A visual servoing system for tracking of moving object

Abstract:

The robotic visual servoing system related to robot control and machine vision issue has received many attentions from researchers. The system introduced in this paper consists of a pan/tilt robot with 2 degree of freedom that controls a videocamera. The aim of system is to move robot in such a way that the image of an unknown moving object attains the center of camera. We propose a kinematic model that relates the position of object's centroid in the image plane with the pan and tilt rotation angles. In order to maintain the object in the field of vision of videocamera, a prediction algorithm for position and velocity is used. This paper also briefly describes image processing, auto detection and recognition of moving objects. The experiment with VICON system presents good results.

1. INTRODUCTION

Applications of visual servoing may have many attentions in the industry and military related fields. Visual servoing, can be classified into 2 categories, belongs to system's structure, the fixed-camera system and eye-in-hand system.

In fixed-camera robotic systems, multiple cameras fixed in the world–coordinate frame capture images of both the robot and its environment. The objective of this approach is to make the robot move in such a way that its end-effector reaches a desired object [3].

In the camera-in-hand configuration, a camera is mounted on the robot, which supplies visual information of the environment. The objective of this approach is to move the manipulator in such a way that the projection of either a moving or a static object be always at a desired location in the image captured by the camera [4, 8, 10].

This presentation deals with camera-in-hand robot control. The camera will capture image of environment in the field of view, then digitalizes the image and the information about image feature is used in the feedback control loop, to control the joints of manipulator. For tracking moving object, we use a 2 DOF, pan and tilt robot. Computer is embedded in the control loop to calculate image features, recognize the object, control and predict the target's location. Figure 1 presents the structure of a visual control system.

Vision systems can also be classified as the static scene system and the dynamic scene system according to the characteristics of the obtained scene from the vision sensors.

In the dynamic scene systems, the characteristics of scene are changing. The study of the dynamic scene systems is relatively fallen behind that of the static scene system since the dynamic scene systems are largely influenced by image processing speed as well as accuracy of the image analysis [4]. Therefore, many algorithms to achieve the real time processing of the dynamic scene system through the reduction of the image processing time and the sufficient consideration of the disturbance factor have been suggested [5, 6, 9, 10].

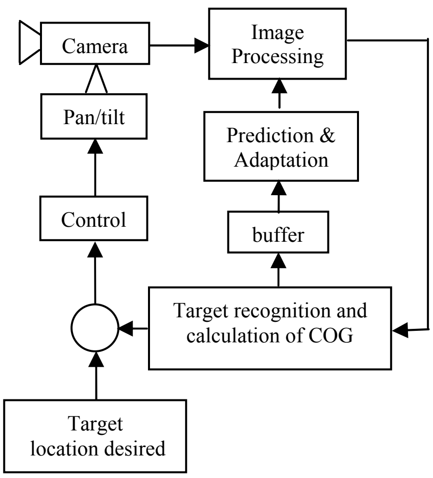


Figure 1. Diagram of a tracking moving system.

Another challenge of visual servoing system is how to recognize an object. A robot may be expected to face variety of objects, but one of them is important to it while others not. Therefore, we have expanded the abilities of our system to include a mechanism for recognition the type of detected object. To obtain the information about the position of the moving object in these dynamic scene systems, the detection of the feature points is important. The hole points, corner points, centroid of the object are used as feature points and the dynamic characteristics of the moving object can be obtained from these feature point analysis. The centroid of the object is easily obtained by the computation of first moment of the entire image, but it is not acceptable in case of the dynamic scene system because computation load is increased. Kalman filter is used to filter noises, predict the target's location of the next frame, to reduce time of image processing.

VICON (*VIsual CONtrol*) system was designed as seen in the figure 1 and experiments in the laboratory gave good results. This paper is organized in the following way: first part describes the tracking moving objects system. The second part presents the system dynamics. The Kalman filter is used for the prediction of the target's position and velocity. The fourth section is about the target recognition algorithm. Experiment of VICON system is shown in the section 5 and the last is the conclusion remarks.

