

Implementation of 3D Moving Target-Tracking System based on MSE and BPEJTC Algorithms

Jung-Hwan Ko, Maeng-Ho Lee, and Eun-Soo Kim*

Abstract

In this paper, a new stereo 3D moving-target tracking system using the MSE (mean square error) and BPEJTC (binary phase extraction joint transform correlator) algorithms is proposed. A moving target is extracted from the sequential input stereo image by applying a region-based MSE algorithm following which, the location coordinates of a moving target in each frame are obtained through correlation between the extracted target image and the input stereo image by using the BPEJTC algorithm. Through several experiments performed with 20 frames of the stereo image pair with 640 x 480 pixels, we confirmed that the proposed system is capable of tracking a moving target at a relatively low error ratio of 1.29 % on average at real time.

Keywords : stereo target tracking, 3D robot vision, pan/tilt system, correlation.

1. Introduction

Through the 3D vision system, three-dimensional images can be perceived as clearly as the human eyes, by using the binocular disparity [1] between the left and right eyes. In this system, if the viewing-point of the target object is not in accord with the focus points of the left and right eyes, it may cause fatigue to the eyes because the target object would be seen as two different objects. Therefore, if there exists stereo disparity in the target object, it should be removed. This process is known as convergence angle control [2]. In general stereo vision and tracking system should control the convergence angle of the camera system according to the distance of object just like as the apple of eye do. That is, by controlling the convergence angle, a stereo disparity between the left and right images is maintained at zero. In addition, the stereo vision and tracking system also requires the ability to pan/tilt control, by which a moving target is always kept at the center of the camera's field of view (FOV).

In general, removing the background noises from the relationship between the previous image and the present image through 2-D sequential input images has been carried out to extract a tracking object in the conventional image processing. There are several methods for extracting a tracking object from a sequential input image: image difference method, model-based method, optical flow method, and block matching method, etc. However, it is very difficult to extract target objects from a moving teleworking system or a moving object tracking system, because normally, there exist complex and changes in the background of an input image. In particular, there are difficulties in coping adaptively with the changes of surrounding environment and conditions.

Therefore, in this paper, a new 3D target-tracking system based on the MSE (minimum square error) and BPEJTC (binary phase extraction joint transform correlator) algorithms are proposed. In the newly proposed system, the target object is adaptively extracted from the input stereo image by using the region-based MSE algorithm and then, the pan/tilt and convergence angle of the stereo camera system is controlled at real time by using these location coordinates of the moving target, which are obtained by applying a correlation between the extracted target object and the input image.

Some experiments were conducted on real-time tracking of a moving target of 'car' using the proposed system. This was carried out under the situation where

Manuscript received January 6, 2004; accepted for publication March 8, 2004.
* Member, KIDS
Corresponding Author : Jung-Hwan Ko
3D Display Reserch Center, National Research Lab. of 3D Media, Dept. of Electronic Eng., Kwangwoon University 447-1 Wolge-Dong, Nowon-Gu, Seoul 139-701, Korea.
E-mail : misagi@explore.kw.ac.kr **Tel** : +2 940-5118 **Fax** : +2 941-5979

there are background noises and two other dummy moving cars with the 20 frames of the input stereo image pairs.

2. Convergence Control by Pan/tilt System Embedded-on Stereo Camera

In the proposed method, a moving-target object can be detected and extracted from the sequential stereo input images by using the MSE algorithm, in which a region-based window mask is employed [5]. That is, by applying the MSE algorithm to the left image of a stereo image pair of *t-1* frame, a window-masked target object of $I_{t-1}(x,y)$ is obtained, which is used as a reference image for obtaining a relative moving-distance of a target object between the *t-1* and *t* frames. Then, the degrees of similarity between this reference images of *t-1* and the left and right images of *t* frame are measured by computing the MSE value on each pixel location using Eq. (1), where the left image of *t* frame, I_{tl} is employed for example.

$$MSE(a,b) = \frac{1}{N_x \cdot N_y} \sum_{y=0}^{(N_y-1)} \sum_{x=0}^{(N_x-1)} \left| I_{t-1}(x,y) - I_{tl}(a+x, b+y) \right|^2 \quad (1)$$

Where, N_x , N_y and $I_{t-1}(x,y)$, $I_{tl}(a+x,b+y)$ mean the matching block sizes in the *x* and *y* directions and the reference image of *t-1* frame and the left image of *t* frame in the coordinates of (x, y) and (*a*+*x*, *b*+y) in which *a* and *b* are given by $a = -2N_x \sim 2N_x$, $b = -N_y \sim N_y$, respectively. Also, in the image processing of the moving object, the shifting component for the horizontal axis (x-axis) is shown much bigger than the one for the vertical axis (y-axis). Thus, the shifting component for the vertical axis can be assumed to be very small. Therefore, when the size for the window mask is N_x and N_y, the MSE applying domain is restricted for x-axis to be $4N_x$ and y-axis to be $2N_y$.

$$I'_{t-1}(x,y) = \begin{cases} Window-masked\ object\ image & : minimum\ MSE\ value \\ 0 & : otherwise \end{cases} \quad (2)$$

Where $I_{t-1}'(x,y)$ represents the target object extracted from the left image by using the MSE algorithm and window-masking scheme.

