Robotica (2003) volume 21, pp. 615–625. © 2003 Cambridge University Press
DOI: 10.1017/S0263574703005332 Printed in the United Kingdom

Implementation of a real-time moving object tracking system using visual servoing

M-C. Tsai*, K-Y. Chen*, M-Y. Cheng† and K.C. Lin‡

(Received in Final Form: March 27, 2003)

SUMMARY

Due to the increasing popularity of surveillance and security systems, the problem of automatically tracking a moving target by visual servoing has become a research topic deserving more investigation. Nonetheless, the success of tracking a moving target in real-time primarily depends on the performance of the motion detection techniques employed. This paper addresses visual tracking control of an unknown target that could be motional arbitrarily in the scene. A pan-tilt mechanism is used to gain the flexibility of tracking, and the so-called region-based matching method and motion energy method are modified and proposed in this study to detect a moving target based on the consecutive images acquired. A visual servo control scheme that adopts proportional control in the visual loop for reducing the servo lagging is proposed using output disturbance feedforward compensation. Experimental results show the superiority of the proposed method in achieving high system bandwidth and tracking accuracy.

KEYWORDS: Visual servoing; Image tracking; DSP.

1. INTRODUCTION

Visual tracking control is an interesting and important research topic in the field of machine vision. The main reason for its increasing importance is due to the rapid growing power of the microprocessor's speed in implementation of servo control and image processing. On the other hand, the booming success of the internet also creates new ideas in the applications of visual servoing, such as web conferencing and e-learning classes. In these applications, automatic vision machines track speakers or specific targets to assure good focus in the broadcast images that are needed. With visual servoing, it is possible to control the pointing direction of camera and to guarantee the capture of the right image in the scene. Achieving such a vision system in a cost-effective way has become more interesting to researchers than ever before.

In addition, the applications of visual servoing technique to the security/surveillance system and military related fields have also received a considerable amount of attention. With the visual tracking ability, a surveillance system will be able to detect an intruder's motion, in which a motorized camera will move dynamically to keep the intruder's image in the center of the image plane and also issue an appropriate warning signal. Also, law enforcement personnel certainly can use this data to identify the intruder.

The idea of using vision as a feedback signal to close the outer (position) loop was proposed by Hill and Park.[1] However, hampered by the heavy computation load required in image processing, there was a lack of breakthrough achievement in the early trials. Nevertheless, as the CPU's computational capability becomes more powerful, the idea of using vision feedback is no longer an unachievable goal. It is no wonder that the application of visual servoing has attracted considerable attention recently.[2–7]

In general, a real-time visual servoing system consists of two parts: "motion detection unit" and "servo control unit". The main task of the motion detection unit is to detect the moving target based on the consecutive images captured, and also to determine the target position in the current image frame. In order to execute all the tasks in real-time, the central issues for the motion detection unit are how to shorten the time spent in detecting a moving target and determining its corresponding position in the image plane. Several approaches on these issues have been proposed. One of the approaches is the "recognition-based tracking method",[3] which is a modification to object recognition. Through the calculations of 3D positions of all objects in the scene, the desired target can be detected. Others like Horn and Schunck,[8] had proposed a "differential method"[9] to compute optical flows[10] based on the consecutive images. A minimum square solution satisfying intensity constancy constraint and global smoothness constraint can be found by using an iteration method. Another popular technique is called "model-based method",[11] where the features of a target (such as human face or body) are extracted in advance to serve as a template for matching in each subsequent image frame.

Nonetheless, the methods discussed above belong to the class of high-level operation that can be very time consuming. Consequently, they may not be suitable for real-time execution. Since this study focuses on implementing real-time visual servoing, thus "region-based matching method"[3] and "motion energy method"[12] are adopted in this paper to detect the motion of an unknown moving target due to their less demand on computing power. A relatively complete study was conducted in this study to compare the performances of these two methods.

The design of servo control unit in a visual servoing system is also an interesting topic for those researchers

* Department of Mechanical Engineering, National Cheng Kung University, Tainan (Taiwan).
† Department of Electrical Engineering, National Cheng Kung University, Tainan (Taiwan)
‡ Department of Mechanical Engineering, Kun San University of Technology (Taiwan)

https://doi.org/10.1017/S0263574703005332 Published online by Cambridge University Press

specializing in the fields of control, automation, and robotics. Corke and Good[4] discussed the issue of dynamic effect in visual servoing problems. Control schemes for visual servoing that include PID control, pole placement control and the feedforward control were explored in their studies. Oh and Allen[7] developed a design methodology for a 5 DOF pan-tilt-gantry system, that can be used to determine which DOF motion should be under visual servoing control.

This paper is aimed to develop a prototype pan-tilt visual tracking system that can automatically track an unknown moving object in real-time, and also keep the object's images in the center of image plane, and is organized as follows: Section 2 provides a brief introduction to the structure of the prototype visual tracking system. Techniques for detecting an unknown object's motion are discussed in Section 3. Section 4 introduces the visual servoing control structure and the camera projection model adopted in this study. Experimental results and Conclusions are given in Section 5 and 6, respectively.

2. REGION-BASED MATCHING METHOD FOR MOTION DETECTION

Figure 1 illustrates a two-degree of freedom visual tracking system developed in this study, where the CCD camera can move in both the pan and tilt directions simultaneously. Since the camera is mounted on a pan-tilt unit, this configuration is of eye-in-hand and the control action is purely based on the image feature of the object. The servo control scheme of the developed visual tracking system hence belongs to the category of image-based where its control block diagram is illustrated in Figure 2. It consists of two parts: (1) **motion detection unit**, i.e. the dash-line box inside of Figure 2 and (2) **servo control unit**. Based on the consecutive images taken by the camera, the motion of a moving object can be detected by the motion detection unit and then, the camera is controlled by the servo control unit such that the image of the moving object is kept in the center of the image plane. In addition, many approaches have been proposed to tackle the motion detection problem recently. In general, they are fundamentally different and deserve further investigation. Among them, the "region-based matching method" and the "motion energy method" are two of the popular approaches. Detailed descriptions about these two methods are given in the section below and Section 3, respectively.

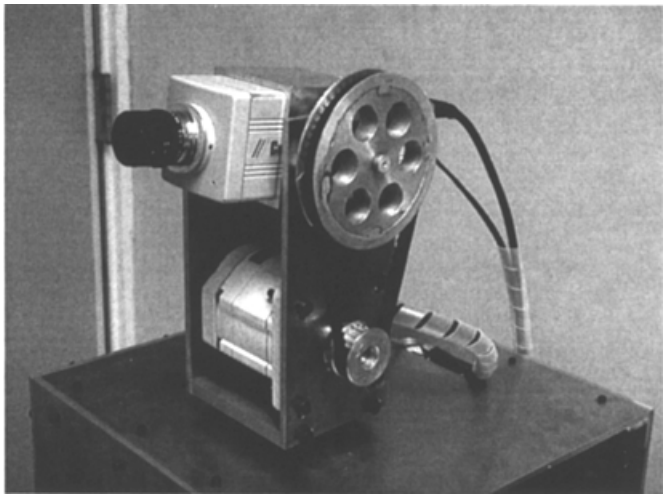


Fig. 1. Real-time pan-tilt visual tracking system.

