department of Systems Innovation, Graduate School of Engineering,
The University of Tokyo
[2] RACE, The University of Tokyo
{egawa,sirayama}@race.u-tokyo.ac.jp

Abstract. Visualized images have always been a preferred method of communication of information contained in complex data sets. However, information contained in the image is not always efficiently communicated to others due to personal differences in the way subjects interpret image content. One of the approaches to solving this issue is to determine high-saliency or eye-catching regions/objects of the image and to share information about the regions of interest (ROI) in the image among researchers. In the present paper, we propose a new method by which an importance map for a visualized image can be constructed. The image is first divided into segments based on a saliency map model, and eye movement data is then acquired and mapped into the segments. The importance score can be calculated by the PageRank algorithm for the network generated by regarding the segments as nodes, and thus an importance map of the image can be constructed. The usefulness of the proposed method is investigated through several experiments.

1 Introduction

As complexity of raw data in computational simulations increases, visualization of the data has played an important role in its analysis, and such visualized images have always been a preferred method of communication. On the other hand, interpretation of these visualized images is becoming more difficult, due to the enormous amount of information contained in them. Therefore, clear interpretation of the images is required in order to share useful information. However, information contained in the image is not always efficiently communicated to others due to personal differences in the way subjects interpret image content. Moreover, the need for collaborative evaluations, second opinions, and third party evaluations makes the problem of conveying information to others correctly even more difficult to solve.

Both formative and summative evaluations of the image are often required in order to solve those issues. In these cases, determining how to estimate high saliency or eye-catching regions/objects of the image is one of the most important problems.

A number of methods for conveying information to others and for determining high-saliency regions have been proposed. One such method is to manifest regions of interest (ROI) [1] and to show the ROI according to level of detail (LOD) [2]. In most cases, particular objects or regions are extracted from the image based on the meaning of the content, after which the LOD is designed. Although this is a powerful method, there are two major problems with this approach.

G. Bebis et al. (Eds.): ISVC 2009, Part I, LNCS 5875, pp. 135–146, 2009.

One is the difficulty involved in separating particular objects or regions from the image. In order to separate these objects/regions from the image, the problem of figure-ground separation has been solved [3], and methods to detect the critical points and particular lines/regions have been developed [4][5]. However, it is not always easy to detect the boundaries in an image, or to analyze the figure-ground relationship, and figure-ground separation remains an unsolved problem. Despite this, methods of detecting the critical points and particular lines have been established for images produced by visualizing fields that have a mathematical structure.

The second problem is that if there is no definitive structure, then the ROI and LOD depend strongly on personal interpretation, and the definition of an ROI or LOD becomes subjective. Therefore, it is difficult to assign a certain quantity of interest to a particular region. This is related to the first. In general, it has been pointed out that images have inherently recognizable areas. Considering the visual characteristics of the human eye, most viewers naturally focus their eyes on regions that contain recognizable areas. Therefore, such regions may be regarded as ROI candidates.

In order to detect such regions, we focused on the following two approaches. The first is a model-based approach that consists of computational models that imitate visual attention based on the human cognitive system. One of the most popular computational models was introduced by Itti et al. [6]. They modeled high-saliency regions that attract the human eye in a task-independent manner. Their model is an extension of the visual attention model proposed by Koch et al. [7]. The model of Itti et al. is referred to herein as the saliency map model. Using the saliency map model, eye-catching areas can be extracted from a source image [8][9]. Furthermore, the validity of this model has been confirmed by eye-tracking experiments [10][11] and in applications [12][13][14].

The second is an eye movement analysis. Although most eye movement data is meaningless, as pointed out by Duchowski et al. [15], the overt response of the subject in visual attention processes, tacit interest, and personal skill or knowledge can be estimated based on the fixation and the saccade [16][17][18].

These two approaches for detecting recognizable areas or ROI have been applied to image processing. Maeder [19] reviewed methods for image quality assessment and argued the usefulness of a perceptual modeling framework known as an image importance map. Although he considered the human cognitive system, his goal was to assess the image quality effectively. Lee et al. [20] proposed methods for detecting manipulated images. Their methods involve segmentation, classification, and common-sense reasoning. An importance map constructed by mean-shift image segmentation [21] and the visual attention algorithm of Itti et al. [6] is used in the processes of segmentation and classification. In these studies, an importance map of an image played an important role in processing the image.

