A ROBUST METHOD FOR STEREO VISUAL ODOMETRY BASED ON MULTIPLE EUCLIDEAN DISTANCE CONSTRAINT AND RANSAC ALGORITHM

Qi Zhou[a], Xiaohua Tong [a*], Shijie Liu [a], Xiaojun Lu [b], Sicong Liu [a], Peng Chen[a],Yanming Jin[a], Huan Xie[a]

[a] College of Surveying and Geo-Informatics, Tongji University, Shanghai, China – xhtong@tongji.edu.cn
[b] China International Engineering Consulting Corporation, Beijing, China

Commission III, WG III/2

KEY WORDS: Visual Odometry, Stereo Vision, Robot Navigation, RANSAC Algorithm

ABSTRACT:

Visual Odometry (VO) is a critical component for planetary robot navigation and safety. It estimates the ego-motion using stereo images frame by frame. Feature points extraction and matching is one of the key steps for robotic motion estimation which largely influences the precision and robustness. In this work, we choose the Oriented FAST and Rotated BRIEF (ORB) features by considering both accuracy and speed issues. For more robustness in challenging environment e.g., rough terrain or planetary surface, this paper presents a robust outliers elimination method based on Euclidean Distance Constraint (EDC) and Random Sample Consensus (RANSAC) algorithm. In the matching process, a set of ORB feature points are extracted from the current left and right synchronous images and the Brute Force (BF) matcher is used to find the correspondences between the two images for the Space Intersection. Then the EDC and RANSAC algorithms are carried out to eliminate mismatches whose distances are beyond a predefined threshold. Similarly, when the left image of the next time matches the feature points with the current left images, the EDC and RANSAC are iteratively performed. After the above mentioned, there are exceptional remaining mismatched points in some cases, for which the third time RANSAC is applied to eliminate the effects of those outliers in the estimation of the ego-motion parameters (Interior Orientation and Exterior Orientation). The proposed approach has been tested on a real-world vehicle dataset and the result benefits from its high robustness.

1. INTRODUCTION

Autonomous navigation is quite significant for many robotic applications such as planetary exploration and auto drive. For these robotic applications, Visual Odometry is the critical method for relative locating, especially in GPS-denied environments. VO estimates the ego-motion of robot using a single or stereo cameras, which is more accurate than the conventional wheel odometry according to Maimone et al. (2007a).

VO is a specific application of Structure From Motion (SFM), which contains the camera pose estimation and 3D scene point reconstruction according to Scaramuzza et al. (2011). Simultaneously, VO differs from the SLAM (Simultaneous Localization And Mapping), which contains the mapping process and loop closure (Engel et al., 2015; Pire et al., 2015).

In the last few decades, VO has been divided into two kinds of method, monocular and stereo cameras. For monocular VO, the main issue is solving the scale ambiguity problem. Some researchers set the translation scale between the two consecutive frames to a predefined value. Ke and Kanade (2003) virtually rotate the camera to the bottom-view pose, which eliminates the ambiguity between the rotation and translation and improves the motion estimation process. On the other side, some researchers assume that the environment around the monocular camera is flat ground and the monocular camera is equipped on a fixed height with fixed depression angle, like the situation in Bajpai et al. (2016a). According to Bajpai et al. (2016a), the advantage of monocular VO method is smaller computational cost compared to the stereo VO, which is quite important for those real-time embedded applications.

For large robotic platforms with strong computational ability like automatic drive platforms and future planetary exploration robots, the stereo cameras perform superior to the monocular one. Because of the certain baseline between the left and right camera, the ambiguity scale problem does not exist in stereo VO. And the Stereo VO can estimate the 6-Degree of Freedom (DOF) ego-motion no matter what kinds of environment the system works in. Currently, there are two kinds of stereo VO methods, 3D-3D method and 3D-2D method.

