2014 IEEE International Conference on Robotics & Automation (ICRA)
Hong Kong Convention and Exhibition Center
May 31 - June 7, 2014. Hong Kong, China

A Two-Point Algorithm for Stereo Visual Odometry in Open Outdoor Environments

Kyohei Otsu[1] and Takashi Kubota[2]

Abstract— **This paper proposes a novel method to estimate relative poses for a calibrated stereo camera. Three corresponding points in 3D space are theoretically required to recover unconstraint motion which has six degrees of freedom. The proposed method solves this problem with only two 3D points by exploiting a common reference direction between poses. Two points are selected in accordance with the distance to the camera: one distant point is used for deriving a reference direction, and one near point is used for estimating accurate translation. The distance is computed by triangulation in stereo vision. The uncertainty of triangulation can be mitigated by the appropriate selection strategy. The experiments using synthetic and real data demonstrate the effectiveness and higher stability of the proposed method against image pixel noise.**

I. INTRODUCTION

This paper describes an efficient solution to the relative pose problem using two image point correspondences in two stereo views. The relative pose problem is one of the major problems in a number of fields, including computer vision, photogrammetry, robotics, augmented reality, etc. This paper assumes a stereo camera rig in an open outdoor environment, which can be applied to SLAM (Simultaneous Localization and Mapping) and VO (visual odometry).

There is a crucial challenge for pose estimation in handling outliers of point correspondences. In outdoor environments, there are a lot of possible causes which make data association problem difficult, e.g., moving objects, occlusions, illumination changes, and image noise. The RANSAC [1] has been established as a standard method for outlier removal. It is a hypothesis-and-test framework which produces hypotheses with randomly sampled data sets and selects the hypothesis that acquires highest consensus with other data. It is important to find the minimal number of data points which can estimate model parameters, since the required number of iterative operations is exponential in the number of necessary data points [1], [2].

Several attempts have been performed to find the minimal solver for the relative pose problem. For a monocular camera, the minimal number of recovering motion parameters is five [3], [4]. It has been reduced by adding cues for the orientation. Naroditsky et al. [5] proposed a three-plus-one (four) algorithm by exploiting a point at infinity. They formulate a close-form solution and a solution using action matrix. Fraundorfer et al. [6] proposed a three-point algorithm based on the formulation of the five-point algorithm [3]. They resolve two orientational angles with inertia measurements. Kalantari et al. [7] also showed a three-point algorithm with different formulation using Gröbner basis. They used inertial measurements or a vertical vanishing point to reduce the number of motion parameters. Kneip et al. [8] further reduced it to two, by computing full rotation matrix from an IMU rigidly attached to the camera. A single-point algorithm was proposed by Scaramuzza [2] for a planar motion of a nonholonomic wheeled vehicle. These monocular methods can also be applied for a stereo camera.

For a stereo camera, three correspondences are required to recover motion parameters. The solvers may be divided into two approaches. The 3D-to-2D methods use the correspondences between 3D space points and 2D image points. These methods are often referred as the perspective-three-point (P3P) solvers [1], [9], [10]. The 3D-to-3D methods compute poses using 3D space point correspondences. The solution can be obtained algebraically by least squares [11], [12] and weighted least squares [13], or solved by optimization procedure [14]. The number of data points can be reduced by exploiting the projective geometry. The one-point solver for a stereo rig is proposed by Ni and Dellaert [14], which resolves the camera orientation using the infinite homography. Their follow-up work in [15] solves the orientation using two distant points and the translation using one near points.

Recent approaches for reducing the number of data points are summarized in Table I. A major approach to reduce the number of data points is to use common directional references between two camera poses. The directional correspondences are often computed by inertial measurements or points at infinity. However, such measurements are not always available particularly for off-road vehicles in an outdoor environment. The inertial measurements may be unstable for running vehicle on uneven terrain, and vanishing points are hard to be found on environments without man-made objects.

