

A 4-point Algorithm for Relative Pose Estimation of a Calibrated Camera with a Relative Rotation Angle

Abstract— We propose an algorithm to estimate relative camera pose using four feature correspondences and one relative rotation angle measurement. The algorithm can be used for relative pose estimation on a vehicle platform mounted with camera and a relative rotation angle sensor, either an odometer or an IMU. This algorithm exploits a fact that relative rotation angle of odometer is invariant under the rigid transform between camera and the rotation sensor. Therefore, extrinsic calibration between camera and sensor is not required. We show that the proposed algorithm improves relative pose estimation compared with the well known 5-point and 1-point algorithm.

I. INTRODUCTION

Vehicle platform mounted with camera and either a odometer, IMU, or yaw rate sensor has been widely used in research fields of computer vision, robotics and automatic control. For years research has been focused on using these economic sensors to locate the vehicle as well as reconstruct the environment. This research is also referred as Visual SLAM in Robotics.

The key step of Visual SLAM is to estimate relative camera pose between frame pair. One commonly used method is feature based estimation. A subset of image feature correspondences are selected to estimate the Fundamental Matrix or Essential Matrix between two frames. Relative rotation and translation can then be extracted from the estimation result. A series of “n-point” (n-correspondences) algorithms are proposed for this aim. If the camera has with unknown intrinsics, the Fundamental Matrix can be estimated by 8-point or 7-point [1] algorithm. If the camera has calibrated intrinsics, then 6-point [2], [3] or 5-point [4], [5] can be used to the Essential Matrix. Based on these algorithms, robust estimation methods like RANSAC or LMedS are used to generate best estimation from a set of points with both inliers and outliers. The result of “n-point” algorithm is significantly affected by the feature correspondences detected from images. It is well admitted that the algorithm using fewer points has lower estimation error. For calibrated camera, 5-point is the minimal case to solve the 5 DoF relative pose. In $[x, x]$, 5-point algorithm shows the best estimation performance compared with 6, 7, 8-point algorithms. For Visual SLAM or Structure from Motion problem, 5-point algorithm is also the most commonly used algorithm if the camera is calibrated.

However, 5-point algorithm is not guaranteed to always have a stable estimation. On a vehicle with camera mounted in its front, relative pose estimation is especially a difficult case for 5-point algorithm. The difficulties include the following: 1) 5-point algorithm less stable in forward motion compared with sideways motion. 2) Features seen by a front camera are mainly far away from the vehicle, which

also make the estimation unreliable. Readers can find some discussion about the performance of 5-point algorithm in [4], [5]. To improve the estimation quality, research has been drawn on exploiting extra information from other sensors or from specific motion model. For example, [6] obtains two rotation angle from IMU and used a 3-point algorithm to estimate the relative pose of a micro aerial vehicle. This method requires that the extrinsic calibration between camera and IMU is known. In [7], a very neat 1-point algorithm was proposed. The algorithm assumes that vehicle platform motion follows the General Ackermann steering model. The 1-point algorithm can work fast and stable for vehicle platform but requires that camera is well align on the rear axis of the vehicle.

If the vehicle platform has an odometer and its extrinsic parameter is available, it is well known that camera relative pose can be directly obtained from odometer pose H_o as $H^{-1}H_oH$, where H is the transform between camera and odometer. However, in practice estimating H is not easy. This problem is known as Hand-eye (or Head-eye) calibration [8], [9]. Moreover, the success of these hand-eye calibration algorithms also requires accurate visual odometry estimation. This visual odometry estimation also requires pure feature based algorithm like 5, 6, 7, 8-point algorithms.

Similarly to [6], [7], in this paper we propose an algorithm to improve the relative pose estimation using information from extra sensors. The algorithm uses 1 rotation angle from any relative rotation sensor including odometer, IMU, and 4 point pairs from camera. In the algorithm, camera can be mounted anywhere on the platform and no extrinsic calibration is required. Since odometer reading can be much more stable and accurate, the proposed can largely improve relative estimation compared with existing methods.

