Stereo Camera-based Intelligence Surveillance System

Junghwan Ko[1] and Jungsuk Lee[2]
[1]3D Display Research Renter, Korea Ministry of Information and Communication, Korea
[2]Department of Mechatronics, Inha Technical College, Korea.
Email: jhko@inhatc.ac.kr; ungbolee@inhatc.ac.kr

***Abstract*—In this paper, a stereo camera-based video surveillance system using pan/tilt controller is suggested and implemented. Some experiments with video images for 3 moving persons show that a person could be identified with these extracted height and stride parameters. Some experiments with video images for 3 moving persons show that a person could be identified with these extracted height and stride parameters.**

***Index Terms*—stereo camera, surveillance, pan-tilt.**

I. Introduction

So far most of the conventional video surveillance systems have been developed basing on the monocular camera system so that, gathering 3-dimensional information of a moving target and highly accurate and robust tracking a moving target from the stream of monocular image are known to be a very hard to be achieved in these systems [1]-[3]. These limitations of the conventional surveillance and tracking system make it difficult to measure and estimate the human behavior and movement under tracking [4]. Typically, video surveillance system which has already been applied to supermarkets, ATM rooms, airports, stations, parking lots, and other public places, monitor an environment with CCTV(closed circuit television) cameras using the perceptual capabilities of a human operator to detect and identify target objects moving within the camera's FOV(field of view). But, these systems have a drawback to keep operators to watch the monitors, so that this manual task is so labor-intensive and inefficient.

Recently the wide spread of CCTV cameras and the advances in digital information technologies now allow robot and computer vision researchers to develop various automatic video surveillance systems for the purpose of monitoring and security [1]-[3]. The basic task of a video surveillance system is to accurately detect a moving target from the sequential input images regardless of complex background and track the moving target by controlling the surveillance camera with the obtained coordinate values of the target on the frame basis [4]. That is, if full information for the moving target can be extracted from the sequential input video images, it can be used not only for detection of the target object, but also for obtaining its real three-dimensional position data. Moreover, by calculating the changes of the location values for the target object between two consecutive frames, the moving target can be tracked or monitored.

Accordingly, in this paper, In the proposed method, face area of the moving target person is extracted from the left image of the input stereo image pair by using a threshold value of YCbCr color model [5] and by carrying out correlation between the face area segmented from this threshold value of YCbCr color model and the right input image, the location coordinates of the target face can be acquired, and then these values are used to control the pan/tilt system through the modified PID-based recursive controller. In addition, the proposed real-time stereo target tracking system is implemented and some experimental results with this system by using a sequence of 780 frames of stereo input image are also included.

II. The Proposed Surveillance System

Fig. 1 shows an operational flowchart of the proposed stereo security surveillance system, which is largely consisted of 2 stages. At the 1st stage, target's face region and location coordinates in the left image plane are detected by use of YCbCr color model and centroid method and then, location coordinates of a moving target in the right image plane are obtained through correlation between the target face extracted from the left image and the right image by using the BPEJTC algorithm. Then, displaced distances from the centers of the left and right image planes to the centers of the target face are calculated and used to control the pan/tilt angles not only for positioning the target face in the center of the camera's FOV and but also for making the focusing points of the right and left cameras coincided on the target face, which is known as a function of convergence control.

At the 2nd stage, using a triangulation method with rotated pan/tilt angles and geometry of the pan/tilt system, distances from the left and right cameras to the target face can be computed and then, finally 3D location data of a target person in the world space can be obtained. Moreover, by computing the displacement distance of the target person between two consecutive frames, moving trajectory of a target person can be also obtained.

