3D Robot Vision System using the Hierarchical Opto-Digital Algorithm

Jung-Hwan Ko, Eun-Soo Kim
National Research Laboratory of 3D Media, School of Electronic Engineering, Kwangwoon University, 447-1 Wolge-Dong, Nowon-Gu, Seoul 139-701, Korea, misagi@explore.kwangwoon.ac.kr

Abstract

In this paper, a new 3D robot vision system using the hierarchical opto-digital algorithm is proposed and implemented. From some experimental results with the 20 frames of the stereo input image pairs, the proposed system is found to be able to effectively extract the area where the target object is located from the stereo input image regardless of the background noises.

1. Introduction

The 3D vision system can perceive three-dimensional feeling, just as the structure of the human eyes, by using the binocular disparity [1] between the left and right eyes. In this system, in case the viewing-point of the target object is not in accord with focusing points of the left and right eyes, it gives fatigue to the eyes because the target object is seen as two different objects. Therefore, when there exists stereo disparity in the target object, it has to be removed, and this process is called the convergence angle control [2]. Generally the stereo vision system have to control the convergence angle of camera according to the distance of object just like as the apple of eye do at the same time track the object. Convergence angle control keeps the stereo disparity of between the right and left images zero.

In this paper, a new 3D tracking algorithm to control the convergence of moving object in the 3D robot vision system is proposed. The proposed algorithm, firstly, extracted the target object by removing the background noises through the block-based MSE (mean square difference) of the sequential left images and then, controlled the pan/tilt and convergence angle of the stereo camera by using the coordinates of the target position obtained from the optical BPEJTC(binary phase extraction joint transform correlator) which are executed between the extracted target image and the input image. Also, a novel 3D vision system is proposed, in which it can detect and segment the moving object adaptively through controlling the pan/tilt system of stereo cameras by using optical BPEJTC.

2. Convergence Control by using Pan/Tilt System

This proposed system can adaptively track a moving target by controlling the convergence of stereo camera. Firstly, extracted the target object by removing the background noises through the block-based MSE of the sequential left images. Secondly, the location's coordinate of the moving target object for each of the sequential input frames can be extracted through carrying out optical BPEJTC between the reference image of the target region mask and the stereo input image. Finally, the convergence and pan/tilt of stereo camera can be sequentially controlled by using these target coordinate values and the target can be kept in tracking.

The newly proposed algorithm in this paper consists of the 3 steps.

2.1 Step.1

In this paper, the tracking object is perceived and extracted through the MSE algorithm, in which the region-based window mask is used [5].

The MSE algorithm is applied to the reference image of the window mask obtained from the previous frame and the input left image. Just like the Eq. (1), in the MSE algorithm, the degree of similarity between the reference image of the window mask and the input image is measured and the MSE value on each pixel location is calculated. That is, in case the MSE algorithm as shown in Eq.(1) is applied to the reference image of the window mask [$I_{Ref}(t-1)$] and the left image [$I(t)$], the minimum value is calculated on the very identical pixel domain.

$$MSE(a,b)=\frac{1}{N_x \cdot N_y}\sum_{j=0}^{(N_y-1)}\sum_{i=0}^{(N_x-1)}\left| I(a+i,b+j)-I_{Ref}(i,j)\right|^2 \quad (1)$$

Where and are the search range of the MSE algorithm and given by $a=-2N_x \sim 2N_x$, $a=-N_y \sim N_y$ respectively. Then, by applying the MSE algorithm, find the pixel domain of the minimum value. Then, a window mask is formed on the pixel domain having the minimum value as shown in Eq. (2), and the tracking object can be extracted by using this mask.

$$
l'(t) = \begin{cases} \text{Window Mask}, & MSE = \text{the minimum} \\ 0, & \text{the otherwise} \end{cases} \qquad (2)
$$

Here, $l'(t)$ means the left image, in which the tracking object is just separated from the background in the $l(t)$ image. Then, by extracting the location coordinates [$(\Delta x_l, \Delta y_l)$] of the tracking object and moving to the center of the image, a new image is obtained. This image is used as an updated reference image in the next frame's MSE algorithm. And, this moving value of $(\Delta x_l, \Delta y_l)$ is used to control the pan/tilt of the left camera.

