論文2002-39SP-1-7

로그폴라 변환과 ZDF를 이용한 이동 물체 추적 스테레오 시스템

(Stereo System for Tracking Moving Object using Log-Polar Transformation and ZDF)

尹 鍾 建 *, 崔 一 **, 李 容 範 ***, 秦 成 一 ****

(Jong-Kun Yoon, Il Choi, Yong-Bum Lee, and Sung-Il Chien)

요 약

능동 스테레오 비젼 시스템은 목표물 판별을 위해서 간단한 연산만으로 작은 시차값을 가지는 특징들만을 통과시킴으로써 목표물 위치 측정을 가능케 한다. 그런데 이 간단한 방법은, 복잡한 배경이 포함되거나 시차가 영인 영역에 다른 물체들도 동시에 존재하면 적용이 어렵게 된다. 이러한 문제점들을 해결하기 위하여, 본 논문에서는 시야 중심 영역의 해상도는 높이고 일반적으로 중요도가 낮은 주변의 해상도는 감소시키는 포비에이션 기법을 필터링 방법과 결합시켰다. 포비에이트 영상 표현을 위하여 영상 피라미드 또는 로그폴라 변환을 사용하였다. 또한 추적이 수행되는 동안에 스테레오 시차가 작은 값이 유지되도록 투영을 이용하여 스테레오 시차를 추출하였다. 실험 결과는 로그폴라 변환이 복잡한 배경으로부터 목표물을 분리하는 제안된 방법이 영상 피라미드 또는 기존 방법보다 우수하며, 추적성능을 상당히 개선함을 보여주고 있다.

Abstract

Active stereo vision system allows us to localize a target object by passing only the features of small disparities without heavy computation for identifying the target. This simple method, however, is not applicable to the situations where a distracting background is included or the target and other objects are located on the zero disparity area simultaneously. To alleviate these problems, we combined filtering with foveation which employs high resolution in the center of the visual field and suppresses the periphery which is usually less interesting. We adopted an image pyramid or log-polar transformation for foveated imaging representation. We also extracted the stereo disparity of the target by using projection to keep the stereo disparity small during tracking. Our experiments show that log-polar transformation is superior to either an image pyramid or traditional method in separating a target from the distracting background and fairly enhances the tracking performance.

* 正會員, LG電子
(LG Electronics)
** 正會員, 龜尾1大學 精報通新傳工
(Information & Communications, Gumi College)
*** 正會員, (株)후후
(HUHU Co. Ltd.)
**** 正會員, 慶北大學校 電子電氣工學部
(School of Electronic and Electrical Engineering Kyungpook National University)
接受日字:2000年6月21日, 수정완료일:2001年10月19日

I. Introduction

It was claimed that visually guided behavior can be facilitated by utilizing visual cues commonly found in biological visual system[1,2]. For example, if our eyes (stereo camera) are verged to common fixation, fixation target lies near the horopter and points near by tends to have small disparity. Horopter is the set of world points whose disparity is zero, passing through the two nodal points of

stereo camera and the fixation point as shown in Fig. 1.

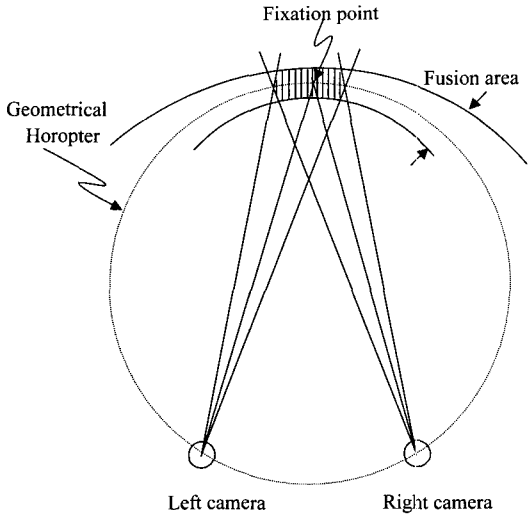


그림 1. 호롭터
Fig. 1. Horopter where fixation target is located with zero disparity.