2. VICON TRACKING MOVING OBJECT SYSTEM

Figure 2 shows the tracking moving object system presented in this paper. It includes a camera mounted on a 2 DOF pan-tilt robot.

Figure 2. VICON Pan/tilt-camera system.

The visual servoing consists of 2 main parts. The first is image processing to calculate and recognize the target's location. The second is control section to keep the target's image in the middle of the image plane.

	Precision deg/step	Range deg	Max velocity deg/s
Pan	0.0129^0	$-159^0 - +158^0$	60
Tilt	0.0129^0	$-30^0 - +41^0$	60

Table 1. Main characteristics of the robot.

Basic parameters of the robot is presented in the Table 1. Because the control signal based on image processing, it is necessary to define the kinematic model between the image plane and robot's coordinates. The image preprocessing, target recognition and estimation of position and control of VICON system are discussed below.

3. KINEMATIC MODEL

Let variables x_p and y_p be the target's centroid position in the image plane (IP), and variables ϕ and θ be the pan/tilt angles of the robot.

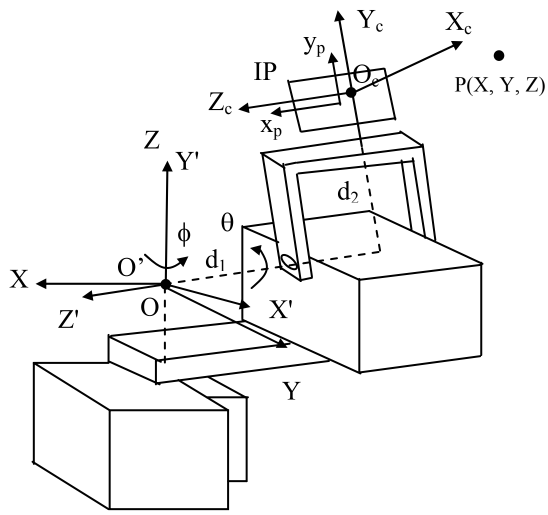


Figure 3. Kinematic model geometry.

In Fig. 3 we can see that ϕ and θ represent the amount to rotate about the OZ and OZ', axes respectively, so that, after the movement, the target's centroid P'(x_p, y_p) located at P(X, Y, Z) be projected at the center of the IP.

The kinematics model will be constructed in two parts. First is shown that it is possible to build an approximate model that relates the control ϕ and θ and observable variables x_p and y_p. Afterwards, using that approximation, the kinematic model is constructed.

3.1. Geometric Approximations.

Suppose that OXYZ is presented the base coordinate of the system and $OX_cY_cZ_c$ is the coordinate fixed to the camera. P(X, Y, Z) is the coordinate of the target in the base coordinate. The projection to the IP is P'(x_p, y_p). From Fig 3, we can deduce the Fig 4, which presents the amount of rotation ϕ angle of X axis so that the target's centroid lies on O_cX_c.

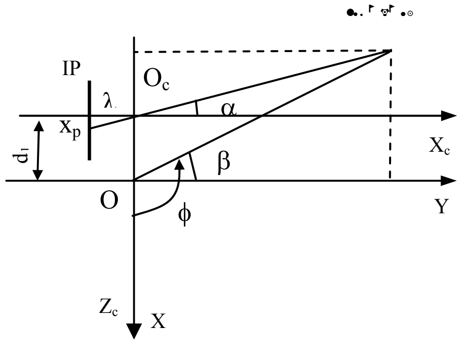


Figure 4. Rotation about OZ axis projection.

Define β is the angle between OP and OY axis as in the Fig 4, so that: $\phi = \beta + 90^0$. We can see that:

$$tan(\phi) = -\frac{Y}{X} \tag{1}$$

where '-' means the algebraic sign.

$$tan(\beta) = -\frac{X}{Y}$$

and

$$tan(\alpha) = -\frac{Z_c}{X_c} = -\frac{X + d_1}{Y} = \frac{x_p}{\lambda} \tag{2}$$

where X and Y represent the target's centroid position in the robotic arm reference system, d_1 is the distance from the OY axis to the O_cX_c axis, and λ is the focal length.

