Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Pattern Recognition Letters 25 (2004) 733–741
Pattern Recognition Letters
www.elsevier.com/locate/patrec

Two quantitative measures of inlier distributions for precise fundamental matrix estimation

Jung-Kak Seo [a], Hyun-Ki Hong [a,*], Cheung-Woon Jho [a], Min-Hyung Choi [b]

[a] *Department of Image Engineering, Graduate School of Advanced Imaging Science Multimedia & Film, Chung-Ang University, 221 Husuk-Dong, Dongjak-Ku, Seoul 156756, South Korea*
[b] *Department of Computer Science and Engineering, University of Colorado at Denver, Denver, CO, USA*

Received 16 April 2003; received in revised form 19 September 2003

Abstract

Because the estimation of a fundamental matrix is much dependent on the correspondence, it is important to select a proper inlier set that represents variation of the image due to camera motion. Previous studies showed that a more precise fundamental matrix can be obtained if the evenly distributed points are selected. When the inliers are detected, however, no previous methods have taken into account their distribution. This paper presents two novel approaches to estimate the fundamental matrix by considering the inlier distribution. The proposed algorithms divide an entire image into several sub-regions, and then examine the number of the inliers in each sub-region and the area of each region. In our method, the standard deviations are used as quantitative measures to select a proper inlier set. The simulation results on synthetic and real images show that our consideration of the inlier distribution can achieve a more precise estimation of the fundamental matrix.

Keywords: Stereo vision; Fundamental matrix; Epipolar geometry; Correspondence; Inlier set

1. Introduction

Stereo vision, a useful technique for obtaining 3-D information from 2-D images, has many practical applications including robot navigation and realistic scene visualization. Given a point in the one image, we find the corresponding point in the other image so that the two points are the projections of the same physical point in space. In this process the fundamental matrix representing succinctly the epipolar geometry of stereo vision is estimated. The fundamental matrix contains all available information on the camera geometry and it can be computed from a set of point correspondences (Faugeras, 1992; Trucco and Verri, 1998).

Several methods to estimate the fundamental matrix have been proposed for over two decades. The main difficulty in estimating the fundamental

* Corresponding author.
E-mail addresses: pikei24@hotmail.com, jkseo@imagelab.cau.ac.kr (J.-K. Seo), honghk@cau.ac.kr (H.-K. Hong), blue@cglab.cau.ac.kr (C.-W. Jho), minchoi@acm.org (M.-H. Choi).

0167-8655/$ - see front matter
doi:10.1016/j.patrec.2004.01.014

matrix stems from the unavoidable outliers inherent in the given correspondence matches. Because the outliers by a false match can severely disturb estimation of the fundamental matrix, we should identify and then reject them. The robust algorithms to solve the problems of errors on the point locations and mismatches have been actively studied up to now (Zhang et al., 1994). However, these methods based on the aleatory way of selecting the points are much affected by which the corresponding points are selected (Rousseeuw, 1987; Fischler and Bolls, 1981). In general the evenly distributed corresponding points can thoroughly represent the variation of the image due to the camera motion, so it is important to select a proper inlier set for a more precise fundamental matrix. However, no previous method has taken into account the distributions when the inliers are detected.

This paper presents two novel quantitative measures to select a proper inlier set by considering the inlier distributions. Our method is based on least-median-squares (LMedS), which calculates the median of distances between points and epipolar lines for each fundamental matrix (Zhang et al., 1994). In this process LMedS eliminates many outliers, and then inliers and the fundamental matrix are obtained. The first method divides an image uniformly into several sub-regions based on the number of the inliers. We calculate the standard deviation of the number of the inliers in each subregion and the average value, which is the total number of the inliers divided by the number of subregions. The obtained standard deviation represents the degree of the point distribution in each sub-region relative to the entire image. The obtained information is used as a quantitative measure to select the evenly distributed point set. The second method uses the Delaunary triangulation connecting the inliers to decompose an image into non-uniform sub-regions, and the area of each subregion is computed. We calculate the average area by dividing the area of the image by the total number of the triangles. Then the standard deviation of the area of each sub-region and the average is used as a quantitative measure. The experimental results on synthetic and real images show that our consideration of the inlier distribution can provide a more precise estimation of the fundamental matrix.

