ACCV'98: Third Asian Conference on Computer Vision (Hong Kong, China), Vol.I, pp.256-263: 1998/1/8-10

Foveated Vision for Scene Exploration

Naoki Oshiro[1], Atsushi Nishikawa[2], Noriaki Maru[3] and Fumio Miyazaki[2]

[1] University of the Ryukyus, 1 Senbaru, Nishihara-cho, Okinawa 903-01, JAPAN
[2] Osaka University, 1-3 Machikaneyama-cho, Toyonaka city, Osaka 560, JAPAN
[3] Wakayama University, 930 Sakaedani, Wakayama 640, JAPAN

Abstract. In this paper, foveated vision for scene exploration is implemented. The peripheral and central vision are the basic capabilities of foveated vision. The informations obtained from the peripheral vision are used to determine the next gaze point. Due to the low resolution of the periphery, however, the determination is not always appropriate. To solve this problem, we propose to evaluate the target object by the central vision after gazing. We implement foveated vision based on the Log Polar Mapping (LPM) and construct an evaluation scheme of the target object in the central vision using LPM rotational-invariance. The peripheral vision is realized by Zero Disparity Filter for LPM stereo images. Some experimental results are also shown to demonstrate the effectiveness of the proposed method.

1 Introduction

A human possesses foveated vision and the retina has a non-uniform structure. The retina's resolution is high in the central area and this resolution continuously drops as you go into the peripheral area. As detailed information can be obtained only in the central area (*central vision*), in order to search in the scene, gaze process becomes necessary. Meanwhile, the information from the peripheral area is used to determine the next gaze point (*peripheral vision*). That is, using the peripheral vision, we determine the gaze point and the target object will be analyzed based on the central vision. These are indispensable elements once foveated vision is used, and both coordinated movements are required.

In light of this, a method for scene exploration by foveated vision is discussed. In the previous works, we have proposed a method for using Log Polar Mapping (LPM), which is a good approximation of the non-uniform structure of the retina, to extract and track a gaze object by the vision system [4]. In this paper, we implement foveated vision based on the LPM and apply to scene exploration.

Our scene exploration algorithm by foveated vision consists of two parts: (i) feature extraction process (*peripheral vision*) and (ii) feature evaluation process (*central vision*). We first detect rapid changes in image intensity levels as features. Then, the features in the neighborhood of the gaze point are picked up from them; this is realized by Zero Disparity Filter for LPM stereo images [4]. These features are selected as the candidate of the next gaze point. Due to the low resolution of the information obtained from the peripheral vision, however, the next gaze point is not always appropriate. To solve this problem, we propose

to evaluate the feature after gazing. We construct this evaluation scheme of the target object by making use of the LPM rotational-invariance. As a result, each point selected as the next gaze point through peripheral vision is evaluated as being suitable or not for fixation. Some experimental results are also shown to demonstrate the effectiveness of the proposed method.

2 Log Polar Mapping in Our Implementation

In primate visual systems, the retina has a space variant structure; the resolution is high at the central region and becomes decreasingly low toward the periphery. The mapping of the retina into the visual cortex can be approximately represented by "Log Polar Mapping" (henceforth LPM) [7]. Formulation of LPM as follows:

$$w = \ln(z), \quad z = x + iy, \quad w = u + iv \tag{1}$$

where z is a point on the original image plane Z , and w is a point on the LPM image plane W . Denoting the image resolution of Z-plane by (w, h), and that of W-plane by $(u_{\max}, v_{\max})$ respectively, the transformation from $z = (x, y)$ to $w = (u, v)$ is represented by

$$u = u_{\max} \times \left(\frac{\rho - \rho_{\min}}{\rho_{\max} - \rho_{\min}} \right) = u_{\max} \times \frac{\rho}{\rho_{\max}}, \quad v = v_{\max} \times \frac{\text{atan2}(y, x)}{2\pi} \tag{2}$$

where $\rho_{\max} = \ln(\max(w, h)/2), \rho_{\min} = \ln(1) = 0, \rho = \ln((x^2 + y^2)^{\frac{1}{2}})$. In this paper, LPM is performed in software (look up table form) [4], [8]:

$$w = \text{LUT}(z) \tag{3}$$

where LUT means the look up table for the original image $Z(x, y)$.

