UNSYS digital

International Journal of Robotics and Mechatronics

Monocular Vision Based Simultaneous Localization and Mapping (SLAM) Technique for UAV Platforms in GPS-denied Environments

Harsha Subramanya
Point Parking, Sydney, Australia

***Abstract*—**For Unmanned Air Vehicles (UAV) to operate autonomously in unstructured and GPS-denied environments, requisite information about the surroundings with adequate detailing needs to be generated and Simultaneous Localization and Mapping (SLAM) is among the most preferred protocols for fulfilling these needs. This study presents a SLAM-based linear optimal control approach for UAV navigation with limited sensor resources. The suggested protocol utilizes 1-Point RANSAC and an Extended Kalman Filter (EKF) for SLAM from a 6 degree-of-freedom motion monocular image sequence. Output of this research effort includes the estimated camera motion and a sparse map of salient point features through sensory representation. In this study we present a unique combination of 1-Point RANSAC (Random Sample Consensus) in conjunction with EKF with innate focus on reducing of the computational complexity. Contemporary studies by UAV research groups have successfully demonstrated the usefulness of algorithms based on 1-Point RANSAC and also provided comparison of algorithmic results with those of visual odometry. The present study further extends the scope by evaluating the algorithm with data generated from an input device mounted on a custom developed UAV. Employment of SLAM mission for mapping the areas that are prone to mining and land exploration is among the important outcomes of the proposed study.

***Keywords*—**Efficient algorithms, Unmanned Air Vehicles, RANSAC, filters, mapping, localization.

I. INTRODUCTION

In the field of mobile robotics, one of the prime objectives is the development of an autonomous robot capable of operating in real world situations. A pre-requisite to such a system is the ability to answer characteristic questions about its position (pose) and the movement with respect to the environment. Building of global maps with spatial consistency of the environment is essential to compute courses or paths and to enable mission control.

Micro aerial vehicles—and particularly multi-rotor helicopters—have several advantages compared to fixed-wing micro aerial vehicles: they are able to take off and land vertically, hover on a spot, and even dock to a surface. This capability allows them easily to work in small indoor environments, pass through windows, traverse narrow corridors, and even grasp small objects. For mechatronics research, quadrotors are a popular platform because of their maneuverability, stability, and comparatively low mechanical intricacy. An Extended Kalman Filter (EKF) was successfully used for real-time SLAM on a quadrotor platform, which avoids the problems associated with least-squares method altogether [1].

A vital problem in aerial-vehicle navigation is the stabilization and control in six degrees of freedom (6 DOF), that is, attitude and position control. Over time, drifting becomes a major issue with regards to navigation accuracy. In indoor and GPS denied environments, motion capture systems, laser range finders, cameras etc. can be used to address the issues of navigation, localization and mapping.

In this paper, we present a modular implementation of a SLAM system. We discuss the effectiveness of 1-Point Random Sample Consensus (RANSAC) over standard RANSAC algorithm and its integration with an Extended Kalman Filter (EKF) with innate focus on reducing the computational complexity. The developed system is tested on board an AR parrot drone in various datasets and real world simulations.

II. BACKGROUND AND MOTIVATION

The approach used in this paper has been motivated by the research work of Civera et al. [2] and their demonstration of the viability and effectiveness of the application of 1-Point RANSAC to solve the SLAM problem. Additionally, the works of Scaramuzza et al. [3] on the real time monocular visual odometry of a vehicle over 3 km of urban area and the use of 1-Point RANSAC and histogram voting for the removal of outliers has provided a foundation for the work carried out.

RANSAC was first introduced to visual geometry in 1993 [4] and has since become a standard outlier elimination tool. The early detection and elimination of bad data sets have since been the focus of several researchers including [5] and [6]. Steder [7] presented a system to study large visual maps of the ground using flying vehicles. The set up contained an inertial sensor and a low-quality camera. Steder formulates the SLAM

Corresponding author: Harsha S. (e-mail: harshasubramanya@y7mail.com).
This paper was submitted on February 7, 2014; revised on October 22, 2015; and accepted on November 25, 2015.

problem with a graphs based approach with the nodes of the graph describing the pose of the vehicle. A combinatorial usage of RANSAC and EKF filter was proposed by [8].