In both kinds of stereo VO method, feature detection and matching great influence both the accuracy and speed issues. In feature point matching field, there are many feature point detectors and descriptors having been presented in last twenty years. Scale-Invariant Feature Transform (SIFT) invented by Lowe (1999) is the most famous one because of its excellent detecting accuracy and robustness. Bay et al. (2006) presents the Speeded Up Robust Feature (SURF), which is an improved version of SIFT. It uses Haar wavelet to approximate the gradient method in SIFT, using integral image technology at the same time to calculate fast. In most cases, its performance can reach the same level precision compared to SIFT, with 3-7 times faster. For those cases with very fast speed issue, Oriented FAST and Rotated BRIEF (ORB) is employed by Rublee et al. (2011).

* Corresponding author

3D-3D method treats the stereo cameras as the point cloud generator, which make use the stereo cameras to generate 3D point cloud and estimates the rotation and translation between two consecutive frames using 3D point cloud registration method like Iterative Closest Point (ICP) algorithm in Balazadegan et al. (2016a). ICP algorithm can only converge to the local minimum, which differs from VO propose. Therefore, we must obtain a good initial value for the VO motion estimation parameters according to Hong et al. (2015a).

On the other side, the aim of the 3D-2D VO method is to solve the Perspective-n-Point (PnP) problem. According to Scaramuzza et al. (2011), 3D-2D method is more accurate than the 3D-3D method, therefore 3D-2D method has received attention in both Photogrammetry like McGlove et al. (2004a) and Computer Vision like Hartley and Zisserman (2000a). The least feature points needed in PnP problem are 3, which called P3P problem. Gao et al. (2003a) presents a solution to the 3point algorithm for P3P problem. For n>3 points, some more accurate but slower methods based on iteration exist presented by Quan and Lan (1999a) and Ansar and Daniilidis (2003a).

In 3D-2D method, the outlier elimination is quite important because the precision of feature matching impacts highly the result of motion estimation. Kitt et al. (2010) presented an outlier rejection technique combined with the RANSAC (Random Sample Consensus). Talukder et al. (2004) initially estimate the motion by all the matching points, then eliminate outliers using the initial motion parameters, with the iterative least mean-square error estimation at last. Fanfani et al. (2016a) combines the tracking and matching process with the key frame selection, i.e. based on motion structure analysis to improve the system robustness. There are some researches presenting their robust method in specific scenes (Musleh et al., 2012a; Min et al., 2016a) or using other sensors (He et al., 2015a). Musleh et al. (2012a) presented an inliers selection scheme in urban situations that only the feature points on the ground will be chosen to estimate the motion. He et al. (2015a) fused an inertial sensor to estimate the rotation of the robot, which can compare to the measurement of VO and eliminate outliers or feature points on dynamic objects.

2. PROPOSED METHOD

2.1 Overview of the Proposed Method

In this work, the VO method we choose is 3D-2D method because of its precision. Firstly, the feature points are detected by the chosen feature detector both in the left and right image. The feature point detector is SURF and the descriptor employing ORB, considering the accuracy and speed issue as a trade-off. After getting the correspondences between the two images, BF matcher is performed for the initial mismatches elimination. Then the EDC and RANSAC algorithms are carried out to eliminate mismatches whose distances are beyond a predefined threshold (the average of all feature points). Similarly, when the left image of the next time matches the feature points with the current left image, the EDC and RANSAC are iteratively performed. The depth information of the feature points is obtained by the Space Intersection or the Triangulation after the first matching process. Finally, we employ the EPnP method provide by Moreno-Noguer et al. (2007) to solve the PnP problem, which is non-iterative and lower computational complexity but almost accurate compared to the iterative ones, with the RANSAC algorithm.

In generally, we employ three times outlier elimination scheme. The RANSAC algorithms in the first and second time eliminate outliers which offset the epipolar line beyond 1 pixel. The EDC scheme drops the other outliers which are close to the epipolar line but far Euclidean Distance calculated by the feature descriptor. The third time RANSAC algorithm estimate the motion parameters, which differs from the effect of the first and the second time RANSAC. The procedure is shown in Figure 1.