LPM has the following characteristics: (i) space-variant structure (ii) functional vision (iii) precessing data reduction. Mathematically, LPM is (iv) rotation-invariant and (v) scale-invariant, so it is useful for pattern recognition [2]. Notice that these two properties ((iv) and (v)) are satisfied only in case that the projection of the target object is at/near the image center. Thus, if we want to apply these properties to the target object, we must gaze it. In this paper, we propose a "scene exploration" method by making use of these characteristics. We find a target object by using the functional vision of LPM (*peripheral vision*) and gaze the object (*central vision*). Furthermore, we evaluate the uniqueness of the gaze object by utilizing the rotation-invariance of LPM.

Next, we present the determination method of the border between the fovea and the periphery. Eq. 3 represents many-to-one correspondence. Therefore, when doing the LPM transformation (Eq. 2) by Eq. 3, sparsely transformed region ($u < u_{\text{fovea}}$) exists in the fovea. To overcome this problem, the following methods have been proposed: (a) Separately generate the fovea [9], (b) Fill up the value obtained by the inverse LPM, (c) Formulate LPM as $w = \ln(\text{z} + \alpha)$ [8] where α is the singularity avoidance parameter. We cope with this by only using the peripheral LPM image data. To do this, we determine the border between the fovea and the periphery as follows (see Fig. 1):

1. Initially, we decide the resolution of Z-plane, and that of W-plane.
2. The minimal number of partitions in the rotational direction of the Z plane corresponds to $v_{\max}$, we have Therefore, letting r_{fovea} be the radius such that this partition are individually done, we have $r_{\text{fovea}} \approx v_{\max} \times \sqrt{2}/8$. This means that the diagonal length of the square whose peripheral length is $v_{\max}$ is regarded as the diameter of the circle.
3. Denoting the border line between the fovea and the periphery on the LPM image by $u = u_{\text{fovea}}$, we obtain

$$u_{\text{fovea}} = \text{int}\left(\frac{1}{u_{\max}} \times \frac{\ln(r_{\text{fovea}})}{\ln(w/2)}\right) \tag{4}$$

   where "int" indicates the rounding operator.

We only use the LPM image region such that $u \geq u_{\text{fovea}}$.

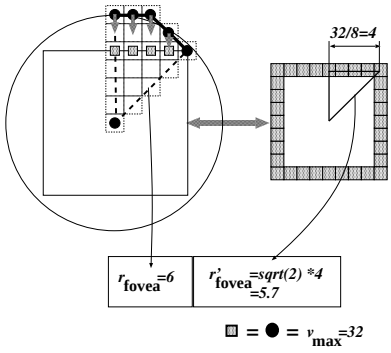


Fig. 1. Determination of the border between the fovea and the periphery

3 Peripheral Vision for Scene Exploration: Feature Extraction

In this section, we explain the function of *peripheral vision* for scene exploration. This corresponds to the feature extraction process for selecting the next gaze.

3.1 Periphery Filtering and Feature Extraction

The extracting method of scene features in our system based on the Zero Disparity Filter (ZDF) [3] which is a well-known method of using a binocular visual system to extract objects on the zero disparity area (*horopter*). The practical problem in applying the ZDF method is how to estimate binocular disparities.

To do this, a few methods have been proposed [1], [6], [10] . However, these are only applicable to uniform sampling images.

In our previous work, we have developed the LPM-ZDF method [4], which is the ZDF method applicable to LPM images. The interest points in the LPM-ZDF method are as follows: (i) stereo correspondence errors are reduced, (ii) the horopter of LPM spreads from the gaze point toward the periphery. (iii) wide and multiple disparity estimators can be worked in real-time. Details of the LPM-ZDF method can be found in [4].

