Simulating binocular eye movements based on 3-D short-term memory image in reading

Satoru Morita
Faculty of Engineering,Yamaguchi University
E-mail: smorita@yamaguchi-u.ac.jp

Abstract

We simulate binocular eye movements in reading. We introduce the 3-D edge features reconstructed from the binocular foveated vision to determine the next fixation point in reading. The next fixation point is determined statistically from the feature points in the 3-D short-term memory edge image. We show the effectiveness of simulating eyes movement based on 3-D short-term memory image to realize humanlike robots.

1 Introduction

It is important to determine eye movement to realize the vision of autonomous robot such as human vision. It is proposed to determine tasks and the viewpoint according to it in the field of computer vision[1][2]. On the other hand, it is based on the psychological experiments[3]that the viewpoint determine according to the given tasks. On the other hand, it is one of important problem that many psychologist studies[4]. It aims at realizing the viewpoint movement in the task of reading. The vision system used CCD device is realized based on the foveated vision that the resolution in the retina center is high and the resolution in the retina periphery is low[5].

But it is not argued for the binocular eye movements in reading. Human understands the 3-D world by computing the depth from the binocular foveated vision. Thus, we realize the viewpoint movement in reading based on the 3-D world derived from binocular eye movements. In this paper, we realize binocular eye movements that viewpoint moves in the wide region such as to search the next line and word. We use the 3-D edge features reconstructed from the foveated vision of binocular eyes to determine the next fixation point in reading. It is argued that human behaviour is related to human memory[6]. In this paper, we introduce the 3-D short-term memory related to binocular vision. The next fixation point is determined statistically from the feature points saved on the short-term memory. We show the effectiveness of simulating eyes movement based on 3-D short-term memory image to realize humanlike robots .

2 Foveated Vision

The center of the retina is called the fovea. Vision in which resolution is low at the periphery of the retina and high at the center of it is called foveated vision. Because the log-polar mapping model varies its scale in rotation, it is used widely as a image sampling model. Wilson proposed the arrangement of the receptive field according to the mapping model and explained the human sensing facility related to the contrast[7]. The receptive field is located on circles whose center is the center of the retina. In order to realize the types of vision, it is necessary that the resolution diminishes as the radius of the circle grows larger. The eccentricity R_n of the nth circle is defined in the following:

$$R_n = R_0(1 + \frac{2(1-Ov)Cm}{2-(1-Ov)Cm})^n \qquad (1)$$

R_0 is the radius of foveated vision, and C_m is the rate between the eccentricity and the distance from the retina center to the center of the receptive field. Ov is the overlapping rate between the receptive fields in the neighbor.

3 Calculation of determining the camera directions based on stereo vision

We describe the necessary technique to reconstruct the 3-D world based on binocular eyes.

3.1 Calculation of camera matrix

Camera matrix M is the translation matrix used to translate from world coordinates to camera display. It is defined using camera position, direction and focus and image size. Camera matrix is calculated as the products of the following four matrix.

$$M = T \cdot M_1 \cdot M_2 \cdot M_3 \qquad (2)$$

T is the translation matrix used to translate from the point coordinate to the homogeneous coordinate . M_1 is the perspective translation matrix used to translate from the homogeneous coordinate to the display in the 3-D coordinate on the display. M_2 is the matrix used to translate from the 3-D coordinate on 3-D display to the 2-D

coordinate on the display. M_3 is the matrix used to translate from the 2-D coordinate to the 2-D display. Camera matrix of left camera and right camera is calculated in the following.

$$M_L = T_L \cdot M_{L1} \cdot M_{L2} \cdot M_{L3} \quad (3)$$

$$M_R = T_R \cdot M_{R1} \cdot M_{R2} \cdot M_{R3} \quad (4)$$

We calibrate using the method of Tsai[9] in the case of simulating the eye movement using the real pan-tilt camera.