The structure of the paper is as follows. Section 2 introduces the epipolar geometry and reviews previous work on robust estimators for the fundamental matrix. Section 3 discusses the effect due to the inliers distribution in estimating the fundamental matrix. After details of two novel algorithms to consider the inliers distribution are presented in Section 4, we show the experimental results for synthetic and real scenes in Section 5. Finally, the conclusion is described in Section 6.

2. Epipolar geometry

Epipolar geometry is a fundamental constraint used whenever two images of a static scene are to be registered. Given a point in one image, corresponding point in the second image is constrained to lie on the epipolar line. All the epipolar geometry is contained in the fundamental matrix (Hartley, 1997). The epipolar constraint can be written as:

$$u'^{\mathrm{T}}Fu = 0, \tag{1}$$

where u and u' are the homogeneous coordinates of two corresponding points in the two images. F is the fundamental matrix that has rank 2, and since it is defined up to a scale factor, there are seven independent parameters. From Eq. (1), the fundamental matrix can be estimated linearly, given a minimum of eight corresponding points between two images (Hartley, 1997; Longuet-Higgins, 1981). Because the fundamental matrix contains the intrinsic parameters and the rigid transformation between both cameras, it is widely used in various areas such as stereo matching, image rectification, outlier detection and computation of projective invariants.

Several algorithms for the estimation of a fundamental matrix are categorized into three methods: the linear, the iterative and the robust (Salvi et al., 2001). The linear and the iterative methods use some points to estimate the fundamental matrix. First, the linear approaches, such as eight-point algorithm, estimate the fundamental matrix by using eight corresponding points. With more than eight points, a least mean square minimization is used, followed by the enforcement of the singularity constraint so that the rank of the resulting matrix can be kept in 2. These approaches have been proven to be fast and easy to implement, but they are very sensitive to image noise. Secondly, the iterative methods are based on optimization criteria, such as the distance between points and epipolar lines, or the gradient-weighted epipolar errors (Hartley and Zisserman, 2000). Although these methods are more accurate than the linear, they are time consuming and much affected by the unavoidable outliers inherent in the given correspondence matches and the error on the point locations. Finally, the robust methods such as LMedS and random sampling consensus (RANSAC), can cope with either outliers or bad point localization. RANSAC uses a minimum subset for parameter estimation and the solution is given by the candidate subset that maximizes the number of consistent points and minimizes the residual (Fischler and Bolls, 1981). It is computationally infeasible to consider all possible subsets, since the computation load grows exponentially according to the number of the inliers. Therefore, additional statistical measures are needed to derive the minimum number of sample subsets. In addition, because of the restrictive way of sampling the points randomly, the obtained fundamental matrix can be much changed by which points are selected.

3. Consideration of the point distribution

In order to cope with the unavoidable outliers inherent in the given correspondence matches, our methods are based on LMedS that calculates the median of distances between points and epipolar lines for each fundamental matrix. RANSAC paradigm finds a significant set of data points consistent with consensus group, and the remaining data points are rejected as outlier. Instead of finding a large consensus group, LMedS estimates the model parameters by minimizing the median of the squared residuals computed from the entire data set (Zhang et al., 1994). The residual error is defined as:

$$r = d(u, F^{\mathrm{T}}u') + d(u', Fu), \tag{2}$$

where $d(x, l)$ is the distance between a point x and a line l.

RANSAC method selects the biggest consensus set of inliers that the residual error is smaller than a user defined threshold. The main difficulty with RANSAC is that a compatibility threshold has to be preset. On the contrary, the LMedS method does not require such a threshold. The estimates must yield the smallest value for the median of squared residuals computed from the entire data set. RANSAC depends on the number of inliers and the LMedS method does the least median of residual. However, these approaches are insufficient for detection of a proper inlier set for the estimation of a precise fundamental matrix. In other words, because the fundamental matrix contains the relative orientation and position between both cameras, the inliers should reflect sufficiently 3D structure of the scene as well as the variation of the image due to camera motion. If the proper inlier sets are selected based on the point distribution, we can achieve a more precise estimation of the fundamental matrix.