We utilize the LPM-ZDF method to extract features from the periphery of LPM stereo images. It is summarized as follows: (i) Consider that stereo cameras gaze at a given point, (ii) Edges of the features located on the zero disparity area are picked up from LPM images by using the LPM-ZDF method, (iii) Continuous edge region are detected as the candidates of the next gaze point. They are selected in order of the amount of their area in terms of the pre-transformed original image.

3.2 Related Works

Generally speaking, gaze control algorithms are classified into two types: (i) context-dependent [5] and (ii) context-free. The latter uses general image features. For example, Yamamoto et al. [9] used an interest operator to detect geometrical image features such as corner or symmetry for selecting a sequence of gaze points. Our method also falls into the latter one. The major difference between Yamamoto et al.'s method and ours is that we only detect rapid changes in image intensity levels as features.

4 Central Vision for Scene Exploration: Feature Evaluation

In this section, we present the function of *central vision* for scene exploration. In our system, this corresponds to the feature evaluation process based on the rotational-invariance of LPM images.

The reasons of feature evaluation need as follows: we can select the next gaze point from the feature extraction results described in Section 3. However, there are the following problems: (a) the *ambiguity* of the extracted feature description due to the low resolution of the periphery (b) the *uncertainty* of the extraction results due to the same reason This means that we cannot always select the "appropriate" gaze point by the peripheral vision function. In foveated vision, the high resolution can be automatically realized by gazing process (*central vision*). Therefore, it is very reasonable to evaluate the feature object after gazing.

4.1 Feature Evaluation using Rotational-Invariance of LPM

We evaluate features by utilizing the rotational-invariance of LPM images. At first, we calculate the autocorrelation value for each image obtained by rotating the LPM image as follows:

1. Calculate the mean intensity μ:

$$\mu = \frac{1}{n} \sum_{u=u_{\text{fovea}}}^{u_{\max}-1} \sum_{v=0}^{v_{\max}-1} L_{u,v}, \quad n = (u_{\max} - u_{\text{fovea}}) \times v_{\max}. \tag{5}$$

2. Calculate the autocorrelation value c_i:

$$c_i = \sum_{u=u_{\text{fovea}}}^{u_{\max}-1} \sum_{v=0}^{v_{\max}-1} (L_{u,v} - \mu) \times (L_{u,v-i} - \mu) \qquad (i = 0, 1, \ldots, c_{\max} - 1), \tag{6}$$

   where $c_{\max} = v_{\max}/2$, $L_{u,v}$ is the intensity value at $W(u, v)$ on the LPM image plane.
3. Normalize c_i by the maximal value $\max_i(c_i)$.

Then, we can judge the *uniqueness* of the feature by using the resulting values $\{c_i\}(i = 0, 1, \ldots, c_{\max} - 1)$. That is, the more biased the distribution of these autocorrelation values are, the more *unique* in the scene the feature on the LPM image is. Notice that the value c_0 is always 1 and the other values c_i are zero if the feature is the most *unique*. Thus, we use the sum of the autocorrelation values, denoted by S_c, as the criterion for *uniqueness*.

$$S_c = \frac{1}{c_{\max}} \sum_{i=0}^{c_{\max}-1} c_i \tag{7}$$

The minimal value of S_c is $1/c_{\max}$. The less *unique*, the greater S_c.

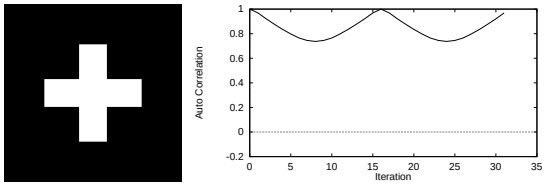
Figure 2 shows some examples. Each test images, we calculated their autocorrelation values (graphs in Fig. 2) and the criterion for *uniqueness* S_c. Original image size $(w, h) = (256, 256)$, and LPM image size $(u_{\max}, v_{\max}) = (34, 64)$ in these examples. (a) cross: the autocorrelation values take the shape with periodic iteration in the rotational direction. It can be regarded as the *non-unique* feature. (b) ellipse: almost all the autocorrelation values are near 1. It can be also regarded as the *non-unique* feature. (c) mnote: the autocorrelation values relatively take the shape of impulse. It is considered as the *unique* feature.