Civera et al. proposes a more efficient method with the hypotheses generator size set definitively to one. Additionally the EKF algorithm is described for application in real time by splitting the computationally expensive EKF covariance update in two stages. This builds upon the more restrictive models like [3] and thereby offers better motion estimation. Further information for the predicted camera motion comes from the probability distribution function that the EKF naturally propagates over time. Therefore, this method is suitable for use in 6 degrees of freedom estimation. Furthermore the application of SLAM technique to AR drone has been successfully demonstrated by Nick Dijkshoorn [9].

III. VISION BASED SLAM

In the absence of an initial map or an outright localization means, a mobile robot needs to concomitantly solve the localization and mapping problems. For this desired outcome, vision based sensing is an appropriate protocol, since it offers data from which stable features/landmarks can be extracted and matched as the robot moves.

The implementation of a feature-based SLAM approach involves four functional requirements namely environment features selection, relative measures estimation, data association and estimation [10]. Vision based sensors are also particularly versatile due to their applicability in various other areas of mobile robotics such as obstacle detection and target tracking, whilst being compact enough to be mounted on unmanned aerial vehicles and having low computational complexity.

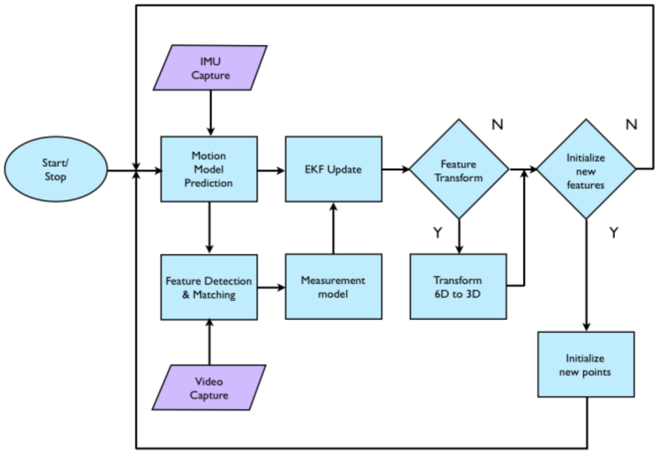


Figure 1 Flowchart of a typical monocular SLAM system [1]

The formation of consistent correspondences from the vision based sensor data is fundamental for most approximation algorithms in robotics. Therefore data association or feature matching is generally based on likening local descriptors of salient features in the sensor data. The presence of outliers in the sensor data and their accumulation over time leads to inaccuracies and inconsistencies. Robust methods operate by checking the consistency of the data against the global model assumed to be generating the data, and discarding as spurious any that does not fit into it [2]. Random Sample Consensus (RANSAC) [11] stands out as one of the most successful and widely used, especially in the Computer Vision community. Figure 1 shows a flowchart illustrating the typical SLAM system used in conjunction with an MAV.

The formation of consistent correspondences from the vision based sensor data is fundamental for most approximation algorithms in robotics. Therefore data association or feature matching is generally based on likening local descriptors of salient features in the sensor data. The presence of outliers in the sensor data and their accumulation over time leads to inaccuracies and inconsistencies. Robust methods operate by checking the consistency of the data against the global model assumed to be generating the data, and discarding as spurious any that does not fit into it [2]. Random Sample Consensus (RANSAC) [11] stands out as one of the most successful and widely used, especially in the Computer Vision community. Figure 1 shows a flowchart illustrating the typical SLAM system used in conjunction with an MAV.

