行政院國家科學委員會專題研究計畫成果報告

機器臂視覺伺服控制之可觀性
Observability of the Depth Estimation for Robot Visual Servo Control

計畫編號: NSC88-2213-E-009-124
執行期限: 87年08月01日至88年07月31日
主持人:林錫寬教授　交通大學電機與控制系

1 Abstract

This report investigates the observability problem of the visual system. We find that the linear velocity of the camera must satisfy some constraints for a success of a depth estimation method. Some simulation results show that the components of the linear velocity also affect the performance of the depth estimation. This fact indicates that using a modified camera motion control can improve the depth estimation.

Keywords: perspective observability, depth estimation, visual servo control

中文摘要

本計畫主要探討視覺系統的可觀性。計畫中我們發現攝影機的平移速度必須滿足許多限制才能成功估測深度，模擬並顯示攝影機的線性速度分量會影響深度估測性能。此結果可用在速度控制器上，改善深度的估測性。

關鍵字:可觀性、深度估測、視覺伺服控制

2 Introduction

Depth recovery is a very important problem in 3-D visual applications, such as object tracking and motion estimation. A hand-eye system uses a CCD mounted on the end-effector of the manipulator to enhance the tracking accuracy. In this report, we will discuss on the depth estimation using a single vision sensor via the dynamic motion.

Pose estimation by a moving camera have been studied recently [1]. The depth variable was seen as an unknown state, then the depth estimation problem became an observer design problem and the observability needed checking first. In this report, we relate the observability to the linear velocity of the camera by considering the visual dynamic as a nonlinear system. Some authors use the extended Kalman filter (EKF) to estimate the range data [2]. However, little attention has been paid to the effect of the velocity of the vision sensor on the observability. Matthies and Kanade [2] first mentioned that EKF estimates more effectively when the camera moves almost parallel to the image plane. Their experimental suggestion motivates us to investigate the influence of the direction of the camera velocity on the depth estimation.

In our previous work on visual servoing [3], a feature-based controllers had been proposed with EKF used to estimate the depthes of the feature points under assumption of all optical parameters given. We conclude that the components of the linear velocity of the camera will entirely determine the observability

of the visual system. We also attempt to establish the relationship between the linear velocity of the camera and the performance of the depth observer. Some simulation results substantiate our previous statements.

3 Pose Estimation

Consider a pinhole camera model [3]. There is a 3-D point P with coordinates (X, Y, Z) with respect to the camera frame E_{XYZ}. The value of Z is referred as the depth of point P to the camera lens. The image of point P projected onto the image plane is denoted by p. The coordinates of p with respect to E_{xyz} are $(x, y, 0)$. The projection relationships state

$$x = \gamma_x \frac{X}{Z}, \quad y = \gamma_y \frac{Y}{Z} \tag{1}$$

where $\gamma_x = f_e/S_x$ and $\gamma_y = f_e/S_y$, in which f_e is the *effective focal length*, S_x, S_y are, respectively, the horizontal and vertical lengths per pixel on the camera sensing array. Equation (1) is the so-called *perspective projection equation* [3]. Let the translation velocity and the angular velocity of the camera be represented by $\mathbf{v} \triangleq [v_x,\ v_y,\ v_z]^T$ and $\boldsymbol{\omega} \triangleq [\omega_x,\ \omega_y,\ \omega_z]^T$, respectively. The dynamic of (x, y) is the so-called *optic flow-motion equation* [3]:

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \mathbf{J}\mathbf{u} \tag{2}$$

where $\mathbf{u} \triangleq [\mathbf{v}^T,\ \boldsymbol{\omega}^T]^T$ and

$$\mathbf{J} = \begin{bmatrix} -\frac{\gamma_x}{Z} & 0 & \frac{x}{Z} & \frac{xy}{\gamma_y} & -\frac{\gamma_x^2 + x^2}{\gamma_x} & \frac{\gamma_x}{\gamma_y} y \\ 0 & -\frac{\gamma_y}{Z} & \frac{y}{Z} & \frac{\gamma_y^2 + y^2}{\gamma_y} & -\frac{xy}{\gamma_x} & -\frac{\gamma_y}{\gamma_x} x \end{bmatrix}$$

