Voting based Cue Integration for Visual Servoing

Che-Seung Cho* and Byeong-Mook Chung**

* School of Mechanical Engineering, Yeungnam University, Korea
(Tel : +8253-811-2569; E-mail: okrobo@yumail.ac.kr)
** School of Mechanical Engineering, Yeungnam University, Korea
(Tel : +8253-810-2569; E-mail: bmchung@yu.ac.kr)

Abstract: The robustness and reliability of vision algorithms is the key issue in robotic research and industrial applications. In this paper, the robust real time visual tracking in complex scene is considered. A common approach to increase robustness of a tracking system is to use different models (CAD model etc.) known a priori. Also fusion of multiple features facilitates robust detection and tracking of objects in scenes of realistic complexity. Because voting is a very simple or no model is needed for fusion, voting-based fusion of cues is applied. The approach for this algorithm is tested in a 3D Cartesian robot which tracks a toy vehicle moving along 3D rail, and the Kalman filter is used to estimate the motion parameters, namely the system state vector of moving object with unknown dynamics. Experimental results show that fusion of cues and motion estimation in a tracking system has a robust performance.

Keywords: : Kalman Filter, Visual Tracking, Moving Object, Visual Cue, Cue Integration

1. INTRODUCTION

Object tracking is an important task in robot applications such as navigation of a mobile robot, object recognition, estimate of a target's speed, etc. The reason that a vision system is used very much for object tracking is that it provides useful information about dynamic environment. So it is received more attention from researchers, and the various studies and efforts are performed recently. The overview of a visual servo system is shown in Fig. 1. Generally, intertwined processes are needed for target tracking, composed with tracking and control process. Each of these processes can be studied independently, but the actual implementation must consider the interaction between them to achieve robust performance. The key issue for tracking algorithm is the robustness and reliability. Tracking is responsible for maintaining the target's position, while servo control reduces the error between the current and the desired position of the target. Tracking of a target can be divided into the following subtasks: 1) detection of the target, 2) matching across images in an image stream, and 3) estimation of the motion of the target. If the position and the velocity of the target are achieved during tracking process, they are fed into a control loop.

The tracking can either be performed in image (i.e., image coordinates of the tracked features are estimated) or in world coordinates (a model of the target/camera parameters are used to retrieve the 3-D pose of the tracked features). Image-based servo control uses image coordinates of the features directly in the control loop. If the control is performed in the 3D Cartesian space, it is called position-based servo control. In this paper, position-based servo control is studied. The reported literature on visual tracking and servoing is extensive[1-2].

Robust target detection is often achieved through the use of artificial markers that are easy to segment. An alternative approach is the use of CAD models, for example, as demonstrated by H. Kollnig, et al [3] and Hirzinger, et al. [4]. Such an approach is particularly relevant for tracking of well-known/well-defined objects that can be modeled *a priori*. For general objects of complex shapes, an approach to increased robustness may be an integration of multiple visual cues. Integration methods of visual cues are various. For example, neural network, fuzzy logic, probabilistic fusion, and voting, etc. A lot of work has been reported on fusion of visual cues [5]–[6]. I.Boch[7] shows common operator in fusion theories like probability, fuzzy set, Bayesian decision in classification. Danica K. et al.[8] integrates visual cues using voting and fuzzy logic method to track the

Several authors have applied Kalman filter theory to estimate 3D motion parameters and depth[9-10]. P. Allen[11] demonstrates tracking a toy train which moves along the 2-D rail according to the estimated position from Kalman filter. W. Y. Jae[12] estimates 3-Dmotion and depth of 3-D moving points using of stereo motion sequence and Kalman filter.

Visual servoing requires techniques that are suited for real-time implementation. To achieve this, one often has to resort to simple visual cues.

In this paper, we propose the robust visual tracking algorithm and demonstrate through real experiments in which on x-y Cartesian robot tracks a toy train moving along 3D rails. The target is extracted from complex background using integration of visual cue, such as, color, shape moment, disparity, SSD(Sum of Squared Differences). Voting-based fusion of cues is adapted. In voting, a very simple or no model is used for fusion. Kalman filter is used to estimate the motion parameters : position and velocity of target. And the Cartesian robot continues to track according to the estimated position.