Objects that stay in this horopter can be easily extracted by suppressing features at large disparities. This is the principle of zero disparity filtering (ZDF) and can be implemented via logical AND of stereo vertical-edge images[3]. We can estimate the location of target by computing center of gravity mass of ZDF output image. However, ZDF may produce false outputs in a complex scene due to AND results of two different objects. Tanaka et al.[4] proposed disparity-based segmentation method to track an object that is moving in a complicated scene. However, they assume only one object is located on the common space of both cameras field of view and the horopter. Target windowing can be a solution[5]. Fixed-size window deteriorates tracking performance when the size of a target varies, and thus target window needs to be resized adaptively.

Usually, the central area is our concern. If we locate high-resolution portion in the center of the visual field, the target object can be dominant over background. So, the target can be effectively separated from the distracting background without considering target windowing. Oshiro et al.[6] used log-polar mapping in ZDF to spread the zero disparity area from the gaze point towards the periphery so that a target in the periphery could be found. In our case, we map each pixel exclusively in the log-polar transformation[7~10] especially intended for target enhancement over the background while avoiding target windowing, rather than large detectable range of disparities of the target object. Besides log-polar transformation multiresolution image pyramid[11,12] was also considered for foveation and the ZDF outputs of each method was compared in this paper. Both have some important differences, however, share high central resolution features that are the main reason for their adoption in this paper.

As the centroid of ZDF output gives only rough measurement of a target location, the stereo disparity should be extracted and then cancelled to fixate on the target object with the stereo camera. Various disparity extraction methods have been introduced in binocular tracking, cepstrum operator[13], virtual horopter[5], phase-based approach[14], and correlation[4]. While most of the above-mentioned require heavy computation or arbitrary shifting, the projection of reference edge images we used can assure correct steering direction and stereo disparity of the target with less computation cost. The proposed stereo tracking algorithm is implemented on our active vision system and the experimental results are demonstrated by actual tracking experiments.

Ⅱ. TARGET EXTRACTION FROM THE BACKGROUND

In the following, we describe both log-polar transformation and foveal image generation using an image pyramid and show that ZDF with log-polar transformation is very efficient in target extraction compared to either traditional ZDF or ZDF with an image pyramid.

1. Log-Polar Transformation

Log-Polar transformation is a space-variant

sampling method that decreases sampling density from the center to the periphery. A point in the Cartesian coordinates (x, y) can be mapped to the log-polar coordinates (u, v) using the following equation.

$$(u, v) = (\log_k r, \theta) \tag{1}$$

where r and θ are given by $r=\sqrt{x^2+y^2}$, $\theta = \tan^{-1}\frac{y}{x}$. Here, k is scaling factor determined by the log-polar image size. Fig. 2 shows a log-polar mapping between the Cartesian coordinates and the log-polar coordinates.

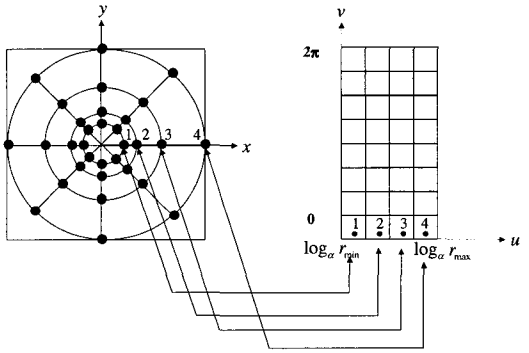


그림 2. 로그폴라 변환
Fig. 2. Cartesian coordinates and corresponding log-polar coordinates from space-variant sampling.

A log-polar image whose size is $E \times A$ pixels can be generated as follows. E and A represent the number of rings and that of the cells per ring, respectively. Let the minimum and maximum radius be $\rho_{\min}$ and $\rho_{\max}$. The u and v are defined in the following domain:

$$u \in [\log_k \rho_{\min}, \log_k \rho_{\max}]$$
$$v \in [0, 2\pi] \tag{2}$$

In the u direction, there exist pixels.

$$E = \log_k \rho_{\max} - \log_k \rho_{\min} \tag{3}$$
$$\ln k = \ln(\rho_{\max}/\rho_{\min})/E \tag{4}$$

Using (1) and (4), we can map the Cartesian coordinatesinto the corresponding u. Also we can obtain v direction value, considering that the spacing angle of each cell is $2\pi/A$. From above, we obtain the following relations.

$$u = \frac{E}{\ln(\rho_{\max}/\rho_{\min})} \times \ln\sqrt{x^2+y^2} \tag{5}$$
$$v = \tan^{-1}(y/x) \times A/2\pi \tag{6}$$

Inverse mapping leading to the Cartesian coordinates is given by

$$x = \exp\left[\frac{\ln(\rho_{\max}/\rho_{\min})}{E} \times u\right]\cos(v \times 2\pi/A) \tag{7}$$
$$y = \exp\left[\frac{\ln(\rho_{\max}/\rho_{\min})}{E} \times u\right]\sin(v \times 2\pi/A) \tag{8}$$

Fig. 3 shows the space-variant images from the log-polar transformation.

