SICE Annual Conference in Fukui, August 4-6, 2003
Fukui University, Japan

Comparison of the Feature Point Tracking Method in Image Sequences

Y. Naito[1], T. Okatani[2] and K. Deguchi[3]

Tohoku University, Miyagi sendai, Japan
1 naito@fractal.is.tohoku.ac.jp
2 okatani@fractal.is.tohoku.ac.jp
3 kodeg@fractal.is.tohoku.ac.jp

Abstract: In order to reconstruct 3D structure from image sequences we have to do tracking feature points on the sequence. There are some methods to track feature points on image sequence. One is to track the feature points and remove outliers by the geometric constraints between successive images. Another method is a camera optical system applying is approximated by the affine camera model and tracking feature points. In this paper we took some image sequences and compared the results of experiments of the feature point tracking. Consequently, we obtained correct correspondences with high probability in methods by using geometric constraint. But the number of correspondences was small. On the other hand, we obtained many correct correspondences and some incorrect correspondences in method by using the affine space constraint.

Keywords: Computer vision, Feature point tracking, Image sequence, 3D reconstruction, Trifocal tensor

1. Introduction

There have been developed many methods to reconstruct 3D structure from image sequences, such as the the factorization method and the method which we carry out strict iteration based on perspective transformation and estimate camera parameters. Commonly to these methods, we must take point correspondences between frames. That is, we have to do tracking feature points in the image sequence. The tracking fails when the points go out of the field of view or behind other objects or the extracting of the feature point itself fails. Although when to reduce the number of frames prevents it, we need a large number of images taken from different view points to reconstruct 3D structure.

We carry out matching between successive images to track feature points. One of the method of image matching is: first we select candidates of correspondences using local correlation. Then, we remove the incorrect matches using the geometric constraints on the corresponding points. This method can be expected to acquire the correct corresponding points with high probability.

In another method, a camera optical system is approximated by the affine camera model. Then we extend interrupted tracking by imposing the constraint that under the affine camera model feature trajectories should be in an affine space in the parameter space. In this method the number of corresponding points can be increased, but if approximation of the camera model breaks down depending on a scene, the number of incorrect correspondences may increases by the incorrect estimation.

In this paper we carried out three feature points tracking methods. The first one is the method tracking the feature points, and removing incorrect correspondences by the geometry constraints described with fundamental matrix. The second one is the method tracking the feature points, and removing incorrect correspondences by using the geometry constraints of trifocal tensor. The third one is the method extending interrupted tracking by imposing the affine constraint.

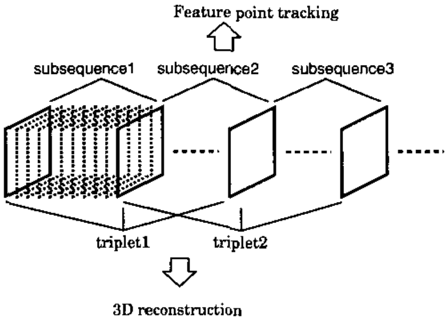

Figure 1: Image sequence

Consequently, we obtained correct correspondences with high probability in methods by using geometric constraint. But the number of correspondences was small. On the other hand, we obtained many correct correspondences and some incorrect correspondences in method by using the affine space constraint. We discuss the difference of the results based on the tracking method and the scene.

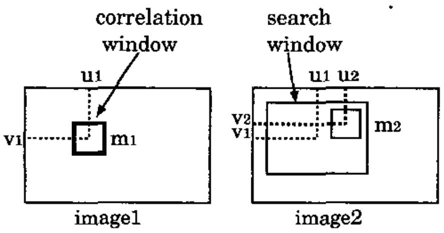

Figure 2: Selection of Candidate Match

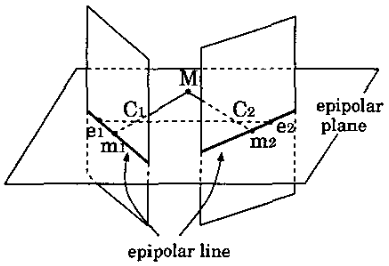

Figure 3: epipolar geometry

2. Tracking Feature Point

In this paper, as shown in Figure 1, an image sequence is divided into some subsequences. The feature points are tracked within subsequences to obtain correspondences. From these correspondences, a trifocal tensor is calculated, and the camera self-calibration is carried out. Then we reconstruct the 3D scene. We applied three methods described above for tracking feature points within the subsequences. Hereinafter we call the first method removing incorrect correspondences by the geometry constraint of fundamental matrix "fundamental matrix tracking", and the second method removing incorrect correspondences by the geometry constraint of trifocal tensor "trifocal tensor tracking", and we call the third method extending the interrupted tracking by imposing the constraint that under the affine space constraints "affine space tracking".