The rest of the paper is organised as follows. Sec.II establishes notations and formulas used in the proposed methods. Sec.III presents the form of 4-point relative pose problem and the algorithm to solve it. The performance of the algorithm is studied in Sec.IV, where it is compared with the well known 5-point algorithm on simulated cases to explore how much the algorithm improves 5-point algorithm. Furthermore, the algorithm is compared with 5-point and 1-point algorithm on a real dataset obtained from our vehicle platform mounted with camera and odometer.

II. PRELIMINARIES

Image points from the first and second frames are denoted by homogeneous vectors $p_1 = (x_1, y_1, 1)^\top$ and $p_2 = (x_2, y_2, 1)^\top$, resp. Intrinsic matrix of the camera is denoted

as K . Since the proposed algorithm requires K to be known, we hereby assume that p_1 and p_2 are always premultiplied by K^{-1} .

Denote R and t as the relative rotation and translation between the first and second frame. The Essential Matrix corresponding to R and t can be denoted as

$$E = [t]_{\times} R \quad (1)$$

where $[t]_{\times}$ denotes the skew symmetric matrix

$$[t]_{\times} \equiv \begin{pmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{pmatrix} \quad (2)$$

For correspondence p_1, p_2 , it is well know that

$$p_2^{\top} E p_1 = 0 \quad (3)$$

Rodrigues' rotation formula Given a 3D unit rotation axis vector $r = (r_x, r_y, r_z)^{\top}$ and a rotation angle θ , it is easy to find the the corresponding rotation matrix using Rodrigues' rotation formula.

$$R(\theta, r) = \cos \theta I + (1 - \cos \theta) r r^{\top} + \sin \theta [r]_{\times} \quad (4)$$

where I is a 3×3 identity matrix.

Theorem The rotation angle of camera and odometer is always the same.

This is a very well known fact for rigid motion. Detailed proof can be found in many textbooks on mechanics. This fact means that relative camera rotation is the same value measured by the relative rotation angle sensor. Note that this angle is the only invariant measurement from the sensor to camera. Without extrinsic calibration, we do not know the rotation axis in the camera frame, though in the world frame, it has the same direction with the rotation angle of the sensor.

III. ALGORITHM

A. Problem Formation

Substituting Eq.(4) into Eq.(1). We can denote the Essential Matrix as

$$E(\theta, r, t) = [t]_{\times} (\cos \theta I + (1 - \cos \theta) r r^{\top} + \sin \theta [r]_{\times}) \quad (5)$$

where r is a 3D unit vector and t can also be assumed to be unit since it is up to scale. Since we assume that we know θ from odometer reading, this means that the 5DoF camera relative pose now only have 4 DoF. By using 4 image points correspondences, we can solve r and t .

Thus the equation system for solving relative pose can be form as follows.

$$p_2^i{}^{\top} E(\theta, r, t) p_1^i = 0, \quad i = 1, 2, 3, 4 \quad (6)$$

$$\|r\|^2 = 1 \quad (7)$$

$$\|t\|^2 = 1 \quad (8)$$

where $r = (r_x, r_y, r_z)^{\top}$ and $t = (t_x, t_y, t_z)^{\top}$ are six unknowns.

B. Solution

Equation system (6) includes 4 polynomials with the highest monomial in the form of $t_{\star} r_{\star} r_{\star}$, where \star denotes any arrangement of 'x', 'y', 'z'. Eq.(7) and Eq.(8) are two quadratic polynomials.

The degree and the number of equations for this equation system is very high compared with other normal minimal solution problem. In our experiment, we attempted to reduce it using Resultant Method in Maple. However, this attempt failed as it exceeded the PC memory capacity after one hour running. A working way to solve the system is to use Groebner basis. In our experiment, we used the automatic generator of Groebner solver developed by [10]. After possible simplification, the solver still required executing a Gauss-Jordan elimination on a 270×290 matrix, which is very computationally expensive. In this paper, we proposed a method to use gradient descent method to solve the above equation system.

