Feature Tracking from an Image Sequence Using Geometric Invariants

H. T. Tsui, Z. Y. Zhang, and S. H. Kong
Department of Electronic Engineering
The University of Hong Kong, Shatin, N. T., HK
Tel: (852) 2609 8256, Fax: (852) 2603 5558, Email: httsui@ee.cuhk.edu.hk

Abstract

In this paper two new feature tracking algorithms are proposed. In the first algorithm, a perspective camera model is used. Making use of the projective invariant of Barrett, and assuming the image feature points corresponding to 8 general points in space are tracked by a conventional method in the image sequence, the other feature points in the sequence can be tracked using a Hough technique. Correspondence between two reference images as required by the original Barrett's invariant is not necessary. In the second algorithm, an affine camera model is assumed and the image feature points corresponding to 4 non-coplanar points in space are assumed tracked in the image sequence using a conventional method. These image points form the basis of affine coordinates in each image. After the correspondence of a fifth point is established between the first two images, the affine coordinates of all image points in the first images existence can be computed. As far as we know, this is the only algorithm which can transfer a point knowing only a single image. Experiments showed that both algorithms gave highly accurate tracking results.

Keywords : Feature tracking, affine invariant, perspective invariant, Hough transform

1 Introduction

Recently, a number of papers have discussed the problem of computing the information of an image using the information from two or more other images of the same scene [1, 2, 3]. This problem is strongly related to the tracking problem. Many of the methods proposed assume the knowledge of the epipolar geometry or the cameras are weakly calibrated [1, 8]. However, without camera calibration, some important results [7, 6, 3] have been obtained by a number of researchers using geometric invariance to map image features from two reference images to a third image. Shashua [3] shows the of a trilinear function between three perspective views and that the coefficients of the function can be recovered linearly without establishing the epipolar geometry first. Hartley [11] proposed a trifocal tensors method for transferring lines and points corresponding within two images into a third image. The parameters of the trifocal tensors can be computed if seven point correspondences are established. Barrett [7] has proposed a perspective invariant based on the correspondence of eight points in general positions in three views. If the correspondence of a new feature point, not included in the previous eight points, can be established between the two reference images, this point can be mapped to the third image using the above invariant. This method was termed by Mundy and Zisserman as *point transfer* [7]. Reid and Murray [10] applied the concept to the design of a real-time gaze control system.

In this paper, we propose two tracking algorithms with uncalibrated camera: the first one uses a perspective invariant of Barrett [7] and the second takes advantage of affine invariants. Following the same line of thinking as the algorithm reported in [4], assuming eight image feature points are assumed tracked correctly by a conventional technique. The proposed algorithm will then establish the correspondence of any image feature through out an image sequence by a Hough technique using the Barrett's invariant. In the second algorithm, an affine camera model is assumed. Unlike the method of Shashua and the trifocal tensor method, we can transfer a point from a first image to a third image without having to find its correspondence in a second image in advance. As far as we know, ours is the only method which can do this. In our algorithm, a set of four images features corresponding to four non-coplanar points in space is assumed tracked correctly throughout the image sequence. The correspondence of a fifth image feature between the first

1063-6919/97 $10.00 © 1997 IEEE

two images is also assumed established. The affine coordinates of every image point in the first image can then be computed using the five image features. Thus all image points in the first image, including non-detectable image points, can be transferred to any target image in the sequence. This is true even if the location of the image point at the target image happened to be occluded provided the 4 basis points are detectable in the target image. In [10] Reid and Murray independently proposed a point transfer scheme. Unlike our algorithm, correspondence between a first and a second image must be established before the corresponding point in a third image can be computed by transfer. We provide a formal proof that 5 corresponding image points in two images is sufficient for computing the invariant affine coordinates of any image point.

2 Feature Tracking Using Perspective Invariant

This section depicts our work on the tracking of sparse feature points under the circumstance of perspective camera.

