Fast Outlier Rejection by Using Parallax-Based Rigidity Constraint for Epipolar Geometry Estimation *

Engin Tola[1] and A. Aydın Alatan[2]

[1] Computer Vision Laboratory, Ecóle Polytechnique Fédéral de Lausanne (EPFL)
CH–1015, Lausanne, Switzerland
engin.tola@epfl.ch
WWW Homepage: http://cvlab.epfl.ch/∼tola
[2] Dept. of Electrical & Electronics Eng. Middle East Technical University (METU)
TR–06531, Ankara, Turkey
WWW Homepage: http://www.eee.metu.edu.tr/∼alatan/
alatan@eee.metu.edu.tr

Abstract. A novel approach is presented in order to reject correspondence outliers between frames using the parallax-based rigidity constraint for epipolar geometry estimation. In this approach, the invariance of 3-D relative projective structure of a stationary scene over different views is exploited to eliminate outliers, mostly due to independently moving objects of a typical scene. The proposed approach is compared against a well-known RANSAC-based algorithm by the help of a test-bed. The results showed that the speed-up, gained by utilization of the proposed technique as a preprocessing step before RANSAC-based approach, decreases the execution time of the overall outlier rejection, significantly.
Key words: Outlier removal, Parallax-based rigidity constraint, RANSAC

1 Introduction

Epipolar geometry computation is a fundamental problem in computer vision. Most of the state-of-the-art algorithms use a statistical iterative algorithm, Random Sample Consensus(RANSAC) [3], in order to select the set of correspondences between frames, required for determining the epipolar geometry. This simple and powerful approach is practically a brute force model estimation algorithm. The random samples are selected from the input data and model parameters are estimated from these subsets, iteratively [3]. The subset size is chosen to allow the estimation of the model with minimum number of elements, while this model is tested with the whole input data at each iteration and at the end, the one with the largest consensus set is selected as the output. The iterations are usually stopped, when determination of a better model has statistically a very low probability. Although the results of RANSAC are quite acceptable, it takes considerable amount of time to find this result due to

* This research has been partially funded by EC IST 3DTV NoE

the iterative nature of the algorithm, when the input data contamination level is high. Due to these reasons, several schemes have been proposed to accelerate RANSAC. In [10], the computation time is reduced by evaluating the model only on a sample set of the inliers, as an initial test and the models passing this test are evaluated over the whole data set. In a different approach [2], a local optimization is applied in order to find a better estimate around the current guess of the model, stemming from the fact that a model estimated from an outlier-free subset is generally quite close to the optimal solution. The method in [4] uses weak motion models that approximate the motion of correspondences and by using these models the probability of correspondences being inliers or outliers are estimated. These probability values are then used to guide the search process of the RANSAC. Another type of approach is to reduce the ratio of outliers to inliers by eliminating some of the outliers before using iterative procedures. In [1], the possibility of rotating one of the images to achieve some common behavior of the inliers is utilized in order to speed up the process. However, it requires camera internal parameters to be known a priori. In this paper, a non-iterative algorithm to reduce the ratio of outliers to inliers without any knowledge of the camera geometry or internal parameters is proposed. This algorithm is intended to be used as a post-processing step after any point matching algorithm and is tested for scenes containing independently moving objects.

2 Proposed Algorithm

Typical scenes consist of independently moving objects (IMO) as well as a stationary background. In order to extract 3-D structure of such a complex scene from multi-views, the first subgoal is to determine that of the stationary background. Hence, after finding correspondences between views of the scene, the resulting data should be examined to discriminate between correspondences due to the background and moving objects, as well as the outliers. It is possible to differentiate between the stationary background motion vectors and the remaining ones (whether they are outliers or they belong to IMOs) by using a constrained, called parallax-based rigidity constraint (PBRC) [9]. PBRC is first proposed for segmenting IMOs in environments, containing some parallax, via 2-D optical flow [9]. The utilization of PBRC strictly requires the knowledge of at least one vector that belongs to the background. However, the method in [9] does not suggest any automatic selection mechanism for such a purpose. In this paper, initially, an automatic background vector selection algorithm is proposed. Based on this seed vector, a robust outlier rejection technique is developed via PBRC. The robustness of the method is guaranteed by using more than one vector from the background to calculate the PBRC scores of the motion vectors. These supporting vectors are also determined from the seed background vector.

