A Novel Algorithm for Refinement of Vision-based Two-View Pose Estimates

S. S. Mehta, P. Barooah, S. Susca, W. E. Dixon

***Abstract*— A novel method is developed to obtain a refined estimate of relative position and orientation (pose) from two views captured by a calibrated monocular camera. Due to the typically large number of matched pairs of feature points available, many estimates of the pose are possible by taking minimal sets of feature points. Among such estimates, the proposed method selects a subset that has low estimation error, and averages them in an appropriate manner to provide a refined estimate. Preliminary experiments on the image data show that the proposed method provides a more accurate pose estimate than refinement using a least squares method.**

I. INTRODUCTION

Estimation of position and orientation (pose) from two views is an important problem that is relevant to a number of autonomous guidance, navigation, and control applications. For vision-based control as well as visual odometry, the relative motion between a camera and a target is estimated from the images.

Feature points in a sequence of images can be detected and tracked using various methods [1]–[4]. Given a set of matched feature points between the two images several methods can be used for pose estimation such as homography computation [5], [6], fundamental/essential matrix computation [7], [8], trifocal tensor computation, etc. [7]. However, the presence of noise in the matched feature points makes accurate pose estimation a challenging problem. The noise could be due to pixelization of the image, feature localization errors, feature drift, change in illumination, moving objects, etc. The resulting inaccurate feature point matches, typically referred to as *outliers*, could cause large errors in the pose estimate. A common approach to remove outliers is to adopt a hypothesis-and-test framework, such as RANSAC and its variants [9], [10] in conjunction with a algorithm for pose hypothesis generation, such as a 5, 6, 7 or 8 point algorithm [7], [11], [12]. RANSAC provides a pose estimate and a set of matched feature point data, called *inliers*, that is consistent with that estimate. Usually the outlier rejection is followed by a refinement process, in which the feature points in the residual inlier set are used to compute a refined pose estimate. A typical feature-based framework for relative pose estimation therefore consists of three components: feature point correspondence, two-view pose estimation with outlier rejection, and a final pose refinement.

Even for the feature point data without outliers, an accurate pose estimation – the refinement process – is non-trivial, especially when real-time requirements prohibit computationally complex algorithms. A simple approach is to use a least-squares method to perform the refinement. At the heart of the 5, 6, 7, and 8 point algorithms is a least-squares method, since these methods compute the essential matrix by minimizing a quadratic cost function [7], [11], [12]. Therefore, one could provide all the inlier feature points to, say, the 8-point algorithm, and obtain a least squares estimate. However, such an approach may lead to large errors since the least-squares estimates are highly sensitive to image noise [11], [12]. It is noted in [12] that pose estimates become poor when measurement errors enter the input data, and "even the usage of far more data than is minimally required – the approved recipe: 'redundancy combats noise' – does not significantly defuse this situation".

Despite the above mentioned limitation, the least squares method is routinely used for pose refinement due to its simplicity. A more accurate refinement method is the bundle adjustment [13]. However, being a non-linear optimization problem, the bundle adjustment is computationally expensive and also needs a good initial condition.

In this paper we proposed a novel refinement method for estimating the relative pose between two images captured by a calibrated camera. The key idea behind the proposed method is, if there are M matched pairs of feature points between two views, one can compute a maximum of $\binom{M}{P}$ possible estimates of the relative pose by using a P-point algorithm. By considering these estimates as "noisy measurements" of the actual pose, an appropriate average of these measurements should lead to an estimate with higher accuracy than that of the individual measurements. To carry out this idea, a pdf (*probability distribution function*) of the pose is empirically estimated from the measurements, which is used to find a subset of "good" data by selecting the measurements within a small distance of the mode, where the pdf attains its highest value. These estimates are then averaged using a method by Moakher [14] to obtain a refined pose estimate.

Preliminary experimental results demonstrate that the proposed method obtains markedly more accurate estimates than the least squares method. The proposed method can therefore be used as a refinement process after outlier rejection. It can also be used as a method for initial pose refinement to provide a good initial guess for bundle adjustment.

S. S. Mehta, P. Barooah, and W. E. Dixon are with the Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL 32611. S. Susca is with Honeywell Advanced Technologies, Minneapolis, MN

This research is supported in part by the NSF CAREER award 0547448. The authors would like to acknowledge the support of the Department of Energy, grant number DE-FG04-86NE37967. This work was accomplished as part of the DOE University Research Program in Robotics (URPR).