2.1 Point Transfer

Barrett's perspective invariant [7] is described as below. Here, the case of perspective viewing at two positions is considered and the world coordinate frame is supposed to coincide with that of the left camera. Figure 1 gives the illustration for the geometry. Let $\boldsymbol{P}$ be an arbitrary point in $\mathcal{P}^3$, i.e., 3D projective space, $\boldsymbol{P}_l$ and $\boldsymbol{P}_r$ its images in the left and right image planes, respectively. In Euclidean space coordinate frame, $\boldsymbol{P}_l$ and $\boldsymbol{P}_r$ possess the following relationship:

$$\boldsymbol{P}_r = \boldsymbol{R}\boldsymbol{P}_l + \boldsymbol{T} \tag{1}$$

where $\boldsymbol{R}$ and $\boldsymbol{T}$ are the rigid rotational matrix and translational vector, respectively.

As shown in Figure 1, the following equation obviously holds:

$$\boldsymbol{P}_r \cdot (\boldsymbol{T} \times \boldsymbol{P}_r) = 0 \tag{2}$$

Equation (2) can be rewritten as,

$$\boldsymbol{P}_r^t \boldsymbol{M} \boldsymbol{P}_l = 0 \tag{3}$$

where $\boldsymbol{M} = \boldsymbol{\tau}\boldsymbol{R}$ is called the *essential matrix* for the two view geometry, $\boldsymbol{\tau}$ is theequivalent skew symmetric matrix of $\boldsymbol{T}$. Equation (3) reveals the epipolar geometric relation between the left and the right image planes in Euclidean space.

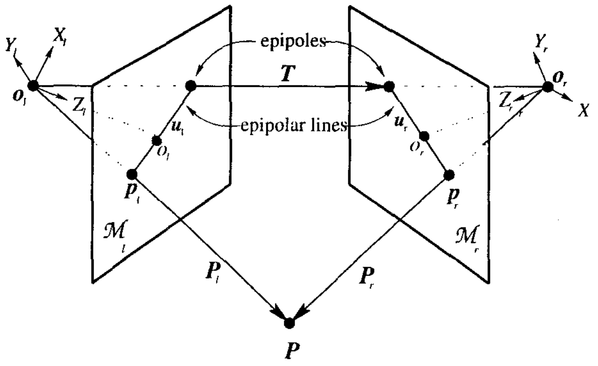


Figure 1: The geometry of two perspective views.

By substituting Equation (3) with the projective relationship between the image and world coordinate frames, $\boldsymbol{p} = (f/Z)\boldsymbol{P}$, where f is the focal length of the camera, we have

$$\boldsymbol{p}_r^t \boldsymbol{M} \boldsymbol{p}_l = 0 \tag{4}$$

where $\boldsymbol{p}_r = (x_r, y_r, 1)^t$ and $\boldsymbol{p}_l = (x_l, y_l, 1)^t$ are the right and left homogeneous image coordinates, respectively.

Assume that 8 point correspondences over 3 images π_1, π_2, π_3 are established. From a general form of Equation (4),

$$\begin{aligned} \boldsymbol{p}_r^t \boldsymbol{M} \boldsymbol{p}_l &= \begin{pmatrix} x_l \\ y_l \\ 1 \end{pmatrix}^t \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{pmatrix} \begin{pmatrix} x_r \\ y_r \\ 1 \end{pmatrix} \\ &= 0 \end{aligned}$$

an expanded form of above equation can be obtained:

$$\boldsymbol{b}\boldsymbol{m} = 0 \tag{5}$$

where $\boldsymbol{b} = (x_l x_r, x_l y_r, x_l, y_l x_r, y_l y_r, y_l, x_r, y_r, 1), \boldsymbol{m} = (m_{11}, m_{12}, m_{13}, m_{21}, m_{22}, m_{23}, m_{31}, m_{32}, m_{33})^t$.