This paper proposes a new scheme to use two corresponding points, instead of three, obtained from a stereo camera rig. It estimates the orientation and translation separately, similar to [14], [15]. This strategy can greatly improve the computational efficiency. The two corresponding points are selected in accordance with the distance to the camera. One corresponding point is a distant point used for orientation estimation. As demonstrated in [5], the points which are sufficiently far can be used as reference direction. While [5] uses the RANSAC to select distant points, the proposed method simply uses stereo triangulation to know the depths

[1]K. Otsu is with Department of Electrical Engineering and Information Systems, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo, Japan. kyon@ac.jaxa.jp

[2]T. Kubota is with Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency, 3-1-1 Yoshinodai, Chuo, Sagamihara, Kanagawa, Japan. kubota@isas.jaxa.jp

TABLE I
RECENT APPROACHES TO REDUCE DATA POINTS

Reference	Camera(Min)	Points	Clues	Natural Scene	Driving	6DoF	Scale	Solutions
Naroditsky et al. [5]	Monocular(5)	3+1	Point at infinity	✓	✓	✓		4
Fraundorfer et al. [6]	Monocular(5)	3	IMU	✓		✓		4
Kalantari et al. [7]	Monocular(5)	3	Vanishing point/IMU	(✓)	(✓)	✓		4
Kneip et al. [8]	Monocular(5)	2	IMU	✓		✓		1
Scaramuzza [2]	Monocular(5)	1	Planar, nonholonomic	✓	✓			1
Ni and Dellaert [14]	Stereo(3)	1	Infinite homography		✓	✓	✓	1
Proposed method	Stereo(3)	2	Distant point	✓	✓	✓	✓	2

of points. The other corresponding point is a near point used for translation estimation. Near space points are accurately triangulated and therefore less uncertain. A single point in 3D space is sufficient to recover a translational motion with known camera orientation.

The proposed two-point method is quantitatively evaluated through the experiments with synthetic and real data. The impact of distance for the accuracy of pose estimation is studied. The proposed method outperformed other methods in noisy situation which often appears in outdoor fields.

II. STEREO CAMERA MODELS

To understand the error model in stereo vision, the geometry of stereo camera is reviewed in this section.

A. Front-Parallel Configuration

A standard configuration of stereo cameras is using two horizontally displaced cameras viewing the same direction (Fig. 1). The cameras are calibrated, i.e., with known intrinsic and extrinsic parameters, which together form a projection matrix P. The position of a point $\boldsymbol{X} \in \mathbb{R}^3$ is estimated by the triangulation using two set of projected points $\boldsymbol{u}_L, \boldsymbol{u}_R \in \mathbb{R}^2$. The ideal case of this specific configuration derives the simple stereo equations

$$\boldsymbol{X} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \frac{b}{d} \begin{bmatrix} u \\ v \\ f \end{bmatrix} \tag{1}$$

where b denotes the length of the baseline, f denotes the focal length of cameras, and $d = u_L - u_R$ denotes the disparity of a projected feature point. However, the presence of noise complicates the problem. Under the noisy situation, two rays from the cameras are not guaranteed to cross, thus the stereo equations do not hold. Several methods have tackled with this problem, some of which can be found in [16]. A common method is the linear triangulation algorithm. In this paper, the Linear-LS (Least Square) method is used which is fast to compute and affine-invariant.

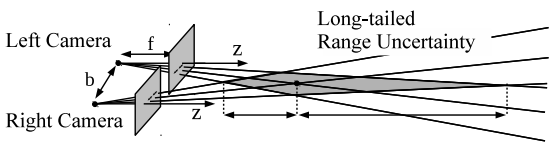


Fig. 1. Geometry of front-parallel stereo camera

B. Range Uncertainty and Statistical Bias

The accuracy of motion estimation is degraded by the range estimation uncertainty derived from triangulation [17]. The uncertainty grows larger along the depth direction.

A study has conducted to reveal how the pixel uncertainty of feature detectors (e.g. corner detectors) translates into the range estimation uncertainty in long-range stereo measurements [18]. The pixel uncertainty is commonly modeled as a gaussian distribution [18], [19]. As one can see in Fig. 1, the pixel uncertainty in an image direction forms a quadrilateral uncertainty in 3D space (gray region). Analytically, this can be shown in Fig. 2 as the range probability density function. Note that the range uncertainty derived from the pixel uncertainty has a heavy-tailed distribution. Supposing the Gaussian uncertainty in pixels, the range uncertainty can be modeled as

$$\frac{bf}{\sqrt{2\pi}\sigma_d z^2} \exp\left(\frac{-(bf/z - \mu_d)^2}{2\sigma_d^2}\right) \tag{2}$$

where μ_d and σ_d denotes the mean and standard deviation of disparity measurements. The numerical analysis in [18] shows that the mean of (2) statistically overestimates the true mean due to the heavy tail. Such uncertainty increases as a strong function of range and the bias does as well.