A. *Standard RANSAC*

RANSAC has been established as the standard method for model estimation in the presence of outliers. RANSAC works by generating model hypothesis from an arbitrarily sampled minimal data set and substantiating it on the unabridged data set. It is an iterative method to estimate parameters of a mathematical model from a set of observed data which contains outliers. It is a non-deterministic algorithm in the sense that it produces a reasonable result only with a certain probability, with this probability increasing as more iterations are allowed. Hypotheses are randomly generated based on the minimum number of points necessary to calculate the model parameters, which, in the case of line estimation, is two. Provision for each hypothesis can be calculated by counting the data points inside a threshold, although more sophisticated methods have been used [12]. The number of hypothesis (iterations) N that is necessary to guarantee that a correct solution is found can be computed by

$$N_{hyp} = \frac{\log(1-p)}{\log(1-(1-\in)^{s})} \qquad (1)$$

Where s is the number of minimal data points, ε is the percentage of outliers in the data points and *p* is the requested probability of success [8]. Since *N* is exponential in data points needed for accurate modelling, it is imperative to find the minimal parameterization of the model.

Data sets/points that elected for the most supported hypothesis are considered clear inliers, as is seen in figure 2. In a second step, these inliers are used to evaluate the model parameters. Specific compatibility is checked for each one of the rest of the points against the evaluated model. If any of them is recognized as inlier, the model parameters are re-evaluated again in a third step [2].

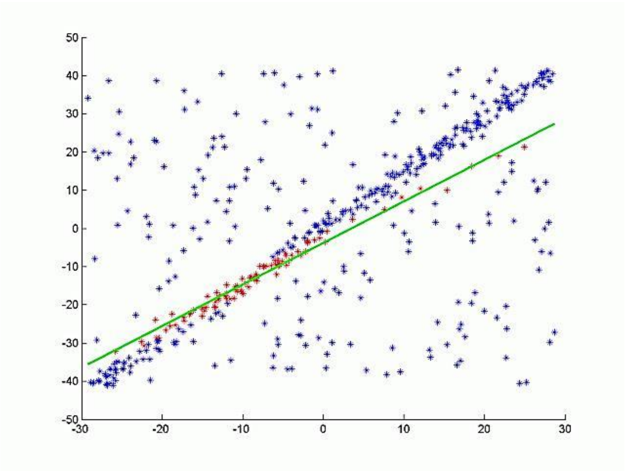


Figure 2 RANSAC algorithm

B. 1-Point RANSAC

The crucial difference in 1-Point RANSAC is that the starting point is a data set and its underlying model, along with a previous probability distribution over the model parameters.

TABLE I ITERATION COMPARISON BETWEEN MINIMUM NUMBER OF DATA POINTS FOR RANSAC

Min. data points:	6 points	5 points	2 points	1point
No. of iterations:	292	145	16	7

As can be seen in **TABLE 1**, the number iterations required based on Equation 1 increase greatly with the number of minimum data of points. Since the aim here is to create an approach suitable for affordable applicability on aerial platforms, it is important to minimize the parameterization model i.e. 1-Point RANSAC. Using a single feature correspondence for motion estimation is the lowest model parameterization possible and results in the most efficient RANSAC algorithm [2]. The use of prior information reduces the size of the data set that reverts the model to the minimum size of one point. This aspect makes 1-Point RANSAC computationally less intensive and therefore best suited for the needs of this paper. Since it is unusual to have prior probability information in classical Structure from Motion (SfM), due to assumed widely separated views [13], methods like standard RANSAC are mandatory. But in sequential SfM from video [5], [14], [15] smooth inter frame camera motion can be assumed within limit and this could be used to generate a prior probabilistic information of the motion. For the specific EKF implementation of sequential SfM used in this paper, this prior probability is naturally propagated by the filter developed by a research team at Universidad de Zaragoza, Spain [2].

C. Extended Kalman Filter based Monocular SLAM

A typical EKF based monocular SLAM system is first given by Andrew. J. Davison [5]. Generally for most visual SLAM systems, the camera state is comprised of position, rotation, velocity and angular velocity of the camera.

