A Kalman-Filter-Based Method for Real-Time Visual Tracking of a Moving Object Using Pan and Tilt Platform

B.Torkaman, M.Farrokhi

Abstract— The problem of real time estimating position and orientation of a moving object is an important issue for vision-based control of pan and tilt. This paper presents a vision-based navigation strategy for a pan and tilt platform and a mounted video camera as a visual sensor. For detection of objects, a suitable image processing algorithm is used. Moreover, estimation of the object position is performed using the Kalman filter as an estimator. The proposed method is implemented experimentally to a laboratory-size pan and tilt platform. Experimental results show good target tracking by the proposed method in real-time.

Index Terms— Visual servoing, Vision-based navigation, Target tracking, Estimation, Pan and tilt platform, Kalman filter, Image processing.

1 Introduction

Target tracking is the process of locating a moving object throughout a sequence of sequential frames. The objective of visual target tracking, referred to as "visual servoing", is to estimate the position and velocity of an object that is observed with video camera. Target tracking, in its current day form, is the combination of tracking techniques employed for a multitude of purposes, like air-surveillance or spy satellites used to track the movements of certain targets, many applications including, e.g., robotics, human-machine interfaces, motion capture, and many others [1--4].

Visual sensors provide the necessary information to achieve position of targets. In recent years, different kind of visual servos have been developed [5--8].

Depending on the space in which the visual error is calculated, visual servo controllers can be classified into two categories 1) Pose Based (PB) systems and 2) Image Based (IB) systems. PB systems define the goal of the motion and the current end-effector pose in 3D space and use these values to compute the error in this space. IB systems define the goal and the end-effector pose in terms of some features that can be deduced from the image. Usually they are 2D positions of certain distinguished points visible in the image.

Successful target tracking depends on the extracted useful information about the target state from the observed data. In order to achieve this goal, one needs a useful target model. The white-noisy acceleration model is the simplest model for a target. This model is used when the target manoeuvre is small or random [9], [10]. The other simple model is the Wiener-process acceleration model that is referred to as the constant-acceleration model [11]. Another target model, called the Singer acceleration model, is a standard model for targets with high manoeuvres [12]. In this model, the target acceleration is a zero-mean stationary first-order Markov process. A jerky model is proposed for high manoeuvring targets by Mehrotra and Mahapatra in [13]. In this method, the jerk is modeled as a zero-mean first-order Markov process, in the same way as the Singer acceleration model. Because of limitations of the laboratory pan and tilt movement, the Wiener-process acceleration model, that is referred to as the constant-acceleration model, and the Kalman filter for real-time estimation of the target position and velocity is considered in this paper.

This paper is organized as follows. Section 2 describes the pan and tilt platform and the target tracking algorithm. Section 3 explains the image processing algorithms to detect target to obtain the target position and then using pinhole model of camera to achieve absolute meter position of the target. Section 4 presents the Kalman filter algorithm. Section 5 shows experimental results, followed by conclusion in Section 6.

- *Bahareh Torkaman, Department of Electrical Engineering, Iran University of Science and Technology, Iran, Tehran.*
  E-mail: bahareh.torkaman@gmail.com
- *Mohammad Farrokhi, Department of Electrical Engineering, Iran University of Science and Technology, Iran, Tehran.*
  E-mail: farrokhi@iust.ac.ir

2 Pan and Tilt Platform

Pan and tilt is a 2 degrees-of-freedom platform. The goal of controlling the pan and tilt is to keep the target in camera's field-of-view and try its best to keep the target at the center of camera's field-of-view to avoid lost of the target. In Other words, the pan and tilt platform rotates its direction in a way such that the object is projected at the center of the image. The camera is mounted on a pan and tilt platform, which is actuated by two electric motors. Two Step motors are driven by a motion card installed in a PC. Like many commercial motion cards, the PID control gains are factory set, balancing transient response with minimal overshoot. When designing a motion controller for the pan and tilt, it is assumed that an angular velocity controller is available that can control the angular velocity of the platform.