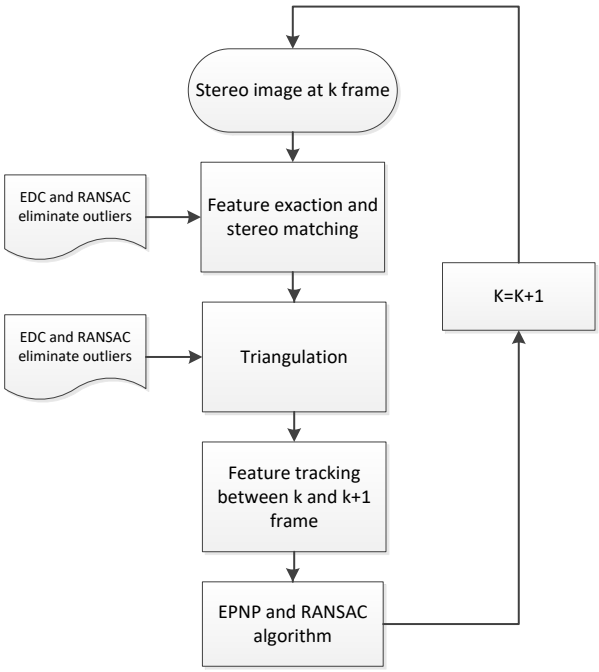

Figure 1. Overview of the proposed method

2.2 Euclidean Distance Constraint

To compare the correspondence degree of each pair tie points, the Euclidean Distance between every pair matched points is calculated, defined in Equation (1):

$$ED=\sqrt{\sum_{i=1}^{n}(x_{left}^{i}-x_{right}^{i})^{2}} \quad (1)$$

Where: n = dimensions of the ORB descriptor
x^{i}_{left} = ith dimension of the descriptor of the tie point in the left image
x^{i}_{right} = ith dimension of the descriptor of the tie point in the right image

When all EDs of the tie points have been calculated by the Equation (1), computing the average value of all ED following the Equation (2):

$$ED_{a}=\frac{\sum_{j=1}^{m}ED_{j}}{m} \quad (2)$$

Where: ED_a = the average value of all ED
m = the number of the tie points

In this work, we consider all the correspondence points whose ED is beyond ED_a as the outliers. There is an example shown in Figure 2.

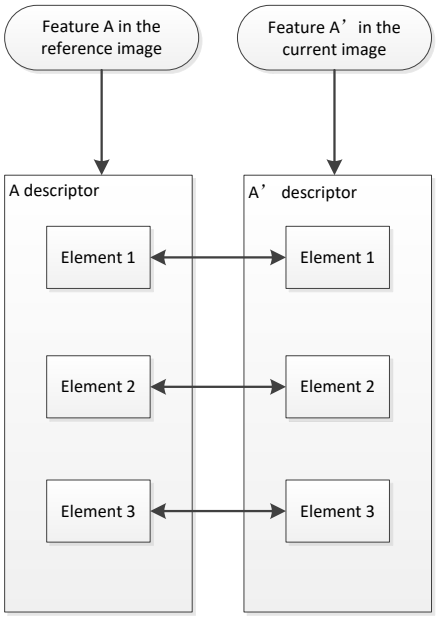

Figure 2. the process of EDC calculation

2.3 RANSAC Algorithm

There are two main impacts of RANSAC algorithm in our work, assisting the EDC to eliminate the outliers and removing the effect of the possible remaining outliers in the PnP process. RANSAC algorithm has been proven that it is an effective tool to extract the optimal subset from a huge data set with some errors. Firstly, three points have been chosen randomly for the initial ego-motion parameters estimation and all tie points calculate the reprojection error according to the ego-motion parameters. If the reprojection error of the tie point is under the threshold, which in this work is 3 pixels, it will be seen as an inlier. When the inliers are beyond 90% of all the tie points, the model will be accepted and all inliers recalculate the ego-motion parameters using EPnP method. The process is shown in Figure 3.

