Servo Tracking of Three-Dimensional Motion by the Parallel Trinocular

CHI-CHENG CHENG[1], LAN-YUAN HSU[2], GWO-LONG LIN[3], and CHIEN-HUNG CHIANG[4]
[1,2,4]Department of Mechanical and Electro-Mechanical Engineering
[1]Institute of Underwater Technology
National Sun Yat-Sen University
70 Lian-Hai Road, Kaohsiung, TAIWAN 80424
REPUBLIC OF CHINA
[1]chengcc@mail.nsysu.edu.tw [1]http://www.mem.nsysu.edu.tw/~CCCheng

[3]Department of Digital Game Design
Kao Fong College
38 Sin-Sing Road, Chang-Jhih, Pingtung, TAIWAN 90841
REPUBLIC OF CHINA

Abstract: - This study develops a novel least-squares algorithm for recovering the translational motion parameters of a moving object using a parallel trinocular vision system. Although the proposed approach overcomes the matrix singularity problem inherent in binocular observers, its implementation is somewhat complex. Accordingly, a compact closed-form version of the algorithm is proposed to facilitate real-world visual imaging applications. The closed-form scheme not only resolves the matrix singularity problem, but also avoids the requirement for matrix manipulation and is therefore computationally efficient. The validity of the closed-form scheme is verified by comparing the known translation displacements of a target object with the estimated values. The results demonstrate that the proposed scheme accurately recovers the translational motion parameters provided that the movement of the target in the depth direction is of limited magnitude only. To verify the practical applicability of the proposed scheme, a servo tracking experiment is performed in which a parallel trinocular system mounted on a servo-driven positioning platform is used to track the motion of a target object as it is moved through a series of 3D displacements. The experimental results demonstrate that the closed-form scheme enables the parallel trinocular vision system to track the target effectively.

Key-Words: - Machine vision, Optical flow, Signal processing, Tronocular, Visual servo.

1 Introduction

Estimating three-dimensional (3D) motion parameters from a sequence of images is an important task in a wide variety of image processing applications, ranging from robot vehicle guidance, to obstacle avoidance, target tracking, and the reconstruction of 3D models. However, a number of major problems remain in such active vision systems. For example, monocular observers have a simple hardware configuration, but suffer a depth-speed ambiguity, which prevents the 3D coordinates of features within the viewed scene from being easily recovered [6]. Since such systems are unable to derive absolute translation and depth values, 3D interpretation of a scene can be achieved only by applying an arbitrary scaling factor to the detected values of the translational motion and depth [1]. In addition, very few of the parameter reconstruction techniques presented in the literature are capable of providing reliable optical-flow-based reconstruction results due to the difficulties involved in extracting flow fields of sufficient accuracy from realistic scenes [6]. Moreover, many of the schemes designed to recover translational motion parameters by solving the corresponding equations of motion involve time-consuming iterative processes, which not only provide no absolute guarantee of convergence to the correct solution, but also suffer problems of nonlinearity, and thus always lead to the possibility of multiple results [5, 34].

Stereo vision systems have been widely used in many applications. The applications involve industrial detection, pattern recognition, and reverse engineering, etc. Zhang *et al*. presented an on-line detection system of crystal oscillator shell based on machine vision [36]. Machinery structure of the detecting system and the periphery control system were designed and implemented according to the requirement of product detecting. The detection of the surface quality was realized based on the Euler

number and binary area method. An object recognition technique based on human perception was also proposed recently [17]. This work was achieved through features extraction and pattern analysis using contouring and template matching approaches, respectively. In order to perform classification for moving objects, statistical bins were employed for each pattern of subparts of the moving object. Each bin contains generic and specific measures and every measure posses some weight that contributes in analyzing each bin for making the decision about similarity or dissimilarity. Besides, Caleanu *et al*. introduced a mesh adaptive direct search algorithm for the Gabor filters frequency and orientation optimization [10]. The automatic feature selection method was applied to the problem of facial recognition. A humanoid robot system that could recognize human's expression by detecting human's face and emotion was developed by Shin *et al*. [30]. This technique can be applied to an intelligent machine vision module for future robotic applications.

In a binocular vision system, the 3D position of any feature in the viewing field can be derived from two sets of image pixels corresponding to identical points on the feature. This property of binocular systems has been widely exploited in the literature to recover 3D motion and depth parameters from known optical flows [21] or to compute 3D structural information from image velocity differences in the left and right flow fields, respectively [28]. Barron and Eagleson proposed a recursive estimation scheme based upon a Kalman filter in which the time-varying motion and structure parameters of a scene were recovered by exploiting the correlation between the optical flows and the corresponding 3D physical structure [5]. Recently, Ogale and Aloimonos demonstrated the feasibility of introducing horizontal and vertical slants into the scene in order to create a first-order approximation to piecewise continuity such that shape effects could be resolved when establishing stereo correspondence in the left and right images of a binocular system [25]. However, resolving the correspondence problem in binocular systems using schemes such as those presented above is a highly complex and time-consuming process; particularly for sophisticated real-time, real-world applications such as mobile robot navigation, for example [9, 27]. The matching-phase condition used in most stereo vision techniques is generally achieved by applying certain geometrical constraints. However, while such constraints increase the amount of available image information, to facilitate a detailed reconstruction of the scene, the constraints applicable to binocular stereo vision systems are insufficient to enable the construction of a unique solution [16, 22]. Some researchers have demonstrated that trinocular vision systems, in which an additional image acquisition device is added to the binocular configuration, enable the correspondence problem to be readily resolved [14, 32] and overcome many of the limitations inherent in binocular schemes [2, 12, 27].

Recently, a significant improvement in the capabilities of integrated electronic circuits at an ever-decreasing cost has been witnessed. As a consequence, CCD cameras have emerged as a particularly attractive solution for active vision systems such as the binocular and trinocular observers described above. Of these two structures, the trinocular system is generally preferred since the addition of a third camera significantly increases the amount of image information available and therefore reduces the complexity of the motion parameter estimation task. Broadly speaking, trinocular observers can be classified as either parallel [26], surrounding [22], divergent [29], right triangular [2, 3, 26, 32], orthogonal [33], or arbitrary [7, 15, 16]. The first three configurations are coplanar, while the latter three are non-coplanar. As shown in Fig. 1, the parallel, surrounding and divergent configurations are coplanar in the *XZ* plane. Each configuration has its own unique set of geometrical constraints governed by the particular arrangement of the three CCD cameras.