Expanding Eq.(6), we can denote the equations in the following form.

$$f_1^i(r_x, r_y, r_z) t_x + f_2^i(r_x, r_y, r_z) t_y + f_3^i(r_x, r_y, r_z) t_z = 0 \quad (9)$$

where $i = 1, 2, 3, 4$. f_{\star}^i is a polynomial of r_x, r_y and r_z . We can further stack f_{\star}^i into a matrix as follows.

$$F(r_x, r_y, r_z) t \equiv \begin{pmatrix} f_1^1 & f_2^1 & f_3^1 \\ f_1^2 & f_2^2 & f_3^2 \\ f_1^3 & f_2^3 & f_3^3 \\ f_1^4 & f_2^4 & f_3^4 \end{pmatrix} \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} = 0 \quad (10)$$

Since we assume that $\|t\|^2 = 1$, the rank of F must be 2. This means that the determinant of all 3×3 submatrices must be 0. We form this as follows.

$$\begin{vmatrix} f_1^1 & f_2^1 & f_3^1 \\ f_1^2 & f_2^2 & f_3^2 \\ f_1^3 & f_2^3 & f_3^3 \end{vmatrix} = 0 \quad (11)$$

$$\begin{vmatrix} f_1^2 & f_2^2 & f_3^2 \\ f_1^3 & f_2^3 & f_3^3 \\ f_1^4 & f_2^4 & f_3^4 \end{vmatrix} = 0 \quad (12)$$

Combining Eq.(11), Eq.(12) and Eq.(7), we have a new equation system on r_x, r_y, r_z . Eq.(11) and Eq.(12) are of 5 degree. Notice that Eq.(8) implies that $(r_x, r_y, r_z)^{\top}$ is on a unit sphere. We can parametrize it as 2D point and then use gradient descend method like Powell's Hybrid method [11] to find a root from an initial guess. In our experiment, we used 100 initial guesses drawn by uniform sampling on a unit sphere. This setting could always reaches a proper solution and fails very rarely. After r_x, r_y, r_z are solved, F is then obtained and t can be solved as the nullvector of F .

Note that multiple roots r exist for the equation system and for each possible solution (r, t) , $(r, -t)$ is also a possible solution. By checking reprojection error e.g. Sampson's error, extra solutions can be rejected till only two solutions (r, t) and $(r, -t)$ remain. The extra solution can then be rejected by checking if the reconstructed 3D points have positive depth. This is also called cheirality check in [4].

For a set of correspondences detected from an image pair, methods like RANSAC or LMedS can be used to find the optimal estimation and remove outliers.

IV. EXPERIMENTS

A. Implementation details and timing issues

As mentioned in Sec.III-B, we implemented the 4-point algorithm using Powell's Hybrid method from General Scientific Library (GSL). For each initial guess, the method reached a root or failed within 20 iterations. For the minimal case with four points, this implementation takes 5 ms on average. In our experiment on a Core i7 desktop, this is enough to process one RANSAC estimation in 0.4s.

The computational complexity can be further optimized. The calculation of coefficients in Eq.(11) and Eq.(12) involves a series extremely large polynomials generated by Maple, which is the computational bottleneck of the current implementation. The computation can be much accelerated by carefully rearrange the polynomial calculation. In addition, the Groebner based solver generated by [10] can also be accelerated with sparse matrix decomposition algorithm.

B. Performance under noise

We use simulation data in this section to test the performance of the algorithm. For related algorithms including 8-point, 7-point, 6-point and 5-point algorithms, it is well admitted that 5-point algorithm outperforms the others for calibrated camera. Therefore, in this section we only compare our algorithm with 5-point algorithm. Detailed comparison between 8, 7, 6, 5-point algorithms can be found from [4], [5].

