

16-662 Robot Autonomy Project 2B

William Ku (Andrew ID: wku)
03/23/2015

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.



(a) Full correspondences.



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

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

- c) Assuming a pinhole camera, triangulation of the stereo correspondences was performed using Direct Linear Transform (DLT) on the camera projection relationship, $x \times PX = 0$, where P is the camera matrix, X the 3-D world point, and x the 2-D projection image point of X . The 3-D plots of the feature locations in Figure 1a is 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 the Direct Linear Transform. 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 outlier correspondences, the competition 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 was needed to compute the 11 degrees-of-freedom for the 3-by-4 camera matrix (12-1 dof for scale ambiguity). For 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:

$$\begin{aligned} P_n &= [P_{n,1-3}|P_{n,4}] \\ &= K[R_n|t_n], \end{aligned}$$

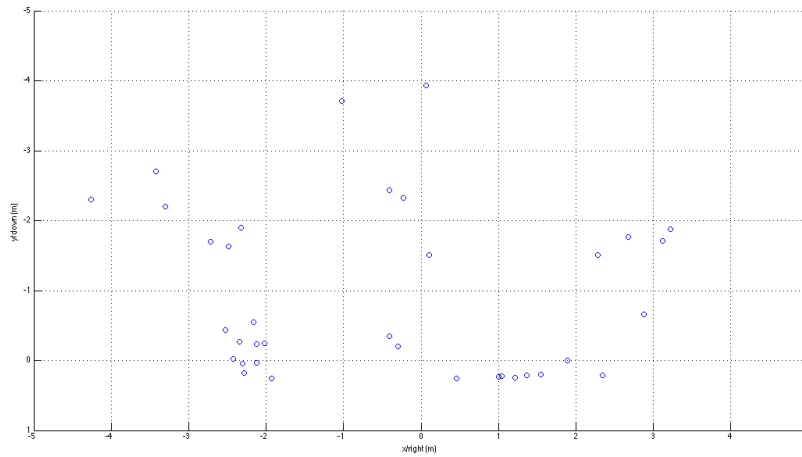
where R_n and t_n are the incremental rotation and translation between camera frames $n-1$ and n . $P_{n,1-3}$ was decomposed using RQ decomposition to obtain K and R_n . Translation could be calculated using $t_n = K^{-1}P_{n,4}$.

- 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 . Please note the frame convention aligns with that of the MATLAB image coordinates: x-right, y-down, z-forward. It is noticeable from Figure ?? that the end position falls below the starting position. This drift 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 iteration numbers and thresholds.

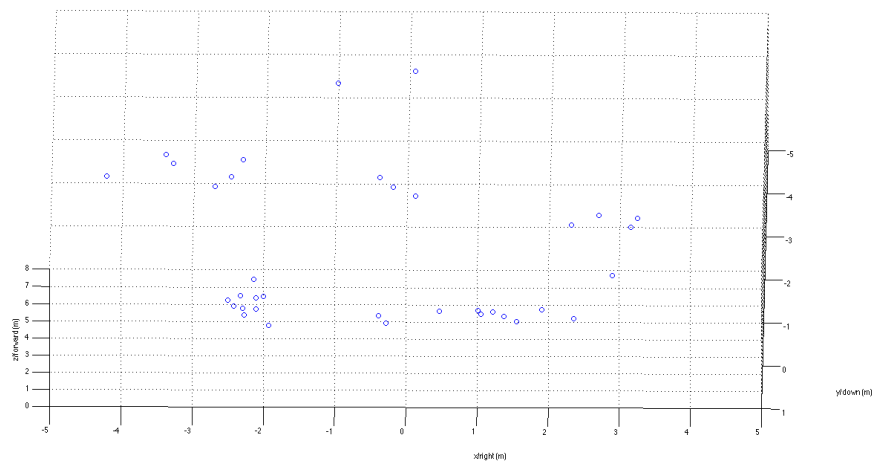
- Decrease the number of iterations by a factor of 10:
The trajectory is shown in Figure 6 and the position and orientation plots are shown in Figures 7 and 8. The decrease in the number of iterations resulted in additional drifts due to less optimal camera matrix estimations.
- Increase the threshold 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.

- g) The trajectory computed did not return exactly to the starting position (see Figure 3c). 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.

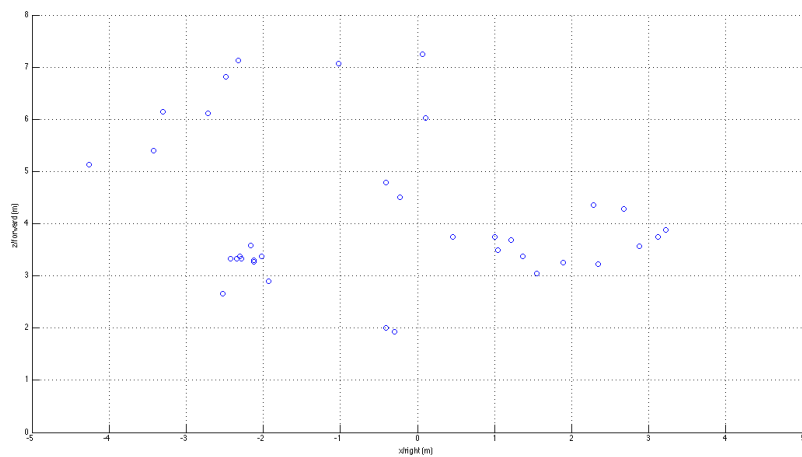
Knowing that such drift was proportional to the image frames calculated, a new trajectory was generated using only every seven frames, as shown in Figure 14. The position and orientation are shown in Figure 15 and 16. This interval of seven images was determined empirically. Intuitively, a smaller interval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



(a) Front view.