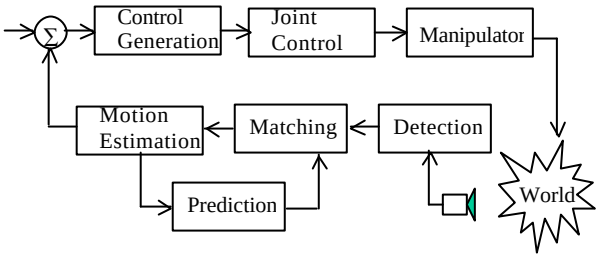


Fig. 1 Major components of a visual servo system

2. VISUAL CUE

The following section presents visual cues used in the experiment. Fig.2 shows the overview of tracking mechanism.

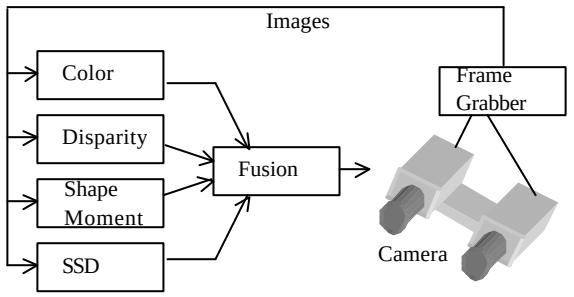


Fig. 2 Overview of tracking mechanism

2.1 Color based segmentation

Color detection is based on the *hue* (H) and *saturation* (S) components of the color histogram values color training was performed off-line, i.e., the *a priori* known color is used to compute its distribution in the – plane. In the segmentation stage, all pixels whose hue and saturation values fall within the set defined during off-line training and whose brightness value is higher than a threshold are assumed to belong to the tracked object.

$$H = \arccos\left[\frac{\frac{1}{2}\left[(R-G)+(R-B)\right]}{\left[\sqrt{((R-G)^2+(R-B)(G-B))}\right]}\right]$$

$$S = 1-\frac{3}{(R+G+B)}\min(R,G,B) \qquad (1)$$

2.2 Shape moment

Moment can be applied usefully to classify the object shape. The feature of Moment has constant values without regard to the extension, reduction, or rotation of the object. Moment m_{ij} is defined as follows.

$$m_{ij} = \sum_x \sum_y x^i y^j f(x,y) \qquad (3)$$

where, i, j represents the order of moment, and x, y is the position value of horizontal and vertical of the image. $\bar{x}, \bar{y}$ is the center of the gravity, and m_{ij} is the central moment.

$$\bar{x} = \frac{m_{10}}{m_{00}} \qquad \bar{y} = \frac{m_{01}}{m_{00}}$$

$$m_{ij} = \sum_x \sum_y (x-\bar{x})^i (y-\bar{y})^j f(x,y) \qquad (4)$$

The length of major and minor axis of the color bar is represented in (5).

$$a \approx \sqrt{2}\sqrt{(m_{20}+m_{02}+\sqrt{((m_{20}-m_{02}))^2+4m_{11}^2))}\,/\,m_{00}}$$

$$b \approx \sqrt{2}\sqrt{(m_{20}+m_{02}-\sqrt{((m_{20}-m_{02}))^2+4m_{11}^2))}\,/\,m_{00}}$$

The ratio of the length between major and minor axis is as follows [9].

$$Axis\ \ ratio = \frac{b}{a} \qquad (5)$$

2.3 Disparity

Fig.3 shows the parallel stereo geometry. m_1, m_2 is the image point, v_1, v_2 is the distance between m_1, m_2 and image center, f is focal length of the camera, Z is the distance between object and camera, and b is the base line between two cameras. Disparity is defined as the difference between v_1 and v_2, and represented in (6). If we know the disparity, we can determine the depth information.

$$d = v_2 - v_1 = \frac{bf}{z} \qquad (6)$$

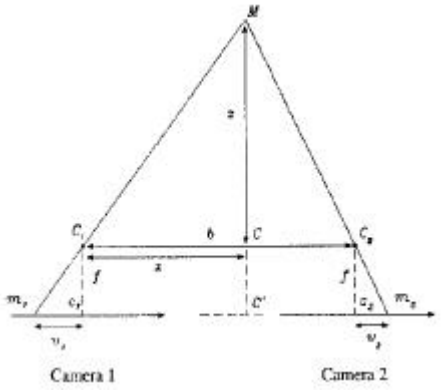


Fig. 3 Parallel stereo camera setup

The fundamental problem of disparity computation is finding the corresponding elements between two or more images. And it is necessary to reduce a calculation time in matching. A lot of work has been reported on this problem. First, the matching problem is reduced to 1-D(x-direction) search by the use of parallel binocular configuration. Namely, the epipolar line is set to parallel with a scanning line to reduce matching scope.