Let $\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_8$ denote the 8 point correspondences between π_1 and π_3, $\boldsymbol{b}_9 = \boldsymbol{b}(\boldsymbol{p}_1, \boldsymbol{x})$ the correspondence between an arbitrary point $\boldsymbol{p}_1$ in π_1 and $\boldsymbol{x}$, its counterpart in π_3, and $\boldsymbol{B} = (\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_9)^t$, then the following equation holds:

$$\boldsymbol{B}\boldsymbol{m} = 0 \tag{6}$$

In order for Equation (6) to have a non-trivial solution for $\boldsymbol{m}$, the determinant of $\boldsymbol{B}$ must be identically zero, i.e., $|\boldsymbol{B}| = 0$. Since this condition holds for any positions of the camera and any selections of the set of the points, it is an invariant for the two-views imaging process. A linear equation defined in the third image plane is obtained by expanding the invariant condition $|\boldsymbol{B}| = 0$,

$$ax_x + by_x + c = 0 \tag{7}$$

where $(x_x, y_x, 1)^t$ is the homogeneous coordinates of $\boldsymbol{x}$ in π_3, a, b, and c are determined by $\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_8$ and $\boldsymbol{p}_1$. From the correspondences between π_2 and π_3, a similar equation can be derived,

$$a'x_x + b'y_x + c' = 0 \tag{8}$$

The intersection of the two lines is the position of $\boldsymbol{x}$.

2.2 Sparse Point Identification

Given an image sequence $\{f_i \mid i = 1, 2, \cdots, M\}$, the correspondences of 8 control points are assumed to be established over the image sequence. Let $\{\boldsymbol{p}_{ij} \mid j = 1, 2, \cdots, N_i\}$ denote a set of points other than the 8 control points visible in the ith image frame. For a frame f_i $(i \geq 2)$, each point $\boldsymbol{p}_{ij}$ is mapped into a line in the the first image. For a certain point $\boldsymbol{p}_j$, repeating the mapping operation throughout the image sequence except for the first image would result in up to M lines intersected where $\boldsymbol{p}_{1j}$ locates in the first image. Figure 2 depicts the mechanism of multiple mapping, which is essentially coincident with Hough transformation.

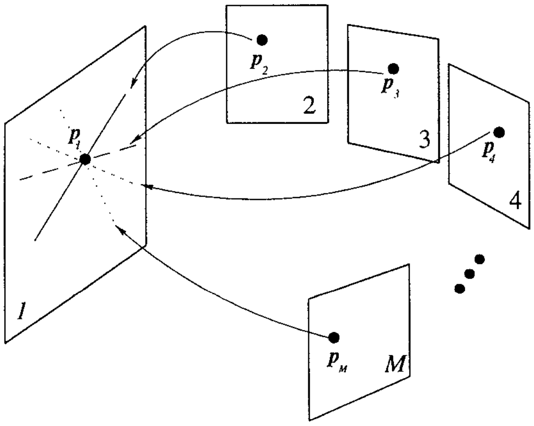


Figure 2: Hough transformation performed over the image sequence.

To accomplish the trajectory of the point over the image sequence, all points of $\{\boldsymbol{p}_{ij}\}$ for each frame f_i $(i \geq 2)$, are mapped onto the the first frame to draw out their corresponding lines on it, with each line being assigned with a unique label for its identification. A 2D accumulating array associated with the first image is created to count the times the lines intersect in an accumulator cell of finite size. By locating the local maxima of the counts over the accumulator array, the positions of the feature points in the first image are estimated. Trajectory over the image sequence for every feature point is therefore found out by tracing back through the labeling clues recorded during the processes of point-to-line mapping.

3 Feature Tracking Using Affine Invariants

In this section, we discuss a novel approach to feature point tracking which is based on affine coordinate transfer.

3.1 Linear Representation for Affine Points

Given four non-coplanar points in $\mathcal{P}^3$, $\boldsymbol{P}_i$ $(i = 0, 1, 2, 3)$, the vectors ε_i defined below are obviously linearly independent:

$$\varepsilon_i = \boldsymbol{P}_i - \boldsymbol{P}_0, \quad i = 1, 2, 3 \tag{9}$$

Thus ε_i $(i = 1, 2, 3)$ compose a basis in $\mathcal{P}^3$. Therefore, any other point $\boldsymbol{P}_i \in \mathcal{P}^3$, $i = 4, 5, \cdots$, can be represented in terms of ε_i $(i = 1, 2, 3)$ by the following linear formula:

$$\boldsymbol{P}_i = \boldsymbol{P}_0 + \alpha_i\varepsilon_1 + \beta_i\varepsilon_2 + \gamma_i\varepsilon_3 \tag{10}$$

where α_i, β_i, and γ_i are termed as the *affine coordinates* of point $\boldsymbol{P}_i$.