The camera state is given by

$$X_c = (p^T, r^T, v^T, \omega^T)^T \tag{2}$$

The feature state is given by

$$f_i = (x_i, y_i, z_i)^T \tag{3}$$

Since the AR drone comes with an Inertial Measurement Unit (IMU), we can use the acceleration model to get one step prediction of the EKF state [16].

$$x_{c,t+1} = \begin{pmatrix} x_{t+1} \\ r_{t+1} \\ v_{t+1} \end{pmatrix} = \begin{pmatrix} p_t + v_t\Delta t + \frac{1}{2(a_t+n_a)\Delta t^2} \\ r_t^w * r\big((w_t + n_w)\Delta t\big) \\ v_t(a_t + n_a)\Delta t \end{pmatrix} \tag{4}$$

D. Aerial Platform – Parrot Drone

The Parrot AR.Drone (quadrotor) was selected after taking into account its robustness, affordability, low mechanical complexity and superior indoor and outdoor maneuverability. The AR.Drone is equipped with a front-camera and a down-looking camera that provide live video streaming. Two complementary algorithms make use of the down-looking camera for enhanced stabilization. Also, the AR.Drone is equipped with an ultrasound sensor and an IMU that measures pitch, roll, yaw and accelerations along all axes. The drone is controlled by sending instructions over a Wi-Fi connection.

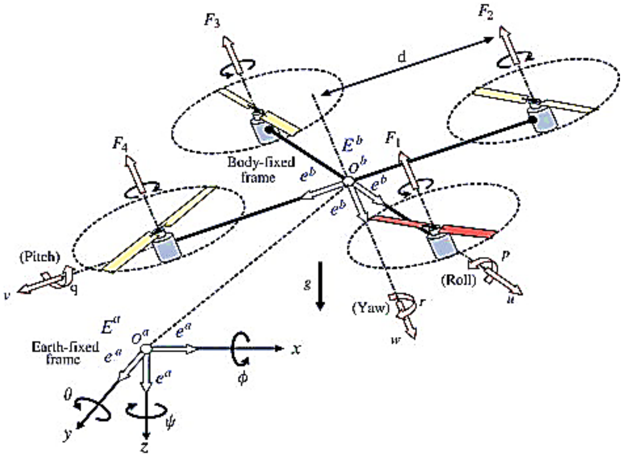
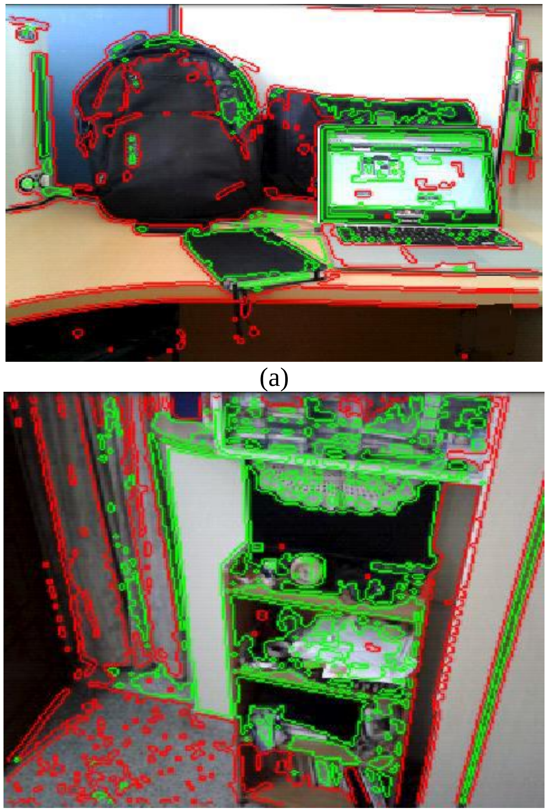


Figure 3 Free body diagram of quadrotor [17]

The sensors onboard the Drone provide the information for usage in the Extended Kalman filter in order to estimate the current pose. The new estimate is calculated based on (4). Once the Drone's pose has been estimated, a map can be generated to aid localization and to ease the error rate in pose estimation. A feature map can be generated to aid in the localization of the Drone and a Visual map for human understanding [9].