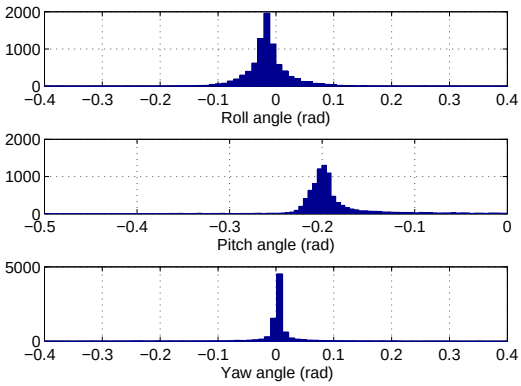


Fig. 1: Histogram of the Euler angle data obtained from rotation measurements $\hat{R}_i$, $i = 1, \ldots, \binom{21}{8}$ from two images.

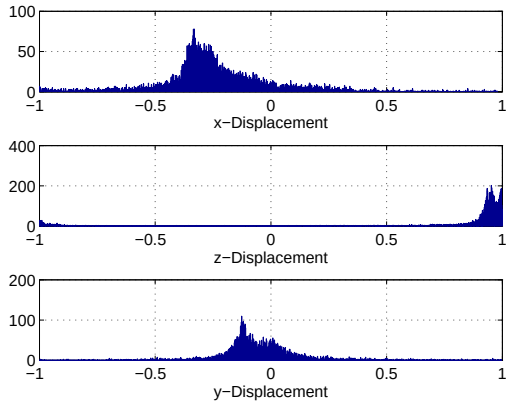


Fig. 2: Histogram of the unit translation data (displacements along x, y, and z-axis) $\hat{t}_i$, $i = 1, \ldots, \binom{21}{8}$ from two images.

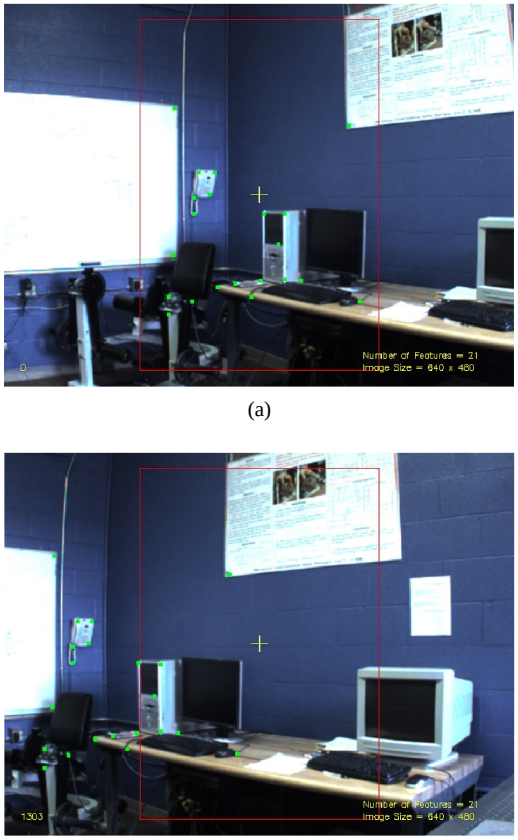
The rest of the paper is organized as follows. Section II explains the approach behind the proposed algorithm. The algorithm is described in detail in Section III. Experimental results from image data are presented in Section IV. Concluding remarks with a discussion on future work is presented in Section V.

II. Problem Statement and Approach

Let R denote the *true rotation* between two views and t be the *true unit translation*, since using a monocular camera the translation can be determined only up to a scale.

If there are M matched pairs of feature points between two views and a P-point algorithm is used for pose estimation, then $n := \binom{M}{P}$ distinct pose estimates are possible, which we denote as $\hat{R}_i, \hat{t}_i, i = 1, \ldots, n$. Each $\hat{R}_i$ and $\hat{t}_i$ are "noisy measurements" of the true rotation R and true unit translation t, respectively. Since typically M is large, there exists a large number of noisy measurements of the true rotation and unit translation.