In the present paper, a method for conveying quantitative information about both subjective/ personal ROI favored by ordinary or skilled viewers and also an objective view of ROI in a visualized image is considered. For this method, we use the saliency map model and eye movement analysis to extract an ROI from an image and then assign a quantitative index to the ROI. We refer to a distribution of such ROI as an Image Importance Map (IIM). A method for constructing an IIM from the visualized image is described below.

2 Method

First, a source image and eye movement data for a visualized image are provided as the input data. Second, the source image is segmented by the following three steps:

Step 1. A saliency map for the image is created according to the model proposed by Itti et al. [6].
Step 2. Regions of high salience in the map are clustered.
Step 3. The image is segmented based on the clusters produced in Step 2.

Third, two types of Image Importance Map (IIM) are constructed. The IIM is constructed by assigning a quantitative index to the segments. The quantitative index for the first IIM is calculated using the attention shift model of Itti et al. [6]. The index for the second IIM is calculated by eye movement data using the PageRank algorithm [22]. Finally, two IIMs are integrated into one map using the Biased PageRank algorithm [23]. The details of this process are presented in the following sections.

2.1 Segmentation of the Source Image

In a task-dependent manner in particular, it may be considered that conveying information related to subjective/objective ROI in the image to others is based on important objects. In order to identify the objects in an image, an image segmentation method for objects that have discernable or presumed contours has been used [24]. Therefore, object-based segmentation is used in the importance map of Lee et al. [20].

However, considering the visual characteristics of the human eye, it is assumed that both ordinary and skilled viewers naturally train their eyes on regions that contain recognizable areas, even while working to complete a task. Such regions or areas do not always have clearly identified boundaries.

Accordingly, we propose a new image segmentation method based on the saliency map model.

Creating saliency map. First, the image is decomposed into three feature channels: intensity, color (red/green and blue/yellow), and orientation (0°, 45°, 90°, 135°). The orientation channel is then calculated using the Gabor filter and nine spatial scales are created from these three channels using Gaussian pyramids.

Second, differences between the fine and coarse scales in each channel are calculated based on a center-surround operation that imitates the human visual system.

Third, for each channel, three conspicuity maps that highlight the parts that differ noticeably from their surroundings are created using across-scale combinations. Finally, a saliency map is created as a linear combination of the three conspicuity maps. The saliency map is composed of $lmax \times mmax$ cells. Let (l,m) be the coordinates of a cell in the map. The saliency at (l,m) is denoted by $s(l,m)$. In our implementation, the scale of the saliency map is 1/16 the size of the source image (one cell on the saliency map corresponds to a 16×16 pixel region in the source image). Figure 1 shows the source image (left figure) and the saliency map (right figure). White and black areas indicate high- and low-saliency regions, respectively.

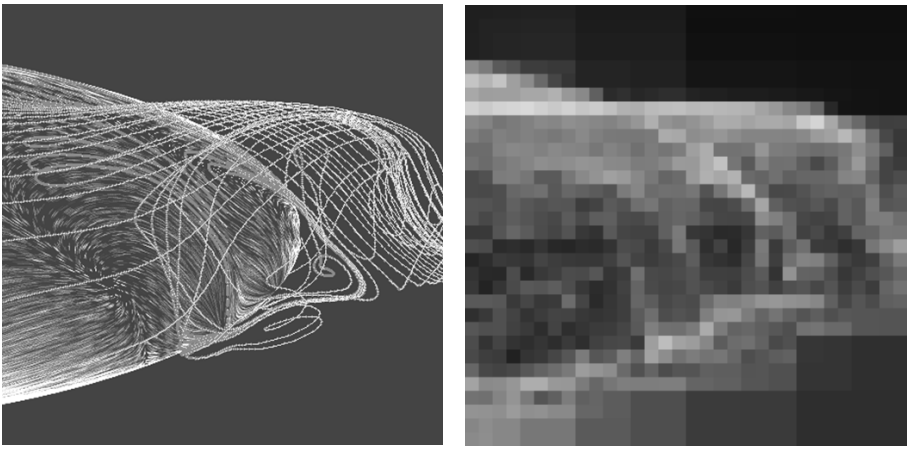

Fig. 1. Source image (left) and its saliency map (right)

