

## Dynamic Visual Tracking Using SDG-Like Matching\*

MING-YANG CHENG, MI-CHING TSAI<sup>1</sup> AND CHUN-JEN CHEN<sup>2</sup>

*Department of Electrical Engineering*

<sup>1</sup>*Department of Mechanical Engineering*

*National Cheng Kung University*

*Tainan, 701 Taiwan*

<sup>2</sup>*Optical Storage BU*

*MediaTek Inc.*

*Hsinchu, 300 Taiwan*

Motion detection/estimation plays a crucial role in dynamic visual tracking. Whether a dynamic visual tracking system can successfully track a moving target closely depends on the quality of motion detection/estimation results. In dynamic visual tracking, the camera used to capture images is not stationary, so any slight vibration of the camera motion or the target motion can lead to a blurry image causing the visual tracking performance to be deteriorated. To cope with this difficulty, a motion detection/estimation approach consisting of a region-based spatial distribution of Gaussians (SDG)-like matching algorithm, a template-update-with-memory algorithm, and a template mask is developed in this study. Moreover, linear interpolation on vision commands is performed to improve the tracking performance. A dynamic visual tracking system designed for locking the target's image in the center of the image plane is used as the test platform. Experimental results demonstrate the effectiveness of the proposed approach.

**Keywords:** dynamic visual tracking, region-based matching, spatial distribution of Gaussians, motion detection/estimation, template mask

### 1. INTRODUCTION

Applications of dynamic visual tracking [1, 2] can be found in many areas, *e.g.* military, health care, video conferencing, distance learning, security and surveillance [3, 4], *etc.* In general, vision is used to provide position and/or velocity information about the target in dynamic visual tracking problems. It acts like a feedback sensor used to close the outer loop of the overall visual tracking system. In order to provide reliable and accurate vision information about the moving target in real time, efficient and robust motion detection/estimation techniques are needed.

There are many existing motion detection/estimation techniques, *e.g.*, differential techniques [5], temporal differencing methods [6], and region-based matching methods [5, 7], *etc.* The demand for massive computing power makes the differential technique unsuitable for real-time visual tracking. On the other hand, although the temporal differencing method is efficient and easily implemented, its applications are restricted to visual tracking with stationary cameras. In order to extend the use of temporal differencing methods to the scenarios of moving cameras, Murray and Basu [6] proposed an approach that combines the idea of background compensation with morphological operation. How-

---

Received March 9, 2006; revised June 20, August 25 & October 27, 2006; accepted November 8, 2006.

Communicated by Jenq-Neng Hwang.

\* This paper was partially supported by the National Science Council of Taiwan, R.O.C., under grant No. NSC 91-2213-E-006-122.

ever, as pointed out in [6], due to the synchronization error that is inevitable in dynamic visual tracking with moving cameras, in practice, the performance of temporal differencing with background compensation is limited. An alternative approach to Murray and Basu's method was proposed by Ren *et al.* [8, 9], in which the spatial distribution of Gaussians (SDG) model is used to reduce the errors resulting from background compensation.

In general, the region-based matching method consists of a similarity measure (*e.g.*, sum of absolute difference (SAD), sum of square difference (SSD) [5, 7]) and a search algorithm (such as the renowned three-step hierarchical search (3SHS) [10]). Given a target template, the region-based matching method attempts to find a displacement vector such that the resulting similarity measure assumes the minimum. However, if a fixed target template is used throughout the tracking process, a tracking failure may occur if the shape or appearance of the moving target changes over time. In order to cope with this difficulty, Lipton *et al.* [11] proposed a template update approach. However their method may fail if the moving target is occluded by other objects.

Moreover, in many dynamic visual tracking applications, the camera used to capture images is mounted on a motorized platform. The dynamic response of the motorized platform must be quick enough to attain good tracking performance. However, when tracking fast moving objects, sometimes the target image captured by the frame grabber may become blurry due to slight vibrations of the camera motion or the target motion. As a result, the visual tracking performance will be seriously deteriorated.

This study is aimed at developing a motion detection/estimation method that can provide effective and reliable information about the moving target. The proposed motion detection/estimation method consists of three parts — a region-based SDG-like matching algorithm, a template-update-with-memory algorithm, and a template mask. Moreover, linear interpolation on vision commands is also performed to improve the tracking performance. The dynamic visual tracking system developed in our lab is used as the test platform and several experiments have been conducted to evaluate the performance of the proposed approach.

The rest of the paper is organized as follows. Section 2 gives a brief introduction to the modified temporal differencing method using an SDG-model. In section 3, a novel motion detection/estimation approach consisting of a region-based SDG-like matching algorithm, a template-update-with-memory algorithm, and a template mask is proposed. Section 4 addresses the issues concerning the visual loop controller design and vision command interpolation. Experimental results and conclusions are included in section 5 and section 6, respectively.

## 2. MODIFIED TEMPORAL DIFFERENCING METHOD USING AN SDG MODEL

One of the most popular and simple motion detection/estimation techniques is the temporal differencing method. Based on the absolute value of the intensity difference between two consecutive image frames and the edge information of the current frame, the temporal differencing method can extract the moving object in the current image frame. However, the temporal differencing method is only suitable for scenarios using

static cameras. To overcome this difficulty, Murray and Basu [6] employed the background compensation technique to extend the temporal differencing method to applications with motorized cameras. However, due to the synchronization error which is inevitable in dynamic visual tracking problems with moving cameras, a noisy image may be obtained. Although morphological operations can be used to filter out the undesired noise, the size of the morphological filter is case dependent. If it is not selected properly, the likelihood of yielding false detection results cannot be overlooked. Moreover, applying morphological operations will filter out not only the noise but also the image contents for slow target motions. In order to obtain more accurate and reliable motion detection/estimation results, Ren *et al.* [8] proposed a statistics-based approach — the SDG model. Brief reviews on the SDG model are given in the next subsection.

**2.1 Brief Review on the SDG Model**

The SDG model considers the neighborhood of every pixel in the image after performing background compensation as shown in Fig. 1, where  $x_c$  represents the pixel location in the current image  $I_c$  (before performing background compensation),  $\theta_c$  denotes the area containing all of the possible corresponding positions of  $x_c$  in the background image  $I_b$  (after performing background compensation), and  $\hat{x}_b$  is the predicted pixel location of  $x_c$  after performing background compensation.