Since the unknown depth Z ($Z \neq 0$) is involved in (2), the dynamic equation of Z needs to be considered together. Let $\boldsymbol{\xi} \triangleq [x,\ y,\ Z]^T$ and $\boldsymbol{\psi} \triangleq [x,\ y]^T$. We describe the present system by the nonlinear system of

$$\begin{aligned} \dot{\boldsymbol{\xi}} &= \begin{bmatrix} & & \mathbf{J} & & & \\ 0 & 0 & -1 & -\frac{yZ}{\gamma_y} & \frac{xZ}{\gamma_x} & 0 \end{bmatrix} \mathbf{u} \\ \boldsymbol{\psi} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \boldsymbol{\xi} \end{aligned} \tag{3}$$

Since Z should be greater than zero, we consider $\boldsymbol{\xi} \in \mathcal{M}$, where $\mathcal{M}$ is an open set, and defined as

$$\mathcal{M} = \left\{ \boldsymbol{\xi} : \boldsymbol{\xi} = [x,\ y,\ Z]^T, \forall x,\ y \in \Re,\ Z > 0 \right\} \tag{4}$$

The *pose estimation* problem is to estimate the state vector $\boldsymbol{\xi}$. In the state vector $\boldsymbol{\xi}$, x and y are the the measured values, so the only unknown is the depth Z. Therefore, the state estimation of the system of (3) is also called the *depth estimation*. Since the system (3) is nonlinear, the input $\mathbf{u}$ could affect the observability. Therefore, we are then interested in knowing what kind of an input $\mathbf{u}$ can make the output $\boldsymbol{\psi}$ different for different $\boldsymbol{\xi}$. For this purpose, we examine the local observability at a certain state $\boldsymbol{\xi}_0$ while $\mathbf{u}(t)$ satisfies some special forms, which is the topic of the next section.

4 Observability of the Pose Estimation

According to the definitions and theorems on observability for nonlinear systems [4], we obtain the following theorem. Due to the limitation of the space, the proof is omitted.

Theorem 1 *Consider the system of* (3)*, and the open set* $\mathcal{M}$ *defined in* (4)*. The system is locally observable in* $\mathcal{M}$ *if and only if the projection of* $\mathbf{v}(t)$ *onto the distribution* $span\{\mathbf{v}_1(\boldsymbol{\xi}), \mathbf{v}_2(\boldsymbol{\xi})\}$ *is nonzero for some time interval, where*

$$\mathbf{v}_1(\boldsymbol{\xi}) = [-\gamma_x,\ 0,\ x]^T,\ \mathbf{v}_2(\boldsymbol{\xi}) = [0,\ -\gamma_y,\ y]^T$$

■

Let $\mathbf{v}_3(\boldsymbol{\xi}) = \mathbf{v}_1(\boldsymbol{\xi}) \times \mathbf{v}_2(\boldsymbol{\xi})/(\gamma_x \gamma_y) = [x/\gamma_x,\ y/\gamma_y,\ 1]^T$. Then $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ are linear independent and the linear velocity $\mathbf{v}(t)$ can be expressed as

$$\mathbf{v}(t) = \mathbf{v}_s(t) + \alpha_3(t)\mathbf{v}_3(\boldsymbol{\xi}) \tag{5}$$

where $\mathbf{v}_s(t) = \alpha_1(t)\mathbf{v}_1(\boldsymbol{\xi}) + \alpha_2(t)\mathbf{v}_2(\boldsymbol{\xi})$ and $\mathbf{v}_s(t)\cdot\mathbf{v}_3(\boldsymbol{\xi}) = 0$, for a given $\boldsymbol{\xi}$. The condition in Theorem 1 is equivalent to that the input satisfies $\|\mathbf{v}_s(t)\| > 0$ for some time interval.

Remark: Theorem 1 provides only the *local* observability of (3). It applies only to a slowly varying system or a system with quick sampling rate, since the local observability is valid only in the neighborhood of a state $\boldsymbol{\xi}_0$.In pratice, the camera moves slowly to constrain the interesting features inside the field of view, thus the local observability is sufficient for the success of the depth estimation.

4.1 Performance of Depth Estimation

Theorem 1 has indicated that $\|\mathbf{v}_s\| > 0$ is necessary for the estimation of depth. We are also interested in the influence of $\mathbf{v}_s$ on the convergence performance of the depth estimator. Since x and y are measurable, the estimation errors of them could be ignored. If the variation rate of Z is also small enough to be negligible, we can use the least squares estimation theory to analyze the relation of $\mathbf{v}$ to the convergence performance of the depth estimation.