To make our experiment comparable with previous research, we kept our simulation setting close to the challenging condition in [4]. The setting is summarized in Table I and Figure 1. The two algorithms are tested under the forward and sideways motions. Relative pose estimation error is measured by the included angle between groundtruth translation and estimated translation. This is based on the fact that translation estimation is much more sensitive than rotation estimation, see e.g. [12]. Normally for both 4-point and 5-point algorithm, the rotation error is less than 0.1 deg. Image feature noise is generated as zero mean Gaussian noise with increasing standard deviation. For the minimal case test, four and five points are generated for 4-point and 5-point algorithms, resp. For the RANSAC case test, we used 50 point to run a RANSAC scheme to generate the best estimation. 1000 tests were executed for each noise level in the minimal case test. 100 tests were executed for each noise level in the RANSAC test. Fig.2 shows the plot of estimation error against increasing feature noise standard deviation. In Fig.2a and 2b, we plotted the lower quartile of error at each noise level for tests of minimal case. In Fig.2c and 2d, we plotted the mean error at each noise level for the RANSAC tests. The estimator were selected to be consistent with those in [4]. From the plots, we can clearly see the improvement of 4-point algorithm over 5-point.

TABLE I
EXPERIMENT SETTING FOR SIMULATION DATA.

Minimal Distance	10
Depth	10
Baseline	1
Image Size	350×350
Field of View	60 deg
Error Measurement	Translation deviation angle
Error Estimator	Lower quartile (minimal case) Mean (RANSAC case)
Tests per Noise Level	1000 (minimal case) 100 (RANSAC case)

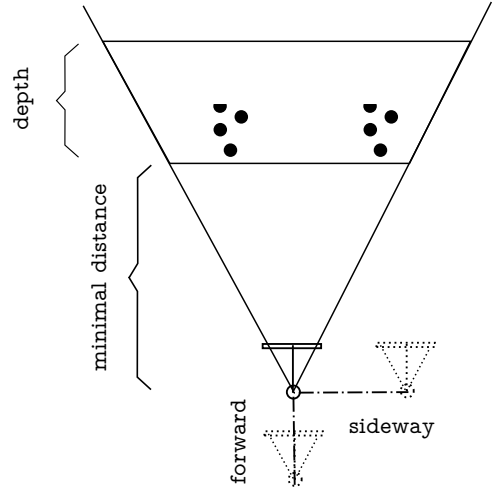


Fig. 1. Experiment setting for simulation data. Two dashed cameras show how camera is moved under forward and sideways motion.

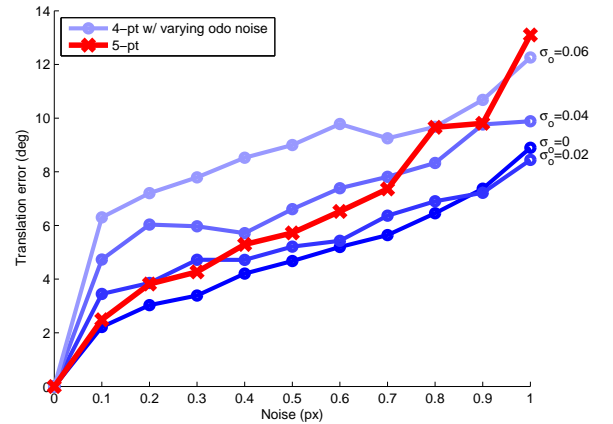


Fig. 3. Translation error in deg against noise standard deviation in px, under rotation angle standard deviation σ_o of 0 \sim 0.06. 50 points are used to execute a RANSAC scheme for each test.