The 3D affine transformation is given by

$$\boldsymbol{P}'_i = \boldsymbol{A}\boldsymbol{P}_i + \boldsymbol{T} \tag{11}$$

where $\boldsymbol{A}$ is a 3×3 transformation matrix and $\boldsymbol{T}$ a 3-vector of translation.

Substituting Equation (10) into Equation (11) results in a transformed version of Equation (10):

$$\boldsymbol{P}'_i = \boldsymbol{P}'_0 + \alpha_i\varepsilon'_1 + \beta_i\varepsilon'_2 + \gamma_i\varepsilon'_3 \tag{12}$$

From Equation (10) and (12) we know that the affine coordinates α_i, β_i, and γ_i are geometrically invariant.

3.2 Affine Coordinates Computation

By substituting the Equation (10) into the following 3D to 2D projection equation for an affine camera model:

$$\boldsymbol{p} = \boldsymbol{M}\boldsymbol{P} + \boldsymbol{t} \tag{13}$$

where $\boldsymbol{M}$ is a general 2×3 matrix and $\boldsymbol{t}$ a general 2-tuple vector, we have the counterpart of Equation (10) on the image plane:

$$\boldsymbol{p}_i - \boldsymbol{p}_0 = \alpha_i \boldsymbol{e}_1 + \beta_i \boldsymbol{e}_2 + \gamma_i \boldsymbol{e}_3 \tag{14}$$

where $\boldsymbol{e}_j = \boldsymbol{M}\boldsymbol{\varepsilon}_j,\ j = 1, 2, 3$. Similar equation holds for any other view:

$$\boldsymbol{p}'_i - \boldsymbol{p}'_0 = \alpha_i \boldsymbol{e}'_1 + \beta_i \boldsymbol{e}'_2 + \gamma_i \boldsymbol{e}'_3 \tag{15}$$

So if the affine coordinates α_i, β_i, and γ_i for an individual point are known, its trajectory over the entire image sequence is determined, provided the affine basis is also known. The problem we are facing now is how to compute the three invariant parameters from multiple views.

Assertion 1. *Five non-coplanar points visible in two views with the points' correspondences between the two views being known are sufficient to determine the affine coordinates of any other point visible at least in one of the two images.*

Proof. Given a point $\boldsymbol{P} \in \mathcal{P}^3$. $\boldsymbol{P}$ is projected onto the first image plane with a manner depicted by Equation (13). Moving the camera to a new position, the image of $\boldsymbol{P}$ on the second image plane is given by

$$\boldsymbol{p}' = \boldsymbol{M}'\boldsymbol{P} + \boldsymbol{t}' \tag{16}$$

By eliminating $\boldsymbol{P}$ in Equation (13) and (16) an implicit form of the relationship between $\boldsymbol{p}$ and $\boldsymbol{p}'$ is obtained:

$$ax' + by' + cx + dy + e = 0 \tag{17}$$

where a, b, c, d, and e are unknowns, which depend only on the camera geometry and motion parameters in the 3D space. Equation (17) is known as *affine epipolar constraint equation* [8].

By setting $e = 1$ without loss of generality, a nontrivial solution of Equation (17) is obtained from four point correspondences between the two views:

$$ax'_i + by'_i + cx_i + dy_i = -1, \quad i = 1, 2, 3, 4 \tag{18}$$

Since a, b, c, d are related only to the poses of the affine camera at the two viewing positions in the world frame, they are the same for any pair of corresponding points in the two image planes. Note that the condition that the four control points are not coplanar is necessary. Without it, the determinant of the coefficient matrix will be zero because of the linear dependence of the column vectors of the coefficient matrix.

