EVOLUTUS*

2014-15 MRSD PROJECT TEAM H

CHIMP GRASPING STRATEGY

16-662 ROBOT AUTONOMY

PROJECT 2B: REPORT

Team Evolutus
William Ku, Yuhan Long, Yu-Te Cheng, Dawei Wang.

March 19, 2015

1 Stereo Visual Odometry

a) The Harris feature detector function in MATLAB, `detectHarrisFeatures`, was used to extract features in the first pair of stereo images, `left000.jpg` and `right000.jpg`.

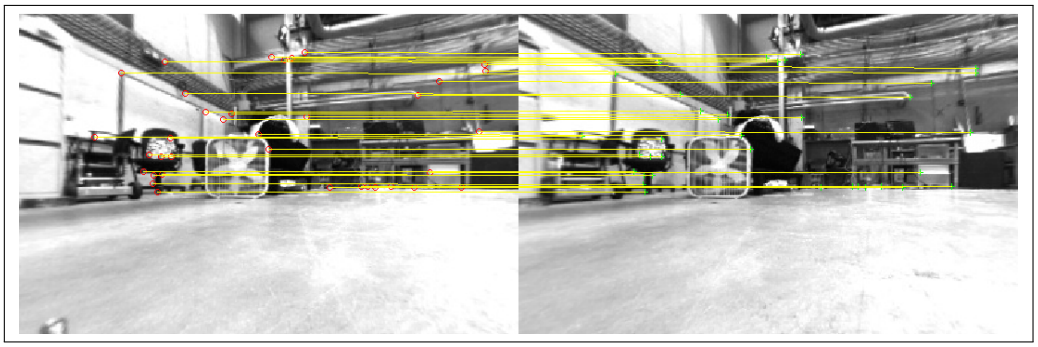
b) Figure 1a shows the stereo correspondences found using MATLAB's `matchFeatures()` function.

c) Assuming a pinhole camera model, triangulation of the stereo correspondences was performed using Direct Linear Transform (DLT) on the camera projection relation, $x \times PX = 0$, where P is the camera matrix, X the three-dimensional world point, and x the two-dimensional image point of X. Plots for the feature locations in Figure 1a are shown in Figure 2.

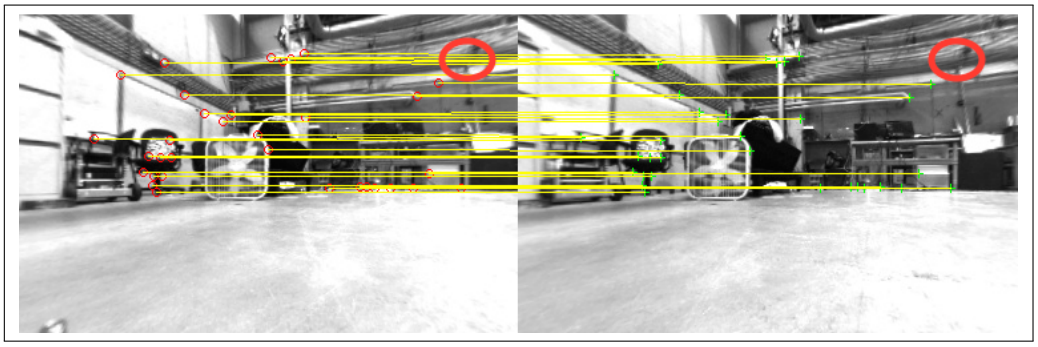
d) Correspondences between successive left images, `left000.jpg` and `left001.jpg` were found using the same feature detection and matching procedure described in parts (a) and (b).

Let x_n and P_n be the correspondence points and the camera matrix for the n-th image, then x_1 and the 3-D triangulated points, X, could be used to estimate the camera matrix, P_1 using Direct Linear Transform (DLT). We assume the origin was at the camera center of the previous image, `left000.jpg` (i.e. $P_0 = K[I_3|\mathbf{0}_3]$, where K is the 3-by-3 upper-triangular camera intrinsic matrix (identical for all images), I_3 the 3-by-3 identity matrix, and $\mathbf{0}_3$ the 3-by-1 zero vector).

To reject outliers, the computation of P_n was wrapped within a RANSAC routine thresholded by the camera reproduction error, $e = \|x_n - P_n X\|$. A minimum of 5.5 such correspondences were needed to compute the 11 degrees-of-freedom for the 3-by-4 camera matrix (minus one for scale ambiguity). For

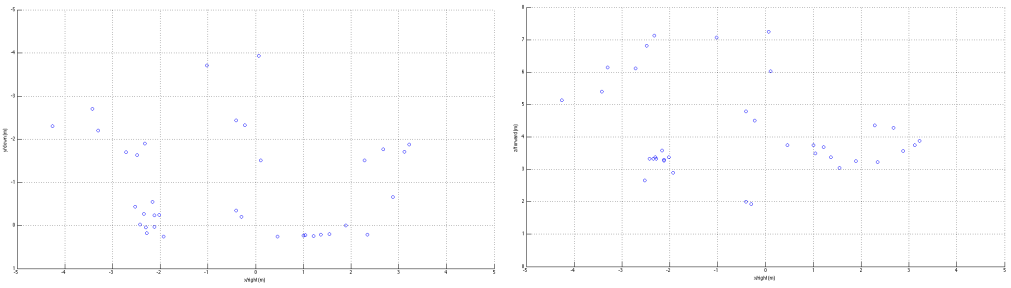


(a) Full correspondences.