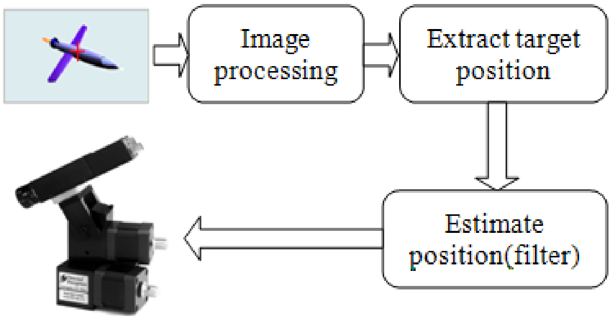


Fig. 1. Block diagram of the visual servoing loop.

In this way, there is no access to the voltage applied to the servo motors of the platform, rather only set points for the angular position, velocity and acceleration of the system can be given. Since the target is moving with high velocities, these set points must be estimated by some appropriate algorithms, which will be explained in Section 4.
The block-diagram of the visual servoing loop is depicted in Fig. 1. When the target is identified as a positive detection, its position is sent to an algorithm for the estimation of the new pan and tilt position. The pan and tilt system is moved using the extracted information from a position estimator, which is a Kalman filter is this paper.

3 Target Detection

In order to detect an object in the image acquired by a camera, first the size of the image needs to be reduced. To reduce the size of the image, the margins of the image must be cropped. The result is a centroid area of the image. Then, the image is resized with a factor of 50%. After binerization, some image processing is employed for noise reduction of the images. Many different approaches toward noise reduction are known, including optimal linear filtering, nonlinear filtering, scale-space processing, and Bayesian techniques [14]. In this work, a median filter is used. Median filtering is a nonlinear operation often used in image processing to reduce "salt and pepper" noises. The overall goal of the image processing module is to determine the coordinates of the target from a noised-free image. The MATLAB function developed for this purpose takes the individual frames from camera as input and outputs the x-y coordinate of the centroid of the image.

Fig. 2. Show a frame of the image and the result of applying the aforementioned image processing algorithm. In this image, the center of the target is labeled with a plus sign.

The result of this step is the target position in the image plane. For the pan and tilt movement, the angular position of the target is required. Using the pinhole model of the camera, the metric position of the target can be achieved. Fig. 3 shows the pinhole camera model.

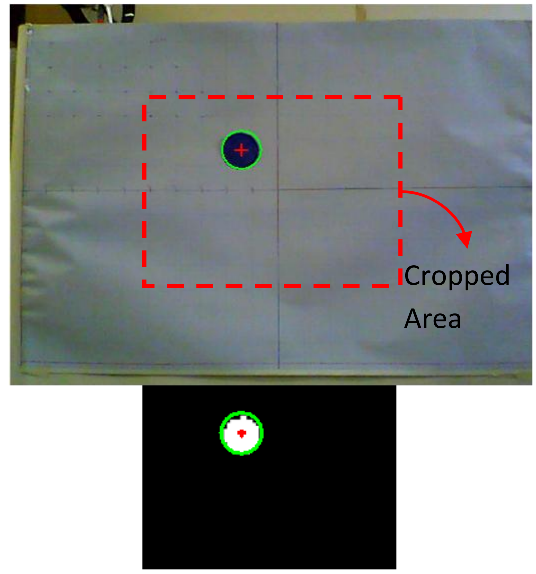


Fig. 2. Result of applying image-processing algorithms to an image frame.

Based on this figure, the Pinhole equations can be calculated as

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \Rightarrow \begin{pmatrix} fX \\ fY \\ Z \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (1)$$

$$\begin{aligned} u_{(\text{pixel})} &= f\,\frac{x}{z} \Rightarrow \frac{u}{f} = \frac{x}{z} \\ v_{(\text{pixel})} &= f\,\frac{y}{z} \Rightarrow \frac{v}{f} = \frac{y}{z} \end{aligned} \quad (2)$$

where (x, y, z) and $(u_{(\text{pixel})}, v_{(\text{pixel})})$ are the target position in the metric coordinate and in the image plane, respectively and f is the focal length of the camera[15]. Using the focal length of the experimental camera employed in this paper, it gives

$$\begin{aligned} f_x &= 944(pixel) = 1.06(mm) \\ f_y &= 928(pixel) = 1.07(mm) \end{aligned} \quad (3)$$

Here we use $f = 1.065(mm)$ for both axis.