The Sum-of-Square Difference (SSD)[2,13] method is one of the efficient and popular region-based matching methods. Under the assumptions that the moving object should be almost rigid (non-deformable) and its image intensity remains constant over time (temporal consistency), the SSD method can be applied to detect the motion of a moving object. The principle of the SSD method is mainly to search for the most similar region in each image frame by using a template N of the intended target. That is to find a displacement vector $d=[\Delta u \quad \Delta v]^T$ such that the similarity error measure E defined below has a minimum[2]:

$$E(u,v)=\sum_{m,n\in N}(I(u+m+\Delta u,v+n+\Delta v,t)-I(u+m,v+n,t_0))^2 \tag{1}$$

where m and n represent the pixels in the region of interest N, i.e. "template" of the intended target and $I(\cdot)$ is the image intensity of a pixel.

To apply the SSD method, the region of interest N must be determined in advance. Suppose that there is no background motion and the CCD camera is still initially. The intensity difference between two consecutive image frames can be used to determine which portion belongs to a moving target. After the moving target is detected, a rectangular window N around the center of the moving target in the image plane is selected as the "template" for the SSD method.

To find the displacement vector in the image plane that yields least SSD measurement, most of the previous studies adopted a sequential search from top-left to bottom-right (i.e. full search as shown in Figure 3). An alternative approach – wave-like search method[14] shown in Figure 4, starts the search from the center position of the target in the previous image frame. Nevertheless, if the displacement is large, then both approaches are not efficient enough. In this study, to speed up the searching process, the three-step hierarchal search (3SHS) method[15] as shown in Figure 5, is employed to search for the displacement vector of the moving target in the image plane. In a coarse-to-fine manner, nine sparsely located candidates are evaluated and the one with the minimal SSD is chosen as the winner of this step. In the second step, the search is focused on the area surrounding the winner of Step 1. Similarly, SSD of nine sparsely located candidates are evaluated, but the distances between candidate locations are shortened by half compared with that of Step 1. Again, the one with the minimal SSD is chosen as the winner of this step. In the same manner, Step 3 calculates SSD of nine locations around the winner of Step 2 and gives the final displacement vector.

https://doi.org/10.1017/S0263574703005332 Published online by Cambridge University Press

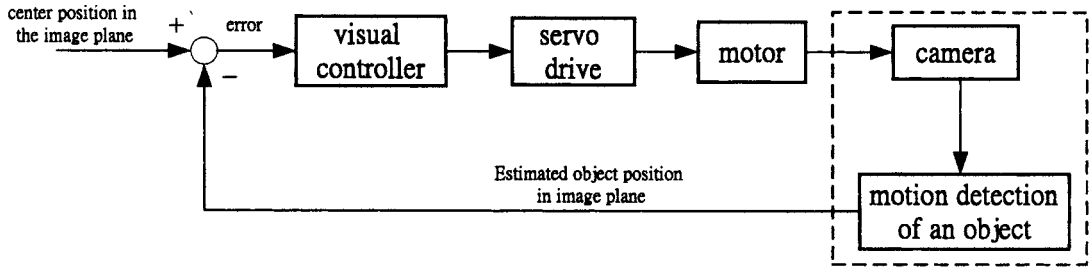


Fig. 2. Image-based visual servoing control structure.

3. MOTION ENERGY METHOD FOR MOTION DETECTION

Another popular approach to the problem of motion detection is the so-called "Motion Energy Method". By thresholding the absolute value of the intensity difference between two consecutive image frames at a suitable level, one can obtain a binary subtracted image that includes static and motion regions in both image frames (the previous and the current ones). To extract the motion region of current image frame, an edge detection technique is applied to acquire a binary edge image.[11] Then by performing a logical "AND" operation between the binary subtracted image and the edge information of current image frame, the edges of motion region can be obtained.

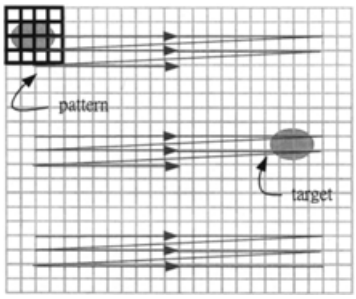


Fig. 3. Full search of displacement vector in the image plane.

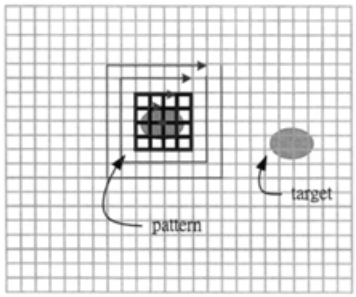


Fig. 4. Wave-like search of displacement vector in the image plane.

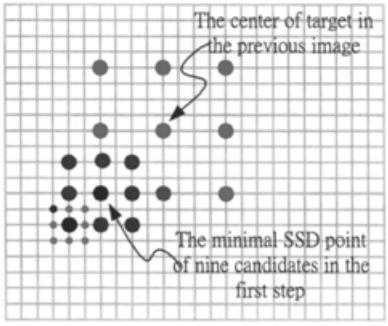


Fig. 5. 3-step search of displacement vector in the image plane.

In this study, rather than performing the edge detection method, three consecutive image frames and a simple subtraction operation are used. The details of our modified approach are summarized as follows.

(A) Let $I(t)$, $I(t-1)$ and $I(t-2)$ denote the current image frame, previous image frame and the one before the previous image frame, respectively. Use equation (2) to perform subtraction operation between $I(t)$ and $I(t-1)$, also between $I(t)$ and $I(t-2)$ to obtain the corresponding binary images D.

$$D_{i,j}=\begin{cases}1 & \text{if } |\, I(i,j,t-k)-I(i,j,t)\,| \geq \varepsilon \\ 0 & \text{otherwise}\end{cases} \tag{2}$$

where $k=1$ or 2, and ε is the predefined threshold.

(B) Apply the logical "AND" operation on the two binary images obtained in Step **(A)** to determine the motion region in the current image frame.

For a static camera, the method described in steps **(A)** and **(B)** can work effectively, but it is not the case for a motorizing camera. The apparent motion of background caused by camera rotation will make the tracking task of true motion fail. Therefore the background compensation technique (BCT)[11] is adopted in this study to overcome this difficulty. A brief introduction to the BCT technique is presented as follows:

Assume that a camera rotates about its lens center and the pinhole camera model is used. Then the relationship between every pixel position in two images, taken from different rotation angles can be expressed as:[11]

$$\begin{bmatrix} u(t-1) \\ v(t-1) \\ f \end{bmatrix} = s \begin{bmatrix} 1 & 0 & \theta_p \\ 0 & 1 & \theta_t \\ -\theta_p & \theta_t & 1 \end{bmatrix} \cdot \begin{bmatrix} u(t) \\ v(t) \\ f \end{bmatrix} \tag{3}$$

where $u(t)$, $v(t)$ are the pixel positions of the current image, and $u(t-1)$, $v(t-1)$ are the pixel positions in the previous image. In addition, s is the scaling factor, f is the focal length, θ_p is the angle in the pan direction and θ_t is the angle in the tilt direction.