2.2 Step.2

For the right camera's object tracking, the optical BPEJTC(Binary Phase Extraction Joint Transform Correlator) is applied to the image [$l'(t)$] from the left image and the right image [$r(t)$]. Then, the relative distance between two objects can be calculated and this value is used to control the pan/tilt of the right camera.[5]

In input plane of the optical BPEJTC shown in Fig. 1, the image on the spatial light modulator (SLM1) is Fourier-transformed by the lens and detected on CCD1 as the form of JTPS (joint transformed power spectrum).

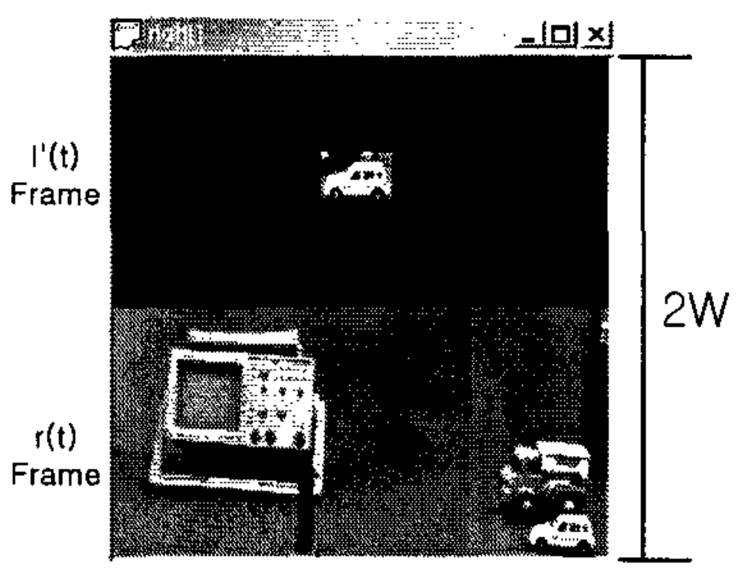


Fig. 1. The input plane of optical BPEJTC

The reference image and the right image which are given by Eq. (3), are positioned in the upper and lower half of the spatial light modulator (SLM1), respectively.

$$
l'(x-\Delta x_l,\ y-[\Delta y_l+\frac{w}{2}]),\quad r(x-\Delta x_r,\ y-[\Delta y_r-\frac{w}{2}]) \qquad (3)
$$

Where 2w is the height of the display unit. $(\Delta x_l, \Delta y_l)$ and $(\Delta x_r, \Delta y_r)$ are the distance coordinates to the tracking object from the screen center of the $l'(t)$ and $r(t)$ image, respectively.

Equation (4) shows the correlation result between the modified JTPS and the image of the SLM2.

$$
\begin{aligned}
C_{BPEJTC}(x,y) &= \Im^{-1}\{E_{BPEJTC}(u,v)\} \qquad (4)\\
&= [l'(x,y)\otimes r(x,y)] * \delta[x+(\Delta x_l-\Delta x_r),\ y+(\Delta x_l-\Delta x_r+w)]\\
&\quad + [r(x,y)\otimes l'(x,y)] * \delta[x-(\Delta x_l-\Delta x_r),\ y-(\Delta x_l-\Delta x_r+w)]
\end{aligned}
$$

The correlation peak point on the correlation plane is detected on CCD2. This peak point is the relative distance between the reference image and the right image and given by Eq.

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

Consequently, using this correlation peak point, the location coordinates $(\Delta x_r, \Delta y_r)$ of the tracking object on the right image can be obtained.