The location coordinates of the target object, $(\Delta x_l, \Delta y_l)$ is then computed, and based on these values, the target object is moved to the centre of the image plane. This target image is then used as a reference image in the next process of joint-transform correlation.

In case of the right camera's tracking operation, the correlation between the reference target image of *t*-1 frame, $[I_{t-1}'(x,y)]$ and the right image of *t* frame, $[r_t(x,y)]$ is performed using the BPEJTC system after which the relative distance between the two objects can be calculated from the location where the correlation peak occurs, which is consequently used to control the pan/tilt system embedded-on the right camera for tracking a moving target [5].

In other words, the reference target image and the right image are given by Eq. (3) in the BPEJTC.

$$I_{t-1}'(x-\Delta x_l,\ y-[\Delta y_l+\frac{w}{2}]),\ r_t(x-\Delta x_r,\ y-[\Delta y_r-\frac{w}{2}]) \quad (3)$$

Where $2w$ is the height of the input display unit and $(\Delta x_l, \Delta y_l)$ and $(\Delta x_r, \Delta y_r)$ are the distance coordinates to the target from the screen centre of the $I_{t-1}'(x,y)$ and $r_t(x,y)$ images, respectively.

Equation (4) shows the correlation output of the BPEJTC system.

$$\begin{aligned} c_{BPEJTC}(x,y) &= \Im^{-1}\{E_{BPEJTC}(u,v)\} \\ &= [I'_{t-1}(x,y) \otimes r_t(x,y)] * \delta[x+(\Delta x_l-\Delta x_r),\ y+(\Delta y_l-\Delta y_r+w)] \\ &+ [r_t(x,y) \otimes I'_{t-1}(x,y)] * \delta[x-(\Delta x_l-\Delta x_r),\ y-(\Delta y_l-\Delta y_r+w)] \end{aligned} \quad (4)$$

This correlation peaks occurred on the correlation plane are detected on CCD2 and from the relative distance between the reference image and the right image, the peak location value can be calculated and expressed as Eq. (5).

$$x_{peak} = \pm(\Delta x_l - \Delta x_r), \quad y_{peak} = \pm(\Delta y_l - \Delta y_r + w) \quad (5)$$

Consequently, by using this correlation peak value, the location coordinates of a target object on the right image can be obtained as shown in Eq. (6).

$$\Delta x_r = \Delta x_l - x_{peak}, \quad \Delta y_r = \Delta y_l + w - y_{peak} \quad (6)$$

Through a USB (universal serial bus) interface, the location coordinates of the target object obtained from the left and right images are transferred to the pan/tilt system on which the stereo camera is embedded, theres coordinates are controlled based on there values.

Fig. 1 shows a target-tracking screen, in which the stereo camera's pan/tilt system is continuously controlled to ensure that the target object is placed at the centre of the camera's field of view (FOV) by using the relative moving distance values of the target between two consecutive frames.

The pan/tilt control system used in this paper consists of PC, USB control board, pan/tilt and PID controller. Here the PID controller is employed to control the motor of the pan/tilt system adaptively as shown in Eq. (7).

$$u(n) = kp * e(n) + ki\sum_{N=0}^{n} e(n) + kd[e(n') - e(n'-1)] \quad (7)$$

Also, its error compensation signal is given by Eq. (8).

$$u(n) = e(n)\cdot(k_p + k_i + k_d) + e(n-1)\cdot(k_i - 2k_d) + e(n-2)\cdot(k_i + k_d) + k_i\cdot(e(n-3) + \cdots e(0)) \quad (8)$$

Each parameter used in Eqs. (7) and (8) is defined as follows.