3. Selection of Candidate Match

We have to obtain candidate match between successive neighboring images to tracking feature points. We extract feature points from images by using Harris corner detector[1)] and the consistent gradient operator[2)]. Next, as shown in Figure 2 we set a correlation window centered at the point m_1 in imagel. We then select a rectangular search area around this point in the image2, and evaluate the correlation on the given window between point m_1 in the first image and all feature points m_2 lying within the search area in the image2. The correspondence which acquired high correlation is selected; and candidate matches are determined considering of other point correspondences the arrangement on each image of corresponding points.

4. Fundamental Matrix Tracking

After selecting the candidate match, we track the feature points by using the constraint of epipolar geometry. As shown in Figure 3, about cameral, object point $\boldsymbol{M}$, the optical center C_1 and the image point $\boldsymbol{m}_1$ are on a straight line. About camera2, object point $\boldsymbol{M}$, the optical center C_2 and the image point $\boldsymbol{m}_2$ are on a straight line. Thus all these points are on a certain plane in space. This plane is called the epipolar plane. Now there is following relation between two images.

$$\boldsymbol{m}_1^\top \mathbf{F} \boldsymbol{m}_2 = 0 \tag{1}$$

Equation (1) is called the fundamental equation, and $\mathbf{F}$ is called fundamental matrix. Suppose that a certain point (u_1, v_1) is given in the first image. Its homogeneous coordinates is $\boldsymbol{m}_1 = (u_1, v_1, 1)$, we set $\boldsymbol{m}_1^\top \mathbf{F}$ to (a_1, b_1, c_1). Then the corresponding point (u_2, v_2) in the second image is on the straight line (2).

$$a_1 v_2 + b_1 v_2 + c_1 = 0 \tag{2}$$

Thus there holds geometric constraint between two images,this means that the coordinates of corresponding points cannot locate arbitrarily each other. Incorrect candidate matches are removed evaluating the next distance of m_2 from the line (2).

$$distance(\boldsymbol{m}_1^\top \mathbf{F}, \boldsymbol{m}_2) = \frac{|\boldsymbol{m}_1^\top \mathbf{F} \boldsymbol{m}_2|}{\sqrt{(\boldsymbol{m}_1^\top \mathbf{F})_1^2 + (\boldsymbol{m}_1^\top \mathbf{F})_2^2}} \tag{3}$$

Calculation of fundamental matrix $\mathbf{F}$ was performed by the method of Zhang et al[8)] which uses RANSAC algorithm and the least median of squares method. We are able to obtain the $\mathbf{F}$, when more than a half of candidate matches are correct. Candidate matches will meet the requirement when difference of the views of two images is small. This is because the change of local image around the feature points is small.

5. Trifocal Tensor Tracking

A triplet images are geometrically constrained by a $3 \times 3 \times 3$ homogeneous tensor $\boldsymbol{T} = [\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3]$[4)]. This tensor $\boldsymbol{T}$ is called the trifocal tensor. The degree of freedom in points constraint by trifocal tensor is lower than that of fundamental matrix. It can be expected that more certainly correct correspondence can be obtained by employing the trifocal tensor. Three points constraint is given as (4).

$$[\boldsymbol{m}_2]_\times \left(\sum_i^3 m_1^i \mathbf{T}_i \right) [\boldsymbol{m}_3]_\times = \mathbf{O}_{3\times 3} \tag{4}$$

where $[\boldsymbol{m}]_\times$ is a skew-symmetric matrix

$$[\boldsymbol{m}]_\times = \begin{bmatrix} 0 & -m_3 & m_2 \\ m_3 & 0 & -m_1 \\ -m_2 & m_1 & 0 \end{bmatrix}$$

$\mathbf{T}_i$ is 3×3 matrix, and m_j^i is i-th element of the homogeneous coordinates $\boldsymbol{m}$ in the j-th image.($j = 1, 2, 3$)

Calculation of the trifocal tensor $\boldsymbol{T}$ was performed by using RANSAC algorithm and least median of squares method[5].

