

Epipolar Geometry Estimation for Wide Baseline Stereo

Zhang Dazhi, Wang Yongtao, Tao Wenbing

*Institute for Pattern Recognition and Artificial Intelligence Huazhong University of Science and Technology
Wuhan 430074, China*

Abstract

The epipolar geometry is the intrinsic projective geometry between two views, and the fundamental matrix is the algebraic representation of epipolar geometry. Recovery of epipolar geometry is a fundamental problem in computer vision. Its importance is due to the fact that it provides relationships between corresponding point in the two images. In this paper, the problem of automatic robust estimation of the epipolar geometry for wide-baseline image pair is addressed. this problem for wide-baseline image pair is difficult because the putative correspondences include a low percentage of inlier correspondences, and it could become a severe problem when the veridical data are themselves degenerate or near-degenerate. Base on our previous work, a topological clustering(TC) is proposed to apply to fundamental matrix estimating. The TC algorithm has been demonstrated to be able to effectively eliminate the mismatches and reserve the correct matches. This advantage is extremely important to speeds up the performance of the epipolar geometry estimation and avoid the degeneracy. First, a set of match clusters are generated from the initial SIFT matches using topological clustering algorithm. Then, all the valid pairs of clusters are used to generate a series of fundamental matrix estimates and the best estimate is chosen as the solution. Six famous image pairs are used to test the proposed algorithms and the comparison with the related methods has been conducted. The compared experiments emphasize the performance of the proposed CPC algorithms.

Index Terms: Computer vision; Wide baseline stereo; fundamental matrix; topological clustering

© 2012 Published by MECS Publisher. Selection and/or peer review under responsibility of the Research Association of Modern Education and Computer Science.

1. Introduction

Recovery of epipolar geometry is a fundamental problem in computer vision. It is an important building block in many applications such as ego-motion estimation, structure from motion, and visual servoing. Given a pair of images, the fundamental matrix $F \in R^{3 \times 3}$ is defined as the matrix satisfying the relation

* Corresponding author.

E-mail address: zdjoseph@gmail.com

$$x'^T Fx = 0 \quad \forall x', x \quad (1)$$

Where $x, x' \in R^{3 \times 3}$ are the projections expressed in homogeneous coordinates of the same 3D point in the two images. In this paper, we address the problem of robust estimation for wide baseline stereo (WBS). Unlike classical short baseline stereo techniques, wide baseline stereo algorithms can tolerate a large change in viewpoint between the images. Although technically more difficult, wide baseline stereo has a number of advantages, including higher precision of depth measurement and smaller number of images needed to completely cover an object or scene. The price to be paid for these advantages is the need to cope with more frequent occlusions and large geometric and radiometric distortions. As a result, the correspondences are contaminated with a high percentage of outliers, which are inevitably included in an initial fit and can so distort a fitting process that the fitted parameters become arbitrary. This is particularly severe when the veridical data are themselves degenerate or near-degenerate with respect to the model, for then outliers can appear to break the degeneracy.

In [1], traditional statistical estimators have breakdown points that are no more than 50 percent. These robust estimators assume that the inliers occupy the absolute majority of the whole data, which is far from being satisfied with wide baseline. It frequently happens that outliers occupy the absolute majority of the data.

Random sampling and consensus (RANSAC) [2] is the most widely used robust estimator in computer vision today. Rather than using as much data as is possible to obtain an initial solution and then attempting to identify outliers, as small a subset of the data as is feasible to estimate the parameters.

Recently some other methods based on RANSAC has been proposed. The MLESAC [3] (Maximum Likelihood Sample Consensus) evaluates the likelihood of the hypothesis, representing the error distribution as a mixture model in which the inlier error is Gaussian and the outlier error is uniform. Similar methods were suggested in [4] where nonparametric density estimation has been used to the robust regression problem. In [5], a computational definition of rigidity and a probabilistic criterion to rate the meaningfulness of a rigid set were proposed, this approach permits to detect rigidity and to improve the performance of classical RANSAC algorithms.