In order to show whether the evenly distributed point set is effective for the fundamental matrix estimation, we have simulated on two pairs of synthetic images from a same camera motion as shown in Fig. 1(a). In the one pair of synthetic images, feature points are mainly distributed on the left-sided region, and the points are evenly distributed in the other region. We have tested 100 times on each image by the LMedS method. Fig. 1(b) and (c) present distribution of feature points and the distance error graph for the obtained epipole relative to the real at each time, respectively. The dot line and the solid line in (c) represent the distance error of the left and right images in Fig. 1(a). The simulation results show that if the evenly distributed inlier sets are detected, we can estimate a more precise fundamental matrix. This paper presents two quantitative methods to evaluate the distribution of the inliers.

3.1. Point density

The standard deviation of the point density in sub-regions and an entire image can be used to evaluate whether the points are evenly distributed.

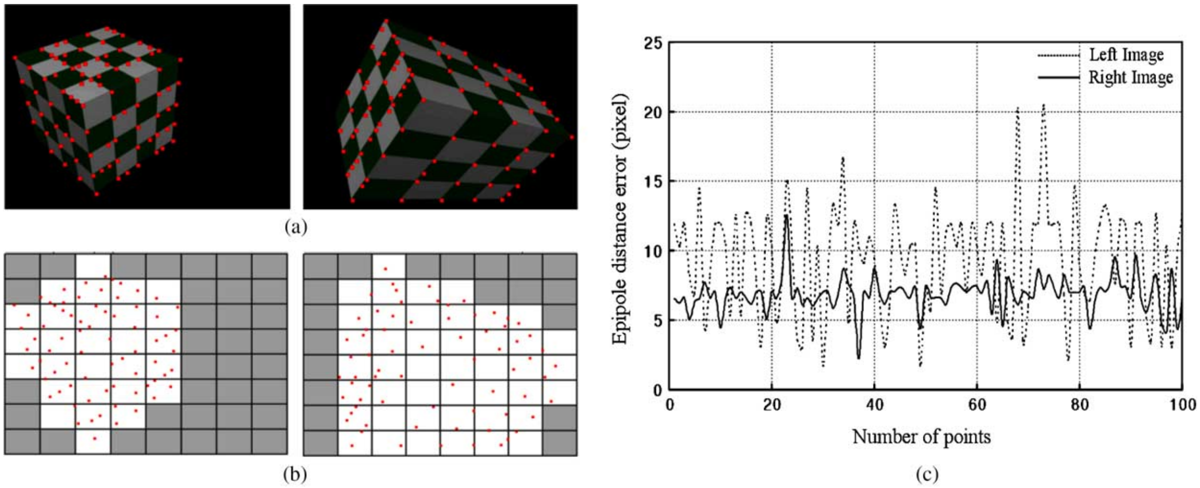

Fig. 1. Effects of the points distribution for the fundamental matrix estimation: (a) synthetic images with a same camera motion, (b) distribution of feature points and (c) distance error graph of LMedS method.

The image is divided into the uniform sub-regions according to the number of the inliers by Eq. (3), and then we calculate the point density of sub-regions and an entire image. The width (W_s) and the height (H_s) of the sub-region can be written as:

$$W_s = W/\mathrm{int}(\sqrt{N}), \quad H_s = H/\mathrm{int}(\sqrt{N}), \tag{3}$$

where N is the number of the inliers, and $\mathrm{int}(\cdot)$ means conversion to integer. W and H are the height and the width of the image. Two left images in Fig. 2 show sub-regions divided uniformly on the synthetic image and the real. The proposed method computes the standard deviation of two densities that represents a degree of the point distribution in each sub-region relative to the entire image. The obtained information is used as a quantitative measure to select the evenly distributed point sets. The standard deviation of the point density is defined as:

$$\sigma_p = \sqrt{\frac{1}{N_s}\sum_{i=1}^{N_s}\left(P_{si} - \frac{N}{N_s}\right)^2}, \tag{4}$$

where N_s is the number of sub-regions, N and P_{si} are the number of inliers and that in the ith sub-region, respectively.