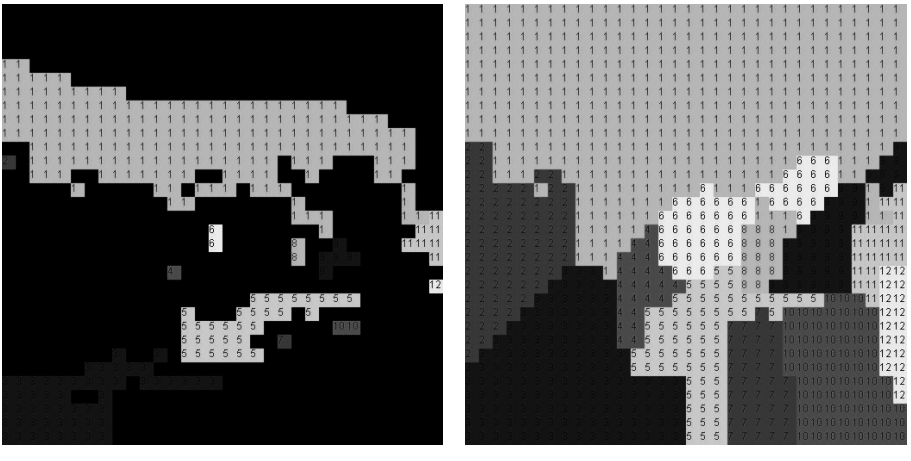

Fig. 2. Saliency Cluster Map. Left and right figures show the core-clusters and all of the other clusters, respectively.

Clustering of the saliency map. Regions of high salience in the map are clustered. The proposed procedure consists of the two steps. The pseudocode for the first step is shown in Table 1, where *imax* and *jmax* denote the width and height, respectively, of the source image. A cluster is identified by index K. The index of the cluster at (l,m) is stored in $c(l,m)$. Let c_K be a set of cells in cluster K. The centroid of cluster K is (x_K, y_K), which is computed as $(x_K, y_K) = \frac{1}{N_K}(\sum_{(l,m)\in c_K} l, \sum_{(l,m)\in c_K} m)$, where N_K is the number of cells in cluster K. The distance between a cell $c(l,m)$ and cluster K is calculated as $\sqrt{(l-x_K)^2+(m-y_K)^2}$. NearestCluster$(l,m)$ is the function that returns the index of the cluster that has the shortest distance between (l,m) and (x_k, y_k) in eight

Table 1. Pseudocode for the first step

```
set 0 to c(l,m)
K ← 0
for l = 0 to [(imax − 1)/16] {
  for m = 0 to [(jmax − 1)/16] {
    if s(l,m) ≥ δ and c(l,m) = 0 {
      if at least one of the eight neighbors of (l,m)
      belongs to any cluster {
        K ← K + 1
        c(l,m) = K
        (x_K, y_K) = (l,m)
      }
      else {
        k = NearestCluster(l,m)
        c(l,m) = k
        update (x_k, y_k)
      }
    }
    if s(l,m) < δ { c(l,m) = 0 }
```

neighbors of (l,m). The threshold value δ is given by the user. We refer to the cluster obtained in the first step as the core-cluster.

In the second step, the cells for which $s(l,m) < \delta$ are assigned to one of the clusters which are closed to the cells. All of the cells on the saliency map belong to the clusters after this clustering.

In this way, all of the cells are eventually assigned to the appropriate cluster. These clusters are referred to collectively as the Saliency Cluster Map (SCM). The SCM for Figure 1 is shown in Figure 2. The core-clusters are shown in the left-hand image, and the right-hand image shows all of the other clusters.

Segmentation using the saliency cluster map. The SCM is applied to the image segmentation in a straightforward manner. A segment consists of the pixel regions that correspond to cells belonging to the same cluster. Note that the segment is composed of two types of pixels. One type exists in the core-cluster. We refer to the region of a segment composed of these pixels as the *core-segment*. The other type exists outside the core-cluster. Let the index of a segment be k. The core-segment is then denoted as R_k, and the region outside the core-segment is denoted as R'_k. An example of image segmentation of a source image is shown in Figure 3.

2.2 First Image Importance Map

Our approach is based on the Focus Of Attention (FOA) and FOA shift proposed by Itti et al. algorithm [6].