As an example, Fig. 1 shows a histogram of the Euler angles obtained from the $\hat{R}_i$'s and Fig. 2 shows a histogram



(a)

(b)

Fig. 3: (a) Initial image and (b) final image in the sequence of images taken by a moving camera with the matching points between the images.

of the unit translations $\hat{t}_i$'s along x, y, and z-directions between two calibrated views that are shown in Fig. 3.

Our approach is to appropriately "average" these noisy measurements to construct estimates of R and t with a much lower error than any of the measurements themselves. There are several, but two main, difficulties in carrying out this idea. The first difficulty is that some of the measurements may have very large noise – as seen from the Figs. 1 and 2 – so averaging all of the measurements may not offer any improvement in accuracy. The wide distribution of the Euler angles and unit translations is a possible explanation of why a direct least squares approach may still result in a poor estimate. The second difficulty is that since the group of rotations (and unit translations) is not an Euclidean space but rather a manifold, it is not clear how to average rotations and unit translations.

The rotation R is a random variable, with a pdf (probability distribution function) f defined over an appropriate probability space. In particular, the n possible measurements $\hat{R}_i$ are samples drawn from the distribution f. The pdf can be empirically estimated from the samples $\hat{R}_i$, $i = 1, \ldots, n$, or, in practice, a subset $i = 1, \ldots, N$ where $N \leqslant n$. A measurement $\hat{R}_i$ of the rotation R obtained from feature points with low levels of noise will have higher accuracy, so

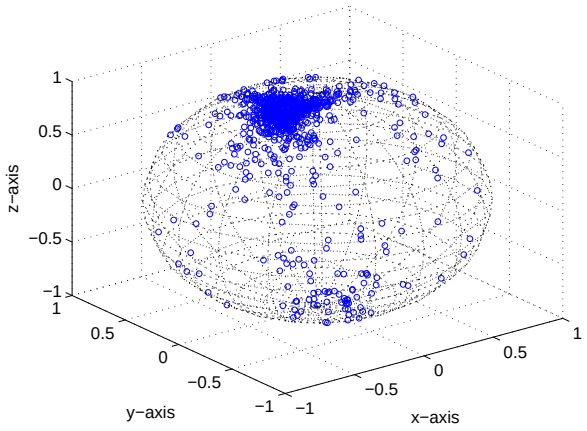


Fig. 4: $\binom{21}{8}$ unit translations, $\hat{t}_i$, estimated from two views with 21 feature points each using an 8-point algorithm projected on a unit sphere in $\mathbb{R}^3$.

that it will be closer to the true value R than the estimates obtained from feature points with larger noise. Assuming that noise is random so that the inaccurate estimates are widely distributed, one can expect the empirically estimated pdf of R to have a *mode* near its true value. A subset Q of the measurements can be identified such that the elements of Q are close to the mode of the empirically estimated pdf. By the argument presented above, these measurements are close to the true estimates R. Averaging the measurements in Q is likely to result in an accurate estimate of R. The second issue of how to average rotations is addressed by using a technique developed by Moakher in [14], which guarantees that the resulting average is a rotation matrix and is closest to all the data in Q in a certain sense.

To facilitate the pdf estimation, the measurements of rotations are expressed in terms of their unit-quaternion representation. Every measurement $\hat{q}_i := q(\hat{R}_i)$ can now be thought of as a unit vector in $\mathbb{R}^4$, where $q(\hat{R}_i)$ denotes a unit quaternion of the rotation $\hat{R}_i$. That is, $\hat{q}_i \in \mathbb{S}^3$, where $\mathbb{S}^3$ (the so-called 3-sphere) is the unit sphere in $\mathbb{R}^4$. The pdf is empirically estimated using the spherical data of the unit-quaternions $\hat{q}_i$.

The same procedure can be applied to the unit translation data. Fig. 4 shows the distribution of the unit translation data, $\hat{t}_i$, on a 2-sphere $\mathbb{S}^2$.

III. Proposed Algorithm

The input to the proposed algorithm is a set of M matched feature points between the two views and a P-point algorithm is used for pose estimation. The pose refinement algorithm is summarized in Algorithm 1 and the steps are detailed in the following sections.