3.2. Area density

In the second method, the Delaunay triangulation connecting the inliers decomposes the image into non-uniform sub-regions. The main rule for the Delaunay triangulation can be formulated in the circle criterion (Delaunay, 1934). The Delaunay triangulation in 2-D consists of non-overlapping triangles where no points in the triangle are enclosed by the circumscribing circles of another triangle. As shown in Fig. 2, the entire image is segmented into several non-uniform sub-regions, and the area of each sub-region is computed. We obtain the average area by dividing the area of the image by the number of the triangles, and the standard deviation of the area of each sub-region and the average is used as the quantitative measure. The area of triangle is defined as:

$$A_\Delta = \tfrac{1}{2}\|e_2\|\left\|v_2 - \frac{e_1^T e_2}{\|e_2\|}e_2\right\|, \tag{5}$$

where v_i is the ith vertex of the triangle, e_1 and e_2 are vectors, $e_1 = v_2 - v_1$, $e_2 = v_3 - v_1$. The standard deviation of the area density is defined as:

$$\sigma_A = \sqrt{\frac{1}{N_T}\sum_{i=1}^{N_T}(A_{\Delta i} - A_{aver})^2}, \tag{6}$$

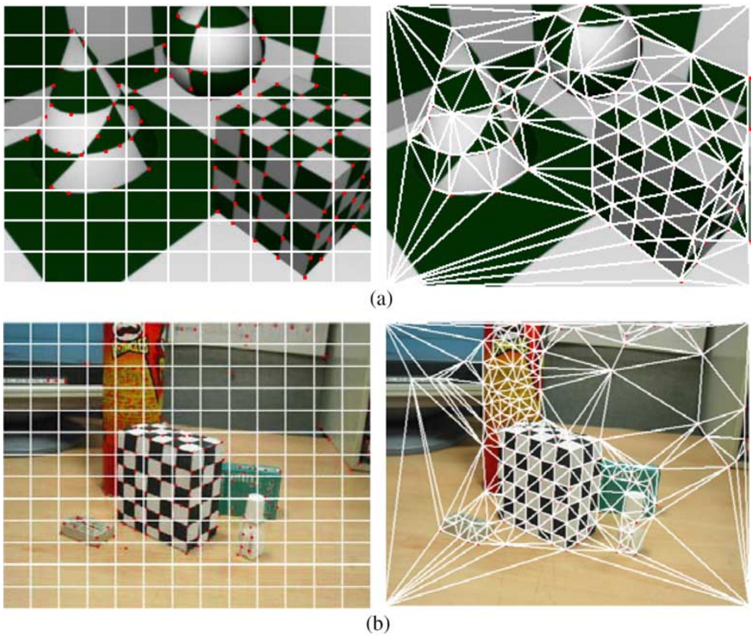

Fig. 2. Uniformly and non-uniformly segmented sub-regions by the proposed methods: (a) synthetic and (b) real image.

where N_T and A_{aver} are the number of triangles by Delaunay triangulation and the average area.

4. Proposed algorithm

LMedS method chooses the fundamental matrix with the least median of residual. As shown in Table 1, however, the median minimization approach always does not guarantee a precise solution. The proposed algorithm combines LMedS method with the quantitative measure to select the evenly distributed point sets. Among the inliers in the range from the least median to the larger medians by 10%, we choose the inlier set with the least deviation of the inlier distribution. Sum of squared differences (SSD) is used as a matching method, and it establishes the correspondence between images. The iterative number is computed by following equation:

$$N = \log(1 - P)/\log(1 - (1 - \varepsilon)^q), \quad (7)$$

where P is the probability that these points are the inliers in sampling q points at N times. For example, in case that every point is the inlier, P is 1. In general, P is near 1. ε is the ratio of the outlier to all points, and q is the number of the sample

Table 1
Distance errors of epipoles with the smallest five medians on the synthetic image (Fig. 3)

	Median value	x	y	Distance error	Proposed method	
					Eq. (4)	Eq. (6)
		380.50	180.20	0.00		
Case 1	0.01385	389.40	189.18	12.64	1.2157	82343.09
Case 2	0.02252	380.12	183.17	2.99	1.1742	81594.95
Case 3	0.02668	382.60	186.87	6.99	1.2278	82778.39
Case 4	0.03349	378.47	184.30	4.57	1.2398	83923.38
Case 5	0.04224	392.00	179.24	11.54	1.2238	82715.67

points. Since the eight point algorithm uses eight points to estimate the fundamental matrix, q is 8. In the LMedS method, a threshold to detect the inliers is calculated by using the median value as the following (Zhang et al., 1994):

$$\tau = 2.5 \times 1.4826[1 + 5/(n - q)]\sqrt{\text{median}}, \quad (8)$$

where n is the number of correspondences.