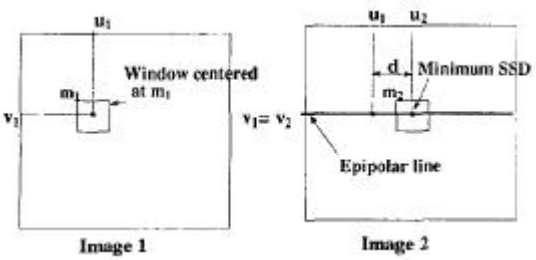


Fig. 4 Intensity based stereo matching

To obtain robust matching between left and right image points without increasing the computational effort, we cover the (16×16) window around the correspondence points. And the coordinate where SSD has a minimum value is selected. From this coordinate, disparity is decided by a difference of x coordinates of two images. The error of SSD is defined as follows.

$$e(u,v,d)=\sum_{m,n}[I_1(u+m,v+n)-I_2(u+m,v+n)]) \quad (7)$$

Where, u,v represent the image position, and m,n represent the window size.

3. INTEGRATION OF VISUAL CUE

Voting has been widely used in machine vision in various forms. The main advantage of voting mechanisms is that they can operate model-free with respect to the individual cues. In probabilistic fusion, a model of the form encodes the relationship between visual cues and particular objects/patterns. In voting, a very simple or no model is used for fusion. A common estimation/classification space Θ is mapped as follows by each cue estimator v_i.

$$v_i:\Theta\rightarrow[0;1] \quad (8)$$

A general class of voting schemes, known as weighted consensus voting, is defined by the following definition.

$$V(q)=\begin{cases}\Lambda(C_1(q),.....\,C_n(q)) & if\ \sum_{i=1}^{n}v_i(q)\geq m\\ 0 & otherwise\end{cases}$$

$$v_i(q)=\begin{cases}1, & if\ \ c_i(q)>0\\ 0, & otherwise\end{cases} \qquad for\ i=1,\ldots,n \quad (9)$$

Where, $v_i(q)$ is voting function, $\Lambda:[0;1]^n\rightarrow[0;1]$ is a function for combining the confidence for each estimator, and n is the number of cue estimators.

4. KALMAN FILTER

In our system, Kalman filter are used to predict 3Dmotion of moving object. The velocity of the target is not constant in practice. However, for simplicity the object velocity is assumed to be constant. State vector x_k is defined as.

$$x_k=[x(k),\dot{x}(k),y(k),\dot{y}(k),z(k),\dot{z}(k)] \quad (10)$$

System model and measurement model are represented in (11) and (12).

$$x_{k+1}=\Phi_k x_k+w_k \quad (11)$$

$$z_k=H_k x_k+v_k \quad (12)$$

where, Φ_k is a (6×6) diagonal matrix of the form and $\Phi_k=diag\{\begin{bmatrix}1 & T\\ 0 & 1\end{bmatrix},....,\begin{bmatrix}1 & T\\ 0 & 1\end{bmatrix}\}$. H_k is output vector and $H_k=I_3$. The observation z is the 3D position from a pair of projection with known stereo camera geometry. Random variables w_k and v_k represent process noise, covariance Q_k and measurement covariance R_k has the system noise and output noise respectively.

$$E[w_k w_i^T]=\begin{cases}Q_k, i=k\\ 0, i\neq k\end{cases} \quad (13)$$

$$E[v_k v_i^T]=\begin{cases}R_k, i=k\\ 0, i\neq k\end{cases} \quad (14)$$

Kalman filter algorithm is composed of the following two parts.