We define a set R as a collection of all segments, where the total number of elements of R is M. Let the saliency of segment be $s_k(k = 1,...,M)$. The saliency s_k is set as the

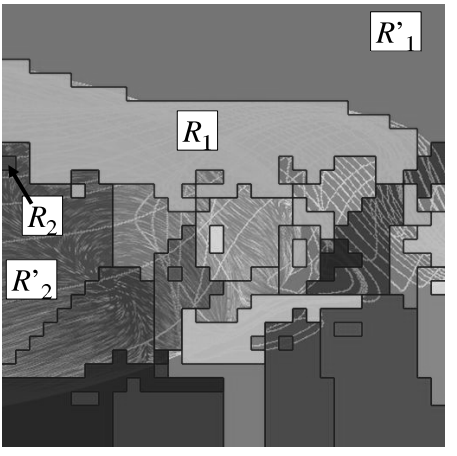

Fig. 3. Example of image segmentation

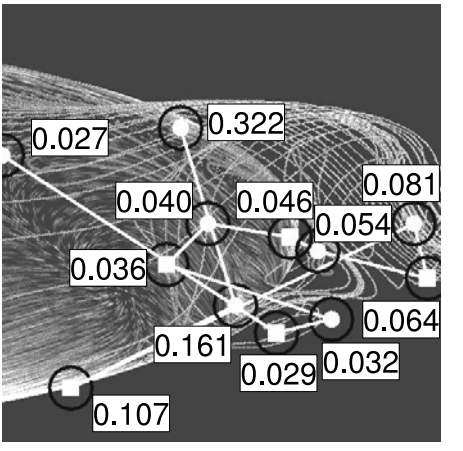

Fig. 4. First Image Importance Map

highest saliency in the cells corresponding to the core-segment R_k. Let (x_k, y_k) be the centroid of segment k, which is defined as the centroid of the cluster corresponding to segment k. In addition, the preferential saliency is denoted as s'_k.

A set F that consists of a sorted FOA is obtained as follows:

Step 1. Let F be an empty set $\emptyset$. Also, let $s'_k = s_k (k = 1, ..., M)$
Step 2. Find the segment that has the highest preferential saliency in $R \setminus F$, and set the index of the segment to i
Step 3. Add i to set F
Step 4. Compute s'_k as follows:
if $\sqrt{(x_i - x_k)^2 + (y_i - y_k)^2} \leq \beta$ {
$s'_k = s_k + \alpha$
}
else {
$s'_k = s_k$
}
Step 5. Repeat Step 2 through 4

where α and β are the parameters of the proximate preference. α and β indicates the strength and range of the proximate preference respectively. We use $\alpha = 50$ and $\beta = 100$ (pixels) in this paper.

We use the set F for the quantitative index assigned to the segment. First, R_k and R'_k are redefined. The value of index k indicates the k-th element of F, i.e., the order of the FOA. Next, the importance of the segment is defined. Let i^1_k be the importance of segment k.

$$i^1_k = \frac{1}{k \cdot \sum_{k=1}^{M} \frac{1}{k}}. \tag{1}$$

The distribution of the segments that have this importance is referred to as the first Image Importance Map (first IIM). An example of a first IIM is shown in Figure 4. In

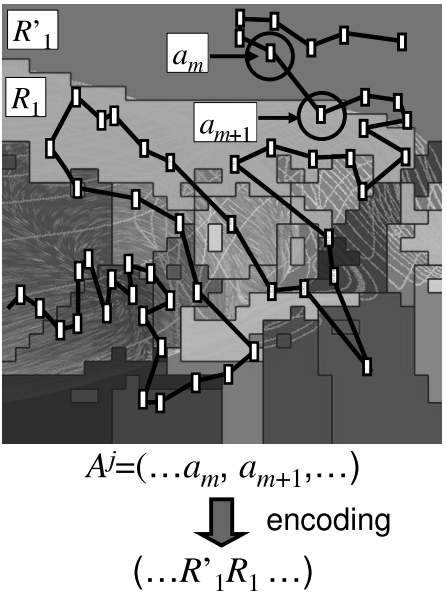

Fig. 5. Scheme of encoding the eye movement dataset

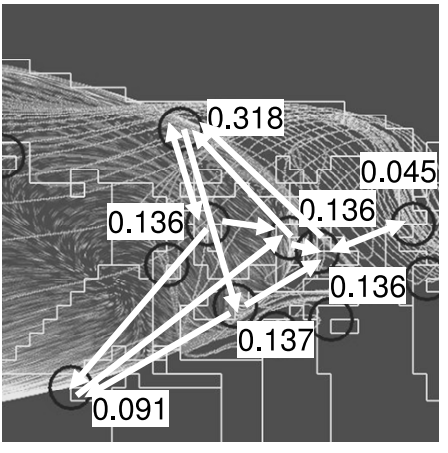

Fig. 6. Second Image Importance Map

this figure, the blue circles are the centroids of the segments, and the numerical values indicate the importance of the segments. In addition, the green filled circle denote R_1, and the red lines denote the FOA shift.