IV. RESULTS

To begin with, the performance of 5-point and 1-point RANSAC was compared, in order to confirm that there was no deterioration of performance associated with the reduced sample size. It can be observed from the figures 5 and 6 that no substantial effect on the camera displacement or the consistency of estimation was noticed despite the reduced sample size. The figure shows 1-point outperforming 5-point RANSAC. This behavior is explained in [2] as the lack of inflation in the theoretical hypothesis as opposed to classic SfM algorithms. The standard deviation of image noise was chosen to be 0.5. While the experimental data concurs with the theoretical calculations, a higher pose uncertainty rate and camera displacement error was found, as seen in the figures below. This was attributed to the reduced stability of the chosen platform as opposed to ground vehicles or hand held input devices. In addition to this, the number iterations required by 5 Point RANSAC to eliminate higher percentages of outliers (and spurious matches) in data sets, was found to be significantly higher than 1 Point RANSAC. Since higher number of iterations leads to greater computational complexity, it was inferred that 1 Point RANSAC would indeed reduce the computational complexity for the SLAM process as required by this paper. In **Figure 4**, the display of feature points as seen during the SLAM process itself is seen.



(a)

(b)

Figure 4 Skeletal display (a and b) of the feature points as seen through the AR drone during the SLAM process

V. CONCLUSION AND FUTURE WORK

In this paper, we tried to establish the efficacy of employing a specific technique to solve the SLAM problem instead of relying on the platitudes of various other algorithms.

- 1 Point RANSAC was shown to be reliable, efficient and computationally inexpensive and therefore readily applicable to the case of a low cost, multi-rotor system for the purposes of solving the SLAM problem.
- Since 1 Point RANSAC is able to deal with large outlier data sets at low computational cost, when used in conjunction with EKF, we are able to achieve high accuracy in estimation as well.
- However, if used with a highly feature rich scene or multiple moving targets, 1 Point RANSAC would not be a suitable protocol.

The AR Drone itself proved to be an effective low-cost tool. The information acquired through the on board sensors were fed into the EKF in order to estimate the pose of the drone. Since, the position of the AR Drone could not be estimated directly it was derived from the velocity estimates from the AR drone. This estimate was based on the data from the IMU, aerodynamic model and visual odometry obtained from the relative motion between camera frames. With the aid of this estimate, feature and texture maps could be generated using the SLAM protocols.

Since, this paper addressed the specific case of 1 Point RANSAC under set boundary conditions,

- The next frame of work would be the study and development of specific algorithms in order to deal with areas that are very rich in features and contain moving targets
- Additionally global map optimization methods would have to be researched to aid loop closure techniques

Commercially available laser scanners such as the Microsoft Kinect are easy to use and affordable. This could be looked into as a potential research tool in order to aid the SLAM technique.

- Depth Maps could be generated using the IR Laser Array and thereby 3D reconstruction of targeted space could be seen as the next outcome of the SLAM process.
- In other words, instead of having feature, texture and elevation maps, we could develop an accurate 3Dmap of the prior unknown targeted space.

REFERENCES

[1] Achtelik, M., Bachrach, A., He, R., Prentice, S., & Roy, N. (2008). Autonomous navigation and exploration of a quadrotor helicopter in gps-denied indoor environments. *Robotics: Science and Systems.*

[2] Civera, J., Grasa, O., Davison, A., & Montiel, J. (2010). 1-Point RANSAC for EKF Filtering: Application to Real-Time Structure from Motion and Visual Odometry. *Journal of Field Robotics*. CrossRef

[3] Scaramuzza, D., Fraundorfer, F., & Siegwart, R. (2009). Real-Time Monocular Visual Odometry for On-Road Vehicles with 1-Point RANSAC. *IEEE International Conference on Robotics and Automation* (pp. 4293-4299). Kobe: IEEE. CrossRef

[4] Torr, P., & Murray, D. (1993). Outlier detection and motion segmentation. *Sensor Fusion VI*, 2059:432-443. CrossRef

[5] Davison, A. J. (2003). Real-time simultaneous localisation and mapping with a single camera. *IEEE International Conference on Computer Vision*, (pp. 1403-1410). IEEE. CrossRef

[6] Chum, O., & Matas, J. (2008). Optimal randomized RANSAC. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(8): 1472-1482. CrossRef

[7] Steder, B., Grisetti, G., Stachniss , C., & Burgar, W. (2008). Visual SLAM for flying vehicles. *Robotics, IEEE Transactions on, volume 24(5)*, 1088–1093. CrossRef

[8] Vedaldi, A., Jin, H., Favaro, P., & Soatto, S. (2005). KALMANSAC: Robust filtering by consensus. *Proceedings of the International Conference on Computer Vision (ICCV)*, (pp. volume 1, pages 633-640. CrossRef

[9] Dijkshoorn, N. (2012). *Simultaneous localization and mapping with AR Drone*. Amsterdam: Universiteit van Amsterdam.

