Journal of Advanced Engineering Vol. 5, No. 3, pp. 205-212 / July 2010

用於即時立體視覺系統之改良匹配方法
An Improved Correspondence Method Used in Real-Time Stereo System

方晶晶* 郭泰宏
Jing-Jing Fang*, Tai-Hong Kuo

摘要

特徵辨識與匹配運算掌握了大部分即時立體視覺系統的效率，如何正確地使用極限幾何求得共軛對成為了匹配運算法優良與否的關鍵，但是假若拍攝影像的深度與範圍差異不是很大，例如手術房應用或平台量測，這類小範圍內的眾多匹配點的拍攝影像將會導致系統效率不彰。本研究透過紅外線攝影方式簡化特徵辨識的複雜度，並針對有限深度區域提出了影像匹配方法的改良，保留了影像扭正的優點，利用影像套合技術提升傳統扭正的效率以進行快速的匹配工作，根據實驗結果，本研究所提出的改良方法可成功地降低匹配候選點的數量，約可將匹配準確度提升 50%，效率亦可增進 100%。
關鍵詞：點匹配，影像扭正，影像套合

Abstract

Both feature detection and point correspondence algorithms always heavily influence the performance of real-time tracking in general binocular systems. In this paper, we develop an infrared tracking system to reduce the deficiencies of feature detection. When this system is implemented in the operating room, light reflection from the metal instruments causes computational redundancy during point correspondence. General epipolar rectification with image registration techniques improves the efficiency of correspondence dramatically. The total amount of candidate correspondences can effectively be reduced by using the image registration method for point correspondence. Our algorithm takes the advantages of both rectification and image registration to achieve real-time tracking. Moreover, the proposed method improves accuracy by approximately 50% and efficiency by 100%, in comparison to general epipolar rectification.
Keywords: point correspondence, epipolar rectification, image registration

I. INTRODUCTION

The aim of general stereopsis is to reconstruct 3D positioning from images obtained from multiple pinhole cameras in different locations. There are three major phases in the process of stereopsis: preprocessing, correspondence, and depth recovery. In the preprocessing stage, the specific features of each image are preliminarily identified. The associated strategy is usually classified into several methods, such as area-based, feature-based, or linear edge segmentation, etc. The correspondence stage, in which correspondence points are pairs of image points that correspond to the same point of an image, is considered to pose a well-known problem in stereopsis. The depth of the scene point is obtained in the final stage. There have been a considerable number of studies in the past two decades. A comprehensive survey by Barnard and Dhond [1, 2] investigated these major procedures in stereo analysis and the evaluation criteria for stereopsis.

In this study, the self-developed optical tracking system consists of a pair of charge-coupled device (CCD) cameras to track spatial coordination formed by light-emitting diode (LED) markers. Complex pattern recognition operations are unnecessary because an infrared filter is used to block the visible rays. Prior to this study, the device had been rectified to 0.2 mm RMS accuracy within a 1.5 meter distance. The mean error and standard deviation of the developed system are 177 and 96 μm, respectively, under a sample rate of 30 fps. The system is applicable to surgical navigation in the operating room; however, light reflection off of metal instruments used in

國立成功大學機械工程學系
*Corresponding author. E-mail: fangjj@mail.ncku.edu.tw
Department of Mechanical Engineering, National Cheng Kung University, Tainan, Taiwan, R.O.C.
Manuscript received 15 May 2009; revised 28 September 2009; accepted 8 October 2009

the operating room generates noises that increase the number of correspondence candidates and also derive system latency.

Therefore, we focus on the promotion of the correspondence method in real-time. For the purpose of increasing the efficiency of corresponding points, geometric and physical constraints [3] are used to reduce the complexity of the potential correspondences for any given pair of image points. Based these constraints, many different approaches have been developed to solve the correspondence problem, for instance, area-based corresponding [4, 5], relaxation labeling [6, 7], dynamic programming [8], hierarchical algorithms [9, 10], structural corresponding [11], etc. However, most of the stereo correspondence algorithms described above make the assumption that conjugate epipolar lines are collinear by epipolar rectification. These algorithms perform well and quickly enough for real-time applications under certain conditions, including good camera calibration, to obtain accurate outcomes.