2.1 Parallax-Based Rigidity Constraint (PBRC)

PBRC primarily depends on the decomposition of translational (with non-planar structure) and rotational (with planar structure) components of 2-D displacements between frames. This decomposition is achieved by removal of rotation+planar effects through affine model fitting to the displacements between two frames; hence, the remaining components are due to parallax effects. The relative 3D projective structure of two points, p_1 and p_2, on the same image is defined as the following ratio [9]:

$$\frac{\overline{\mu}_2^T(\Delta\overline{p}_w)_\perp}{\overline{\mu}_1^T(\Delta\overline{p}_w)_\perp} \tag{1}$$

where, $\overline{\mu}_1$, $\overline{\mu}_2$ are the parallax displacement vectors of these two points between two frames. In Equation 1, $\Delta\overline{p}_w = \overline{p}_{w2} - \overline{p}_{w1}$, where $\overline{p}_{w1} = p_1 + \overline{\mu}_1$ and $\overline{p}_{w2} = p_2 + \overline{\mu}_2$ ($v_\perp$ denotes a vector perpendicular to v). It has been proven that relative 3D projective structure of a pair of points does not change with respect to the camera motion [9]. Hence, between different views, PBRC is formally defined as:

$$\frac{\overline{\mu}_2^{jT}(\Delta\overline{p}_w)_\perp^j}{\overline{\mu}_1^{jT}(\Delta\overline{p}_w)_\perp^j} - \frac{\overline{\mu}_2^{kT}(\Delta\overline{p}_w)_\perp^k}{\overline{\mu}_1^{kT}(\Delta\overline{p}_w)_\perp^k} = 0 \tag{2}$$

where $\overline{\mu}_1^k$, $\overline{\mu}_2^k$ are the parallax vectors between the reference frame and k^{th} frame, and $(\Delta\overline{p}_w)^j$, $(\Delta\overline{p}_w)^k$ are the corresponding distances between the warped points. By using this constraint, it is possible to discriminate between the background and foreground vectors in three frames, by the help of a motion vector, which belongs to the background.

2.2 Automatic Background Seed Selection Algorithm

Background seed selection is a critical step for eliminating IMO and outlier contributions from the correspondence set. PBRC can be utilized for this purpose, since it puts an explicit constraint on the 3-D structure of all stationary background points. PBRC, although, forces the change in the relative 3-D structure to remain zero, this constraint does not always hold due to noise. Therefore, as a simple method, only choosing a random vector and counting the number of vectors that obey this exact constraint, should not solve the problem of the background vector selection. Moreover, the errors in the parallax-based rigidity constraint differ, when one changes the *support vector*(background vector) of the constraint ($\overline{\mu}_1$ in Equation 2). Therefore, simple thresholding will also not be the solution to this problem, since such a threshold should also be adapted for different scenes.

The proposed novel solution to this problem can be explained as follows: N different random vectors are chosen as *candidate support vectors* and the number of vectors, which are outside a certain neighborhood around one of these N candidate vectors, that obey the rigidity constraint within a small threshold, are counted. After testing all candidate vectors in this manner, the vector yielding the maximum number of *supports*, is chosen as the *background seed*.

For robustness, the candidate vectors are selected according to the magnitude of the residuals (distance of the points found by plane registration to their correct locations). The magnitude range of the residual vectors is divided into N equal intervals and a support vector is selected for every interval. This selection method is adopted due to the fact that the plane registration step usually leaves behind vectors with small residuals from the dominant plane. Therefore, the vectors on this dominant plane should not be selected, since their small norm is due to noise. On the other hand, the vectors with large residuals are not reliable, since they might be outliers. Hence, in order to cover the whole range of vectors, the above procedure is proposed.