The proposed method chooses each inlier set by using the computed median of residual, and computes the standard deviation of the distribution of each inlier set by Eqs. (4) and (6). Then we find the inlier set with the least standard deviation. In the final step, we re-estimate the fundamental matrix from the selected inlier set by minimizing a cost function that is the sum of residual of the inliers. The proposed algorithm is summarized as follows:

1. Select a random sample of eight correspondence points and compute the fundamental matrix (F matrix) by using the selected point set, then calculate the median of residual.
2. Repeat 1 for N samples (Eq. (7)) and store the median and F matrix set.
3. Select the least median and the larger medians than that by 10%.
4. Detect each inlier set by using threshold (Eq. (8)) and compute the standard deviation of the distribution of the inlier set (Eqs. (4) and (6)).
5. Re-estimate F matrix from the inlier set with the least standard deviation by minimizing a cost function that is the residual sum of the inliers.

5. Experimental results

We have compared the experimental results of previous methods such as normalized eight-point, RANSAC and the LMedS method with the proposed algorithm on synthetic and real images. In order to check whether an optimal solution is obtained in terms of the median error minimization, we compute a real epipole on the synthetic image (Fig. 3), and obtain the distance error between the real and the epipole by the LMedS method. In this simulation, we can freely control the camera parameters and motion, and the real epipole and the fundamental matrix of the synthetic images are computed precisely. Table 1 shows the obtained epipoles with the smallest five medians on the synthetic image pair (Fig. 3). In Table 1, x and y represent the epipole position, and "distance error" describes a distance error between the real epipole and the computed. The simulation results show that the median minimization approach always does not guarantee a precise solution. In other words, as shown in Table 1, although the least median value in x and y is obtained in case 1, we cannot always obtain a precise fundamental matrix. Table 1 shows the proposed methods obtain the smallest values in case 2, so we can compute a more precise epipole in terms of the distance error.

Fig. 3(a) shows the inlier set and the obtained epipolar lines on the synthetic image pair, and (b) represents the accumulation distance errors between the real epipole and the computed on 100 iterations. The proposed algorithm is compared with previous methods, RANSAC and LMedS method, when the sampling times and the number of inliers is 500 and 120, respectively. In (b), the proposed method obtains the best performance over previous methods because of our consideration of the inlier distribution. (c) provides the comparison of the epipole distance error according to the number of the matching points. As the number of inliers becomes large, all methods show better performances. Normalized eight points algorithm, one of the direct methods, obtains the worst results over other methods. In addition, the proposed method gives relatively better results over LMedS when the number of points goes more than 80. The epipole distance errors according to the sampling time are analyzed in (d) and the number of the matching point is 120. Because many points satisfying the specified condition are used, RANSAC gives a relatively stable performance independent of the sampling time. Although LMedS method shows better results over RANSAC, its performance is irregularly and largely affected by the sampling, and a large sampling numbers do not always produce a good performance. On the contrary, the proposed methods