In this paper, we present a new method for point correspondence by employing an image wrapping technique. The proposed method establishes correspondence from binocular images and is applicable to coplanar and non-coplanar calibration. Experimental results and comparisons of epipolar rectification methods are also presented. Finally, we suggest a correspondence procedure which adopts both rectification and our proposed correspondence methods in larger area tracking for future applications.

II. IMAGE REGISTRATION

Rectification is a classic problem for stereo vision. Some researchers in the field report rectification under several restrictive assumptions. Papadimitriou [12] proposed a rectification method for an un-calibrated camera system. The researcher assumed that the vertical axis of the reference frame is under a parallelized constraint. Therefore, the angle between the optical axes of the two cameras is within $10°$ in order to avoid performance decline. Previously, some researchers [13-15] presented refined algorithms for rectification in a stereo rig with weak calibration. In their methods, only point correspondences between images were needed. Most recent research [16-18] concentrated on the issue of minimizing the rectified image's distortion.

The advantage of rectification is that the computation of stereo correspondences [2] is rather simple, because feature corresponding is performed by aligning the horizontal lines of the rectified images. Given a pair of stereo images R_1 and R_2 (see Fig. 1), the epipolar rectification method determines a transformation of each image plane such that pairs of conjugate epipolar lines become collinear and have parallel image axes. The rectified images can be thought of as a new stereo rig that is obtained by rotating the paired camera planes from R_1 and R_2 to R'_1 and R'_2 separately. The conjugate pair (M_1, M_2) of a spatial point M demonstrates the same height h on both image planes R'_1 and R'_2. Therefore, it is unnecessary to search the 2D image plane and is sufficient to only focus on the vertical lines. Under these circumstances, the working space can be reduced to one dimension.

III. CORRESPONDENCE ALGORITHM

Image registration [19] is the process of overlaying two images of the same situation taken at different time periods, from different viewpoints, or by different sensors. It is one of the most important tasks that integrates and analyzes, for example, familiar colors from various sources. In other words, image registration is usually applied to adjust two associated images by geometric transformation. This process is widely used in remote sensing, medical imaging, computer vision, etc. Image registration is often required when investigating multiple pictures of the same scene. In this paper, we take advantage of image registration to reduce the difficulty caused by the corresponding problem.

The mapping function in image registration is used to promote efficiency by the epipolar rectification method. As shown in Fig. 2(a), a stereo rig is set up to focus on an array of coplanar dots. Fig. 2(b) illustrates the calibration pattern taken by the left and right cameras on a stereo rig. By using the epipolar rectification technique, every corresponding row of rectified images is aligned to the epipolar line as shown in Fig 2(c). Select a singular point P in the left image of Fig. 2(b), for example. Based on the epipolar constraint, each pixel representing the center of every dot of the same ordinate as point P' in Fig 2(c) is located on the same epipolar plane. Therefore in this example, 10 candidates in the right image of Fig. 2(b) are extracted for further correspondence. The total number of candidate matches is reduced from 100 to 10, in this case, by using the rectification method in point correspondence.

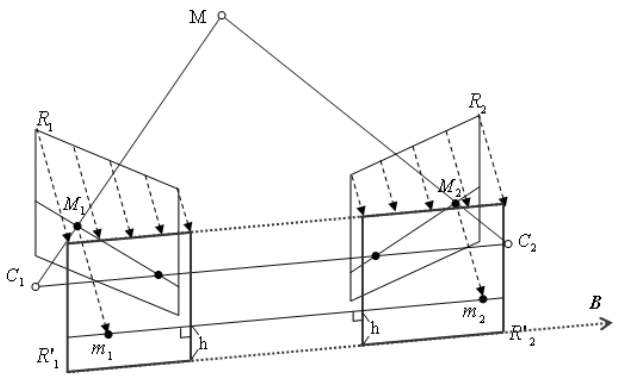

Fig. 1 Epipolar geometry and rectification

As shown in Fig. 2(a), the mapping function between two cameras at depth *h* is known. It is emphasized that the mapping function associated with a specific depth *h*, or very nearly *h*, will benefit the correspondence operation. By using this mapping function, the selected point P in the left image is transformed into P'', as shown in Fig. 3. Given a horizontal length d as a predefined searching domain along the abscissa of point P'', the dots in the area defined by the dash line are potential candidates obtained by the epipolar rectification method. In the previously chosen example, to invoke the registration method for point correspondence, the number of corresponding candidates can further be reduced from 10 to 3.