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

2.3 Step.3

Through the use of the USB(Universal Serial Bus) interface, the location coordinates of the left and right images obtained at the step1 and step 2 is transferred to the pan/tilt of the stereo camera system, then it can be controlled according to this values. Fig. 2 shows the tracking screen, in which the camera's pan/tilt system centers the tracking object on the camera's field of view(FOV). Therefore, the stereo object tracking is possible through the continuous control of the stereo camera's pan/tilt by the detected values of $(\Delta x_l, \Delta y_l)$ and $(\Delta x_r, \Delta y_r)$.

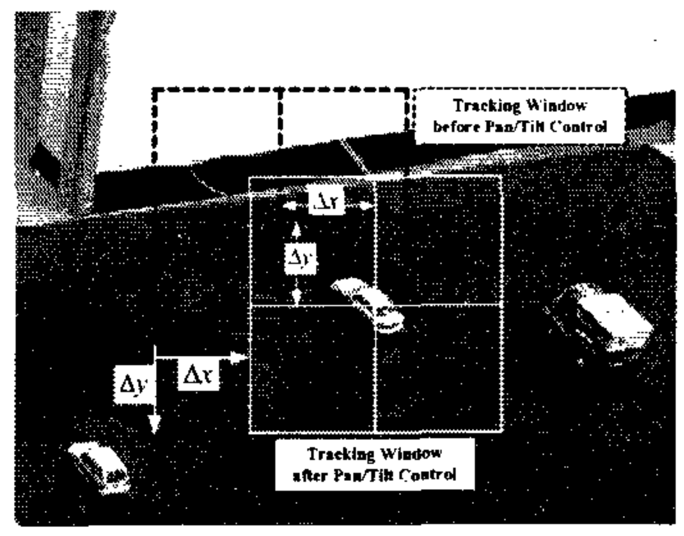


Fig. 2. Tracking screen by pan/tilt camera system

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

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

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

$$
\begin{aligned}
u(n) &= e(n)*(k_p+k_i+k_d) + e(n-1)*(k_i-2k_d)\\
&+ e(n-2)*(k_i+k_d) + k_i*(e(n-3)+\cdots e(0)) \qquad (8)
\end{aligned}
$$

In Eq. (7) and (8), each parameter is defined as following, respectively.

$u(n)$ = 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

In this paper, the pan/tilt simulator is implemented, in which the PID control variables can be given and PID can be tuned.

3. Stereo Tracking Experiments and Results

In this experiments, the cameras of Dong Kyoung Electronics, Inc. (Model: CS-82393BS) was used as the left and right cameras with configuration of crossing stereo camera and two pan/tilt control systems of Hanwool Robotics Co. (Model : HWR-PT1) are also used as the pan/tilt control devices. The distance between the left and the right cameras was 22cm, and 30 image frames were saved and used at the experiments. The Meteor II/4 and Meteor II MC/2 of Matrox, Inc. were used as the image saving frame grabber, and Pentium (1.2GHz, 512MB) was used as the digital image processing system.