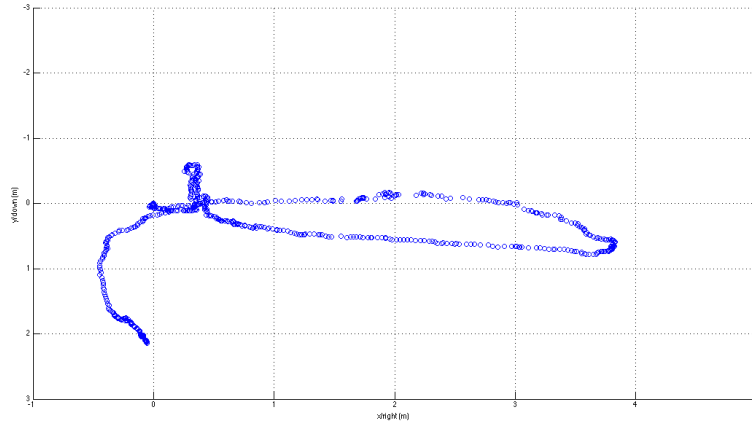


(b) Perspective view.

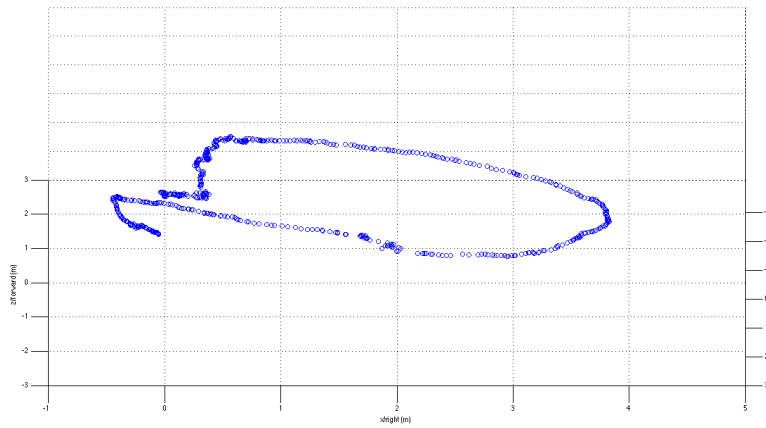


(c) Top view.

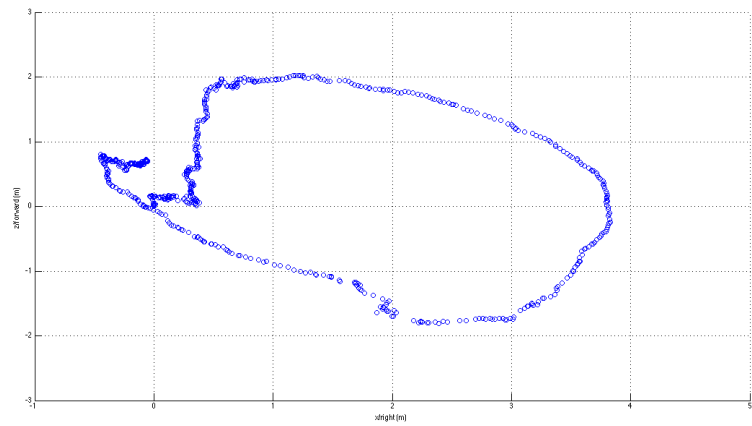
Figure 2: The triangulated 3-D points of stereo correspondences.



(a) Front view.



(b) Perspective view.



(c) Top view.

Figure 3: The trajectory generated using incremental rotations and translations between successive images.

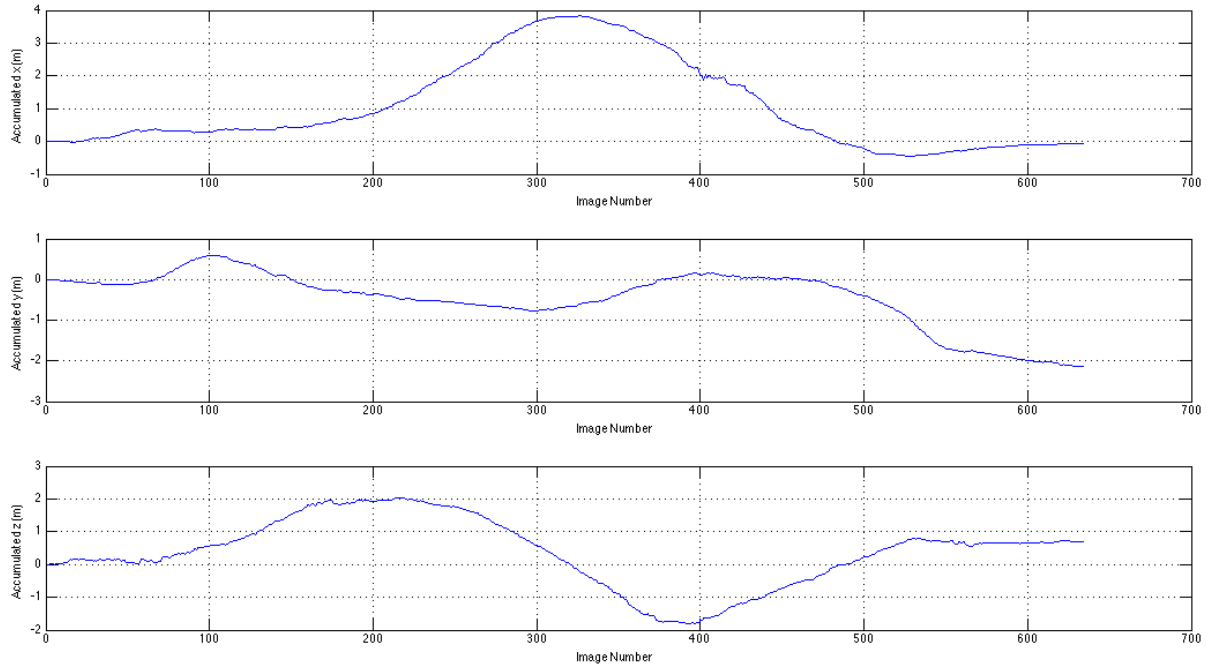


Figure 4: The accumulated position of the camera frame across all successive images.