6. Affine Space Tracking

N feature points are tracked over K frames. The n-th point in the α-th image $\{p_n\}$ is expressed as $(x_{\alpha n}, y_{\alpha n})$, $\alpha = 1, ..., K$, $n = 1, ..., N$. The movement trajectory of the point is expressed with the following $2K$-dimensional vector, and it is called a trajectory vector.

$$\boldsymbol{p}_n = (x_{1n}y_{1n}x_{2n}y_{2n} \cdots x_{Kn}y_{Kn})^\top \qquad (5)$$

In this section we consider that scene moves relatively to camera which is fixed to the world coordinates. We set 3D coordinates to (X_n, Y_n, Z_n), the origin of scene coordinates in α-th image and basis vectors of each coordinates axis to, respectively, $\boldsymbol{t}_\alpha$, and $\{\boldsymbol{i}_\alpha, \boldsymbol{j}_\alpha, \boldsymbol{k}_\alpha\}$ which are represented in world coordinate system. The 3D coordinates of the feature point $\boldsymbol{p}_n$ in α-th image $\boldsymbol{M}_{\alpha n}$ is written as

$$\boldsymbol{M}_{\alpha n} = \boldsymbol{t}_\alpha + X_n \boldsymbol{i}_\alpha + Y_n \boldsymbol{j}_\alpha + Z_n \boldsymbol{k}_\alpha. \qquad (6)$$

Affine camera model which models weak-perspective and para-perspective, is assumed so that a 3-dimensional point $\boldsymbol{M}_{\alpha n}$ is projected on a image as,

$$\begin{pmatrix} x_{\alpha n} \\ y_{\alpha n} \end{pmatrix} = \mathbf{A}_\alpha \boldsymbol{M}_{\alpha n} + \boldsymbol{b}_\alpha. \qquad (7)$$

Here, $\mathbf{A}_\alpha$,and $\boldsymbol{b}_\alpha$ are 2×3 matrix and 2-dimensional vector, respectively, which are determined by camera location and intrinsic parameters in α-th image. Substituting (6) to (7) we have,

$$\begin{pmatrix} x_{\alpha n} \\ y_{\alpha n} \end{pmatrix} = \tilde{s}_{0\alpha} + X_\alpha \tilde{s}_{1\alpha} + Y_\alpha \tilde{s}_{2\alpha} + Z_\alpha \tilde{s}_{3\alpha} \qquad (8)$$

where $\tilde{s}_{0\alpha}, \tilde{s}_{1\alpha}, \tilde{s}_{2\alpha}$ and $\tilde{s}_{3\alpha}$ are 2-dimensional vectors which depend on location of camera in α-th image and camera intrinsic parameters. Aligning them according to $\alpha = 1, ..., K$, trajectory vectors $\boldsymbol{p}_n$ in (5) is written as

$$\boldsymbol{p}_n = \boldsymbol{s}_0 + X_n \boldsymbol{s}_1 + Y_n \boldsymbol{s}_2 + Z_n \boldsymbol{s}_3 \qquad (9)$$

where $\boldsymbol{s}_i, (i = 0, 1, 2, 3)$ are $2K$-dimensional vectors which $\tilde{s}_{i\alpha}$ are aligned $\alpha = 1, ..., K$ (9) represents that all the trajectory vectors $\boldsymbol{p}_n$ are included into "4-dimenisonal subspace" spanveel by $\{\boldsymbol{s}_0, \boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3\}$.

	zoom		wide	
	correspondence	incorrect	correspondence	incorrect
fundamental	163	1	97	0
matrix	169	0	117	1
tracking	172	0	114	0
trifocal	55	0	45	0
tensor	76	0	45	0
tracking	64	0	54	0
affine	380	3	408	8
space	375	2	409	5
tracking	413	3	446	15

Table 1: room sequence

	zoom		wide	
	correspondence	incorrect	correspondence	incorrect
fundamental	127	0	81	0
matrix	125	0	95	0
tracking	147	0	92	0
trifocal	42	0	33	0
tensor	47	0	34	0
tracking	40	0	34	0
affine	309	6	227	4
space	303	5	230	6
tracking	285	5	261	8

Table 2: dolls sequence

This is called "subspace constraints". But coefficient of $\boldsymbol{s}_0$ is 1 in common with all n. Then trajectory vectors $\boldsymbol{p}_n$ are included within "3-dimensional affine space" which is constraint 4-dimensional subspace[6]. This is the affine space constraint.