For three consecutive images taken from different rotation angles, BCT can compensate the backgrounds of

https://doi.org/10.1017/S0263574703005332 Published online by Cambridge University Press

the first two frames such that their backgrounds are the same as that of the current image frame. After completing the background compensation, the three consecutive images can be seen as the static images, so that the motion energy method (namely steps **(A)** and **(B)**) can be applied to detect the motion region. In an ideal situation, the true motion region can be detected accurately if exact background compensation is applied. However in reality, several factors such as inaccuracies in the inputs to the compensation algorithm and an approximation that was used in deriving the algorithm all contribute to the compensation error. As a result, a noisy image and a narrow false motion region might be obtained. To overcome these drawbacks, a "morphological operation" is performed to filter out these undesired noises. The primary morphological operations include erosion and dilation, which are expressed as:

$$\begin{aligned} &\text{Dilation: } \mathbf{X} \oplus \mathbf{B} = \{p \in \varepsilon^2 : p = x + b, x \in \mathbf{X} \textit{ and } b \in \mathbf{B}\} \\ &\text{Erosion: } \mathbf{X} \ominus \mathbf{B} = \{p \in \varepsilon^2 : p + b \in \mathbf{X} \textit{ for every } b \in \mathbf{B}\} \end{aligned} \quad (4)$$

where **X** and **B** are the pixel coordinate set with value equal to 1 in a binary image and an operating mask, respectively; x and b are the certain pixel in **X** and **B**, respectively.

Using morphological erosion and dilation, the narrow parts of false motion regions due to inaccurate background compensation will be eliminated, while the original size and shape of the wide parts (the real motion region) will be preserved. The modified motion energy method for motorized cameras is illustrated in Figure 6, and its off-line test results are shown in Figure 7. In Figure 7, $I^*(t-2)$ is the image $I(t-2)$ after background compensation, and $I^*(t-1)$ denotes the image $I(t-1)$ after background compensation. In addition, $D(1)$ is the image intensity difference between $I(t)$ and $I^*(t-2)$, while $D(2)$ represents the image intensity difference between $I(t)$ and $I^*(t-1)$. A logical "AND" operation was performed on $D(1)$ and $D(2)$, where the obtained results are illustrated in Figure 7(h). To reduce the noise effect, a morphological operation is performed on the obtained results in Figure 7(h), and the final results are illustrated in Figure 7(i). Clearly, the moving object has been detected successfully.

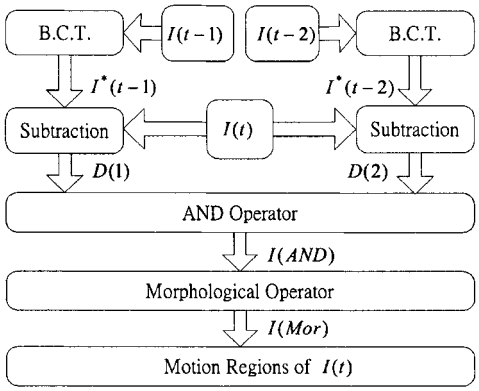


Fig. 6. Illustration of the modified motion energy method.

4. CAMERA MODEL AND VISUAL SERVOING CONTROL SCHEME

Using the modified SSD approach or modified motion energy method, the position of a moving target in the image plane can be estimated. Based on the estimated target's position in the image plane, the servo control task is to drive the pan-tilt camera such that the moving object's image is kept in the center of the image plane.

4.1. Camera projection model

Before investigating the visual servoing control, the camera projection model should be derived first. Since the camera used in this study (Figure 1) is mounted on a pan-tilt unit, it can rotate in both the pan and tilt directions. Nonetheless, to obtain the relations that describe the projection of the target

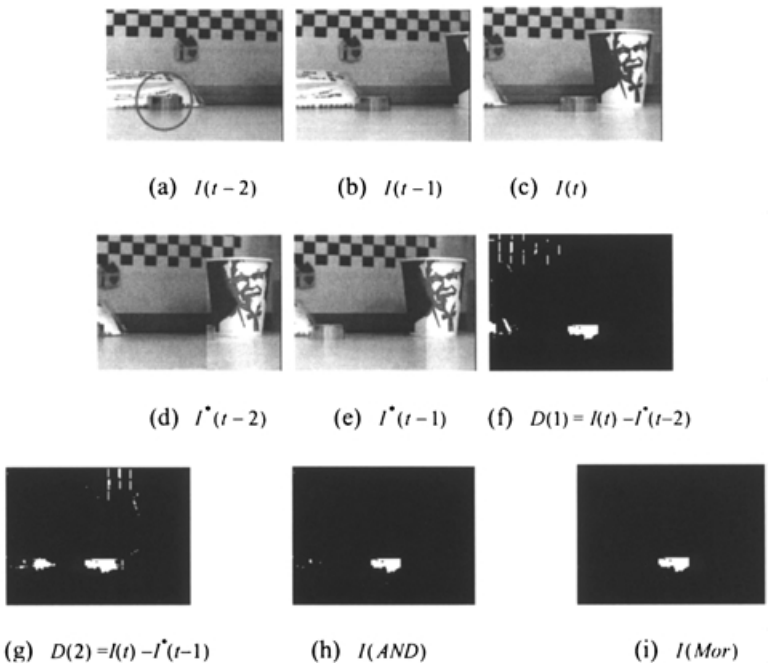


Fig. 7. Off-line test results of the modified motion energy method.

https://doi.org/10.1017/S0263574703005332 Published online by Cambridge University Press

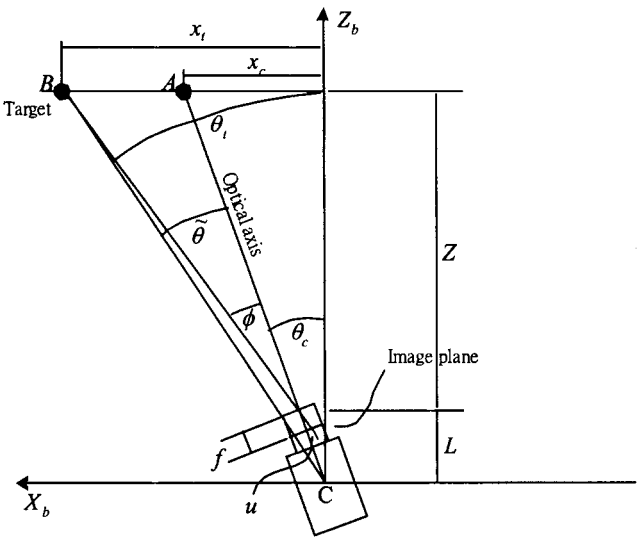


Fig. 8. Geometric relations for the camera projection model in the pan direction.