Assume that the data satisfy

$$\mathbf{Y}_o = \boldsymbol{\Phi}\boldsymbol{\theta} + \boldsymbol{\rho} \tag{6}$$

where $\mathbf{Y}_o \in \Re^m$, $\boldsymbol{\Phi} \in \Re^{m\times n}$, and $\boldsymbol{\theta} \in \Re^n$ is a parameter vector to be estimated. The last term is a stochastic vector with variance matrix $\mathbf{S}$. The *best unbiased linear estimator* [5] is $\hat{\boldsymbol{\theta}}^* = \mathbf{S}^{-1}\boldsymbol{\Phi}(\boldsymbol{\Phi}^T\mathbf{S}^{-1}\boldsymbol{\Phi})^{-1}\mathbf{Y}_o$. Then the corresponding minimal value of $\mathrm{Cov}(\hat{\boldsymbol{\theta}})$ is $\mathrm{Cov}(\hat{\boldsymbol{\theta}}^*) = (\boldsymbol{\Phi}^T\mathbf{S}^{-1}\boldsymbol{\Phi})^{-1}$. Since $\mathrm{Tr}[\mathrm{Cov}(\hat{\theta})] \triangleq E[(\theta-\hat{\theta})^T(\theta-\hat{\theta})]$, where $\mathrm{Tr}(\cdot)$ is a trace operator, an increase in $(\theta-\hat{\theta})^2$ leads to an increase in $\mathrm{Tr}[\mathrm{Cov}(\hat{\theta})]$ which can serve as an accuracy index of the parameter estimation.

Define $a_x(\boldsymbol{\xi},\boldsymbol{\omega}) = (xy/\gamma_y)\omega_x - (\gamma_x^2 + x^2/\gamma_x)\omega_y + (\gamma_x y/\gamma_y)\omega_z$ and $a_y(\boldsymbol{\xi},\boldsymbol{\omega}) = (\gamma_y^2 + y^2/\gamma_y)\omega_x - (xy/\gamma_x)\omega_y - (\gamma_y x/\gamma_x)\omega_z$, then equation (2) can be reformed in the form of (6) in which $\hat{\boldsymbol{\theta}} = 1/\hat{Z}$, $\mathbf{Y}_o = [\dot{x} - a_x(\boldsymbol{\xi},\boldsymbol{\omega}),\ \dot{y} - a_y(\boldsymbol{\xi},\boldsymbol{\omega})]^T$, and

$$\boldsymbol{\Phi} = \begin{bmatrix} \mathbf{v}\cdot\mathbf{v}_1 \\ \mathbf{v}\cdot\mathbf{v}_2 \end{bmatrix}, \quad \boldsymbol{\rho} = \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} \tag{7}$$

ρ_1 and ρ_2 are two zero-mean Gaussian noise terms. The variance matrix $\mathbf{S}$ is then $\mathrm{diag}(q_{11}, q_{22})$, where q_{ii} is the variance of ρ_i, $i = 1, 2$. The unknown parameter $1/Z$ can be solved by statistics. Note that $\mathbf{v}_3\cdot\mathbf{v}_1 = \mathbf{v}_3\cdot\mathbf{v}_2 = 0$. If we substitute (5) and (7) into $\mathrm{Cov}(\hat{\boldsymbol{\theta}}*)$, then the covariance of $1/\hat{Z}$ is

$$\mathrm{Cov}\Big(\frac{1}{\hat{Z}}\Big) = \left[\mathbf{v}_s^T\left(\frac{\mathbf{v}_1\mathbf{v}_1^T}{q_{11}} + \frac{\mathbf{v}_2\mathbf{v}_2^T}{q_{22}}\right)\mathbf{v}_s\right]^{-1} \tag{8}$$

Although $\mathrm{Cov}(1/\hat{Z}) \neq \mathrm{Cov}(\hat{Z})$, both variances can be used to measure the accuracy of the estimated depth. However, this is a linear result, when Z is almost unchanged. It is difficult rigorously to obtain a nonlinear version of this property. Fortunately, this property is still retained for the nonlinear system (3) according to our simulations and is summarized as a conjecture in the following.

Conjecture 2 *Consider the system* (3) *with input* $\mathbf{u} = [\mathbf{v}^T,\ \boldsymbol{\omega}^T]^T$. *Suppose that* (x, y) *and* $\|\mathbf{v}(t)\|$ *are given. Then the depth estimation has a faster convergent rate, if the linear velocity of the input leads the greater* $\mathcal{J}_o(\boldsymbol{\xi},\mathbf{v})$, *where* $\mathcal{J}_o(\boldsymbol{\xi},\mathbf{v}) = \mathbf{v}_s^T(\mathbf{v}_1\mathbf{v}_1^T/q_{11} + \mathbf{v}_2\mathbf{v}_2^T/q_{22})\mathbf{v}_s$. ■

In the next section, two simulation examples are used to verify Theroem 1 and support Conjecture 2.