Manuscript received May 22, 2014; revised July 30, 2014.

doi: 10.12720/joace.3.3.253-257

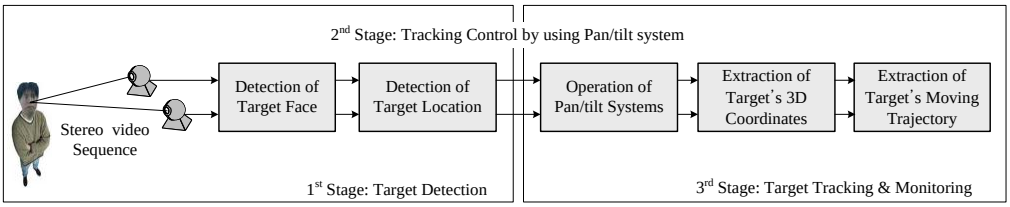

Figure. 1. Operational flowchart of the proposed video surveillance system

A sequence of stereoscopic video image pairs is captured by two cameras and potential face areas of the target person in the sequential left images are detected by using a skin color model [10]. That is, this algorithm can classify the input left image into the skin and non-skin color pixels in the YCbCr color space [5], in which Y, Cr and Cb represents the luminance and chrominance of the image, respectively.

Because this technique makes use of intensity values of both luminance and chrominance components of the image, it can detect the potential face areas by exploiting the distribution property of these components as shown in Eq. (1) [5].

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -37.797 & -74.203 & 112 \\ 112 & -93.786 & -18.214 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

Also, in this paper, as a method to extract the position data by which a change of relative position between two input images can be obtained, the phase-type correlation method is used, in which relative position between the segmented reference frame and the moving target frame can be detected by calculating the correlation peak value from sequential frames. That is, the reference target image and the right image are given by Eq.(1) in the phase-type correlation [3].

$$l_t'(x-\Delta x_l,\ y-[\Delta y_l+\frac{w}{2}]),\ r_t(x-\Delta x_r,\ y-[\Delta y_r-\frac{w}{2}]) \quad (2)$$

where, 2w is the height of the input display unit $(\Delta x_l, \Delta y_l)$ and $(\Delta x_r, \Delta y_r)$ are the distance coordinates to the target from the screen center of the $l_t'(x,y)$ and $r_t(x,y)$ images, respectively. In this paper, the crossing camera geometry is employed, in which the cameras' optic axis coincides with the convergence point of a target face and the optic axis of the cameras are moved in proportion to the distance of a target face. Here, the distance between the target and stereo camera can be calculated by using the triangulation method [3].

According to the moving distance values obtained in the 1st stage on the frame basis, the pan/tilt angles can be controlled. That is, moving distances in the x, y direction between two consecutive frames are just given by the values of $\Delta x_l, \Delta y_l$ and these are used for controlling the pan/tilt system to track a moving target. For this purpose, in this paper, moving distance data obtained in each frame are converted into the pan/tilt control angles by using the Kanatani's background compensation algorithm [6], which could apply to two consecutive frames of the current image (x_t, y_t) and the previous image (x_{t-1}, y_{t-1}) as shown in Eq. (3).

$$x_{t-1} = f \cdot \frac{x_t + \theta \sin\alpha \cdot y_t + f \cdot \theta \cos\alpha}{-\theta \cos\alpha \cdot x_t + \phi \cdot y_t + f},$$
$$y_{t-1} = f \cdot \frac{-\theta \sin\alpha \cdot x_t + y_t - f \cdot \phi}{-\theta \cos\alpha \cdot x_t + \phi \cdot y_t + f} \quad (3)$$

where, f means a focal length of the stereo camera and α, θ, and ϕ represents an initial inclination, pan and tilt rotation angles of the pan/tilt system, respectively. From Eq. (3), pan and tilt rotation angles can be derived as shown in Eq. (4).

$$\theta = \frac{f x_{t-1} y_t (y_{t-1} - y_t) + f(f^2 - y_{t-1} y_t)(x_{t-1} - x_t)}{f(y_t \sin\alpha - f\cos\alpha) + x_{t-1} x_t (y_{t-1} y_t \cos\alpha - y_t \sin\alpha + \cos\alpha)} \quad (4)$$

$$\phi = \frac{f x_t \cos\alpha (x_{t-1} - x_t)(f - y_{t-1}) + f(y_{t-1} y_t) x_{t-1} x_t \cos\alpha + f(y_t \sin\alpha + f\cos\alpha)}{(f^2 - y_{t-1} y_t)(x_{t-1} x_t \cos\alpha + f y_t \sin\alpha + f^2 \cos\alpha) - (1 - y_{t-1})(x_{t-1} x_t^2 \cos\alpha)}$$

Accordingly, if the location data of the current and previous images are detected then, the corresponding pan and tilt rotation angles of and can be calculated by using Eq. (8) and finally motor control angle can be derived through an encoder of the pan/tilt controlling system.