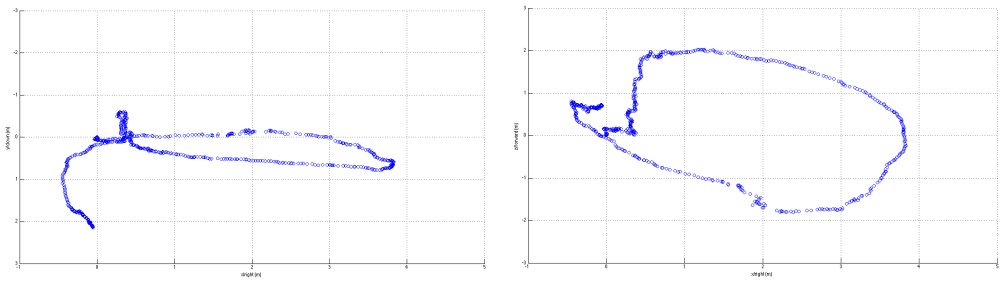
(b) Outliers removed (thick red circle) using RANSAC on reprojection errors.

Figure 1: Point correspondences between `left000.jpg` and `right000.jpg` using a Harris feature detector.



(a) Front view (b) Top view

Figure 2: The triangulated 3D points of stereo correspondences.



(a) Front view (b) Top view

Figure 3: The trajectory generated with incremental rotations and translations

convenience, six correspondences were used for every iteration of the RANSAC procedure. The inlier correspondences are shown in Figure 1b.

For the computed camera matrix, P_n, the following relationship holds:

$$P_n = [P_{n,1-3}|P_{n,4}] = K[R_n|t_n], \tag{1}$$

where R_n and t_n are the rotation and translation between camera frames $n-1$ and n. $P_{n,1-3}$ was decomposed using RQ decomposition to obtain K and R_n. Translation was compuled with $t_n = K^{-1}P_{n,4}$.

e) Parts (a) to (d) were repeated to find R_n and t_n for every successive image pair. Assuming that the camera started at the origin of the world frame, a trajectory, shown in Figure 3, was generated by composing sequential transformations of R_n and t_n. Note that in Figure 3a the end position falls below the starting position. This will be discussed in a later section. The position and orientation, in ZYX Euler angles, over time are shown in Figures 4 and 5. To increase readability, the y values in all of the position plots in this document were negated due to a downward-pointing y-axis.

f) Part (e) was repeated with different RANSAC numbers of iterations and thresholds.

- Decrease the number of iterations by a factor of 10:

  The trajectory is shown in Figure 6 and the position and orientation plots are shown in Figures 7 and 8. The decrease in the number of iterations resulted in additional drifts due to less optimal camera matrix estimations.

