Three-Dimensional Motion Tracking by the Parallel Trinocular

CHI-CHENG CHENG[*], GWO-LONG LIN[†] and CHIEN-HUNG CHIANG[†]
[*†]Department of Mechanical and Electro-Mechanical Engineering
[*]Institute of Underwater Technology
National Sun Yat-Sen University
70 Lian-Hai Road, Kaohsiung, TAIWAN 80424
REPUBLIC OF CHINA

Abstract: - Incorporating a third CCD camera into a conventional binocular is verified to be very helpful to solve translational motion. The extra device not only provides more image information, but also plays a significant role regarding the solution issue. In this paper, a novel algorithm to recover parameters for translational motion using such a parallel trinocular system is presented. This approach overcomes the difficulty of matrix singularity happening in binocular. In order to fit into application requirements, a compact closed-form solution is also derived. This solution owns some important features, such as no matrix manipulation, no danger of matrix singularity, and easy to apply. To validate this closed-form solution, extensive experiments were conducted. It is concluded that the movement magnitude in the depth direction has great influence on the estimation performance of the translational motion. Experiments with implementation environment and an image tracking situation are also performed to validate the feasibility of compact form solution. During the experiments, the presented parallel trinocular system can demonstrate excellent performance on recovering parameters of translational motion under the circumstance of limited motion along the depth direction.

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

1 Introduction

Recovering 3-D motion parameters from 2-D displacements is a difficult task, given the influence of noise contained in these data, which correspond at best to a crude approximation of the real motion field. Especially, due to the depth-speed ambiguity experienced by monocular observers, translation motion and 3D coordinates in the viewed scene cannot be easily recovered [3]. Besides, monocular motion analysis involves the solution of nonlinear equations. Thus, the 3-D interpretation can only be specified by an arbitrary scale factor since only a single camera is used. Whereas, most algorithms solving motion of equations involve some types of nonlinearity. In general, difficulties exist in handling such nonlinearities. Possible solutions usually require iterations from some initial guess with no guarantee of convergence and the nonlinearity also brings about the possibility of multiple results arises [2][15]. Li and Duncan (1993) used disparity in left and right images to determine the translational motion, which cannot be recovered in binocular systems, by adding an additional constraint [7]. Barron and Eagleson (1996) developed the relationship between time-varying optical flow and physical structure to resolve motion parameters by applying the Kalman filter to integrate motion and structure parameters over time [2]. Nevertheless, those approaches involve abstruse mathematics. Besides, it has been shown that the difficulty of matching problem in binocular is much more complicated than in trinocular [5][12][15].

Owing to rapid advance of integrated electronics, the CCD cameras become more and more popular and inexpensive. This improvement on electronic technology also promotes possible researches on multiple-camera vision framework such as the trinocular system. Trinocular can be divided into six arrangement manners, such as right triangular [1], parallel [10], surrounding [9], divergent [11], orthogonal [13], and the arbitrary sets [6]. The former four types belong to the coplanar category, and others are non-coplanar. The right triangular arrangement is plane *XY* coplanar, while the parallel, the surrounding, and the divergent schemes are plane *XZ* coplanar as illustrated in Fig. 1.

Each configuration has its own geometric constraints based on specific arrangement of multiple CCD cameras. The focus of this paper is on investigating 3-D translational motion parameters via the parallel trinocular stereo vision.

Ohya et al. (2001) indicated that binocular stereo captured targets with triangulation [10]. This

approach appears to be time exhaustive for pattern matching problems, especially in finding correlations. Therefore, taking the advantage of adding one more CCD can significantly shorten matching time. They established a parallel trinocular stereo vision system on a robot and applied the technique of teaching and playback to achieve autonomous navigation with landmarks in indoor environment. Recently, Lin and Cheng (2006) presented a trinocular approach for recovery of the translational motion [8]. 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. More important is that the proposed approach owns significant computation efficiency by reducing mathematical complexity and avoiding difficulty from nonlinearities which is very available in low end programming environment for actual operation. The measurement validations and related experiments of implementation are conducted in this paper.

2 Parallel Trinocular and Translational Motion

Given *O* is the origin of a spatial reference system, assume three identical CCD cameras are arranged parallel in the *X* axis. The distances between adjacent cameras are h_1 and h_2 respectively as illustrated in Fig. 2.

For any point (*X*, *Y*, *Z*) in 3-D space, its projections on the image planes of the left, middle, and right CCD cameras lead to the following expressions:

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

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

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 the 3-D point and is usually unknown.