A. Step 1 & 2: Computing N rotations and unit translations

Given M matched pairs of feature points between two images and using a P-point pose estimation algorithm, $n = \binom{M}{P}$ pose estimates can be obtained. Typically, n would be large, hence $N \leqslant n$ is chosen based on computational constraints. The N pose estimates, $\hat{R}_i, \hat{t}_i, i = 1, \ldots, N$, are the samples of the random variables R and t (rotation and unit translation), respectively.

Algorithm

1. Randomly select N samples of P-points from available M feature points such that $N \leqslant \binom{M}{P}$.
2. Compute N rotations and unit translation vectors using a normalized P-point algorithm.

Rotation Estimation

3. Obtain a set $D_q = \{\hat{q}_i \in \mathbb{S}^3 | i = 1, \ldots, N\}$ of unit quaternions representations of the N rotation estimates.
4. Project the quaternions on a $\mathbb{S}^2$ using stereographic projection. Empirically estimate the mode of the distribution of rotation measurements on $\mathbb{S}^2$. The mode is denoted by $\bar{\hat{q}}$.
5. Average the rotations that fall within a Riemannian distance of ε_q from the mode $\bar{\hat{q}}$ using the method of [14].

Unit Translation Estimation

6. Obtain a set $D_t = \{\hat{t}_i \in \mathbb{S}^2 | i = 1, \ldots, N\}$ of unit translation estimates.
7. Empirically estimate the mode of the distribution of unit translation measurements on $\mathbb{S}^2$. The mode is denoted by $\bar{\hat{t}}$.
8. Average the unit translations that fall within a spherical distance of ε_t from the mode $\bar{\hat{t}}$, and project the average onto $\mathbb{S}^2$.

Algorithm 1: A pose refinement algorithm using M matched feature points between two images.

B. Step 3 & 4: Estimating mode of quaternion distribution

The rotations $\hat{R}_i \in \mathbb{R}^{3\times3}, i = 1, \ldots, N$ can be represented by the unit quaternions $\hat{q}_i \in \mathbb{S}^3, i = 1, \ldots, N$. A histogram based density estimator is used to empirically estimate the pdf of the quaternion q from the measurements $\hat{q}_i$. The choice for the histogram based estimator, instead of other estimators such as a Kernel-based one, is due to its simplicity.

1) Review of histogram estimator: Suppose $X_i, i = 1, \ldots, N$ are N samples of a random variable X with a distribution f with support S. Let S_j, $j = 1, \ldots, K$ be a partition of the domain S. We denote by p_j the probability of X lying in S_j:

$$p_j = P(X \in S_j) = \int_{S_j} f(X)dX, \tag{1}$$

where $P(\cdot)$ denotes probability. Then, p is a pmf (probability mass function). An estimate of p_j can be computed from a histogram as follows:

$$\hat{p}_j = \frac{\#\text{ of } X_i\text{'s in } S_j}{\text{total }\#\text{ of } X_i\text{'s}} = \frac{1}{N}\sum_{i=1}^{N} I_{S_j}(x_i), \tag{2}$$

where $I_{Sj}(\cdot)$ is the indicator function of the set S_j:

$$I_{S_j}(X_i) = \begin{cases} 1 & \text{if } X_i \in S_j \\ 0 & \text{otherwise .} \end{cases} \tag{3}$$

The key point about the histogram estimator is that $\hat{p}_j$ is an unbiased estimator of p_j even if the samples X_i are correlated.

2) Density estimation of noisy rotations: The density estimation consists of two steps: (a) projection from 4-dimensions (4-D) to 3-dimensions (3-D) and (b) density estimation in 3-D. This results in the identification of $Q \subset D$ that consists of data that lies close to q.

a) Projection onto 3-D: In principle, one can form a partition of the "surface" of $\mathbb{S}^3$ into cells and count the number of data points $\hat{q}_i$ to compute the histogram estimates. However, such calculations in 4-D are purely algorithmic and are difficult to visualize. Therefore the 4-D quaternion data is projected onto the 3-D surface passing through the equator of $\mathbb{S}^3$. This simplifies the computations as well as visualization. At the same time, accuracy of the estimate is not compromised, since the purpose of the estimator is merely to identify those data points that are close to the mode.