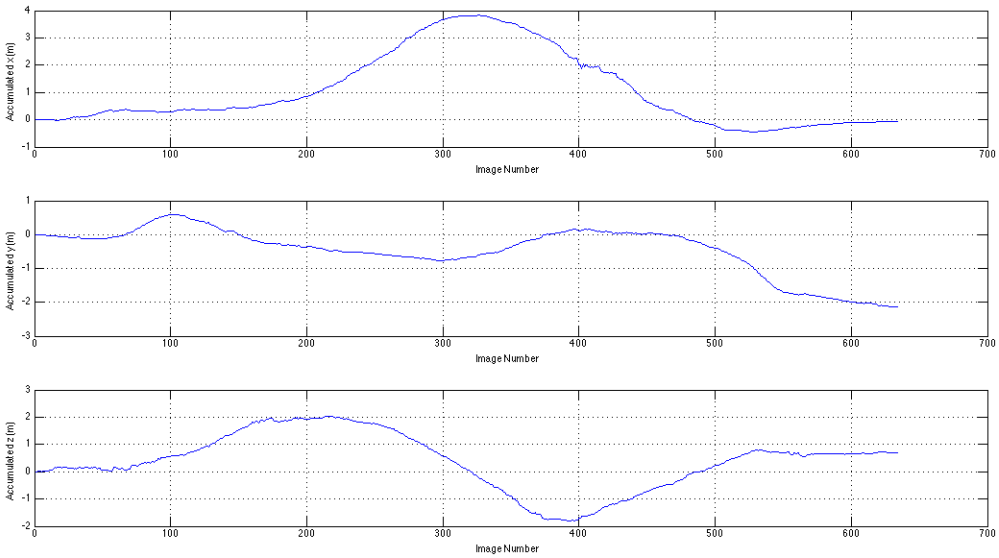


Figure 4: Camera Positions from Concatenated Transformations

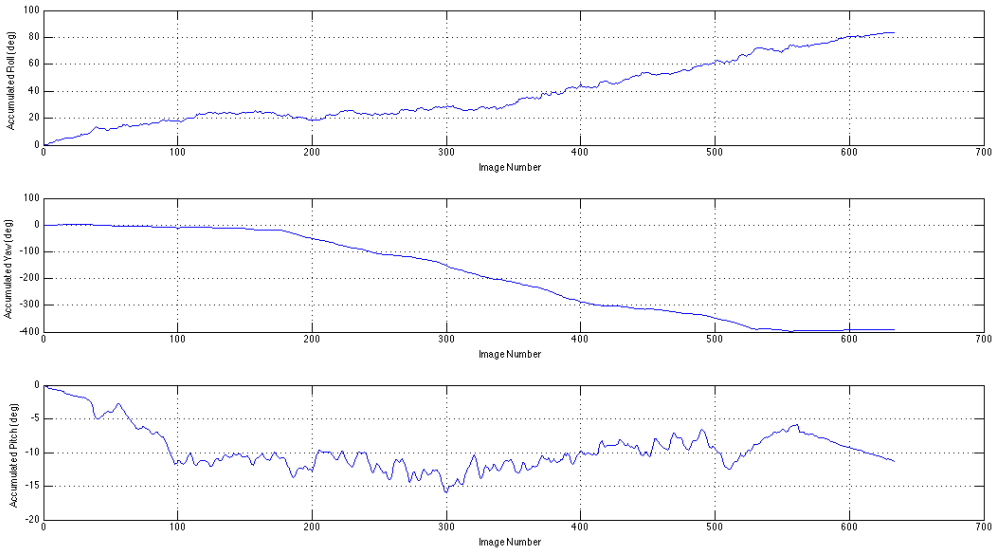
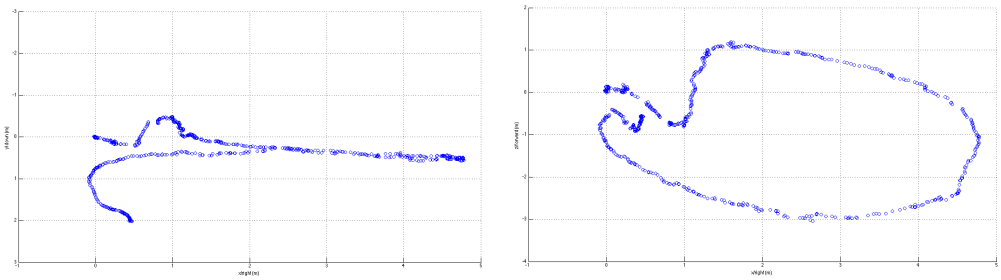


Figure 5: Camera Orientations (ZYX) from Concatenated Transformations

- Increase the threshold by a factor of 10:

  The trajectory is shown in Figure 9 and the position and orientation plots are shown in Figures 10 and 11. A larger threshold potentially allowed outliers to be included as inliers during the RANSAC routine, thus yielding less optimal and noisier camera matrix estimations. As a result, additional errors in drift were introduced to the trajectory.



(a) Front view (b) Top view

Figure 6: The trajectory generated with 10 times fewer RANSAC iterations.

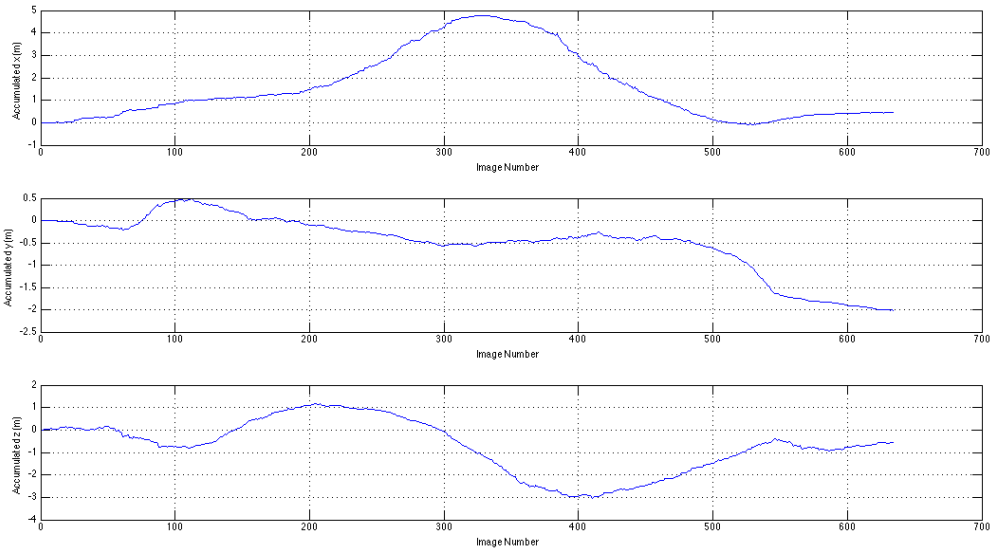


Figure 7: Camera Positions with 10 times fewer RANSAC iterations.