From (1) we can deduce that in order to measure ϕ or $\phi = \beta + 90^0$, we need to know X and Y, which are not observable nor measurable variables, in the case of using a static camera. From (2) we know that by x_P, an observable variable, and by λ, a measurable variable, α can be calculated.

If the error of working with the approximation $\beta \approx \alpha$ is small enough, then α is a valid approximation of β and can be computed from observable and measurable variables.

Let β_e is the error when working with the approximation $\beta \approx \alpha$ then, $\beta = \alpha + \beta_e$. We can write:

$$tan(\beta) = tan(\alpha + \beta_e) = \frac{tan(\alpha) + tan(\beta_e)}{1 - tan(\alpha)tan(\beta_e)} \tag{3}$$

from (1), (2), (3) we get:

$$\beta_e = arctan\left(\frac{d_1 Y}{X^2 + d_1 X + Y^2}\right) \tag{4}$$

From (4) we can see that the error tends to zero in the following cases:

- $\lim_{Y \to 0} \beta_e = 0$. When the control system follows accurately the moving target.
- $\lim_{d1 \to 0} \beta_e = 0$. When the origins O and O_c are the same point.
- $\lim_{X \to \infty} \beta_e = 0$. When the coordinate X is big compared to d_1 and Y, that is $X >> d_1 Y$.

Once know β, we can calculate $\phi = \beta + 90^0$. The reasoning for θ is similar to the one for ϕ. So we can conclude that, when calculating ϕ and θ, it is admissible assuming that O and O_c are the same point. With this approximation we can compute ϕ and θ from observable x_P, y_P and measurable variable λ.

3.2. Simplified Kinematics Model.

In order to find a kinematics model that relates x_p and y_p with ϕ and θ, first step is to find an homogeneous transformation matrix $R_{O_c}^{O}$ when rotating the camera ϕ and θ angles.

The reference coordinates for robot system depicted in the fig 3, based on David – Hetenberg method.

Transformation matrix between O'X'Y'Z' and OXYZ:

$$R_{O'}^{O} = Rotate(Z, \varphi)Rotate(X', 90^0)$$

$$R_{O'}^{O} = \begin{pmatrix} cos(\phi) & 0 & sin(\phi) & 0 \\ sin(\phi) & 0 & -cos(\phi) & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Transformation matrix between O_cXYZ and O'X'Y'Z' when camera rotate around O'Z' angle θ:

$$R_{O_C}^{O'} = \begin{pmatrix} cos(\theta) & -sin(\theta) & 0 & 0 \\ sin(\theta) & cos(\theta) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Homogeneous transformation matrix that models rotations ϕ and θ angles is $R_{O_C}^{O} = R_{O'}^{O} R_{O_C}^{O'}$. So we have:

$$R_{O}^{O_c} = \left(R_{O_C}^{O}\right)^{-1} = \left(R_{O_C}^{O}\right)^{T}$$

then,

$$R_{O}^{O_c} = \begin{pmatrix} cos(\phi)cos(\theta) & sin(\phi)cos(\theta) & sin(\theta) & 0 \\ -cos(\phi)sin(\theta) & -sin(\phi)sin(\theta) & cos(\theta) & 0 \\ sin(\phi) & -cos(\phi) & & \\ 0 & 0 & & \end{pmatrix}$$

Figure 3 and 4 show that the points P(X, Y, Z) are transformed in the middle of the image plane after the ϕ and θ rotations. So once known $R_{O}^{O_c}$, the following calculation can be written:

$$\begin{pmatrix} c \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} cos(\phi)cos(\theta) & sin(\phi)cos(\theta) & sin(\theta) & 0 \\ -cos(\phi)sin(\theta) & -sin(\phi)sin(\theta) & cos(\theta) & 0 \\ sin(\phi) & -cos(\phi) & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{5}$$

Where (c, 0, 0, 1) and (X, Y, Z, 1) are the homogeneous coordinates of point P in the Oc and O respectively.