A stereographic projection is used for projecting the 4-D quaternion data onto a 3-D sphere [15]. Specifically, a point with coordinate x, y, z, w in $\mathbb{R}^4$ is projected into a point in $\mathbb{R}^3$ with coordinates x', y', z' by the stereographic projection

$$x' = \frac{x}{1-w}, \qquad y' = \frac{y}{1-w}, \qquad z' = \frac{z}{1-w}. \tag{4}$$

Using the quaternion properties, it can be ensured that every unit quaternion data $\hat{q}_i$ lies in the lower "hemisphere" of $\mathbb{S}^3$. Note that corresponding to every rotation matrix R, there are two possible quaternions, since q and $-q$ express the same rotation. Among these two, a quaternion that has a negative 4-th coordinate is selected to ensure that it lies in the lower hemisphere of $\mathbb{S}^3$. The quaternion q is now imagined as a point in $\mathbb{R}^4$, which is then projected into a point $\underline{q}$ in $\mathbb{R}^3$ using the stereographic projection (4). Since q lies in the lower hemisphere, $\underline{q}$ lies inside the unit sphere in $\mathbb{R}^3$, namely $\mathbb{S}^2$ (see Fig. 5), which can be verified from (4) as follows:

$$\begin{aligned} x'^2 + y'^2 + z'^2 &= \frac{x^2+y^2+z^2}{(1-w)^2} \\ &= \frac{1-w^2}{(1-w)^2} = \frac{1+w}{(1-w)} \leq 1, \end{aligned} \tag{5}$$

where the inequality results from the fact that $-1 \leq w \leq 0$.

b) Density estimation in 3-D: The volume of the 2-sphere $\mathbb{S}^2$ is divided into a number of cells of equal volume V_j, $j = 1, \ldots, K_{\mathrm{q}}$, where K_{q} is the (appropriately chosen) number of cells. A histogram estimate of the pmf $p_j = \mathrm{P}(\underline{q} \in V_j)$ can be computed by counting the number of points $\underline{q}_i$ inside V_j. The mode of the distribution is the value of q at which the pmf takes the maximum value, which is denoted as $\bar{\hat{q}}$.

C. Step 5: Computing average of low-noise quaternion data

a) Extracting low-noise measurements: A sample X_i drawn from a distribution with mode $\bar{X}$ is classified as a

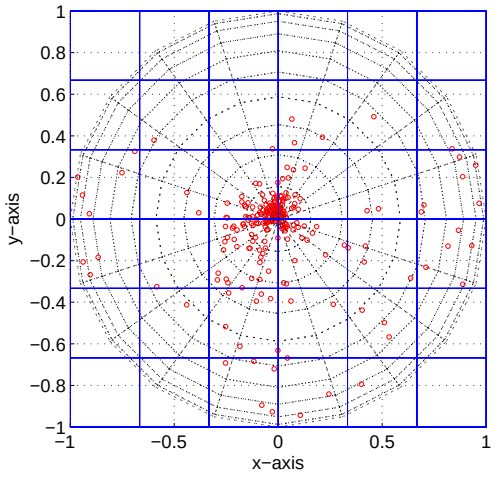


Fig. 5: A slice of the $\mathbb{S}^2$ sphere showing the projected quaternion data on a 3-D sphere along with the cells.

low-noise sample[1] with respect to the mode $\bar{X}$ if

$$d(\bar{X}, X_i) < \epsilon, \tag{6}$$

where $d(\cdot, \cdot)$ represents a suitable distance, and ϵ is an appropriately chosen positive scalar (a design parameter). Once the mode of the rotation q is identified as described in Section III-B, the low-noise data set Q_q is selected by choosing all those $\hat{q}_i \in D_q$ that satisfies

$$d_q(\bar{\hat{q}}, \hat{q}_i) < \varepsilon_q, \tag{7}$$

where the Riemannian distance between the rotation matrices $R(\bar{\hat{q}})$ and $R(\hat{q}_i)$ is used for the metric $d_q(\cdot)$. The Riemannian distance between two matrices $R_1, R_2 \in SO(3)$ is given by

$$d(R_1, R_2) = \frac{1}{\sqrt{2}} \| \log(R_1^T R_2) \|_F, \tag{8}$$

where the subscript F refers to the Frobenius norm. The principal logarithm $\log(R)$ of a matrix $R \in SO(3)$ is a matrix in a skew-symmetric matrix space $so(3)$ given by