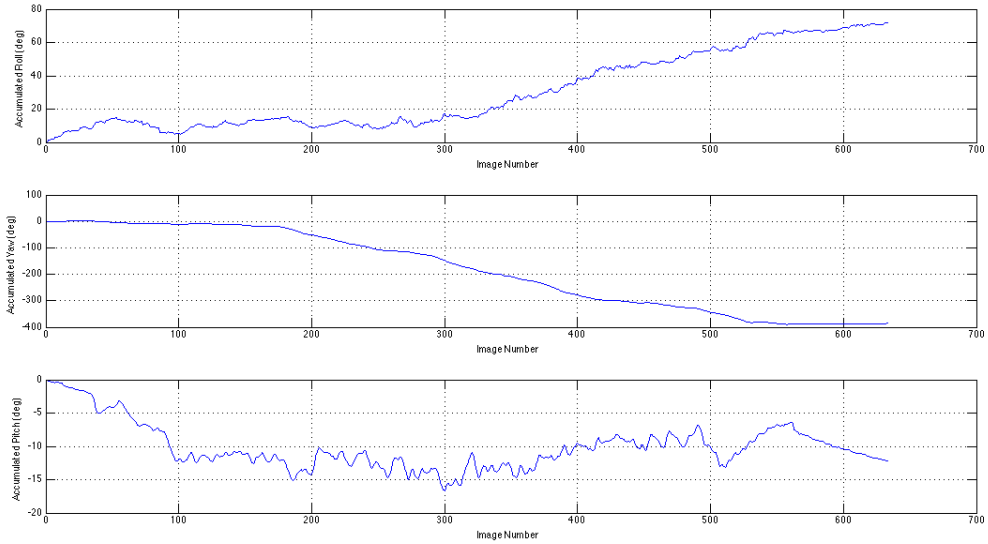
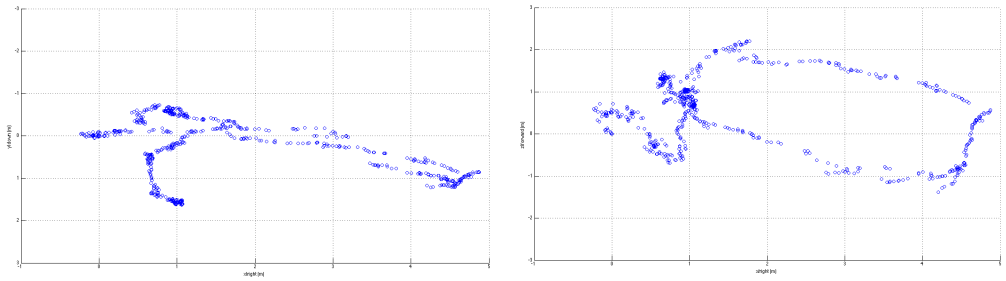


Figure 8: Camera Orientations with 10 times fewer RANSAC iterations.



(a) Front view (b) Top view

Figure 9: The trajectory generated with 10 times RANSAC threshold.

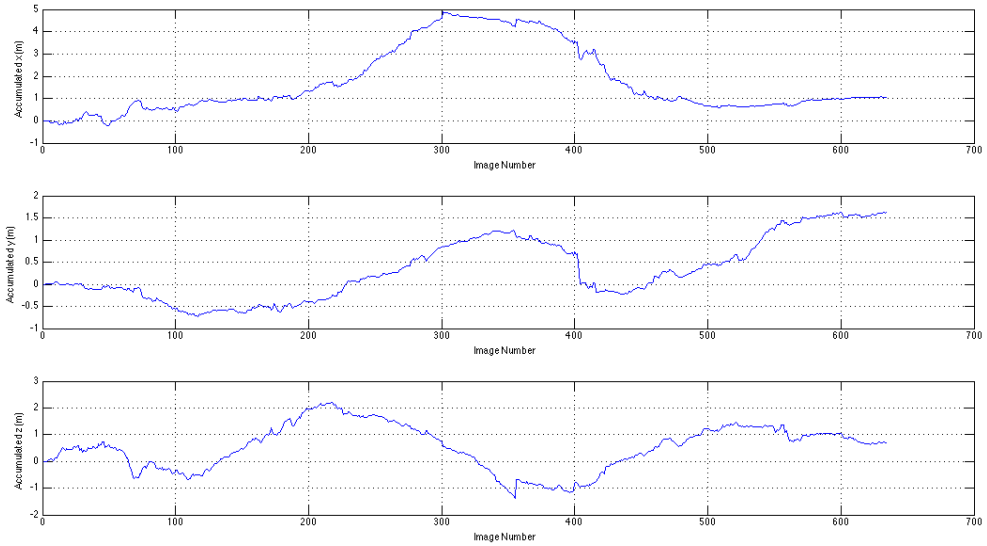


Figure 10: Camera Positions with 10 times RANSAC threshold.