$u(n)$ = output of motor control signal at sampling time n

$e(n)$ = position error at sampling time n

n' = derivative sampling rate

kp = proportional term

ki = integral term

kd = derivative term

il = integration limit

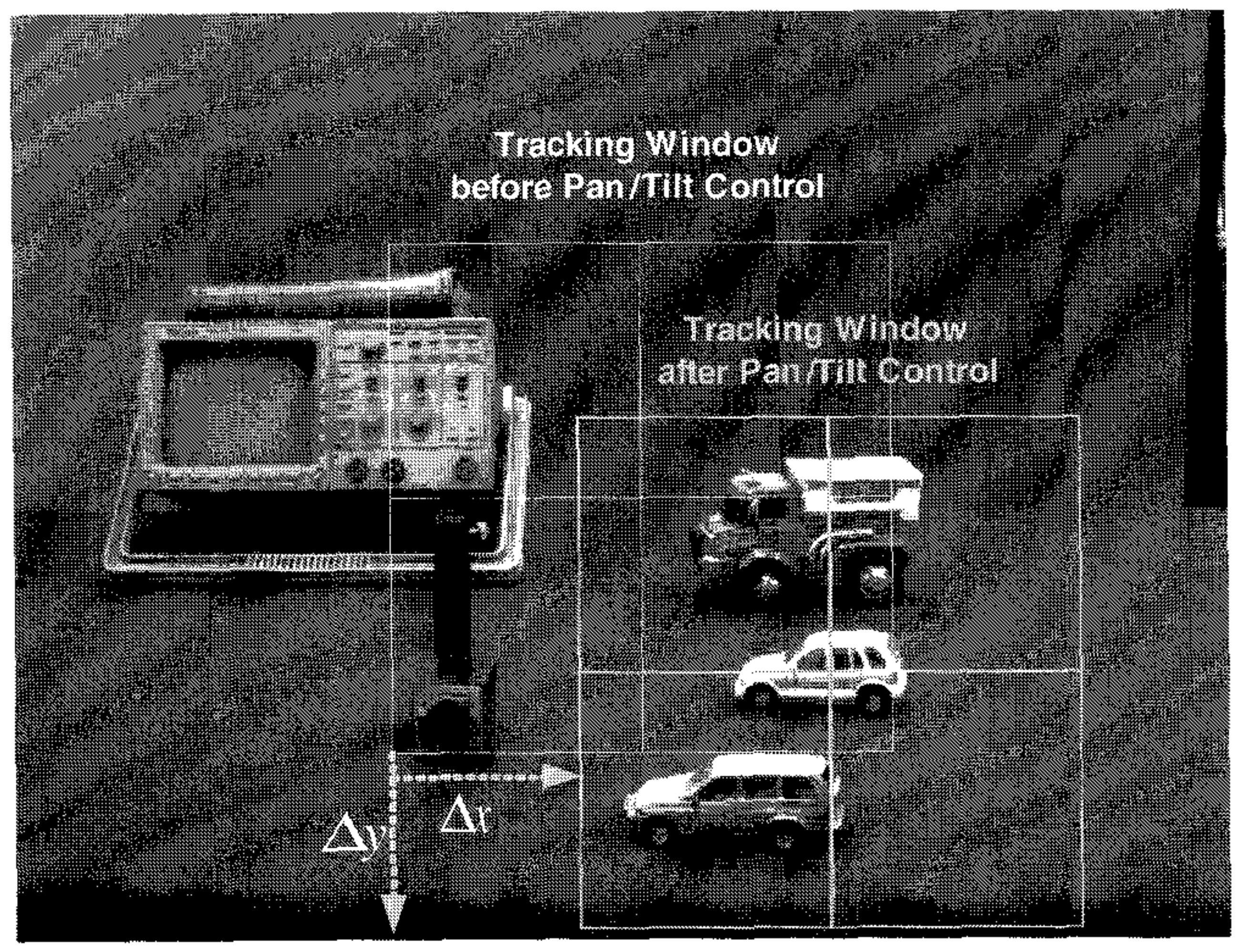


Fig. 1. Tracking screen by pan/tilt-embedded stereo camera system.