If the mapping function of depth *h* is known, then the performance of image registration can be improved. Two problems arise when applying the image registration method in stereopsis. The first is what kind of mapping functions should be used for registration. The second is how to acquire benefic mapping functions for the tracking range and how to decide sufficient corresponding points as the control points in the left and right images to generate the relationships of mapping.

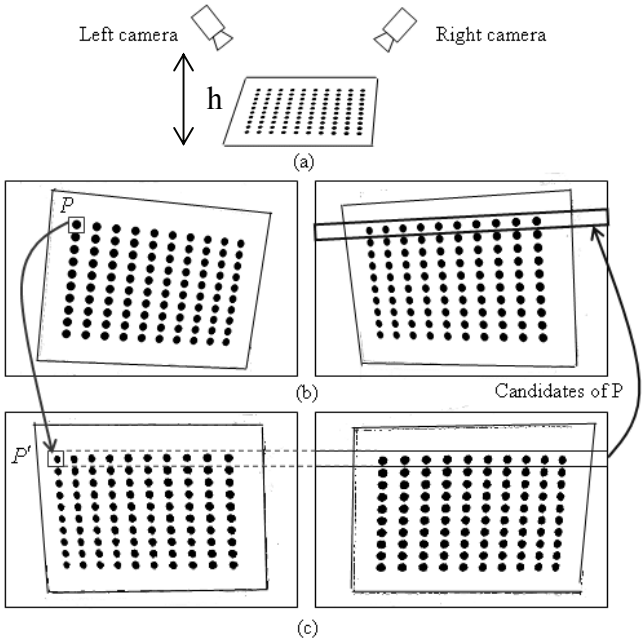

Fig. 2 Epipolar rectification for point correspondence: (a) a stereo rig; (b) a stereo pair of an image; (c) the rectified stereo pair

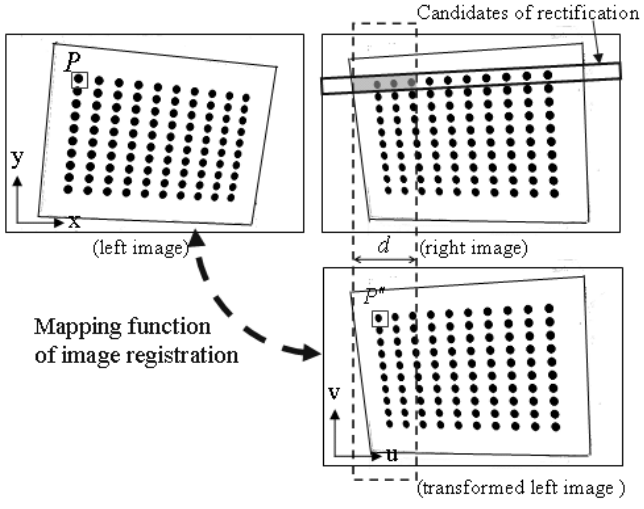

Fig. 3 Image registration for point correspondence

1. Mapping functions

To apply the mapping function, a transform model is used for multi-view registration. Our selection of the transform model should meet several conditions. First, it must be responsive to the spatial position of the cameras. Second, it is assumed that the distance from the cameras to the target is much greater than the entire range of the scene/situation. Third, a flat scene taken by a pinhole camera that produces a distorted image is also considered. The most widely used mapping model is the perspective projection model. In Fig. 3, for example, the mapping from two images to the transformed model is:

$$u = \frac{a_0 + a_1 x + a_2 y}{1 + c_1 x + c_2 y}$$
$$v = \frac{b_0 + b_1 x + b_2 y}{1 + c_1 x + c_2 y} \quad (1)$$

where (u, v) and (x, y) are the coordinates after and before transformation, In (1), a_0. a_1, a_2, b_0, b_1, b_2, c_1, c_2 are projection parameters.