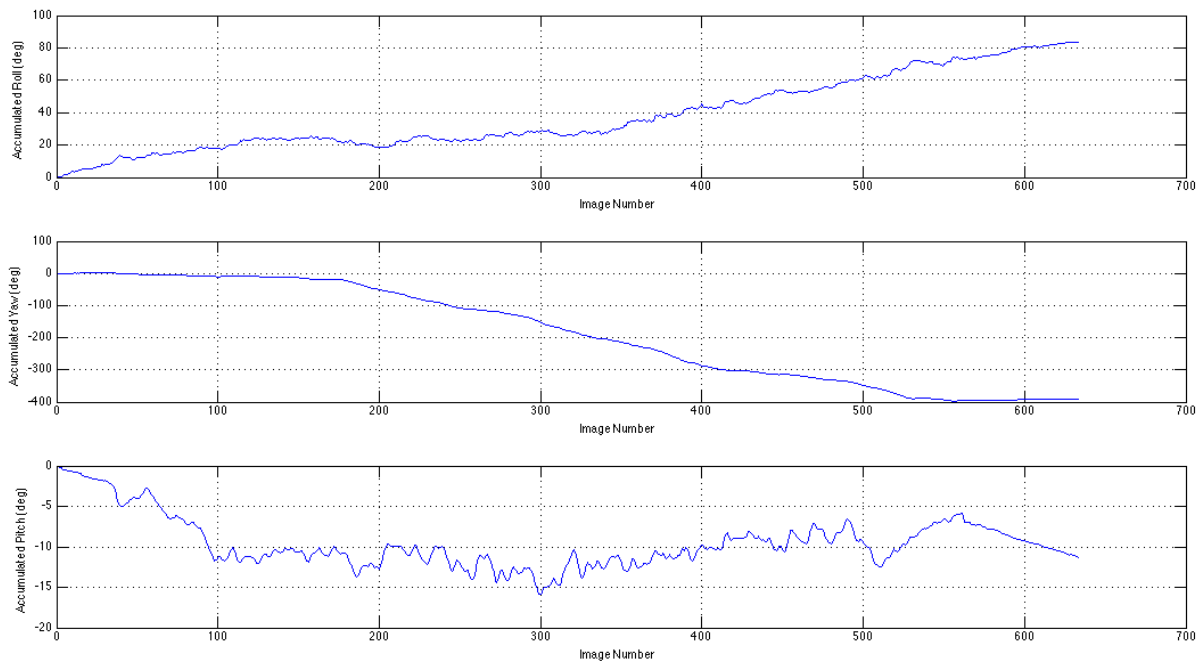
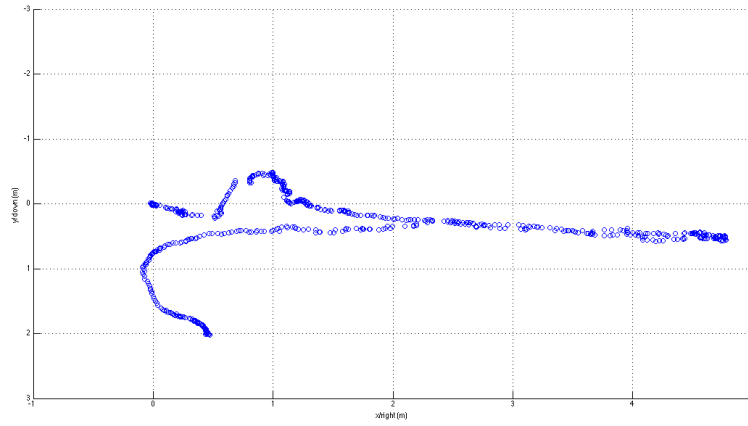
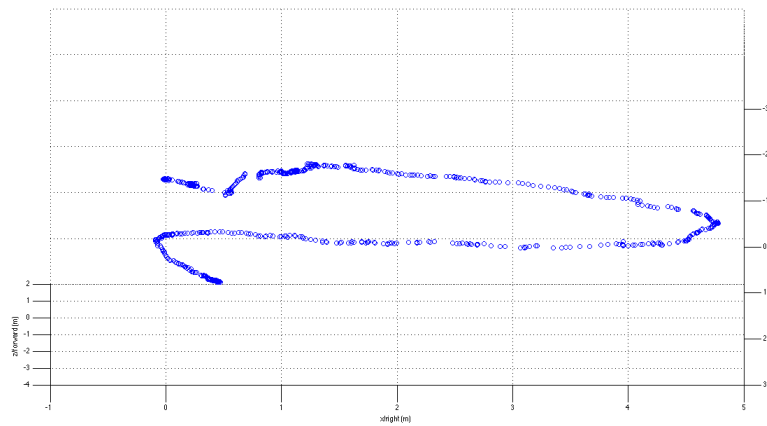


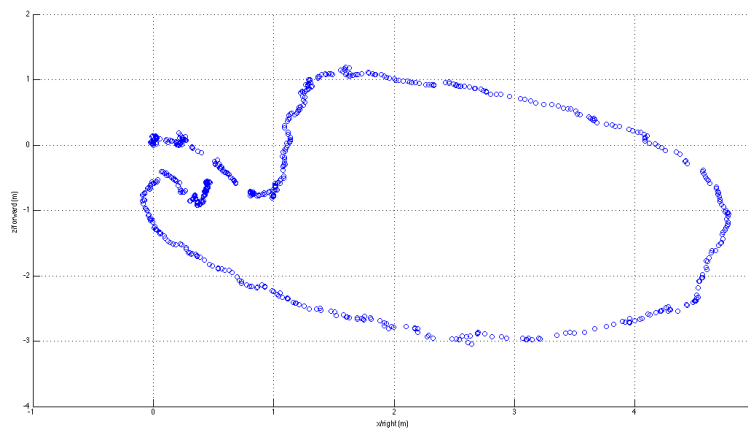
Figure 5: The accumulated orientation (in ZYX Euler angles) of the camera frame across all successive images.