Another important aspect of the proposed selection criteria is the elimination of the vectors within the neighborhood of the candidate support vector, while calculating the number of vectors that obey the rigidity constraint. In this manner, it is possible to eliminate some points, belonging to an IMO, which should mostly has its support vectors within its neighborhood. If this constraint is not used, one might find the change in the rigidity constraint still a small number and erroneously declare an IMO point as a background seed, while unfortunately, most of the vectors in the supporting set are belonging to the IMO itself. On the other hand, this constraint reduces the number of the consistent vectors to an IMO-belonging candidate vector. This situation is not a problem for the background vectors, since they are not confined (i.e. localized) to a single region.

2.3 Application of PBRC by Selected Background Seed

At this stage, all the correspondence vectors are tested by using PBRC with the previously selected background seed pixel. In order to increase the robustness of the algorithm, more than one representative background pixel can be used to discriminate between background and other vectors. In this scenario, a vector is decided to belong to a background point, if, out of M different background supports, it is within the first p-percent of the sorted cost, which is calculated according to Equation 2 at least K times. ($K < M$ and K is larger than some threshold). Hence, the following algorithm is obtained for rejecting IMO contributions, as well as any kind of outliers, in the correspondence set.

Algorithm

1. Apply plane registration to the motion vectors between the first two frames as well as the second and third frames and determine residual motion vectors. Dominant plane estimation and registration is accomplished by the use of a RANSAC based method to have robustness (Since this problem has a small number of freedom, it is quite fast.)
2. Find the *background seed* as explained in Section 2.2
   (a) Sort the residual motion vectors according to their norms.
   (b) Choose N *candidate support vectors* with equal distance from each other in terms of their norm values.
   (c) Calculate the number of vectors that obey PBRC within threshold T for each of the candidate vectors. Do not consider vectors within d distance to the candidate vector.

(d) Choose the maximally supported vector as the background seed.
3. Select M vectors yielding the smallest error with the background seed and calculate the PBRC errors of the rest of the vectors with respect to each of these *support vectors*.
4. Sort the elements of these sets according to their errors and select the vectors that are within the first p-percent of the sets.
5. Choose the vectors that are selected more than K times ($K < M$) as background pixels and discard the rest.

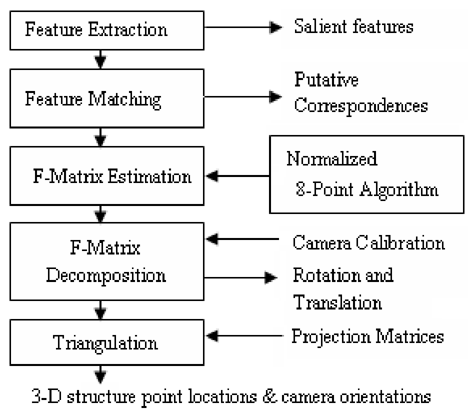

Fig. 1. Test Bed Flow Chart

3 System Overview

A test-bed is prepared in order to compare the operation speeds of three different algorithms. In fact, this test-bed is a standard structure from motion algorithm, in which 3-D structure points and motion parameters are estimated (see Figure 1). It is assumed that the camera intrinsic parameters are known a priori. This assumption is necessary only for the triangulation stage, but not in the outlier rejection step.
The process starts first by finding point matches between two images. In order to match points for different images, it is necessary to extract salient features from these images. For this purpose a modified Harris corner detector [5] is used. The modification is such that the extracted features are determined in sub-pixel resolution. This is achieved by bi-quadric polynomial fitting [12]. After the extraction of salient features, a moderately simple algorithm (in terms of computational complexity) is used in order to match the features. The matching is performed by examining two main criterions: normalized cross correlation (NCC) and neighboring constraint (NC) [13]. NCC is used to measure the similarity of image patches around the feature positions and NC is used to introduce smoothness to motion vectors by neighborhood information.