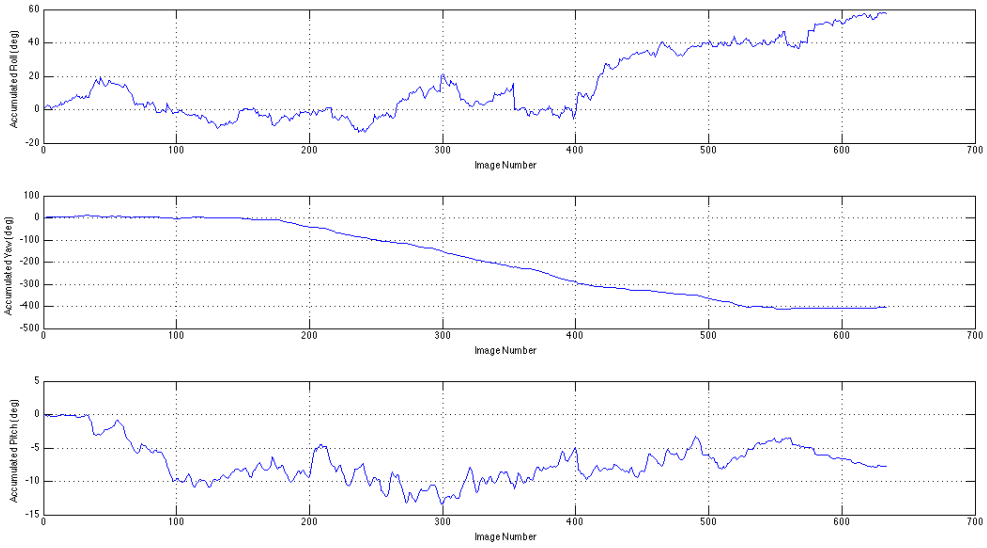


Figure 11: Camera Orientations with 10 times RANSAC threshold.

g) The trajectory computed did not return exactly to the starting position (see Figure 3b). As seen in Figure 3a, the height of the final position was even below ground level. This drift could be explained by the accumulated errors in estimations across each pair of successive frames. After hundreds of estimations, the errors became observably significant in the position estimates. Figures 12 and 13 display the ”drift profile” in position and orientation obtained by calculating the estimation errors between each image frame and its own.

Knowing that such drift was proportional to the image frames calculated, a new trajectory was generated using only every seven frames, as shown in Figure 14. The position and orientation are shown in Figure 15 and 16. This interval of seven images was determined empirically. Intuitively, a smaller interval would account for more image drifts while a larger interval would start to exhibit unstable estimations due to drastic scene (and hence, features) changes.

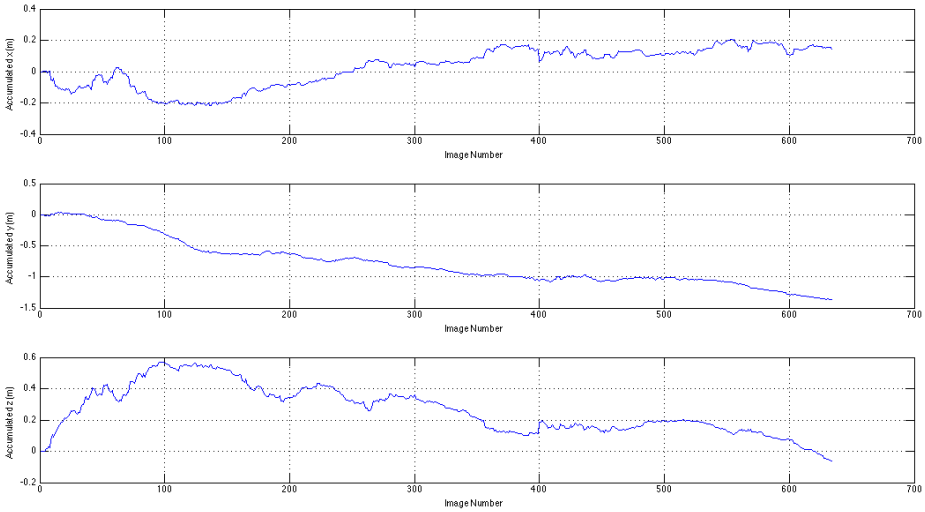


Figure 12: Position Drift Across All Successive Images

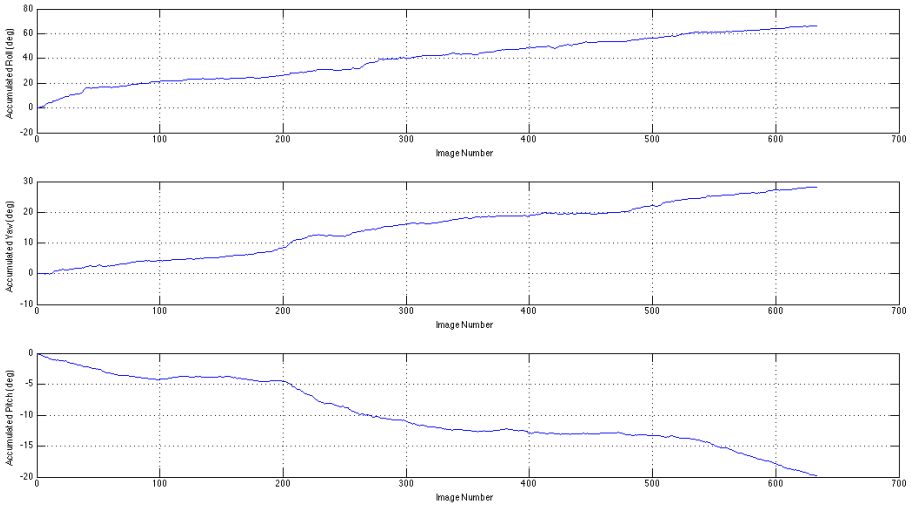
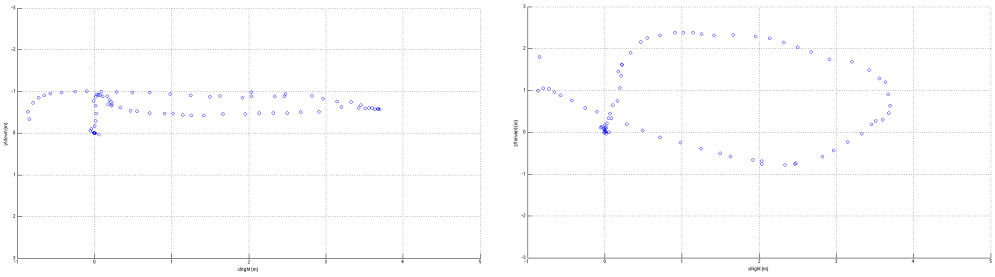


Figure 13: Orientation Drift Across All Successive Images



(a) Front view. (b) Top view.