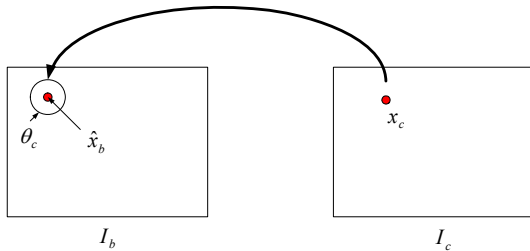


Fig. 1. The SDG model considers the neighborhood of every pixel in the image after background compensation.

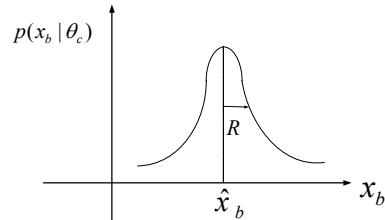


Fig. 2. Corresponding position  $x_b$  in the background map for the pixel  $x_c$  in the current image is assumed to be a Gaussian distribution about  $\hat{x}_b$ .

For the pixel  $x_c$  in the current image, its corresponding position  $x_b$  in the background map is assumed to be a Gaussian distribution about  $\hat{x}_b$  (Fig. 2) and is expressed as Eq. (1) [8, 9]

$$p(x_b | \theta_c) = \frac{1}{\sqrt{2\pi} |R|^{1/2}} \exp\left(-\frac{1}{2} (x_b - \hat{x}_b)^T R^{-1} (x_b - \hat{x}_b)\right) \tag{1}$$

where  $R$  is the covariance of the position errors.

Similarly, the background intensity distribution at  $x_b$  can be expressed as

$$p(I | B_{x_b}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \frac{(I - \bar{I}(x_b))^2}{\sigma^2}\right) \tag{2}$$

where  $\bar{I}(x_b)$  and  $\sigma$  are the mean and the standard deviation of the background distribution at  $x_b$ , respectively.

## 2.2 Modified Temporal Differencing Method Based on a Spatial Gaussian Filter

Although the performance of the SDG-based temporal differencing method seems to be promising [8, 9], however it needs a sophisticated procedure to determine proper parameter values and may require a considerable computation time. One way to speed up the computation process rather than evaluating the Gaussian distribution for position and intensity (Eqs. (1) and (2)), is to use a lowpass spatial Gaussian filter to approximate the effect of the SDG model. That is, every pixel in the previous image after background compensation is compared with a local average of intensities of all the pixels in the neighborhood of its corresponding pixel in the current image, where it can be expressed as

$$E(\hat{x}_b) = \left| I_b(k-1; x_b) - \sum_{|x-x_c| \leq d_s} w(x) I_c(k; x) \right| \quad (3)$$

where  $k$  is the frame index,  $d_s = \sqrt{2}$ ,  $w(x)$  represents a  $3 \times 3$  lowpass Gaussian spatial filter that may have the form of

1/14	1/7	1/14
1/7	1/7	1/7
1/14	1/7	1/14

In Eq. (3), if  $E(\hat{x}_b)$  is smaller than a pre-determined threshold, then pixel  $x_c$  is considered to belong to the background, otherwise it belongs to the moving target.

## 3. MODIFIED REGION-BASED SDG-LIKE MATCHING METHOD

Another popular motion detection/estimation technique is the region-based matching method, in which it consists of a similarity measure and a search algorithm. It is well known that the region-based matching methods have attractive features such as low computation load and robust tracking performance, suggesting that they are suitable for dynamic visual tracking applications. In [12], SSD was adopted as the similarity measure, and the three-step hierarchical search (3SHS) which is very popular in video compression, was chosen as the search algorithm. However, there are some drawbacks when integrating SSD with 3SHS. In particular, if the intended moving target does not locate within a three-step search range, SSD + 3SHS will still obtain a search minimum. However, this search minimum causes the algorithm to search in the wrong direction and eventually lead to a search failure. A possible approach to overcoming this difficulty is to check whether the search minimum is smaller than a prescribed threshold or not. However, the target template often contains some sort of background content. In addition, the value of the prescribed threshold is case dependent. These observations suggest that the determination of the prescribed threshold is very difficult. Moreover, if there is a shape change in the moving target during tracking, the likelihood of yielding a tracking failure

would be even larger. To cope with these difficulties, a motion detection/estimation technique that uses an SDG-like matching in replacing the SSD and also combines with a template-update-with-memory algorithm and a template mask, is developed in this study. Details concerning the proposed approach are provided in the following subsections.

### 3.1 Derivation of the Moving Target Template

The template of the moving target plays a crucial role in region-based matching. An accurate target template increases the chance of a successful tracking. Generally, the process of deriving a target template is similar to that of the temporal differencing method. Fig. 3 shows a typical example of the target template and the flowchart for deriving a target template.

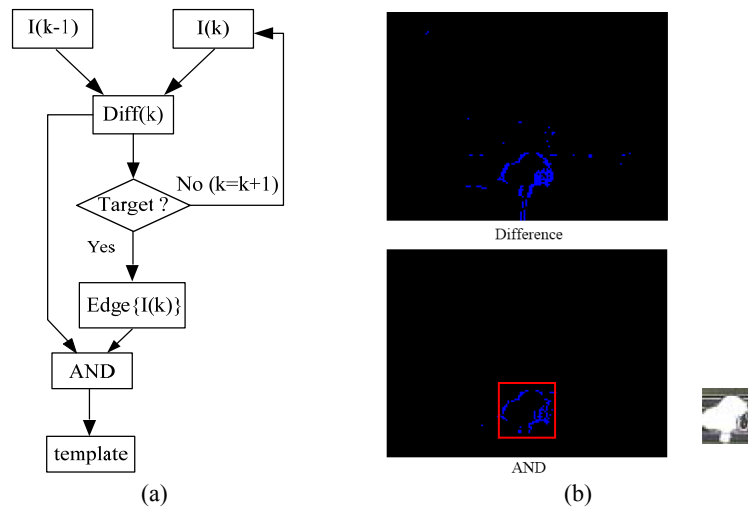


Fig. 3. (a) Flowchart for deriving the template of a moving target; (b) Typical example of the target template.

In Fig. 3 (a), the absolute value of the intensity difference between two consecutive image frames is thresholded at a suitable level to obtain a binary subtracted image that contains static and motion regions. If there is no motion region detected, then repeat the above procedure. If there are motion regions, to extract the motion region of the current image frame, an edge detection technique is applied to acquire a binary edge image. Then by performing a logical “AND” operation between the binary subtracted image and the edge information of the current image, the edges of the motion region can be extracted. The target template can be obtained by choosing a window of a suitable size that contains the motion region of the current image.