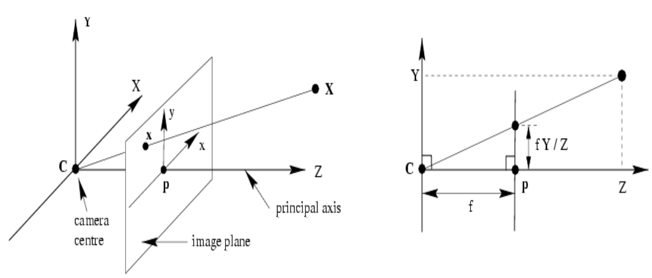


Fig. 3. Pinhole camera model.

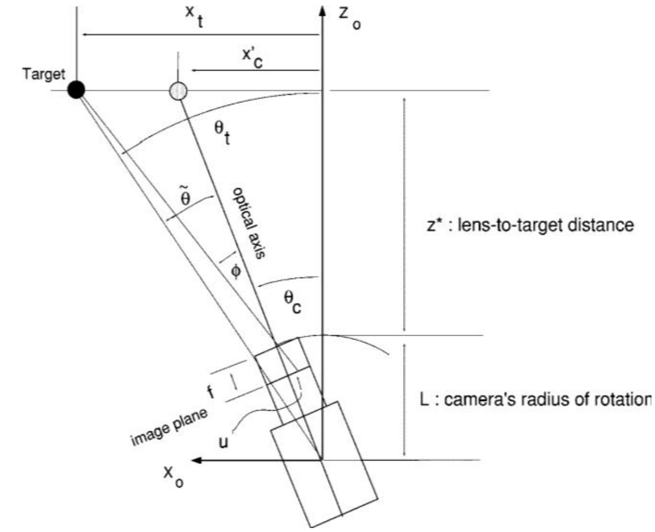


Fig. 4. Geometry of target coordinate.

Fig. 4. show geometry of the target coordinates. Base on this figure, the required pan and tilt angle can be calculated as follow:

$$\theta_{pan} = \arctan\left(\frac{x}{z^{*}+L}\right) = \arctan\left(\frac{u}{f}\right)$$
$$\theta_{tilt} = \arctan\left(\frac{y}{z^{*}+L}\right) = \arctan\left(\frac{v}{f}\right) \quad (4)$$

Then, the scenario of tracking without any estimation filter is as follow:

- Acquire one frame,
- Use the image processing algorithms to detect the target in the image sequence,
- Find the centroid of the target,
- Use (2) to find the x and y position of the target,
- Use (4) to calculate the desired angular position of the pan and tilt,
- Move the pan and tilt platform to the desired position

4 KALMAN FILTER

In target tracking applications, the most popular methods for estimating target positions is the Kalman filter. The Kalman filter assumes that the dynamics of the target can be modeled and that the noise affecting the target dynamics and the sensor data is stationary with zero mean. The Kalman filter is a recursive estimator. This means that only the estimated states from the previous time step and the current measurements are needed to estimate the current states [16], [17].

The state model for the target motion is defined by the following vector-matrix equations [18]:

$$x_{\mathbf{k}} = \mathbf{A}x_{k-1} + \mathbf{B}u_k + w_{k-1} \quad (5)$$
$$z_{\mathbf{k}} = \mathbf{H}x_k + v_k \quad (6)$$

where $x = [x \quad y \quad \dot{x} \quad \dot{y}]^T$ is the state vector, y is the output, and w and v are white noises denoting the process and the measurement noises, respectively, all with appropriate dimensions. Moreover, the transition matrix and the observation vector are

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & T & 0 \\ 0 & 1 & 0 & T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad (7)$$

The equations for the Kalman filter fall into two groups: the time update equations and the measurement update equations. The Time Update or “Predict” step projects the current state estimate ahead in time to obtain ‘a priori’ estimate for the next time step. The Measurement Update or “Correct” step adjusts the projected estimate by an actual measurement at that time. The measurement update equations are responsible for the feedback. for incorporating a new measurement into the a priori estimate to obtain an improved a posteriori estimate.

