

# Simulating Eye Movement in Reading Using Short-term Memory

Satoru Morita

Faculty of Engineering, Yamaguchi University  
2557 Tokiwadai Ube 755-8611 Japan

## Abstract

We propose the computation model of the eye movement based on the short-term memory. We also applied this model to simulating the human eye movement in reading. We use the foveated vision that the resolution is high in the center of the retina and is low in the periphery of it to simulate human eye movement. It is reported that the viewpoint moves to the first word of next line from the end of line in a moment. But we cannot identify the first word of the next line in the end of line, because the resolution is too low in the periphery of the retina. It is reported that only four characters can be identified in the case of using the foveated vision. We introduce the short-term memory to identify the first word of the next line in the end of line. The viewpoint can be moved to the first character of next line using the image which is integrated the foveated images saved from the first word identified in the line. We call it the short-term memory image in this paper. These mechanism is also used the eye movement that we see a word more than two times. If the edge features derived from the short-term memory image in the current word is over the threshold, the viewpoint moves to next word. If not, the viewpoint is in the current word once again. The human eye movement that the viewpoint moves to next word, moves the viewpoint to the first word of next line and sees a word more than two times is simulated using this model.

**keywords:** Short-term memory, Foveated Vision, Reading

## 1 Introduction

The pixel density is high around the center of the retina and low in the peripheral of it. The vision such as this is called the foveated vision. The CCD devices are developed to generate foveated vision[1][2][3]. Some researchers propose the model of the eye movement based on foveated vision[4]. We uses the foveated vision to simulate the human vision. When human feels the movement of an object, the viewpoint moves to the next viewpoint reflectively. It is difficult to real-

ize the eye movement in wide region using the computation model, as the eye based on the foveated vision is controlled in the region of a foveated vision[4]. The short-term memory and task model are not discussed in the model of these eye movements.

In the other hand, the macro model consisted of the sensory register, the short-term memory, and the long-term memory is proposed in the psychological field [5]. The information derived from the sensory register is saved in the short-term memory. We introduce the short-term memory to the eye movement to realize the eye movement in wide region using the computation model. While the viewpoint moves quickly, we do not have the feeling that eye moves quickly. We show that the viewpoint moves to wide region by moving the viewpoint based on the short-term memory image generated using low level features. It is reported that the viewpoint moves to the first word of next line from the end of line in a moment. But we cannot identify the first word of the next line in the end of line, because the resolution is too low in the periphery of the retina. We introduce the short-term memory to identify the first word of the next line in the end of line. The viewpoint can be moved to the first character of next line using the image which is integrated the foveated images saved from the first word identified in the line. We call the image the short-term memory image in this paper. These mechanism is also used the eye movement that we see a word more than two times. If the edge features derived from the short-term memory image in the current word is over the threshold, the viewpoint moves to next word. If not, the viewpoint is in the current word once again.

It is difficult to realize the eye movement in the computation model because eye movement depends on tasks, while the psychological aspects in the eye movement is studied[6]. In the computer vision, the problem that defines the next viewpoint according to tasks is discussed [7][8][9][10][11]. In these studies, the viewpoint is defined according to tasks. But the eye movement based on the images received in the receptive field is not discussed. On the other hand, the

subsumption architecture that does not use the serial execution but the parallel execution of modules in the perception and control[12]. The parallel object-oriented model is studied in the field of the distributed architecture of the computer network[13]. It is discussed that it is important to define the viewpoint by detecting the edge features independently[11][14].

We propose the eye movement based on the parallel object-oriented model. We realize the eye movement in the small region that the viewpoint moves to the next word and in the wide region that the viewpoint moves to the next line in order to realize the eye movement in reading using the computation model by changing the range of the attention. We show the effectiveness by comparing the proposed model to the human eye movement [6][16].

We explain the foveated vision in section 2 and the short-term memory image in section 3 and the eye movement based on the task model and the short-term memory in section 4. In section 5, we show the effectiveness by introducing the eye movement in the computation model based on the task model and the short-term memory to the eye movement in reading.