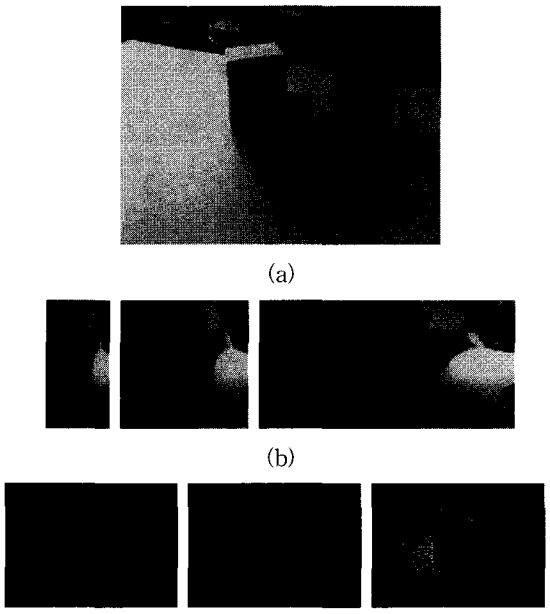


그림 3. 로그폴라 변환 예. (a) 원영상: 160×120. (b) 로그폴라 영상: 좌로부터, 32×64, 64×64, and 128×64. (c) 역변환된 로그폴라 영상.
Fig. 3. Example of log-polar transformation and inverse mapping. (a) Original image: 160×120. (b) Log-Polar images with various image size: from the left, 32×64, 64×64, and 128×64. (c) Images after inverse mapping.

The difference of resolution shown in Fig. 3c

originates from that the distance between the concentric ring is determined by the log-polar image size which imposes a constraint on the value of the scaling factor. Here, we can see that log-polar transformation keeps shape of an object that is currently being fixated mostly unaffected by the size of log-polar image.

2. Image Pyramid

The generation of foveal image using the image pyramid is well explained in[12]. They extracted equal-sized central image patches from the each level of the image pyramid, performed successive expanding, smoothing, and inserting the finer level of the patch into the center of the present level. Fig. 4 shows the foveal images generated from the four level in the image pyramid.

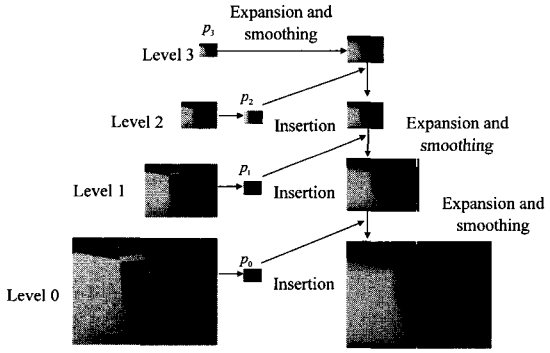


그림 4. 영상 피라미드를 이용한 포비에이션 예
Fig. 4. Foveal image pyramid from successive low pass filtering and sub-sampling.

Here, we deal with the ZDF outputs from log-polar transformation and the image pyramid, and traditional method. Firstly, in case of log-polar transformation, vertical edge operation is performed for the left and right stereo images and those edge images are transformed to the log-polar coordinates. We then binarize both log-polar images by an appropriate threshold value. Then we obtain log-polar ZDF output by performing logical AND of both thresholded log-polar images. ZDF output by the image pyramid is obtained by applying the traditional ZDF operation to the foveal image. ZDF outputs are shown in Fig. 5.