Once a set of correspondences are obtained, the next step is the estimation of the fundamental matrix, *robustly*. In this step, in order to introduce robustness, 3 different algorithms are tested in terms of their performances: fast outlier rejection algorithm proposed in this paper (denoted as IMOR), a RANSAC based iterative solution [12] and the concatenation of these two algorithms. The estimation of the fundamental matrix is performed by using the Normalized 8-Point algorithm proposed in [6] for all of these methods. The estimated F-matrix is then refined by using non-linear minimization, namely Levenberg-Marquardt, with Sampson Error [8] as its quality metric. For visual inspection of the results, 3-D structure points are also estimated by using the computed fundamental matrix. In order to achieve this aim, the essential matrix is computed by using the camera calibration information and the computed fundamental matrix. Then, this essential matrix is decomposed into its rotation and translation components [11]. This way, projection matrices for the views are acquired and using the triangulation algorithm proposed in [7], 3-D point locations are determined.

4 Results

In this section, comparison of the proposed algorithm, IMOR, and a robust method, based on RANSAC [12], is presented. In the implementation of the IMOR algorithm N (the number of candidate support vectors) is chosen as 10, d (the distance of the vectors to the candidate) is chosen as 30 pixels, PBRC threshold T is chosen as $1e^{-4}$, M is 5, K is 3 and p is 70%. The value for d depends highly on the input data. For large images and large IMOs d must be set higher. M and K are purely for robustness purposes and they may be set to 1 if the input vector set is known to have low outlier to inlier ratio.

The results, which are summarized in Table 1, are performed over different data sets. The data set (7 different image triplets) contains real and synthetic outdoor scenes. The presented results are obtained by simple averaging. "Wrong Rejections" column in the table refers to the number of true inliers that are labeled as outliers by the algorithm whereas "Inlier Number" column refers to the number of correspondences, algorithms declare as inliers. As it can be observed from these results, IMOR algorithm gives comparable results with the algorithm based on RANSAC, although it cannot always eliminate all of the outliers. However, IMOR is clearly advantageous compared to RANSAC, due to its shorter execution time. It should be noted that RANSAC is iterative and its number of iterations is not fixed, whereas IMOR is a single step approach. Hence, it is possible to utilize IMOR before RANSAC to eliminate most of the outliers and then use this powerful iterative algorithm to refine the results. In this manner, with a small number of iterations, a comparable reconstruction quality may be achieved in less time. The results for this approach are presented in the third row of Table 1. Some typical results for an image triplet is also shown in Figure 2 with motion vector elimination, as well as 3-D reconstruction results.