Solve (5) we get:

$$Xsin(\phi) - Ycos(\phi) = 0 \tag{6}$$
$$-Xcos(\phi)sin(\theta) - Ysin(\phi)sin(\theta) + Zcos(\theta) = 0$$

so the equation searched are:

$$\phi = arctan\left(\frac{Y}{X}\right)$$
$$\theta = -arctan\left(\frac{Z}{Y}cos(\phi)\right) \quad (7)$$

from approximately (1) and (2) we have:

$$\phi = -arctan\left(\frac{\lambda}{x_p}\right)$$
$$\theta = arctan\left(\frac{y_p}{\lambda}cos(\phi)\right) \quad (8)$$

Once we established the equations relating the current coordinates of the moving target's centroid in the digital image with the two joint variables of the robotic arm, we can drive robot so that the target's centroid reaches to the middle of the image. Before presenting about image processing, we introduce an method to predict the target's location, target's speed so that target's centroid maintain in the middle of the image frame, despite moving target, and this method also reduce time consumption.

4. PREDICTION OF TARGET'S LOCATION IN THE DISTURBANCE ENVIRONMENT BY KALMAN FILTER

To track unknown moving object it is necessary to predict the location and speed in the next image frame. Kalman filter is used to estimate position using image information at the current time. The Kalman filter is defined as observer for the linear discrete system. It has simple filter structure and good convergence and ability to remove high frequency noise [9, 10].
The target prediction problem can be described by the following discrete time state and output equations:

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{G}\boldsymbol{\omega}_k$$
$$\mathbf{y}_{k+1} = \mathbf{C}\mathbf{x}_k + \zeta_k$$

where **A** is the system matrix, **G** is the control input matrix, **C** is the output matrix, $\mathbf{x}\in \boldsymbol{R}^n$ is the state vector and $\mathbf{y}\in \boldsymbol{R}^m$ is the output vector, 'k' and '$k+1$' are the vector indexes at the current time and the next time.
With the estimation of image feature, we have:

$$\mathbf{x}_k = [u_k \ \dot{u}_k \ v_k \ \dot{v}_k]^T ;$$

$$\mathbf{y}_k = [u_k \ v_k]^T ; u_k = x_p(k), v_k = y_p(k)$$

where (u_k, v_k) and $(\dot{u}_k, \dot{v}_k)$ are the position and velocity of the target in the image plane at time k.

The noise ω_k represents disturbance noise or modeling inaccuracy and is assumed stationary white noise process with zero mean and known covariance matrix **Q**. Measurement noise ξ_k is due to sensor inaccuracy and also assumed white noise process with zero mean and known covariance matrix **R**. Also it is assumed that $\mathbf{Q} \geq \mathbf{0}$, $\mathbf{R} > \mathbf{0}$ and both are positive definite symmetric matrices.

We assume that velocity of moving target during sample time T is constant. The state-transition matrix is:

$$\mathbf{A} = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad \mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$\mathbf{G} = \mathbf{I}$$

The Kalman filter structure using in VICON system is:

<u>Time update step</u>

$$\hat{x}_{k+1}^- = A\hat{x}_k$$
$$P_{k+1}^- = AP_kA^T + GQG^T$$

<u>Kalman filter gain</u>

$$K_{k+1} = P_{k+1}^- C^T [CP_{k+1}^- C^T + R]^{-1}$$

<u>Measurement update step</u>

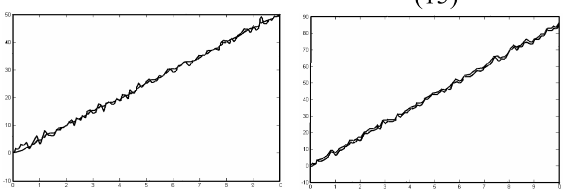
$$\hat{x}_{k+1} = \hat{x}_{k+1}^- + K_{k+1}(y_{k+1} - C\hat{x}_{k+1}^-)$$
$$P_{k+1} = P_{k+1}^- - K_{k+1}CP_{k+1}^-$$

<u>Initial state</u>

$$x_0 \sim (\bar{x}, P_0)$$

where **P** is the error covariance matrix, that denotes estimated state vector, superscript "-" denotes priori value before the measurement update step and subscript k denotes time step.
The output of the Kalman filter gives us the position and velocity of the image feature.
Simulation of Kalman filter presents in figure 5, when target's image motion is linear.