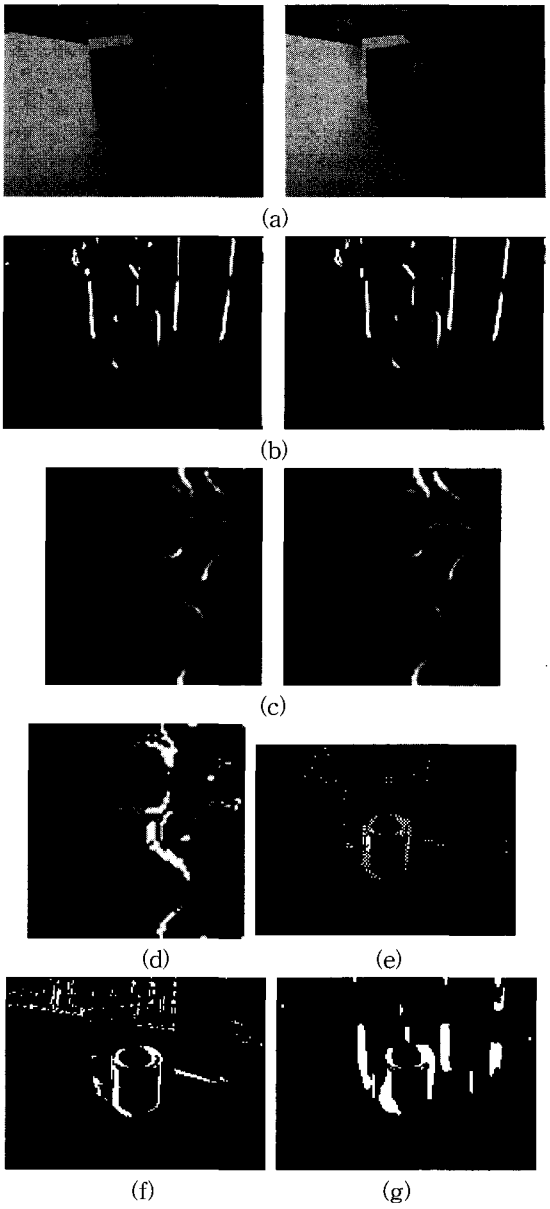


그림 5. 관측물체 추출 (a) 좌우 영상. (b) 수직 에지 영상. (c) 로그폴라 영상. (d) 로그폴라 ZDF 영상. (e) ZDF와 로그폴라 변환을 이용하여 추출한 관측물체. (f) ZDF만을 이용하여 추출한 관측물체. (g) ZDF와 영상 피라미드를 이용하여 추출한 관측 물체
Fig. 5. ZDF outputs. (a) Original images. (b) Edge images. (c) Log-Polar images. (d) Log-Polar ZDF. (e) Inverse log-polar ZDF. (f) Traditional ZDF output. (g) Image pyramid ZDF.

We can observe that log-polar transformation is quite effective, whereas traditional ZDF produces background outputs and the image pyramid ZDF shows strong vertical edges due to blurring. Fig. 6 surely shows its superiority over them.

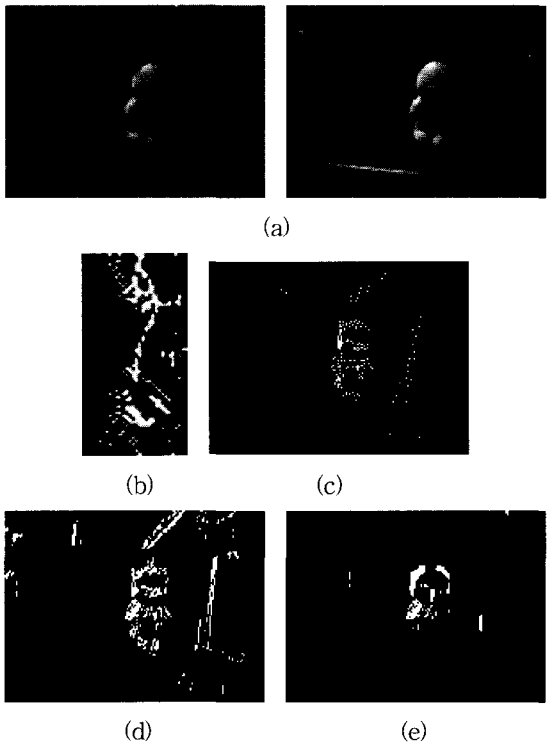


Fig. 6. ZDF outputs. (a) Original images. (b) Log-Polar ZDF. (c) Inverse log-polar ZDF. (d) Traditional ZDF output. (e) Image pyramid ZDF.