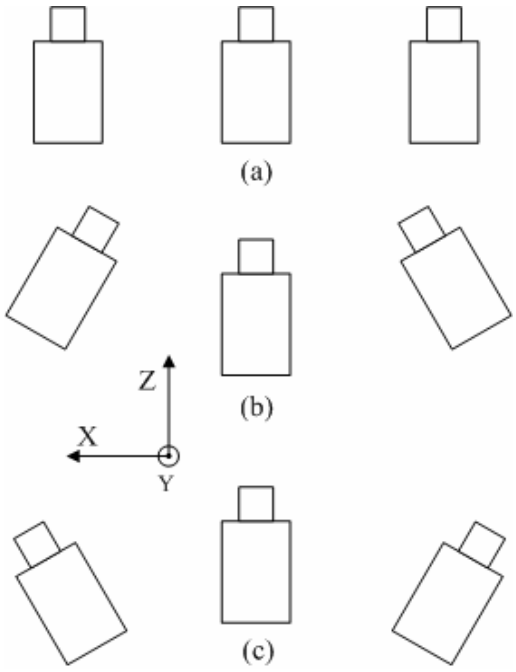

Fig. 1 Arrangements of the trinocular vision system: (a) parallel, (b) surrounding, and (c) divergent.

Ohya *et al*. [27] presented a triangulation technique for the stereo capture of targets using a binocular system. However, determining the correlation between the two images has been proved to be a complex and highly time-consuming operation. Accordingly, a third camera was added to the vision system and the resulting trinocular scheme yielded a significant reduction in the pattern matching time. The vision system was implemented on a robot, which was then trained using standard teaching and playback techniques. The experimental results showed that the robot had the ability to perform autonomous navigation based on the detection of pre-learned landmarks. Li and Duncan used the image flow fields captured by two parallel stereo cameras to determine the 3D translational motion parameters of a moving object and to establish the correspondence between equivalent features in the left and right image planes [19]. However, the disparity between the left and right images in the binocular system was insufficient to fully determine the translational motion, and hence they were obliged to introduce an additional geometrical constraint within the system to establish a phase-matching condition.

Recently, a parallel trinocular scheme for recovery of the translational motion was presented [20]. It has been indicated that the third camera not only provides more image information, but also significantly improves both efficiency and accuracy for estimation of motion parameters. The extra device not only provides more image information, but also plays a significant role regarding the solution issue. Ayache and Lustman [4] indicated the main advantages of trinocular versus binocular stereo are simplicity, reliability, and accuracy. It is believed that these advantages make trinocular stereo vision of practical use for many robotics applications. This technique has been successfully applied to indoor and industrial environments. Williamson and Thorpe [35] further presented a multi-baseline stereo using in highway obstacle detection. One of their main ideas is that the importance of the third camera to their system can be shown that additional measurements allow more accurate results by averaging noise; in the case of a large number of cameras, outliers can be rejected by voting or statistics.

As an alternative to introducing an additional constraint, the current study adds a third camera to the conventional binocular system to form a parallel trinocular vision system. A novel least-squares algorithm is then proposed to recover the translational motion parameters of an object within the system's field of view. It is shown that the proposed algorithm successfully resolves the matrix singularity problem inherent in binocular vision systems. However, the computational procedure is somewhat complex. Accordingly, the algorithm is derived in a compact closed form such that it is capable of meeting the requirements of complex, real-world machine vision applications in a real-time manner. The validity of the closed-form scheme is verified by comparing the estimated translational motion parameters of a moving object with its specified displacement settings. The applicability of the closed-form scheme is then demonstrated by performing a tracking experiment in which the parallel trinocular system is mounted on a servo-driven platform and used to track the motion of a target object in real-time as it is moved in a piecewise fashion along a 3D path.

2 Parallel Trinocular and Translational Motion

As shown in Fig.2, the parallel trinocular vision system constructed in this study comprises three identical CCD cameras arranged parallel to the *X*-axis of a spatial reference system with origin *O*. The three cameras are spaced at intervals of h_1 and h_2, respectively.

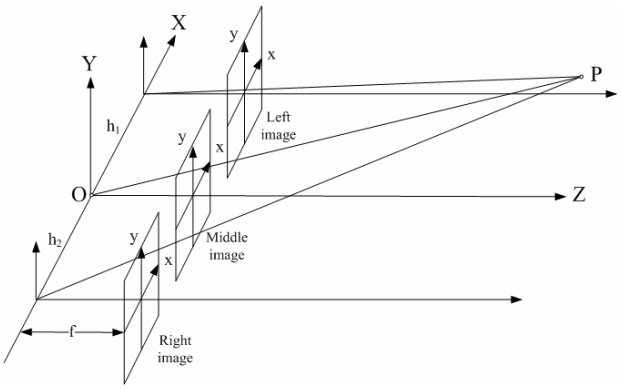

Fig. 2 Schematic illustration of parallel trinocular configuration.

For any point (*X*, *Y*, *Z*) in the spatial reference system, its projections on the image planes of the left, middle, and right CCD cameras hold the following relationships:

$$x_m - x_l = \frac{h_1 f}{Z_m},\ x_r - x_m = \frac{h_2 f}{Z_r},\ x_r - x_l = \frac{h_3 f}{Z_r}, \quad (1)$$

$$y_l = y_m = y_r = y, \quad (2)$$

where $(x_l, y_l), (x_m, y_m)$, and (x_r, y_r) are pairs of matching points on the left, middle, and right image planes respectively; $h_3 = h_1 + h_2$, f is the focal length, and Z is the depth of point P and is usually unknown.