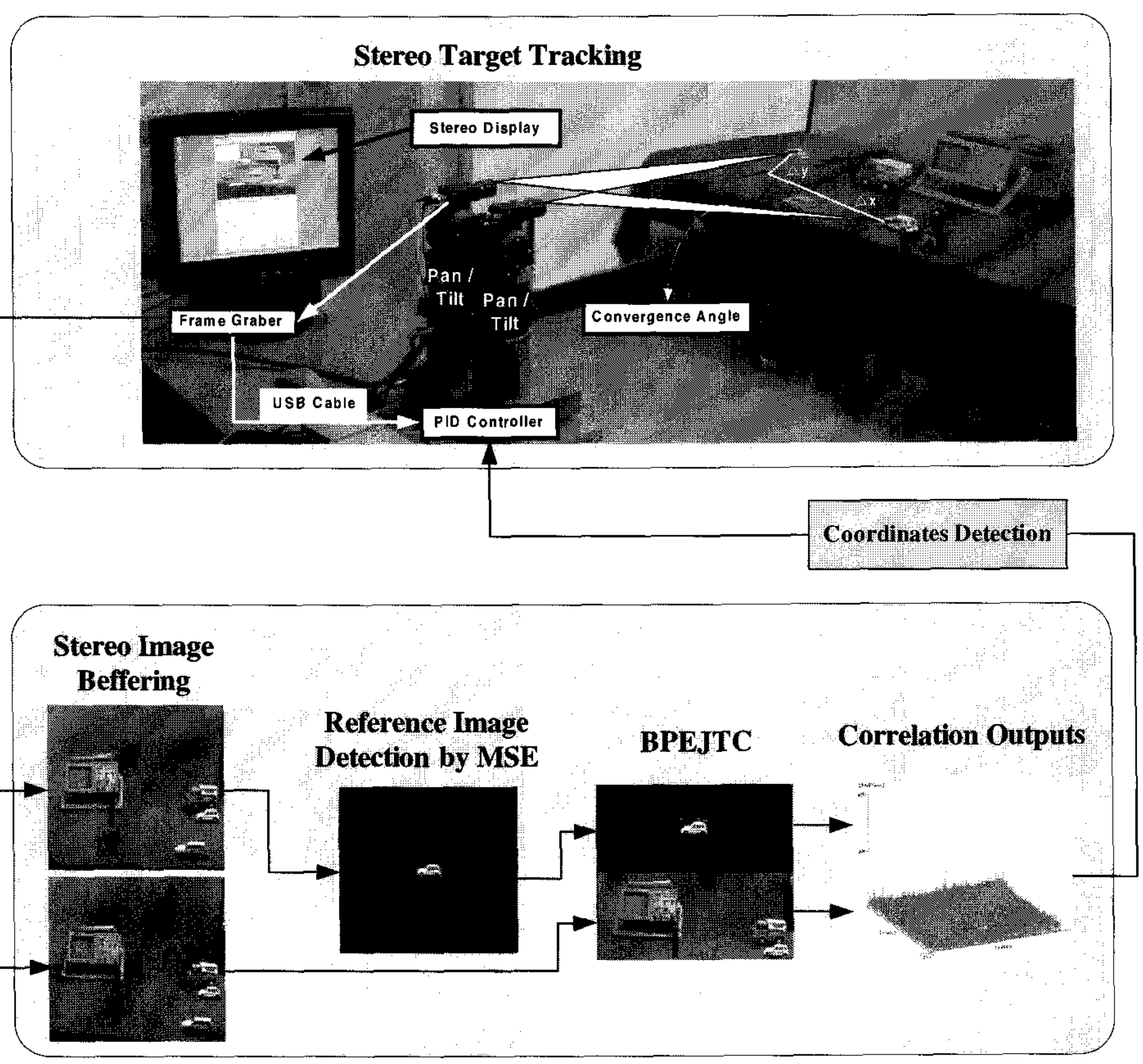


Fig. 2. Experimental set-up for the proposed stereo target-tracking system.

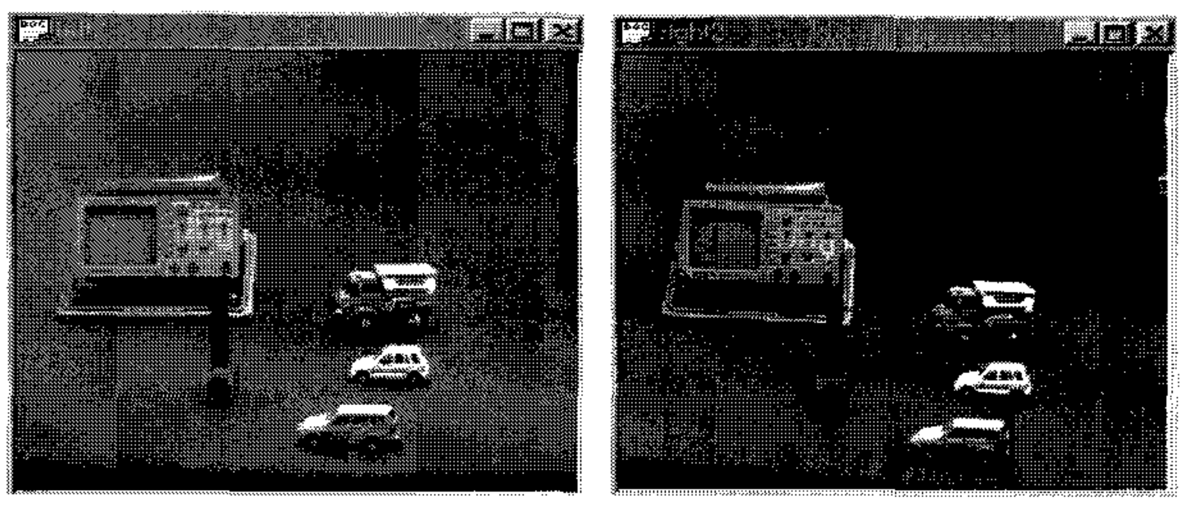
In Eq. (8), error of n^{th} can know how variable values of each kp, kd, ki influence to error compensation u(n) of system as time passes. K_p can know that influence present error, and influences about all errors of 0~n^{th} in the case of k_i, and influences to 2 points below n^{th} in the case of k_i. Therefore, point to keep in mind when adjust each PID variable must control considering present acceleration and velocity of motor, inertia of motor and sampling interval in system composed not that the overshoot rate is decreased even though raise k_d rashly. However, inertial side can ignore because can ignore inertia of motor in case of motor that have high gear ratio usually.

In this paper, we composed pan/tilt system such that can input PID control variable and do PID Tuning, considering effect that each PID variable about error of $n, n-1, n-2$ times get to $u(n)$ as Eq. (8).

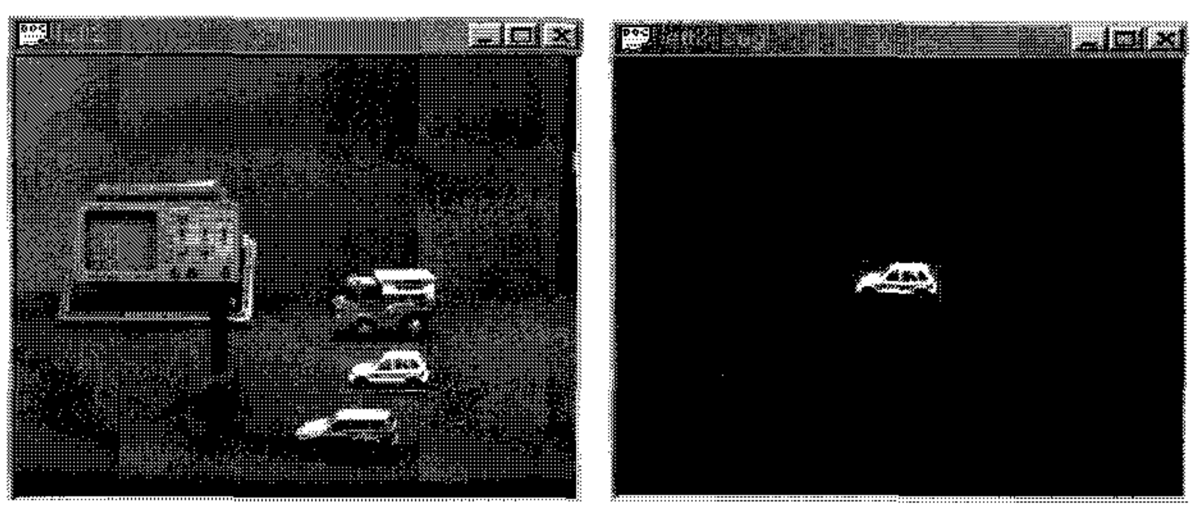
3. Experiments on StereoTarget-tracking and Results