### 3.2 Replacing SSD with SDG-like Matching

Without loss of generality, consider the case that the current image is obtained from a moving camera, in which the captured image often contains some sort of noise. In this

study, to reduce the noise effect, the idea of motion detection based on an SDG model is exploited in a region-based motion detection approach. The proposed region-based SDG-like matching approach will be elaborated in the following.

For a pixel  $(u, v)$  in the template, define

$$E_{\Delta u, \Delta v}(u, v) = \left| I_t(u, v) - \sum_{m, n \in d} w(u+m, v+n) I_c(u+m+\Delta u, v+n+\Delta v) \right| \quad (4)$$

where  $E_{\Delta u, \Delta v}(\cdot)$  is the matching error,  $I_t(\cdot)$  the template,  $I_c(\cdot)$  the current image,  $w(\cdot)$  the low-pass spatial Gaussian filter,  $d$  the neighborhood of pixel  $(u, v)$ , and  $(\Delta u, \Delta v)$  the displacement vector. Eq. (4) gives the absolute intensity difference between pixel  $(u, v)$  in the template and a local average of pixel  $(u + \Delta u, v + \Delta v)$  in the current image  $I_c(\cdot)$ . In other words, in the proposed approach, Eq. (4) is used as the similarity measure and is only performed on the pixels inside the candidate window rather than the entire image.

Based on the value of  $E_{\Delta u, \Delta v}(\cdot)$ , Eq. (5) is used to determine whether the pixel  $(u + \Delta u, v + \Delta v)$  in the current frame belongs to the moving target or not

$$\hat{I}_{\Delta u, \Delta v}(u, v) = \begin{cases} 0, & E_{\Delta u, \Delta v}(u, v) > \kappa \\ 1, & E_{\Delta u, \Delta v}(u, v) \leq \kappa \end{cases} \quad (5)$$

where  $\kappa$  is a prescribed threshold.

In Eq. (5), if  $\hat{I}_{\Delta u, \Delta v}(u, v) = 1$ , the pixel  $(u + \Delta u, v + \Delta v)$  belongs to the moving target, otherwise it belongs to the background. Eq. (5) is applied to every pixel in the candidate window. The total number of the pixels inside the candidate window that belong to the moving target is defined as the similarity between the template and the candidate target region in the current image frame, which is described by

$$S(\Delta u, \Delta v) = \sum_{u, v \in I_t} \hat{I}_{\Delta u, \Delta v}(u, v). \quad (6)$$

Using the 3SHS algorithm, the distance vector  $(\Delta u, \Delta v)$  that results in the maximum similarity between the template and the candidate target region of the third step can be obtained. If the maximum similarity is larger than a prescribed threshold, then one can conclude that the moving target in the current image frame is found. Otherwise, it suggests that the moving target does not lie within the current search range. Hence a larger search range may be needed.

### 3.3 Template Update with Memory

Generally, the template used in a region-based matching method is fixed. However in many scenarios, the shape of the moving target or the illumination condition may change over time. If that is the case, then a fixed template may lead to a search failure. Lipton *et al.* [11] proposed a template update approach to deal with this problem. Consider Eq. (7),

$$P_{k+1}(u, v) = \eta I_k(u, v) + (1 - \eta)P_k(u, v); 0 < \eta < 1 \quad (7)$$

where  $P_{k+1}(u, v)$  denotes the template after updating,  $P_k(u, v)$  the template before updating,  $I_k(u, v)$  the obtained image of the moving target in the current frame, and  $\eta$  the weighting factor. Suppose the shape of the moving target undergoes some changes at the  $j$ th time instant, namely  $I_j(u, v) \neq I_{j-1}(u, v)$ . According to Eq. (7), it is easy to find that the template will converge to the image of the moving target in the current image frame if the moving target maintains its shape. In other words, “template update” provides an effective method to cope with the problem of shape changes during tracking. However, Eq. (7) may end up with a false result when the moving target is occluded by other objects. For instance, an updating sequence of the template of the moving target (Snoopy) using Eq. (7) is shown in Fig. 4. Due to the fact that the moving target is partially occluded by other objects at the time instant 30 sec, there exists a large difference between the actual target and the template as illustrated in the rightmost image of Fig. 4. This large difference may lead to a tracking failure.

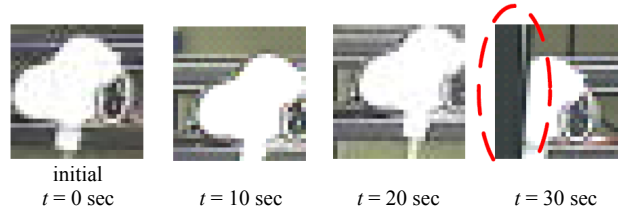


Fig. 4. Updating sequence of the template of the moving target (e.g., the Snoopy) using the Lipton’s approach. Since the moving target is partially occluded by other objects at the time instant 30 sec, there exists a large difference between the actual target and the template shown in the rightmost image.

In this study, when an object occludes the moving target, to avoid tracking failures, the Lipton’s approach is modified by adding an additional term — the initial template  $I_0$  to Eq. (7) to get

$$P_{k+1}(u, v) = \frac{1}{2} [\lambda I_0(u, v) + (1 - \lambda)I_k(u, v) + P_k(u, v)]; 0 < \lambda < 1 \quad (8)$$

where the value of  $\eta$  in Eq. (7) is chosen to be 0.5, and  $\lambda$  is referred to as the similarity coefficient in this study.

Eq. (8) can be rewritten as

$$\begin{aligned} P_{k+1}(u, v) &= \frac{1}{2} [\lambda I_0(u, v) + (1 - \lambda)I_k(u, v) + P_k(u, v)] \\ &= \lambda I_0(u, v) + (1 - \lambda) \sum_{n=1}^k \frac{1}{2^{k-n+1}} I_n(u, v). \end{aligned} \quad (9)$$

Clearly, if  $\lambda = 0$ , Eq. (9) has the same form as Eq. (7) for the case of  $\eta = 0.5$ . In contrast, if  $\lambda = 1$ , Eq. (9) degenerates into a special case — a fixed template. Therefore, by

adjusting the value of  $\lambda$ , one can determine the degree of similarity between the current template and the initial template. Due to the fact that the initial template  $I_0(u, v)$  is included in the template updating process, the proposed approach is referred to as the “template-update-with-memory” algorithm.