in the image plane, without loss of generality, consider the geometric relations in the pan direction as shown in Figure 8 only.[7] Let the target move perpendicularly to Z_b axis. Then when the target is at point A, its projection locates at the center of the image plane and in contrast, if the target moves to the point B, its projection in the image plane will be u pixels away from the center position. Based on Figure 8, one can obtain the following relations:

$$\theta_t = \tan^{-1}\frac{x_t}{Z+L} \text{ and } \theta_c = \tan^{-1}\frac{x_c}{Z+L},$$

where x_t is the current target location; θ_t is the angle between $\overline{CB}$ and Z_b axis; x_c the previous target location; θ_c is the angle between $\overline{CA}$ and Z_b axis; f: the focal length; Z the depth of target; and L the distance between lens center and rotational center C of the pan-tilt unit.

In addition, if θ_t and θ_c are assumed to be small, then one will have:[7]

$$\tilde{\theta} = \theta_t - \theta_c \approx \frac{x_t}{Z+L} - \frac{x_c}{Z+L} \tag{5}$$

Moreover, if the pinhole camera projection model[3] is adopted, one will obtain:

$$\phi \approx \frac{x_t - x_c}{Z} \approx \frac{u}{f} \tag{6}$$

where ϕ represents the current target's angle with respect to the lens center; u is the difference (in pixel) between the target projection and the center in the image plane. Note that it is assumed that the angle ϕ is small.

According to equations (5) and (6), the relations describing u and $\tilde{\theta}$ can be expressed as:[7]

$$u = \frac{f(x_t - x_c)}{Z} = \frac{f(Z+L)}{Z}\tilde{\theta} \tag{7}$$

By combining the camera projection model of equation (7), a visual servoing control scheme illustrated in Figure 9 is proposed in the following.

4.2. Visual servoing control structure

The surveillance camera of Figure 1 is mounted on a pan-tilt unit that is driven by two AC-servomotors. In general, commercially available servomotor drives can be operated in three different modes – position, velocity and torque. Corke and Good claimed that a visual servoing system could exhibit satisfactory performance if the velocity mode is adopted.[4] Therefore, the velocity mode of the servomotor drives was used in this study.

As we have mentioned previously, the target's motion is unknown, therefore the target location x_t in Figure 9 can be considered as an output disturbance.[4] The servo system design is to find an appropriate visual controller such that

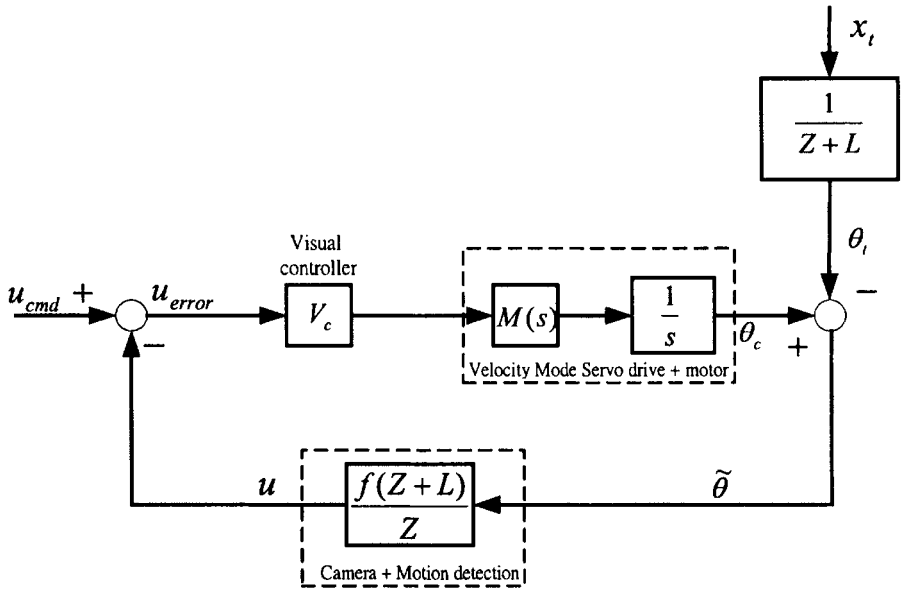


Fig. 9. Visual servoing control structure for the case that the servo drives of AC-servomotors are set to the velocity mode.

https://doi.org/10.1017/S0263574703005332 Published online by Cambridge University Press

the target's projection will be kept in the center of image plane. This implies that the reference command u_{cmd} in Figure 9 will be 0 (for simplicity, only motion in the pan direction is considered here). In addition, the bandwidth of the velocity loop (inner loop) for the used servomotor drive is often much larger than that for the vision loop (outer loop). Thus, the transfer function, denoted $M(s)$, of the velocity loop in the desired bandwidth can be simplified as a gain constant K_v so that the servo control system of Figure 9 can also be simplified (i.e. $M(s)=K_v$) for controller design and analysis. Now let

$$u_{error}=u-u_{cmd} \tag{8}$$

From $u=\dfrac{f(Z+L)}{Z}\,\tilde{\theta}$, letting $u_{cmd}=0$ will yield

$$u_{error}=\frac{f(Z+L)}{Z}\,\tilde{\theta} \tag{9}$$

In addition, by letting $M(s)=K_v$, the transfer function between u_{error}, and x_t in Figure 9 is given by

$$\frac{u_{error}}{x_t}=\frac{fs}{Zs+f(Z+L)V_CK_v} \tag{10}$$

Based on equation (10), one can design an appropriate visual controller, denoted V_c, for satisfying desired performance specifications, e.g. the steady state error. Since the visual servo system of Figure 9 is of type I, the position loop controller V_c can simply be a gain constant (i.e. only a P-gain $V_c=k$) in practice as used in many industrial servo systems. This setting is to reduce the servo lag and also to avoid the overshooting problem. The corresponding steady state error can be derived as shown in the following subject to two different types of target's motion trajectory.