$$\log(R) = \begin{cases} 0 & \text{if } \theta = 0 \\ \frac{\theta}{2\sin\theta}(R - R^T) & \text{if } \theta \neq 0, \end{cases} \tag{9}$$

where θ corresponds to the angle of the Euler axis and angle representation, which can be computed from

$$tr(R) = 1 + 2cos(\theta) \quad \text{and} \quad |\theta| < \pi \tag{10}$$

b) Averaging low-noise data: Let N_1 be the number of elements in the low-noise data set Q_q of rotations obtained as described above. The *optimal* mean of $R_1 \ldots R_{N_1}$ in the Euclidean sense is given by [14]

$$\bar{\hat{R}} = \underset{R \in SO(3)}{\operatorname{argmin}} \sum_{i=1}^{N_1} ||R_i - R||^2{}_F, \tag{11}$$

[1]Perhaps "inlier" is more appropriate; though we refrain from using that to avoid confusion with accepted terminology.

in which R_i denotes the rotation matrix corresponding to $\hat{q}_i \in D_q$. The rotation specified by (11) can be computed by the orthogonal projection of the arithmetic average $\bar{R} = \sum_{i=1}^{N_1} \frac{R_i}{N_1}$ onto the special orthogonal group $SO(3)$, as

$$\bar{\hat{R}} = \bar{R}Udiag(\frac{1}{\sqrt{\Lambda_1}}, \frac{1}{\sqrt{\Lambda_2}}, \frac{s}{\sqrt{\Lambda_3}})U^T, \tag{12}$$

where the orthogonal matrix U is such that

$$\bar{R}^T\bar{R} = U^T DU \text{ and } D = diag(\Lambda_1, \Lambda_2, \Lambda_3). \tag{13}$$

In (12), $s = 1$ if $det\bar{R} > 0$ and $s = -1$ otherwise.

The matrix $\bar{\hat{R}}$ computed using (12) is the desired estimate of R.

D. Step 6 & 7: Estimating mode of unit translation distribution

The estimation scheme for unit translation is very similar to that for the rotation. The unit translation data $\hat{t}_i, i = 1, \ldots, n$ are 3-D samples drawn from the distribution of the true unit translation t which represents points on $\mathbb{S}^2$, as shown in Fig. 4.

The surface of the $\mathbb{S}^2$ is divided into a number of regions of equal area [16] A_j, $j = 1, \ldots, K_\mathrm{t}$, where K_t is the (appropriately chosen) number of regions. A histogram estimate of the pmf $p_j = \mathrm{P}(\hat{t} \in A_j)$ can be computed by counting the number of points $\hat{t}_i$ inside A_j.

The mode of the unit translation distribution is the value of t at which the pmf takes its maximum value, which is denoted as $\bar{\hat{t}}$.

E. Step 8: Computing average of low-noise unit translation data

Once the mode $\bar{\hat{t}}$ of the unit translation t is identified, as described in Section III-D, the low-noise data set Q_t is selected by choosing all those $\hat{t}_i \in D_t$ that satisfies

$$d_t(\bar{\hat{t}}, \hat{t}_i) < \varepsilon_t, \tag{14}$$

where the spherical distance between the unit translation vectors $\bar{\hat{t}}$ and $\hat{t}_i$ is used for the metric $d_t(\cdot)$.

Let, N_2 be the number of elements in the low-noise data set Q_t of the unit translations obtained as described above. The mean of the unit translations in the set Q_t can be obtained as follows:

$$\bar{t} = \frac{\sum_{i=1}^{N_2} \frac{t_i}{N_2}}{\left|\left|\sum_{i=1}^{N_2} \frac{t_i}{N_2}\right|\right|}, \tag{15}$$

which is the desired estimate of t.

IV. Performance evaluation

The performance of the proposed algorithm is evaluated by testing it on video images captured by a moving monocular camera and then comparing it to the ground truth. The pose estimation results are then compared with the least squares method by computing the pmf of the estimation error.