그림 6. 관측물체 추출. (a) 좌우 영상. (b) 로그폴라 ZDF 영상. (c) ZDF와 로그폴라 변환을 이용하여 추출한 관측물체. (d) ZDF만을 이용하여 추출한 관측물체. (e) ZDF와 영상 피라미드를 이용하여 추출한 관측 물체.

Foveation by the image pyramid can suppress the background, but part of the target object is missing as shown in Fig. 6e. This implies that the image pyramid structure is not suitable for representing more exactly the central region of gaze compared to space-variant resolution structure. The location of target is calculated by reconverting log-polar ZDF output to the Cartesian image and computing its centroid. Due to subsampling in log-polar processing, the centroid tends to be squeezed toward the image center. When the active camera is controlled in a closed loop, the controller will try to minimize the error every frame so that this problem does not usually cause much problem.

Ⅲ. STEREO DISPARITY EXTRACTION

Let the left and right projections beand $p_l(i)$ and $p_r(i)$ and the centroid of ZDF output be (x_f, y_f). SSD is used for disparity extraction. The sum of squared difference (SSD) of the intensity values is the simplest and most effective criterion in measuring similarity between images[15]. The SSD of each searching point k in a search area s is defined as

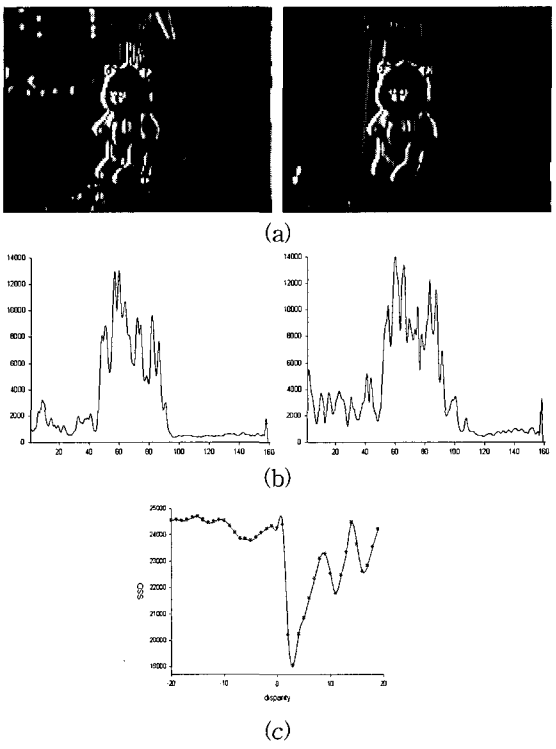


그림 7. 스테레오 시차 추출. (a) 수직 에지 영상. (b) 투영 결과. (c) 제곱차합 값

Fig. 7. Projection of stereo image and extraction of stereo disparity from the lowest SSD value. (a) Edge images. (b) Projection results. (c) SSD values.

$$e(k)=\sum_{i=-r/2}^{i=r/2}[P+l(x_f+i)-P_r(x_f+i+k)]^2 \quad (9)$$

where $k=-s/2,\cdots,s/2$. Here, r is the size of windowing for computing SSD. Fig. 7 shows the projection data and their SSD results.

Stereo disparity d of a target can be determined by finding a searching point whose SSD is minimum.

$$d=\arg\min[e(k)] \quad (10)$$

After the stereo disparity is calculated, the desired values for controlling active stereo system are determined so that the target is located in the image center with the small stereo disparity. The entire tracking algorithm is described in Fig. 8.

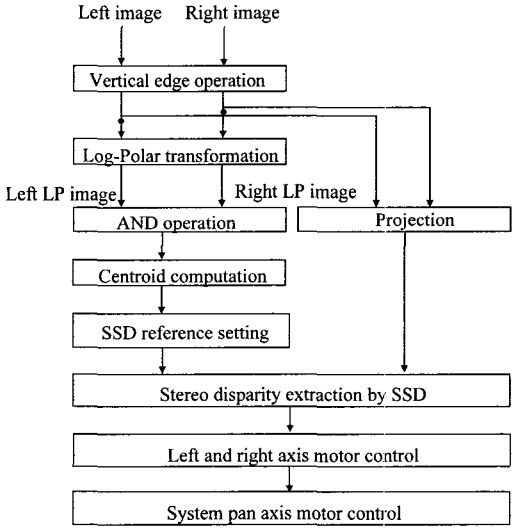


그림. 8. 제안된 스테레오 추적 알고리듬의 데이터 흐름도
Fig. 8. Dataflow of proposed stereo tracking algorithm.