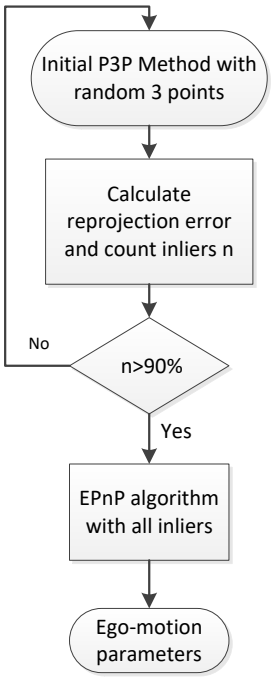

Figure 3. RANSAC process

3. EXPERIMENT AND RESULT

3.1 Experiment Data

In this section, we evaluate the proposed method compared to the VO method only with RANSAC algorithm. We use the publicly available KITTI Benchmark Suite provided by Geiger et al. (2013a), which includes stereo image sequences on an autonomous vehicle platform.

The grayscale stereo cameras equipped on the platform are Point Grey Flea 2(FL2-14S3M-C) with a baseline of 53cm and 1226×370 pixels. All KITTI cameras were synchronized at 10Hz and carefully calibrated.

For the position ground truth, KITTI provides RTK GPS/IMU ground truth with open sky localization errors about 5cm.

In order to evaluate the methods from the perspective of both robustness and precision, we use Tracking Success Proportion (TSP) and Average Distance Error(ADE), calculated as follow Equations (3) ~ (4):

$$ADE = \frac{\sum_{i=1}^{N} \sqrt{(x_i - x_i^{GPS})^2 + (y_i - y_i^{GPS})^2}}{N} \quad (3)$$

Where: N = the number of image frames
x_i, y_i = ith frame position evaluated by the algorithm
x_i^{GPS}, y_i^{GPS} = ground truth at the ith frame.

$$TSP = \frac{n_s}{n_{total}} \quad (4)$$

Where: n_s = the number of successful frames
n_{total} = the number of all frames

The experiment we using is the synchronized dataset "2011_10_03_drive_0027", whose duration is about 445 s and the whole trajectory length is about 3.5km.

In the experiment, the average speed is about 7m/s, the maximum speed is under 10m/s, so we consider the VO result as failure if the translation between two consecutive frames beyond 1.5m (about 15m/s).

3.2 Result and Discussion

We first present the typical matching result between first frame left image and second frame left image using the proposed method and RANSAC based outlier elimination in Figure 4a. and Figure 4b., respectively.

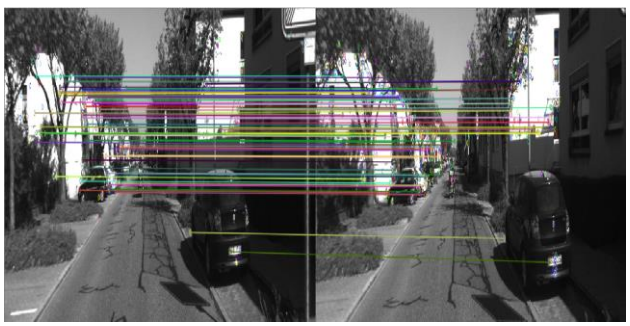

Figure 4a. The matching result by the proposed method

Figure 4b. The matching result by the RANSAC only method

From Figure 4a. and Figure 4b. we can see clearly that the matching result presented by the proposed method exists rare outliers and shows better correspondence compared to the RANSAN based outlier elimination method.

Then we present the whole results obtained by the proposed method and RANSAC based Visual Odometry (RVO) in the Table 1:

	Proposed method	RVO
Number of fail frames	12	24
TSP	99.7%	99.4%

Table 1. TSP result for "2011_10_03_drive_0027"

We can see the result from the Table 1 that the robustness of VO benefits from the employment of EDC. For total 4540 frames, the number of fail tracking frames is decreased by 50% compared with the RANSAC only method, from 24 to 12.

The whole trajectory of data set "2011_10_03_drive_0027" is shown in Figure 5, presented by the GPS ground truth, RVO and the proposed method. The GPS ground truth is described by the blue solid line. Red dot dash line and green dot dash line represent the ROV method and proposed method, respectively.