When the three cameras in Fig. 2 are arranged in a parallel trinocular configuration, the three projected images will lie on the same XZ plane. Thus, the heights, the y coordinate values, of corresponding projection points on each image plane are all the same. Considering a special case with $h_1 = h_2$, both the y coordinate values of the corresponding projection points and the image disparities are identical in all three image planes. However, it is interesting to note that regardless of the interval distances between adjacent cameras, i.e., h_1 and h_2, the products of the image disparities in X direction always meet the following constraint:

$$(x_{mi} - x_{li})(x_{rj} - x_{mj}) = (x_{ri} - x_{mi})(x_{mj} - x_{lj}), \quad (3)$$

where i and j denote two different image pixels. (Refer to Appendix A for a detailed derivation of (3)).

Define that $\mathbf{V} = (V_x, V_y, V_z)$ is the relatively translational motion of the moving object with respect to the camera setup. The relationships between image flow velocities in the three image planes, i.e., (v_x^i, v_y^i), and the translational motion parameters can be written as:

$$\begin{bmatrix} v_x^i \\ v_y^i \end{bmatrix} = \frac{1}{Z_i} \begin{bmatrix} x_i V_z - f V_x \\ y_i V_z - f V_y \end{bmatrix}, \quad (4)$$

where the superscript i indicates the left camera (l), or the middle camera (m), or the right camera (r), respectively.

Selecting the left and middle images, a pair of optical flow equations can be established by combining velocity components in (4) in X and Y directions over all pairs of corresponding points in the two images. Assuming that the moving object lies at an equal distance from the two image planes, the resulting optical flow equations can be expressed by:

$$f\left(\sum_{\Omega_l} x_l - \sum_{\Omega_m} x_m\right) V_x + \frac{1}{2}\left(\sum_{\Omega_m} x_m^2 - \sum_{\Omega_l} x_l^2\right) V_z = \frac{h_1 f}{2}\left(\sum_{\Omega_l} v_x^l + \sum_{\Omega_m} v_x^m\right), \quad (5)$$

$$f\left(\sum_{\Omega_l} x_l - \sum_{\Omega_m} x_m\right) V_y + \left(\sum_{\Omega_m} x_m y_m - \sum_{\Omega_l} x_l y_l\right) V_z = \frac{h_1 f}{2}\left(\sum_{\Omega_l} v_y^l + \sum_{\Omega_m} v_y^m\right) \quad (6)$$

where Ω_l and Ω_m respectively represent corresponding image regions in the left and middle image planes.

It appears that the translational motion is not a function of the depth since the depth term disappears in the binocular scheme described in (5) and (6). In order to recover the three unknown translational motion parameters, i.e., V_x, V_y, and V_z, additional constraint equations have to be included. Consequently, by choosing the left-middle and middle-right image pairs, the final equations can be expressed in the following matrix form:

$$\mathbf{Ax} = \mathbf{d},$$

where $\mathbf{x}$ is the unknown translational motion parameters which are to be recovered, $\mathbf{A}$ represents the image matrix, and $\mathbf{d}$ is the optical flow vectors of the trinocular system, i.e.,

$$\mathbf{x} = (V_x, V_y, V_z)^T,$$

$$\mathbf{A} = \begin{bmatrix} f\left(\sum x_l - \sum x_m\right) & 0 & (1/2)\left(\sum x_m^2 - \sum x_l^2\right) \\ 0 & f\left(\sum x_l - \sum x_m\right) & \sum x_m y_m - \sum x_l y_l \\ f\left(\sum x_m - \sum x_r\right) & 0 & (1/2)\left(\sum x_r^2 - \sum x_m^2\right) \\ 0 & f\left(\sum x_m - \sum x_r\right) & \sum x_r y_r - \sum x_m y_m \end{bmatrix},$$

$$\mathbf{d} = f \begin{bmatrix} (h_1/2)\left(\sum v_x^l + \sum v_x^m\right) \\ (h_1/2)\left(\sum v_y^l + \sum v_y^m\right) \\ (h_2/2)\left(\sum v_x^m + \sum v_x^r\right) \\ (h_2/2)\left(\sum v_y^m + \sum v_y^r\right) \end{bmatrix}.$$

Therefore, the translational motion can be quickly recovered with the standard least squared approach:

$$\mathbf{x} = \left(\mathbf{A}^T \mathbf{A}\right)^{-1} \mathbf{A}^T \mathbf{d}.$$

Generally speaking, the above derivation can be employed to both point-to-point correspondence and patch-to-patch matching cases.

In order to ensure the applicability of the above proposed algorithm, the singularity problem of the image matrix $\mathbf{A}$ needs to be further investigated. For the purpose of simplicity and clarity, let the matrix $\mathbf{A}$ be defined as

$$\mathbf{A} = \begin{bmatrix} fa_1 & 0 & a_2 \\ 0 & fa_1 & a_3 \\ fb_1 & 0 & b_2 \\ 0 & fb_1 & b_3 \end{bmatrix}, \quad (7)$$

where

$a_1 = \sum x_l - \sum x_m$,
$a_2 = (1/2)\left(\sum x_m^2 - \sum x_l^2\right)$,
$a_3 = \sum x_m y_m - \sum x_l y_l$,
$b_1 = \sum x_m - \sum x_r$,
$b_2 = (1/2)\left(\sum x_r^2 - \sum x_m^2\right)$, and
$b_3 = \sum x_r y_r - \sum x_m y_m$.

The determinant of $\mathbf{A}^T\mathbf{A}$ to recover translational motion parameters in the proposed parallel trinocular can be derived as

$$\det(\mathbf{A}^T\mathbf{A}) = f^4(a_1^2 + b_1^2)\left[(a_1b_2 - a_2b_1)^2 + (a_1b_3 - a_3b_1)^2\right] \quad (8)$$

The above formulation is fully derived in Appendix B. Since three CCD cameras are arranged in a parallel trinocular configuration, disparities in *X* direction exist in the three images so that a_1 and b_1 are non-zero values. Moreover, it can be verified that a_1b_2 is not equal to a_2b_1 and $a_1b_3 - a_3b_1$ diminishes to zero. As a result, the determinant of $\mathbf{A}^T\mathbf{A}$ appears to be always positive. Thus, it can be concluded that the least squares solution of the parallel trinocular configuration resolves the matrix singularity difficulty associated with binocular vision systems.