In this paper, stereo camera is constructed with a crossing geometry, in which the cameras' optic axes coincide with the convergence point of a target face and the optic axes of the cameras are moved in proportion to the distance between the target and stereo camera, which can be computed by using a triangulation of the crossing camera as shown in Fig. 2.

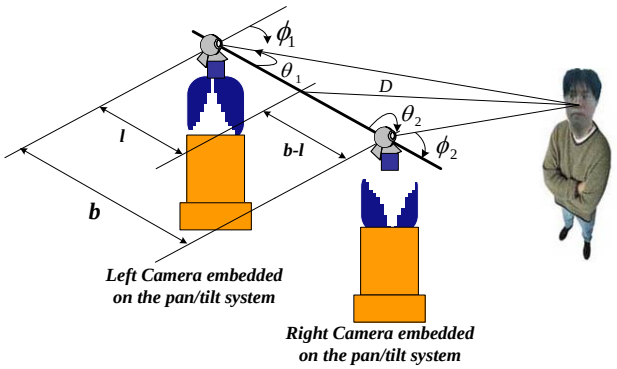

Figure. 2. A crossing stereo camera geometry

where, b, l and D represents a baseline between the left and right cameras, a distance from a point in the baseline to the left camera and a distance from a point in the baseline to the target face, respectively. Then, the distance D can be obtained by using Eq. (5) under the condition l value is given.

$$D = \left(\frac{\tan\theta_1 \cdot \tan\theta_2}{\tan\theta_1 + \tan\theta_2} \right) \cdot d \qquad (5)$$

Fig. 3 shows a space model of the proposed pan/tilt-based stereoscopic video surveillance system to extract target's 3D location coordinates and moving trajectory. Some geometric parameters used in the space model are summarized in Table I.

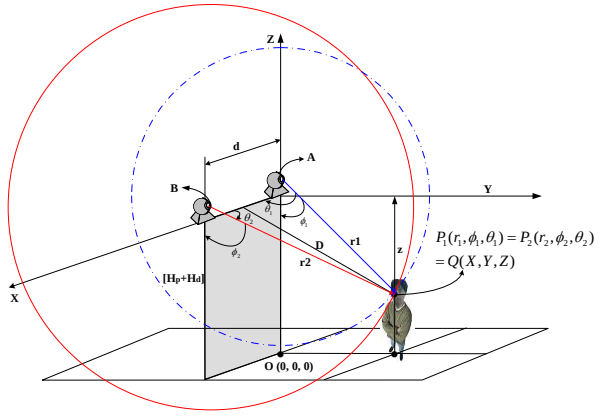

Figure. 3. Space model of the surveillance tracking system

where, x and y means the practical position of the target image, which can be estimated from the moving trace of the target image. Also, the height estimation of target person can be calculated by using a method as shown in Fig. 2.

TABLE I: GEOMETRIC PARAMETERS USED IN THE SPACE MODEL OF FIG. 3

Parameter	Definition of the parameters
θ	Pan angle of the stereo camera
ϕ	Tilt angle of the stereo camera
d	Distance between the left and right camera
D	Distance from the stereo camera to the target face
H_P	Physical height of the pan/tilt system
H_d	Physical height of the desk on which the pan/tilt systems are positioned
A	Location of the left camera in the Cartesian coordinates (0, 0, $[H_p+H_d]$)
B	Location of the right camera in the Cartesian coordinates (d, 0, $[H_p+H_d]$)
P,Q	Center location of the target face in the 3D world coordinates