## 2 Foveated vision

Wilson proposed the space variant receptive field. As the scale and rotation are variant, the log-polar mapping is used as the image sampling model of the machine vision. It is necessary to change the radius of the receptive field, as the radius in the center of the retina is bigger than the radius in the periphery of the retina. So the eccentricity of the  $n$ th ring  $R_n$  is the following:

$$R_n = R_0 \left( 1 + \frac{2(1 - Ov)Cm}{2 - (1 - Ov)Cm} \right)^n, \quad (1)$$

where  $R_0$  is the radius of the fovea.  $Cm$  is the ratio of the diameter of the receptive field to the eccentricity of that receptive field from the center of the fovea.  $Ov$  is the overlap factor. If the receptive fields touch each other,  $Ov = 0$ . Also, the radius of the receptive field on the  $n$ th ring is  $\frac{CmRn}{2}$ , the number of receptive fields is  $\frac{2\pi}{Cm(1-Ov)}$  per ring, and the angle of  $m$ th receptive field is  $\theta_m = Cm(1 - Ov) * m$ .

the retina and the sparse information in the peripheral vision. Figure The horizontal axis is  $m$  in  $\theta_m$  and the vertical axis is  $n$  in  $R_n$ .

## 3 Short-term memory image

In the short-term memory, the information can be saved in the short term by about 20 seconds. The two roles are to send the information received from the sensory register to the long-term memory. and

to encode the information according to the request. We pay attention to the relation between vision and short-term memory.

Though human recognize the space while the viewpoint moves quickly, human does not have the consciousness. We generate two dimensional image that human feels in brain, while the viewpoint moves quickly based on the short-term memory.

At first, we explain the algorithm generating of the short-term memory. The short-term memory image is the image which human feels in brain. We compare the  $n$  value in the previous image to the  $n$  value in a foveated image. If the  $n$  value is smaller than  $n$  value in the previous image, we renew the  $n$  value and RGB value in the pixel. The constraint of the capacity in the short-term memory is discussed. The time value is set in the pixel of the short-term memory to introduce the concepts of the capacity to the short-term memory. The pixel value that has been seen before, can be cleared from the short-term memory image using time value.

Next, we explain the algorithm moving eye based on the short-term memory image. The RGB value is detected in the position of the receptive field.  $R\theta$  image is generated from the information of the receptive field. The low-level features such as brightness, edge and corner are detected.

$$f_i = l(\sum_{j=0}^p (\alpha_j g(n_j))) \quad (2)$$

$f_i$  is the evaluation for the features of the  $p$  number such as brightness, edge and corner.  $l(\cdot)$  is the function.  $g(\cdot)$  is the function to encode the resolution information corresponding to the eccentricity of the receptive field.  $\alpha_j$  is the weight of feature  $j$ .

$$ff_i = \frac{f_i}{\sum_i f_i} \quad (3)$$

$ff_i$  is the probability moving the viewpoint to the feature  $i$ . The viewpoint is defined using the probability  $ff_i$ .  $g(\cdot)$ , the number  $p$  of the features, the features and the weight  $\alpha_j$  change according to tasks. We use the edge feature for the experiment in reading and the step function for the function  $g$ .

## 4 Algorithm of eye movement based on task model and short-term memories

In general, a task consists of tasks, as the task of the eye movement in reading is divided into the the tasks that the viewpoint moves to the next line and to the next word. In this paper, we propose the model that a task is divided into tasks, and the task is executed based on the event which is the trigger to execute.

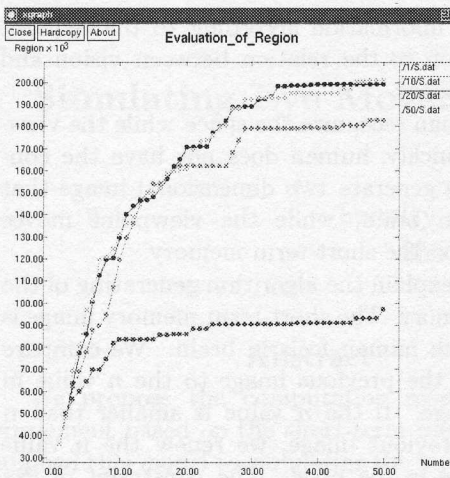


Figure 1: The relation between the size and the number of observation (in the case of the short-term memory which can save 1, 10, 20 and 50 foveated visions.)