[10] Lemaire, T., Berger, C., Jung, I.-K., & LACROIX, S. (2007). Vision-Based SLAM: Stereo and Monocular Approaches. *International Journal of Computer Vision*, 343–364. CrossRef

[11] Fischler, M., & Bolles, R. (1981). RANSAC random sampling concensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of ACM 26*, 381–395. CrossRef

[12] Torr, P., & Zisserman, A. (2000). MLESAC: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 138-156. CrossRef

[13] Hartley, R., & Zisserman, A. (2004). *Multiple View Geometry in Computer Vision*. Cambridge University. CrossRef

[14] Klein, G., & Murray, D. (2008). Improving the Agility of Keyframe-Based SLAM. *10th European Conference on Computer Vision: Part II* (pp. 802-815). Springer. CrossRef

[15] Mouragnon, E., Lhuillier, M., Dhome, M., Dekeyser, F., & Sayd, P. (2009). Generic and real-time structure from motion using local bundle adjustment. *Image and Vision Computing*, 27(8): 1178-1193. CrossRef

[16] *Monocular SLAM*. (2012). Retrieved from Vision. VIEW

[17] Derafa, L., Benallegue, A., & Fridman, L. (2012). Super twisting control algorithm for the attitude tracking of a four rotors UAV. *Journal of the Franklin Institute, Volume 349, Issue 2*, 685-699. CrossRef