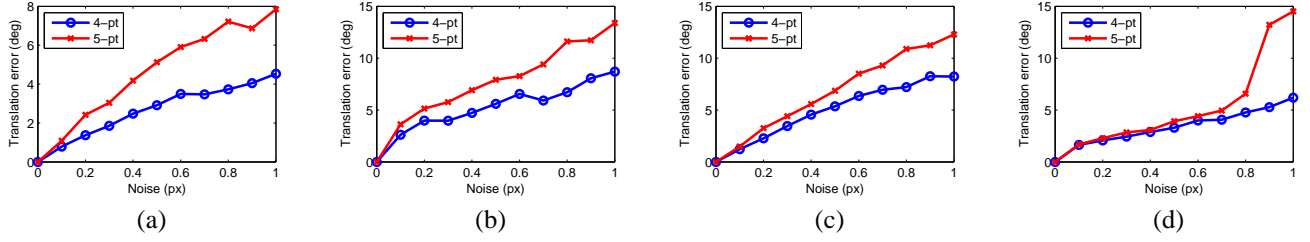


Fig. 2. Translation error in deg against noise standard deviation in px. (a) Minimal cases, forward motion. (b) Minimal cases, sideways motion. (c) 50 points, forward motion. (d) 50 points, sideways motion.

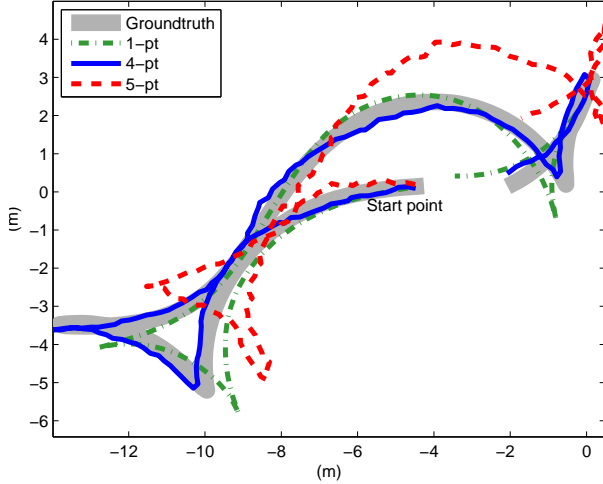


Fig. 4. Visual odometry generated using 1-point, 4-point and 5-point and corresponding groundtruth.

In addition to the above tests, we provide a test here to show how the 4-point algorithm is influenced by the noise from relative rotation angle sensor. Generally it is not straightforward to model the noise of relative rotation angle sensor. Since the rotation angle is usually obtained by integrating the angular velocity, we assume here that a relative rotation angle θ is measured as $(1 + e)\theta$ by sensor, where e follows a zero mean Gaussian distribution. This test, we selected the standard deviation σ_e as 0, 0.02, 0.04, 0.06, and generated 4 plots between translation error and noise level. In this tests, we used the same setting with the above RANSAC tests, except that translation direction is arbitrary and rotation angle is within $-10 \sim 10$ deg. The plot is shown in Fig.3. We can find that with error rate e smaller than 0.04, 4-point algorithm can give a better result than 5-point algorithm. In practice, the error rate of rotation sensor is much smaller than that, which guarantees 4-point to have a solid performance.

C. Performance for real data

In this section, we compare the 4-point algorithm with 5-point and 1-point algorithms, which are two well known algorithm for relative pose estimation. Our platform is built on a VW Golf, with a odometer and a camera mounted at the back of the vehicle. The camera intrinsic is well calibrated.

Note we put the camera at the back so that it is roughly on the rear axis of the vehicle. This condition is required by the 1-point algorithm. Further details about the constraints 1-point algorithm needs can be found in [7].