As discussed in [20], this range measurement bias has a strong influence on motion estimation. Their experimental result claims that the improper selection of feature points may cause the bias error in motion estimation. A bias reduction scheme is proposed in [18] based on Taylor series expansion of (1) under the assumption that the variation of disparity measurements around the true value is small.

The proposed method, however, does not compensate the bias errors. Rather, it selects points in a proper way so that the motion estimation error can be reduced. The idea is to

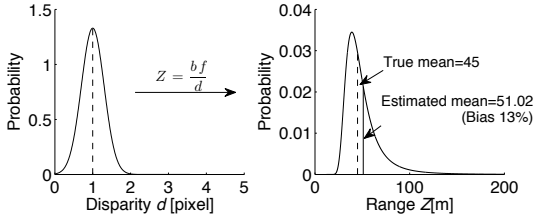


Fig. 2. Statistical bias in stereo triangulation (Parameters: $\sigma = 0.3$, $d = 1$, $bf = 45$).

use a distant point as a reference direction since it is less uncertain in location but has non-biased Gaussian distribution in direction. The translation is estimated with a near point which is accurately triangulated in 3D space.

III. RELATIVE POSE ESTIMATION

The mathematical formulation of the relative pose solver is described in this section. The solution mentioned here uses only two point correspondences. The key technique is estimating two rotation angles using an approximated reference direction derived from a distant point correspondence, and three translation parameters and the rest one rotation angle from a near point correspondence.

A. Notations

Let $\mathcal{F}_{(\cdot)}$ denote a three dimensional coordinate frame. The relation between the frames $\mathcal{F}_s$ and $\mathcal{F}_d$ is expressed by two elements $\{{}^d\boldsymbol{R}_s, \boldsymbol{t}_s^d\}$, where ${}^d\boldsymbol{R}_s \in \mathbb{R}^{3\times3}$ is the rotation matrix which rotates vectors expressed in $\mathcal{F}_s$ to $\mathcal{F}_d$, and $\boldsymbol{t}_s^d \in \mathbb{R}^{3\times1}$ is the vector in $\mathcal{F}_s$ to the origin of $\mathcal{F}_d$. A 3D point in $\mathcal{F}_s$ can be transformed to $\mathcal{F}_d$ by

$$\boldsymbol{X}_d = {}^d\boldsymbol{R}_s(\boldsymbol{X}_s - \boldsymbol{t}_s^d) + \boldsymbol{N}(z) \tag{3}$$

where $\boldsymbol{X}_{(\cdot)} \in \mathbb{R}^{3\times1}$ is the vector specifying a point in $\mathcal{F}_{(\cdot)}$ and $\boldsymbol{N}(z) \in \mathbb{R}^{3\times1}$ is the distance-dependent noise vector.

Four coordinate frames $\mathcal{F}_L, \mathcal{F}_R, \mathcal{F}_{L'}, \mathcal{F}_{R'}$ are defined for the left and right camera frames at time t and $t+1$. The origin of the camera frames are set to the optical centers. The relations of these frames are described by the known extrinsic parameters $\{{}^R\boldsymbol{R}_L, \boldsymbol{t}_L^R\} = \{{}^{R'}\boldsymbol{R}_{L'}, \boldsymbol{t}_{L'}^{R'}\}$ and the unknown relative pose $\{{}^{L'}\boldsymbol{R}_L, \boldsymbol{t}_L^{L'}\}$. See Fig. 3 for the relation of frames.

B. Recovering Two Angles from a Distant Correspondence

Consider unit vectors to a distant point $\boldsymbol{d}_L$ and $\boldsymbol{d}_R$ as shown in Fig. 4. Since these two vectors are not guaranteed to cross, compute the midpoint of two rays given by

$$\alpha_M \boldsymbol{d}_M = \frac{\alpha_L \boldsymbol{d}_L + \boldsymbol{t}_L^R + \alpha_R {}^L\boldsymbol{R}_R \boldsymbol{d}_R}{2} \tag{4}$$

where $\alpha_{(\cdot)}$ denotes the length of each ray. The normalization of the vector gives the unit vector $\boldsymbol{d}_M$. Two vectors $\boldsymbol{d}_M$ and $\boldsymbol{d}_{M'}$ should be aligned in the ideal case.