(a) Front view.



(b) Perspective view.



(c) Top view.

Figure 6: The trajectory generated using incremental rotations and translations between successive images for 10 times fewer RANSAC iterations.

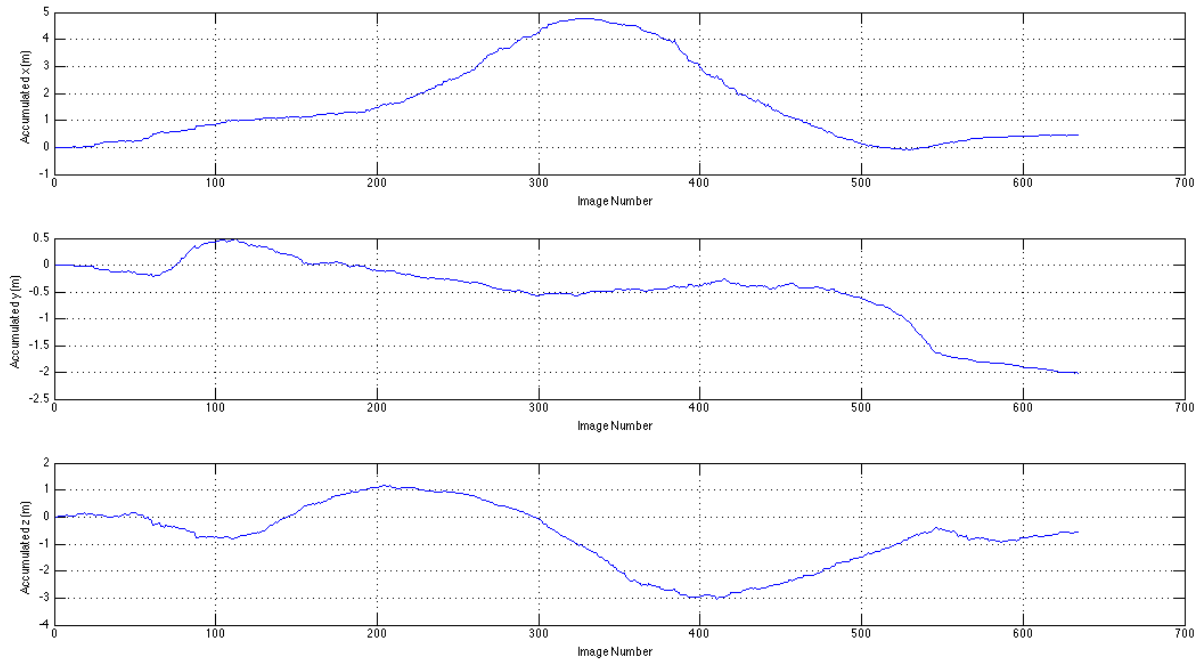


Figure 7: The accumulated position of the camera frame across all successive images for 10 times fewer RANSAC iterations.

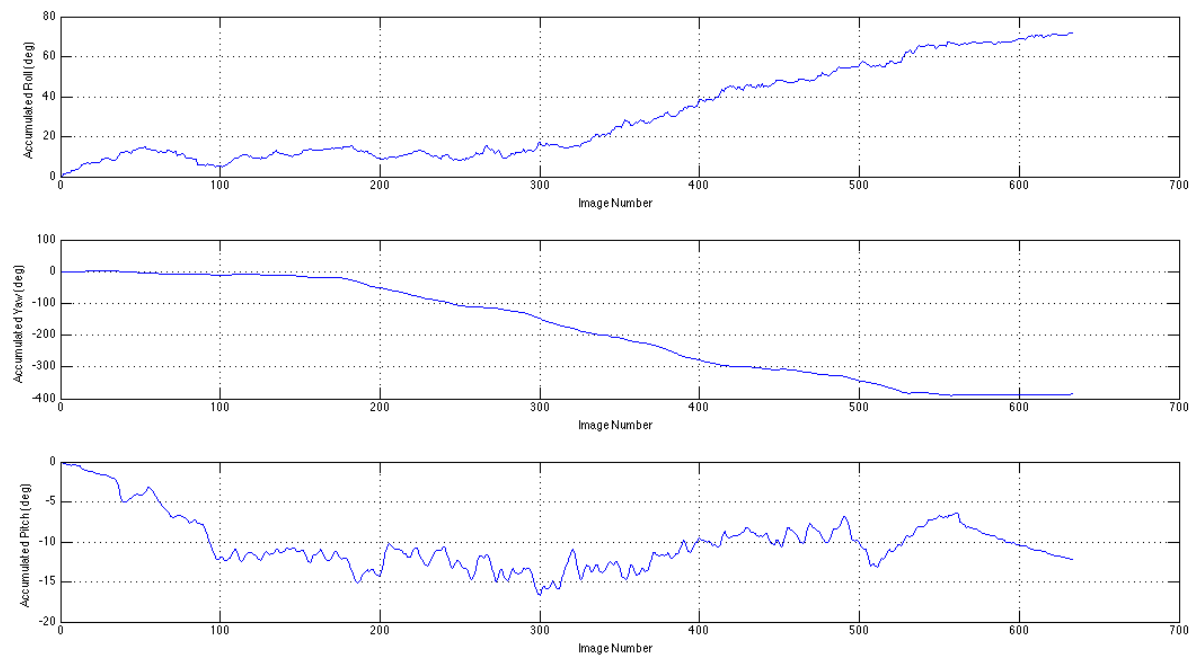
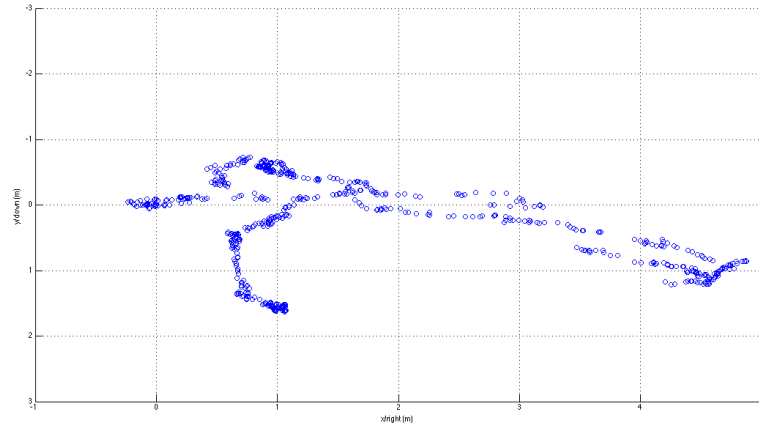
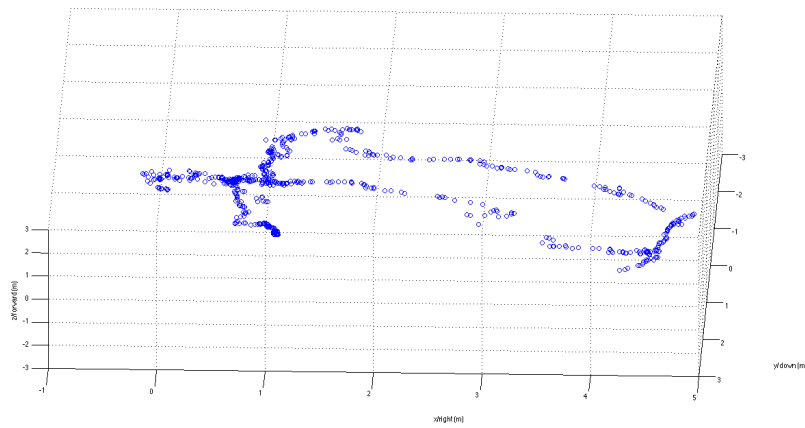