To calculate the affine coordinates α_i, β_i, and γ_i in Equation (14), the equation in combination with Equation (15) and (17) are used to form following simultaneous equations:

$$\left.\begin{array}{l} ax'_i + by'_i + cx_i + dy_i = -1 \\ x_i = x_0 + \alpha_i e_{1x} + \beta_i e_{2x} + \gamma_i e_{3x} \\ y_i = y_0 + \alpha_i e_{1y} + \beta_i e_{2y} + \gamma_i e_{3y} \\ x'_i = x'_0 + \alpha_i e'_{1x} + \beta_i e'_{2x} + \gamma_i e'_{3x} \\ y'_i = y'_0 + \alpha_i e'_{1y} + \beta_i e'_{2y} + \gamma_i e'_{3y} \end{array}\right\} \tag{19}$$

In Equation (19), $\alpha_i, \beta_i, \gamma_i, x'_i$, and y'_i are unknowns. From Equation (19) we can solve for the unknowns.

The control points respectively used in the computations of both the affine epipolar geometric parameters and the affine invariant coordinates should be different each other at least for one point. Otherwise, there will be no solution to Equation (19) because of the linear dependence of the component equations. Therefore, a minimum of *five* non-coplanar control points are *necessary* and *sufficient* to compute out the invariant affine coordinates. □

3.3 Point Transfer

Given an image sequence, $\{f_i \mid i = 1, 2, \cdots, M\}$. Let $\boldsymbol{p}_{1j}$ be an arbitrary point other than the five control points in the first image of the image sequence. If the affine coordinates of $\boldsymbol{p}_{1j}$, i.e., $\alpha_j, \beta_j, \gamma_j$, has been computed out with the method we proposed in Section 3.2, then the counterpart of $\boldsymbol{p}_{1j}$ in any other image $f_i,\ (i = 2, 3, \cdots, M)$ is obtained by

$$\boldsymbol{p}_{ij} = \boldsymbol{p}_0^{(i)} + \alpha_j \boldsymbol{e}_1^{(i)} + \beta_j \boldsymbol{e}_2^{(i)} + \gamma_j \boldsymbol{e}_3^{(i)} \tag{20}$$

where i denotes the image number in the sequence, j the point number, $(\boldsymbol{e}_1^{(i)}, \boldsymbol{e}_2^{(i)}, \boldsymbol{e}_3^{(i)})$ is the affine basis in ith image and $\boldsymbol{p}_0^{(i)}$ is the origin of the basis. Please note that the implicit assumption that the set of 4 points forming the affine basis are correctly tracked throughout the image sequence beforehand. Obviously, Equation (20) accomplishes a method for point feature tracking. Since the image numbering is arbitrary, the method permits the selection of the original point in any image. Therefore, points invisible in some images can still be tracked over the entire sequence.

4 Experiments

In this section we present the experimental results showing the performances of our methods for feature point tracking.

4.1 Tracking by Point Transfer

An image sequence of a model is acquired by a CCD camera mounted on an active vision platform with multiple translating and rotating axes [5]. Figure 3 is the first image in the sequence. Twenty-two black dots are marked on the face of the model to serve as feature points in the experiments. The marks are labeled with numbers for their identification.

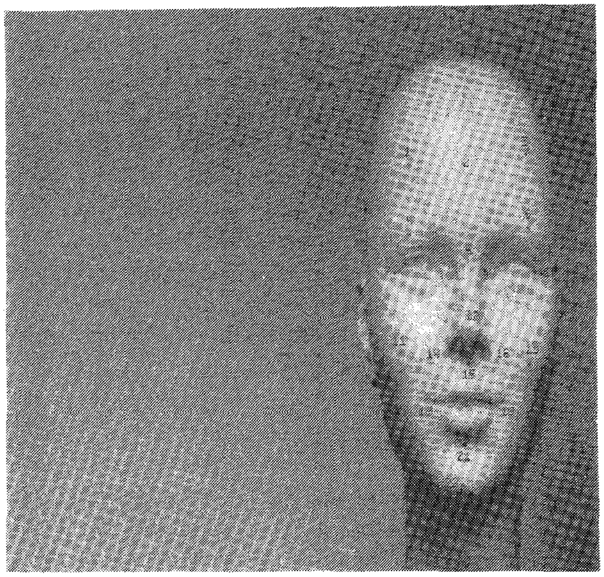


Figure 3: The first image of the image sequence. Numbers serve as the labels for the feature points.