2.3 Second Image Importance Map

In the first IIM, top-down (task-dependent manner) factors are not considered. Previous researches have improved this model by adding top-down factors such as eye movement data [25] and a stochastic factor [26]. As in those studies, in the present study, the top-down factors are considered in the construction of the IIM. In our method, we use the same image segmentation of the first IIM and an IIM with a top-down factor is constructed as follows.

First, we obtain eye movement data from the subjects using a head mounted eye-tracking system. Let $a_n = (x_n, y_n, t_n)$ be the n-th eye movement data. The data a_n is composed of (x_n, y_n) (the coordinates in the image) and t_n (the recorded time). An eye movement dataset for person j is denoted by $A^j = (a_1, a_2, ..., a_N)$, and N is the total number of eye movement data.

Second, the eye movement dataset A^j is mapped into the segmented source image and encoded using the identifier of the segment that contains the eye position. The scheme for encoding the eye movement dataset is shown in Figure 5.

Third, a network for the eye movements is generated by regarding the segments as nodes. We denote a node as v_i. The number of nodes is the same as the number of segments (M). The node corresponds to the centroid of the segment that contains the eye

position. For simplicity, in the present paper, we do not discriminate between the core-segment and the region outside the core-segment. A link (an arc) is generated by the eye movement. The encoded eye movement dataset A^j is transformed into a network.

In this way, one network is generated from one eye movement dataset. This network is represented by a weighted adjacency matrix **A** in which the weight is given by the number of links.

The score of importance is obtained by the PageRank algorithm [22]. First, the matrix **A** is converted to a transition probability matrix **T**. For example, in the case of the eye movement dataset $R_2R_4R_1R_1R_1R_2R_4R_3$, the matrix **T** is obtained by

$$\mathbf{T} = \begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \\ v_4 \end{array} \begin{array}{c} \begin{array}{cccc} v_1 & v_2 & v_3 & v_4 \end{array} \\ \begin{pmatrix} 2/3 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix} \end{array}. \tag{2}$$

Second, the PageRank of each node is calculated by the following iterative process:

$$\mathbf{p}^{\nu+1} = \mathbf{T}^t \mathbf{p}^{\nu}, \tag{3}$$

where $\mathbf{p} = (i_1^2, ..., i_i^2, ..., i_M^2)$, i_i^2 represents the PageRank of the v_i, and ν is the iterative number.

The importance of the segment k is i_k^2. The distribution of the segments that have this importance is designated as the second Image Importance Map (second IIM). An example of a second IIM is shown in Figure 6. In this figure, the blue circles are the centroids of the segments, and the numerical values indicate the importance of the segments. In addition, the orange arrows indicate the links of the network.

2.4 Integration of Image Importance Maps

A second IIM is obtained for each eye movement dataset. In order to integrate the second IIMs, the weighted adjacency matrix **A** is composed of the eye movement datasets that we attempt to integrate.

For the case in which the first IIM and the second IIM are integrated, we use the biased PageRank algorithm [23]. First, a vector **i** that denotes the importance of the segment in the first IIM is calculated. Second, the weighted adjacency matrix **A** is obtained and converted to the transition probability matrix **T**.

The first IIM and the second IIM are integrated by the following iterative process:

$$\mathbf{p}^{\nu+1} = \omega \mathbf{T}^t \mathbf{p}^{\nu} + (1-\omega)\mathbf{i}, \tag{4}$$

where $\mathbf{p} = (i_1^s, ..., i_i^s, ..., i_M^s)$, i_i^s represents the integrated importance of segment i, and ω ($0 \leq \omega \leq 1$) is the user parameter. In the present paper, we use $\omega = 0.5$.