For realizing efficient computation, let us define the intermediate frame $\mathcal{F}_Z$ as in Fig. 5. $\mathcal{F}_Z$ has the same origin with $\mathcal{F}_L$, and its z-axis is aligned with $\boldsymbol{d}_M$ in $\mathcal{F}_L$. The

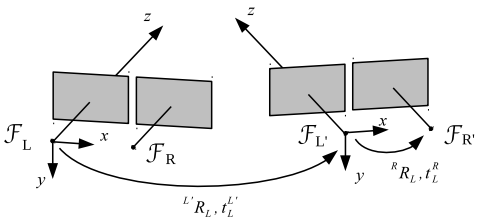


Fig. 3. Coordinate frames

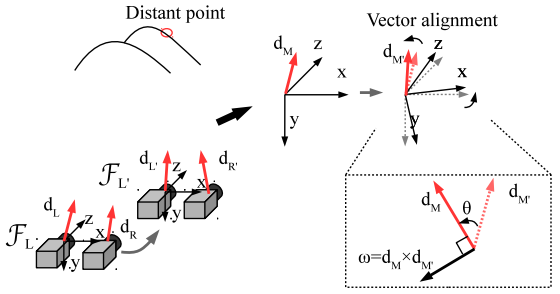


Fig. 4. Distant point correspondence

transformation between two frames are described only with rotation and given by

$$\boldsymbol{X}_Z = \boldsymbol{R}_\omega(\theta)\boldsymbol{X}_L \tag{5}$$

where

$$\boldsymbol{R}_{\boldsymbol{\omega}}(\theta) = \mathbf{1} + [\boldsymbol{\omega}]_\times \sin\theta + (1-\cos\theta)[\boldsymbol{\omega}]_\times^2 \tag{6}$$

$$\boldsymbol{\omega} = \mathbf{e}_z \times \boldsymbol{d}_M \tag{7}$$

$$\theta = \arccos(\mathbf{e}_z \cdot \boldsymbol{d}_M) \tag{8}$$

which is derived from Rodrigues' formula. Here, $\mathbf{1}$ denotes the identity matrix and $\mathbf{e}_z$ denotes the unit vector aligned with z-axis.

The frame $\mathcal{F}_{Z'}$ can also be introduced for the frame $\mathcal{F}_{L'}$. If the point is sufficiently far, $\boldsymbol{d}_{M'} \simeq \boldsymbol{d}_M$ and thus the z-axes of $\mathcal{F}_Z$ and $\mathcal{F}_{Z'}$ are approximately aligned. The rotation matrix between two poses can then be expressed as

$${}^{L'}\boldsymbol{R}_L = {}^{L'}\boldsymbol{R}_{Z'}{}^{Z'}\boldsymbol{R}_Z{}^Z\boldsymbol{R}_L \tag{9}$$

$$\simeq \boldsymbol{R}_{\boldsymbol{\omega}'}(\theta')^\top \boldsymbol{R}_{\mathbf{e}_z}(\psi)\boldsymbol{R}_{\boldsymbol{\omega}}(\theta) \tag{10}$$

where ψ is the rotation angle for solving the rotational ambiguity around z-axis.

C. Recovering One Angle and Three Translation Parameters from a Near Correspondence

Next, the remaining parameters are solved by a near point correspondence. In (3), the triangulation uncertainty $\boldsymbol{N}(z)$ can be negligible where the point is close to the camera. Thus, the translation is computed as

$$\boldsymbol{t}_L^{L'} \simeq \boldsymbol{X}_L - {}^{L'}\boldsymbol{R}_L^\top \boldsymbol{X}_{L'} \quad . \tag{11}$$

From (10) and (11), the translation vector between two intermediate frames is derived as

$$\boldsymbol{t}_Z^{Z'} \equiv \boldsymbol{X}_Z - \boldsymbol{R}_{\mathbf{e}_z}(\psi)^\top \boldsymbol{X}_{Z'} \quad , \tag{12}$$