To verify the accuracy of the algorithm on tracking by transfer using affine invariants proposed in Section 3.3, points 2, 11, 13, 21 are selected for the epipolar parameters computation, while points 2, 11, 13, 12 are for the linear basis construction. Table 1 gives the statistics for transfer accuracy in three typical frames. About 90% of the transfers has subpixel accuracy and the average transfer error is also less than one pixel. Such accuracy is sufficient for many visual tasks, such as shape from motion.

As a comparison, an experiment of feature point transfer using the method of Barrett described in Section 2.1 is described below. It should be noted that this is not a tracking algorithm and the experiment on the practical tracking algorithm derived from the Barrett invariant is described in Section 4.2. Assuming a perspective camera model, points 1, 2, 3, 4, 5, 6, 18, 19 are taken as the set of eight control points with the rest as test points on the head model shown in Figure 3. The mean absolute error is about 0.58 pixels with 80% of all transfer are having subpixel errors. This accuracy is close to that of the affine case above.

Table 1: Statistics for Affine Transfer

Statistics	*Image 2*	*Image 3*	*Image 4*
Arith Mean (pxls)	0.15	-0.17	0.14
Absol Mean (pxls)	0.53	0.48	0.50
MSV (pxls)	0.62	0.58	0.61
Absol Max (pxls)	1.57	1.54	1.45
Absol Min (pxls)	0.07	0.02	0.03
% of Subpxl	87.5%	90.6%	87.5%

To further demonstrate the performance of the affine transfer method, edges obtained by zero-crossing technique in the first image are tracked over the image sequence using the proposed method. Figure 4 illustrates the tracking results. From the pictures in the figure we can see that the overall accuracy is very good. However, the five control points *must* be visible in every images in the method.

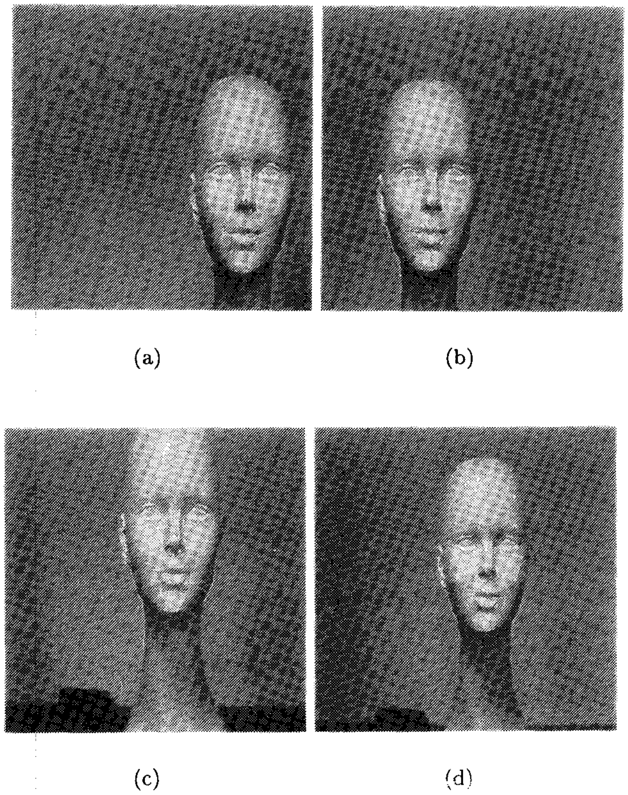
4.2 Tracking by Point Identification

To demonstrate the method for identifying the points sparsely distributed in sequential images, the same image sequence is applied here. Points 1, 2, 3, 4, 6, 12, 18, and 19 are picked out to serve as the control points. Except for Point 17 and 20, the remaining 12 points are correctly tracked over the image sequence by the method described in Section 2.2. In other words, these points are correctly identified in the sequence.