In general, the capacity saving the memory differs according to the tasks. The region of the attention is controlled according to tasks. The event is the trigger to execute the tasks. The task is executed on the event while the short-term memory related to the task is referred. When the task is finished, the short-term memory image is cleared, and the next task is executed.

We explain the algorithm using figure 3. At first, we explain the algorithm generating the short-term memory image on the sensory register and the short-term memory. We explain the algorithm in the sensory register as shown in figure 3(a).

- $R\theta$  image is generated from the features gotten in the receptive field.
- The edge, corner and color features are detected from the  $R\theta$  image.
- $R\theta$  image, features and the candidates of the next viewpoint is sent to the short-term memory.

We explain the algorithm in the short-term memory as shown in figure 3 (b).

- The information of  $R\theta$  image derived from the receptive field is sent to the short-term memory .
- Short-term memory images based on  $n$  value, RGB value and feature are rewritten by recovering a  $R\theta$  image.
- The pixels are cleared from the short-term memory image, if the time is over the constant time.

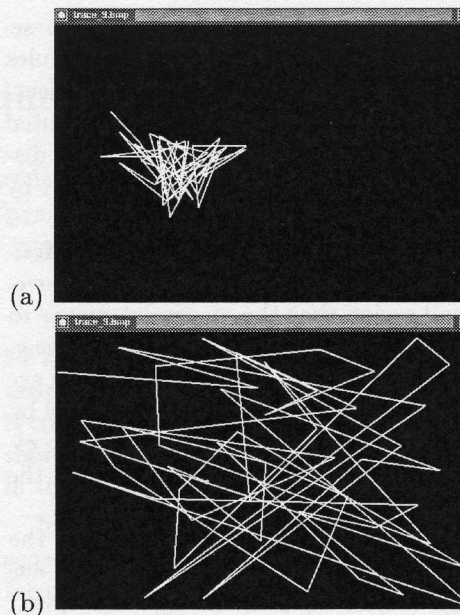


Figure 2: The viewpoint movements using the short-term memory which saves a image(Upper figure) and the short-term memory which saves 50 images (Lower figure).

- The next viewpoint is defined by evaluating the pixel density distribution.

Next, we explain the algorithm of the eye movement based on the task model and the short-term memory as shown in figure 3 (c ). As soon as the event is detected, the task corresponding to the event is executed. The short-term memory image is generated by integrating RGB values in the receptive field for the all tasks using the short-term memory images as shown in figure 3 (d ). These tasks are designed based on the object-oriented model. A task is defined as an object composed of a short-term memory feature image,  $n$  value image, time value image and a event.

## 5 Evaluations

### 5.1 Evaluation of the short-term memory

The region size of the short-term memory changes by the eye movements as the capacity of the short-term memory changes. We estimate the capacity of the short-term memory which can be saved the foveated images of 1, 10, 20 and 50 viewpoints. Figure 1 shows the graph that x axis is the number of eye movement, and y axis is the capacity of the short-term memory. Figure 2 (a) and (b) show the track of the eye movement generated using the computation model

of the short-term memory which can be saved a viewpoint and 50 viewpoints. Though the viewpoint moves in the small region in figure 2 (a), the viewpoint moves in the wide region in figure 2 (b). From these results, it is found that the region of the eye movement changes by changing the capacity that the short-term memory can be saved. It is possible to change the attention region according to tasks by defining the capacity of the short-term memory corresponding to a task.

## 5.2 Evaluation of the task model

To realize the eye movement in reading, it is required that the viewpoint moves to the next word in the small region and to the next line in the wide region. We introduce two short-term memories for the eye movement in reading in order to change the attention region according to tasks. The task that the viewpoint moves to the next word is executed on the line and the task that the viewpoint moves to the next line is executed on the end of line. The priority of the task that the viewpoint moves to the next line is higher than the priority of the task that the viewpoint moves to the next word. When the end of line is found, the task that the viewpoint moves to the next line is executed referring to the short-term memory. After that, the short-term memory related to the task is cleared. When the viewpoint is on the line, the task that the viewpoint moves to the next word is executed referring to the short-term memory and the short-term memory related to the task is cleared. After that, the short-term memory is saved again. These tasks are executed by computing the density of the short-term memory image.