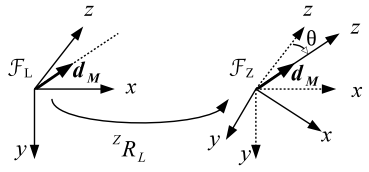


Fig. 5. Transformation to intermediate frame

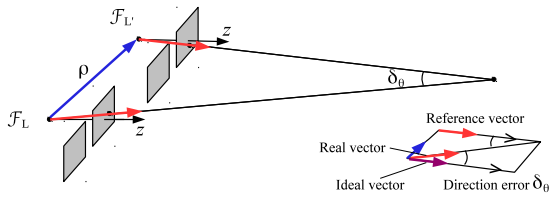


Fig. 6. Directional error imposed by using a point at finite distance

which is three equations with four unknown parameters. Notice that the vectors are transformed to the intermediate frames. By introducing the epipolar constraint as

$$\tilde{\boldsymbol{m}}_Z^\top [\boldsymbol{t}_Z^{Z'}]_\times \boldsymbol{R}_{\mathbf{e}_z}(\psi)^\top \tilde{\boldsymbol{m}}_{Z'} = 0 \tag{13}$$

all unknown parameters can be recovered by solving polynomial systems. Note that (12) and (13) are incompatible in the ideal case (i.e., without any noise), although it does not cause big problems in most real world cases. See the numerical accuracy in the simulation.

Supposing $\boldsymbol{X}_Z = \begin{bmatrix} x & y & z \end{bmatrix}$ and $\boldsymbol{m}_Z = \begin{bmatrix} u & v & 1 \end{bmatrix}$, substitution with (12) and (13) results in the following polynomials:

$$a_2 c + a_1 s + a_0 = 0 \tag{14}$$
$$c^2 + s^2 - 1 = 0 \tag{15}$$

where $c = \cos\psi$ and $s = \sin\psi$, and the coefficients

$$a_2 = uy' - vx' - v'x + u'y - (uv' - vu')(z' - z) \tag{16}$$
$$a_1 = ux' + vy' - u'x - v'y - (uu' + vv')(z' - z) \tag{17}$$
$$a_0 = vx - uy + v'x' - u'y' \tag{18}$$

Now, solve for c and s using (14) and (15). These quadratic equations gives at most two roots for c and s that cannot be neglected. Finally, the two set of c and s yield two solutions for ψ, which means all unknowns are solved. The relative pose $\{{}^{L'}\boldsymbol{R}_L, \boldsymbol{t}_L^{L'}\}$ can be recovered using (10) and (11).

D. Extension to Over-determined Case

The method can be extended to the over-determined case, where more than two point correspondences can be obtained, by returning least squares solution.

In the step of orientation estimation, two corresponding common directions disambiguates the parameter ψ, and thus uniquely determines the orientation. In the case of more than two correspondences, the solution can be obtained by the classic least squares fitting [11].

The least squares solution for the translation parameters is given by replacing $\boldsymbol{X}$ in (11) with mean vector $\bar{\boldsymbol{X}} = \sum_i \boldsymbol{X}_i$. It will be shown in the experiments that the stability and accuracy are improved by adding extra points.

E. Thresholds

In the proposed algorithm, there are two distances to be concerned: the thresholds for distant and near. The simple model of orientation error is depicted in Fig. 6. The two reference vectors are ideally parallel. However, if the point is at finite distance, the vectors form an acute-angled triangle. The orientation error δ_θ can be expressed by the theorem of cosines:

$$\delta_\theta = \arccos\left(\frac{\alpha_M^2 + \alpha_{M'}^2 - \rho^2}{2\alpha_M \alpha_{M'}}\right) \tag{19}$$
$$\simeq \arccos\left(1 - \frac{\rho^2}{2\alpha_M^2}\right) \tag{20}$$

where $\rho = |\boldsymbol{t}|$ and $\alpha_M \simeq \alpha_{M'}$. Hence, the threshold for distant points can be determined in function of the motion size and affordable error:

$$\tau_{dist} \geq \frac{\bar{\rho}}{\sqrt{2(1 - \cos\bar{\delta}_\theta)}} \tag{21}$$