5 Experiment

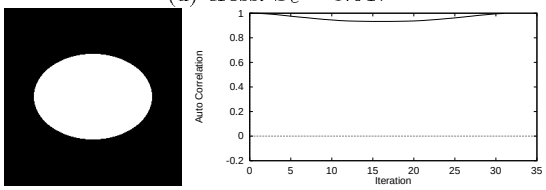
5.1 System Configurations and Experimental Setup

The cameras tilt up and down together on a common platform, and pan independently from side to side, driven by three motors: one for the tilt platform and twin pan motors. Besides the cameras are mounted on a head which pans and slides independently, driven by two motors: one for the pan of the head and one slide motor. In the experiment, we use and control the three pan motors (eyes and neck) and tilt motors.

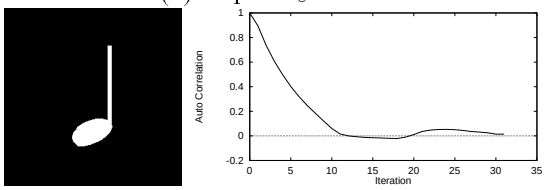
Stereo images sequences taken by the two CCD cameras are digitized into 512×480 pixel, 8 bit gray scale images and processed by the host computer (Sun



(a) cross: $S_c = 0.847$



(b) ellipse: $S_c = 0.961$



(c) mnote: $S_c = 0.169$

Fig. 2. Results of Feature Evaluation (Test Images)

SPARC station 10). The baseline length is 120 (mm) and focal length is 4.8 (mm) respectively. Details in relation to "Active stereo vision system" (motor control system, performance specifications etc.) can be found in [4].

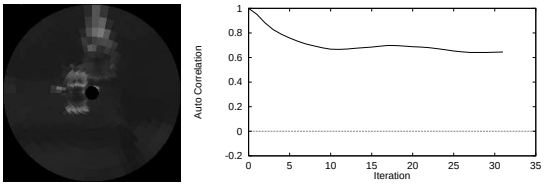
In experimental setup, we assume the black background for simplicity. Initial gaze point was set right in front of the Vision Robot at a distance of 1000 mm, and three objects were put in the vicinity of the gaze point. Figure 3 shows the stereo images obtained by initial gazing process. Original image size $(w, h) = (256, 256)$, LPM image size $(u_{\max}, v_{\max}) = (34, 64)$ in this experiment.

5.2 Results

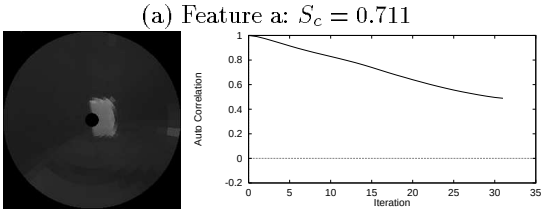
The proposed method has been applied to the actual scene. Figure 3 shows the feature extraction result; '+' indicates the feature points. We can see that all of the objects are appropriately picked up from the scene. As a result, three region (numbered as 1, 2, 3) were selected as the candidate of the next gaze point.

Figure 4 shows the inverse LPM images obtained after gazing at the selected feature region, in order of (a), (b), (c). The autocorrelation values (graphs in Fig. 4) and the criterion of *uniqueness* S_c are also shown. From these results, the feature region 1 (corresponding to Fig. 4(a)) is considered as the most *unique* feature.