Basically, when the three cameras are arranged as be a parallel trinocular, those three images will lie in the same *XZ* plane. Thus, the *y* distance for the projections on the image planes will be all the same. A special case that three images have the same *y* direction and same disparity exists when $h_1 = h_2$. However, it is interesting to note that regardless of the difference between h_1 and h_2, the products of the image disparities in *X* direction 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* indicate two image pixels.

Suppose the $\mathbf{V} = (V_x, V_y, V_z)$ is the relatively translational motion of the camera setup with respect to the moving object. The relationships between image flow velocities, (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*).

If both left and middle images are selected, a pair of optical flow equations can be established. Sum up velocity components in (4) in *X* and *Y* directions over all pairs of corresponding points. Assuming that each point has a match, and due to the depth for those two cameras are the same, the concluding results can be derived as follows:

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

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

where Ω_l and Ω_m respectively represent corresponding image regions.

When only two sets, left-middle and middle-right, are selected, the final equations can be expressed by matrix form. Therefore, the translational motion can be quickly recovered with the standard least squared approach as : $\mathbf{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{d}$, where $\mathbf{x}$ is the recovering translational motion, $\mathbf{A}$ represents the image matrix, and $\mathbf{d}$ is optical flow vectors of the trinocular, i.e.,

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

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

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

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

For the purpose of simplicity and clarity, define

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

where

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

The determinant of $\mathbf{A}^T\mathbf{A}$ recovering translational motion in parallel trinocular is expressed by

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

Due to the fact of three CCD cameras arranged as a parallel trinocular, 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 which derived from (3). The determinant of $\mathbf{A}^T\mathbf{A}$ therefore appears to be always positive. As a result, it is concluded that the least squared solution of the parallel trinocular will not encounter the singularity difficulty.

3 Compact Closed-Form Solution

This section is devoted to find the compact closed-form solution, not the estimation by the least squared method, for the translational motion. Define

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

where

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

Then the three components of the translation are:

$$V_x = \frac{b_2d_1 - a_2d_3}{a_1b_2 - a_2b_1}$$
$$V_y = \frac{a_1d_2 + b_1d_4}{a_1^2 + b_1^2} + \frac{(a_1a_3 + b_1b_3)(b_1d_1 - a_1d_3)}{(a_1^2 + b_1^2)(a_1b_2 - a_2b_1)} \tag{8}$$
$$V_z = \frac{f(a_1d_3 - b_1d_1)}{a_1b_2 - a_2b_1}$$

The least square estimation method usually suffers from computational inefficiency of the inverse matrix, especially in actual implementation. The proposed compact closed-form solution gets rid of complex matrix manipulation and makes this technique to be realizable and efficient for applications.

4 Tracking Experiments

To demonstrate feasibility of conducting image following 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 schematic of parallel trinocular tracking configuration shown in Fig. 3. Both stations A and B have identical structure, an image poster and a video camera set which equipped a fixture which holds three 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 camera. The acquired images are with resolution of 320x240 pixels and color levels. 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. 4. It is noted that in the following task, the distance between the poster and lens is unknown.

In general, the camera calibration method often adopts the algorithm presented by Zhang (2000) to estimate the focal length of the X, Y axes and principal point of each camera [16]. The three CCDs of the parallel trinocular used in translational motion could be estimated via standard calibration toolbox proposed by Bouguet (2005) [4]. The estimated

intrinsic parameter of each CCD is shown in Table 1.

The focal length (FL) of all three CCDs set the mean value of six items in Table 1 as 533.32 pixels used in our algorithm. The principal point (PP) after calibration cannot be directly used in *X*, *Y* direction of image to adjust the centre of image. Due to the parallel trinocular is a special configuration, how to adjust to meet the requirement of (1) and (2), it is a practical consideration.

In general condition, the image height of y axis and distance of *x* axis can be corrected if we put these three CCDs on the horizontal plane and adjust the *x*, *y* axis of image based on calibrated principal point. However, upon these three CCDs installed on the camera stand of platform, the alignment in *x*, *y* axes of image will soon be missing, just because of the insufficient precision of platform. Since the calibrated principal point could not be used directly, and the datum of *x*, *y* axes need to be adjusted respectively. To meet the actual situation in the parallel trinocular platform, the strategy of modified principal point in *X*, *Y* axes is suggested as follows:

1) The image of *X* axis is improved based on that the disparities between left-middle and middle-right CCD will be as close as possible.
2) The image of *Y* axis is amended according to that the image centre of the middle CCD will be near to the calibrated data, the height of all three images will be the same, and to meet the actual error of platform.