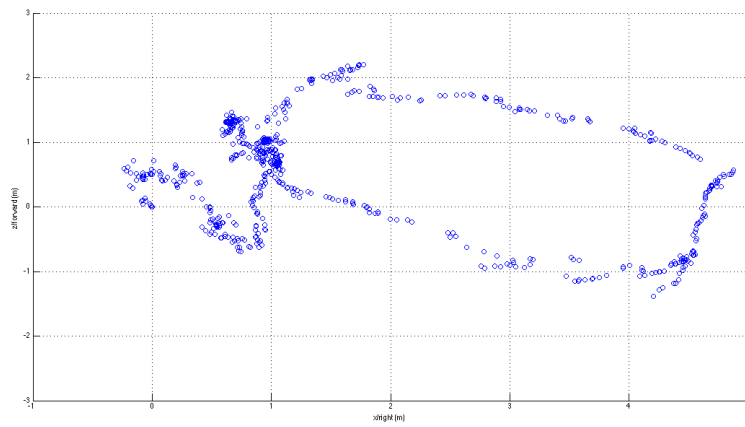
Figure 8: The accumulated orientation (in ZYX Euler angles) of the camera frame across all successive images for 10 times fewer RANSAC iterations.



(a) Front view.



(b) Perspective view.



(c) Top view.

Figure 9: The trajectory generated using incremental rotations and translations between successive images for 10 times RANSAC threshold.

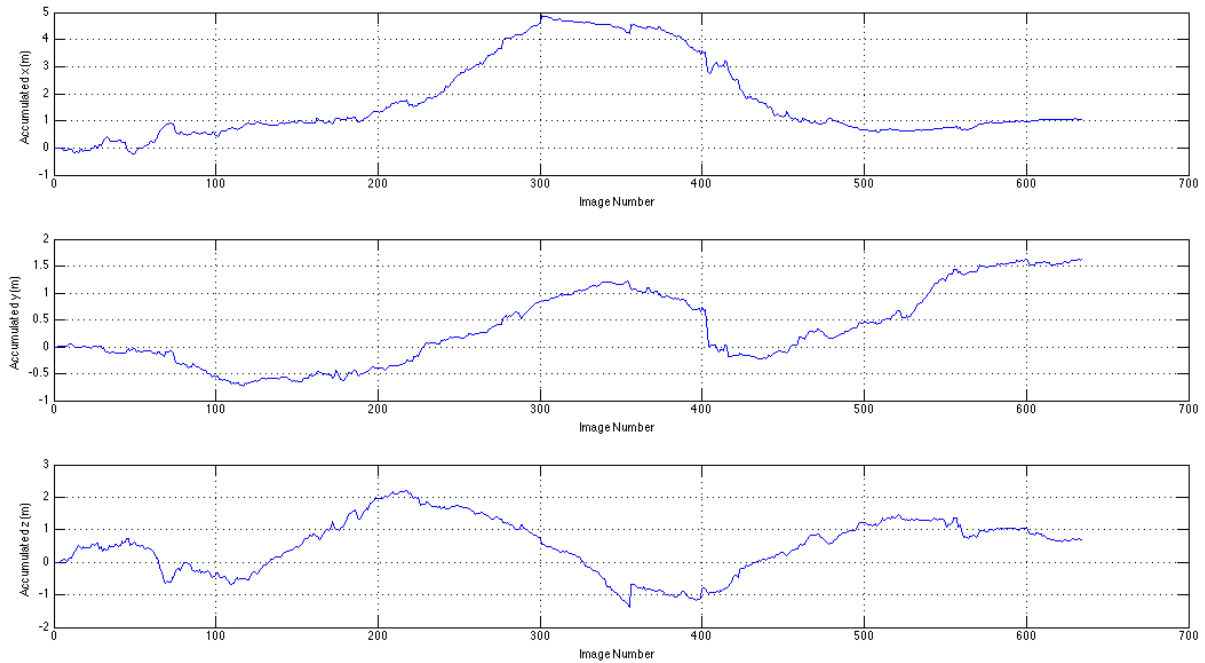


Figure 10: The accumulated position of the camera frame across all successive images for 10 times RANSAC threshold.