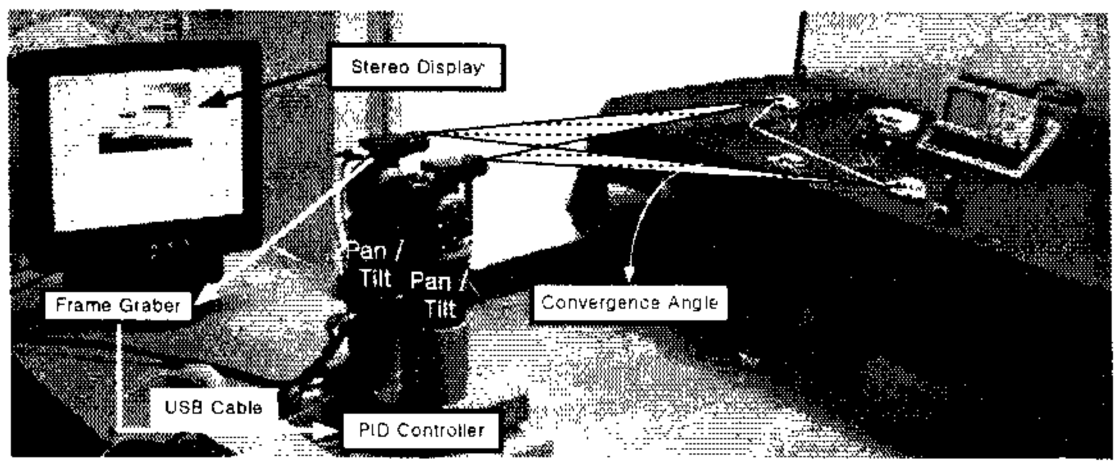
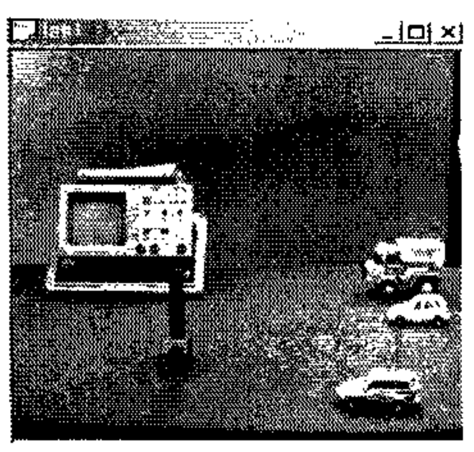
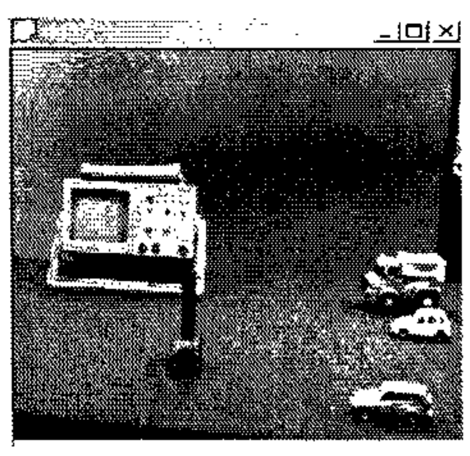


Fig. 3. Pan/Tilt control system with crossing stereo camera by using experiment

3.1 Step.1

In case the camera and the tracking object move simultaneously, the tracking result for two frames of the 20 images are shown in Fig. 4.

Frame 1	
Left Image	Right image
	
Target mask by MSE :	Reference image :

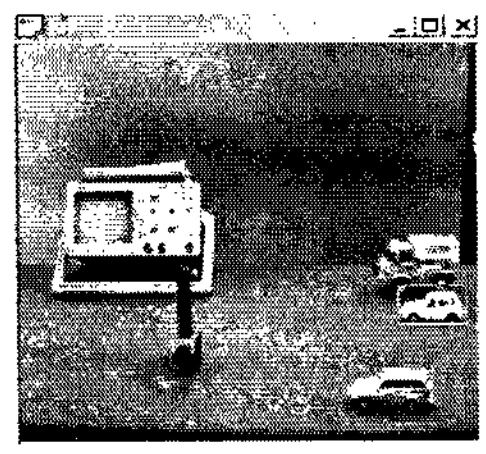
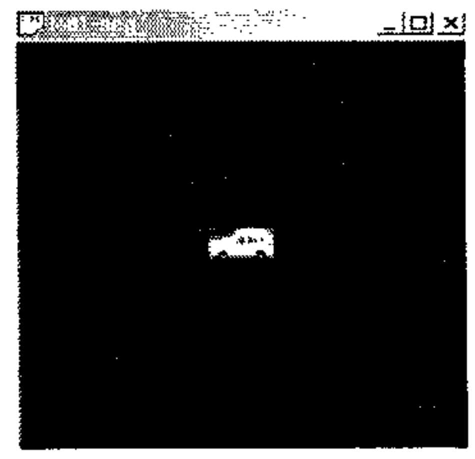
$I_{REF}(t)$	$I'(t)$
	

Fig. 4. Stereo input image and extracted tracking object

3.2 Step.2

Fig. 5 shows the input plane to execute the optical BPEJTC between the segmented reference image $l'(t)$ and the right input image $r(t)$ and the correlation peak point on the correlating plane after performing the optical BPEJTC. The distance, $(\pm\Delta x_l, \pm\Delta y_l)$ and $(\pm\Delta x_r, \pm\Delta y_r)$, from the center of a screen to the tracking object of the left and the right images can be acquired by using the location coordinates of the correlation peak points.