In spite of their success, these random sampling algorithms are limited by their drawbacks. First, as a stochastic algorithm, none of them can generate a determinate solution. Second, most of these algorithms are no mechanism for detecting degeneracy.

Recently, in [6], we develop a novel topological clustering algorithm based on the topological relationship consistence between matches, which performs very well in discarding wide baseline mismatches and outperforms the earlier methods for discarding wide baseline mismatches .

Furthermore, we develop a novel method to estimate the fundamental matrix for the wide-baseline image pair. The essential of this method is that pairs of match clusters instead of match samples are used to generate a series of fundamental matrix estimates. Experimental result shows that the drawbacks of random sampling algorithms are well overcome by our method.

The remainder of the paper is organized as follows. In section II we review our earlier work about “topological clustering”. Section III describes our novel method for estimating the fundamental matrix for the wide-baseline image pair. The data set and evaluation criterion as well as implementation details for the methods used in our comparison are introduced in section IV, and the experimental results are given in this section too. Finally, section V closes the paper with conclusions.

2. Topological Clustering

A novel algorithm called “topological clustering” is invented in our recent work [6]. Using this algorithm, an initial region match set can be automatically partitioned into a set of match clusters which are topologically invariable between frames.

Define I as a given image in image domain Ω and $R_i(i=1,2,\dots,N)$ is the regions in image domain Ω . Then, there are four topological relationships between two regions R_i and R_j : (a) they are disjoint, i.e. $R_i \cap R_j = \phi$, (b) they intersect, i.e. $R_i \cap R_j \neq \phi$, $R_i \not\subset R_j$, $R_j \not\subset R_i$, (c) R_i includes R_j , i.e. $R_i \supset R_j$, (d) R_i is included in R_j , i.e. $R_i \subset R_j$. However, only the last three relationships (b)-(d) are taken into account, and Fig. 1 illustrates three kinds of connected region match pairs. Based on the preceding definition of the connectedness for two region matches, any region match set can be divided into a collection of connected components namely match clusters.

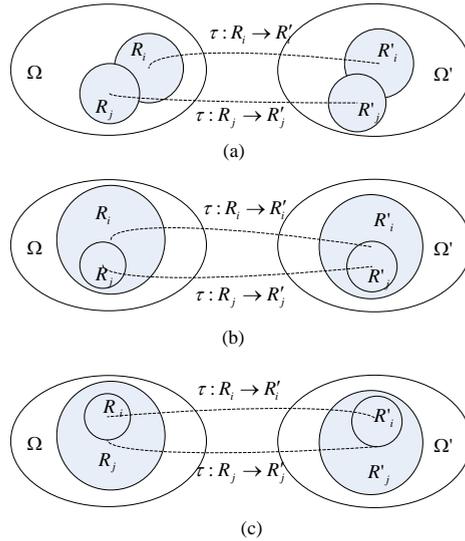


Fig. 1 Three kinds of connected wide baseline match pairs.

Although there are many methods to extract interest regions and compute descriptors, we concentrate on the SIFT method in Ref. [7], because it had the best performance. This method combines a scale invariant region detector and a descriptor called “scale invariant feature transform” (SIFT). A SIFT feature comprises four components: keypoint location p , scale σ , orientation o and descriptor f , which represents a circle patch of the image with center p and radius $r = c\sigma$. It’s easy to determine the relationship between two SIFT regions R_1 and R_2 by considering them as circles:

- (i) If $\|p_1 - p_2\|_2 > c(\sigma_1 + \sigma_2)$, then $r(R_1, R_2) = a$.
- (ii) If $\|p_1 - p_2\|_2 > c|\sigma_1 - \sigma_2|$ and $\|p_1 - p_2\|_2 \leq c(\sigma_1 + \sigma_2)$, then $r(R_1, R_2) = b$.
- (iii) If $\|p_1 - p_2\|_2 \leq c|\sigma_1 - \sigma_2|$ and $\sigma_1 > \sigma_2$, then $r(R_1, R_2) = c$.