The thin plate splines (TPS) [20] method is one of the radial basis functions used for image registration. It minimizes the potential energy that reflects the amount of function variation; the potential energy should be as small as possible to be a good mapping function. An infinite surface obtained under the imposition of the control point load is known as the surface spline, and its equation is given as:

$$u = f(x,y) = a_0 + a_1 x + a_2 y + \sum_{i=1}^{N} F_i r_i^2 \ln r_i^2$$
$$v = g(x,y) = b_0 + b_1 x + b_2 y + \sum_{i=1}^{N} G_i r_i^2 \ln r_i^2 \quad (2)$$

N is the number of control points, $r_i^2 = (x - x_i)^2 + (y - y_i)^2$, (x_i, y_i) is the position of the i^{th} point load, and $f(x,y)$ is the surface value, or the elevation, of the surface at point (x,y). The obtained surface represents the x-component of the mapping function and its corresponding y-component. The parameters a_0, a_1, a_2, F_i, $i = 1,...,n$ are determined by substituting the coordinates of the corresponding control points and solving the linear equations of the obtained system:

$$\sum_{i=1}^{N} F_i = 0$$
$$\sum_{i=1}^{N} x_i F_i = 0$$
$$\sum_{i=1}^{N} y_i F_i = 0 \qquad (3)$$
$$f(x_1, y_1) = a_0 + a_1 x_1 + a_2 y_1 + \sum_{i=1}^{N} F_i r_{i1}^2 \ln r_{i1}^2$$
$$\vdots$$
$$f(x_n, y_n) = a_0 + a_1 x_n + a_2 y_n + \sum_{i=1}^{N} F_i r_{in}^2 \ln r_{in}^2$$

As shown in Fig. 4, the TPS registration yields good results. However, the computations can be very time consuming and involve a large number of control points. Therefore, how to define a sufficient number of control points while simultaneously maintaining efficient performance is a critical issue in our applications.

2. Control points for registration

The parameters of the mapping functions are obtained by least-square fit such that the polynomials minimize the total of squared errors at the control point. For that reason, there must be a sufficient amount of selected control points in the two images.

Correspondence is unnecessary in a single camera calibration. In this study, we invoke the non-planar camera calibration method [21], which is a single camera model. Due to the coordinate consistency during calibration of the system, the relationship of bi-camera is obtained. Fortunately, during calibration, images from the bi-camera are