In case a target person stands at an arbitrary point Q in the 3D space as shown in Fig. 4, the left and right pan/tilt are rotated in proportion to the displacement distances obtained in the 1st stage to adaptively track a target person, in which pan/tilt angles are given by, and, respectively. Using these angle values and structure of a stereo camera system, the distance data from the stereo camera to the target face can be calculated using Eq. (5) and the distance data from the left and right cameras to the target face, r1 and r2, in Fig. 4 can be also obtained from Eq. (6).

$$r_1 = \frac{D}{\sin\theta_1}, \qquad r_2 = \frac{D}{\sin\theta_2} \qquad (6)$$

Accordingly, using pan/tilt angles and distances from the left and right cameras to the target face, 3D location coordinates of a target person in the spherical coordinates can be obtained on the frame basis. That is, two kinds of the spherical location coordinates for the same target person can be obtained; one is originated from the left camera, which is given by , and the other from the right camera and given by , so that the real 3D locations of the target in the Cartesian coordinates can be finally obtained from them through coordinate conversion. As these target's 3D location data can be extracted on the frame basis, not only its moving trajectory but also some target information such as height, stride and moving velocity of a target person useful for target identification can be estimated from them.

Accordingly, the height of the target person can be estimated by using Eq. (7).

$$M_h = (P_h + D_h) - z + f_h \qquad (7)$$

III. EXPERIMENTS AND RESULT

In the experiments, two USB-PC cameras of Chung Mack Electronics (MPC-M55) with 320 240 pixels are used for capturing input stereoscopic video image pairs at a speed of 30 frames/s and two pan/tilt systems of Hanwool Robotics (HWR-PT1), on which the left and right cameras are embedded, are also used for tracking a target person. Here, a crossing camera geometry is employed for constructing stereo camera and the distance between the left and right cameras is fixed by 12 cm. Also, the physical heights of the pan/tilt system and the desk on which two pan/tilts are located are measured to be 46cm and 179cm, respectively. A user interface for the proposed stereoscopic video surveillance system has been developed to see an entire operational situation of stereoscopic target detection and tracking processes on the screen. Fig. 4 shows an implemented user interface for the proposed pan/tilt-based stereoscopic video surveillance system. As shown in Fig. 4, the sequential stereoscopic image pairs caught up by stereo camera are transferred to the host computer through a general graphic card. In the host computer, position values of the target face in each frame are detected through execution of YCbCr skin color and centroid algorithms and then, transmitted to the pan/tilt control board, in which the control signals for pan/tilt systems are generated using the micro controller (89C51) and by using these signals the pan/tilt systems are finally controlled through the motor controller (LM629). The user interface for the proposed stereo video surveillance system is developed in visual C++.

As shown in Fig. 4, the sequential input stereo image pairs are caught up by the stereo camera embedded on the pan/tilt system and transferred to the host computer through a general graphic card and then, transmitted to the pan/tilt control board, in which the feedback control signals for the pan/tilt are generated using micro controller (89C51) and finally the pan/tilt is controlled by using these signals through the motor controller (LM629). For the sequential 80 frames of stereo input images having a

resolution of 640*480 pixels and a 30 frames/sec, randomly selected 3 frames are shown in Fig. 5.

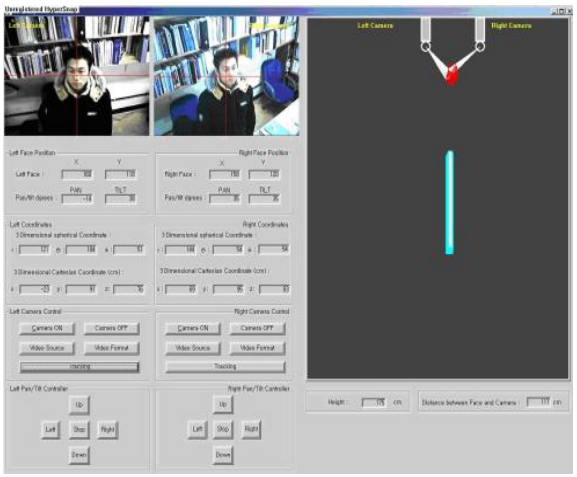

Figure. 4. Experimental setup for pan/tilt-based stereo target surveillance-tracking

Figure. 4. Examples of the captured stereo input images

In Fig. 4, spherical coordinates(r, Φ, θ) composed by using the extracted parameter r, Φ and θ are shown Table 1, and then spherical coordinates(r, Φ, θ) are converted to Cartesian coordinates (x, y, z) to extract the current position of the target human. The subject's face has been detected within a few frames after entering the field of view (FOV) of the camera.