We took a image sequence with 200 frames and plot the visual odometry generated by the three algorithms. We on purpose made the vehicle trajectory very curvy such that the the estimation deviation can be obvious even in a short sequence. The groundtruth of camera motion is generated using wheel odometry and the extrinsics between camera and odometer. Since in this paper we only need compare the relative pose estimation, the scale information between each frames is directly obtained from the groundtruth for the three algorithms. Furthermore, only the relative pose between consecutive frames is estimated and no bundle adjustment is used. The plotted trajectory is in Fig.4. We can clearly see that 4-point algorithm generates a trajectory most close to the groundtruth. 5-point and 1-point algorithm have large drift in their visual odometry. For 5-point algorithm, if the image feature quality is low, the relative pose can be significantly deviated. This leads to the sawtooth shape of the trajectory. For 1-point algorithm, the trajectory can be very smooth due to the Ackermann steering assumption. However, the assumption is only an approximation. This leads to a continuous bias for each frame. Comparing to 5-point and 1-point algorithm, the proposed 4-point algorithm provides a much stabler estimation than 5-point, while it requires no special approximation for the extrinsics like 1-point algorithm.

V. CONCLUSION

In this paper, we showed that by using the relative rotation angle from some relative rotation sensor, the relative camera pose can be estimated from four image feature correspondences. The algorithm does not require any information about the extrinsics. Therefore, it can widely be used for normal vehicle platform with camera and some rotation sensors. Furthermore, we would like to point here the advantage of the algorithm in the scenario of hand-eye calibration, since in this case no extrinsic information is available.

In both the simulated and real experiments, the algorithm has shown obvious improvement compared with the well known 5-point and 1-point algorithm. The complexity of this algorithm is lower compared with 5-point and 1-point algorithm, but still potential for further optimization. An open source release of the algorithm will be soon available.

REFERENCES

- [1] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge Univ Press, 2000, vol. 2.
- [2] H. Stewénius, D. Nistér, F. Kahl, and F. Schaffalitzky, "A minimal solution for relative pose with unknown focal length," *Image and Vision Computing*, vol. 26, no. 7, pp. 871–877, 2008.
- [3] H. Li, "A simple solution to the six-point two-view focal-length problem," *Computer Vision–ECCV 2006*, pp. 200–213, 2006.
- [4] D. Nistér, "An efficient solution to the five-point relative pose problem," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 6, pp. 756–770, 2004.
- [5] H. Stewénius, C. Engels, and D. Nistér, "Recent developments on direct relative orientation," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 60, no. 4, pp. 284–294, 2006.
- [6] F. Fraundorfer, P. Tanskanen, and M. Pollefeys, "A minimal case solution to the calibrated relative pose problem for the case of two known orientation angles," *Computer Vision–ECCV 2010*, pp. 269–282, 2010.
- [7] D. Scaramuzza, "1-point-ransac structure from motion for vehicle-mounted cameras by exploiting non-holonomic constraints," *International journal of computer vision*, vol. 95, no. 1, pp. 74–85, 2011.
- [8] R. Y. Tsai and R. K. Lenz, "A new technique for fully autonomous and efficient 3d robotics hand/eye calibration," *Robotics and Automation, IEEE Transactions on*, vol. 5, no. 3, pp. 345–358, 1989.
- [9] Y. C. Shiu and S. Ahmad, "Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form $\begin{bmatrix} a_{ij} \\ x_{ij}/e_{ij} \end{bmatrix} = \begin{bmatrix} e_{ij} \\ x_{ij}/e_{ij} \end{bmatrix}$," *Robotics and Automation, IEEE Transactions on*, vol. 5, no. 1, pp. 16–29, 1989.
- [10] Z. Kukelova, M. Bujnak, and T. Pajdla, "Automatic generator of minimal problem solvers," *Computer Vision–ECCV 2008*, pp. 302–315, 2008.
- [11] M. J. Powell, "An efficient method for finding the minimum of a function of several variables without calculating derivatives," *The computer journal*, vol. 7, no. 2, pp. 155–162, 1964.
- [12] T. Y. Tian, C. Tomasi, and D. J. Heeger, "Comparison of approaches to egomotion computation," in *Computer Vision and Pattern Recognition, 1996. Proceedings CVPR'96, 1996 IEEE Computer Society Conference on*. IEEE, 1996, pp. 315–320.