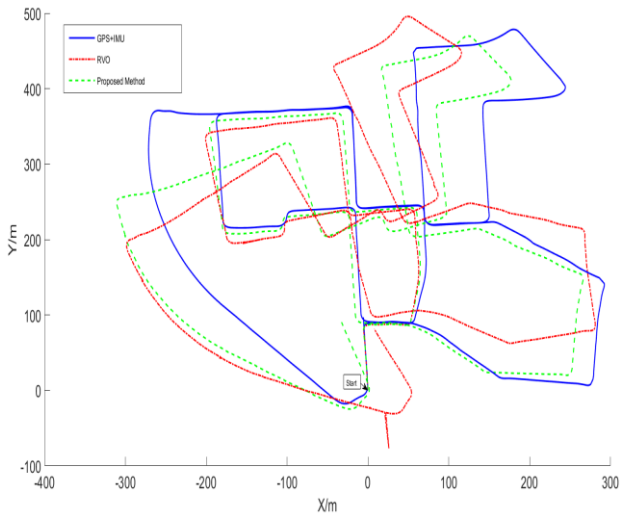

Figure 5. The trajectory result of "2011_10_03_drive_0027".

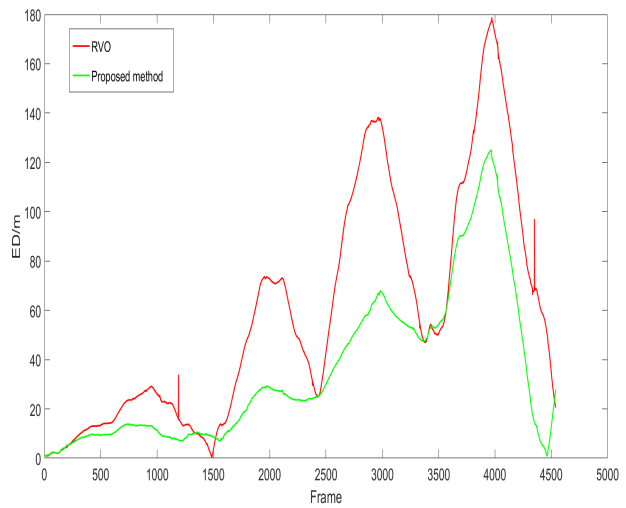

Figure 6. The distance error from frame 1 to frame 4540.

	Proposed method	RVO
AED/m	35.2392	59.3825

Table 2. The AED result of the proposed method and RVO

In our experiment, when kth frame is proven a fail tracking frame, we just reserve the result for raw data recording. But take the next k+1th frame into account, we calculate its rotation and translation from the k-1th frame to maintain the precision of VO method.

From Figure 5 and Table 2 we can see clearly that the proposed method performs superior to the RVO both in rotation and translation. The AED of proposed method is 35.2392m, reducing by 40% error compared to the RVO method. Besides, the distance error of the proposed method is slighter than the RVO at almost frames, which displays notably at the Figure 6.

4. CONCLUSION

This paper presents a novel robust method for 3D-2D VO method. The EDC and RANSAC algorithm are employed in the proposed method, the former dealing with outlier elimination primarily and the latter removing the effect of the remaining outliers.

For evaluating our method, we employ the KITTI dataset, which can obtain from the Internet. From the experiment result using the synchronized dataset "2011_10_03_drive_0027" we can see that the fail tracking rate reduce from 0.6% to 0.3%, which means the improvement of robustness, benefiting from the proposed method. The precision of the trajectory improves from 59m to 35m due to the EDC at the mismatch elimination process. Our quantitative analysis shows that our robust method is superior to the RANSAC only method by 40%~50%.

ACKNOWLEDGEMENTS

This paper was substantially supported by the National Natural Science Foundation of China (Project nos. 41631178, 41325005 and 41401531) and the Fundamental Research Funds for the Central Universities.

REFERENCES

Ansar, A., Daniilidis, K., 2003a. Linear pose estimation from points or lines. IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(5), 578–589.

Bajpai, A., et al., 2016a. "Planetary Monocular Simultaneous Localization and Mapping." Journal of Field Robotics 33(2): 229-242.

Balazadegan Sarvrood, Y., et al., 2016a. "Visual-LiDAR Odometry Aided by Reduced IMU." ISPRS International Journal of Geo-Information 5(1): 3.