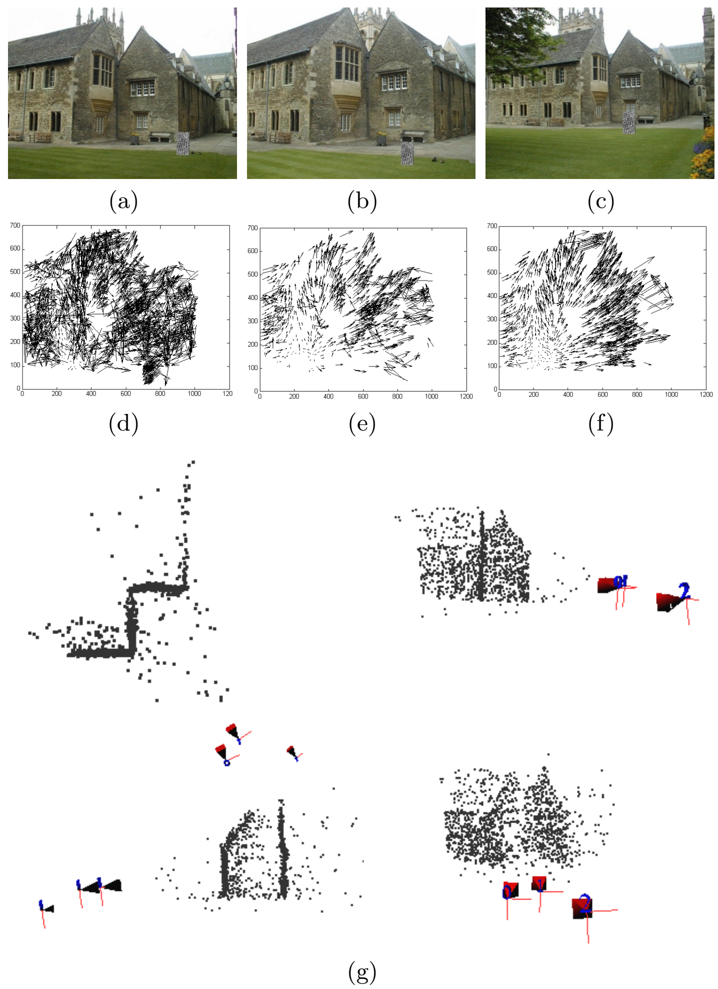

Fig. 2. Test results: (a-c) input images, (d) computed motion vectors. Results of (e) proposed IMOR algorithm & (f) RANSAC-based algorithm. (g) Reconstruction results of the cascaded system for various camera locations.

Table 1. Matching performance and execution time comparisons: RANSAC-based approach vs. the proposed algorithm (IMOR) and cascaded application of these two algorithms

	Iter No.	Time (ms)	Wrong Reject	Missed Outlier	Inlier	Total Vector#
RANSAC	1626	4968	11	3	971	1651
IMOR	-	31	156	33	856	1651
IMOR+RANSAC	21	112	158	1	824	1651

5 Conclusions

It can also be inferred from Table 1 that the IMOR algorithm cannot detect significant amount of outliers, and therefore, the fundamental matrix estimate, computed by using this contaminated set, will give inferior results. As expected, the reconstruction by only using IMOR algorithm has been unacceptable during the performed tests. Although, the results of RANSAC alone yields very accurate reconstruction results, utilization of IMOR as a preprocessing step before RANSAC decreases the execution time of the overall outlier rejection algorithm considerably, approximately 40 times (averaged over all image triplets). Therefore, it is proposed to jointly utilize the outlier rejection algorithms in a cascaded manner (IMOR+RANSAC). This combination yields significant improvement for the execution time without losing from 3-D reconstruction quality.

References

1. A. Adam, E. Rivlin, and I. Shimshoni. Ror: Rejection of outliers by rotations. In *IEEE Trans. PAMI*, Jan. 2001.
2. O. Chum, J. Matas, and J. V. Kittler. Locally optimized ransac. In *Proc. of the 25th DAGM Symp.*, volume 2, pages 236–243, Sep. 2003.
3. 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*, 24(6), 1981.
4. L. Goshen and I. Shimshoni. Guided sampling via weak motion models and outlier sample generation for epipolar geometry estimation. In *Proceedings of the CVPR*, 2005.
5. C. Harris and M. A. Stephens. Combined corner and edge detector. In *In Proc. 4th Alvey Vision Conference*, pages 147–151, 1988.
6. R. I. Hartley. In defense of the eight-point algorithm. *IEEE Transactions on PAMI*, 19(6), June 1997.
7. R. I. Hartley and P. Sturm. Triangulation. *IEEE Trans. on PAMI*, 1998.
8. R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004.
9. M. Irani and P. Anandan. A unified approach to moving object detection in 2d and 3d scenes. *IEEE Trans. on PAMI*, June 1998.
10. J. Matas and O. Chum. Randomized ransac with td,d test. British Machine Vision Conference, Sept. 2002.
11. M. Tekalp. *Digital Video Processing*. Prentice Hall, 1995.
12. E. Tola. Multiview 3d reconstruction of a scene containing independently moving objects. Master's thesis, METU, August 2005.
13. Z. Zhang, R. Deriche, O. Faugeras, and Q. T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. Technical Report 2273, INRIA, 1994.