On the other hand, the threshold for near points is given by several tangled factors: the location uncertainty, the directional error in (20), and the triangulation bias derived from (2). Under the assumption of $z >> b$, the location uncertainty is formulated as

$$\sigma_z = \frac{\sigma_d z^2}{bf} \quad . \tag{22}$$

According to [18], the bias can be computed by the second-order Taylor series expansion as

$$\delta_z = \left(\frac{\sigma_d}{bf}\right)^2 z^3 \quad . \tag{23}$$

The theorem of cosines gives the bound to determine the threshold as

$$\tau_{near}\sqrt{p^4\tau_{near}^4 + p^2 q \tau_{near}^2 + q} + \frac{\sqrt{2}}{p}\tau_{near}^2 \leq \bar{\delta}_{\boldsymbol{t}} \tag{24}$$

where $p = \sigma_d / bf$ and $q = 2(1 - \cos\delta_\theta)$. The proper threshold τ_{near} should be selected so as to satisfy (24).

IV. SIMULATION STUDY

The proposed two-point algorithm is studied using synthetic data. The numerical accuracy and stability are statistically evaluated in comparison with three-point methods.

A. Setup

In the simulation, reference points in the 3D space are randomly generated within a visible range of cuboid. Considering the properties of camera projection and feature point extraction, the distribution of points is clearly the function of distance. A heavy-tailed distribution is assumed in this simulation through observation of real outdoor data. The points are then projected into 1024×768 (XGA) image planes using the camera pose and camera parameters. Some of the projected points are eliminated with the non-maximum suppression. The image points are perturbed by pixel noise, modeled by zero-mean Gaussian distribution with the standard deviation σ_p which is specified in the experiments. One hundred points are generated for each camera pose.

The camera poses are synthetically generated assuming general outdoor vehicles. The translation of the camera system is specified by unit vectors for forward and sideways motion. The orientation of the camera system is randomly

generated within reasonable bounds. The focal length is set to 900 and the principal point to the center of the image. The baseline of the stereo camera is set to 0.85.

B. Distance vs Estimation Accuracy

Firstly, the impact of distance against estimation accuracy is studied. The orientation estimation is evaluated by the smallest rotation angle of relative rotation between the true and estimate orientation. The translation estimation is evaluated by the euclidean distance between the true and estimate position.

The distant points at infinity, i.e., vanishing points, are not always available in natural scenes. The impact of using points at finite distance is shown in Fig. 7. The error decreases exponentially within the bound given by (21). Therefore, the rotation error can be suppressed given a certain level of estimation accuracy.

On the other hand, the near points should be close to the camera to suppress the growing triangulation uncertainty. The translation error with respect to the distance of near points is shown in Fig. 7. This simulation uses a fixed distant point locating at 100 m distance (=7.65 pixel). The error for most data points increases linearly, while the upper bound is given by the exponential function. It is important to choose sufficiently near points for better accuracy.

In the later experiments, the thresholds are set to 10 to 40 m for near points and $>$100 m for distant points as is otherwise stated.

C. Numerical Stability

The proposed algorithm can be adopted for overdetermined cases. The number of points greater than two can exhibit higher stability in numeric accuracy. Fig. 8 shows the numerical stability of the method with respect to the number of near point correspondences for four different noise levels. The error metric is the Frobenius norm of the estimated and true pose matrices over 10^5 trials.

The straightforward result suggests using more points results in better numerical accuracy. However, the solver with only one near point shows relatively good performance in the presence of noise.

D. Comparison with Three-point Methods

The proposed method is compared with the classic 3D-to-3D least squares estimation [11] and a recent P3P solver

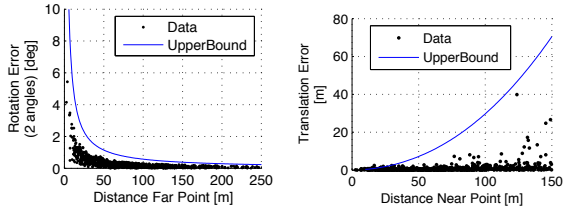


Fig. 7. Estimation error w.r.t. point distance. Left: Rotation error using points at finite distance. Right: Translation error using a near point and a distant point at 100 m. The pixel noise level is set to $\sigma_p = 1.0$.