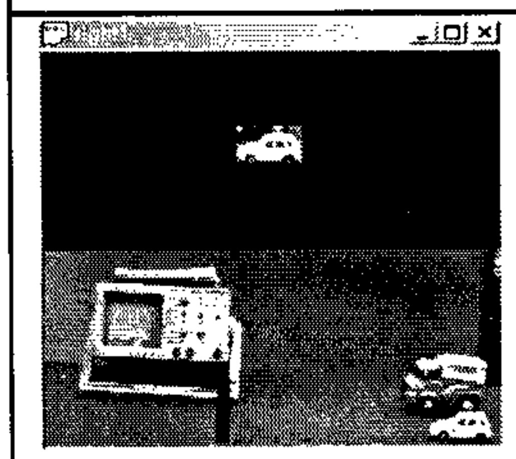
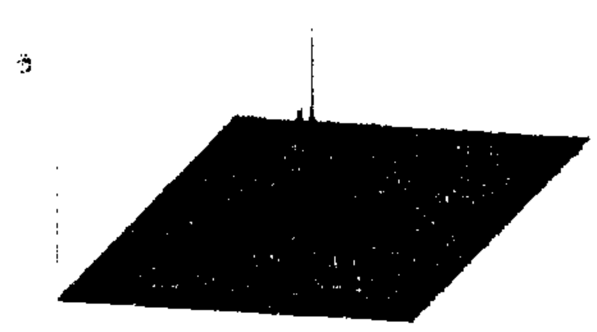
Frame 1	
BPEJTC input plane	Correlation peak point (Location's coordinate)
	 (265, -103)

Fig. 5. Simulation results of optical BPEJTC

Here, this coordinate system shows the location coordinates of the tracking object in the (x,y) axis having the central coordinates in (0,0) of the input image. This extracted position coordinates are used to control the pan/tilt control of right camera.

3.3. Step.3

Figure 6 shows the images after controlling the pan/tilt by using the moving distance for two frames of the input images [$l(t)$ and $r(t)$]. Here, the moving distance is converted to the degree for the pan/tilt control variables.