### 3.4 Template Mask

According to previous analysis, it is found that by combining the proposed region-based SDG-like matching method with the template-update-with-memory algorithm, robust visual tracking can be achieved even if the moving target is occluded by other objects or if it undergoes a shape change. However, since the template used for tracking consists of a moving target and some sort of background contents, if these background contents occupy too big a portion of the template, it may jeopardize the visual tracking process. One possible approach to tackling this problem is to use a template mask, where the template used for tracking is the moving target itself. This way, not only can the background contents be eliminated, but also the computing efficiency can be improved. Details concerning the template mask are elaborated in the following.

Consider two consecutive templates  $P_k(u, v)$  and  $P_{k+1}(u, v)$  at the  $k$ th and  $(k + 1)$ th time instants, respectively. According to Eq. (9), the image difference between  $P_{k+1}(u, v)$  and  $P_k(u, v)$  can be expressed as

$$\begin{aligned} m(u, v) &= P_{k+1}(u, v) - P_k(u, v) = (1 - \lambda) \left\{ \sum_{n=1}^k \frac{1}{2^{k-n+1}} I_n(u, v) - \sum_{n=1}^{k-1} \frac{1}{2^{(k-1)-n+1}} I_n(u, v) \right\} \\ &= (1 - \lambda) \left\{ \frac{1}{2} I_k(u, v) - \frac{1}{2} \sum_{n=1}^{k-1} \frac{1}{2^{k-n}} I_n(u, v) \right\}. \end{aligned} \quad (10)$$

Eq. (10) suggests that if the images of pixel  $(u, v)$  in each frame are almost the same, namely  $I_1(u, v) \approx I_2(u, v) \approx \dots \approx I_k(u, v)$ , then  $m(u, v) \approx 0$ . In other words, if  $m(u, v) \approx 0$ , then pixel  $(u, v)$  belongs to the moving target, otherwise it belongs to the background. The above observation can be expressed as

$$P_{mask}(u, v) = \begin{cases} 1, & m(u, v) \leq \varepsilon \\ 0, & m(u, v) > \varepsilon \end{cases} \quad (11)$$

where  $P_{mask}(u, v)$  is the binary image referred to as the template mask, and  $\varepsilon$  is a prescribed threshold. This way, we can segment an image frame into two parts. If  $P_{mask}(u, v) = 1$ , pixel  $(u, v)$  belongs to the moving target, otherwise pixel  $(u, v)$  belongs to the background. Since  $P_{mask}(u, v)$  can be used to identify the portion of the current template that is considered to belong to the background, then only those pixels with  $P_{mask}(u, v) = 1$  will need to perform the proposed region-based SDG-like matching approach. Therefore Eqs. (4), (5), and (7) are constrained by the condition:  $\forall (u, v) \in \{(u, v) | P_{mask}(u, v) = 1\}$ . A typical updating sequence of the template using the proposed approach is illustrated in Fig. 5. Since the initial template is included in the updating process, even though the moving target may experience shape change temporarily, the current template contains contents of the initial template so that the likelihood of misdetection will be reduced. In

addition, Fig. 6 shows the results of a dynamic visual tracking experiment using the proposed approach. In the experiment, the target undergoes a 1.0 Hz repetitive motion, in which it may cause a slight vibration of the camera motion and lead to a blurry image as shown in Fig. 6 (d). However, the visual tracking system that uses the proposed approach can still lock the target's image in the center of the image plane. The flowchart of the proposed region-based SDG-like matching approach is shown in Fig. 7.

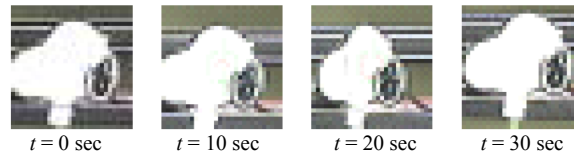


Fig. 5. A typical updating sequence of the template of the moving target (*e.g.*, the Snoopy) using the proposed approach.

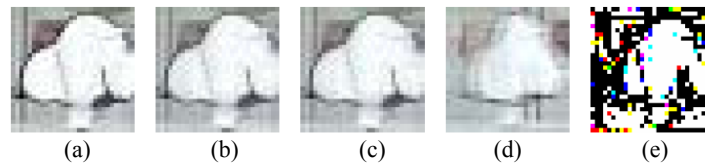


Fig. 6. A dynamic visual tracking experiment using the proposed approach. In the experiment, the target undergoes a 1.0 Hz repetitive motion. (a) Initial template; (b) Template at  $k$ th time instant; (c) Template at  $(k + 1)$ th time instant; (d) Current target image; (e) Current template mask.

#### 4. VISUAL LOOP CONTROLLER DESIGN AND COMMAND INTERPOLATION

The function of a vision system is to provide the position and/or velocity information about the moving targets to close the outer loop of the overall visual tracking system. Therefore the performance of a dynamic visual tracking system is limited by the bandwidth of the motion detection/estimation unit. Franklin and Powell [13] suggested that the sample rate of a digital control system should be 4 ~ 20 times of the designed bandwidth of a digital control system. Since most commercial frame grabbers are capable of achieving a 30 Hz frame rate, a dynamic visual tracking system should at best be capable of achieving a bandwidth between 1.5Hz and 7.5Hz [14]. However, if the target performs a high frequency motion, very likely the motion detection/estimation algorithm will fail to detect the moving target. In [12], a visual control scheme with feedforward compensation was employed to improve the tracking performance. Although the visual control scheme with feedforward compensation can indeed improve the tracking performance, the sampling time of the outer visual loop is equal to 33 ms, while the inner loop (servo loop) is implemented using a DSP-based motion card that adopts a 1 ms sampling time. This kind of multi-rate setup [15, 16] may result in unsatisfactory transient performance for the overall visual tracking system.

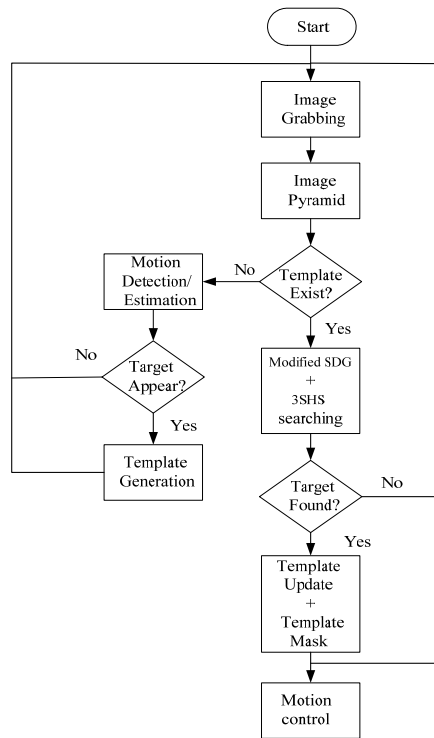


Fig. 7. Flowchart of the proposed region-based SDG-like matching method.