1) Time update equations

$$\hat{x}_{k+1}^-=\Phi_k\hat{x}_k^- \quad (15)$$

$$P_{k+1}^-=\Phi_k P_k\Phi_k^T+Q_k \quad (16)$$

2) Measurement update equations

$$K_k=P_k^-H_k^T(H_kP_k^-H_k^T+R_k)^{-1} \quad (17)$$

$$\hat{x}_k=\hat{x}_k^-+K_k(z_k-H_k\hat{x}_k^-) \quad (18)$$

$$P_k=(I-K_kH_k)P_k^- \quad (19)$$

We define $\hat{x}_k^-$ to be our a priori state estimate at step k given knowledge of the process prior to step , and $\hat{x}_k$ to be our a posteriori state estimate at step k given measurement z_k. K_k in (18) is chosen to be the gain that minimizes the a posteriori error covariance. P_k^- and P_k represent a priori and posteriori estimate error covariance respectively. z_k and $H_k\hat{x}_k^-$ represent actual measurement and a measurement prediction. The difference $z_k-H_k\hat{x}_k^-$ in (18) is called the measurement innovation or the residual. The residual reflects the discrepancy between the predicted measurement and the actual measurement. A residual of zero means that the two are in complete agreement. $K_k(z_k-H_k\hat{x}_k^-)$ updates $\hat{x}_k$ from $\hat{x}_k^-$. Output measurement covariance matrix R_k has been set to $R_k=0.1I_3$, and process noise covariance Q_k has been set set to $Q_k=daig\{20,30,20,30,20,30\}$.

5. EXPERIMENT RESULT

Fig.5 shows an overall system for target tracking. It is composed of Cartesian robot, stereo camera, a toy train, and PC(Pentium-3) with a frame grabber. Stereo camera attached to the robot arm observes a target motion. A toy train attached color bar moves along the 3-D shaped rail. The robot continues to track according to the position vector estimated from the prediction stage of Kalman filtering as shown in Fig.4. The ratio of major and minor axis of the color bar is 0.6(1.5/2.5).

Fig. 5 Experimental Setup

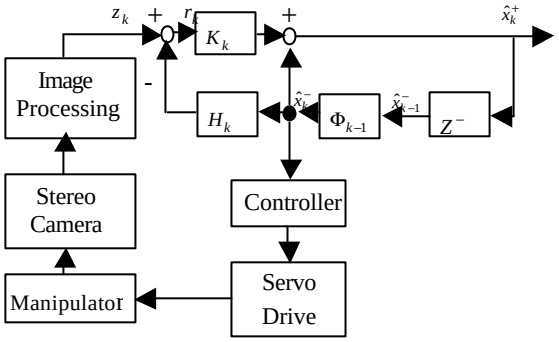
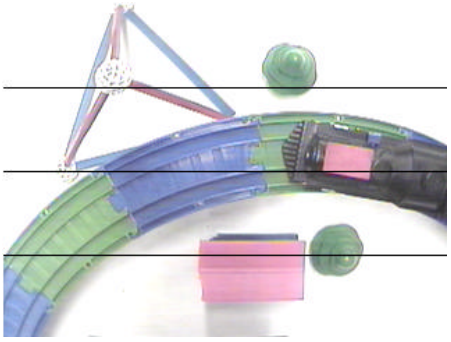
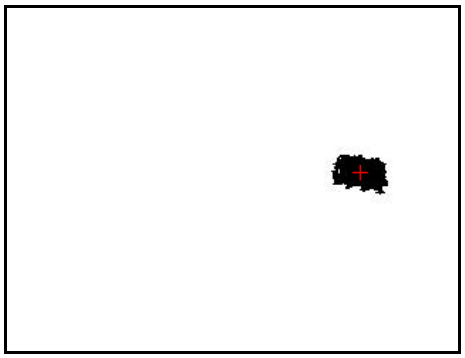