3 Compact Closed-Form Solution

Although the least-squares estimation method presented above is capable of recovering the translational motion parameters of an object in the viewing field of a parallel trinocular system, its implementation is somewhat complex. Accordingly, this section develops a compact closed-form version of the estimation algorithm in order to reduce the computational complexity of the parameter recovery process.

Let the optical flow vectors of the parallel trinocular scheme be defined as

$$\mathbf{d} = f[d_1 \quad d_2 \quad d_3 \quad d_4]^T,$$

where

$d_1 = (h_1/2)\left(\sum v_x^l + \sum v_x^m\right)$,
$d_2 = (h_1/2)\left(\sum v_y^l + \sum v_y^m\right)$,
$d_3 = (h_2/2)\left(\sum v_x^m + \sum v_x^r\right)$, and
$d_4 = (h_2/2)\left(\sum v_y^m + \sum v_y^r\right)$.

The translational components in the X-, Y- and Z-axis directions can then be derived from:

$$V_x = \frac{b_2d_1 - a_2d_3}{a_1b_2 - a_2b_1}$$
$$V_y = \frac{a_1d_2 + b_1d_4}{a_1^2 + b_1^2} + \frac{(a_1a_3 + b_1b_3)(b_1d_1 - a_1d_3)}{(a_1^2 + b_1^2)(a_1b_2 - a_2b_1)} \quad (9)$$
$$V_z = \frac{f(a_1d_3 - b_1d_1)}{a_1b_2 - a_2b_1}$$

In contrast to the least-squares estimation procedure, this compact closed-form scheme avoids the requirement for complex matrix manipulation and is therefore more computationally efficient. As a result, it represents an ideal solution for the estimation of translational motion parameters in complex, real-world visual imaging applications.

To validate the closed-form algorithm, three identical CCD cameras with a focal length of 530 pixels were arranged in the parallel trinocular configuration shown in Fig. 2 with the middle camera positioned on the *Z*-axis. The left and middle cameras were spaced at an interval of 90 mm, and the middle and right cameras at an interval of 70 mm. In the resulting trinocular configuration, any point in the 3D space projects simultaneously onto the left, middle and right image planes. For the purpose of future applications in the validation procedure, the translational motion was calculated using a unit sampling time. In this way, the translational velocity is equivalent to the relative displacement between successive sampling events and the ideal image flow velocity can be determined. As shown in Fig. 3, fifty non-coplanar points were randomly selected in the 3D space for reference purposes during the parameter reconstruction process. Figure 4 shows the corresponding projections of these points on the left, middle and right image planes.

To simulate the performance of the closed-form algorithm, 11 different translational motions were simulated involving displacements of 0.1 mm, 3mm, 5mm, 7mm, 10 mm, 15mm or 20 mm in the *XY*, *Z* or *XYZ* planes, and the corresponding motion parameters were then recovered. The translational motions are summarized in Table 1 together with the corresponding motion parameter estimates.

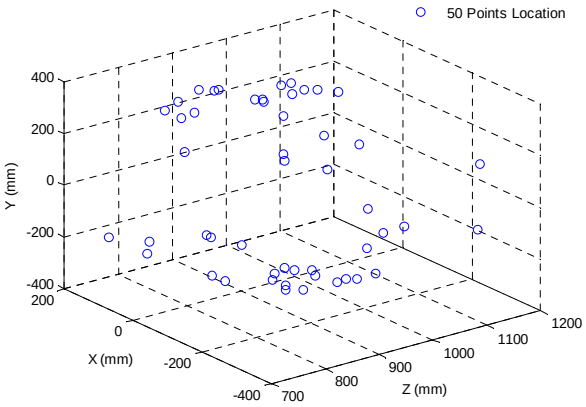

Fig. 3. 50 non-coplanar points used for reference purposes during validation experiments.

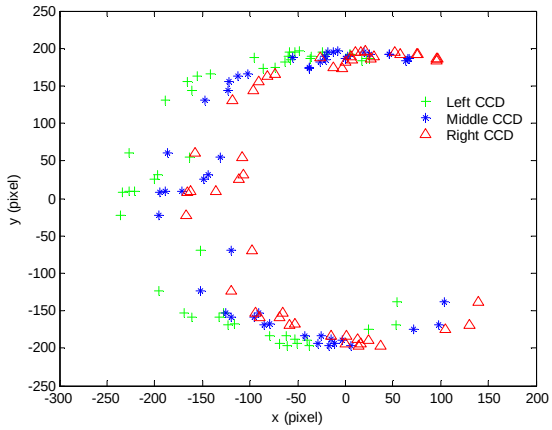

Fig. 4. Projections of 50 test points on left, middle and right image planes.

Table 1
Performance of closed-form scheme in recovering translational motion parameters (unit: mm)

Translational motion parameters			Recovered translational motion parameters		
X	*Y*	*Z*	*X*	*Y*	*Z*
0.1	0.1	0	0.1	0.1	0
0	0	0.1	0	0	0.1
0.1	0.1	0.1	0.1	0.1	0.1
10	10	0	10	10	0
15	15	0	15	15	0
20	20	0	20	20	0
0	0	3	0.0000	0.0000	2.9898
0	0	5	0.0001	0.0000	4.9718
0	0	7	0.0001	0.0000	6.9449
15	15	15	14.7529	14.7524	14.7494
20	20	20	19.5632	19.5622	19.5570

Observing the results, a number of conclusions can be drawn, namely:

1) For very small displacements (i.e. 0.1 mm) in the *XY*, *Z* or *XYZ* planes, the recovered motion parameters are identical to the actual values.

2) Provided that no motion takes place along the *Z*-axis, the recovered parameter values correspond exactly to the actual displacements along the *X*- and *Y*-axis directions, respectively, irrespective of the magnitude of the displacements.