The time update equations are:

$$\hat{x}_k^{-} = \mathbf{A}\hat{x}_{k-1} + \mathbf{B}\mathbf{u}_\mathbf{k} \quad (8)$$
$$\overline{\mathbf{P}}_\mathbf{k} = \mathbf{A}\mathbf{P}_{\mathbf{k-1}}\mathbf{A}^\mathbf{T} + \mathbf{Q} \quad (9)$$

And the measurement update equations are

$$\mathbf{K}_\mathbf{k} = \overline{\mathbf{P}}_\mathbf{k}\mathbf{H}^\mathbf{T}\left(\mathbf{H}\mathbf{P}_\mathbf{k}^{-}\mathbf{H}^\mathbf{T} + \mathbf{R}\right)^{-1} \quad (10)$$
$$\hat{\mathbf{x}}_\mathbf{k} = \hat{\mathbf{x}}_\mathbf{k}^{-} + \mathbf{K}_\mathbf{k}\left(\mathbf{z}_\mathbf{k} - \mathbf{H}\hat{\mathbf{x}}_\mathbf{k}^{-}\right)\mathbf{P}_\mathbf{k}^{-} \quad (11)$$
$$\mathbf{P}_\mathbf{k} = \left(\mathbf{I} - \mathbf{K}_\mathbf{k}\mathbf{H}\right)\mathbf{P}_\mathbf{k}^{-} \quad (12)$$

Feedback cycle of the Kalman filter is shown in Fig. 5.

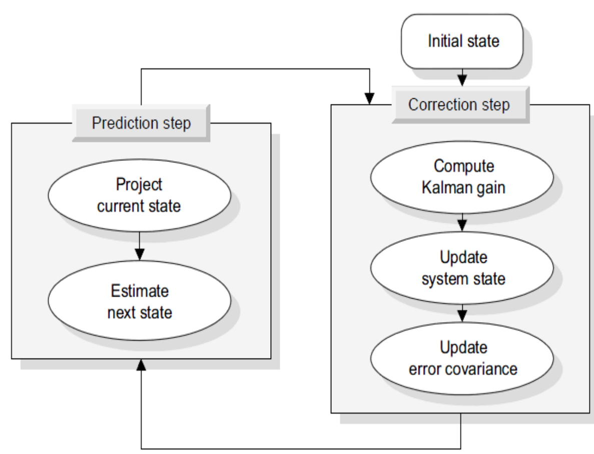


Fig. 5. Feedback cycle of kalman filter.

5 Experimental Result

The pan and tilt platform PTU-D46-17.5made by FLIR Company is used in this paper for target tracking (Fig. 6). The pan and tilt angle limitations are $\pm 150^{\circ}$ for the pan axis and -47° to $+31^{\circ}$ for the tilt axis. Maximum speed of both axis is $300^{\circ}/\mathrm{sec}$. In this project, the target is a dark circle, which is generated using MATLAB and is reflected by a video projector on a screen. The target is moves rectangular and circular paths. A white noise with 10% amplitude of the main signal is added to the measurements. Figs. 7 to 9 show the performance of the closed-loop system for tracking a target moving with constant speed in a rectangular path. Fig. 7 illustrates the target tracking path in the x - y plane with and without the Kalman filter. Figs. 8 and 9 illustrate the x and y positions. Fig. 10 shows the angular position of the pan and tilt. From this figure, the range of the pan and tilt axis movement is about $\pm 17^{\circ}$ for the pan axis and -20° to -2° for the tilt axis. As this figure shows, during the first 5 sec, the tracking error is relatively high. This is due to the fact that at the starting point, the target position is out of the desired path.

Fig. 6. PTU-D46-17.5.

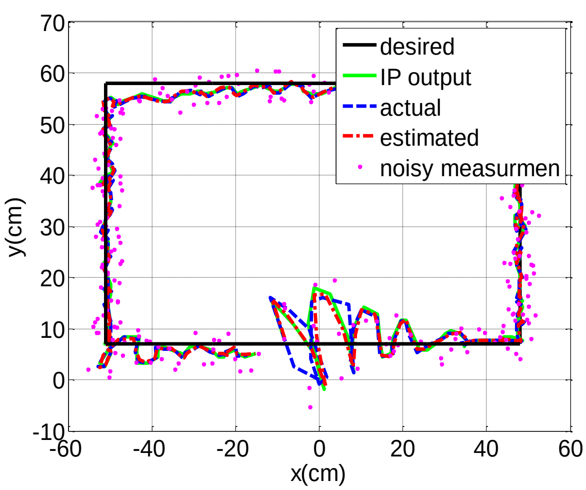


Fig. 7. Target tracking in rectangle path.