5 Conclusions

In this paper, the geometric invariants both in perspective and affine projective cases are considered in the application of point feature tracking from an image sequence. As for the perspective case, Barrett's invariant method is extended in terms of Hough transformation technique to the application case where no correspondences over the image sequence are known for a set of detectable points except for the eight control points. The trajectories of the spares points are reliably determined by the extended method.

A novel method of image point transfer using affine invariants is proposed. This is a very efficient and reliable algorithm for tracking not only detectable feature points, but also every image point of an image (denoted as the reference image) in an image sequence.



(a) (b)

(c) (d)

Figure 4: Feature point transfer over the image sequence. (a), (b), (c), (d) are the first four images of the sequence. The edge points in the first image are correctly transfered into the other images in the sequence.

The transfer of an image point to a target image can still be made even if the corresponding real image point is occluded in the target image. This is extremely useful for shape from motion or model building of 3D objects. For this algorithm to work, 4 points in general positions in space must first be tracked correctly throughout the image sequence a priori by a conventional technique. Further, the correspondence of an additional fifth feature point must be established between two images. These two conditions are very modest as these 5 points have to be tracked anyway if another method is used for tracking. Significantly, 90% of the locational accuracy of the image point transfers is within one pixel with the maximum error less than 1.6 pixel. This is a practical method which can do accurate and reliable dense image points tracking by transfer.

Acknowledgment

The authors are grateful to the support from the RGC Earmarked Research Grant, UPGC Ref No. CUHK 507/95E.

References

[1] O. Faugeras and L. Robert, "What Can Two Images Tell Us about a Third One?" *Int. J. of Computer Vision,* 18, 1996.

[2] S. Ullman and R. Basri, "Recognition by Linear Combination of Models," *IEEE Trans. on PAMI,* 13(10), 1991.

[3] A. Shashua, "Algebraic Functions for Recognition," *IEEE Trans on PAMI,* vol. 17, August 1995.

[4] H. T. Tsui, S. H. Kong and C. W. Chan, "Feature Tracking from an Image Sequence Using Affine Invariance and Hough Transform," *Intelligent Robots and Computer Vision XV, SPIE's Photonics East'96,* Nov., 1996.

[5] H. T. Tsui and Z. Y. Chen, "A New Tracking Method for Shape from Motion Using an Active System," *Proc. of Asian Computer Vision Conference (ACCV'95),* Singapore, December, 1995.

[6] L. Quan, "Invariants of 6 Points from 3 Uncalibrated Images," *IEEE Trans. PAMI,* Vol. 17, No. 1, January 1995.

[7] J. L. Mundy and A. P. Zisserman, *Geometric Invariance in Computer Vision,* MIT Press, Cambridge, MA, 1992.

[8] L. S. Shapiro, A. Zisserman and M. Brady, "Motion from Point Matches Using Affine Epipolar Geometry," *Lecture Notes in Computer Science,* Vol. 801, J.-O. Eklundh (Ed.), Computer Vision–ECCV'94, Springer-Verlag, Berlin Heidelberg, 1994.

[9] P. A. Beardsley, A. Zisserman and D. W. Murray, "Navigation Using Affine Structure from Motion," *Lecture Notes in Computer Science,* Vol. 801, J.-O. Eklundh (Ed.), Computer Vision–ECCV'94, Springer-Verlag, Berlin Heidelberg, 1994.

[10] I. D. Reid and D. W. Murray, "Active Tracking of Foveated Feature Clusters Using Affine Structure," *Int. J. Computer Vision,* 18, 1996.

[11] R. Hartley, "A linear method for reconstruction from lines and points," *Proc. of The Fifth Int'l. Conf. on Computer Vision (ICCV'95),* MIT, Cambridge, Massachusetts, USA, June, 1995.