Bay, H., et al., 2006. SURF: Speeded Up Robust Features. Computer Vision – ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I. A. Leonardis, H. Bischof and A. Pinz. Berlin, Heidelberg, Springer Berlin Heidelberg: 404-417.

Fanfani, M.; Bellavia, F.; Colombo, C, 2016a. Accurate keyframe selection and keypoint tracking for robust visual odometry. Mach. Vis. Appl. 2016, 27, 833–844

Gao, X. S., Hou, X. R., Tang, J., & Cheng, H. F., 2003a. Complete solution classification for the perspective-three-point problem. IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(8), 930–943.

Geiger, A., et al., 2013a. "Vision meets robotics: The KITTI dataset." The International Journal of Robotics Research 32(11): 1231-1237.

Hartley, R., & Zisserman, A., 2000. Multiple view geometry in computer vision. Cambridge: Cambridge University Press

He, H.; Li, Y.; Guan, Y.; Tan, J, 2015a. Wearable Ego-Motion Tracking for Blind Navigation in Indoor Environments. IEEE Trans. Autom. Sci. Eng. 12, 1181–1190.

Hong, S.; Ko, H.; Kim, J, 2008a. Improved motion tracking with velocity update and distortion correction from planar laser scan data. In Proceedings of the International Conference Artificial Reality and Telexistence; pp. 315–318.

J. Engel, J. Stueckler, and D. Cremers, 2015."Large-scale direct SLAM with stereo cameras," in IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).

Ke, Q.; Kanade, T. Transforming camera geometry to a virtual downward-looking camera: Robust ego-motion estimation and ground-layer detection. In Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'03), Madison, WI, USA, 18–20 June 2003; pp. 390–397.

Kitt, B.; Geiger, A.; Lategahn, H, 2010. Visual odometry based on stereo image sequences with RANSAC-based outlier rejection scheme. In Proceedings of the 2010 IEEE Intelligent Vehicles Symposium, San Diego, CA, USA, 21–24 June 2010; pp. 486–492.

Lowe, David G., 1999. "Object recognition from local scale-invariant features". Proceedings of the International Conference on Computer Vision 2, pp. 1150-1157. doi: 10.1109/ICCV.1999.790410

Maimone, M., et al., 2007a. "Two years of Visual Odometry on the Mars Exploration Rovers." Journal of Field Robotics 24(3): 169-186.

McGlove, C., Mikhail, E., & Bethel, J. (Eds.), 2004. Manual of photogrametry.American society for photogrammetry and remote sensing (5th edn.).

Min, Q.; Huang, Y, 2016a. Motion Detection Using Binocular Image Flow in Dynamic Scenes. EURASIP J. Adv. Signal Process. 49

Moreno-Noguer, F., et al., 2007. Accurate Non-Iterative O(n) Solution to the PnP Problem. 2007 IEEE 11th International Conference on Computer Vision.

Musleh, B.; Martin, D.; De, L.; Armingol, J.M, 2012. Visual ego motion estimation in urban environments based on U-V. In Proceedings of the 2012 Intelligent Vehicles Symposium, Alcala de Henares, Spain; pp. 444–449.

Quan, L., & Lan, Z., 1999a. Linear N-point camera pose determination. IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(7), 774–780.

Rublee, E., et al., 2011. ORB: An efficient alternative to SIFT or SURF. 2011 International Conference on Computer Vision.

Scaramuzza, D.; Fraundorfer, F. Tutorial, 2011: Visual odometry. IEEE Robot. Autom. Mag. 18, 80–92.

Talukder, A.; Matthies, L, 2004. Real-time detection of moving objects from moving vehicles using dense stereo and optical flow. In Proceedings of the 2004 IEEE/RSJ International

Conference on Intelligent Robots and Systems, Sendai, Japan; pp. 3718–3725.

T. Pire, T. Fischer, J. Civera, P. De Crist´oforis, and J. J. Berlles, 2015. “Stereo parallel tracking and mapping for robot localization,” in IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1373–1378.

Weiss, S, 2009. Visual SLAM in Pieces; ETH Zurich: Zurich, Switzerland.