Fig. 2 shows the experimental set-up for the proposed stereo object tracking system. In the experiment, 20 frames of the input stereo image pair having a resolution of 640 x 480 pixels were used. The stereo camera (CS-82393 BS, Dong-Kyung Electronics) and the pan/tilt control system (HWR-PT1, Hanwool Robotics) were also used in the experiment.

As shown in Fig. 2, the sequential input stereo image pairs are caught up by using the stereo camera embedded on the pan/tilt system and transferred to the host computer through the frame grabbers (Metero II/4 & Metero II MC/2). In the host computer, the location coordinates of a target object in each frame were detected by executing the proposed algorithm and then, transmitting to the pan/tilt control board, in which the feedback control signals for the

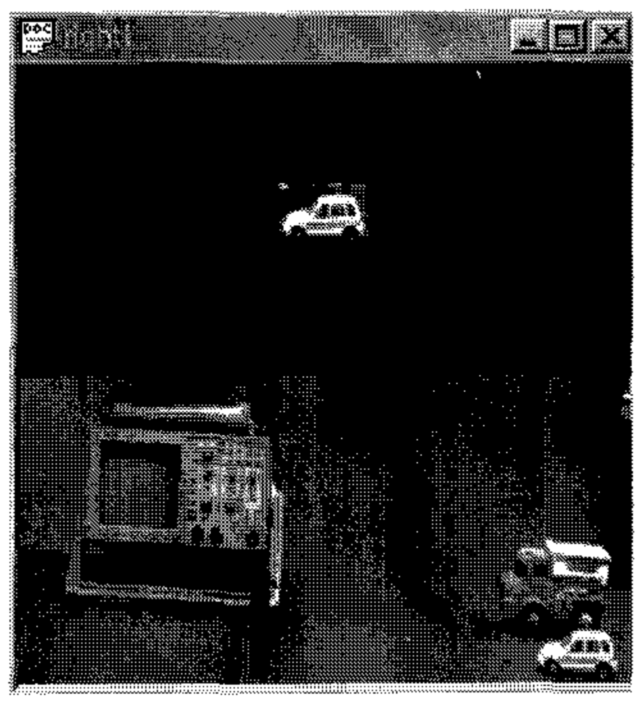


(a) Left image (b) Right image

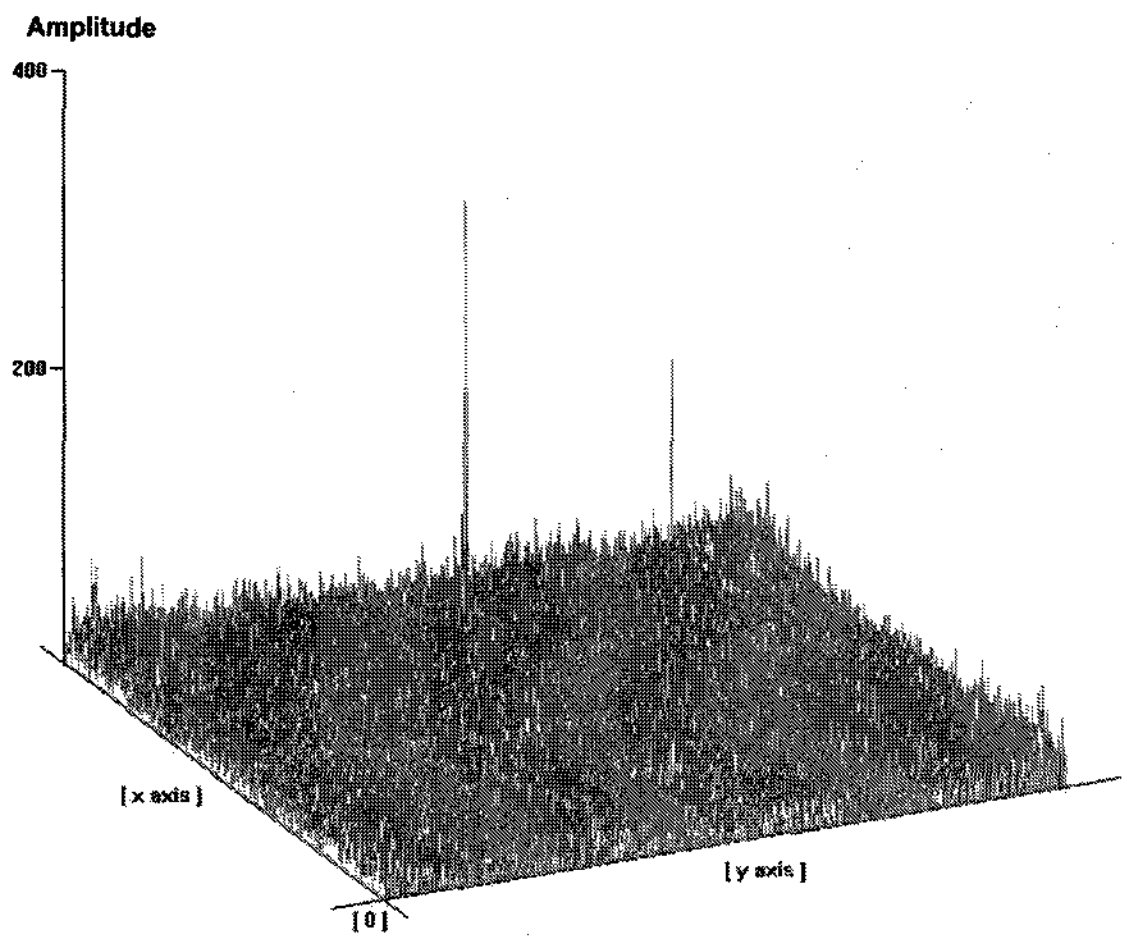


(c) Masked target by MSE (d) Reference target image

Fig. 3. Stereo input image pair and extracted target image by MSE.



(a) Correlation input plane



(b) Correlation peak point (Location's coordinate)

Fig. 4. Input plane of the BPEJTC and its correlation outputs.

Table 1. Extraction of location coordinates of the target object

[unit : pixel]

Frame	1		2		3		4	
	Left	Right	Left	Right	Left	Right	Left	Right