Figure 14: The trajectory generated with every seven images

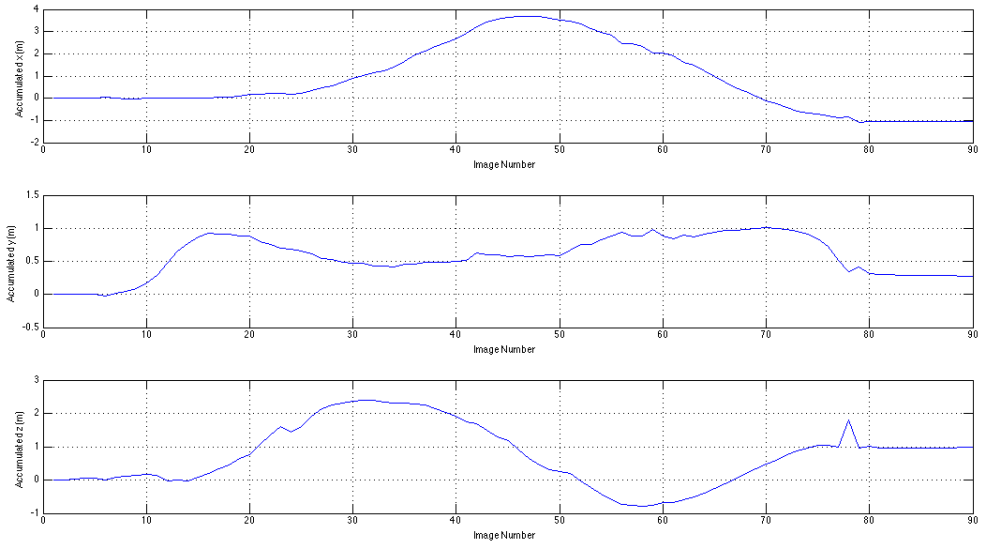


Figure 15: Camera Positions with every seven images

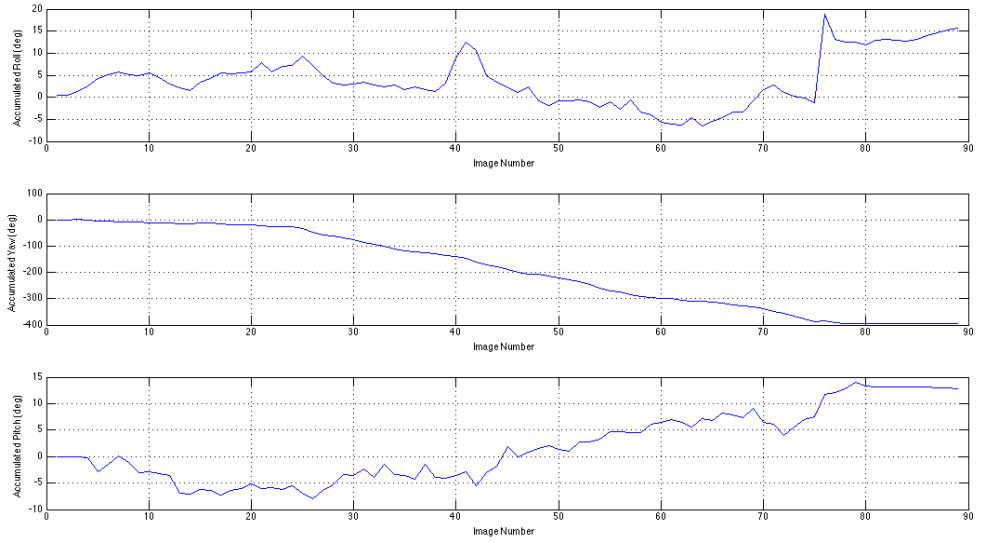
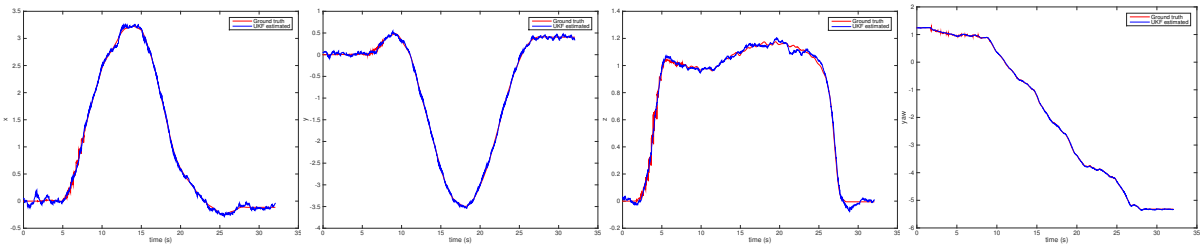