According to equation (10), if x_t is a step-function-like disturbance, the steady state error in the u direction of image plane can be derived theoretically as:

$$u_{error}=e_{ss}=\lim_{s\to 0} s\times\frac{fs}{Zs+f(Z+L)V_CK_v}\times\frac{1}{s}$$

$$=\lim_{s\to 0}\frac{sf}{Zs+f(Z+L)kK_v}=0 \tag{11}$$

Similarly, if x_t is a ramp-function-like disturbance, the steady state error in the u direction is given by

$$u_{error}=e_{ss}=\lim_{s\to 0} s\times\frac{fs}{Zs+f(Z+L)V_CK_v}\times\frac{1}{s^2}$$

$$=\lim_{s\to 0}\frac{f}{Zs+f(Z+L)kK_v}=\frac{1}{(Z+L)kK_v} \tag{12}$$

Obviously, there exists a steady state error if the moving target to the visual tracking system is of the ramp-like functions. To eliminate this steady state error, a disturbance feedforward term, i.e. a derivative gain of F_c for the disturbance input that is similar to that of Corke and Good,[4] is introduced as shown in Figure 10(a). Although Corke and Good[4] proposed a similar structure, however in their study, the Kalman filter and $\alpha\beta$ filter were used to predict the object's velocity $\dot{x}_t$. In contrast, this study uses the LSF method[16] to calculate the object's velocity $\hat{\dot{\theta}}_t$ in the image plane directly based on previous four consecutive image frames. The present scheme illustrated in Figure 10(b) is using simpler computations when compared with that of Corke and Good[4] and hence is numerically stable. The resulting transfer function of equation (10) between u_{error} and x_t now becomes

$$\frac{u_{error}}{x_t}=\frac{(1-F_CK_v)fs}{Zs+f(Z+L)V_CK_v} \tag{13}$$

The steady state error in the u direction due to the ramp disturbance input is then given by

$$u_{error}=e_{ss}=\lim_{s\to 0}\frac{(1-F_CK_v)f}{Zs+(Z+L)kK_v}=\frac{(1-F_CK_v)f}{(Z+L)kK_v} \tag{14}$$

Obviously, if F_c is chosen such that $F_cK_v=1$, theoretically the steady state error for the ramp disturbance input will be eliminated completely.

5. EXPERIMENTAL RESULTS

The control experimental setup for the prototype pan-tilt visual tracking system of Figure 1 mainly consists of a frame grabber of MV-500 (30 frame/sec; 640×480 pixels); a CCD camera of CPT-8000 (811×508 pixels); a PMC32–6000 motion control card that is equipped with a high performance DSP, the TI TMS320C32 floating point processor, is responsible for real-time control of the pan-tilt servomechanism. The pan-tilt unit is driven by two AC servomotors with built-in incremental encoders (2500×4 pulses/rev) for providing motor position feedback. The servo drives of AC servomotors are set to the velocity mode throughout this study. All the image-related computations such as image pyramid, subtraction and morphological operation are performed on a 1 GHz IBM compatible PC.

5.1. Off-line tests for modified SSD and motion energy method

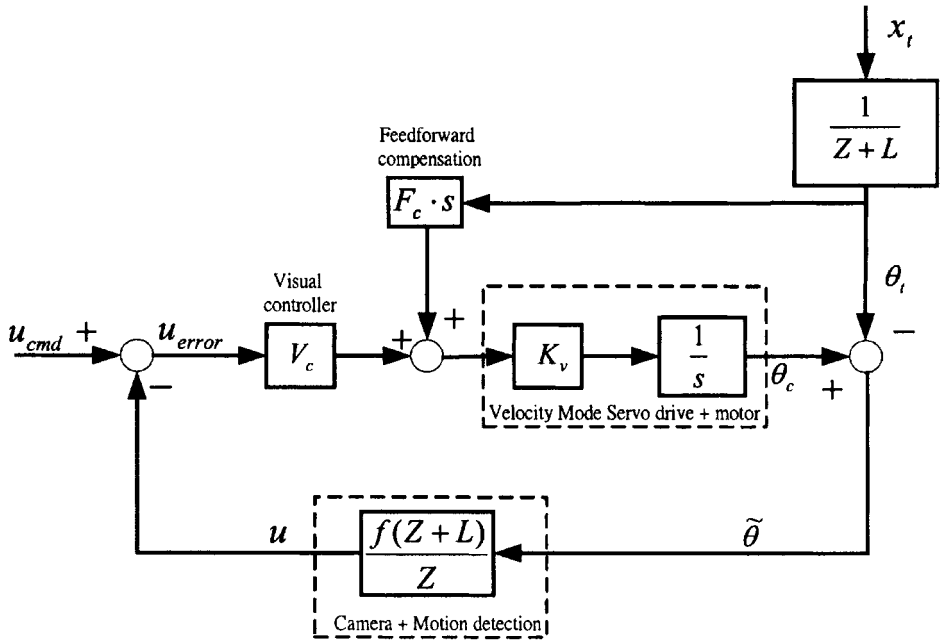
To compare the performance of the modified SSD method with that of the motion energy method, three experiments are conducted off-line in the following.

Experiment #1 (single object, simple background) – Tested image (Figure 11(a)): an object (blue ellipse) moves to right and to top with optical flow 6 pixel/frame.

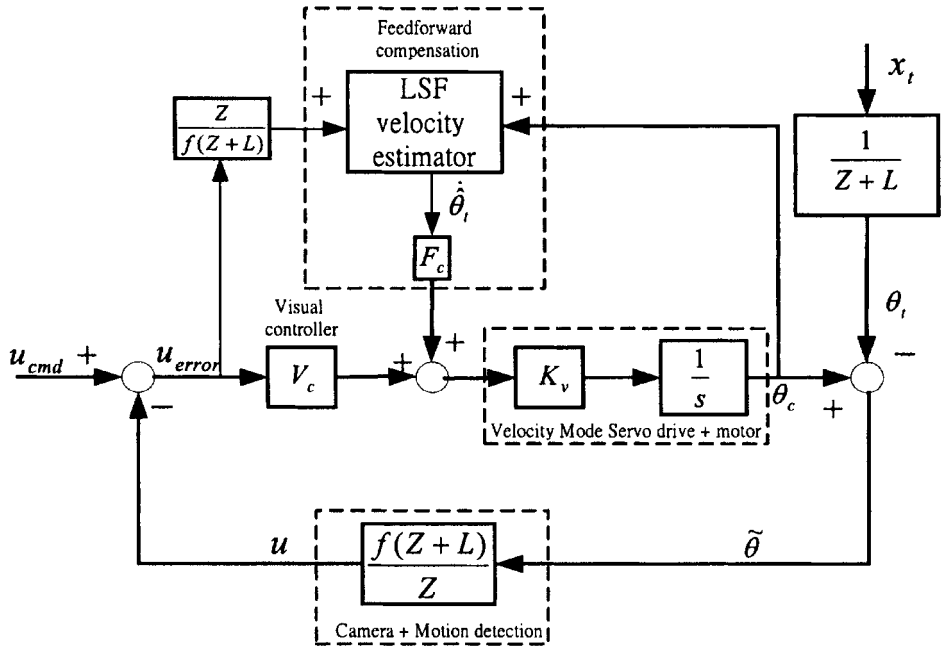
Experiment #2 (single object, complex background) – Tested image (Figure 11(b)): an object (white car) moves to right with optical flow 6 pixel/frame, and moves downward with optical flow 1 pixel/frame.

Experiment #3 (two objects, simple background) – Tested image (Figure 11(c)): an object (blue ellipse) moves to right and upward with optical flow 6 pixel/frame, while the other object (red square) is at rest.

https://doi.org/10.1017/S0263574703005332 Published online by Cambridge University Press



(a)



(b)

Fig. 10. Visual servoing system with feedforward compensation.