Coordinates	(28,-24)	(21,-22)	(103,-98)	(120,-116)	(-13,-90)	(-10,-94)	(-255,-105)	(-256,-109)

pan/tilt were generated using the micro controller (89C51). Finally, the pan/tilt was controlled by using these signals through the motor controller (LM629).

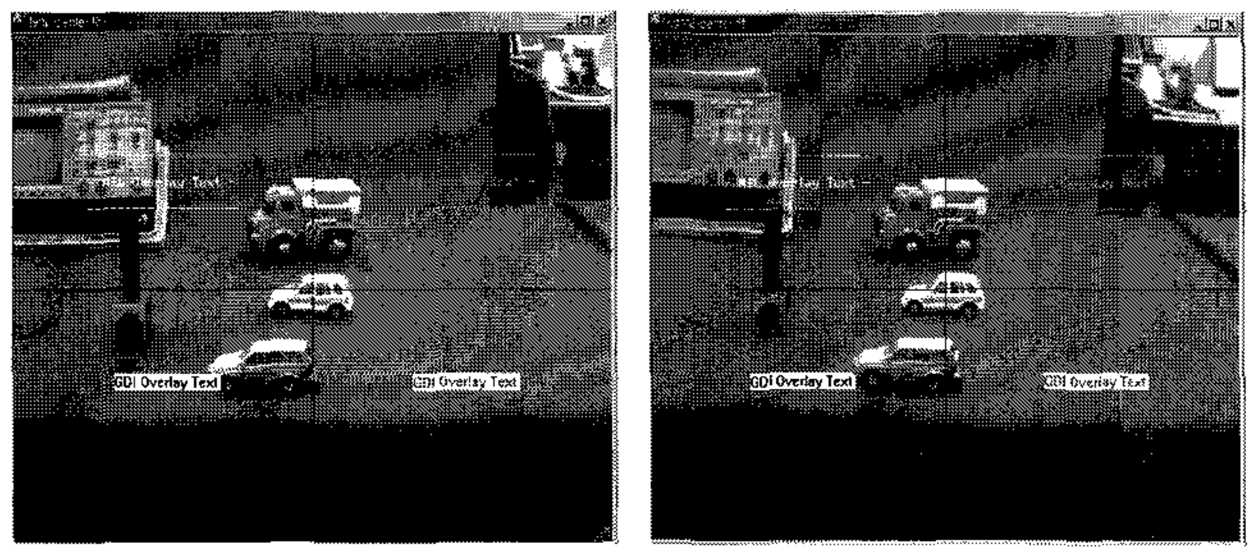
Fig. 3 shows a stereo image pair, in which a small car located between the two other dummy cars is the moving-target and the segmented target object with a window-mask and its shifted one to the center of the image plane.

Fig. 4 shows an input plane of the BPEJTC in which the reference target image of $l_{t-1}'(x,y)$ and the right image of $r_t(x,y)$ are simultaneously loaded on the upper and lower half plane, and its correlation outputs on the correlation plane. The moving distances, $(\pm\Delta x_l,\pm\Delta y_l)$ and $(\pm\Delta x_r,\pm\Delta y_r)$, from the centre of the image screen to the target object in the left and right images can be obtained by using these location coordinates of the correlation outputs, in which the centre coordinates of the input image is regarded as a reference coordinate of (0,0). These extracted position coordinates are then used to control the pan/tilt system embedded on the right camera.

Table 1 present the extracted location coordinates of a moving-target object in the left and right images of the four consecutive stereo input image pairs.