Figure 16: Camera Orientations with every seven images

2 Unscented Kalman Filter

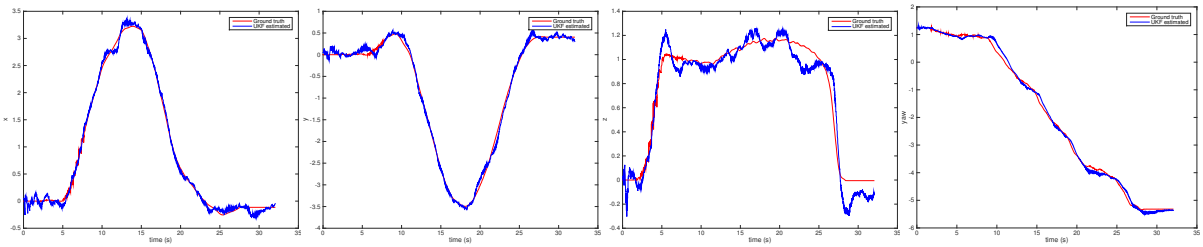
Presented here are four sets of our Unscented Kalman filter results. Figure 17 shows the result of our baseline test, in which all the noise covariances are set as 0.001. To demonstrate the effects of changing the initial noise sigma, IMU noise and observation noise, we have made changes to each group of sigma values to make them much larger than the baseline.

Figure 18 shows the filter output when the initial noise sigma is increased by 100 times. A large error is observed initially, but it then tracks properly after several seconds. Note that the filter needs longer time to reduce the estimation error for some fast-changing states (z in this case). In Figure 19, the IMU noise is increased by 50 times. Larger IMU noise and bias noise will result in increased filter error in the Model Update process; however, it will be corrected by the observations in the Correction process. In conclusion, adding IMU noise will make the filter output noisy, without creating a large bias.



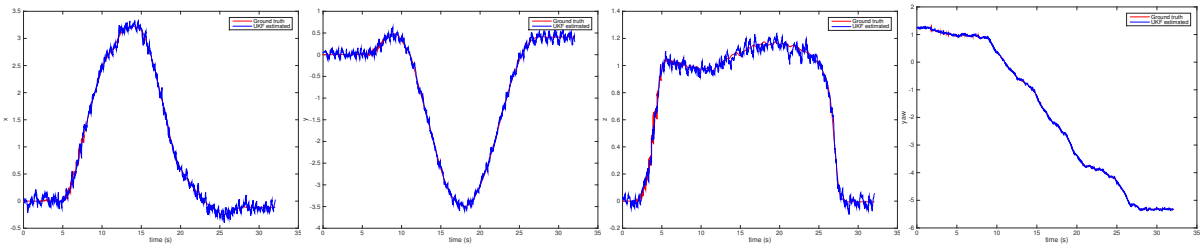
(a) x (b) y (c) z (d) Yaw

Figure 17: Baseline results with all parameters set to 0.001.



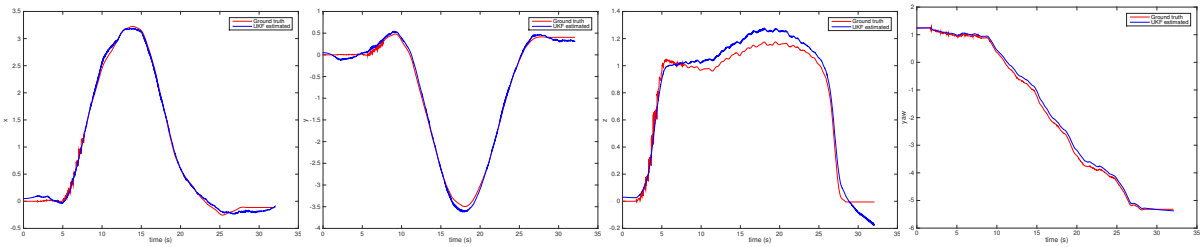
(a) x (b) y (c) z (d) Yaw

Figure 18: Results with 100 times initial noise coefficients.



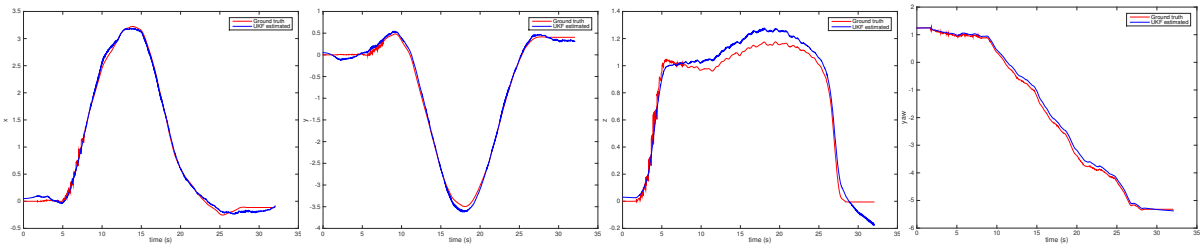
(a) x (b) y (c) z (d) Yaw

Figure 19: Results with 50 times IMU noise coefficients.



(a) x (b) y (c) z (d) Yaw

Figure 20: Results with 100 times observation noise coefficients.