Frame 1	
Left image after pan/tilt control	Right image after pan/tilt control

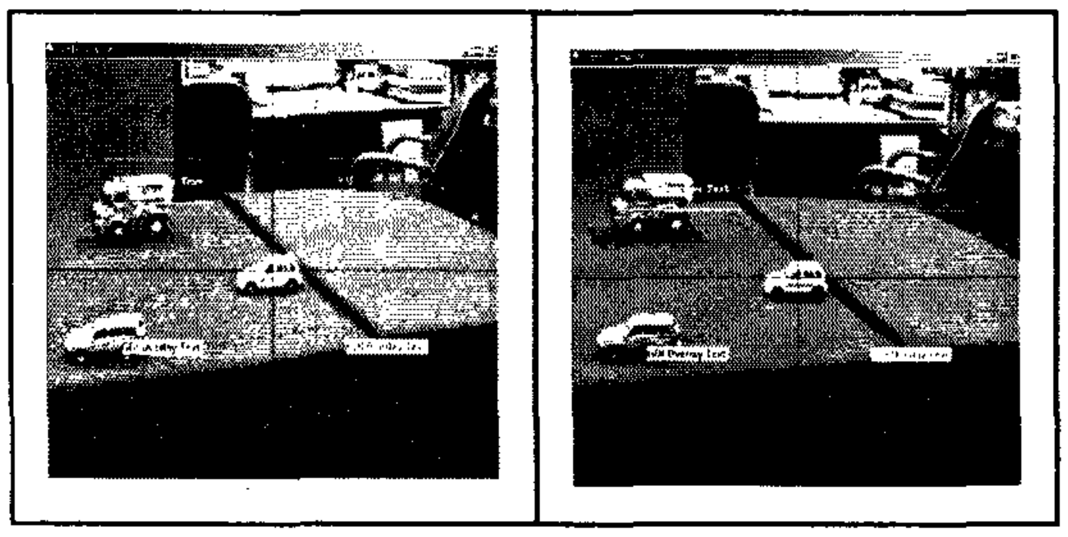


Fig. 6. Image after pan/tilt control

Table 1 shows the moving degree and the simulation results of tracking object two frames of 20 frames on the stereo input image.

Table 1. Moving degree of pan/tilt between pixels

Degree / Frame	Left		Right	
	Pan	Tilt	Pan	Tilt
Frame 1	11.625	-.3281	12.4218	-2.5781

Fig. 7 shows the stereo composite image before tracking and the stereo composite image after tracking by suing table 1 on input image about the two frames of $l(t)$ and $r(t)$

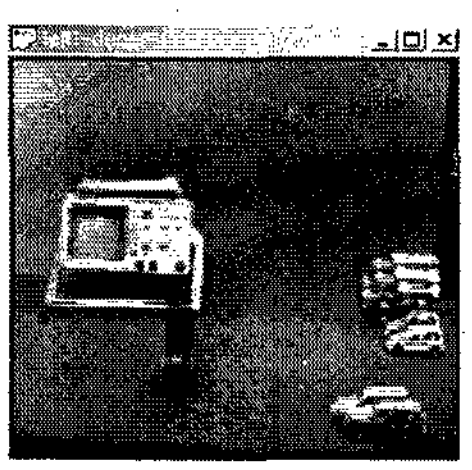
Composite image before tracking	Composite image after tracking
	

Fig. 7. Composite image after convergence and pan/tilt control

Table 2 shows that there is only an error of 0~1 pixels between the shifted location of the moving object and the actual location. Through experimental results for 20 frames, we prove that the proposed system efficiently tracks moving objects.

Table 2. The center coordinates of target object

Position / Frame	Left image	Right image	Error	
			△x	△y
Frame 1	(2, -1)	(3, -1)	1	0

>From Fig. 8, the tracking object (car) is shown, in which the car is always at the center of the camera's field of view. Also, it is found that in case of the stereo composite images after stereo tracking the convergence points are controlled at the same time.

4. Conclusion

In this paper, a new stereo tracking system using the hierarchical opto-digital algorithms is proposed and implemented. From some experimental results with the 20 frames of the stereo input image pairs, the proposed system is found to be able to effectively extract the area where the target object is located from the stereo input image regardless of the background noises. With the location values of the tracking object obtained from the execution of the optical BPEJTC, the convergence angle and the pan/tilt of the stereo cameras are found to be successfully controlled. Therefore, in this paper, a feasibility test for implementing the stereo tele-working system or the stereo robot vision system using the proposed algorithm is suggested.

5. Acknowledgement

This work has been conducted by National Research Laboratory (NRL) Program of the Ministry of Science and Technology of Korea.

6. References

[1] Rafael C. Gonzalez and Richard E. Woods, "Efficient block motion estimation using integral projections", Digital Image Processing, Addison-Wesley, 1992

[2] Stephen T. Barnard, and William B. Thompson, "Disparity analysis of images", IEEE Trans. PAMI, vol. PAMI-2, no.4, pp.333-340, July, 1980

[3] R.C Jane, "Segmentation of frame sequences obtained by a moving observer," IEEE Trans. PAMI, vol.6, no.5, pp.624-629, 1984

[4] M.K. Leung, "Human body motion segmentation in a complex scene," SPIE Optical Pattern Recognition, vol.20, no.1, pp.55-64, 1987

[5] J. S. Lee, J. H. Ko, E. S. Kim, "Real-time stereo object tracking system by using block matching algorithm and optical binary phase extraction joint transform correlator", Optics Communications, vol. 191, pp.191-202, 2001