Ⅳ. ACTIVE STEREO TRACKING SYSTEM

Our head/eye platform shown in Fig. 9 has 5 degrees of freedom (DOF) and is equipped with DC motors and harmonic driver minimizing the backlash.

The maximal pan/tilt angle of each axis is 360°. The focal length of the camera is 25mm and the

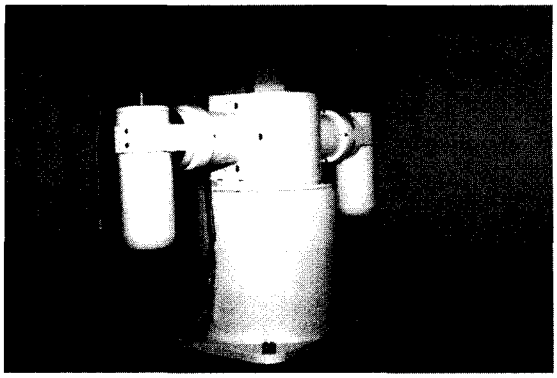


그림 9. 능동 스테레오 추적 시스템
Fig. 9. Active stereo tracking system.

baseline length is 470mm. The stereo images from the two CCD cameras are digitized into the frame memory by the PCI Matrox Meteor image board. Fig. 10 shows the flow of data to control active stereo system.

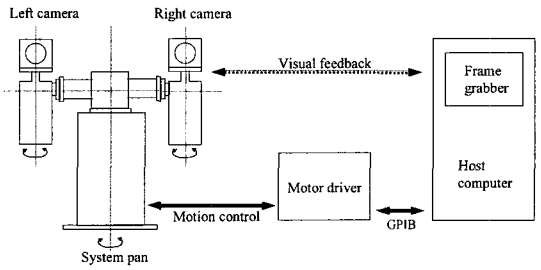


그림 10. 스테레오 추적 시스템 구성
Fig. 10. Configuration of the stereo tracking system.

The servo loop is controlled by a classic proportional control and communication between host and active stereo system for camera control is implemented via GPIB interface. Symmetric pan control is implemented to put the stereo disparity to zero since the positioning mechanism uses an additional pan control system that tries to locate a target in the image center.

Ⅴ. EXPERIMENT RESULTS

The size of input image is 160×120 pixels and transformed into 32×64 log-polar images. We employed a bear doll rotating round the tripod in the

experiment. On top of the tripod, a step motor was equipped and rotated a bear doll at 1.2 cm/sec. Stereo cameras was initially pointed at a target 1.5m away from the bear doll. The tracking environment is shown in Fig. 11 and the tracking performance was evaluated from the difference between actual camera position and target location.

Figs. 12a and 12b shows the tracking responses of three methods, log-polar ZDF, image pyramid ZDF, and traditional ZDF. Sinusoidal response shown in Fig. 12a is the rotation angle of the left camera and corresponds to the camera trajectory following the moving target. The rotation angle of the right camera is the same and only different in sign due to symmetric rotation. In case of the image pyramid ZDF and traditional ZDF, the tracking system began to deviate from the desired trajectory at 75th frame. The reason for this is directly traced to the situation that the doll and tripod are located on the same horopter. Still, the log-polar transformation kept the doll dominant over the distracting background so that the tracking was successfully pursued. Occlusion was occurred at around 90th frame, in part, and led to system fixation on random position. However, its effects were slight in our case and the tripod lying in the zero disparity area was the main cause of deviations. The responses of system pan axis were shown in Fig. 12c and the performance found to be