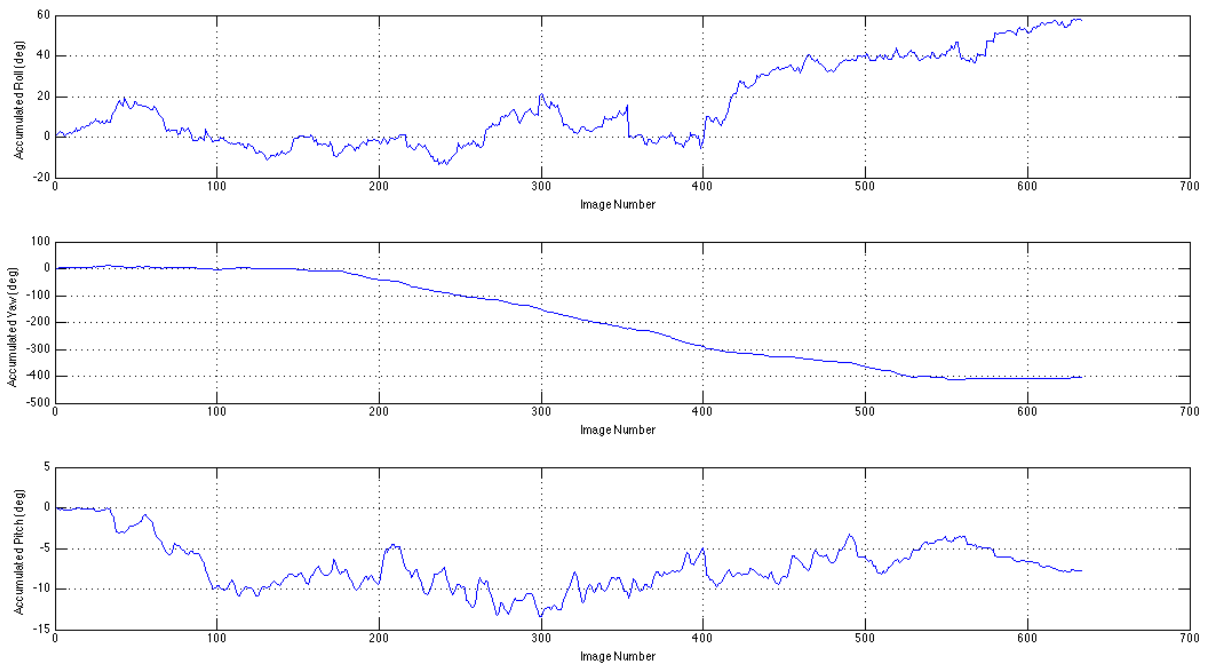


Figure 11: The accumulated orientation (in ZYX Euler angles) of the camera frame across all successive images for 10 times RANSAC threshold.

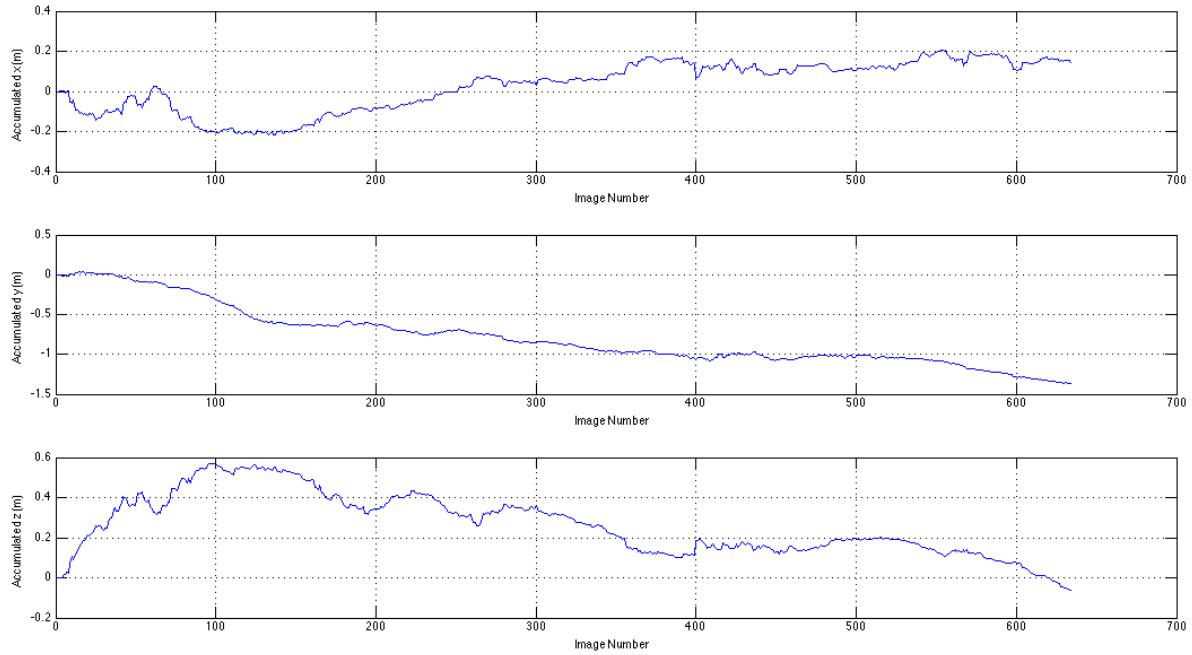


Figure 12: The accumulated position drift of the camera frame across all successive images.