Table II also shows the extracted center locations of the target face for the above 4 sample frames. Moreover, moving distance of a target person between two consecutive frames can be obtained in each left and right image by computing the location shift of the target face from the center of the input image plane. Here, the center location of (160, 120) in each left and right image plane with 640*480 pixels is taken as the reference coordinates for calculating the moving distance of the target between two consecutive frames. This target's moving distance between two consecutive frames can be used for controlling the left and right pan/tilt systems to track a target person under tracking.

TABLE II. EXTRACTED CENTER COORDINATES OF THE TARGET FACE FOR FOUR SAMPLE FRAMES

Center coordinates of the target face (*x*, *y*)		
Frame	Left image	Right image
1st	(225,39)	(280,39)
240th	(37,123)	(39,136)
540th	(112,142)	(100,141)
810th	(226,146)	(239,137)

The target's moving distance between two consecutive frames is given by number of pixels in the x and y directions, so that moving distances in the x and y direction can be converted into the corresponding pan/tilt control angles of and through encoders of the pan/tilt systems by use of Eq. (8). Finally, these pan/tilt angles are used for controlling the pan/tilt systems to keep the target face at the center of the camera's FOV and to make the focusing points of the right and left camera coincided on the target face, as a result the moving target can be kept to be under tracking.

In the stereo camera system, a distance from the base line of the stereo camera system to the target face, D can be found by using a triangulation method with the left and right pan angles and the distance between the left and right cameras, and distances from the left and right camera to the target face, r1, r2 can be also computed by using Eq. (6). Those values are illustrated in Table III for 4 sample frames. Finally 3D location coordinates of a target person in the spherical coordinates system can be obtained from the pan/tilt angles pointing to the target from the left and right camera and the distances from the left and right camera to the target mentioned above. Here two kinds of the spherical location coordinates for a target person can be derived; one is originated from the left camera and represented by $P_1(r_1,\phi_1,\theta_1)$, and the other from the right camera and represented by $P_2(r_2,\phi_2,\theta_2)$ as shown in Table III.

TABLE III. TARGET'S 3D LOCATION IN THE SPHERICAL COORDINATES SYSTEM

Coordinates / Frame	Left Camera		Right Camera		Distances from the target to the cameras (D) [cm]
	Optic axis distance (r) [cm]	Pan/tilt angle (Φ, θ)	Optic axis distance (r) [cm]	Pan/tilt angle (Φ, θ)	
# 3	204	(91, 61)	208	(60, 69)	205.4
# 48	181.1	(83, 77)	182	(84, 82)	180.5
# 77	274.1	(96, 79)	275	(69, 82)	273.4

As this data can be extracted on the frame basis, not only the target's 3D location coordinates at each frame and its moving trajectory can be obtained, but also some personal information such as the height, stride and moving velocity of a target can be also estimated from them. Moreover, as shown in Table III, Z means the height from the top of the camera to the center coordinates of the target face, so that the height of a target person can be calculated by subtracting Z from the physical heights of the pan/tilt and the desk on which two pan/tilts are located. Fig. 5

illustrates the target's moving trajectory in (X, Y, Z) for calculating the height and moving velocity of a target person. From Fig. 5, it is found that the height (Z) of a target person is estimated to be about 177cm. and the target person might be moving almost upright in the room.