3) When the displacements of the translational movements exceed 0.1 mm and include a motion component along the *Z*-axis, the recovered motion parameters for all three axes deviate from their actual values. The degree of the estimation error is dependent on the magnitude of the displacement in the *Z*-axis direction.

It appears that the parameter reconstruction errors are crucially dependent upon the magnitude of the displacement component in the *Z*-axis direction. Therefore, motion along the *Z*-axis should ideally be suppressed.

As commented above, minimizing the translational movement along the *Z*-axis enables the translational motion parameters to be more accurately determined. The parameter estimates for a translational motion of (15 mm, 15 mm, 2 mm), (20 mm, 20 mm, 2 mm) are shown in Table 2.

Table 2
Performance of closed-form scheme in recovering translational motion parameters when Z-axis displacement is suppressed (unit: mm)

Translational motion parameters			Recovered translational motion parameters		
X	*Y*	*Z*	*X*	*Y*	*Z*
15	15	2	14.9665	14.9665	1.9955
20	20	2	19.9553	19.9553	1.9955

It can be seen that the recovery errors are reduced to less than 0.22%. In other words, the results confirm that suppressing the magnitude of the *Z*-axis translational motion ensures that the displacement parameters can be recovered with an improved degree of accuracy.

4 Tracking Experiments

To demonstrate feasibility of conducting servo tracking tasks using the proposed parallel trinocular approach, an image tracking experimental system was established. The visual C++ is adopted in this

trinocular stereo visual system to control the process with a parallel trinocular tracking configuration shown in Fig. 5. Both stations A and B have identical structure, but respectively holding an image poster and a video camera set comprising of three parallel CCDs. These two stations are driven by stepping motors for three dimensional translation motions. The objective is to have the station B follow the behavior of the station A in order to maintain constant relative position based on the image patterns acquired by the cameras. The acquired images are with resolution of 320x240 pixels. The distance between the left and the middle cameras was 70 mm, whereas the distance between the right and the middle ones was 68 mm. The arrangement of parallel trinocular is shown in Fig. 6. It is noted that in the following task, the distance between the poster and lens is unknown.

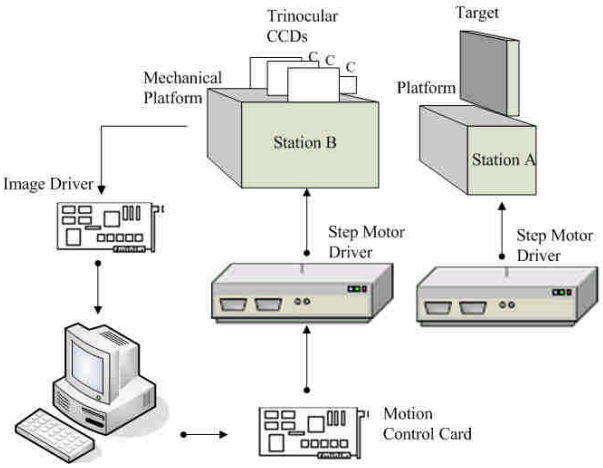

Fig. 5 Schematic illustration of parallel trinocular tracking configuration.

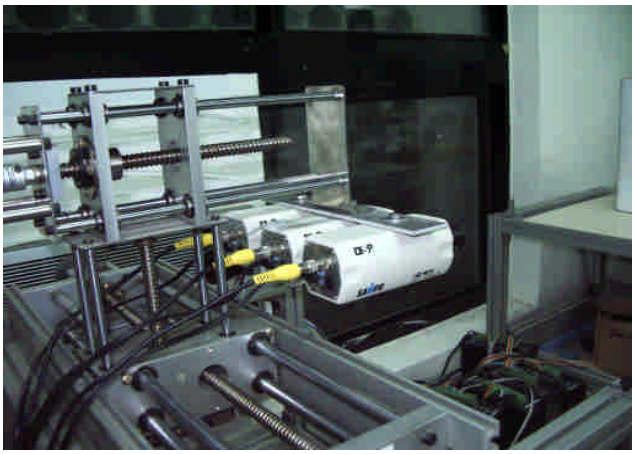

Fig. 6 Arrangement of three CCD cameras in the parallel trinocular vision system.

4.1 Camera calibration

The focal length (FL) and principal points (PP) of the three CCD cameras were estimated using the calibration algorithm presented by Zhang [37] and the standard MATLAB-based calibration toolbox developed by Bouguet [8]. The estimated intrinsic parameters of all three CCDs are summarized in Table 3.

Table 3
Estimated intrinsic parameter of three CCD cameras in the parallel trinocular vision system (unit: pixel)

	X			*Y*		
CCD	Left	Middle	Right	Left	Middle	Right
FL	533.16 ±1.39	538.14 ±1.65	534.55 ±1.62	529.12 ±1.32	534.18 ±1.57	530.76 ±1.54
PP	154.43 ±1.62	170.59 ±1.65	160.48 ±1.59	123.00 ±1.23	111.18 ±1.22	117.76 ±1.20

In implementing the closed-form solution, the focal lengths of the three CCD cameras were assumed to be identical and were assigned to the mean value of the FL results given in Table 3, i.e. 533.32 pixels. In practice, the calculated values of the principal points (PP) of the three cameras cannot be used directly for alignment of the three images in the *X*- and *Y*-axis directions due to the special configuration of the parallel trinocular. As a result, an alignment procedure needs to be considered to meet the requirement of Eqs. (1) and (2).

In theory, the image height of *y* axis and distance along *x* axis can be corrected by placing these three CCDs on a perfectly horizontal plane and adjusting the *Y*- and the *X*-axis coordinates of the three images in accordance with the calibrated values of the principal points of the three CCDs. However, in practice, this is impossible due to the comparatively poor positioning performance of the stepping motor driving system. Since the calibrated principal points could not be used directly, the datum points of *X*- and *Y*-axes need to be adjusted respectively. In the current experimental setup, this is achieved as follows:

1) The position of images along the *X* axis is adjusted so that the difference in the disparities between the left-middle and middle-right CCD image sets, respectively, is minimized.
2) The *Y*-axis height of the CCD images is modified by setting the center of the image of the middle CCD to the calibrated PP value and then adjusting the heights of the left and

right images, respectively, such that they are equal to that of the central image.