To cover a word with the attention region, the attention region is rotated around a current fixation position and the number of edge pixels is calculated in the attention region with every direction. The edge density is measured the every direction in the forward and backward direction of reading. The word under the current viewpoint is extracted using the regions with high edge density in the forward and backward direction of reading (figure 4(a)) and the next word is extracted in the forward direction of reading (figure 4(b)). If the edge density of the attention region including the word under the current viewpoint is over the threshold, the next viewpoint is determined from the attention region covered the next word. If not, the next viewpoint is determined from the attention region covered the next word. The next viewpoint is defined from the edge features existed under space after the position which is the first character on the line is found (figure 4(c)). We can find the next viewpoint from the edge features existed under the first word in

the current line. If the short-term memory is used, an eye movement is only required to find the edge features existed under the first word of the current line. If not, many times of eye movements are required to find the edge features existed under the first word of the current line.

We compare the track of the eye movement using the short-term memory image to the track of the eye movement not using the short-term memory image in this experiment. We use  $R_0 = 0$ ,  $Cm = 0$  and  $Ov = 0$  parameters to generate the foveated vision. Figure 5 is the example of the text image using the experiment. Figure 6(a) and (b) show the short-term memory image for the task that the the viewpoint moves to the next word and to the next line. The short-term memory and the task model are not used in figure 6(c), but in figure 6 (d).

In the case of not using the short-term memory and task model, eye moves to the top of the line step by step and moves to the top of the next line. when the viewpoint is on the end of the line, In the case of using the short-term memory and task model, eye moves to the top of the next line when the viewpoint is on the end of the line. The difference between figure 6 (c) and (d) is the number of the eye movement. The reported human eye movement is not similar to the eye movement in the former computation model but similar to the eye movement in the latter computation model.

We simulate the eye movement in reading the folded text image to show the result clearly. Figure 7(a) is the example of the folded image used in the experiment. Figure 7 (b) and (c) show the short-term memory image for the tasks that the viewpoint moves to the next word and to the next line. The short-term memory and the task model in the computation model of the eye movement are not used in figure 7 (d) but in figure 7(e). Though the difference between figure 6 (c) and (d) is the number of the eye movement, the difference between figure 7 (d) and (e) is not the number of the eye movement but also the track of the eye movement. The human eye movement is measured using the eye movement detection machine, while human read the folded text image shown in figure 7(a). We use EMR-8 made by NAC company to measure it. The measured human eye movement is not similar to the eye movement in the former computation model but similar to the eye movement in the latter computation model.

It is easy to generate the computation model of the eye movement for the new task using the short-term memories and task model. As a task is defined as an object in the object-oriented model which composed

of a short term memory feature image, an event, the  $n$  value image and time value image, it is easy to add and delete a task to the computation model.

## 6 Conclusions

In this paper, we propose the eye movement in the computation model based on the short-term memory. We simulate the eye movement in reading that consists of the tasks that the viewpoint moves to the next line and the next word. The eye movement that the viewpoint moves to next line is simulated by introducing the short-term memory image, though the first word of next line can not be identified in the case of using the foveated vision because the resolution is too low in the periphery of the retina.

## Acknowledgements

The author thanks Prof. J. J. Koenderink for the discussions. Human eye movement was measured in Hakodate Future University with the machine detecting the human eye movement supplied by Prof. T. Kawashima.