3 Construction of the Image Importance Maps

First, an image produced by the visualization of instantaneous streamlines in a flow past a spheroid was used for the experiment. The source image and the first IIM are shown

in the left part of Figure 1 and in Figure 4, respectively. Three subjects participated in the experiment. One (subject A) was an expert in fluid dynamics, whereas the others (subjects B and C) were novices. The second Image Importance Maps (IIMs) and the integrated IIMs are shown in Figure 7 and Figure 8, respectively.

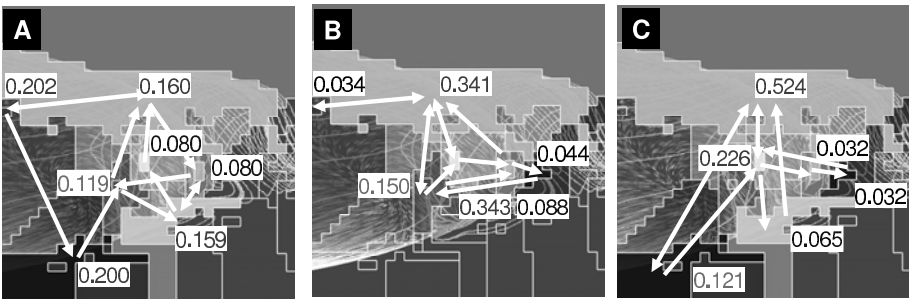

Fig. 7. Three subjests' second IIMs of the first source image

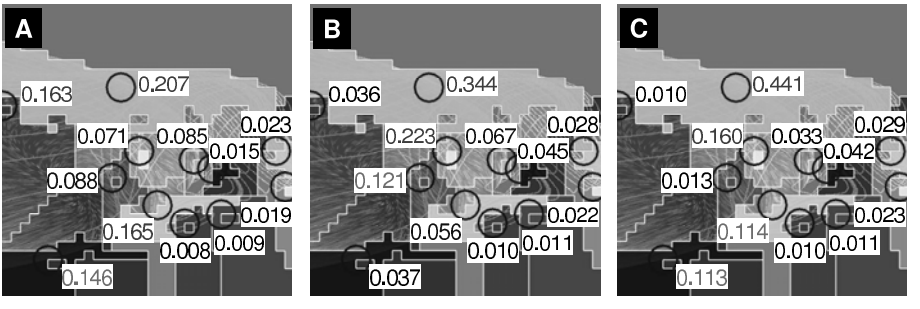

Fig. 8. Three subjests' integrated IIMs of the first source image

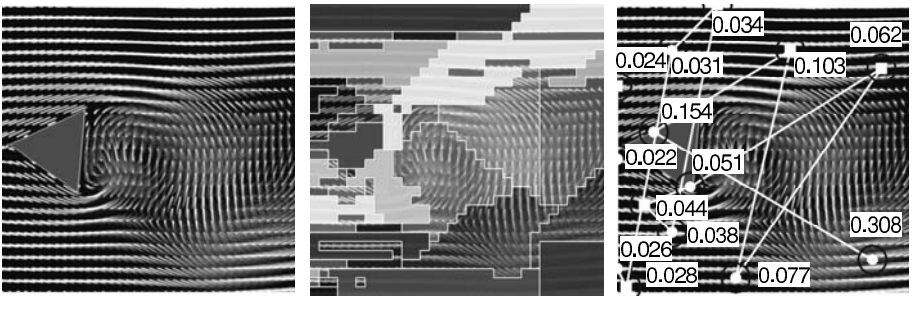

Fig. 9. A source image for the second experiment (left), the corresponding segments of the image (center) and the first IIM (right)