(iv) If $\|p_1 - p_2\|_2 \leq c|\sigma_1 - \sigma_2|$ and $\sigma_1 < \sigma_2$, then $r(R_1, R_2) = d$.

Where p_i , σ_i are the location and the scale of the SIFT region R_i , $i=1,2$, and $r(\dots)$ is the relationship function for two sets, that is, if R_i and R_j are disjoint then $r(R_i, R_j) = a$ and so on.

This topological clustering algorithm has two parameters n_t and c . We set $n_t=3$ i.e. the small clusters with less than 3 matches are eliminated, and c corresponds the radius of the support region from which the SIFT descriptor is computed. Fig.2 illustrates the topological clustering results for the image pair ‘‘Mex’’ with the value of the parameter $c=3$. In part (a) we can see that the initial matches are messily distributed, while in part (b) we can see that the initial matches are regularly organized to form a set of clusters.

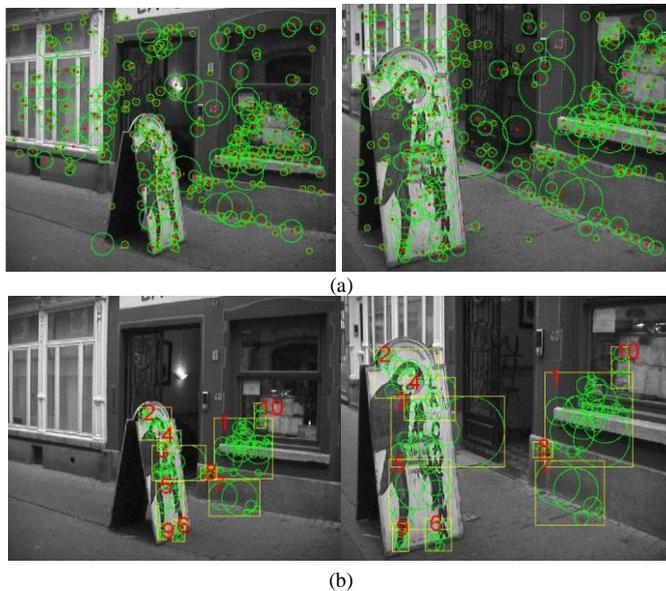


Fig 2. The topological clustering results for the image pair ‘‘Mex’’ with the value of the parameter $c=3$: (a) the initial SIFT matches; (b) the match clusters with no less than 3 matches, each cluster is bounded by a rectangle

3. Estimating the Fundamental Matrix

As shown in [6], the topological clustering is crucial to improve the correct-ratio of matches. Some correct matches are lost, but many more mismatches are removed. At the same time, the rest matches, including an arbitrary amount of mismatches, can be partitioned into match clusters. A cluster is a set of aggregated matches between two images (figure 2).

Based on our previous work, we develop a novel method called Clusters Pairing Consensus (CPC) for estimating the fundamental matrix. In this method, region matches are no longer seen as atomic unit, instead, match clusters are used to generate a series of fundamental matrix estimates. Then, clusters enjoy two fundamental properties. First, the matches in a cluster are most often all correct, or all incorrect. Secondly, as shown as fig3, a set of region matches in a cluster is more likely to lying on a same surface in scene. The two properties are obvious. Mismatches have independent, random affine transformations, uniformly spread in the large 6D affine transformation space. So it’s unlikely to happen that a mismatch satisfied the topological constraint with other correct matches, because the mismatch has little chances to offer a suitable affine transformation. On the other hand, in the wide-baseline stereo context, a set of correct matches respecting

topological constraint means they satisfied the similar affine transformations and suggested they were lying on the same surface.