(a) (b) (c)

Fig. 11. Test images (a) single object (white car), simple background; (b) single object, complex background; (c) two objects, simple background.

The original size of each tested image is 640×480 pixels, and is reduced to 160×120 pixels after performing image pyramid operation. The results for **Experiments #1, #2** and **#3** are provided in Tables I, II and III, respectively. Based on the experimental results provided in Tables I–III some observations are summarized:

(i) Since the modified SSD method adopts a 5-step search, thus the search region will be limited to a rectangular region within a distance of 31 pixels to the estimated center of the moving target in the previous image

https://doi.org/10.1017/S0263574703005332 Published online by Cambridge University Press

Table I. Initial search point for modified SSD: (86,34); template size: 30×30 pixels.

	Modified SSD		Modified motion energy	
Theoretic value	Estimated value	| Error |	Estimated value	| Error |
(80,40)	(80,40)	(0,0)	(82,37)	(2,3)
(74,46)	(74,46)	(0,0)	(76,43)	(2,3)
(68,52)	(68,52)	(0,0)	(70,49)	(2,3)
(62,58)	(62,58)	(0,0)	(64,55)	(2,3)
(56,64)	(55,3)	–	(58,61)	(2,3)
(50,70)	(55,3)	–	(52,67)	(2,3)

frame. If the target is moving too fast such that its new center position in the current image frame exceeds the search region, then the modified SSD will fail to find successful results. In contrast, the motion energy method can successfully detect the moving target as long as the target stays in the field of image plane.

(ii) According to the experimental results, the accuracy of the modified SSD approach is better than that of the motion energy method.

(iii) If two or more moving objects appear in the field of image plane simultaneously, often the motion energy method may give inaccurate results. In contrast, once the "template" is captured for the modified SSD method, even more than one object may appear at the same time, the modified SSD method still provides accurate results.

5.2.Real-time Visual Tracking Experiments

(A). Experiment for tracking a human object in motion. In this experiment, the prototype type pan-tilt visual tracking system is used to track a human object in motion. The visual servoing control structure illustrated in Figure 9 was adopted, where the visual controller V_c is chosen to be a simple P type controller. In addition, both the modified SSD and motion energy methods were employed to estimate the motion of the human object, respectively. Both methods exhibited satisfactory performance, for simplicity, only the results for using the modified SSD method are shown in Figure 12. Clearly, the prototype type pan-tilt visual tracking system is able to track a moving human object successfully.

(B). Experiments for tracking a moving target in the pan direction. To further explore the dynamic issues, the developed visual tracking system is controlled to track a moving target, i.e., the "Snoopy" inside of the circle shown in Figure 13. For simplicity, the moving target is mounted on a one-dimension servomechanism driven by a linear servomotor as shown in Figure 13, in which the linear servomotor is controlled to generate step-function-like and triangular-function-like motions. In other words, the output disturbance x_t to the visual tracking system that can be easily generated by the linear motor for performance evaluation is either a step function disturbance or a ramp

Table II. Initial search point for modified SSD: (83,71); template size: 60×30 pixels.

	Modified SSD		Modified motion energy	
Theoretic value	Estimated value	| Error |	Estimated value	| Error |
(84,77)	(84,79)	(0,2)	(81,80)	(3,3)
(85,83)	(84,85)	(1,2)	(81,84)	(4,1)
(86,89)	(84,91)	(2,2)	(81,91)	(5,2)
(87,95)	(85,98)	(2,3)	(82,98)	(5,3)
(88,101)	(89,89)	–	(83,104)	(5,3)
(89,107)	(89,89)	–	(83,111)	(6,4)

Table III. Initial search point for modified SSD: (86,34); template size: 30×30 pixels.

	Modified SSD		Modified motion energy	
Theoretic value	Estimated value	| Error |	Estimated value	| Error |
(80,40)	(80,40)	(0,0)	(78,94)	(2,54)
(74,46)	(74,46)	(0,0)	(62,106)	(12,60)
(68,52)	(68,52)	(0,0)	(49,103)	(19,51)
(62,58)	(62,58)	(0,0)	(53,84)	(9,26)
(56,64)	(55,3)	–	(43,101)	(13,37)
(50,70)	(55,3)	–	(38,100)	(12,30)

https://doi.org/10.1017/S0263574703005332 Published online by Cambridge University Press

function disturbance. Hence in the following experiments, the linear motor is controlled in a fashion such that the "Snoopy" target is moving with a periodic step function of 0.5 Hz and a periodic triangular function of 0.5 Hz, respectively. Throughout the experiments, $Z=85$ cm, $L=11$ cm, and effective focus length f is 1624 pixel. In addition, the control structure illustrated in Figure 10(b) is adopted, in which $K_v=56.32$ and $V_c=1/2000$.[17]

Figure 14 illustrates the experimental results for a step-function-like disturbance, in which $F_c=0$ (i.e. without feedforward compensation) and the moving distance of "Snoopy" is 15 cm. Based on the corresponding tracking error illustrated in Figure 15, it is found that the prototype real-time pan-tilt visual tracking system can track an object with a periodic-step-function-like motion effectively. Figure 16 illustrates the experimental result for a triangular-function-like disturbance, where $F_c=0$ and the moving distance of "Snoopy" is 30 cm (with velocity ± 10 cm/sec) in this case. According to its corresponding tracking error illustrated in Figure 17, it is found that without feedforward compensation, there exists a relatively large tracking error for the ramp-function-like disturbance around the place where the moving target changes its motion direction. This phenomenon is not surprising, since the prototype pan-tilt visual tracking system is a type I system. To eliminate this tracking error, a feedforward compensation $F_c=0.0534$ is added to the visual tracking system, in which the tracking performance is improved significantly as shown in Figures 18 and 19.

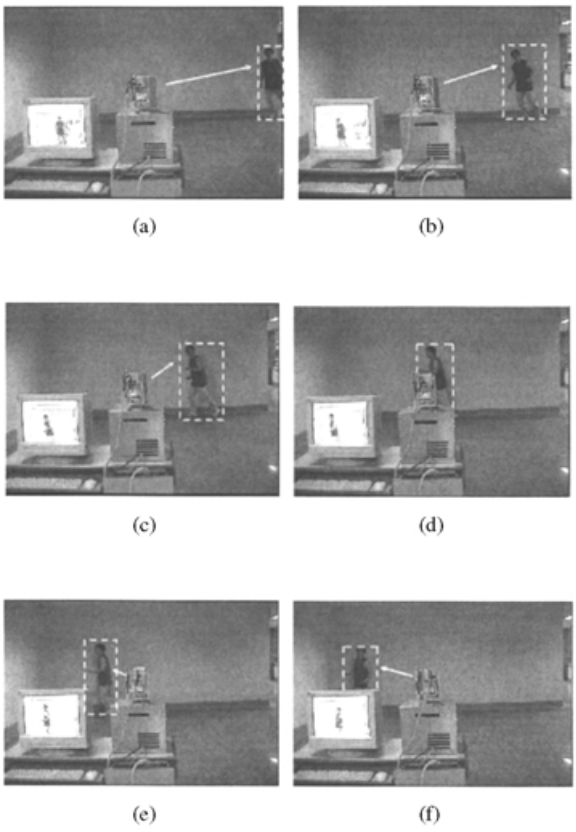


Fig. 12. Sequence of tracking a human object using SSD.