In this paper we extend interrupted tracking by using the method proposed by Tubouchi et al[6]. The method is as follows. First, when complete trajectory vectors are given these are verified statistically by using RANSAC algorithm. We removed outliers. We construct 3-dimensional affine space from inliers of complete trajectory vectors. Then, we verify partial trajectory vectors. Next, 3-dimensional affine space is recalculate from both complete trajectory vectors and partial trajectory vectors. Trajectory vectors are reverified whether they are inliers or outliers. It calculates iteratively until the affine space will not change.

7. 3D reconstruction

After we track feature points through out a subsequence, and triplets of images are made sequentially with the start images of each image subsequences (see Figure 1). We calculate trifocal tensor and perform self-calibration[7].

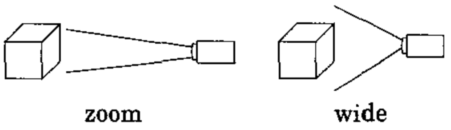

Figure 4: Difference of zoom and wide camera

8. Experimental Results

We took image sequences by a digital camera and used them for the experiments of the feature points tracking. Two scenes, "room" and "dolls", were taken by camera with zoom lens and with wide lens. Figure 4 shows the different of zoom and wide lens camera. We set a subsequences with 10 image frames, then we tracked feature points in each subsequence. The number of the feature points extracted in the first frame of zoom-room sequence was 351,363,374, wide-room sequence was 416,427,442, zoom-dolls sequence was 264,280,274, and wide-dolls sequence was 196,223,237. We show the number of points which are tracked to the final frame of each subsequences, and the number of incorrect correspondence in Table 1 and Table 2, respectively. We show the results of each first subsequence match in Figure 5- 8.

There were few incorrect correspondences in the fundamental matrix tracking and the trifocal tensor tracking. But the number of correspondences was also small. In the trifocal tensor tracking, the constraint was more severer, and many correspondences disappeared.

Meanwhile, we obtained many correspondences in the affine space tracking. The result was better than the other two methods even after the incorrect correspondences were deducted. Moreover there was larger number of incorrect correspondences in the sequence taken with a wide lens rather than the sequence taken with a zoom lens.

This is because the affine camera model was well approximated when the sequence was taken with a zoom lens, which was more close to orthogonal projection (see Figure 4). When we compared two affine space tracking, the difference of a zoom lens and a wide lens appeared in the tracking result more clearly in room sequence. We think this is because of the difference of the depths of scene objects.

Approximated by affine camera model, zoom sequence is better than wide sequence. However, when the depth of the scene objects is small, then the difference of perspective projection and orthogonal projection became small. Herefrom, the effects of lens appeared in the room sequence in which depth is large.

Figure 9 shows the object shapes which are reconstructed from the correspondences. We created surface with triangular patches and put textures on these triangular patches from image sequence. For the trifocal tensor tracking, since the number of correspondence was too small so that self-calibration failed.

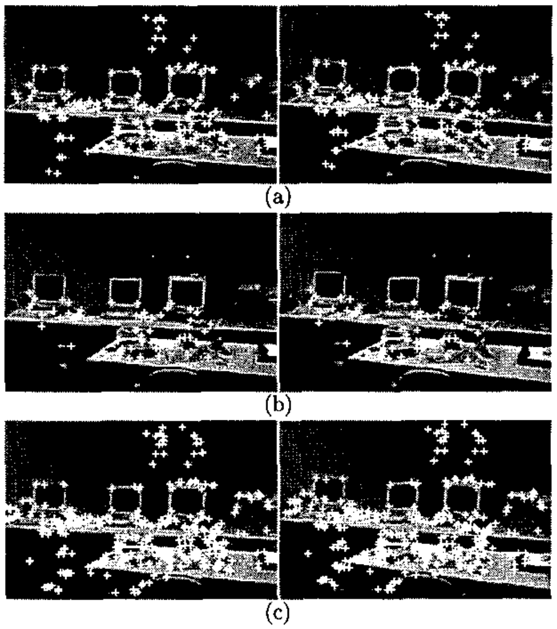

Figure 5: Room sequence taken with a zoom lens. (a):fundamental matrix tracking, (b):trifocal tensor tracking and (c):affine space tracking, Point correspondences between the first and last images in the first subsequence are shown "+"