A monocular camera (make: Matrix Vision GmbH, model: mvBlueFox-120aC) mounted on a 2-link planar manipulator is utilized to record a set of images of a stationary target. The 2-link manipulator is equipped with rotary encoders on each motor with 614,400 readings/revolution. Fig. 3 shows the initial image before the camera motion and the final image of the target after the camera is displaced through a rotation R and translation t.

A multi-scale version of Kanade-Lucas-Tomasi (KLT) feature point tracker is implemented to track the features in successive images as the camera undergoes rotation and translation. 21-feature point correspondences are tracked in the images obtained by the camera. A normalized 8-point algorithm is used to compute a fundamental matrix which can be decomposed to obtain the relative pose between the two frames. From a total $n = \binom{21}{8} = 203490$ possible measurements of rotations and unit translation, $N = 500$ randomly selected measurements are used for pose refinement.

The objective is to estimate rotation and unit translation between the views shown in Fig. 3. The 3-D sphere is then divided into $K_\mathrm{q} = 6\times6\times6$ equal volume cells to estimate the mode of the pdf of q. The design parameter ϵ_q was selected to 0.0121. Fig. 5 shows a slice of the 3-D sphere with the cells for determining the histogram of the quaternion data. The unit translation data on the surface of a 3-D sphere (as shown in Fig. 4) is divided into $K_\mathrm{t} = 7$ equal area regions to estimate the mode of the pdf of t. An estimate of the unit translation is obtained by averaging the unit translation data within $\epsilon_t = 0.0166$ from the mode of pdf.

A. Results

The estimate of the rotation and unit translation obtained by the proposed method, denoted by $\bar{\hat{R}}$ and $\bar{\hat{t}}$, respectively, are

$$\bar{\hat{R}} = \begin{bmatrix} 0.9811 & 0.0223 & -0.1924 \\ -0.0229 & 0.9997 & -0.0009 \\ 0.1924 & 0.0053 & 0.9813 \end{bmatrix} \tag{16}$$
$$\bar{\hat{t}} = [\ -0.2687 \quad -0.0812 \quad 0.9598\ ]^T.$$

The least squares method using the normalized 8-point algorithm yields the following estimate of the rotation and unit translation vector when used with all the 21 pairs of matched feature points:

$$R_{LS} = \begin{bmatrix} 0.9863 & 0.0085 & -0.1646 \\ -0.0104 & 0.9999 & -0.0104 \\ 0.1645 & 0.0119 & 0.9863 \end{bmatrix} \tag{17}$$
$$t_{LS} = [\ -0.5039 \quad 0.0809 \quad 0.8600\ ]^T.$$

The true pose obtained from the high precision rotary encoders mounted on the 2-link planar manipulator is

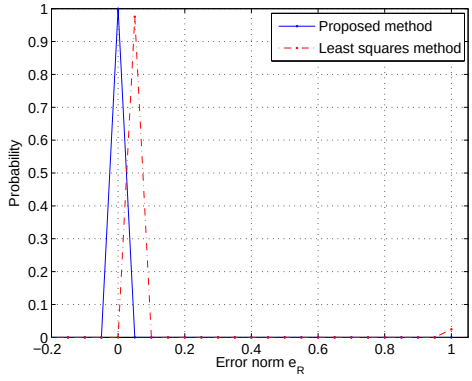
$$R = \begin{bmatrix} 0.9790 & 0 & -0.2036 \\ 0 & 1 & 0 \\ 0.2036 & 0 & 0.9790 \end{bmatrix} \tag{18}$$
$$t = [\ -0.2534 \quad 0 \quad 0.9673\ ]^T.$$

For a clearer performance comparison of the proposed method with the least-squares refinement, we estimate rotations and translations between a number of distinct pairs

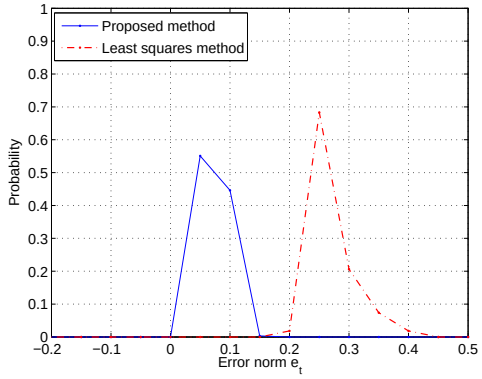
of images from the captured image sequence. Each such image pair is indexed from $i = 1, \ldots, 325$ and the rotation and translation between the frames in the i-th image pair are denoted as $R(i)$ and $t(i)$, respectively. The rotation and translation error norms for the image pair i, defined as e_R and e_t, respectively, are given below