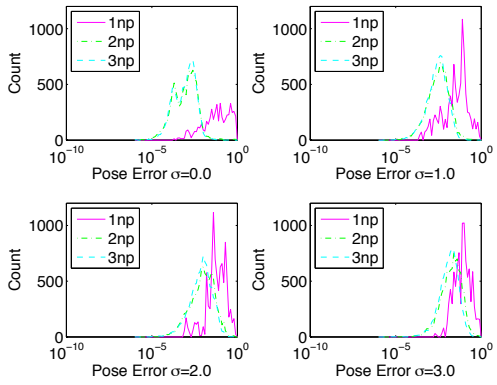
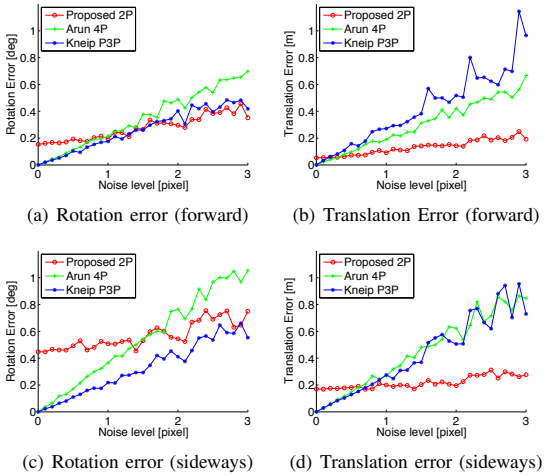


Fig. 8. Distribution of pose error w.r.t. the number of near points at varying pixel noise levels. The error metric is the Frobenius norm of pose matrices over 10^5 trials.



(a) Rotation error (forward) (b) Translation Error (forward)

(c) Rotation error (sideways) (d) Translation error (sideways)

Fig. 9. Method comparison for forward and sideways motion at varying pixel noise levels. (Arun: 3D-to-3D least squares method [11] using four points. Kneip: 3D-to-2D P3P method [10] using three points.)

[10]. Due to the instability of the 3D-to-3D estimation, it uses four points instead of three.

The rotation and translation errors for forward and sideways motion are shown in Fig. 9. Each plot represents the medians over 1000 trials for each noise level. The reason to use the median is that the fraction of good estimates is more important in the RANSAC framework.

Due to the approximation of directional correspondences, the consistent error can be observed in both of the motions. However, it is clearly seen that the proposed method performs stably against the increasing levels of pixel noise compared to the other methods. Specifically, the proposed method outperforms other methods in translation estimation, while it requires less point correspondences.

The reason that the proposed method performs better against pixel noise is that it avoids triangulation uncertainty of distant points. While other methods can also avoid this

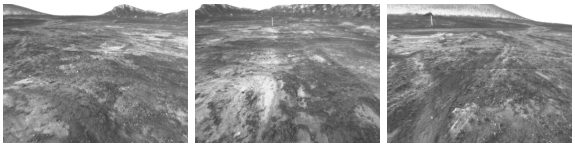


Fig. 10. Typical camera views in a volcanic field

TABLE II
EXPERIMENTAL RESULT

	Time	RMS Error
Arun [11]	0.18 ms	1.061 m (3.16%)
Gao [9]	1.32 ms	2.757 m (8.23%)
Proposed	0.12 ms	1.357 m (4.05%)

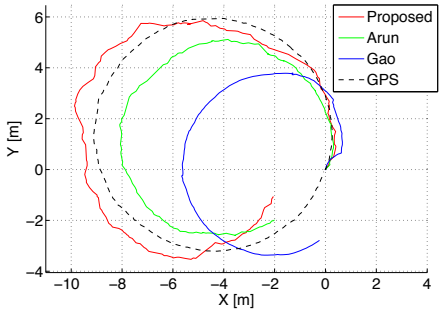


Fig. 11. Estimated trajectories (Arun: 3D-to-3D least squares method [11] using four points. Gao: 3D-to-2D P3P method [9] using three points.)

by filtering points by distance, it may lead excessively small number of feature points in untextured environments, e.g., sandy terrain.

V. EXPERIMENTS

The proposed two-point algorithm is evaluated with the real dataset taken at a volcanic field. A stereo camera rig, which provides 1024×768 grayscale image pairs, is mounted on a four-wheeled vehicle robot. Sample images are shown in Fig. 10.