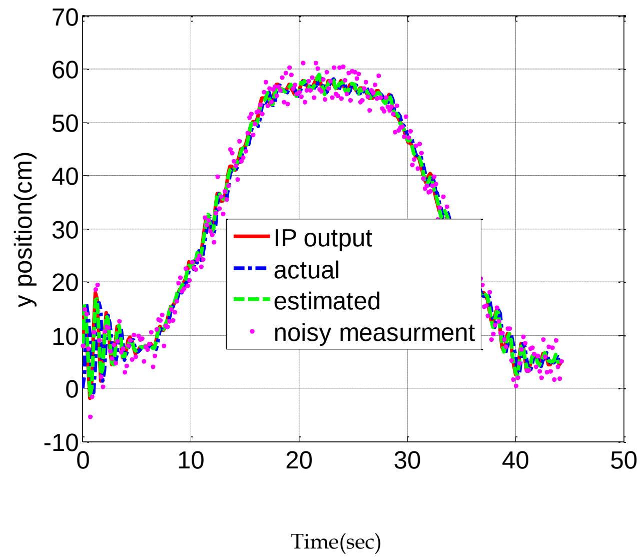


Fig. 8. Target tracking in y axis.

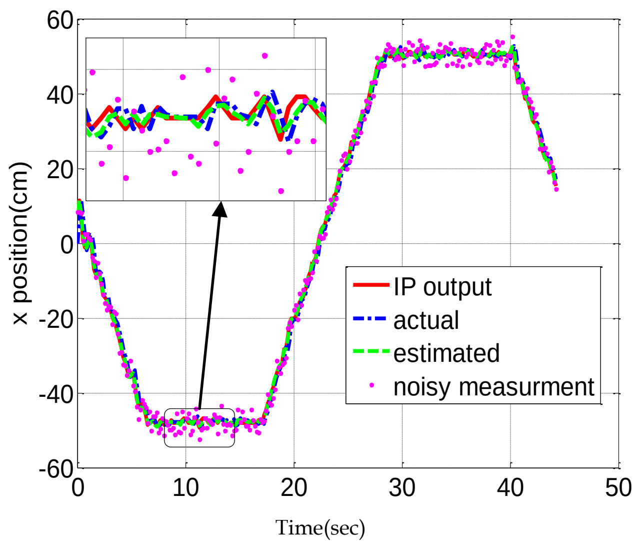


Fig. 9. Target tracking in x axis.

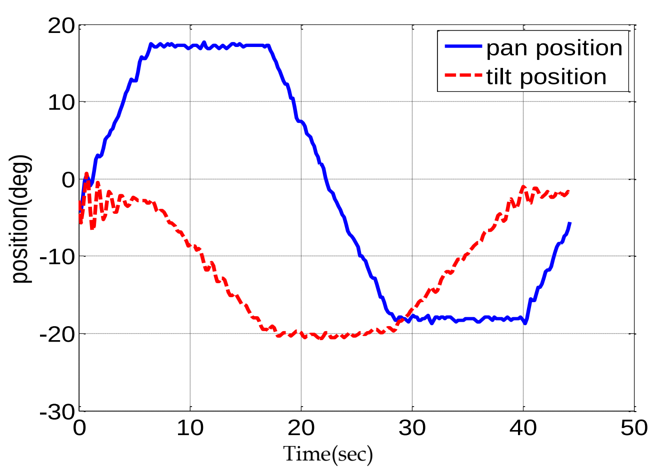


Fig. 10. Angular position of pan and tilt.

The experimental results of tracking a target in a circular path are shown in Figs. 11 to 14. Fig. 11 illustrates the target tracking path in the xy plane with and without the Kalman filter. Fig. 12 shows the total error of tracking for the circular path. The errors of the pan and tilt movements are showed in Fig. 13.