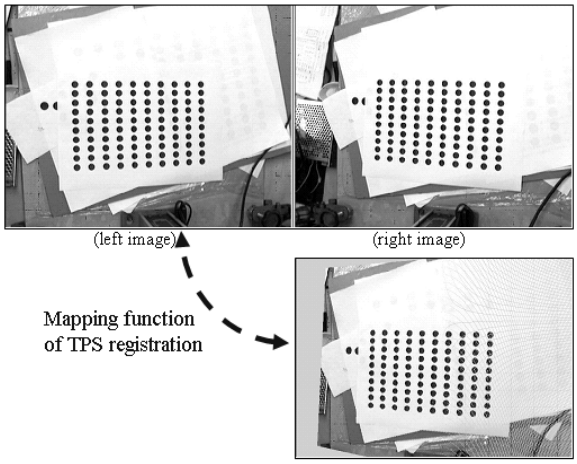

Fig. 4 TPS Image registration for bi-images

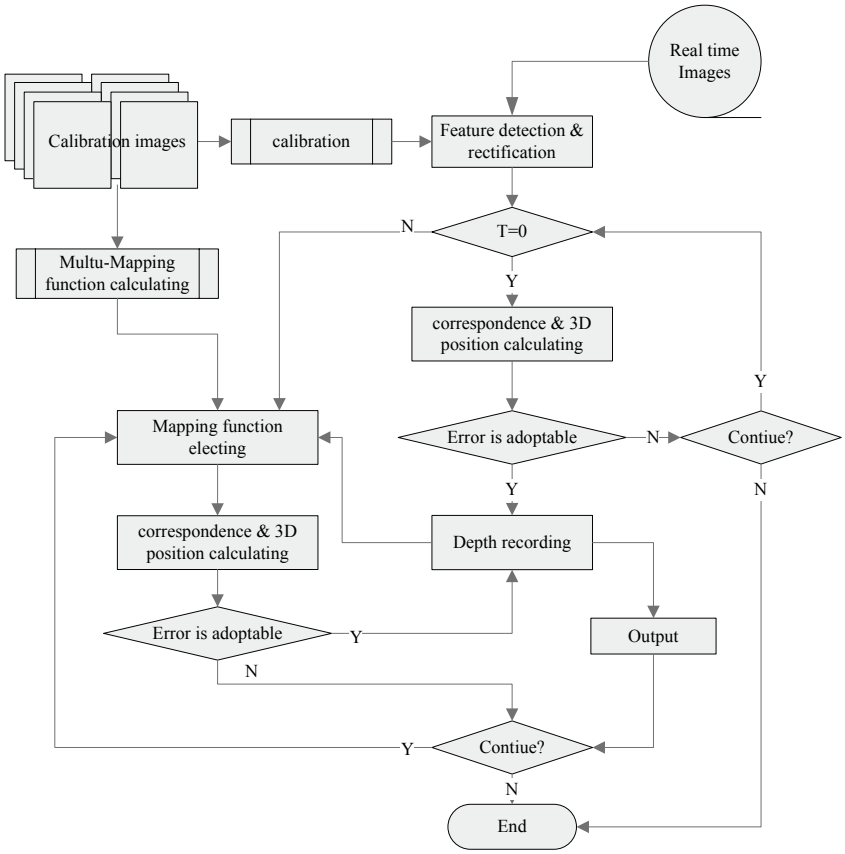

Fig. 5 Flow chart of multi-mapping function corresponding

taken simultaneously. Therefore, the calibration data already embeds correspondence, which is used in our proposed method. The regular calibration data can be used as corresponding points for image registration, manually or semi-automatically. It is expected to be more precise and faster in mapping function evaluation. A pair of co-planar calibration points is only suitable in a small area that is near the calibrated pattern. In order to deal with this limitation, we will use non-coplanar calibration and define the process for choosing feature points from the calibration points that occupy a 3D volume.

Non-coplanar calibration requires several groups of calibration images taken from different depths. According to the groups of correspondence image points taken from the various depths, we can find the mapping functions of these images. As shown in Fig. 5, the system measures the depth value obtained from estimation according to a previous timeline and picks the appropriate mapping functions to conduct the work of point corresponding. It also compensates for mapping function errors caused by depth differences by juxtoposing left-right consistency constraints and setting thresholds for domain searching.

IV. IMPLEMENTATION AND EXPERIMENTAL RESULTS

The research intends to compare the efficiency of image registration for point correspondence on the basis of the nine groups of given coordinates obtained from different depths (see Fig. 6 (a)). The aim of this experiment is to mimic noises to investigate the accuracy and precision of the rapid corresponding method. The corresponding results are known. As shown in Fig. 6 (b), the centroids of the 900 salient points (non-camera calibration points) were photographed by a linear guideway in order to conduct correspondence operations to identify the correspondence candidates in a complex environment.

1. Single mapping function

We use the rectification correspondence method as the comparison group; we record the correspondence success ratio of adding the single registration mapping function and simultaneously obtain the time elapsed. The threshold is set to 0.1 in this operation. As shown in Fig. 3, the search domain is:

$$d = image\ width \times threshold \tag{4}$$

The calibration points, at $Z = 0$, are used as the control points of mapping functions.

Table 1 shows the details of success point correspondences in the single mapping function with the threshold of 0.1. The data is plotted in Fig. 7 to demonstrate the success ratio comparisons between general rectification and our presented method. During the operations on small samples, there are no significant differences in the two methods. By the additional registration operation proposed in this study, of correspondence performance gradually improved as the sample number increased (see Fig. 8). However, as the depth range increased, as well as the sample number, the accuracy ratio experiences only a minor improvement and is apparently unable to be significantly augmented.