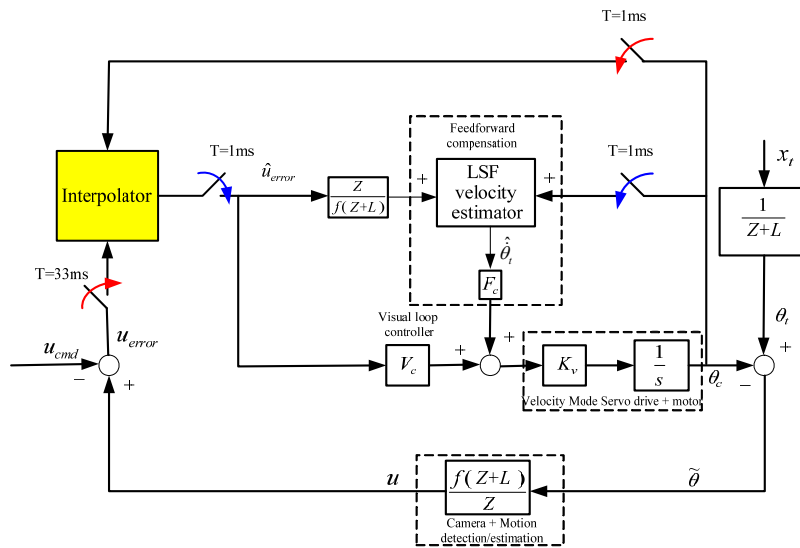


Fig. 8. Visual control scheme of the dynamic visual tracking system developed in this study (only the control in the pan direction is shown).  $x_t$  is the current target location;  $\theta_t$  is the angle corresponding to  $x_t$ ;  $\theta_c$  the camera angle in the pan direction;  $f$  the focal length;  $Z$  the depth of target;  $L$  the distance between lens center and the rotational center.

Fig. 8 illustrates the control block diagram of the dynamic visual tracking system developed in this study, where the visual loop controller is set to the  $P$  type. A 1st order linear interpolator is added to the visual loop to realize a single rate ( $T = 1\text{ms}$ ) visual tracking system. In addition, the servo drives of the servomotors are set to the velocity mode and the target's velocity  $\dot{\theta}_t$  is estimated using the LSF method [17]. In Fig. 8, the 1st order linear interpolator is described by

$$\tilde{u}[k] = u_{error}[k] + \frac{f(Z+L)}{Z} \theta_c[33k] \quad (12)$$

$$\hat{u}_{error}[33k+m] = \frac{m}{33} (\tilde{u}[k] - \tilde{u}[k-1]) + \tilde{u}[k-1] - \frac{f(Z+L)}{Z} \theta_c[33k+m] \quad (13)$$

where  $m = 1, 2, 3, \dots, 33$ , and  $\tilde{u}[\cdot]$  is the estimated target position in the  $u$  direction of the image plane. Similarly, linear interpolation on the vision command in the  $v$  direction can also be performed. Eq. (13) shows an original 33 ms vision signal being sub-sampled into a new 1 ms signal so that the inner visual tracking system has a 1 ms sample rate.

## 5. EXPERIMENTAL RESULTS

The image-based pan-tilt dynamic visual tracking system developed in our lab (Fig. 9) is used to evaluate the performance of the proposed approach, where the motions in the pan and tilt directions are driven by two AC servomotors. Three experiments have been conducted. Throughout these experiments, the visual tracking system is controlled to lock the image of the moving target in the center of the image plane. Two different moving targets — “Garfield” and “Snoopy” are used. In each experiment, a moving target is attached to a slim stick that is driven by a linear servomotor (Fig. 9). When the linear servomotor performs a rapid repetitive motion, the moving target will experience a slight vibration due to the stiffness of the stick being finite. As mentioned previously, this will increase the difficulty of visual tracking. Note that in Experiments 2 and 3, the tracking errors are defined as the distance between the current target position and the center of the image plane, which can be obtained from the motion estimation algorithms.

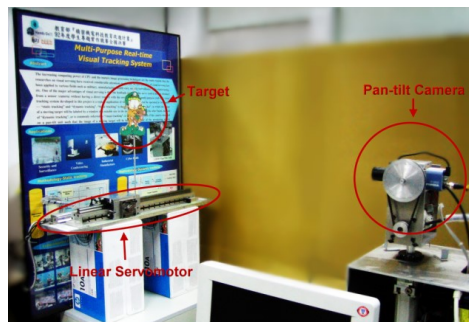


Fig. 9. Experimental setup: the moving target (e.g., “Garfield”) is mounted on a 1-D servomechanism driven by a linear servomotor. Also shown is the pan-tilt visual tracking system on the right side of the picture.

The parameter values in Eqs. (5), (9) and (11) are set to:  $\kappa = 50$ ,  $\lambda = 1/3$ ,  $\varepsilon = 10$  in all three experiments. In addition, in Eq. (4),  $d$  is a  $3 \times 3$  region and the spatial Gaussian filter  $w(x)$  is of the form

1/9	1/9	1/9
1/9	1/9	1/9
1/9	1/9	1/9

**Experiment 1:** Performance evaluation of the “template-update-with-memory” algorithm.

The purpose of this experiment is to evaluate the performance of the “template-update-with-memory” algorithm. In the experiment, the moving target “Garfield” performs a 0.5 Hz repetitive sinusoidal motion. The pan-tilt dynamic visual tracking system is controlled to lock the image of “Garfield” in the center of the image plane, in which the proposed region-based SDG-like matching approach with the template-update-with-memory algorithm and the template mask is employed in this experiment. The image sequences of the experimental results are illustrated in Fig. 10, where the “cross-circle” marker represents the center of the moving target detected by the proposed approach. In the experiments, another moving object (the “hand” in the 541th and 601th frames; “the book held by a hand” in the 1261th and 1411th frames in Fig. 10) appears in the captured images and partially occludes the real moving target “Garfield”. When using the conventional region-based matching method, a tracking failure will occur due to the temporary occlusion by other objects. In contrast, the results in Fig. 10 show that the visual tracking system using the “template-update-with-memory” algorithm can successfully lock the “Garfield” even though it is occluded by other objects.