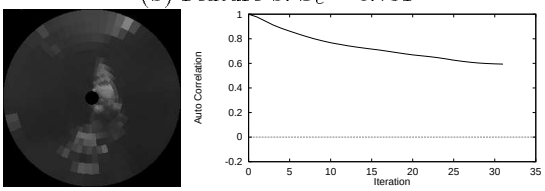
Fig. 3. Stereo Images of after Initial Gaze Process and Results of Feature Extraction by Peripheral Vision



(a) Feature a: $S_c = 0.711$



(b) Feature b: $S_c = 0.731$



(c) Feature c: $S_c = 0.736$

Fig. 4. Results of Feature Evaluation by Central Vision

6 Conclusion

In this paper, we have proposed a method to realize "scene exploration" with an active stereo foveated vision system by making use of functional vision (*peripheral and central vision*). The foveated vision system was implemented in software by using LPM. Our exploration algorithm first detects the feature points based on the peripheral vision and evaluates them based on the central vision. Peripheral

vision were implemented by the LPM-ZDF method that extracts features in the neighborhood of the gaze point. From this, it becomes possible to selectively extract from three dimensional space. By using the rotational-invariance of the LPM image, we calculated the autocorrelation value. As a result, we evaluated the feature by looking at the uniqueness in the rotation of the image taken from the central vision after gazing. In addition, we have shown the effectiveness of our method by the experiment.

Using the functional vision, the vision implementation is divided into two parts: where and what dichotomy. This process is important for solving the problems in the computer vision field such as reducing the amount of process data and process time. By making use of the space variant structure of foveated vision, we proposed a new evaluation method of the gaze object. This method can be used for object tracking problems.

Using stereo cameras, it is possible to selectively extract the feature object in three dimensional space, with respect to the distance. However, the selectiveness that is used is limited to only the process in the peripheral vision, and it is not used in the central vision. Since the relationship with the background can be considered to be a large influence, it is probably efficient to use the selectiveness in this process when evaluating the feature.

References

1. C. Capurro, F. Panerai and G. Sandini, "Dynamic Vergence", *IROS'96*, Vol. 3, pp. 1241–1248, 1996.
2. L. Massone, G. Sandini and V. Tagliasco, "'Form-Invariant' Topological mapping Strategy for 2D Shape Recognition", *Computer Vision, Graphics and Image Processing*, No. 30, pp. 169–188, 1985.
3. T. J. Olson and D. J. Coombs, "Real-Time Vergence Control for Binocular Robots ", *Int. Jounal of Computer Vision*, Vol. 7, No. 1, pp. 67–89, 1991.
4. N. Oshiro, N. Maru, A. Nishikawa and F. Miyazaki, "Binocular Tracking using Log Polar Mapping", *IROS'96*, Vol. 2, pp. 791–798, 1996.
5. R. D. Rimey and C. M.Brown, "Controlling Eye Movements with Hidden Markov Models", *Int. Journal of Computer Vision*, Vol. 7, No. 1, pp. 47–65, 1991.
6. S. Rougeaux, N. Kita, Y. Kuniyoshi, S. Sakane and F. Chavand, "Binocular Tracking Based on Virtual Horopters", *IROS'94*, Vol. 3, pp. 2052–2057, 1994.
7. E. L. Schwartz, "Computational Anatomy and Functional Architecture of Striate Cortex: a Spatial Mapping Approach to Perceptual Coding", *Vision Research*, Vol. 20, pp. 645–669, 1980.
8. R. Wallace, R. W. Ong, B. Bederson and E. Schwartz, "Space Variant Image Processing", *Int. Journal of Computer Vision*, Vol. 13, No. 1, pp. 71–90, 1994.
9. H. Yamamoto, Y. Yeshurun and M. D. Levine, "An Active Foveated Vision System: Attentional Mechanisms and Scan Path Covergence Measures", *Computer Vision and Image Understanding*, Vol. 63, No. 1, pp. 50–65, 1996.
10. Y. Yeshurun, E. L. Schwartz, "Cepstral Filtering on a Columnar Image Architecture: A Fast Algorithm for Binocular Stereo Segmentation", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 11, No. 7, pp. 759–767, 1989.

This article was processed using the LaTeX macro package with LLNCS style