5 Simulation

Establish a distribution spanned by the orthonormal basis $\{\bar{\mathbf{b}}_1, \cdots, \bar{\mathbf{b}}_6\}(\boldsymbol{\xi})$, where $\bar{\mathbf{b}}_i = \mathbf{b}_i/\|\mathbf{b}_i\|$ and

$$\mathbf{b}_1 = \begin{bmatrix}\mathbf{v}_1\\ \mathbf{0}\end{bmatrix},\ \mathbf{b}_2 = \begin{bmatrix}\mathbf{v}_3\times\mathbf{v}_1\\ \mathbf{0}\end{bmatrix},\ \mathbf{b}_3 = \begin{bmatrix}\mathbf{v}_3\\ \mathbf{0}\end{bmatrix},$$
$$\mathbf{b}_4 = \begin{bmatrix}\mathbf{0}\\ \mathbf{v}_1\end{bmatrix},\ \mathbf{b}_5 = \begin{bmatrix}\mathbf{0}\\ \mathbf{v}_3\times\mathbf{v}_1\end{bmatrix},\ \mathbf{b}_6 = \begin{bmatrix}\mathbf{0}\\ \mathbf{v}_3\end{bmatrix} \tag{9}$$

where $\mathbf{0} = [0,\ 0,\ 0]^T$. Note that the distribution is independent of the unknown depth Z. The camera velocity can be spanned by

$$\mathbf{u} = \sum_{i=1}^{6} c_i \bar{\mathbf{b}}_i(\boldsymbol{\xi}) = [\mathbf{b}_1,\ \cdots,\ \mathbf{b}_6] \cdot [\mathbf{u}]_{\mathbf{b}} \quad (10)$$

where $[\mathbf{u}]_{\mathbf{b}} \triangleq [c_1, \cdots, c_6]^T$. It is found that $[\mathbf{v}_s(\boldsymbol{\xi})^T,\ \mathbf{0}^T]^T \in \text{span}\{\bar{\mathbf{b}}_1, \bar{\mathbf{b}}_2\}$, and $\|\mathbf{v}_s\|^2 = c_1^2 + c_2^2$. The roles of factors α_1 and α_2 in the observability test (see Theorem 1) can be replaced by c_1 and c_2, since $\alpha_1^2 + \alpha_2^2 > 0$ is equivalent to $c_1^2 + c_2^2 > 0$.

The model in the following simulation examples is the visual system (3) with $f_e = 16.53$mm, $S_x = 0.0161$mm/pixel, and $S_y = 0.0189$mm/pixel. The states are estimated by EKF [2] with the measurement noises covariance matrices $\mathbf{R} = \text{diag}(2.25\text{pixel}^2, 2.25\text{pixel}^2)$, the system noises covariance matrices $\mathbf{Q} = \text{diag}(2.25\text{pixel}^2, 2.25\text{pixel}^2, 16\text{mm}^2)$, and the image sampling period 200ms. Suppose that an interesting feature point on the screen is initially located at (100,100). The guessed values of the initial depth of the point is 400mm, while its true depth is 450mm, i.e., the initial depth estimation error is 50mm.

Example 1: This example tries to verify Theorem 1, which states that a necessary and sufficient condition for the system (3) to be locally observable is $\|\mathbf{v}_s(t)\|^2 > 0$ for some time interval. Fig. 1 shows some simulations of the depth estimation to verify the sufficiency of Theorem 1. The camera velocities are listed as follows (cf. (10)):

$$\begin{aligned}
[\mathbf{u}_1(\boldsymbol{\xi})]_{\mathbf{b}} &= 8[1,\ 0,\ 0,\ 0,\ 0,\ 0]^T \\
[\mathbf{u}_2(\boldsymbol{\xi})]_{\mathbf{b}} &= 8[0,\ -\sin(0.1t),\ 1,\ 0,\ 0,\ 0]^T \\
[\mathbf{u}_3(\boldsymbol{\xi})]_{\mathbf{b}} &= 6[0,\ 0,\ 1,\ 0,\ 0,\ 0]^T \\
[\mathbf{u}_4(\boldsymbol{\xi})]_{\mathbf{b}} &= 0.2\sin(0.1t)[0,\ 0,\ 0,\ 1,\ 0,\ 0]^T \\
[\mathbf{u}_5(\boldsymbol{\xi})]_{\mathbf{b}} &= 0.2[0, 0, 15, \sin(0.1t), \sin(0.2t), 0]^T
\end{aligned}$$