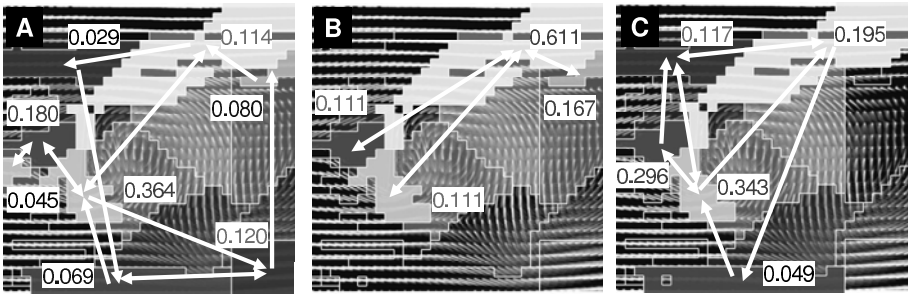

Fig. 10. Three subjests' second IIMs of the second source image

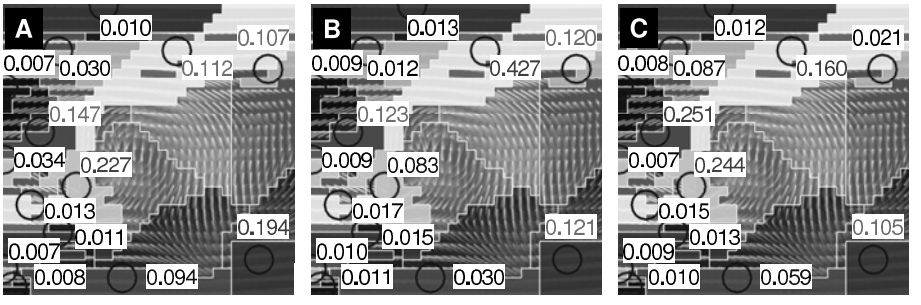

Fig. 11. Three subjests' integrated IIMs of the second source image

Second, a visualized image produced by a tuft visualization for a flow around a trianguler cylinder is used. A source image, the corresponding segments of the image and the first IIM are demonstrated in Figure 9. Figure 10 and Figure 11 shows the second IIMs and integrated IIMs, respectively.

As shown in Figures 7 and 10, all the regions where the expert (subject A) gazed are not always eye-catching for subjects B and C. This means that some important information is lost despite the fact that raw data is visualized and transformed into a recognizable form.

The region which has the highest saliency in the right lower part of the first IIM shown in Figure 9 is remarkable. After the experiment, it was confirmed through interviews with the expert that this region is important to prediction of the far wake region. Actually, the expert (subject A) had focused on the region, and the subject A's score is striking in the integrated IIMs presented in Figure 11. It can be said that the region is inherently recognizable and includes the expert's preference. On the other hand, subjects B and C did not gaze at the region as shown in Figure 10. This also means that some important information will be lost.

From the standpoint of visual communication, it may be considered that the experts' ways of viewing are deliberate and inappropriate since their viewing behaviors are not always correct and are sometimes unnatural. In order to avoid premature judgement, it is important to know the average and personal preferences of the viewers. The

integrated IIMs shown in Figures 8 and 11 are considered to give an immediate indication of the average and personal preferences of the viewers since they are created by combining the importance values of the first and second IIMs. The region that has a higher importance value on the integrated IIM is inherently recognizable; it is the focus of the viewer's gaze, whereas regions that have lower importance are not eye-catching, and so are not gazed at by the viewer. The differences in the evaluations of the subjects can be expressed by the quantitative values. Thus, the integrated IIM provides a more accurate indication of the important regions of the image.

4 Conclusion

We have presented a new method by which an importance map of a visualized image can be constructed. We used a saliency map model for image segmentation and quantitative importance values for the importance maps were obtained based on the attention shift model of the saliency map model and eye movement data. The usefulness of the proposed method was investigated through eye-tracking experiments.

Our method enables researchers to share information about ROIs with quantitative values of importance and indications of differences among individuals. Therefore, this method will help to enhance the interpretability and recognizability of visualized images subsequent to their generation, and can be used in such applications as visual communication for education purposes.

Several extensions can be developed to improve the proposed method. For a more detailed treatment of the eye movement dataset, the regions outside the core-segments can be subdivided into a number of segments in an appropriate manner. In addition, although the saliency map model of Itti et al. is effective for use with the proposed method, other saliency map models that simulate human visual attention may be used as alternatives. Finally, object-based image segmentation may be combined with the proposed method in order to consider eye movement in detail and in a task-dependent manner.

Acknowledgements

One of the authors was supported through the Global COE Program, gGlobal Center of Excellence for Mechanical Systems Innovation,h by the Ministry of Education, Culture, Sports, Science and Technology.