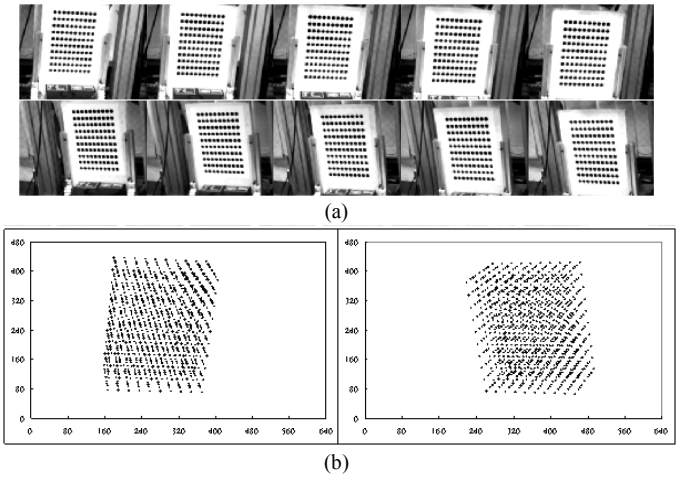

Fig. 6 The input data for the correspondence efficiency experiment: (a) left and right images obtained from different depths; (b) all the centroid of given points are divided into the binocular images

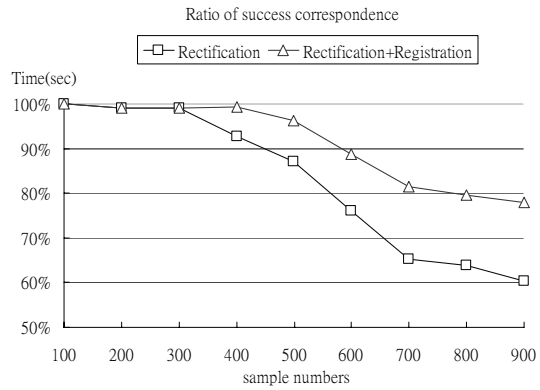

Fig. 7 Accuracy ratio and time contrast for the single mapping function

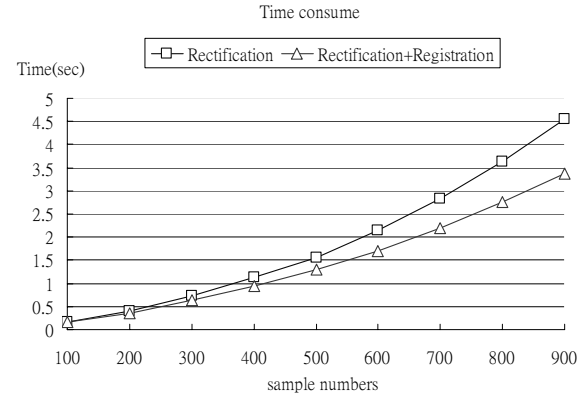

Fig. 8 Time contrast for the single mapping function

Table 1 Correspondence of single mapping function registration (*th*=0.1)

Sample number	Depth range	Rectification			Rectification +Registration (*th*=0.1)		
		No. of success correspondence	Ratio of success correspondence	Time (sec)	No. of success correspondence	Ratio of success correspondence	Time (sec)
100	0 mm	100	100.00%	0.172	100	100.00%	0.172
200	20 mm	198	99.00%	0.406	198	99.00%	0.359
300	40 mm	297	99.00%	0.734	297	99.00%	0.640
400	50 mm	371	92.75%	1.125	397	99.25%	0.937
500	60 mm	435	87.00%	1.563	481	96.20%	1.296
600	70 mm	456	76.00%	2.14	533	88.83%	1.703
700	80 mm	457	65.29%	2.828	571	81.57%	2.203
800	100 mm	511	63.88%	3.625	636	79.50%	2.766
900	120 mm	542	60.22%	4.563	701	77.89%	3.375

Table 2 Accuracy variations via the sample numbers and the threshold settings

Sample number	Depth Range	Threshold										
		0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0 mm	100	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
20 mm	200	99.00%	99.00%	99.00%	99.00%	99.00%	99.00%	99.00%	99.00%	99.00%	99.00%	99.00%
40 mm	300	99.33%	99.00%	99.00%	99.00%	99.00%	99.00%	99.00%	99.00%	99.00%	99.00%	99.00%
50 mm	400	99.50%	99.25%	98.75%	93.25%	92.75%	92.75%	92.75%	92.75%	92.75%	92.75%	92.75%
60 mm	500	98.00%	96.20%	95.20%	87.40%	87.00%	87.00%	87.00%	87.00%	87.00%	87.00%	87.00%
70 mm	600	94.50%	88.83%	87.00%	76.33%	76.00%	76.00%	76.00%	76.00%	76.00%	76.00%	76.00%
80 mm	700	90.14%	81.57%	77.00%	66.43%	65.29%	65.29%	65.29%	65.29%	65.29%	65.29%	65.29%
100 mm	800	77.38%	79.50%	75.25%	64.88%	63.88%	63.88%	63.88%	63.88%	63.88%	63.88%	63.88%
120 mm	900	67.89%	77.89%	71.33%	61.33%	60.22%	60.22%	60.22%	60.22%	60.22%	60.22%	60.22%