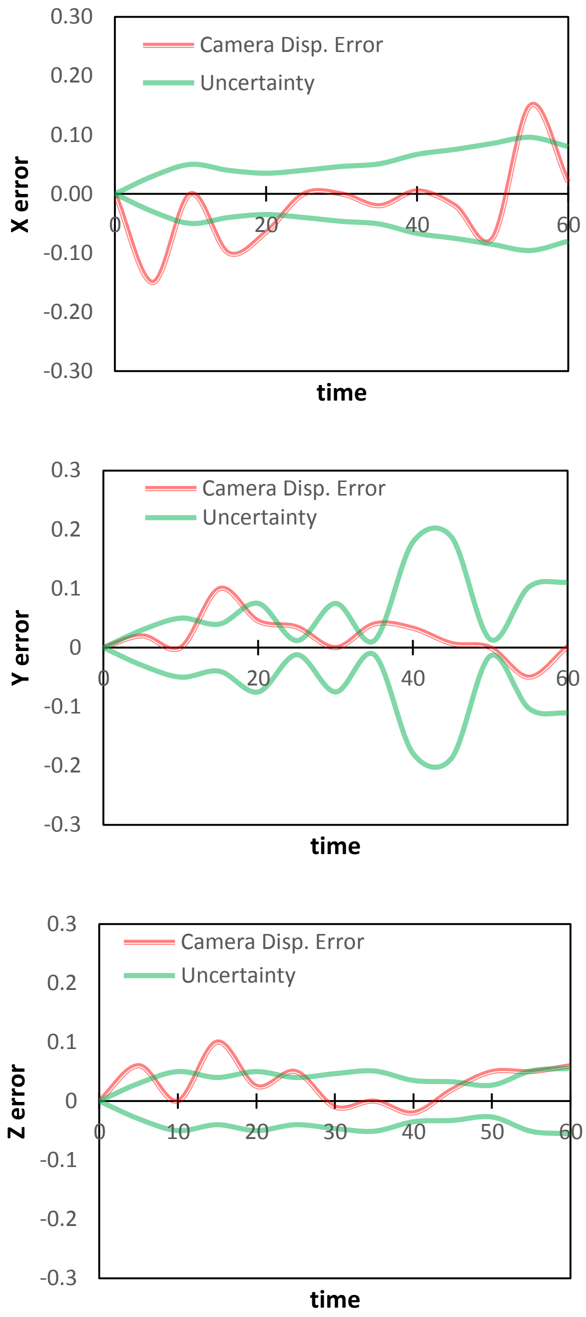


Figure 5 Camera displacement error for 5 points RANSAC

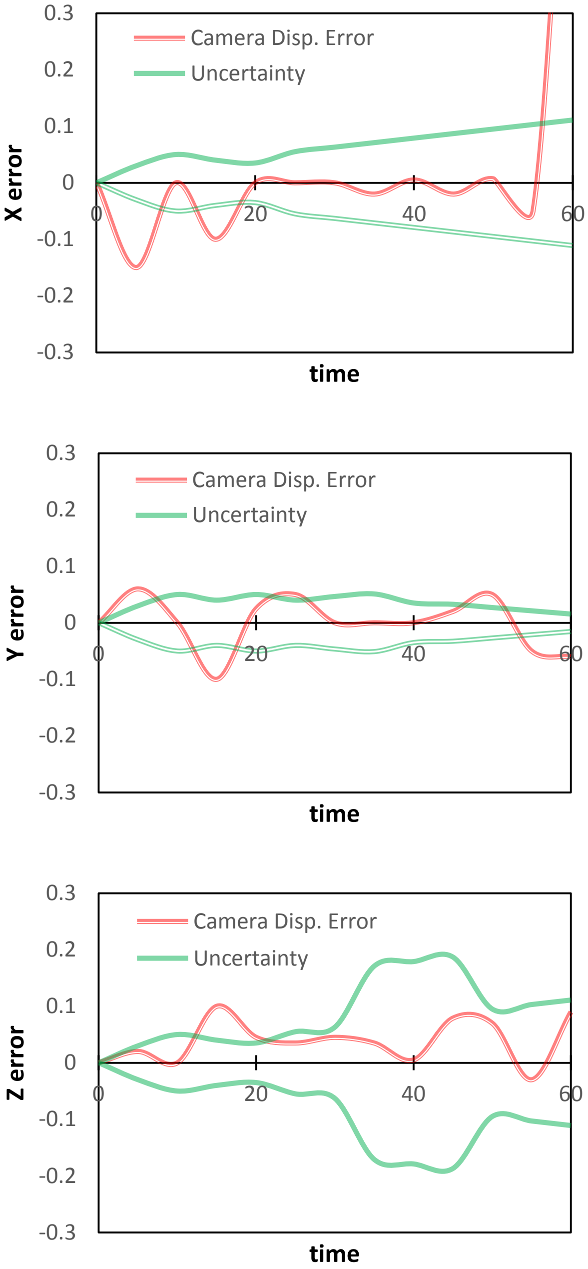


Figure 6 Camera displacement error for 1 point RANSAC