References

1. Kim, H., Min, B., Lee, T., Lee, S., Lee, C., Park, J.: Three-Dimensional Digital Subtraction Angiography. IEEE Transactions on Medical Imaging MI-1, 152–158 (1982)
2. Rauschenbach, U.: Progressive Image Transmission using Levels of Detail and Regions of Interest. In: Proc. of IASTED CGIM 1998, pp. 38–41 (1998)
3. Bhandarkar, S., Zeng, X.: Figure-Ground Separation: A Case Study in Energy Minimization via Evolutionary Computing. In: Pelillo, M., Hancock, E.R. (eds.) EMMCVPR 1997. LNCS, vol. 1223, pp. 375–390. Springer, Heidelberg (1997)

4. Schultz, T., Theisel, H., Seidel, H.: Topological Visualization of Brain Diffusion MRI Data. IEEE Transactions on Visualization and Computer Graphics 13, 1496–1503 (2007)
5. Bürger, R., Hauser, H.: Visualization of Multi-variate Scientific Data. In: Proc. of EuroGraphics 2007 State of the Art Reports, pp. 117–134 (2007)
6. Itti, L., Koch, C., Niebur, E.: A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 20, 1254–1259 (1998)
7. Koch, C., Ullman, S.: Shifts in selective visual attention: towards the underlying neural circuitry. Human Neurobiology 4, 219–227 (1985)
8. Itti, L., Kock, C.: Target Detection using Saliency-Based Attention. In: Proc. of RTO/SCI-12 Workshop on Search and Target Acquisition (Unclassified), pp. 3.1–3.10 (1999)
9. Walther, D., Edgington, D., Koch, C.: Detection and Tracking of Objects in Underwater Video. In: Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 544–549 (2004)
10. Itti, L., Koch, C.: Feature combination strategies for saliency-based visual attention systems. Journal of Electronic Imaging 10, 161–169 (2001)
11. Parkhurst, D., Law, L., Niebur, E.: Modeling the role of salience in the allocation of overt visual attention. Vision Research 42, 107–123 (2002)
12. Kim, Y., Varshney, A.: Saliency-guided Enhancement for Volume Visualization. IEEE Transaction on Visualization and Computer Graphics 12, 925–932 (2006)
13. Davies, C., Tompkinson, W., Donnelly, N., Gordon, L., Cave, K.: Visual saliency as an aid to updating digital maps. Computers in Human Behavior 22, 672–684 (2006)
14. Ma, Y., Lu, L., Zhang, H.J., Li, M.: A User Attention Model for Video Summarization. In: Proc. of the 10th ACM international conference on Multimedia, pp. 533–542 (2002)
15. Duchowski, A.: Eye Tracking Methodology, 2nd edn. Springer, Heidelberg (2007)
16. Yarbus, A.: Eye Movements and Vision. Plenum Press, New York (1967)
17. Russo, J.E., Rosen, L.D.: An eye fixation analysis of multialternative choice. Memory and Cognition 3, 167–276 (1975)
18. Martinez-Conde, S., Macknik, S.L.: Windows on the mind. Scientific American Magazine 297, 56–63 (2007)
19. Maeder, A.: The image importance approach to human vision based image quality characterization. Pattern Recognition Letters 26, 347–354 (2005)
20. Lee, S., Shamma, D., Gooch, B.: Detecting false captioning using common-sense reasoning. Digital Investigation 3, 65–70 (2006)
21. Meer, P., Georgescu, B.: Edge Detection with Embedded Confidence. IEEE Transactions on Pattern Analysis and Machine Intelligence 23, 1351–1365 (2001)
22. Brin, S., Page, L.: The anatomy of a large-scale hypertextual web search engine. Computer Networks and ISDN Systems 30, 107–117 (1998)
23. Kamvar, S., Haveliwala, T., Manning, C., Golub, G.: Exploiting the Block Structure of the Web for Computing Pagerank. Stanford University Technical Report 2003-17 (2003)
24. Ko, B., Byun, H.: FRIP: A Region-Based Image Retrieval Tool Using Automatic Image Segmentation and Stepwise Boolean AND Matching. IEEE Transactions on Multimedia 7, 105–113 (2005)
25. Igarashi, H., Suzuki, S., Sugita, T., Kurisu, M., Kakikura, M.: Extraction of Visual Attention with Gaze Duration and Saliency Map. In: Proc. of IEEE International Conference on Control Applications, 562–567 (2006)
26. Pang, D., Kimura, A., Takeuchi, T., Yamato, J., Kashino, K.: A Stochastic Model of Selective Visual Attention with a Dynamic Baysian Network. In: Proc. of ICME 2008, pp. 1073–1076 (2008)