The algorithm is compared with [11] and [9], all of which are implemented in C++. The trajectories are estimated in Fig. 11 along a 40 m drive on a circular path. The motion is estimated within the RANSAC framework. The two-point algorithm successfully recovers a circular path while it contains small fluctuation in yaw direction. The fluctuation is derived from the approximation using points at finite distance. Even so, this formulation recovers trajectory with sufficiently small RMS error and time (Table II). The remaining error can be eliminated with the optimization procedure such as Bundle Adjustment.

VI. CONCLUSIONS

This paper proposes an efficient method to solve the relative pose problem using only two corresponding 3D points. It gives up to two solutions using one reference direction from a distant point correspondence and one near point correspondence. The experiments using synthetic data show higher stability against image pixel noise, which may be a big problem in real world systems. Furthermore, the algorithm successfully recovers the path from a real video sequence by an outdoor vehicle. This high-speed relative pose solver is suitable for a hypotheses generator of RANSAC systems, which are widely used in robotic localization applications.

REFERENCES

[1] 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.

[2] D. Scaramuzza, "Performance evaluation of 1-point-RANSAC visual odometry," *Journal of Field Robotics*, vol. 28, no. 5, pp. 792–811, 2011.

[3] D. Nistér, "An efficient solution to the five-point relative pose problem," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 6, pp. 756–770, 2004.

[4] 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.

[5] O. Naroditsky, X. S. Zhou, J. Gallier, S. I. Roumeliotis, and K. Daniilidis, "Two efficient solutions for visual odometry using directional correspondence." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 4, pp. 818–824, 2012.

[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," in *European Conference on Computer Vision*, 2010, pp. 269–282.

[7] M. Kalantari, A. Hashemi, F. Jung, and J.-P. Guedon, "A New Solution to the Relative Orientation Problem Using Only 3 Points and the Vertical Direction," *Journal of Mathematical Imaging and Vision*, vol. 39, no. 3, pp. 259–268, 2011.

[8] L. Kneip, M. Chli, and R. Siegwart, "Robust Real-Time Visual Odometry with a Single Camera and an IMU," in *British Machine Vision Conference*, 2011, pp. 16.1–16.11.

[9] X. Gao, X. Hou, J. Tang, and H. Cheng, "Complete solution classification for the perspective-three-point problem," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 930–943, 2003.

[10] L. Kneip, D. Scaramuzza, and R. Siegwart, "A novel parametrization of the perspective-three-point problem for a direct computation of absolute camera position and orientation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 2969–2976.

[11] K. Arun, T. Huang, and S. Blostein, "Least-Squares Fitting of Two 3-D Point Sets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 5, pp. 698–700, 1987.

[12] S. Umeyama, "Least-squares estimation of transformation parameters between two point patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 4, pp. 376–380, 1991.

[13] Y. Cheng, M. W. Maimone, and L. H. Matthies, "Visual odometry on the Mars Exploration Rovers," *IEEE Robotics & Automation Magazine*, pp. 54–62, 2006.

[14] K. Ni and F. Dellaert, "Stereo Tracking and Three-Point/One-Point Algorithms - A Robust Approach in Visual Odometry," in *International Conference on Image Processing*, 2006, pp. 2777–2780.

[15] M. Kaess, K. Ni, and F. Dellaert, "Flow separation for fast and robust stereo odometry," in *IEEE International Conference on Robotics and Automation*, 2009, pp. 3539–3544.

[16] R. I. Hartley and P. Sturm, "Triangulation," *Computer Vision and Image Understanding*, vol. 68, no. 2, pp. 146–157, 1997.

[17] D. Nistér, O. Naroditsky, and J. Bergen, "Visual odometry for ground vehicle applications," *Journal of Field Robotics*, vol. 23, no. 1, pp. 3–20, 2006.

[18] G. Sibley, L. Matthies, and G. Sukhatme, "Bias reduction and filter convergence for long range stereo," *Robotics Research*, 2007.

[19] L. Matthies and S. Shafer, "Error modeling in stereo navigation," *IEEE Journal on Robotics and Automation*, vol. 3, no. 3, pp. 239–248, 1987.

[20] A. J. Lambert, "Visual odometry aided by a sun sensor and an inclinometer," Master Thesis, University of Toronto, 2011.