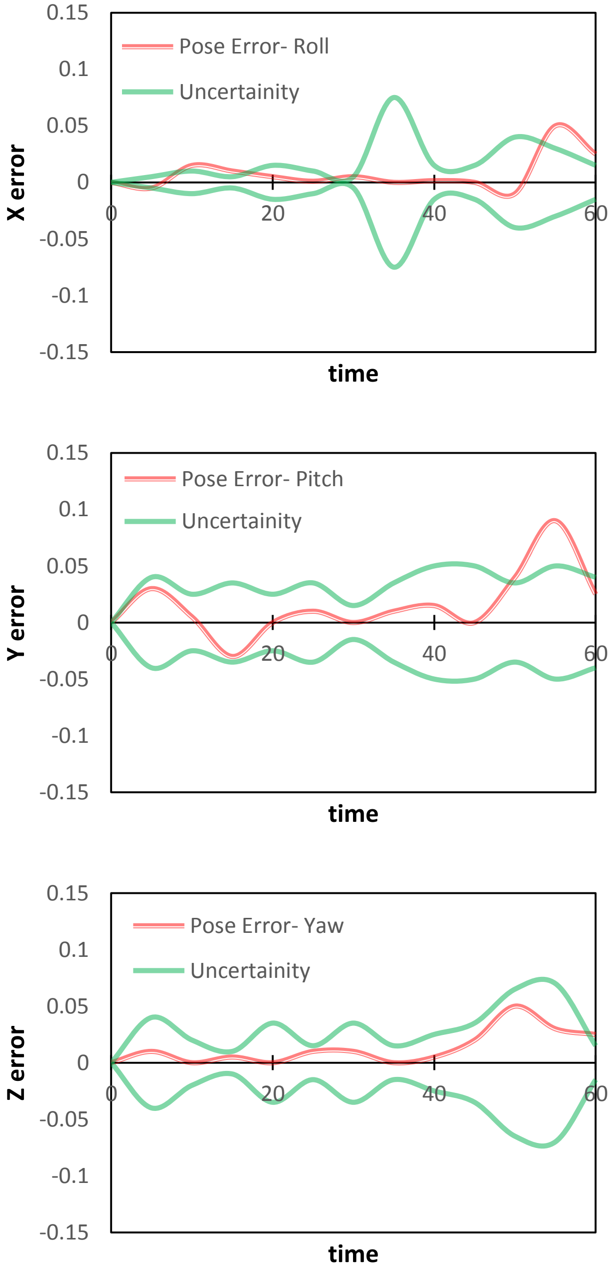


Figure 7 Pose error for 5 points RANSAC

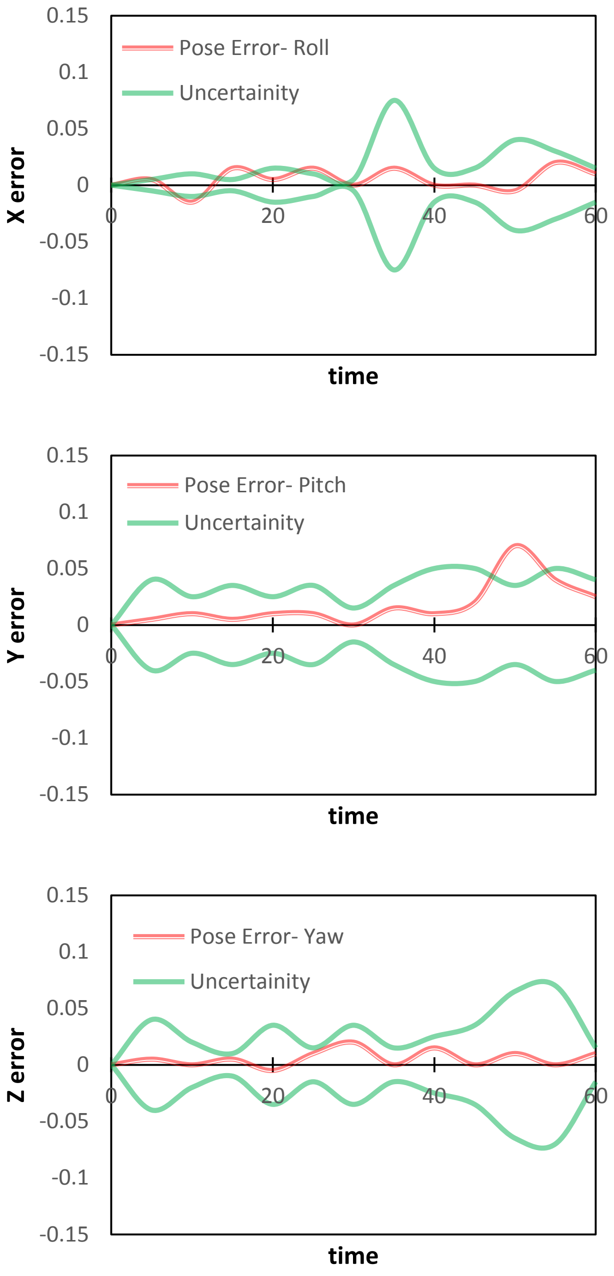


Figure 8 Pose error for 1 point RANSAC