(a) x (b) y (c) z (d) Yaw

Figure 21: Results with tuned noise parameters

Figure 20 shows the filter output when the observation noise is increased by 100 times. Note that there is no noticeable noise, but bias and degraded accuracy are observed. This is because a wrong observation will lead to a bias correction that affects all the state vectors.

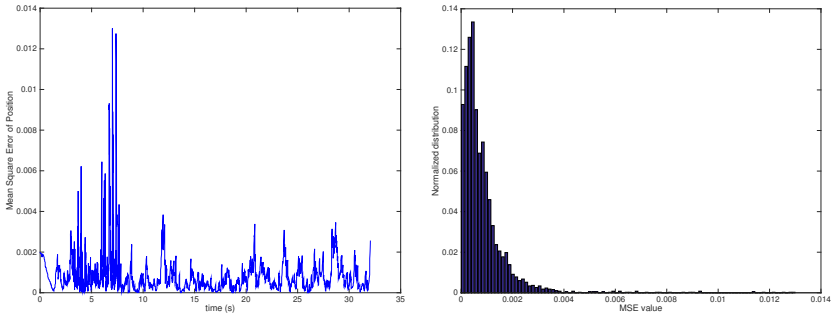
Table 1 shows the noise parameters. The observation noise is estimated by computing the mean and covariance of a series of data derived by subtracting the ground truth from the observation. The IMU bias noise is given as 0.03. The acceleration and angular velocity noises are estimated by computing their covariances when the drone is in a relatively stationary state. The initial sigma value has been manually tuned: it appears that as long as the initial values are kept within a reasonable range, the result will not be affected too much. The filter output after noise parameter tuning is shown in Figure 21.

Table 1: Noise Parameters

`init_pos_sigma`	0.01	`sigma_bw`	0.03
`init_rpy_sigma`	0.001	`sigma_a`	0.08
`init_vel_sigma`	0.001	`sigma_ba`	0.03
`init_bw_sigma`	0.001	`sigma_xy`	0.077
`init_ba_sigma`	0.001	`sigma_z`	0.05
`sigma_w`	0.03	`sigma_ yaw`	0.02

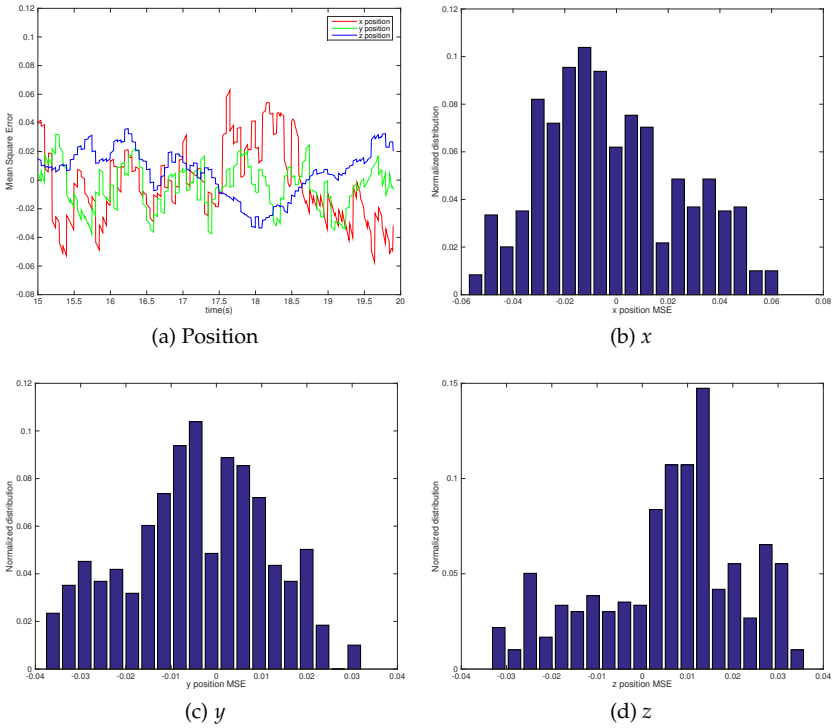
Figure 22 shows the mean square error and its distribution for the observed position against the ground truth. It can be seen from the distribution that after tuning the noise parameter, the largest mean square error is 0.013. The mean error is 0.0008, with most of the errors below 0.004.

The data from 15s to 20s have been selected to compare the empirical standard deviation and to estimate the UKF 1-sigma. Figure 23 shows the position errors and their distributions. The empirical stan-



(a) Mean Square Error Over Time (b) Error Distribution

Figure 22: Position Estimation Error



(a) Position (b) x

(c) y (d) z

Figure 23: Position Error and their Distributions

dard deviations in this time interval are: `std_x`: 0.026569, `std_y`: 0.015594, `std_z`: 0.016420. The estimated 1-sigma values are x: 0.030517, y: 0.030510, z: 0.014851. The empirical standard deviation is smaller than the estimation. This is because the noise of each estimation state will be propagated to the next state. When updating the state, the noise will be accumulated, but the mean state value will generally stay close to the ground truth. In another word, UKF estimates the noise in the process update in a conservative way, resulting in a larger estimated standard deviation than the empirical standard deviation.

EΛOLOTUS*