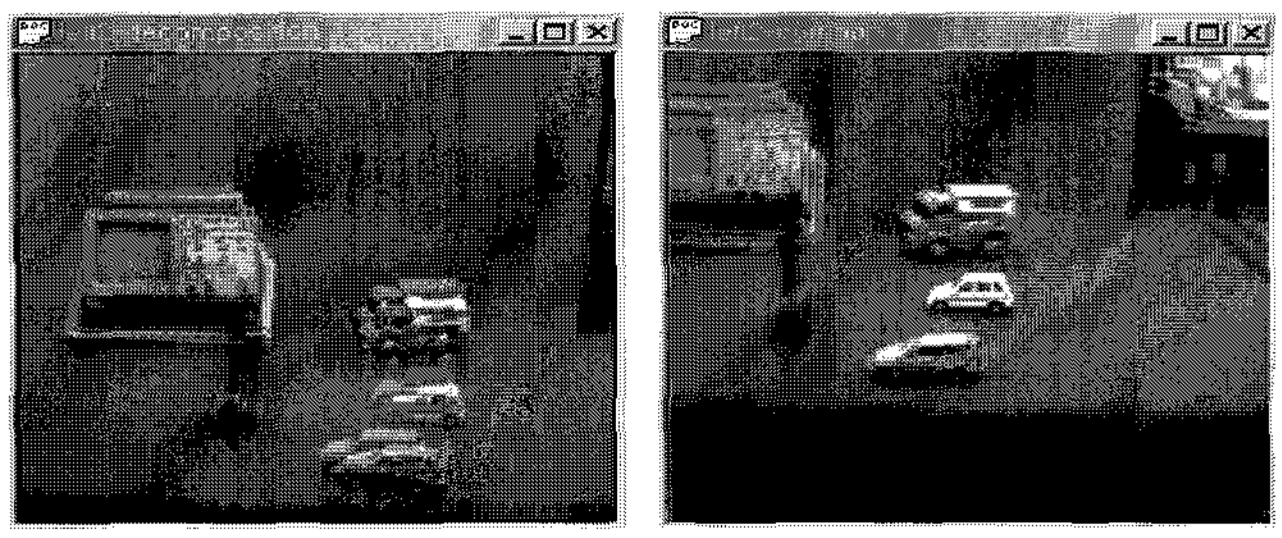
Fig. 5 shows the left and right images after being controlled by the pan/tilt system, using the moving distance values of the target object between two consecutive frames. From Fig. 5, we found that the target object, 'a small car' is always kept at the centre of the camera's field of view in the proposed stereo target tracking system, and also, the convergence points on the target object from the left and right cameras are controlled to be same.

Here, the target's moving distance values are converted into the corresponding degree values to effectively control the pan/tilt's system. Table 2 shows the calculated moving angles to the corresponding moving distances of the target object for four consecutive input stereo image pairs.



(a) Left image after pan/tilt control (b) Right image after pan/tilt control

Fig. 5. Stereo image pair after pan/tilt control.



(a) Composite image before tracking (b) Composite image after tracking

Fig. 6. Composite images before and after being tracked by pan/tilt system.

Table 2. Calculated moving angle of the pan/tilt system

[unit : degree]

Degree / Frame	Left		Right	
	Pan	Tilt	Pan	Tilt
1	11.625	-3.3281	12.4218	-2.5781
2	4.828	-4.594	5.625	-2.906
3	-0.609	-2.250	-0.188	-2.344
4	-11.953	-2.625	-11.719	-5.109

Table 3. Moving distance of the target object

[unit : pixel]

Distance / Frame	Actual distance		Detected distance	
	Relative moving distance	Absolute moving distance	Relative moving distance	Absolute moving distance
1	(0, 0)	(0, 0)	(0, 0)	(0, 0)
2	(62, -6)	(62, 6)	(64, -6)	(64, 6)
3	(63, -2)	(125, 10)	(61, -1)	(125, 11)
4	(70, 3)	(195, 15)	(71, 2)	(196, 14)
5	(65, 8)	(260, 20)	(63, 7)	(259, 19)

Fig. 6 shows the composite images before by after the target object has been tracked by the pan/tilt system using the rotation degree in Table 2. From Fig. 6, we can see that if the moving target is under tracking, the convergence angle on the target object is controlled, so that a stereo disparity between the left and right images is maintained at zero.

Table 3 shows the experimental results for the actual and detected moving distances of a target extracted from the sequential input stereo image. In Table 3, the 1st frame means an initial (reference) frame such that, the moving distance in the 1st frame is given by (0, 0) and for each tracking frame, both of the relative and absolute moving distances are calculated and compared. From the experimental results of Table 3, we can see a very low error of 1~2 pixels between the actual and detected moving distance values of the target object.