The resulting image centers in the parallel trinocular scheme are summarized in Table 4.

Table 4
Image centers in the parallel trinocular vision system (unit: pixel)

	Left	Middle	Right
X axis	+3.95	-7.36	-2.07
Y axis	-0.29	+9.78	-0.43

4.2 Estimation of translation motion

Prior to tracking tests, the validity of the proposed closed-form scheme was verified by driving Station A with displacements of varying magnitudes in the *XY*, *Z* and *XZY* planes, respectively, and then estimating the corresponding translational motion parameters using the parallel trinocular system mounted on Station B. The corresponding results are summarized in Table 5.

Table 5
Comparison between actual translations and estimated translations (unit: mm)

Actual translations			Estimated translations		
X	*Y*	*Z*	*X*	*Y*	*Z*
10	10	0	10.2277	10.4544	0.2204
15	15	0	15.0612	15.2314	0.3384
20	20	0	19.8168	19.2955	1.3064
0	0	3	0.8342	0.3312	2.8683
0	0	5	0.7917	1.3999	5.1455
0	0	7	1.8516	1.3700	6.1966
15	15	2	15.6381	15.7348	1.7697
20	20	2	20.2748	19.6554	1.5016

These results support a number of important conclusions, namely

1) Provided that Station A does not move in the *Z*-axis direction, the parallel trinocular system is capable of estimating the translational motion parameters with an accuracy of approximately 5%, irrespective of the magnitude of the displacements along the *X*- and *Y*-axis directions. The error in the recovered results is most reasonably attributed to noise in the detected images in the parallel trinocular system.
2) When the translational movement comprises a *Z*-axis component, the estimated values of the translational motion parameters deviate notably from the command values. The magnitude of the error varies directly with that of the *Z*-axis displacement, and reaches as much as 35% in the current experiments. Thus, such results indicate that recovery performance of the translation motion is more sensitive to the perturbation in *Z* axis than the perturbations in the *X* and *Y* axes.
3) Comparing the results for the *XY* and *Z* plane displacements in Table 5, it is apparent that the performance of the parallel trinocular system in recovering the motion parameters is significantly more sensitive to perturbations in the displacement in the *Z*-axis direction than in the *X*- or *Y*-axis directions [34].
4) Observing the parameter estimation results obtained for 3D displacements of (15, 15, 2) and (20, 20, 2), respectively, it can be seen that the estimation errors are reduced to 5%. In other words, provided that the magnitude of the displacement in the *Z*-axis direction is restricted to a low value (e.g., 2 mm), the values of the translational motion parameters recovered using the closed-form scheme attain a satisfactory level of accuracy.

4.3 Servo tracking experiments

The results presented above confirm the validity of the proposed closed-form solution scheme. To verify its practical applicability, a servo tracking experiment was performed in which a target (a poster mounted on a board) attached to Station A was moved in accordance with a given set of path commands and was tracked by the parallel trinocular system mounted on Station B. The station B followed the movement information extracted according to the optical flow calculated from consecutive images and relocated the position of the camera set to accomplish the tracking task. The detailed steps in the tracking procedure were as follows:

1) The three CCD cameras captured the target image and a Gaussian filter was applied to the resulting output signals of the parallel trinocular system.
2) Station A was driven through a specified displacement in the *X*-, *Y*- and *Z*-axis directions.
3) The three CCD cameras recaptured the target image and the output signals of the parallel

trinocular system were again processed using a Gaussian filter.

4) The optical flow vector was computed by comparing the two acquired images.
5) The optical flow data were supplied as inputs to the proposed closed-form scheme to estimate the translational motion parameters of Station A.
6) The estimated translational motion parameters were used to compute command signals to drive Station B through appropriate *X*-, *Y*-, and *Z*-axis displacements such that it was realigned with Station A.

In the tracking experiment described above, Station A was driven through five consecutive 3D displacements of (4, 4, 1), where the units are in mm. Table 6 and Figure 7 compare the command path with the estimated path and confirm that a good agreement is obtained between the two 3D trajectories.

Table 6
Comparison between actual translations and estimated translations (unit: mm)

Actual translations			Estimated translations		
X axis	Y axis	Z axis	X axis	Y axis	Z axis
0	0	0	0	0	0
4	4	1	4.1351	4.1724	1.0168
8	8	2	7.7655	7.7374	1.8914
12	12	3	11.8837	11.963	2.8144
16	16	4	15.9714	15.974	4.1645
20	20	5	19.9192	19.7846	4.9699

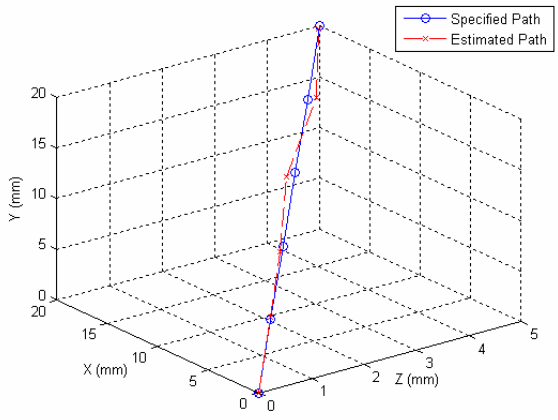

Fig. 7 Tracking paths in three dimensional space.

The servo tracking performance of the parallel trinocular system was further demonstrated by mounting a laser pen pointer on Station B and arranging it such that its projected light spot was incident on a pre-selected point on the poster mounted on Station A. Three images captured by the parallel trinocular system are illustrated in Fig. 8.As shown in the video clip presented in [11], the light spot maintained its preset position on the poster as Station A was displaced through the command path described in Fig. 7.

Fig. 8 Images acquired by the parallel trinocular vision system.