(15)



a) Prediction of u(kT) b) Prediction of v(kT)

Figure 5. Result of Kalman filter.

Selection of the statistical model of Kalman filter largely influences on the performance of the tracking system. Also, if the velocity of moving object is high, and changes quickly, the assumption of constant velocity during sampling time is not valid, then the tracking may fail. We need a systematical approach to the selection of the covariance models such as adaptive approach.

5. RECOGNITION AND TRACKING OF A MOVING OBJECT

Recognition of a moving object is an important task in the tracking system. The aim of this task is to separate the object of interest, determine its location in the current image frame.
There is a wide variety of techniques that could be used for the identification of whether a pixel is part of the object or scene. For example, we could possess a model of known object (such as gray, shape …) and attempt to fit this model to locations within an image. However, identification schemes that are computationally intensive may not be able to complete detection in real-time. Our effective method to process image and recognize an object is presented as figure 6.

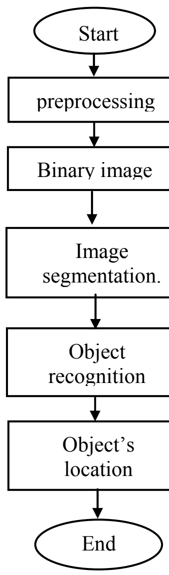
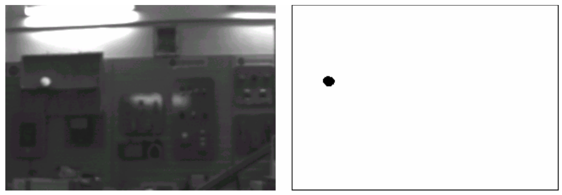


Figure 6. Image processing diagram.

We consider the uninterested objects (such as environment objects) tend to be displayed by pixels whose intensities are constant or very slowly changing over time, while objects of interest (such as moving object to track) tend to be located where pixel intensities have recently changed. Thus, a comparison between pixels in the current frame and threshold, in which pixels larger than threshold will have 1 value, otherwise 0. If the threshold is too large, then pixels belong to objects could be omitted. If the threshold is small, then slight changes in the environment will cause many pixels that don't necessarily belong to important objects. Reasonable threshold can be chosen after using filters and other preprocessing.
Once a binary image has been obtained, we can do analyzing of the objects. We do objects segmentation by the Sequential Labeling Algorithm. Because the algorithm creates segments of hundred objects, several next steps is made through the segment data structure to remove almost undesirable segments. The invariant moments method and Bayes method are the useful methods in the image processing. The centroid of the object is easily obtained by the computation of first moment of the image.
The process is not made in whole the image. We use Kalman filter result and some adaptive computation to reduce the size of processing window. Figure 7 presents the result of image processing and recognition.