The results of the image center on the parallel trinocular are shown in Table 2.

The platform system implement 3-D translation with parallel trinocular, the three CCDs will capture different images at motion before and after. We separate 3 types of directions with *XY*, *Z*, and *XYZ* to make a testing of our algorithm. The results are shown in Table 3.

Based on the above results, a number of observations are worth noting:

1) If there is no movement along the *Z* axis, a little estimation errors on the translational motion about 5% are resulted regardless of the magnitude of the displacements along the *X* and the *Y* axes. That is just because the actual images were interfered by noise during actual situation.
2) When there is a translation movement occurs in *Z* axis, the recovered motion parameters for all three axes will be deviated. The errors appear to have the same tendency with the magnitude of the movement along the *Z* axis. In other words, if the motion in *Z* axis is large, the resulting errors for the three axes also become significant which the error was reaching to 35%. Thus, such results indicate that recovery performance of the translation motion is more sensitive to the perturbation in *Z* axis than the perturbations for the *X* and *Y* axes.
3) Based on the above two types of situations, a suitable adjustment for the translational motion to suppress *Z* axis movement was conducted. When the motion of 3D movement was used (15, 15, 2) and (20, 20, 2), the corresponding results were shown in Table 3. The resulting errors appear to be greatly reduced to 5%. After adjusting the magnitude of *Z* axis movement, the estimation errors for recovering the translational motion can be restricted to a satisfactory level.

Through the above theoretical derivation, image calibration, and test on translational motion with actual image implementation, we have acquired the concrete validation of this algorithm. In application, we would further conduct the tracking experiment of the parallel trinocular. When the target image on the station A moved according to given path commands, the movement information was extracted based on the optical calculated from consecutive images. Then the station B followed the extracted movement and relocated the camera's position. The procedures of this experiment are as follows: (1) The three CCDs captured images of target as input, and a Gaussian filter was applied. (2) After the target moved, the three CCDs do item (1) again. (3) Calculate the features of the target in these three CCDs, and the displacement in motion before and after. (4) Upon getting the feature and optical flow of the target, acquire the translational information via our algorithm to track the target. (5) Finally, a comparison between the actual and estimated movement is conducted.

To demonstrate the tracking situation, a piecewise 3-D movement: (4, 4, 1) with 5 times continuously, where unit was mm, was applied. Both the specified path and the estimated path are demonstrated in Fig. 5. Next, actual tracking experiments using the proposed parallel trinocular scheme were also conducted. The task of image following was implemented by using a light spot shooting from station B onto the poster at station A. Excellent tracking results were .

5 Conclusions

A novel algorithm, which can be formulated as a closed-form solution, for recovering parameters of translational motion by a parallel trinocular is presented. It exploits the available redundancy in multiple images to solve the motion parameters. The proposed approach owns several important features such as no matrix manipulation, no danger of matrix singularity, and easy to apply. By extensive experiments on solution validation in actual image implementation, and tracking experiments, the presented approach demonstrates promising recovery performance of translational motion. Our preliminary progress presented here indicates the feasibility of parallel trinocular in applications. However, in practical implementation, the accuracy of this system for servo tracking of translational motion was inevitably influenced by the environment. Other adaptive approaches to improve the aspect of this drawback will be investigated in the future.

6 Acknowledgment

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

References:

[1] 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, pp. 3-9 (2001).

[2] Barron, J. L., and Eagleson, R., "Recursive Estimation of Time-varying Motion and Structure Parameters," Pattern Recognition, vol. 29, no. 5, pp. 797-818 (1996).

[3] 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, pp. 239-269 (1990).

[4] Bouguet, J. Y., "Complete Camera Calibration Toolbox for Matlab," http://www.vision.caltech.edu/bouguetj/.

[5] Gurewitz, E., Dinstein, I., and Sarusi, B., "More on the Benefit of a Third Eye," in Proc. ICPR, pp. 966-968 (1986).

[6] Hemayed, E., Ahmed, M. T., and Farag, A., "CardEye: A 3D Trinocular Active Vision System," in Proc. IEEE Int. Conf. on Intelligent Transportation Systems, pp. 398-403 (2000).

[7] 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, pp. 657-667 (1993).

[8] Lin, G. L., and Cheng, C. C., "Determining 3-D Translational Motion by the Parallel Trinocular", in IEEE International Conference on Robotics and Biomimetics, pp. 773-778 (2006).

[9] Mulligan, J., and Daniilidis, K., "Trinocular Stereo for Non-parallel Configurations," in Proc. ICPR, pp. 567-570 (2000).