Fig. 6 Block diagram for visual servo



(a) Original Image

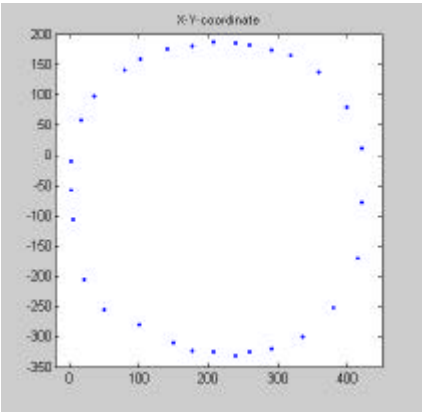


(b) Extracted Target
Fig. 7 Extracted target by cue integration

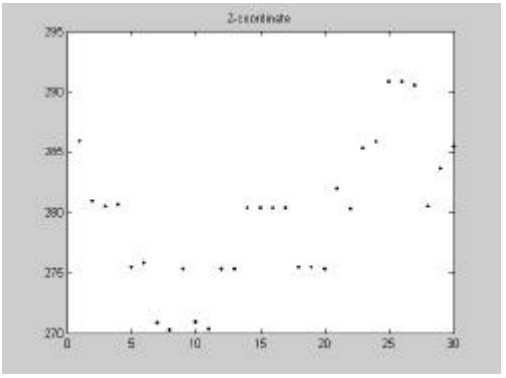
Fig.7 shows overall controller configuration. A robot traces a toy train according to the position presumed from a Kalman filter.

Fig.7(a) shows an example image for moving object in complex background, and Fig.6(b) shows the result of the extracted target from background using cue integration.

In Fig.8 we shows the tracking results for full tracking range. Fig.8(a) shows a x-y tracking position of robot manipulator, and Fig.8(b) shows a tracking position of Z axis for target.



(a) X-Y coordinate



b) Z-coordinate
Fig. 8 Result of tracking position

7. CONCLUSION

We developed a robust tracking algorithm to track a target moving in 3D estimating a position and velocity. The aim of our approach for target tracking is to have the robustness and reliability.

Visual cue was obtained from stereo camera and integrated by voting method. And the target is extracted from integrated cue. For real-time tracking, Kalman filter is used to estimate the motion parameters of the train which moves on the 3-D rail. And the 3-D Cartesian robot can track the target according to the estimated position continuously.

Because the number of visual cue and its complexity restrict a real-time tracking, we need to examine the influence according to the number of visual cue and change of reliability. Experimental results show that fusion of cues and motion estimation in a tracking system has a robust performance.

We expect this work can be applied in many areas of robot vision.

REFERENCES

[1] A. Cretual, "Complex object tracking by visual servoing based on 2D image motion," *Proceedings 14th International Conference on Pattern Recognition*, vol. 2, pp.1251 –1254, 1998.

[2] E. C. Maniere, et al., "Robotic contour following based on visual servoing," *Proceedings of the IEEE/RSJ International Conference on intelligent robotics and system*, pp. 716-722, July, 1993.

[3] H. Kollnig and H. Nagel, "3D pose estimation by directly matching poly-hedral models to gray value gradients," *International Journal of Computer Vision*, Vol. 23, no.3, pp. 282-302, 1997.

[4] G. Hirzinger, et al., "Advanced in robotics: The DLR experience," *International Journal of robotics research*, vol. 18, pp. 1064–1087, Nov. 1999.

[5] H. Borotsching, et al., "A new concept for active fusion in image understanding applying fuzzy set theory," *Proceedings of the Fifth IEEE International Conference on Fuzzy Systems*, vol. 2, pp793 -799, sep. 1996.

[6] B. Parhami, "Voting algorithm," *IEEE Transaction on. Rel.*, Vol.43, no.3, pp. 617-629,1994.

[7] I. Bloch, "Information combination operators for data fusion," *IEEE Transaction on. System. Man and Cybernatics.*, Vol.26, no.1, pp.42-52, 1996.

[8] K. Danica, et al., "Cue integration for visual servoing" *IEEE Transaction on robotics and automation*, Vol.17, No.1, pp. 18-26, 2001.

[9] J. Wang and W. J. Wilson, "3-D relative position and orientation estimation using Kalman filter for robot control," *IEEE Iinternational conference on Robotics and automation*, pp. 2638-2645,1992.

[10] N. P. Papanikolopoulous and P. K. Khosla, "Adaptive robotic visual tracking: Theory and experiments," *IEEE Transaction on Automation and Control*, Vol.38, pp. 429–445, 1993.

[11] P. Allen, "Automated tracking and grasping of a moving object with a robotic hand–eye system," *IEEE Transaction on robotics and automation*, Vol. 9, p. 152, 1993.

[12] W. Y. Jae, et al., "Estimation of depth and 3D motion parameters of moving objects with multiple stereo images by using Kalman filter," *IEEE IECON 21st International Conference on Industrial Electronics, Control, and Instrumentation*, Vol.2, pp.1225 -1230, Nov. 1995.