## References

- [1] G. Sandini, P. Dario and F. Fantini, "A RETINA LIKE SPACE VARIANT CCD SENSOR," SPIE 1242, pp. 133-140, 1990.
- [2] T. Baron, M. D. Levine and Y. Yeshurun, "Exploring with a Foveated Robot Eye System," 12th International Conference on Pattern Recognition, pp. 377-380, 1994.
- [3] Y. Kuniyoshi, N. Kita, S. Rougeaux and T. Suehiro, "A Foveated Wide Angle Lens for Active Vision," Proc. IEEE Int. Conf. Robotics and Automation, pp. 2982-2988, 1996.
- [4] H. Yamamoto, Y. Yeshurun and M. D. Levine, "An Active Foveated Vision System: Attentional Mechanisms and Scan Path Coverage Measures," Computer Vision and Image Understanding, Vol. 63, No. 1, pp. 50-65, 1996.
- [5] R. C. Atkinson and R. M. Shiffrin, "Human memory: A proposed system and its control process," The Psychology of Learning and Motivation, Vol. 2, Academic Press, 1968.
- [6] A. L. Yarbus, "Eye movements and Vision," Plenum Press, 1967.
- [7] J. J. Clark and N. J. Ferrier, "Control of Visual Attention in Mobile Robots," Int. Conf. On Robotics and Automation, pp. 826-831, 1989.
- [8] M. J. Swain, R. E. Kahn and D. H. Ballard, "Low Resolution Cues for Guiding Saccadic Eye Movements," Proc. of CVPR'92, pp. 737-740, 1992.
- [9] L. Bidbaum, M. Brand and P. Cooper, "Looking for Trouble: Using Causal Semantics to Direct Focus of Attention," Proc. of 4th ICCV, pp. 49-56, 1993.
- [10] F. Ennesser and G. Medion, "Finding Waldo, or Focus of Attention Using Local Information," Proc. of CVPR'93, pp. 711-712, 1993.
- [11] R. Milanese et al., "Integration of Bottom-Up and Top-Down Cues for Visual Attention Using Non-Linear Relaxation, Proc. of CVPR'94, pp. 781-785, 1994.
- [12] R. Brooks, "A robust layered control system for a Mobile Robot," IEEE J. Robotics and Automation, vol. RA-2, pp. 14-23, 1986.
- [13] A. Yonezawa and M. Tokoro, "Object-oriented concurrent programming," The MIT Press, 1987.
- [14] A. M. Treisman and G. Gelade, "A Feature-Integration Theory of Attention," Cognitive Psychology, 12, pp. 97-136, 1980.
- [15] S. W. Wilson: "On the retina-cortical mapping," Int. J. Man-Machine Stud. 18, pp. 361-389, 1983.
- [16] R. A. Monthy and J. W. Senders, Eye Movements and Psychological Process, pp. 371-395, Lawrence Erlbaum Associates, 1976.

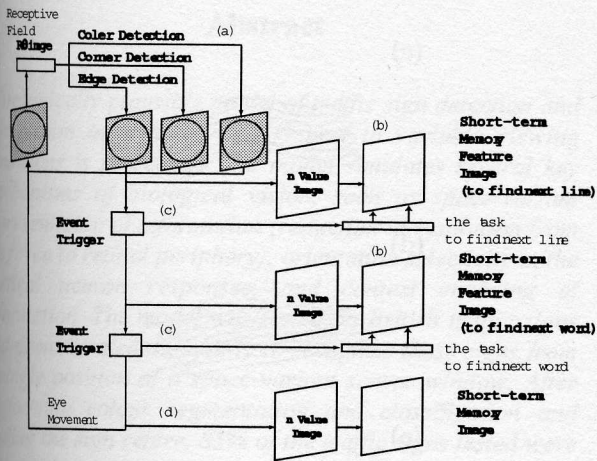


Figure 3: Algorithm of the eye movement based on the task model and the short-term memories in reading.

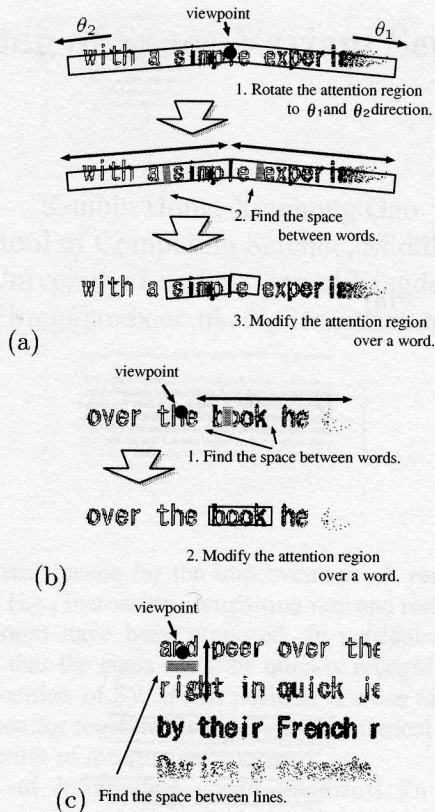


Figure 4: (a)The modification of the attention region. (b)The extraction of the next word.(c)The extraction of next line.