Fig. 7 also shows the 3D plot for the experimental result of Table 3, in which the absolute coordinate values are used. The initial location of the moving target is assumed to be in the origin as given by Table 3 where as the absolute moving distance of the target for each tracking

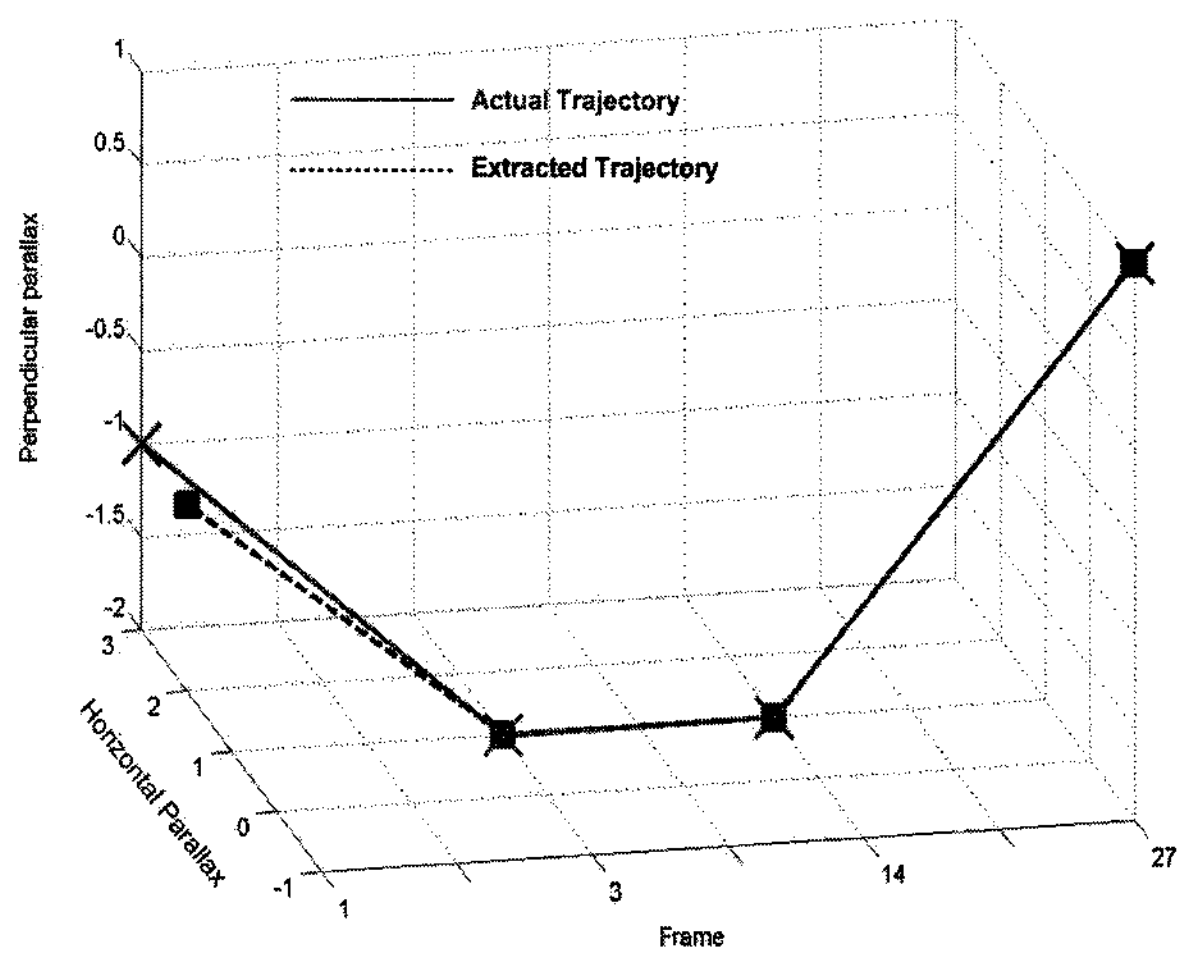


Fig. 7. 3D plot of the target's moving trajectory.

Table 4. Location error ratio of a moving target

[unit: %]

Frame / Error	2	3	4	5
Error ratio	1.21	1.55	1.02	1.41
Average error ratio	1.29			

Table 4 also shows the calculated tracking error ratio between the actual and detected coordinates of the target object for each frame, in which the error ratio is found to be a very low value of 1.29 % on average.

Thus, band on these satisfactory experimental results, we can conclude that the proposed real-time moving-target extraction and tracking system using the MSE and BPEJTC algorithms can be practically implemented on a stereo moving-target tracking system.

4. Conclusions

In this paper, a new stereo 3D moving-target tracking system using the MSE and BPEJTC algorithms is propesed. Based on the experimental results of 20 frames of the stereo image pair, we found that the proposed system is capable of successfully tracking a moving target with relatively low error ratio of about 1.29 % on average at in real time. This good experimental result suggests that a potential of implementing a stereo moving target-tracking system with a high of accuracy and very fast tracking time.

References

[1] Y. Mae, S. Yamamoto, Y. Shirai, and J. Miura, in Proc. of 2nd Japan-France Congress on Mechatronics, **2**, 545 (1994).

[2] J. S. Lee, C. W. Seo, and E. S. Kim, Optics. Communications, **200**, 73 (2001).

[3] C. Tam, T. S. Yu, A. Gregory, and D. Juday, Opt. Eng., **29**, 314 (1990).

[4] B. C. Kuo, Automatic Control Systems, Prentice Hall, (1995), p. 1.

[5] J. S. Lee, J. H. Ko, and E. S. Kim, Optics Comunications., **191**, 191 (2001).

[6] D. Comaniciu and V. Ramesh, Third IEEE International Workshop on Visual Surveillance, **1**, 11 (2000).

[7] T. J. Olson and D. J. Coombs, Intl. J. of Computer Vision, **7**, 67 (1991).