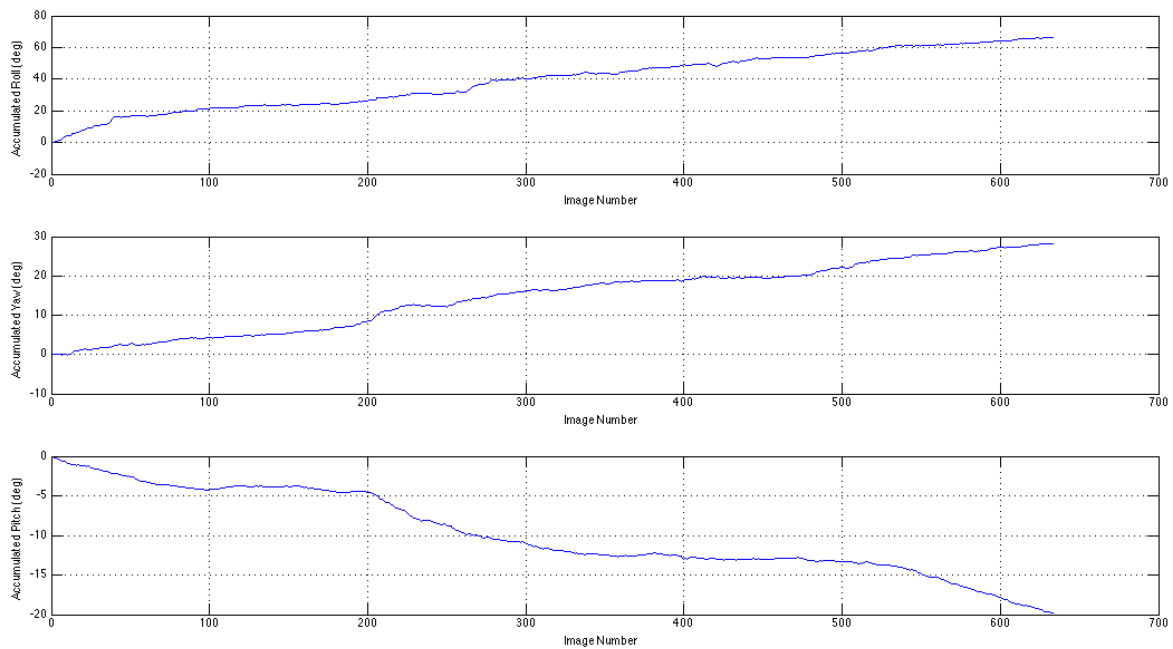
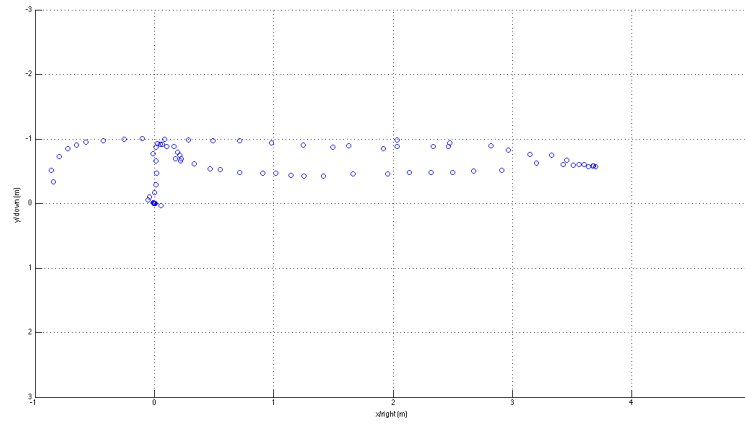
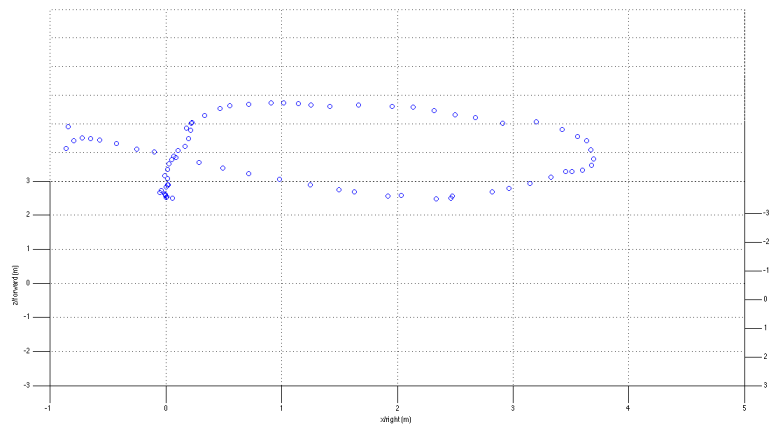


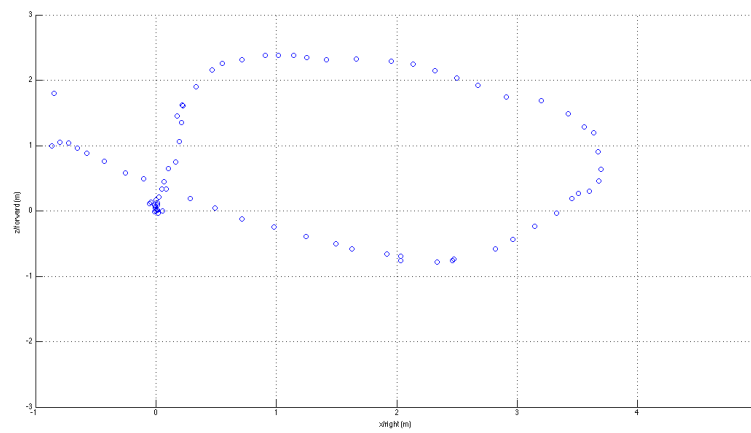
Figure 13: The accumulated orientation (in ZYX Euler angles) drift of the camera frame across all successive images.