3.2 Calculation of the points in 3-D space from binocular images

The point of the 3-D space P is represented using the homogeneous coordinate as $P[p] = P[p_1, p_2, p_3, 1]$. If h_L and h_R are real and the coordinate on left image is $P[p_L] = P[p_{L1}, p_{L2}, 1]$ and the coordinate on right image is $P[p_L] = P[p_{L1}, p_{L2}, 1]$,

$$h_L < p_{L1}, p_{L2}, 1 >= [< p_1, p_2, p_3, 1 > M_L] \quad (5)$$

$$h_R < p_{R1}, p_{R2}, 1 >= [< p_1, p_2, p_3, 1 > M_R] \quad (6)$$

. Thus, the coordinate (p_1, p_2, p_3) is calculated from the right image $< p_{R1}, p_{R2}, 1 >$, the left image $< p_{L1}, p_{L2}, 1 >$ and the camera matrix M_L, M_R.

3.3 Determining of the camera direction from the fixation point

The pan and tilt rotation angle $< \theta_l, \phi_l >$ and $< \theta_r, \phi_r >$ for the left and right camera are calculated from the fixation point (p_1, p_2, p_3) in the 3-D world. The rotation center of the left camera is (cl_1, cl_2, cl_3) and the rotation center of the right camera is (cr_1, cr_2, cr_3). The direction of the left camera is $X' < i_{L1}, i_{L2}, i_{L3} >$ and the direction of the right camera is $X' < i_{R1}, i_{R2}, i_{R3} >$. The direction of the left camera Y' required to look at the fixation point is calculated in the following:

$$A =< p_1 - cl_1, p_2 - cl_2, p_3 - cl_3 > \quad (7)$$

$$Y' =< \frac{p_1 - cl_1}{|A|}, \frac{p_2 - cl_2}{|A|}, \frac{p_3 - cl_3}{|A|} > \quad (8)$$

. The rotation matrix R_L is calculated using $X'Y'$ as

$$Y' = R_L X' \quad (9)$$

$$Y'(X'X'^T)^{-1}X'^T = R_L \quad (10)$$

. After the pan tilt camera rotates as the rotation axis is y axis, rotates as the rotation axis is x axis. The rotation around the y axis shows T_{ry}, and the rotation around the x axis shows T_{rx}. R_L is represented as

$$R_L = T_{rx} \cdot T_{ry} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \quad (11)$$

, T_{ry} and T_{ry} are calculated using ϕ_l and θ_l as

$$T_{ry} = \begin{bmatrix} cos\phi_l & 0 & sin\phi_l \\ 0 & 1 & 0 \\ -sin\phi_l & 0 & cos\phi_l \end{bmatrix} \quad (12)$$

$$T_{rx} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\theta_l & sin\theta_l \\ 0 & -sin\theta_l & cos\theta_l \end{bmatrix} \quad (13)$$

. Therefore, we can get R_L from previous equations:

$$sin\phi_l = r_{13} \quad (14)$$

$$cos\theta_l = r_{22} \quad (15)$$

. In the case of

$$X' =< i_{L1}, i_{L2}, i_{L3} >=< 0, 0, 1 >$$

, as the following equation is gotten:

$$sin\phi_l = \frac{p_1 - cl_1}{|A|} \quad (16)$$

$$cos\theta_l = \frac{p_2 - cl_2}{|A|}/(1 - sin^2\phi_l) \quad (17)$$

, ϕ_l, θ_l can be calculated. In the similar, the right camera direction Y'' required to look at the point (p_1, p_2, p_3) is calculated :

$$Y'' =< \frac{p_1 - cr_1}{|A|}, \frac{p_2 - cr_2}{|A|}, \frac{p_3 - cr_3}{|A|} > \quad (18)$$

.

$$A =< p_1 - cl_1, p_2 - cl_2, p_3 - cl_3 > \quad (19)$$

The rotation matrixR_R is calculated :

$$Y''(X''X''^T)^{-1}X''^T = R_R \quad (20)$$

$$Y'' = R_R X'' \quad (21)$$

In the similar, ϕ_r, θ_r are calculated. Thus, left camera direction $< \phi_l, \theta_l >$ and right camera direction $< \phi_r, \theta_r >$ are calculated from the fixation point (p_1, p_2, p_3).

4 Binocular eye movements in reading based on 3-D short-term memory image

We describe the eye movement in reading based on short-term memory image and foveated vision.

4.1 3-D short-term memory image

Short-term memory saves the information for about 20 seconds[6]. Human does not have the consciousness while the fxation point moves quickly. This is reason why the image saved for the short term does not change while the foveated image of the viewpoint changes quickly. Thus, we realize the observed image that does not change suddenly though the fixation point moves in the wide region. We take the attention to the short-term memory related to vision. In especially, the vision system should not save 2-D image but 3-D image in short-term mermoy image to determine the motion of binocular eye balls.