5 Conclusion and Future Work

This study has developed a closed-form solution procedure for estimating the translational motion parameters of a moving object using a parallel trinocular system. The proposed approach has a simple numerical implementation, avoids the requirement for matrix manipulation and resolves the matrix singularity problem inherent in conventional binocular vision systems. The performance of the proposed approach has been demonstrated by comparing the recovered translational motion parameters with the command values for various displacements of the target object in the *XY*, *Z* and *XYZ* planes. In general, the results have shown that the closed-form scheme yields acceptable estimates of the motion parameters provided that the magnitude of the *Z*-axis component of the displacement is limited to a low value (e.g. 2 mm). Having verified the performance of the closed-form solution scheme, it was employed as the basis for a parallel trinocular servo tracking system. The experimental results confirmed the ability of the parallel trinocular system to follow the target object with a high degree of accuracy.

In practical applications, the accuracy of the tracking system is inevitably degraded to a certain extent by environmental noise. Consequently, a future study will aim to develop an adaptive approach based upon the closed-form solution scheme presented in this study to further improve the reliability of the tracking performance. Although the features themselves can be tracked over time in the image domain, detection of target is often the most computationally expensive and error prone component of the tracking system. Using an extended Kalman filter to reduce the errors of

motion components and developing robust methods to extract features being tracked are also worthwhile to be further investigated.

Furthermore, many researchers put their efforts on the compound eye framework based on inspiration of nature [13, 18, 23, 24, 31]. From the viewpoint of the computer vision, the spherical surface compound-like eye, and the planar array compound-like eye are two major categories [3]. The parallel trinocular stereo vision studied in this paper falls into the type of the planar array compound-like eye. Extension from the parallel trinocular can be incorporated into the planar array compound-like eye and a new trend in compound eye study is therefore highly expected.

6 Acknowledgment

This work was funded by National Science Council, Republic of China under contract NSC 93-2212-E-110-014.

References:

[1] Adiv, G., "Determining three-dimensional motion and structure from optical flow generated by several moving objects," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 7, No. 4, 1985, pp. 384-401.

[2] Agrawal, M. and Davis, L. S., "Trinocular stereo using shortest paths and the ordering constraint," in *Proc. IEEE Workshop on Stereo and Multi-Baseline Vision*, 2001, pp. 3-9.

[3] Aloimonos, Y., "New camera technology: Eyes from eyes," *http://www.cfar.umd.edu/~yiannis/resaerch.html*.

[4] Ayache, N. and Lustman, F., "Trinocular stereo vision for robotics," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 13, No. 1, 1991, pp. 73-85.

[5] Barron, J. L. and Eagleson, R., "Recursive estimation of time-varying motion and structure parameters," *Pattern Recognition*, Vol. 29, No. 5, 1996, pp. 797-818.

[6] Barron, J. L., Jepson, A. D., and Tsotsos, J. K., "The feasibility of motion and structure from noisy time-varying image velocity information," *International Journal of Computer Vision*, Vol. 5, No. 3, 1990, pp. 239-269.

[7] Blake, A., McCowen, D., Lo, H. R., and Lindsey, P. J., "Trinocular active range-sensing, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 15, No. 5, 1993, pp. 477-483.

[8] Bouguet, J. Y., "Complete camera calibration Toolbox for Matlab," http://www.vision.caltech.edu/bouguetj/.

[9] Brown, M. Z., Burschka, D., and Hager, G. D., "Advance in computational stereo," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 25, No. 8, 2003, pp. 993-1008.

[10] Caleanu, C.-D., Gui, V. and Alexa, F., "Direct search optimized feature extraction," *WSEAS Transactions on Systems and Control*, Vol. 1, No. 2, 2006, pp. 113-120.

[11] Cheng, C. C., Hsu, L. Y., Lin, G. L., and Chiang, C. H., "The experiment of servo tracking of translational motion by the parallel trinocular," http://www2.nsysu.edu.tw/MMSL/.

[12] Chiou, R. N., Chen, C. H., Hung, K. C., and Lee, J. Y., "The optimal camera geometry and performance analysis of a trinocular vision system," *IEEE Trans. Systems, Man, and Cybernetics*, Vol. 25, No. 8, 1995, pp. 1207-1220.

[13] Collett, T., "Animal behaviour: Survey flights in honeybees," *Nature*, Vol. 403, 2000, pp. 488-489.

[14] Gurewitz, E., Dinstein, I., and Sarusi, B., "More on the benefit of a third eye," in *Proc. Int. Conf. on Pattern Recognition*, 1986, pp. 966-968.

[15] Hemayed, E., Ahmed, M. T., and Farag, A., "CardEye: A 3D trinocular active vision system," in *Proc. IEEE Int. Conf. on Intelligent Transportation Systems*, 2000, pp. 398-403.

[16] Hemayed, E. E. and Farag, A. A., "A geometrical-based trinocular vision system for edges reconstruction," in *Proc. 10th International Conference on Image Processing*, 1998, pp. 162-166.

[17] Iftikhar, N., Iqbal, S. and Khattak, N. S., "Location and classification of moving objects through template matching," *WSEAS Transactions on Systems and Control*, Vol. 2, No. 2, 2007, pp. 201-205.

[18] Iida, F., "Biologically inspired visual odometer for navigation of a flying robot," *Robotics and Autonomous Systems*, Vol. 44, 2003, pp. 201-208.

[19] Li, L., and Duncan, J. H., "3-D translational motion and structure from binocular image flows," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 15, No. 7, 1993, pp. 657-667.