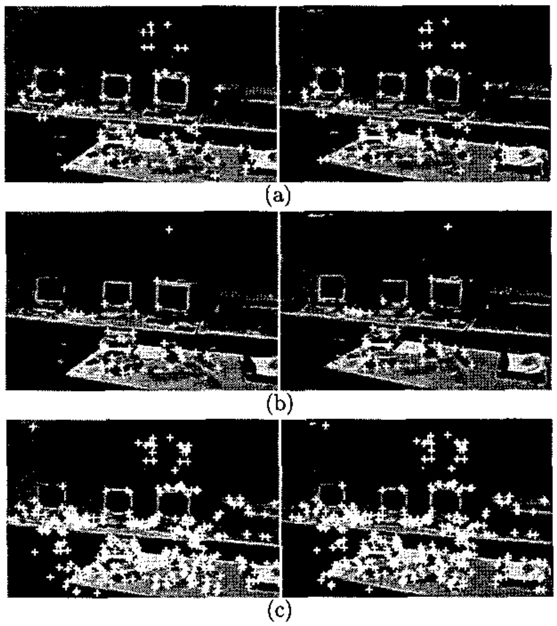

Figure 6: Room sequence taken with a wide lens. (a):fundamental matrix tracking, (b):trifocal tensor tracking and (c):affine space tracking, Point correspondences between the first and last images in the first subsequence are shown "+"

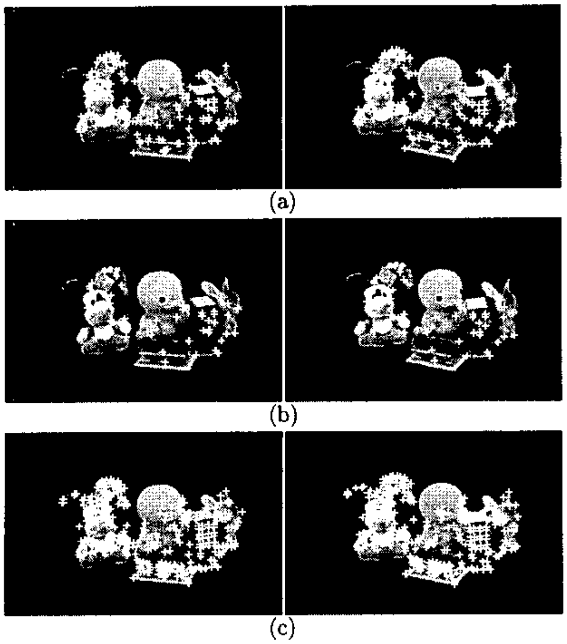

Figure 7: Dolls sequence taken with a zoom lens. (a):fundamental matrix tracking, (b):trifocal tensor tracking and (c):affine space tracking, Point correspondences between the first and last images in the first subsequence are shown "+"

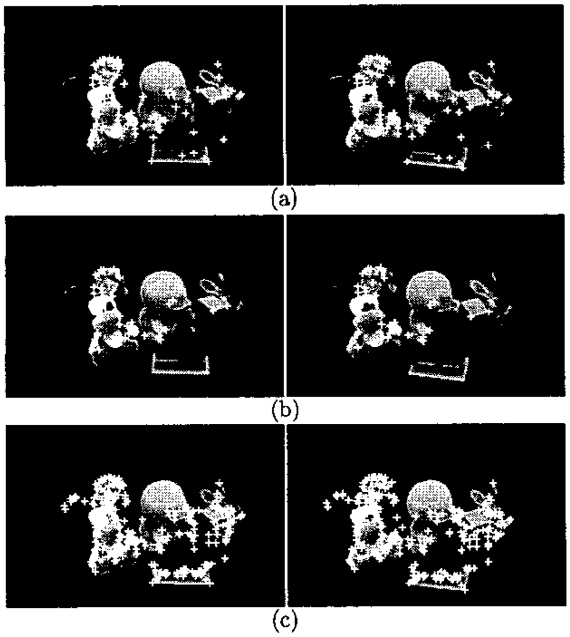

Figure 8: Dolls sequence taken with a wide lens. (a):fundamental matrix tracking, (b):trifocal tensor tracking and (c):affine space tracking, Point correspondences between the first and last images in the first subsequence are shown "+"