The feature is saved in the 3-D short-term memory after 3-D point is recovered from these feature in the case that the feature of right image is same as the feature of left image and the feature is on the epipolar line of left camera corresponding to a feature of right camera. The next fixation point is statistically determined from the feature points of the short-term memory. Thus, 3-D coordinates (x, y, z) are recovered from the two feature points in right and left camera.

4.2 Generating 3D short-term memory feature image

We explains the algorithm of 3-D short-term memory edge feature image.

1) A pixel is selected from the xy coordinate of right foveated vision. A plane determined from the center of two cameras and a pixel on the image plane of the right camera. The epipolar line is generated by intersecting the plane with the left image plane.

2) We determine the rotation that the sum of absolute difference between RGB values is minimum using the correlation method for each pixel, we get the corresponding point on the epipolar line[8].

3) The 3-D point coordinate is recovered from xy coordinate gotten from the foveated vision of right and left camera ((3.1)(3.2)).

4) The color and the time value are saved in the coordinate of the 3-D point in the 3-D short-term memory image. We generate 3-D short-term memory image using voronoi algorithm from 3-D point sets in 3-D short-term memory.

5) The sequences 1)2)3)4) are applied for the xy coordinate corresponding to each pixel of foveated vision and the 3-D points are reconstructed.

6) The next fixation point is determined from the edge feature image included in the attention region in 3-D short-term memory edge feature image(4.3). The motion of the right camera and left camera is determined(3.3).

7) The time value is incremented. The difference between the current time and the time on the short-term memory image is calculated. If the difference is over the constant value, the point is deleted in the short-term memory image.

4.3 Binocular eye movements in reading

In the case that the fixation point is on the line in reading, we use the square mask. The direction that the most edges exist in the square mask is determined while the square mask is rotated. The attention region is composed of the regions to the space from the fixation point backward and forward for the reading direction. If the edge number is not over the constant value, the next fixation point is in the attention region. If the edge number is over the constant value, the next fixation point is in the region from the next space following forward the current attention region to the space after the space. Thus, it is simulated to determine the next fixation point both backward and forward. In the case that a fixation point is on the end of the line, the fixation point moves to the top of the next line through the line saved in the 3-D short-term memory image. At first, the direction that edges are many is determined using the square. In this time, the attention region is selected in the opposite of the reading direction. The top of the line is detected by repeating this. After the space is found under the current line, the next fixation point is determined from the attention region that the edges are many under the space.

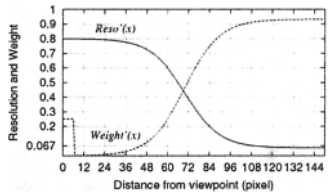

Fig. 1. The change of resolution and weight on the foveated vision. : $Reso'(x)$ and $Weight'(x)$ in the case of $\alpha = 0.8$, $\beta = 0.25$.

α makes the resolution $Reso'(x)$ change. The resolution becomes low, if α decreases. β makes the weight $Weight'(x)$ change. The frequency that the next fixation point is determined to the feature points where the distance between the fixation ooint and the feature point is less than 6 pixels is high, if β is big.

Figure 1 shows the change of resolution and weight on the foveated vision. $Reso'(x)$ and $Weight'(x)$ is determined as $\alpha = 0.8$,

$\beta = 0.25$. The resolution $Reso'(x)$ calculated the feature point number per unit volume from the short-term memory feature image. In this paper, the edge length of a cube is 24 pixels in a unit volume.

$$Reso(x) = \frac{1+\exp(-ab)}{1+\exp(a(x-b))}(1.0-0.067)+0.067 \quad (22)$$

$$Weight(x) = 1.0 - Reso(x) \quad (23)$$

$$Reso'(x) = \alpha Reso(x)$$

$$Weight'(x) = \begin{cases} 1.0 - Reso(x) & if\ x > 6 \\ \beta & otherwise \end{cases} \quad (24)$$

If the edge number including in the attention region is N, the probability f_i to determine the next fixation point to the ith edge is defined in the following.