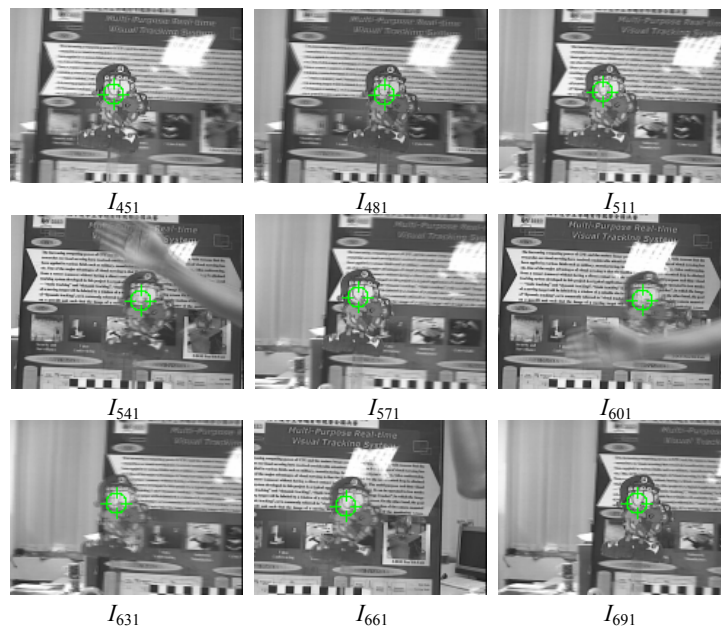


Fig. 10. Experimental results of the “template-update-with-memory” algorithm.

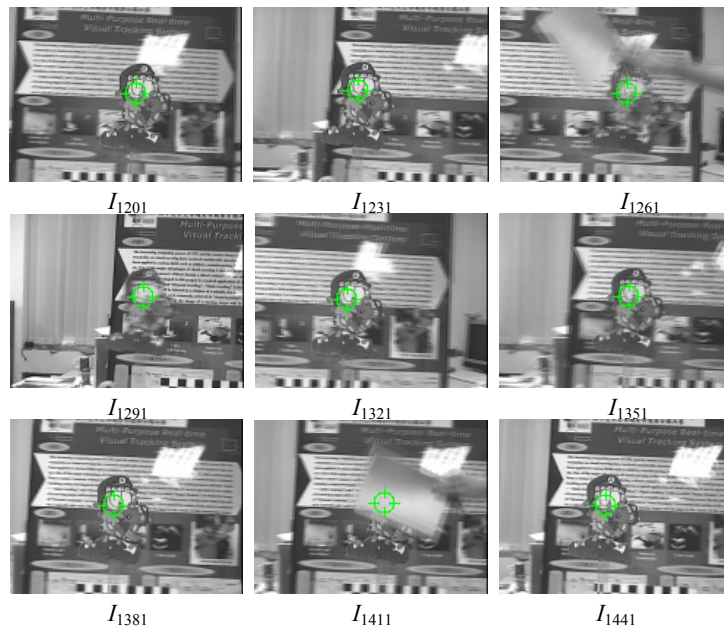
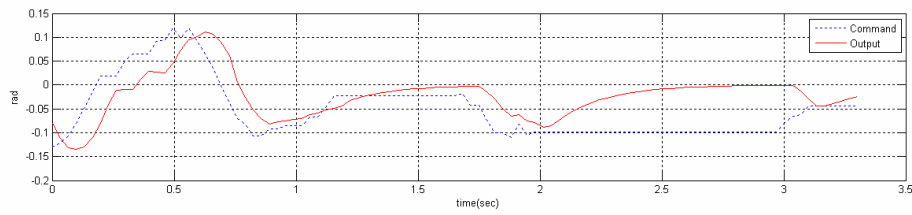


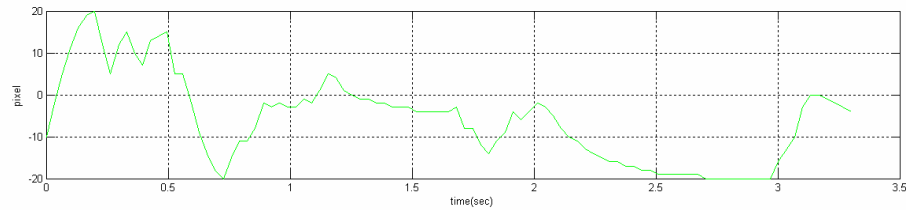
Fig. 10. (Cont'd) Experimental results of the “template-update-with-memory” algorithm.

**Experiment 2:** Performance comparison between the conventional region-based matching method and the proposed region-based SDG-like matching method.

In this experiment, the moving target “Garfield” performs a 1.0 Hz repetitive sinusoidal motion. The pan-tilt dynamic visual tracking system is controlled to lock the image of “Snoopy” in the center of the image plane. Both the conventional region-based matching method (SSD + 3SHS) and the proposed region-based SDG-like matching approach (SDG-like + 3SHS + template-update-with-memory + template mask) are adopted in this experiment. In addition, linear interpolations on the vision commands have been performed. Experimental results are illustrated in Figs. 11 and 12, where the vision command (dash line) is obtained from the first two terms on the right side of Eq. (13), and the output (solid line) is the angular position of the AC servomotor obtained from the encoder. In addition, the tracking errors (in pixel) in the image plane for both approaches are shown in Figs. 11 (b) and 12 (b). If the motion detection/estimation unit were to work ideally, both the dash lines in Figs. 11 (a) and 12 (a) would be perfect sinusoidal curves. Clearly, the dash line in Fig. 12 (a) is much more like a sinusoidal curve than the one in Fig. 11 (a). Moreover, the proposed approach yields smaller tracking errors compared with that of the conventional approach (Figs. 11 (b) and 12 (b)). These facts suggest that the proposed region-based SDG-like matching approach exhibits superior tracking performance compared with the conventional region-based matching method. In addition, between the time instants around 1.15 sec ~ 1.70 sec and 1.95 sec ~ 2.95 sec, the vision command (dash line) in Fig. 11 (a) assumes a constant value. This is because the conventional region-based matching method experienced a tracking failure during that particular period, therefore the vision command is not updated during those two periods.

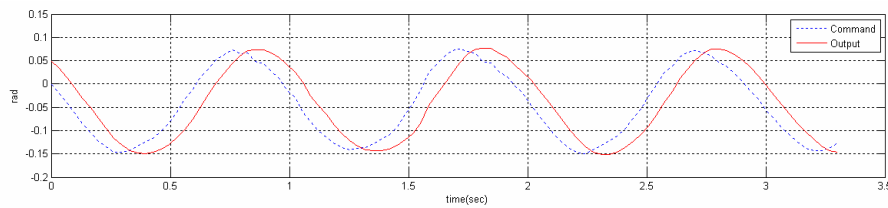


(a) Tracking results.