$$e_R(i) = \|I - R(i)^T \hat{R}(i)\|, \quad e_t(i) = \|t(i) - \hat{t}(i)\|, \quad (19)$$

where $\hat{R}(i)$ and $\hat{t}(i)$ are the estimates of the rotation $R(i)$ and unit translation $t(i)$, and I denotes a $\mathbb{R}^{3\times3}$ identity matrix. Encoder measurements are used to determine the ground truth $R(i), t(i)$. The pmf of the rotation and translation error norms for both the methods are shown in Fig. 6a and Fig. 6b, respectively.



(a)



(b)

Fig. 6: Comparison of the pmf of the (a) rotation error norm and (b) unit translation error norm for the proposed method and least squares method using 325 image pairs.

Results shown in Fig. 6 indicate that the pose estimates generated using the proposed method have a higher probability of less error norm as compared to the least squares method. The mean and standard deviation of the rotation and translation error norms are tabulated in Table I. It can be seen from Table I that the mean and standard deviation of error using the proposed method are less than the least squares method.

Algorithm	Rotation Error Norm		Translation Error Norm	
	Mean	Std. Dev.	Mean	Std. Dev.
A	0.0094	0.0023	0.0852	0.0237
B	0.0739	0.2646	0.2671	0.0377

TABLE I: Statistical comparison between (A) proposed method and (B) least squares method in terms of rotation and translation estimation error norms.

V. Conclusion

A novel method is presented for refining relative pose estimates from two calibrated views with feature point correspondence. Experimental results indicate markedly better performance over the least squares refinement. Another advantage of the algorithm is its simplicity and hence its suitability in real-time applications. There are several avenues of future work that need to be explored, such as comparison of the developed algorithm with bundle adjustment in speed-accuracy trade-off, performance in robust outlier rejection, and comparison with RANSAC-based robust pose estimation methods.

References

[1] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proceedings of the 7th International Conference on Artificial Intelligence*, 1981, pp. 674–679.

[2] J. Shi and C. Tomasi, "Good features to track," in *IEEE Conference on Computer Vision and Pattern Recognition*, 1994, pp. 593–600.

[3] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, p. 91110, 2004.

[4] H. Bay, T. Tuytelaars, and L. V. Gool, "Surf: Speeded up robust features," in *Proceedings of European Conference on Computer Vision*, 2006, pp. 404–417.

[5] O. Faugeras, *Three-Dimensional Computer Vision*. The MIT Press, Cambridge Massachusetts, 2001.

[6] E. Malis, F. Chaumette, and S. Boudet, "2 1/2 d visual servoing," *IEEE Transactions on Robotics and Automation*, vol. 15, pp. 238–250, 1999.

[7] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004.

[8] 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, June 2004.

[9] 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, pp. 381–395, 1981.

[10] P. H. S. Torr and A. Zisserman, "Mlesac: a new robust estimator with application to estimating image geometry," *Computer Vision and Image Understanding*, vol. 78, no. 1, pp. 138–156, 2000.

[11] W. Chojnacki and M. J. Brooks, "On the consistency of the normalized eight-point algorithm," *Journal of Mathematical Imaging and Vision*, vol. 28, no. 1, pp. 19–27, May 2007.

[12] M. Mhlich and R. Mester, "The role of total least squares in motion analysis," in *In Proceedings of European Conference on Computer Vision*. Springer, 1998, pp. 305–321.

[13] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon, "Bundle adjustment a modern synthesis," *Vision Algorithms: Theory and Practice*, 1999.

[14] M. Moakher, "Means and averaging in the group of rotations," *SIAM Journal on Matrix Analysis and Applications*, vol. 24, 2002.

[15] N. I. Fisher, T. Lewis, and B. J. J. Embleton, *Statistical Analysis of Spherical Data*. Cambridge University Press, 1997.

[16] P. Leopardi, "A partition of the unit sphere into regions of equal area and small diameter," *Electronic Transactions on Numerical Analysis*, vol. 25, pp. 309–327, 2006.