Moreover, two planes are sufficient to determine the fundamental matrix F and avoid the degeneracy [8]. It means that we can use a pair of clusters to get a parameter estimate F . Another valuable assumption is that the inlier distribution can be represented as a Gaussian model with a small standard deviation[8]. Then, a simple, redescending M-estimator can be used to improve the accuracy of the fundamental matrix. The M-Estimators are based on solving the following expression:

$$\min_F \sum_i w_i (m_i^T F m_i')^2 \quad (2)$$

In which ω_i is a weight function. In our algorithm, consider r_i the residual of $m_i^T F m_i'$, the samplest weight function can be used with threshold T :

$$\omega_i = \begin{cases} 1 & |r_i| \leq T \\ 0 & |r_i| > T \end{cases} \quad (3)$$

Moreover, the computational complexity of a problem can be reduced, because there are considerably fewer relevant clusters than region matches. So we can test all pair of clusters to get a determinate solution, instead of random sampling.

The flow of the algorithm is summarized in Algorithm 1, and experimental result in following section will show that the solution produced by our method is good and valid for each experimental image pair and each threshold.

1. Clusters $\{C_i\}$ are obtained from the initial SIFT matches by using topological clustering algorithm four times with different parameters $c=3,4,5,6$ and $n_i=3$.
2. For each valid pair of clusters $\langle C_i, C_j \rangle$ i.e. $C_i \cup C_j$ contains 8 matches at least:
 - (a) Estimate the fundamental matrix F with the matches belong to $C_i \cup C_j$ using the normalized 8-point algorithm in [8].
 - (b) Calculate the distance measure d_i for each match.
 - (c) Use the sample M-estimate described in this section and inliers are scored.
3. Choose the F with the largest number of inliers. In the case of ties choose the F that has the lowest standard deviation of inliers.

Algorithm 1. The CPC algorithm for estimating the fundamental matrix

4. Experiment and Result

The proposed algorithm is tested on six famous wide baseline image pairs(see Fig.3), which are experimental image pairs in[9] and are presented in figure3. The putative correspondences were generated using the SIFT keypoint detector. We have compared our CPC algorithm with RANSAC algorithm. The version of RANSAC came from <http://www.csse.uwa.edu.au/~pk/Research/MatlabFns/index.html>.

In order to compare these algorithms, we compute the average distance dt of the correct matches against each estimate F . Obviously, lower dt implies that the fundamental matrix F fits all the correct matches better. For evaluate these algorithms, the true correspondances were selected by manual inspection to obtain the ground truth and presented in Table1 and. The ground truth also had been used in in (Wang et al., 2008).

We tested these algorithms for each image pair with thresholds $t=2$ pixel and the result was shown on Table 1. Because the solution of RANSAC is not determinate, we execute RANSAC has been applied to each image pair 100 times. For each algorithm the following statistics are presented: the number of inliers that have been found with the threshold t , the average distance d_t of all inliners, the run time until the termination of the algorithm. Specially, if the the average distance d_t of all inliners more than 5, we can consider that a bad estimate has occured. For RANSAC, these statistics are the average value of all performance.

From Table 1, CPC generally gave a better fit than RANSAC. Especially, for all the six image pairs the RANSAC algorithm needs more than one hundred seconds. But the CPC algorithm, need only less than one second. Therefore, the computational complexity greatly decreases.

By inspection it can be seen that degeneracies in the fundamental matrix occur frequently with RANSAC. Fig. 4 show two epipolar geometries with the inliers for the image pair “oase”. Part (a) corresponds to a “bad” fundamental matrix estimate obtained using RANSAC, while part (b) corresponds to the fundamental matrix estimate produced by our method CPC. We can observe that all the inliers (black points) in part (a) are on the wall of the building i.e. they are coplanar, and this is the very case of structure degeneracy. However, some inliers in part (b) are on the cover of the tent so the case of structure degeneracy is avoided.



Fig. 3 Experimental image pairs: “church”, “tree”, “mex”, “wash”, “oase”, “auto”.

Table 1 Experiment characteristics and results. T is the number of inlier correspondences. For each algorithm “In.” is the number of inliers that were found and md is the average distance d_t of all inliners.