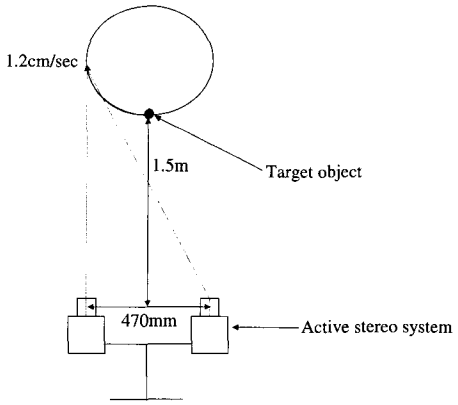
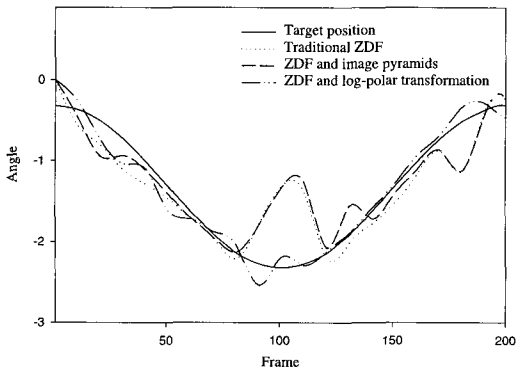
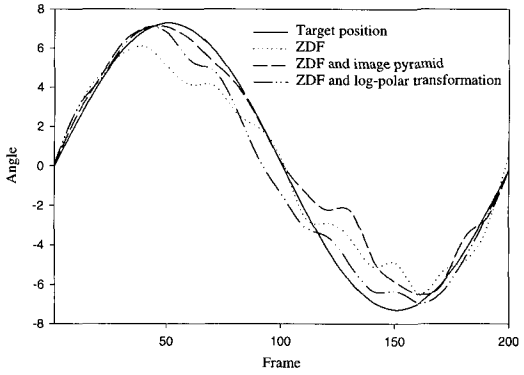


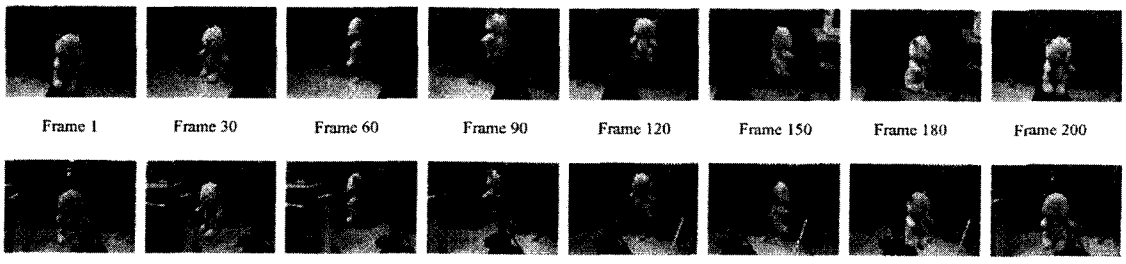
그림 11. 추적 실험 환경 구축
Fig. 11. Experiment setup for tracking



(a)



(b)



(c)

그림 12. 회전 곰인형에 대한 추적 결과. (a) 좌 카메라의 회전각. (b) 시스템 팬 모터의 회전각. (c) 곰인형 추적
Fig. 12. Tracking results for a rotating bear doll. (a) Rotation angle of the left camera. (b) Rotation angle of the system pan motor. (c) Bear doll tracking.

similar to Fig. 12a since its rotation angle is linked with stereo camera movement. Fig. 12a shows the sequences of frames obtained from the log-polar ZDF during tracking. It can be seen that the doll was successfully tracked since it was located in the image center, keeping small disparity. From the experiments we have done, it is evident that log-polar transformation can enhance the tracking performance fairly, providing successful tracking.

VI. CONCLUSIONS

We proposed foveation that was coupled to traditional ZDF for effective target extraction from the distracting background. We chose log-polar transformation and an image pyramid for target enhancement and compared their performances through actual tracking experiment. We have demonstrated that foveation by log-polar transformation is more adequate in tracking mainly *due to effective background suppression* without windowing the target. In addition, we extracted stereo disparity using projection data to fixate on a surface of a moving target and proved its accuracy from the experiment. In fact, our closed-loop controller was somewhat coarse though active stereo system was capable of track a moving target. In the future we plan to design and incorporate reliable PID controller for smoother behavior of active stereo system.

REFERENCES

[1] D. J. Coombs, "Tracking Objects with Eye Movements," *Proc. of the Topical Meeting on Image Understanding and Machine Vision*, Optical Society of America, June 1989.

[2] D. H. Ballard and C. M. Brown, "Principles of Animate Vision," *CVGIP: Image Understanding*, Vol. 56, No. 1, pp. 3-21, 1992.

[3] D. J. Coombs and C. M. Brown, "Real-Time Binocular Smooth Pursuit," *International Journal of Computer Vision*, Vol. 11, No. 2, pp. 147-164, 1993.