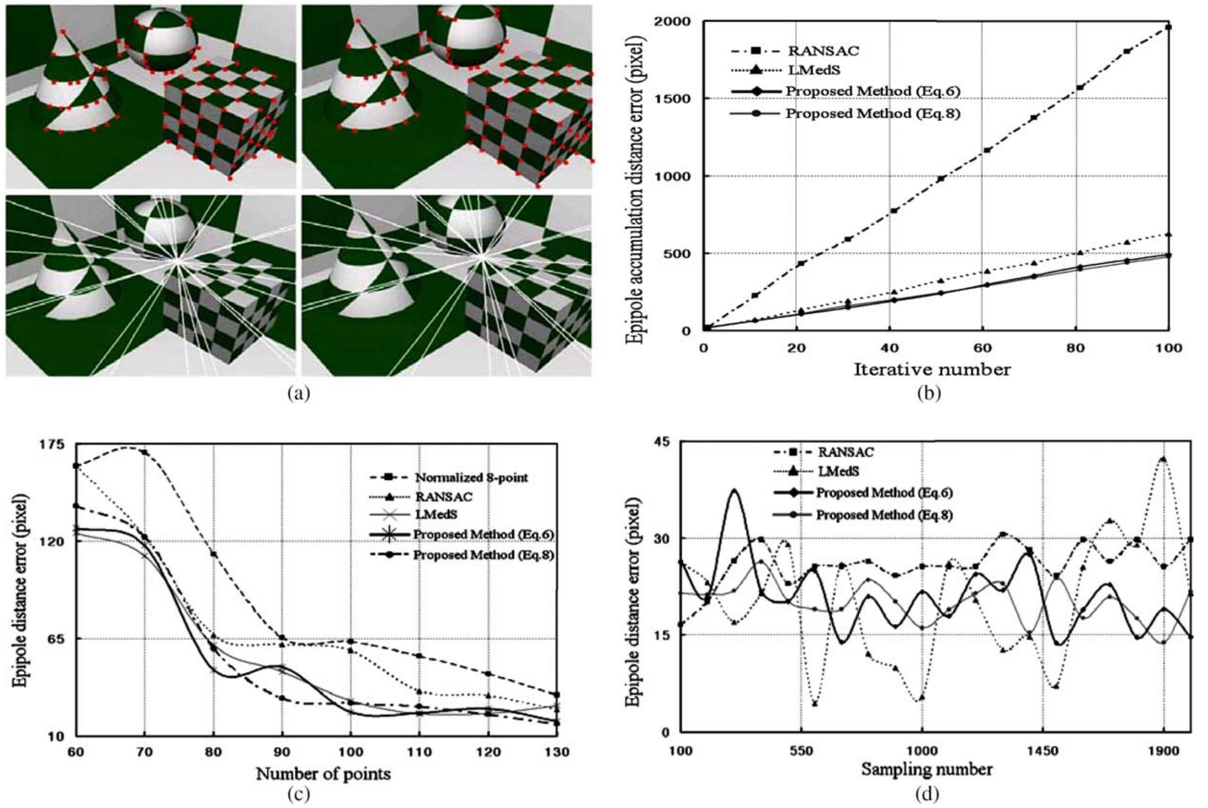

Fig. 3. Experimental results on the synthetic image pair: (a) inliers set and the epipole, (b) accumulation distance error between the real epipole and the computed, (c) the epipole distance error according to the number of the matching points and (d) epipole distance errors according to the sampling time.

obtain relatively better results than previous methods.

Fig. 4 shows the experimental results on real image. (a) shows a calibration rig image (left) and all of the corresponding point flows (right) by matching based on sum of square difference (SSD). (b) represents epipolar lines by camera calibration. Epipolar lines and the selected inlier set flows by LMedS (left) and those by the proposed method (right) are shown in (c). In the experiment, because the LMedS considers only residual error, many points located on the right and upper sides are lost. These points are important for camera motion analysis, so the results by LMedS are not precise as shown in (c). In comparison of the experimental results on the real image, we have ascertained that the proposed method can select the evenly distributed inliers and estimate the fundamental matrix more precisely.

6. Conclusion

For the estimation of an accurate fundamental matrix, it is important to select the inlier sets that reflect the structure and the variation of the image due to camera motion. Previous methods often use the least residual errors to select the inliers. However, the experimental results show that the median minimization does not always guarantee a precise solution. This paper shows that if the evenly distributed inliers are selected, we can estimate a more precise fundamental matrix. Two quantitative measures to select evenly distributed points by considering the inlier distribution in an entire image are presented. The proposed algorithm divides the entire image into the sub-regions, and then examines the number of the inliers in each sub-region and the area of each region. The