6. CONCLUSIONS

This paper has developed a real-time pan-tilt visual tracking system for automatically tracking a moving target. Both the SSD method and the motion energy method are modified and employed in this study to detect a moving target based on the consecutive images acquired. According to the experimental results, it has been found that the modified SSD method results in better and more robust motion detection performance when compared with the modified motion energy method.

Nevertheless, the SSD method demands a significant amount of computational power, hence it is combined with 3SHS in this study to effectively reduce the computation

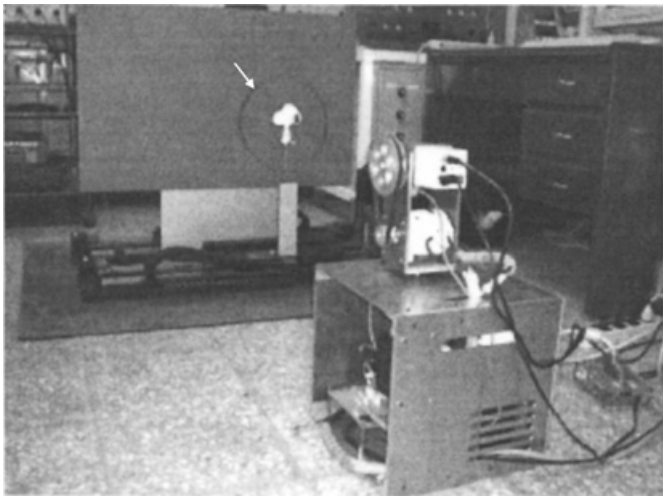


Fig. 13. The "Snoopy" inside of the circle (indicated by the white arrow) is the moving target.

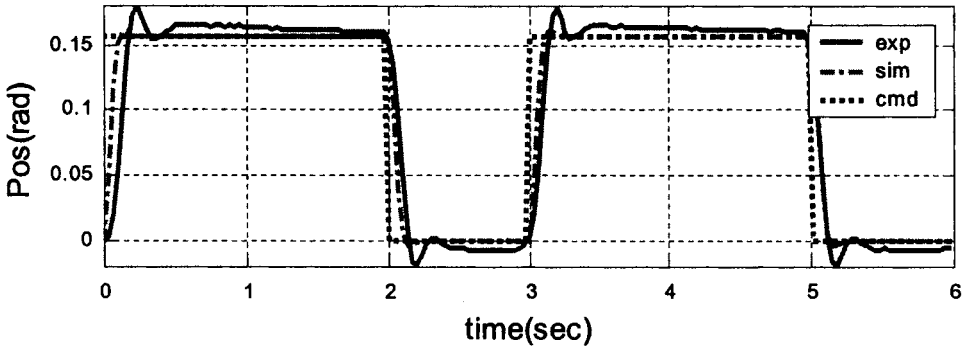


Fig. 14. Experimental results for step-function-like disturbance; θ_i: "cmd"; θ_c: "exp".

https://doi.org/10.1017/S0263574703005332 Published online by Cambridge University Press

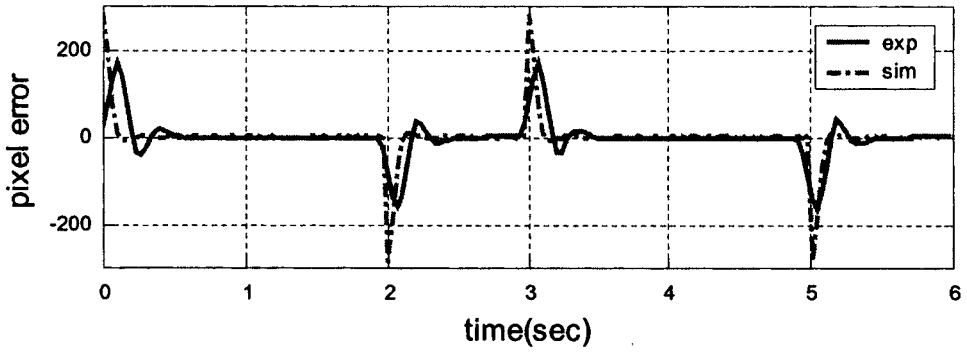


Fig. 15. Tracking error u_{error} for step-function-like disturbance.

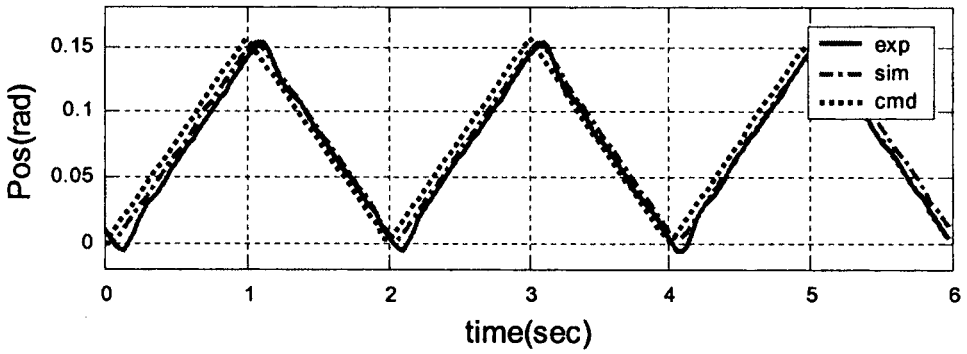


Fig. 16. Experimental results for triangular-function-like disturbance (without feedforward compensation); θ_t: "cmd"; θ_c: "exp".

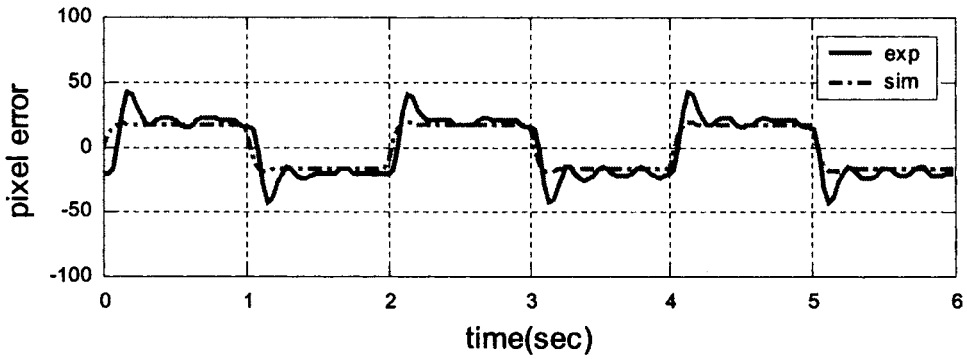


Fig. 17. Tracking error u_{error} for triangular-function-like disturbance (without feedforward compensation).