	T	CPC			RANSAC		
		In.	md	time	In.	md	time
“Church”	101	81	1.0219	0.8125	76.6	1.093	115.83
“Tree”	146	128	0.7724	1.5469	111.3	1.097	178.2
“Mex”	122	93	1.0616	0.7813	89.5	1.096	136.63
“Wash”	112	97	0.8864	0.5469	75.2	2.444	170.09
“Oase”	101	87	0.7456	1.2813	63.8	7.41	225.99
auto	61	44	1.0141	0.3594	26.2	7.961	273.47

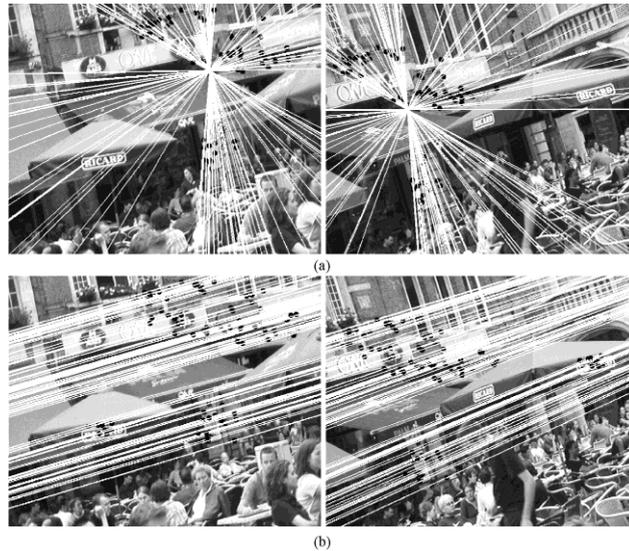


Fig. 4 Two epipolar geometries with the inliers for the image pair “oase”. Part (a) corresponds to a “bad” estimate obtained using RANSAC, while part (b) corresponds to the estimate generated by our method CPC.

5. Conclusions

Within this paper, we develop a novel method for estimating the fundamental matrix for the wide-baseline image pair. The essential of this method is that pairs of clusters instead of match samples are used to generate a series of fundamental matrix hypotheses, so we call this method Clusters Pairing Consensus (CPC). Clusters of matches $\{C_i\}$ are obtained from the initial SIFT matches using our topological clustering algorithm, and the estimate with the largest number of inliers is selected as the solution. The proposed algorithm overcomes the drawbacks of random sampling algorithms such as RANSAC. Experiment result shows that our method can get a good and valid solution for each experimental wide baseline image pair, which demonstrates that our method is more robust than RANSAC.

References

- [1] P.H.S. Torr, and D.W. Murray, “The development and comparison of robust methods for estimating the fundamental matrix,” *International Journal of Computer Vision*, vol. 24, no. 3, pp. 271–300, 1997.
- [2] M. A. Fischler, and R. C. Bolles, “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,” *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [3] P. H. S. Torr, and A. Zisserman, “MLESAC: a new robust estimator with application to estimating image geometry,” *Computer Vision and Image Understanding*, vol. 78, no. 1, pp. 138–156, 2000.
- [4] H. Wang and D. Suter, “Robust adaptive-scale parametric model estimation for computer vision,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 26, no. 11, pp. 1459–1474, 2004.
- [5] L. Moisan, and B. Stival, “A probabilistic criterion to detect rigid point matches between two images and estimate the fundamental matrix,” *International Journal of Computer Vision*, vol. 57, no. 3, pp. 201–218, 2004.
- [6] Y. T. Wang, D. Z. Zhang, J. W. Tian, “Topological clustering and its application for discarding wide-baseline mismatches,” *Optical Engineering*, vol. 47, no. 5, pp. 057202-1- 057202-6, 2008.

- [7] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 2, no. 60, pp 91-110, 2004.
- [8] R. Hartley , A. Zisserman,. *Multiple View Geometry in Computer Vision*, 2nd ed Cambridge University Press, 2000.
- [9] T. Tuytelaars, and L. V. Gool, "Matching widely separated views based on affine invariant regions," *International Journal of Computer Vision*, vol. 59, no. 1, pp. 61-85, 2004.