In regard to the effect of the threshold on the correspondence efficiency, Table 2 and Fig. 9 indicate that in the case of the measurement of ordinary depths and sample numbers, the best situation is that the threshold should be set to be 0.1 or 0.2 times the width of the image. A rather small threshold is likely to eliminate the accurate correspondence point, whereas an increase in the threshold does not produce any obvious effect on the growth of the accuracy ratio in a bigger depth range. Rather, it will increase the calculation load for the system by increasing the quantity of candidate points. Although this data demonstrates that registration can enhance correspondence efficiency, the single mapping function can limitedly improve the correspondence accuracy ratio in photographing practices conducted in bigger depth ranges.

2. Multi-mapping function

We illustrate that the multi-mapping function has a stronger correspondence capability for bigger depth range; we use three groups of calibration points taken from different depths as registered control points. Determine three groups of mapping functions for different depths in advance, make depth estimations, and select mapping functions according to the flowchart illustrated in Section 5. Based on different thresholds, the correspondence efficiency comparisons for single mapping functions are shown in Table 3.

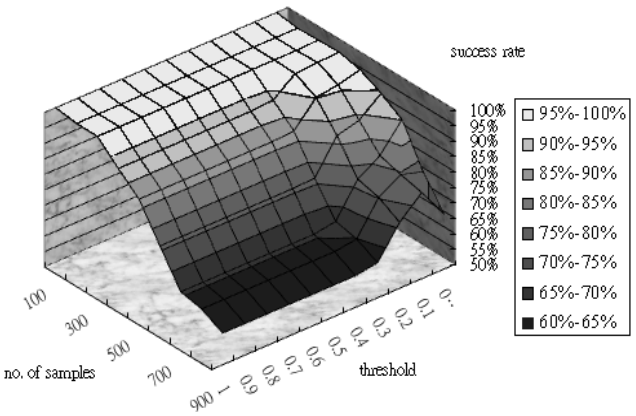

Fig. 9 Accuracy, photographic range and threshold variations for the single mapping function

Table 3 Performance of correspondence for single and multi-mapping functions

Sample number	Threshold	Single mapping function			Multi-mapping function		
		No. of success correspondence	Success ratio of correspondence	Time (sec)	No. of success correspondence	Success ratio of correspondence	Time (sec)
120 mm	0.05	611	67.89%	2.797	697	77.44%	1.750
	0.1	701	77.89%	3.328	885	98.33%	1.875
	0.2	642	71.33%	4.094	885	98.33%	2.094
	0.3	552	61.33%	4.563	817	90.78%	2.187
	0.4	577	64.11%	4.219	817	90.78%	2.141
	0.5	549	61.00%	4.5	817	90.78%	2.187
	0.6	542	60.22%	4.578	817	90.78%	2.219
	0.7	542	60.22%	4.609	817	90.78%	2.188
	0.8	542	60.22%	4.594	817	90.78%	2.219
	0.9	542	60.22%	4.61	817	90.78%	2.235
	1	542	60.22%	4.609	817	90.78%	2.218

This data indicates that multi-mapping functions evidently improve correspondence efficiency in its depth range. They also explain that in the single mapping function, the growth of the threshold value will cause a rise in the quantity of correspondence candidates and then will increase the probability of correspondence error. On the contrary, multi-mapping functions can decrease the quantity of correspondence candidates due to procedural limitations in an attempt to raise the correspondence accuracy ratio. In actual practice, the greater the depth range is or the bigger the angle of a large-angle differentiation camera is, the greater the threshold should be in order to avoid the situation in which accurate correspondence point does not appear in the correspondence search area. From Table 3, it is concluded that the accuracy rate is the best (over 98%) with the threshold value approximately from 0.1 to 0.2. Meanwhile, elapsed time still remains around 50% in comparison to the general method.