## CHAPTER I

A great loser have I been ever since a child, having a rash, daring spirit, a spirit I inherited from my ancestors. When a primary-school boy, I jumped down from the second story of the schoolhouse, and had to lie abed about a week. Some may be curious enough as to ask me why I did such a rash thing. I had no special reason. I had one day been looking down from the window of the second story of the new school, when one of my classmates looked up and said in a taunt.

"You are a great boaster. But you cannot jump down from that height. You poor little creeping thing!" My father, seeing me come home on the back of the janitor, said in an angry tone that no strong boy could be hurt by jumping down from such a low height as from upstairs, and I assured him that he would be proud to see me come out all right next time.

A foreign-made penknife had been given me by one of my relations, and I was showing it proudly to my comrades, the bright blades reflecting the sunlight, when one of the boys said that bright as it shone it was a dull knife

Figure 5: Text image

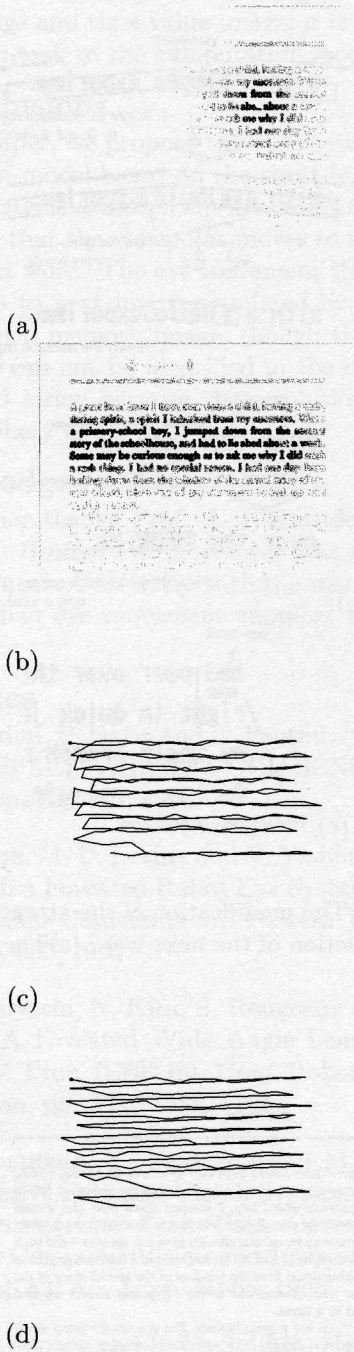


Figure 6: (a)The feature image on the foveated vision for the task that the viewpoint moves to next word along the line. (b)The edge feature image of short-term memory for the task that the viewpoint moves to next line. (c)The track of the eye movement without using the task model and the short-term memories. (d)The track of the eye movement using the task model and the short-term memories.

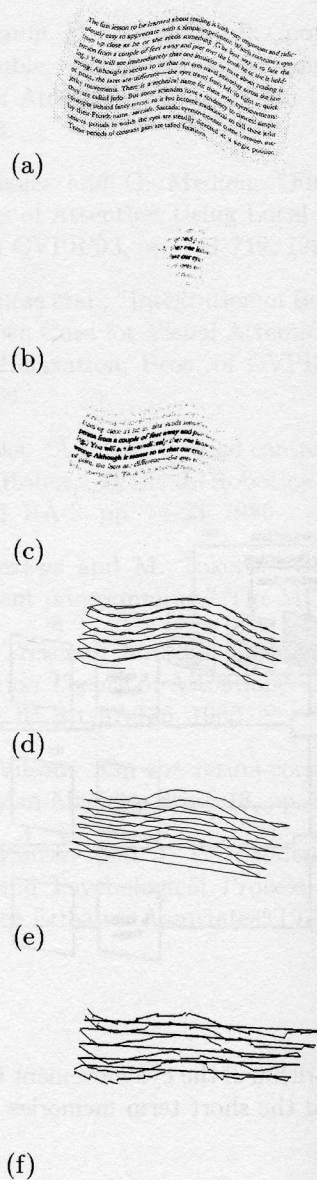


Figure 7: (a)Folded text image.(b)The edge feature on the foveated image for the task that the viewpoint moves to the next word along the line. (c)The edge feature image of short-term memory for the task that the viewpoint moves to the next line. (d)The track of eye movement without using the task model and the short-term memories. (e)The track of the eye movement using the task model and short-term memories. (f)The result measuring human eye movement.