The inputs $\mathbf{u}_1$ and $\mathbf{u}_2$ satisfy the condition in Theorem 1, i.e., $\|\mathbf{v}_s\|^2 > 0$, even $\mathbf{u}_2$ sinusoidally varies along $\mathbf{v}_2$. Fig. 1 shows that the depth estimation errors for the first two inputs both converge to zero. This is consistent with the sufficiency of Theorem 1.

Now, consider the inputs $\mathbf{u}_3$, $\mathbf{u}_4$, $\mathbf{u}_5$ with $c_1 = c_2 = 0$, which do not satisfy the condition in Theorem 1. The sinusoidal functions for the angular velocities are necessary to prevent the values of the depth from approaching zero (i.e., $Z \approx 0$) after a small time interval. Fig. 1 shows that the depth estimation errors for $\mathbf{u}_3$, $\mathbf{u}_4$, $\mathbf{u}_5$ do not tend to converge, which verifies the necessity of Theorem 1. ■

The following example will show that Conjecture 2 is acceptable, at least for $\|\mathbf{v}\| \leq 10$ mm/sec. Consider the input in the form of

$$\mathbf{u} = \|\mathbf{v}\| \cdot [C_\beta(C_\gamma \bar{\mathbf{b}}_1 + S_\gamma \bar{\mathbf{b}}_2) + S_\beta \bar{\mathbf{b}}_3] + [\boldsymbol{\omega}]_{\mathbf{b}}, \quad (11)$$

where $S_\beta \triangleq \sin(\beta)$, $C_\beta \triangleq \cos(\beta)$, $S_\gamma \triangleq \sin(\gamma)$, $C_\gamma \triangleq \cos(\gamma)$, β and $\gamma \in \Re$, and $[\boldsymbol{\omega}]_{\mathbf{b}} \triangleq [0,\ 0,\ 0,\ \boldsymbol{\omega}^T]^T$. It is known that any unit vector in $\Re^3$ can be represented by $[C_\beta C_\gamma,\ C_\beta S_\gamma,\ S_\beta]^T$. If $\|\mathbf{v}_s(t)\|$ is fixed (i.e., β is fixed), $\mathbf{v}_s(t)$ can still be adjusted by γ. By the definition in Conjecture 2, $\mathcal{J}(\boldsymbol{\xi}, \mathbf{v})$ is proportional to C_β^2 when C_γ is fixed. In Example 2, the ratio of $|C_\beta|$ to $|S_\beta|$ will be changed to investigate the relationship between the index function $\mathcal{J}_o(\boldsymbol{\xi}, \mathbf{v})$ and the convergent rate of the depth estimation error.

Example 2: We consider a family of velocities of the camera with fixed C_β. Set $\|\mathbf{v}(t)\| = 10$ mm/sec for all cases. In the first family, $\mathbf{v}_s(t)$ varies with C_β and then $\|\mathbf{v}_s(t)\|$ is proportional to C_β (i.e., $\|\mathbf{v}_s(t)\| = \sqrt{c_1^2 + c_2^2} = 10C_\beta$). Fig. 2 shows the simulation results for $C_\gamma = 1/\sqrt{2}$ and $C_\beta = 1,\ 0.8,\ 0.5,\ 0.3,\ 0.1,\ 0,\ -0.4$, and -0.7, while $\boldsymbol{\omega} = 0.01[2,\ 1,\ 3]^T$ rad/sec. It should be remarked that $\boldsymbol{\omega}$ is arbitrarily assigned and does not affect the property of the simulations. It can be seen from the depth estimation errors in Fig. 2 and the cost functions $\mathcal{J}_o(\boldsymbol{\xi}, \mathbf{v})$ in Fig. 3 that the depth estimation error converges more quickly as $\mathcal{J}_o(\boldsymbol{\xi}_0, \mathbf{v})$ is larger.