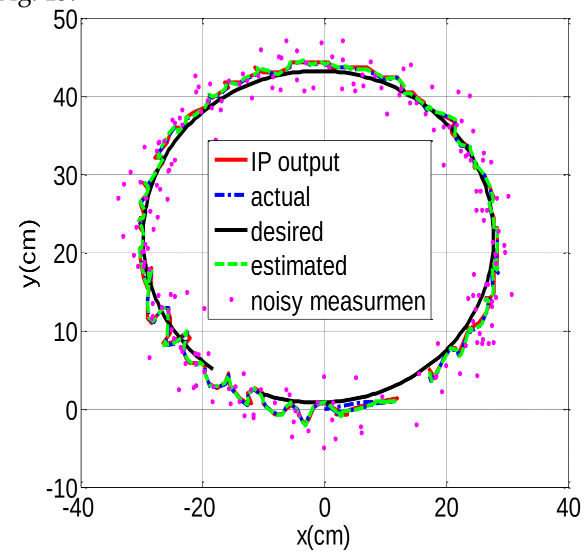
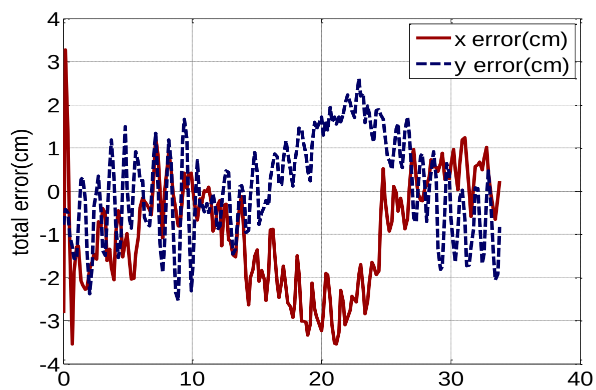
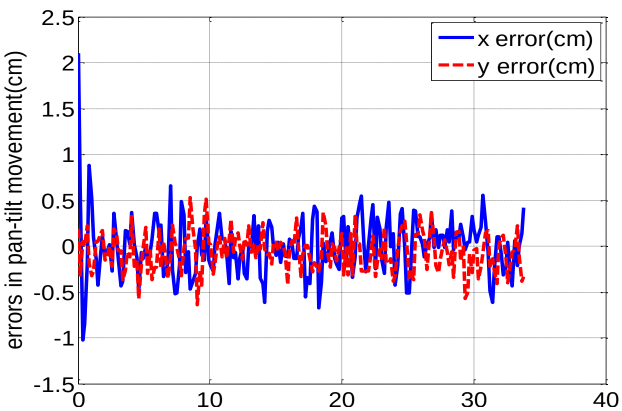


Fig. 11: Target tracking in circle path.



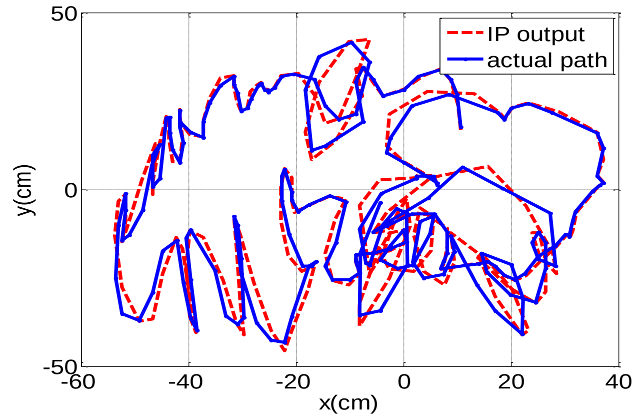
Time(sec)

Fig. 12: Total errors of tracking.



Time(sec)

Fig. 13: Errors in pan and tilt movement.



Time(sec)

Fig. 14: Target tracking in random path.

Fig. 14 shows result of tracking moving target in a random path. As this Figure shows, by using kalman filter, error between the output of image processing algorithms and actual path(that is movement of pan and tilt platform) is negligible.
TABLE I summarizes the Root-Mean-Square Error (RMSE) of target tracking for the rectangular as well as the circular paths in the xy coordinates. As this table shows, by using the Kalman filter, the estimation errors are significantly reduced in both x and y axes. This error is the summation of image processing error and the pan-tilt tracking error. In fact, substantial amount of the error is from the image processing part. In comparison with the rectangular path, the circle path has less error because the circular path is smoother that the rectangular path, which has sharp corners.