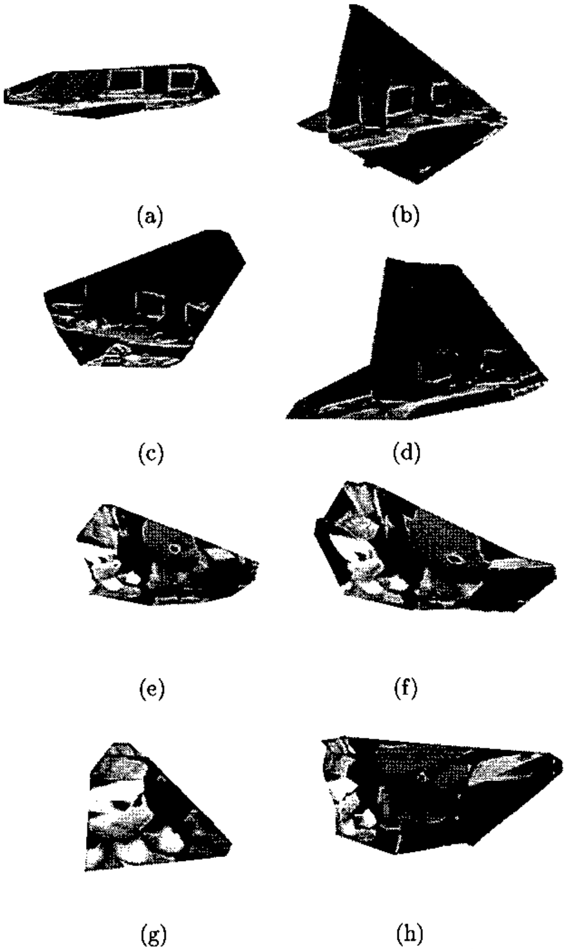

Figure 9: The objects shape which reconstructed from correspondences. (a):room, zoom, fundamental matrix, (b):room, zoom, affine space, (c):room, wide, fundamental matrix, (d):room, wide, affine space, (e):dolls, zoom, fundamental matrix, (f):dolls, zoom, affine space, (g):dolls, wide, fundamental matrix, (h):dolls, wide, affine space

9. Conclusion

In this paper we compared three technique to track feature points on image sequence. Fundamental matrix tracking and trifocal tensor tracking were able to hold correct correspondences in all scene. However the number of correspondences were small. The number of correspondences in trifocal tensor tracking was the least. On the other hand, we obtained many correspondences in affine space tracking.

These affine space tracking results depended on the lens and the depth of scene objects. When the scene objects does not have large depth and the scene was not taken with a wide lens the approximation of an affine camera worked well. There were some incorrect correspondences in affine space tracking, but finally we could remove them by geometric constraints because we made triplets of images in the further steps. So, if we want to obtain many correspondences affine space tracking is effective.

References

[1] C. Harris and M. Stephens. A combined corner and edge detector. Proceedings of the Alvey Conference, pp. 189-192, 1988.

[2] S. Ando, Consistent Gradient Operators. IEEE Trans. Pattern Anal. Machine Intell., vol.22, No.3, pp. 252-265, 2000.

[3] A. W. Fitzgibbon and A. Zisserman, Automatic Camera Recovery for Closed or Open Image Sequences. Proc. European Conference on Computer Vision, Springer-Verlag, pp. 311-326. 1998.

[4] R. I. Hartley and A. Zisserman. Multiple View Geometry in Computer Vision. Cambridge University Press, 2000.

[5] P. H. S. Torr and A. Zisserman. Robust parameterization and computation of the trifocal tensor. Image and Vision Computing, vol. 15, pp. 591-605, 1997.

[6] T. Tsubouchi, Y. Sugaya, K. Kanatani, Extending Interrupted Feature Point Tracking for 3-D Affine Reconstruction, Proceedings of the 9th Symposium on Sensing via Imaging Information, pp. 11-13, 2003.

[7] M. Pollefeys, R. Koch and L. V. Gool, Self-Calibration and Metric Reconstruction Inspite of Varying and Unknown Intrinsic Camera Parameters. International Journal of Computer Vision, vol. 32, No. 1, pp. 7-25, 1999.

[8] Z. Zhang, R. Deriche, O. Faugeras, and Q. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. Arificial Intelligence, vol. 78, pp. 87-119, 1995.