V. CONCLUSION

This paper presents a new method for point correspondence use in real-time stereo system. The proposed method establishes correspondence between binocular images based on an image registration, which is often used in movement tracking. The estimation is expected to be faster than collinear algorithms in a limited region of real-time tracking. We also suggest a correspondence procedure which adopts both rectification and multi-mapping functions in order to apply to larger region tracking. In this manner, the proposed method can be applied to real-time movement tracking because of its high accuracy rate and efficiency. In the operating room, correspondence redundancy due to interfering reflection from the mental instruments can be reduced dramatically.

REFERENCES

[1] S. T. Barnard and M. A. Fischler, "Computational Stereo," *Computer Surveys*, vol. 14, pp. 553-572, 1982.

[2] U. R. Dhond and J. K. Aggarwal, "Structure from Stereo-A Review," *IEEE Transactions on Systems Man and Cybernetics*, vol. 19, pp. 1489-1510, 1989.

[3] B. K. P. Horn, "Robot vision," *MIT Electrical Engineering and Computer Science*, 1986.

[4] D. B. Gennery, "Object detection and measurement using stereo vision," in *Proc. of ARPA Image Understanding Workshop*, College Park, pp. 161-167, 1980.

[5] H. P. Moravec, "Toward automatic visual obstacle avoidance," in *Proc. of 5th Int. Joint Conf. Artificial Intelligence*, pp. 584-590, 1977.

[6] S. T. Barnard and W. B. Thompson, "Disparity analysis of images," *IEEE Pattern Analysis and Machine Intelligence*, vol. 2, pp. 333-340, 1988.

[7] Y. Kim and J. Aggarwal, "Positioning three-dimensional objects using stereo images," *IEEE Journal of Robotics and Automation,* vol. 3, pp. 361-373, 1987.

[8] Y. Ohta and T. Kanade, "Stereo by intra- and inter-scanline search," *IEEE Pattern Analysis and Machine Intelligence*, vol. 7, pp. 139-154, 1985.

[9] A. Hoff and N. Ahuja, "Surfaces from stereo: integrating feature corresponding, disparity estimation, and contour detection," *IEEE Pattern Analysis and Machine Intelligence*, vol. 11, pp. 121-136, 1989.

[10] H. S. Lim and T. O. Binford, "Stereo correspondence: a hierarchical approach," in *Proc. of DARPA Image Understanding Workshop*, Los Angeles, CA, pp. 234-241, 1987.

[11] K. L. Boyer and A. C. Kak, "Structural stereopsis for 3-D vision," *IEEE Pattern Analysis and Machine Intelligence*, vol. 10, pp. 144-166, 1988.

[12] D. V. Papadimitriou and T. J. Dennis, "Epipolar line estimation and rectification for stereo image pairs," *IEEE Transactions on Image Processing*, vol. 5, pp. 672-676, 1996.

[13] R. Hartley and R. Gupta, "Computing matched-epipolar projections," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 549-555, 1993.

[14] R. I. Hartley, "Theory and practice of projective rectification," *International Journal of Computer Vision*, vol. 35, pp. 1-16, 1999.

[15] L. Robert, C. Zeller, O. Faugeras, and M. Hébert, "Applications of non-metric vision to some visually-guided robotics tasks," *Visual Navigation: From Biological Systems to Unmanned Ground Vehicles Lawrence Erlbaum Associates*, Chapter 5, pp. 89-134, 1997.

[16] F. Isgro and E. Trucco, "Projective rectification without epipolar geometry," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 23-25, 1999.
[17] C. Loop and Z. Zhang, "Computing rectifying homographies for stereo vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 125-131, 1999.
[18] M. Pollefeys, R. Koch, and L. Van Gool, "A simple and efficient rectification method for general motion," *ICCV99*, pp. 496-501, 1999.
[19] B. Zitov and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing*, vol. 21, pp. 977-1000, 2003.
[20] A. Goshtasby, "Registration of images with geometric distortions," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 26, pp. 60-64, 1988.
[21] R. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," *IEEE Journal of Robotics and Automation*, vol. 3, pp. 323-344, 1987.