TABLE 1
RMSE OF TARGET TRACKING

RMSE(cm)	path / axis	Rectangular path	Circle path
Noisy measurment	X	3.8032	3.1147
	Y	3.3711	2.6701
with Kalman Filter	X	2.0601	1.5741
	Y	2.0674	1.1785

6 CONCLUSION

In this paper, a real-time target tracking algorithm for pan and tilt platform was presented. The proposed method comprises of an image processing scheme suitable for Kalman filter as the target position estimator. The target was a 2D circle that moves in a circular, rectangular, or random path and was created using a video projector device. Experimental results show good performance of the proposed method in clean as well as noisy measurements of the target.

REFERENCES

[1] J. Zhu, Y. Lao, and Y. Zheng, “Object tracking in structured environments for video surveillance applications,” *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 20, no. 2, pp. 529-551, 2010.

[2] J. Courbon, Y. Mezouar, N. Guenard, and P. Martinet, “Vision-based navigation of unmanned aerial vehicles,” *Control Engineering Particle*, vol. 18, no. 7, pp. 789-799, 2010.

[3] H. Hile, A. Liu, G. Borriello, “Visual navigation for mobile devices,” *IEEE Multimedia*, vol. 17, no. 2, pp. 16-25, 2010.

[4] T. Sasaki, D. Brscic, and H. Hashimoto, “Human-obervation-based extraction of path patterns for mobile robot navigation,” *IEEE Trans. on Industrial Electronics*, vol. 57, no. 4, pp. 1401-1410, 2010.

[5] P.K. Allen, A. Timcenko, B. Yoshimi,and P. Michelman, “Automated tracking and grasping of a moving object with a robotic hand-eye system,” *IEEE Trans. on Robotics and Automation,* vol. 9, no. 2, pp. 152_165, 1993.

[6] N. Hollinghurst, R. Cipolla, “Uncalibrated stereo hand eye coordination,” *Image and Vision Computing,* vol. 12, no. 3, pp. 187_192, 1994.

[7] G .D. Hager, W.C. Chang, A.S. Morse, “Robot hand-eye coordination based on stereo vision,” *IEEE Control Systems Magazine* vol. 15, no. 1, pp. 30_39, 1995.

[8] E. Malis, F. Chaumette, S. Boudet, “visual servoing,” *IEEE Trans. on Robotics and Automation,* vol. 15, no. 2, pp. 238_250, 1999.

[9] S. Reece, and S. Roberts, “The near constant acceleration gaussian process kernel for tracking,” *IEEE Trans. Signal Procesing Letters*, vol. 17, no. 8, pp. 707-710, 2010.

[10] H. Goto, Y. Hasegawa, and M. Tanaka, “A robust moving objects detection based on improved gaussian mixture model ,” *IEEE International Conference on Artificial Intelligence and Computational Intelligence*, Sanya, China, pp. 54-58, 2010.

[11] S. Bogatin, K. Foppe, and P. Wasmeier, “Evaluation of linear kalman filter processing geodetic kinematic measurments,” *Measurement Journal*, vol. 41, no. 5, pp. 561-578, 2008.

[12] Z. Fei, H. Wei-jun, and F. Xin-yue, “Tracking application about singer model based on marginalized particle filter,” *The Journal of China Universities of Posts and Telecommunications*, vol. 17, no. 4, pp. 47-51, 2010.

[13] Mehrotra K. and Mahapatra P. R., “A jerk model for tracking highly maneuvering targets,” *IEEE Trans. Aerospace and Electronic Systems*, vol. 33, no. 4, pp. 1094–1105, 1997.

[14] A. Bovik, *The Essential Guide to Video Processing,* Second ed., Academic Press, New York, 2009.

[15] I. R. Stanciu, *Designing Visually Servoed Tracking to Augment Camera Teleoperators*, PHD Thesis, Drexel University, department of Electrical Engineering, Philadelphia, 2004.

[16] D. Simon, “Kalman filtering with state constraints: a survey of linear and nonlinear algorithm,” *IET Control Theory and Applications*, vol. 4, no. 8, pp. 1303-1318, 2010.

[17] F. Janabi, and M. Marey, “A Kalman filter based method for pose estimation in visual servoing,” *IEEE Trans. Robotics*, vol. 26, no. 5, pp. 939-947, 2010.

[18] D. S. Jang, S. W. Jang, and H. Choi, “2D human body tracking with structural Kalman filter,” *Pattern Recognition*, vol. 35, no. 10, pp. 2041-2049, 2002.