[4] M. Tanaka, N. Maru, and F. Miyazaki, "Binocular Gaze Holding of a Moving Object with the Active Stereo Vision System," *Proc. the 2nd IEEE Workshop on Applications of Computer Vision*, pp. 250-255, Sarasota, Florida, USA, 1994.

[5] 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.

[6] N. Oshiro, N. Maru, A. Nishikawa, and F. Miyazaki, "Binocular Tracking using Log Polar Mapping," *IROS'96*, Vol. 2, pp. 791-798, November 1996.

[7] M. Tistarelli and G. Sandini, "Dynamic Aspects in Active Vision," *CVGIP: Image Understanding*, Vol. 56, No. 1, pp. 108-129, 1992.

[8] A. Bernardino and J. Santos-Victor, "Visual Behaviours for Binocular Tracking," *2nd EuroMicro Workshop on Advanced Mobile Robotics-Eurobot97*, Brescia, Italy, October 1997.

[9] R. A. Messner and H. H. Szu, "An Image Processing Architecture for Real Time Generation of Scale and Rotation Invariant Patterns," *Computer Vision, Graphics, and Image Processing*, Vol. 31, No. 1, pp. 50-66, July 1985.

[10] R. S. Wallace, P. Ong, B. B. Bederson, and E. L. Schwartz, "Space Variant Image Processing," *International Journal of Computer Vision*, Vol. 13, No. 1, pp. 71-90, 1994.

[11] T. J. Olson and R. J. Lockwood, "Fixation-based Filtering," *Proc. of the SPIE Intelligent Robots and Computer Vision XI Conference*, pp. 685-696, 1992.

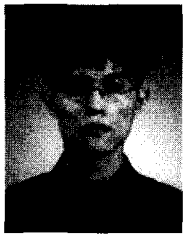
[12] J. R. Taylor and T. J. Olson, "Precise vergence control in complex scenes," *SPIE Vol. 2056 Intelligent Robots and Computer Vision XII*, pp. 20-30, 1993.

[13] T. J. Olson and D. J. Coombs, "Real-Time Vergence Control for Binocular Robots," *International Journal of Computer Vision*, Vol. 7, No. 1, pp. 67-89, 1991.

[14] A. Maki and T. Uhlin, "Disparity selection in binocular pursuit," *Technical report, KTH (Royal Institute of Technology)*, 1995.

[15] M. Okutomi and T. Kanade, "A Multiple-Baseline Stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 4, pp. 353-363, April 1993.

저 자 소 개



尹 鍾 建(正會員)

1968年生. 1995年 2月 경북대학교 전자공학과 졸업(공학사). 1995年 6月~1996年 12月 한국원자력연구소 연구원. 2000年 2月 경북대학교 대학원 전자공학과 졸업(공학석사). 2000年 2月~현재 LG전자 주임연구원. <관심분야> Computer vision

崔 一(正會員)

1964年生. 1986年 2月 경북대학교 전자공학과 졸업(공학사). 1988年 2月 경북대학교 대학원 전자공학과 졸업(공학석사). 2001年 2月 경북대학교 대학원 전자공학과 졸업(공학박사). 1988年 2月~1994年 12月 국방과학연구소 연구원. 1996年 3月~1997年 8月 한국천문연구원 선임연구원. 1998年 3月~1999年 8月 경운대학교 전산공학과 전임강사. 2001年 3月~현재 구미1대학 정보통신전공 전임강사. <관심분야> Computer vision, Pattern recognition, Image processing, Optimal control

李 容 範(正會員)

1962年生. 1985年 2月 경북대학교 전자공학과 졸업(공학사). 1987年 2月 경북대학교 대학원 전자공학과 졸업(공학석사). 1999年 2月 경북대학교 대학원 전자공학과 졸업(공학박사). 1987年~2000年 한국원자력연구소 책임연구원. 2000年~현재 (주)후후 대표이사, 신지식인협회 부회장. <관심분야> Stereo vision, Underwater stereo camera, Robotics, Virtual reality

秦 成 一(正會員)

1955年生. 1977年 2月 서울대학교 전자공학과 졸업(공학사). 1981年 2月 한국과학기술원 전자공학과 졸업(공학석사). 1988年 2月 Carnegie Mellon University (Ph. D). 1977年 1978年 (주)대영전자공업 연구원. 1982年 현재 경북대학교 공과대학 교수. <관심분야> Computer vision, Pattern recognition, Image processing