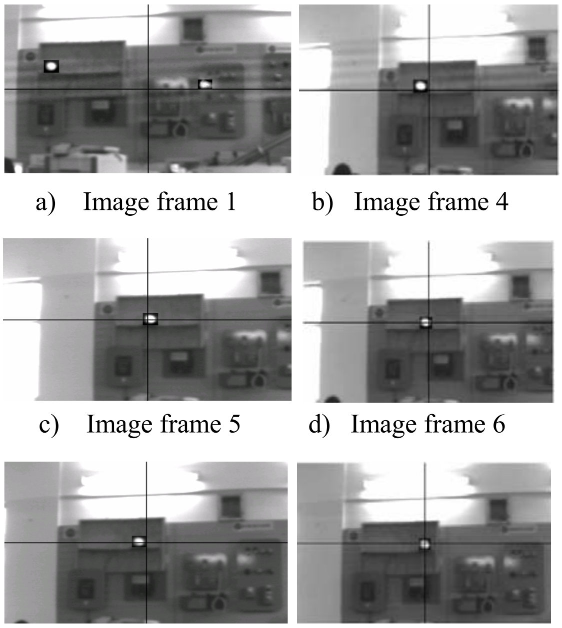


a) initial image b) result image

Figure 7. Recognition result of a test image

6. EXPRIMENTAL RESULTS

The VICON system in this experiment is presented in the following. It consists of a 2DOF robot type pan and tilt by DPerception, as introduced in Figure 1, a camera CV-M50, monochrome CCD Camera by JAI Corporation, with focal length λ = 8mm, with speed of 25 frame/s. The video signal is transmitted from the camera to the PC104 FrameLocker board where it is digitized, low pass filtered and sampled to a resolution of 320 $\times$ 200. All the image-related computations and control operation are performed on a Touchscreeen computer, 800MHz, by Nagasaki IPC Technology. The software of control and image processing is implemented in C in the DOS environment.
To illustrate, we tested the VICON system in a room with a moving object far from system about 7m. Moving object with the speed of about 0.5m/s. For the experiment, the objective of the system is to move camera so that the image projection of moving object remain at the center of the camera's image plane. Figure 8 below shows the tracking results for unknown moving object.
From the experiment, we see that the robot control was shown to perform all of its activities in 50ms per iteration. The VICON tracking system tracks the target object with errors less than $\pm$10 pixels.
This experiment show that when the target move with high speed then image projection of target will escape from the image window and the tracking may be fail. Some image processing algorithm and recognition still need to improve the performance of the system.



a) Image frame 1 b) Image frame 4

c) Image frame 5 d) Image frame 6

e) Image frame 7 f) Image frame 8

Figure 8. Sequence image of tracking an moving object.

7. CONCLUSION

In this paper, we have briefly presented a system for automatic tracking of moving target using computer vision. The system consists of a wrist with two degrees of freedom that drives a camera. Kinematics of overall was given. We used the Kalman filter for prediction of the position of the target's centroid. It also describes the performance of the system through computer simulation and experiments. The tracking system tracked the moving target with time delay of 50msec and errors less than ±10 pixels. Nevertheless, the system reveals a limit to deal with the case of tracking fast moving object. Another study in the future will improve tracking performance of the VICON.

8. REFERENCE.

1. J.Somlo, B.Lantos, P.T.Cat. *Advanced robot control*. Akademiai Kiado press, 1997.
2. Wolfram Stadler. *Analytical robotics and mechatronics*. McGraw-Hill, Inc press, 1995.
3. D. Maravall, L. Baumela. *Adaptive control of a video camera for the automatic detection and tracking of mobiles*. Universidad Politecnica de Madrid
4. Charles A. Rechards, Nicholaos Papanikolopoulos. *The automatic detection and visual tracking of moving objects by eye-in hand robotics system*.
5. Bharat N. Shah, Paul.Y.Oh. *Biomimetic Visual Servoing Using A Partitioned Control Scheme*.Drexel University, PRISM Lab Philadelphia PA USA 19104.
6. Apostolos Dailianas, Robert B. Allen, Paul England. *Comparison of automatic video segmentation algorithms.* Department of Computer Science, Columbia University.
7. Halina Kwa_nicka and Bartosz Wawrzyniak. *License plate localization and recognition in camera pictures*. Faculty Division of Computer Science, Wrocław University of Technology.
8. K. Daniilidis, C. Krauss, M. Hansen and G. Sommer. *Real-Time Tracking of Moving Objects with an Active Camera*. Real-Time Imaging 4, (1998).
9. Kyu Bum Han, Yoon Su Back. *Visual servo tracking strategy using time-varying kalman filter estimation*. Yonsei University, Seoul, Korea.
10. J.A. Piepmeier, G.V. McMuray, H.Lipkin. *Tracking moving target with Model Independent Visual Servoing: A predictive estimate Approach* Proceeding of the 1998 IEEE, Int. Conf. On intelligent robot and system, Leuven, Begium, 1998.
11. Armel Cretual, Francois Chaumette. *Image-based visual servoing by integration of dynamic measurements.* IEEE Int. Conf. Robot. Automat., Vol 3:1994-2001, May, 1998.
12. Venkataraman Sundareswaran, Patrick Bouthemy, Franácois Chaumette. *Visual servoing using dynamic image parameters.* Institute national de recherche en informatique et en automatique.