(a) Front view.



(b) Perspective view.



(c) Top view.

Figure 14: The trajectory generated using incremental rotations and translations between every seven images.

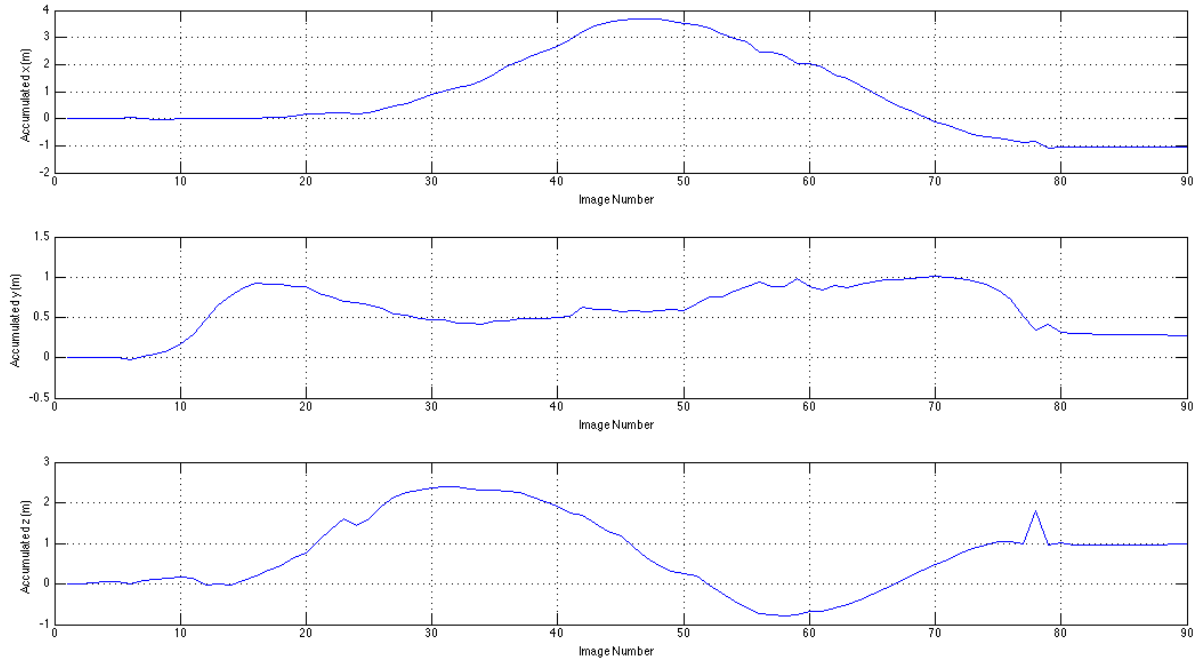


Figure 15: The accumulated position of the camera frame between every seven images.

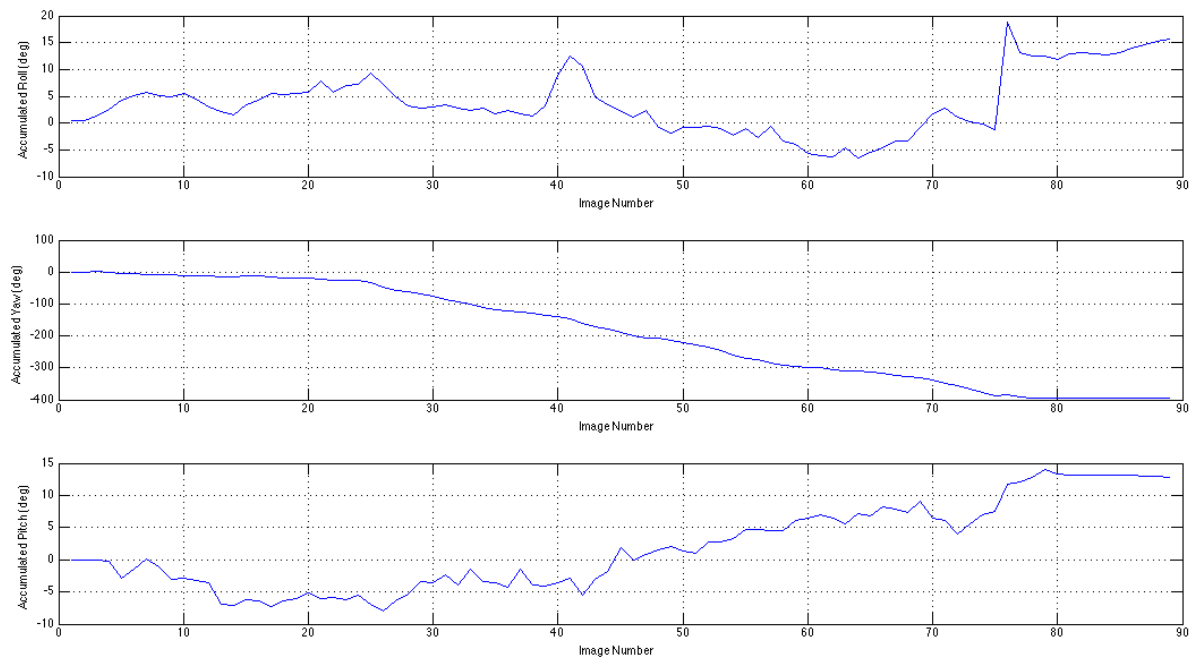


Figure 16: The accumulated orientation (in ZYX Euler angles) of the camera frame between every seven images.