[20] Lin, G. L. and Cheng, C. C., "Determining 3-D translational motion by the parallel trinocular", in *IEEE International Conference on Robotics and Biomimetics*, 2006, pp. 773-778.
[21] Mitiche A., "On kineopsis and computation of structure and motion," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 8, No. 1, 1986, pp. 109-112.
[22] Mulligan, J. and Daniilidis, K., "Trinocular stereo for non-parallel configurations," in *Proc. Int. Conf. on Pattern Recognition*, 2000, pp. 567-570.
[23] Neumann, T. R., "Modeling insect compound eyes: Space-variant spherical vision," in *Proceedings of the 2nd International Workshop on Biologically Motivated Computer Vision*, *LNCS 2525*, H. H. Bülthoff, S.-W. Lee, T. Poggio, and C. Wallraven Eds., Springer-Verlag, Berlin, 2002, pp. 360-367.
[24] Neumann, T. R. and Bulthoff, H. H., "Biologically motivated visual control of attitude and altitude in translatory flight," in *Proc. 3rd Workshop Artificial Intelligence, Dynamische Perzeption*, G. Baratoff and H. Neumann, Eds., 2000, pp. 135-140.
[25] Ogale, A. S. and Aloimonos, Y., "Shape and the stereo correspondence problem," *International Journal of Computer Vision*, Vol. 65, No. 3, 2005, pp. 147-162.
[26] Ohta, Y., Watanabe, M., and Ikeda, K., "Improving depth map by right-angled trinocular stereo," in *Proc. 8th Int. Conf. on Pattern Recognition*, 1986, pp. 519-521.
[27] Ohya, A., Miyazaki, Y., and Yuta, S., "Autonomous navigation of mobile robot based on teaching and playback using trinocular vision," in *Proc. 27th Annual Conf. of the IEEE Industrial Electronics Society*, 2001, pp. 398-403.
[28] Richards, W., "Structure from stereo and motion," *Journal of Optical Society of America*, Vol. 2, 1985, pp. 343-349.
[29] Rieder, A., "Trinocular divergent stereo vision," in *Proc. Int. Conf. on Pattern Recognition*, 1996, pp. 859-863.
[30] Shin, Y.-G., Park, S.-S., Kim, J.-N., and Jang, D.-S., "Development of a humanoid robot for emotion recognition," *WSEAS Transactions on Systems and Control*, Vol. 1, No. 1, 2006, pp. 54-59.
[31] Srinivasan, M. V., Chahl, J. S., Weber, K., Venkatesh, S., Nagle, M. G., and Zhang, S. W., "Robot navigation inspired by principles of insect vision," *Robotics and Autonomous Systems*, Vol. 26, 1999, pp. 203-216.
[32] Stewart, C. V., and Dyer, C. R., "The trinocular general support algorithm: A three-camera stereo algorithm for overcoming binocular matching errors," in *Proc. 2nd Int. Conf. on Computer Vision*, 1988, pp. 134-138.
[33] Tsukune, H., "Trinocular stereo analysis of optical flow," in *Proc. Int. Conf. on Pattern Recognition*, 1990, pp. 43-47.
[34] Weng, J., Huang, T. S., and Ahuja, N., "Motion and structure from two perspective views: algorithms, error analysis, and error estimation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 11, 1989, pp. 451-476.
[35] Williamson, T. and Thorpe, C., "A trinocular stereo system for highway obstacle detection," in *Proceedings of IEEE International Conference on Robotics & Automation*, 1999, pp. 2267-2273.
[36] Zhang, L., Ji, S., Xie, Y., and Yuan, Q., "Auto detecting system of the crystal oscillators cover defects based on machine vision," *WSEAS Transactions on Systems and Control*, Vol. 2, No. 4, 2007, pp. 326-331.
[37] Zhang, Z., "A flexible new technique for camera calibration," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 22, No. 11, 2000, pp. 1330-1334.

Appendix A

$$(x_{li} - x_{mi})(x_{mj} - x_{rj}) = (x_{lj} - x_{mj})(x_{mi} - x_{ri})$$

Proof:
According to (1), for any two sets for matching image pixels on three different image planes, i.e., (x_{li}, x_{mi}, x_{ri}) and (x_{lj}, x_{mj}, x_{rj}),

$$(x_{li} - x_{mi})(x_{mj} - x_{rj}) = \frac{h_1 h_2 f^2}{Z_i Z_j}$$

$$(x_{lj} - x_{mj})(x_{mi} - x_{ri}) = \frac{h_1 h_2 f^2}{Z_i Z_j}$$

Therefore,

$$(x_{li} - x_{mi})(x_{mj} - x_{rj}) = (x_{lj} - x_{mj})(x_{mi} - x_{ri})$$

It should be noted that the above equation is guaranteed no matter whether h_1 is equal to h_2 or not.

Appendix B

$$\det(\mathbf{A}^T\mathbf{A}) = f^4(a_1^2+b_1^2)[(a_1b_2-a_2b_1)^2+(a_1b_3-a_3b_1)^2]$$

Proof:

$$\mathbf{A}^T\mathbf{A}=\begin{bmatrix} f^2(a_1^2+b_1^2) & 0 & f(a_1a_2+b_1b_2) \\ 0 & f^2(a_1^2+b_1^2) & f(a_1a_3+b_1b_3) \\ f(a_1a_2+b_1b_2) & f(a_1a_3+b_1b_3) & a_2^2+b_2^2+a_3^2+b_3^2 \end{bmatrix}$$

$$\begin{aligned}
&\det(\mathbf{A}^T\mathbf{A}) \\
&= f^4(a_1^2+b_1^2)^2(a_2^2+b_2^2+a_3^2+b_3^2) \\
&\quad - f^4(a_1^2+b_1^2)(a_1a_2+b_1b_2)^2 - f^4(a_1^2+b_1^2)(a_1a_3+b_1b_3)^2 \\
&= f^4(a_1^2+b_1^2)[(a_1^2+b_1^2)(a_2^2+b_2^2+a_3^2+b_3^2) \\
&\quad - (a_1a_2+b_1b_2)^2 - (a_1a_3+b_1b_3)^2] \\
&= f^4(a_1^2+b_1^2)[a_1^2b_2^2+a_1^2b_3^2+a_2^2b_1^2+a_3^2b_1^2-2a_1a_2b_1b_2 \\
&\quad - 2a_1a_3b_1b_3] \\
&= f^4(a_1^2+b_1^2)[(a_1b_2-a_2b_1)^2+(a_1b_3-a_3b_1)^2]
\end{aligned}$$