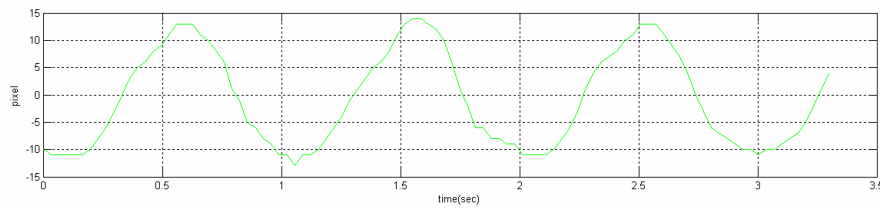


(b) Tracking error in the image plane.

Fig. 11. Results of tracking a 1.0 Hz sinusoidal moving target using the conventional region-based matching method (SSD + 3SHS).



(a) Tracking results.



(b) Tracking error.

Fig. 12. Results of tracking a 1.0 Hz sinusoidal moving target using the proposed region-based SDG-like matching approach.

### Experiment 3: Performance evaluation of the command interpolation algorithm.

In this experiment, the pan-tilt dynamic visual tracking system is controlled to track a “Snoopy” performing a 1.0Hz repetitive sinusoidal motion, so that the image of “Snoopy” will be locked in the center of the image plane. Note that in this experiment, when the moving target “Snoopy” experience a slight vibration, as shown in Experiment 2, a tracking failure will occur if the pan-tilt visual tracking system adopts the conventional region-based matching method (SSD + 3SHS). Hence only the results of the proposed approach will be shown. Two cases are considered.

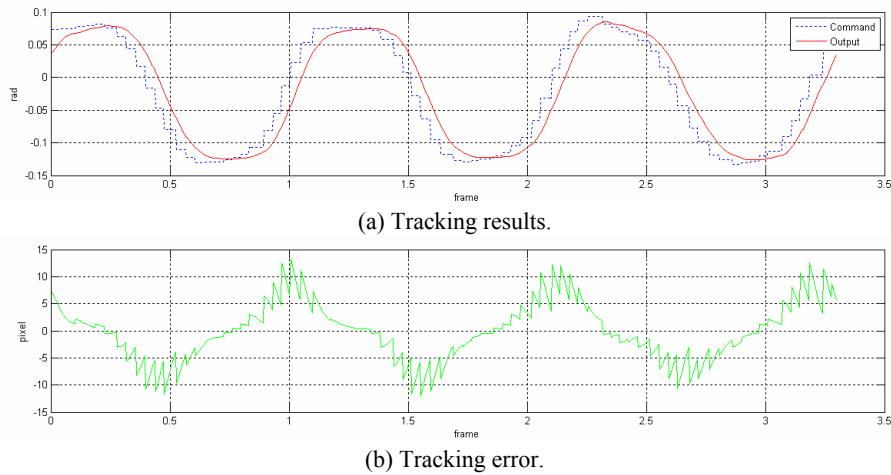


Fig. 13. Results of tracking a 1.0 Hz sinusoidal moving target using the proposed region-based SDG-like matching approach without vision command interpolation.

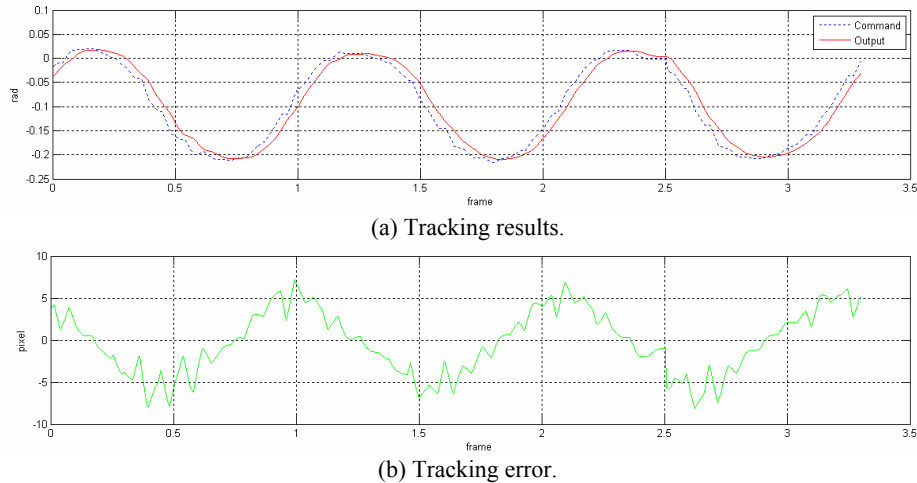


Fig. 14. Results of tracking a 1.0 Hz sinusoidal moving target using the proposed region-based SDG-like matching approach with vision command interpolation.

**Case 1:** dynamic visual tracking without vision command interpolation.

**Case 2:** dynamic visual tracking with vision command interpolation.

Experimental results are illustrated in Figs. 13 and 14, where the dash line represents the vision command obtained from the motion detection/estimation unit, and the solid line is the angular position of the AC servomotor obtained from the encoder. Note that in Case 2, the vision command is obtained from the first two terms on the right side of Eq. (13). According to Figs. 13 and 14, it is found that the tracking performance of Case 2 is better than that of Case 1. This indicates that the visual tracking performance



(a) Original target template.

(b) A portion of a snap shot of the current image.

Fig. 15. Target undergoing a 1.0 Hz repetitive sinusoidal motion and experiencing a slight vibration.

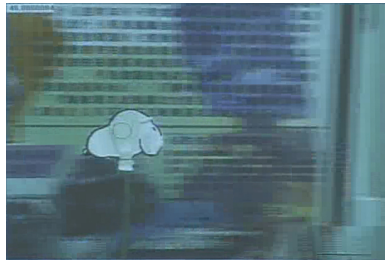


Fig. 16. Snap shot of tracking a 1.0 Hz sinusoidal moving target “Snoopy” using the proposed region-based SDG-like matching approach with vision command interpolation.

can be much improved if interpolations on vision commands are performed. Fig. 15 shows the original target template and a portion of a snap shot of the current image when tracking a 1.0 Hz repetitive sinusoidal motion. When undergoing a rapid repetitive motion, the moving target “Snoopy” attached to a slim stick will experience a slight vibration due to the stiffness of the stick being finite. As a result, the target image captured by the camera becomes blurry. Fig. 16 shows a snap shot of tracking a 1.0 Hz sinusoidal moving target “Snoopy” using the proposed region-based SDG-like matching approach with vision command interpolation. Even though the captured target image is blurry, the pan-tilt visual tracking system is able to lock the image of “Snoopy” around the center of the image plane.