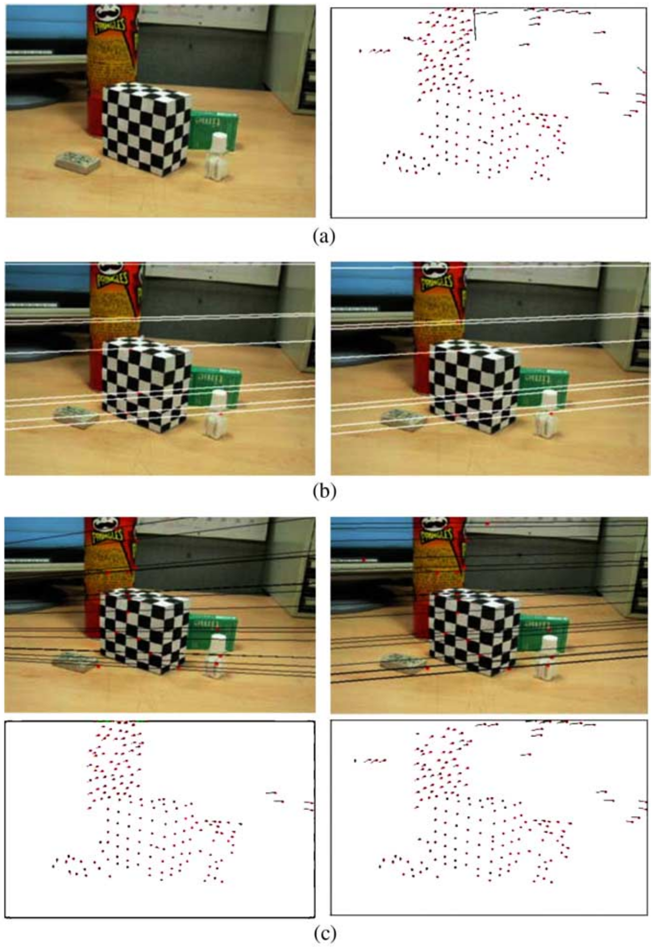

Fig. 4. Experimental results on the real image pair: (a) calibration rig image (left) and all of the corresponding point flows (right), (b) epipolar lines by camera calibration and (c) epipolar lines and the selected inlier set flows by LMedS (left) and the proposed method (right).

experimental results on synthetic and real images show that our consideration of the inlier distribution can achieve provide a more precise estimation of the fundamental matrix. These methods can be used as a base for further work on camera pose estimation and 3D scene geometry recovery.

Acknowledgements

This research has been supported by the Ministry of Education, Korea, and under the BK21 project, and Ministry of Science and Technology, Korea, under the NRL (2000-N-NL-01-C-285) project.

References

Delaunay, B., 1934. Sur la sphere vide. Bull. Academy Sci. USSR, 793–800.

Faugeras, O.D., 1992. What can be seen in three dimensions with an uncalibrated stereo rig? In: G. Sandini, (Ed.), Proceedings of 2nd European conference on Computer Vision. Lecture Notes in Computer Science, vol. 588. pp. 563–578.

Fischler, M., Bolls, R., 1981. Random sample consensus: A paradigm for model fitting with application and automated cartography. Comm. ACM 6 (24), 381–395.

Hartley, R., 1997. In defense of the eight-point algorithm. IEEE Trans. Pattern Anal. Mach. Intell. 19 (7), 580–593.

Hartley, R., Zisserman, A., 2000. Multiple View Geometry in Computer Vision. Cambridge University.

Longuet-Higgins, H.C., 1981. A computer algorithm for reconstructing a scene from two projections. Nature 293 (9), 133–135.

Rousseeuw, P.J., Leroy, A.M., 1987. Robust Regression and Outlier Detection. John Wiley & Sons, New York.

Salvi, J., Armangue, X., Pages, J., 2001. A survey addressing the fundamental matrix estimation problem. In: Proceedings of International Conference on Computer Vision, pp. 209–212.

Trucco, E., Verri, A., 1998. Introductory Technique for 3D Computer Vision. Simon & Schuster, New Jersey.

Zhang, Z., Deriche, R., Faugeras, O., Loung, Q., 1994. A robust technique for matching two uncalibrated images through the recover of the unknown epipolar geometry. Tech. Rep. 2273, Institut National de Recherche en Informatique et Automatique.