$$f_i = \frac{Weight(x_i)}{\sum_{j=1}^{N} Weight(x_j)} \quad (25)$$

5 Simulating binocular eye movements in reading

We simulate binocular eye movements in reading. The pan-tilt cameras that the focus length is $f = 519.615$ are arranged in $(-5, 0, 0)$ and $(5, 0, 0)$. The document size is $512 * 256$ pixels. The center position of the document is fixed in $(0, 0, 700)$. We use the camera that the focus distance f is 519.615. The parameter requiered to generate the foveated image is set in $R_0 = 7$,$C_m = 0.5$,$O_v = 0.9$ and $length = 170$. Figure2(a) shows edges detected

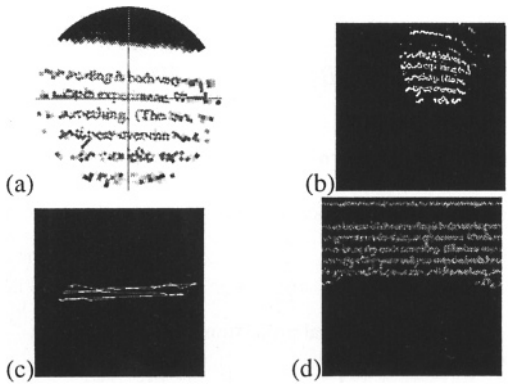

Fig. 2. (a)Edges detected on the right foveated vision at the middle of the second line. (b)The 3-D edge recovered from the foveated vision of the right and left camera(c)The viewpoint movement.(d)The 3-D short-term memory edge image when the fixation point is on the end of second line.

on the right foveated vision at the middle of the second line. Figure2(b) shows The 3-D edge recovered from

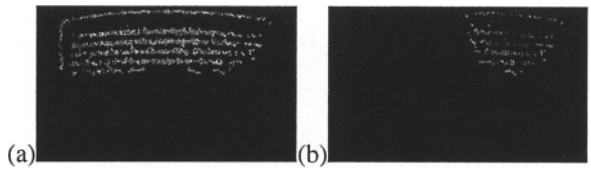

Fig. 3. (a) The 3-D short-term memory edge image for the curved document when the fixation point is on the end of second line.(b)The edge image generated from a foveated vision when the fixation point is on the end of second line

the foveated vision of the right and left camera. Figure2 (c) shows the viewpoint movement. Figure2(d) shows the 3-D short-term memory edge image when the fixation point is on the end of second line. Figure3(a) shows the 3-D short-term memory edge image for the curved document when the fixation point is on the end of second line. Figure3 (b) shows the edge image generated from a foveated vision when the fixation point is on the end of second line. The 3-D short-term memory edge is recovered adequately. It is found that the viewpoint movement to read the document located in the space is realized from these results. Binocular eye movements in reading the curved document in the bound book was simulated using this model.

6 Conclusions

The binocular eye movements in reading based on the 3-D short-term memory are simulated by controlling two pan-tilt cameras.

References

[1] M. J. Swain, R.E. Kahn, and D. H. Ballard (1992) Low resolution cues for guiding saccadic eye movements. CVPR, pp. 737-740

[2] L. Birnbaum, M. Brand, and P. Cooper (1993) Looking for trouble: Using causal semantics to direct focus of attention, ICCV, pp. 49-56

[3] A. Yarbus (1967) Eye movements and vision. Plenum Press

[4] Levy-Schoen, A and ORegan, K. (1979) The control of eye movements in reading, Proc. visible language, Plenum Press, pp. 7-36

[5] G. Sandini, P. Dario and F. Fantini (1990) A RETINA LIKE SPACE VARIANT CCD SENSOR," SPIE 1242, pp. 133-140

[6] R. C. Atkinson and R. M. Shiffrin (1968) Human memory: A proposed system and its control process. The Psychology of Learning and Motivation, Vol. 2, Academic Press

[7] S. W. Wilson, (1983) On the retina-cortial mapping. Int. J. Man-Machine Stud. 18, pp. 361-389

[8] T. Kanade and M. Okutomi, (1994) A stereo matching algorithm with an adaptive window: Theory and experiment, IEEE Trans, PAMI, Vol. 16, No. 9, pp. 920-932

[9] R. Y. Tsai, (1986) An efficeint and accurate camera calibration technique for 3D machine vision, CVPR, pp. 364-374