[10] 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, pp. 398-403 (2001).

[11] Rieder, A., "Trinocular Divergent Stereo Vision," in Proc. ICPR, pp. 859-863 (1996).

[12] 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, pp. 134-138 (1988).

[13] Tsukune, H., "Trinocular Stereo Analysis of Optical Flow," in Proc. ICPR, pp. 43-47 (1990).

[14] 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, pp. 451-476 (1989).

[15] Yachida, M., "3-D Data Acquisition by Multiple Views," in 3rd International symposium on Robotics Research, pp. 11-18 (1986).

[16] Zhang, Z., "A Flexible New Technique for Camera Calibration," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 11, pp. 1330-1334 (2000).

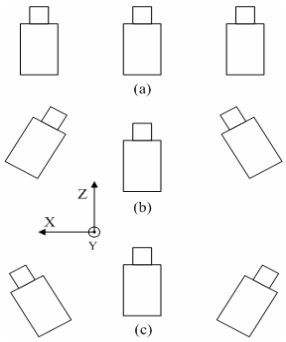

Fig. 1 Arrangements of trinocular: (a) parallel, (b) surrounding, (c) divergent.

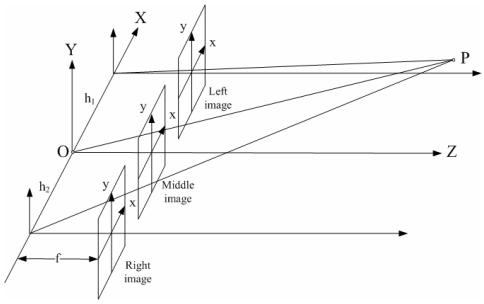

Fig. 2 Schematic parallel trinocular.

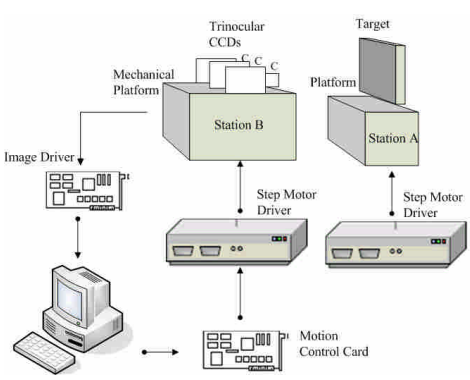

Fig. 3 Schematic of parallel trinocular tracking configuration.

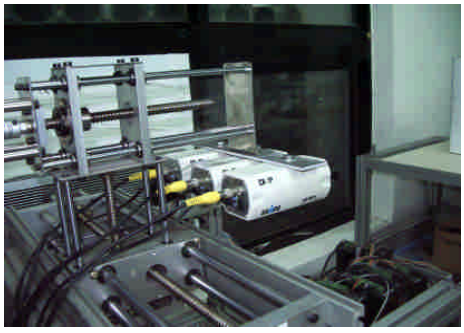

Fig. 4 Arrangement of parallel trinocular.

Table 1
The estimated intrinsic parameter of trinocular

	X (pixel)			*Y* (pixel)		
CCD	Left	Middle	Right	Left	Middle	Right
F L	533.16 ±1.39	538.14 ±1.65	534.55 ±1.62	529.12 ±1.32	534.18 ±1.57	530.76 ±1.54
P P	154.43 ±1.62	170.59 ±1.65	160.48 ±1.59	123.00 ±1.23	111.18 ±1.22	117.76 ±1.20

Table 2
Image centers on the parallel trinocular (unit: pixel)

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

Table 3
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

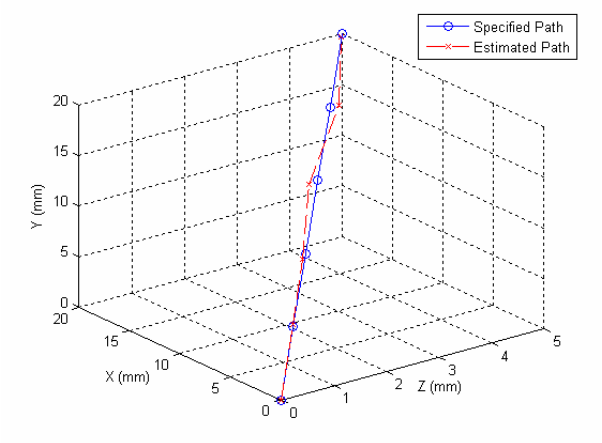

Fig. 5 Tracking path of parallel trinocular in 3-D space.