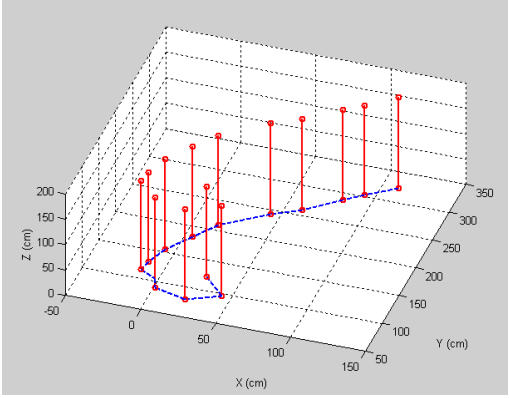

Figure. 5. 3D locations and moving trajectory of a target person

From these good experimental results discussed above, it is analyzed that the proposed pan/tilt-based stereoscopic video surveillance system can effectively track a moving target with very low tracking error and with the proposed method a real-time stereoscopic video surveillance system can be implemented. In addition, in the proposed method, using the target's 3D location coordinates at each frame and its moving trajectory, some personal information such as the height, stride and moving velocity of a target could be also estimated from them. Basically, this paper tried to show a possibility of implementation of a pan/tilt-based stereoscopic video surveillance system under the relatively simple situations of one moving target and slow change of backgrounds. Accordingly, as the future works, a sophisticated target detection and tracking algorithm to exactly extract the convergence point of the stereoscopic images and the location coordinates of a moving target, and to accurately track a moving target in various real situations such as multiple targets, single target in the dynamic background, large and occluded target etc., must be studied. And a method to adaptively compensate the mechanical errors of the pan/tilt systems must be researched as well.

IV. Conclusion

In this paper, a stereo camera-based video surveillance system using pan/tilt controller is proposed. The proposed system can detect a moving human face from the stereo image sequences captured by the stereo camera system using a threshold value of YCbCr color model, measure its distance and 3D coordinates and then, with these values control the stereo camera by using the pan/tilt system for tracking the moving face in real-time. These experimental results suggest a possibility of implementing a new real-time intelligent surveillance system for robust detection and tracking of a moving target person by using the proposed scheme.

References

[1] P. Danielson, "Video surveillance for the rest of us: Proliferation, privacy, and ethics education," in *Proc. International Symposium on Technology and Society*, vol. 1, no. 1, 2002, pp.162-167.
[2] J. Black and T. Ellis, "Multi-camera image measurement and correspondence," *Measu-Rement*, vol. 32, pp. 61-71, 2002.
[3] J. S. Lee, J. H. Ko, and E. S. Kim, "Real-time stereo object tracking system by using block matching algorithm and optical binary phase extraction joint transform correlator," *Optics Communication*, vol. 191, pp. 191-202, 2001.
[4] I. Cohen and G. Medioni, "Detecting and tracking moving objects for video surveillance," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, 1999, pp. 23-25.
[5] D. Chai and A. Bouzerdoum, "A bayesian approach to skin color classification in YCbCr color space," in *Proc. IEEE Region Ten Conference(TENCON' 2000)*, vol. 2, no. 1, 2000, pp. 421-424.
[6] K. Kanatani, "Constraints on length and angle," *Computer Vision Graphics Image Process*, vol. 41, pp. 28-42, 1988.

Junghwan Ko received his BS degree in control & instrumentation engineering from the graduate school of Kwangwoon University,Seoul Korea, in 1999, and his MS and PhD degrees in electronic engineering from the graduate school of Kwangwoon University in 2001 and 2005, respectively. In 2005 he joined the faculty of Kwangwoon University, Seoul, Korea, where he is presently a research professor in the 3D display research center 3DRC-ITRC sponsored by the Korea Ministry of Information and Communication. In 2007 he joined the faculty of Inha Technical College, Incheon, Korea, where he is presently an associate professor in the Department of Mechatronics.
His research interests include 3DTV, 3D robot vision, 3D broadcasting, and automatic target tracking and recognition.

Jungsuk Lee received his BS and MS degree in electrical engineering from the graduate school of Kwangwoon University,Seoul Korea and received his PhD degree in control & instrumentation engineering from the graduate school of Kwangwoon University in 2001. He was a senior research engineer at the Agency for Defense Development from 1990 to 1997. In 2002 he joined the faculty of Inha Technical College, where he is presently a professor in the Department of Mechatronics. His research interests include machine vision, intelligence robot control, and stereo vision