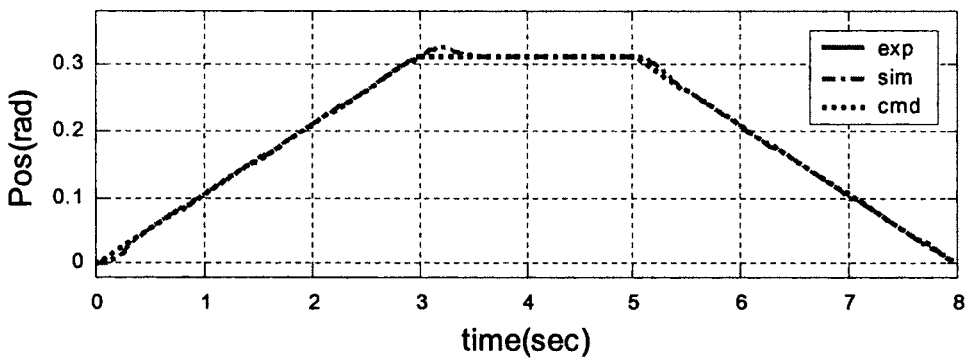


Fig. 18. Experimental results for triangular-function-like disturbance (with feedforward compensation); θ_t: "cmd"; θ_c: "exp".

https://doi.org/10.1017/S0263574703005332 Published online by Cambridge University Press

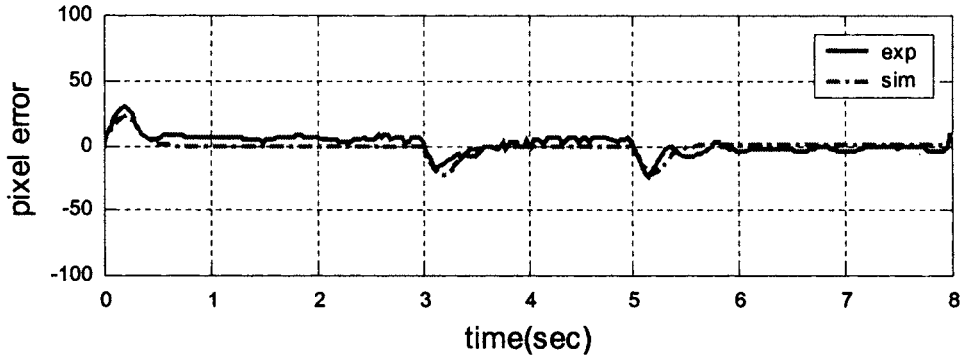


Fig. 19. Tracking error u_{error} for triangular-function-like disturbance (with feedforward compensation).

time so that the task of motion detection can be performed in real-time.

In addition, for the motion in either the pan or the tilt direction, the visual tracking system with the servo drive setting in the velocity mode is a type I system. Experimental results indicated that, due to the latency in vision feedback loop, there exists a relatively large tracking error around the place where the moving target changes its motion direction. To reduce tracking error, a servo control scheme with disturbance feedforward compensation was proposed and experimental results have shown the effectiveness of the feedforward compensation in the tracking performance.

References

1. J. Hill and W.T. Park, "Real Time Control of a Robot with a Mobile Camera", *Proceedings of 9th ISIR*, Washington, D.C. (Mar., 1979) pp. 409–417.
2. 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. on Robotics and Automation* **9**, No. 1, 14–35 (1993).
3. S. Hutchinson, G.D. Hage and P.I. Corke, "A Tutorial on Visual Servo Control", *IEEE Trans. on Robotics and Automation* **12**, No. 5, 651–670 (1996).
4. P.I. Corke and M.C. Good, "Dynamic Effects in Visual Closed-Loop Systems", *IEEE Trans. on Robotics and Automation* **12**, No. 5, 671–683 (1996).
5. J. Gangloff, M.d. Mathelin and A. Gabriel, "6 DOF high speed dynamic visual servoing using GPC controllers", *Proceedings of the 1998 IEEE International Conference on Robotics and Automation*, Leuven, Belgium (1998) pp. 2008–2013.
6. R. Kelly, R. Carelli, O. Nasisi, B. Kuchen and F. Reyes, "Stable Visual Servoing of Camera-in-Hand Robotic Systems", *IEEE/ASME Trans. on Mechatronics* **5**, No. 1, 39–48 (2000).
7. P.Y. Oh and P.K. Allen, "Visual Servoing by Partitioning Degrees of Freedom", *IEEE Trans. on Robotics and Automation* **17**, No. 1, 1–17 (2001).
8. B.K.P. Horn and B.G Schunck, "Determine Optical Flow", *Artificial Intelligence* **17**, 285–204 (1981).
9. J.L. Barron, D.J. Fleet, S.S Beauchemin and T.A Burkitt, "Performance of Optical Flow Techniques", *Proceedings of 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (1992) pp. 236–242.
10. V. Markandey, A. Reid and S. Wang, "Motion Estimation for Moving Target Detection", *IEEE Trans. on Aerospace and Electronic Systems* **32**, No. 3, 886–874 (July, 1996).
11. R. Yang and Z. Zhang, "Model-Based Head Pose Tracking with Stereovision", *IEEE International Conference on Automatic and Gesture Recognition* **20–21**, 242–247 (2002).
12. D. Murray and A. Basu, "Motion Tracking with an Active Camera", *IEEE Trans. on Pattern Analysis and Machine Intelligence* **16**, No. 5, 449–459 (1994).
13. P. Anandan, *Measuring Visual Motion from Image Sequences* (COINS Dept. Univ. of Massachusetts, Tech. Rep. COINS-TR-87-21, 1987).
14. W.J. Jong, "A Real-Time Image Tracking System Based on Optical Flow Computation", *Master Thesis* (Dept. of Elec., NCKU, Taiwan, 2001).
15. H.-M. Jong, L.-G. Chen and T.-D. Chiueh, "Parallel Architectures for 3-Step Hierarchical Search Block-Matching Algorithm", *IEEE Trans. on Circuits and Systems for Video Technology* **4**, No. 4, 407–416 (1994).
16. R.H. Brown, S.C. Schneider and M.G. Mulliden, "Analysis of Algorithms for Velocity Estimation from Discrete Position Versus Time Data", *IEEE Trans. on Industrial Electronics* **39**, No. 1, 11–19 (1992).
17. K.-Y. Chen, "Design and Implementation of a Real-Time Visual Servo Tracking System", *Master Thesis* (Dept. of Mech., NCKU, Taiwan, 2002).

https://doi.org/10.1017/S0263574703005332 Published online by Cambridge University Press