## 6. CONCLUSIONS

This paper explores the problem of dynamic visual tracking. Existing region-based matching methods such as SSD + 3SHS may result in a tracking failure if the moving target experiences changes in its shape over time or is occluded by other objects. To cope with this problem, a visual tracking algorithm that combines region-based SDG-like matching with template-update-with-memory and a template mask is developed to achieve robust visual tracking results. Additionally, in order to improve the tracking performance, a feedforward compensator is incorporated into the visual control loop and linear interpolation on vision commands is also performed. The pan-tilt dynamic visual tracking system developed in our lab is used to evaluate the performance of the proposed approach. Experimental results indicate that the proposed approach indeed exhibits superior tracking performances compared with the conventional region-based matching method.

## ACKNOWLEDGMENT

The authors would like to thank to Mr. H. P. Huang and Mr. C. Y. Sun for their assistances with this work.

## REFERENCES

1. N. P. Papanikolopoulos and P. K. Khosla, "Adaptive robotic visual tracking: theory and experiments," *IEEE Transactions on Automatic Control*, Vol. 38, 1993, pp. 429-445.
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 Transactions on Robotics and Automation*, Vol. 9, 1993, pp. 14-35.
3. R. Cucchiara, G. Grana, A. Prati, and R. Vezzani, "Computer vision system for in-house video surveillance," *IEE Proceedings of Vision, Image and Signal Processing*, Vol. 152, 2005, pp. 242-249.
4. M. Velera and S. A. Velastin, "Intelligent distributed surveillance system: a review," *IEE Proceedings of Vision, Image and Signal Processing*, Vol. 152, 2005, pp. 192-204.
5. J. L. Barron, D. J. Fleet, S. S. Beauchemin, and T. A. Burkitt, "Performance of optical flow techniques," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1992, pp. 236-242.
6. D. Murray and A. Basu, "Motion tracking with an active camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 16, 1994, pp. 449-459.
7. S. Hutchinson, G. D. Hager, and P. I. Corke, "A tutorial on visual servo control," *IEEE Transactions on Robotics and Automation*, Vol. 12, 1996, pp. 651-670.
8. Y. Ren, C. S. Chua, and Y. K. Ho, "Motion detection with nonstationary background," in *Proceedings of the 11th International Conference on Image Analysis and Processing*, 2001, pp. 78-83.
9. Y. Ren, C. S. Chua, and Y. K. Ho, "Motion detection with nonstationary background," *Machine Vision and Application*, Vol. 13, 2003, pp. 332-343.
10. H. M. Jong, L. G. Chen, and T. D. Chiueh, "Parallel architectures for 3-step hierarchical search block-matching algorithm," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 4, 1994, pp. 407-416.
11. A. J. Lipton, H. Fujiyoshi, and R. S. Patil, "Moving target classification and tracking from real-time video," in *Proceedings of the 4th IEEE Workshop on Application of Computer Vision*, 1998, pp. 8-14.
12. M. C. Tsai, K. Y. Chen, M. Y. Cheng, and K. C. Lin, "Implementation of a real-time moving object tracking system using visual Servoing," *Robotica*, Vol. 21, 2003, pp. 615-625.
13. G. F. Franklin and J. D. Powell, *Digital Control of Dynamic Systems*, Addison-Wesley, 1980.
14. P. I. Corke and M. C. Good, "Dynamic effects in visual-loop systems," *IEEE Transactions on Robotics and Automation*, Vol. 12, 1996, pp. 671-683.
15. H. Fujimoto and Y. Hori, "Visual servoing based on multirate sampling control-application of perfect disturbance rejection control," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2001, pp. 711-716.

16. M. Nemani, T. C. Tsao, and S. Hutchinson, "Multi-rate analysis and design of visual feedback digital servo-control system," *Transactions of the ASME, Journal of Dynamic Systems, Measurement, and Control*, Vol. 116, 1994, pp. 47-55.
17. R. H. Brown, S. C. Schneider, and M. G. Mulligan, "Analysis of algorithms for velocity estimation from discrete position versus time data," *IEEE Transactions on Industrial Electronics*, Vol. 39, 1992, pp. 11-19.



**Ming-Yang Cheng (鄭銘揚)** was born in Taiwan, in 1963. He received a B.S. degree in Control Engineering from the National Chiao Tung University, Taiwan, in 1986. He received an M.S. and Ph.D. in Electrical Engineering from the University of Missouri-Columbia, USA, in 1991 and 1996, respectively. From 1997 to 2002, he held several teaching positions at the Kao Yuan Institute of Technology, the Dayeh University, and the National Kaohsiung First University of Science and Technology. In 2002, he joined the Department of Electrical Engineering at the National Cheng Kung University, Taiwan, where he is currently a Professor. His research interests include motion control, motor drives, visual servoing, and biped locomotion.



**Mi-Ching Tsai (蔡明祺)** was born in Taiwan, in 1956. He received both a B.S. and an M.S. degree in Electronic Engineering from the National Taiwan Institute of Technology in 1981 and 1983, respectively. He received his Ph.D. from the Department of Engineering Science at Oxford University in 1990 and became a full Professor in the Department of Mechanical Engineering at National Cheng Kung University (NCKU), Taiwan, in 1996. He was a visiting professor at Engineering Department (control group) of Cambridge University from 2003 to 2004. He is currently the director of the NCKU Electrical Motor Technology Research Center. His research interests include robust control, servo control, motor design, motor control and applications of advanced control technologies using DSP. He is a senior member of IEEE and also a Fellow of the Institution of Electrical Engineers (IEE). In addition, he has served as an Associate Editor of the IEEE/ASME Transactions on Mechatronics.



**Chun-Jen Chen (陳俊壬)** was born in Taiwan, in 1970. He received a B.S. degree in Mechanical Engineering from the National Taiwan Institute of Technology in 1996. He received an M.S. degree in Mechanical Engineering from the National Cheng Kung University, Taiwan, in 2003. He joined MediaTek Inc. in 2003, where he is currently a senior engineer at the Optical Storage BU. His research interests include visual tracking, image processing, and motion control.