We also changed the value of the fixed C_γ in the range $[-1, 1]$ and found that the simulations have the same property of the depth estimation performance as that in Fig. 3. All these simulations in Fig. 2 and 3 believe that $\mathcal{J}_o(\boldsymbol{\xi}, \mathbf{v})$ is a good index for the convergent rate of the depth estimation. ■

6 Conclusion

This report investigated the effect of the camera velocity on the observability of the visual system (3). The results indicate that the observability of (3) is entirely determined by the component ratio of the linear velocity of the camera. By using an index function $\mathcal{J}_o(\boldsymbol{\xi}, \mathbf{v})$ defined in Conjecture 2, the performances of the depth estimations by different velocities can be compared previously. The results of this report may be useful to improve the depth estimation by a modified velocity controller. Specially, $\mathcal{J}_o(\boldsymbol{\xi}, \mathbf{v})$ is described by a quadratic form of $\mathbf{u}$. It is possible to achieve the objective of feature-based control and the improvement of the depth estimation by minimizing $\|\mathbf{J}\mathbf{u} - \dot{\boldsymbol{\xi}}\|^2 - \mathcal{J}_o(\boldsymbol{\xi}, \mathbf{v})$.

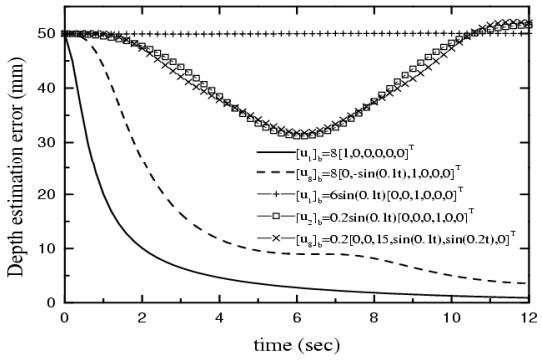

Figure 1: Depth estimation errors by different inputs, for example 1.

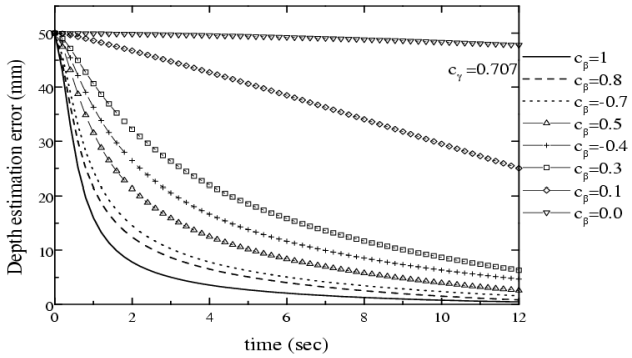

Figure 2: Depth estimation errors by different inputs $\mathbf{u}$ with $\|\mathbf{v}\| = 10$mm/sec, for example 2.

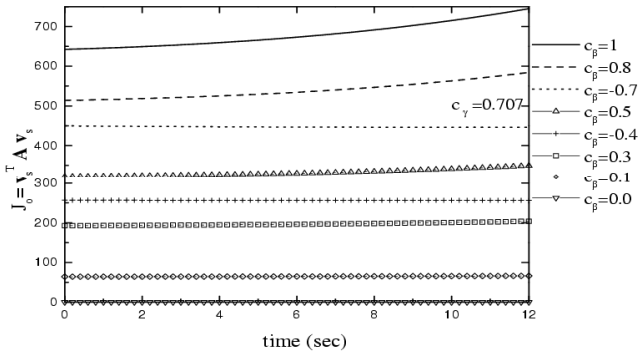

Figure 3: The cost function $\mathcal{J}_o(\boldsymbol{\xi}, \mathbf{v})$ by different inputs $\mathbf{u}$ with $\|\mathbf{v}\| = 10$mm/sec, for example 2.

References

[1] N. P. Papanikolopoulos, P. K. Khosla, and T. Kanade, "Visual tracking of a moving target by a camera mounted on a robot: a combination of control and vision", *IEEE Trans. Robotics Automat.*, vol. 9, no. 1, pp. 14–34, 1993.

[2] L. Matthies and T. Kanade, "Kalman filter-based algorithms for estimating depth from image sequences", *Int. J. Computer Vision*, vol. 3, pp. 209–236, 1989.

[3] C. J. Fang and S. K. Lin, "Image-Feature Based Tracking Control", in *1997 Automatic Control Conference"*, 1997, pp. 646–648.

[4] M. Vidyasagar, *Nonlinear System Analysis.* Prentice-Hall, Englewood Cliffs, 1993.

[5] T. Söderström and P. Stoica, *System Identification.* Prentice-Hall, New York, 1989.