ENEE 739J Project IV

Submitted by Amit Agrawal

1 Problem Statement

Implement Structure and Motion Algorithm using Factorization Method for orthographic projection and para-perspective projection and compare the results with EKF based approach.

2 Implementation

In this project, Structure from Motion algorithm using factorization theorem is implemented for perspective and orthographic projections. The results are compared quantitatively with EKF based fusion approach 1.

2.1 Factorization theorem

Tomasi and Kanade [3] showed that for orthographic projection, the shape and motion of a single 3D object can be recovered from image features using a factorization based approach. They form a $2F \times P$ matrix composed of P features from F frames. This matrix can be decomposed into a shape matrix S and motion matrix M as $W = M \times S$. Poelman [4] extended the factorization theorem to handle para-perspective projection. Para-perspective projection is a better approximation to perspective projection and can better handle the non-linearities of the perspective projection as compared to the orthographic projection which can be used in cases where the object is far away from camera and the depth variation of the object is smaller compared to the distance from the camera.

3 Results

We use a synthetic face data sequence to generate the feature points and analyze the shape and motion parameters recovery. We will consider orthographic, para-perspective and EKF based multi-frame fusion approach under no noise and two pixel random noise in image measurements.

A synthetic face sequence data was generated for 100 frames. Each frame has 150 feature points. Random noise was added to the features. The translation and rotation between frames was assumed to be varying along a constant value. The beta value was set to 0.001 corresponding to focal length = 1000.

For EKF based approach implemented as in [1], the initial batch estimate on the synthetic data accurately estimated the alpha (depth) values and the beta parameters. From the batch estimate, the variances of the state vector noise was calculated. Using the initial batch solution, extended kalman filter was run to estimate the depths and focal lengths. For factorization based methods,

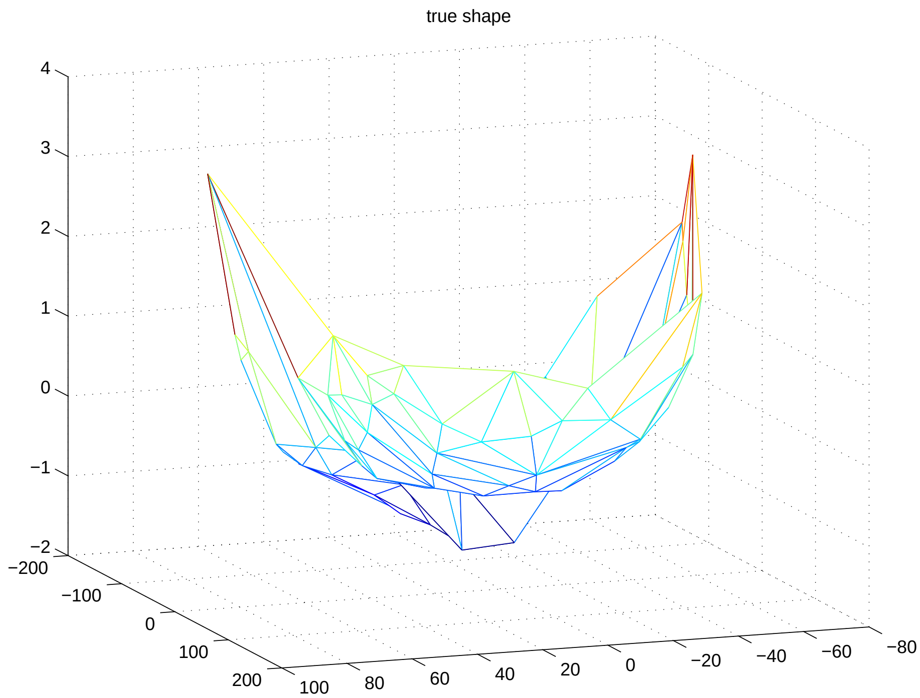

Figure 1: True Shape

SVD was used to decompose the measurement matrix according to the rank constraint. The motion matrix gives the rows of the rotation matrix for each frame which should be orthogonal and orthonormal. Using these constraints, a transformation Q was obtained and the shape and motion matrix were updated as

$$S = S * inv(Q); \tag{1}$$

$$M = Q * M; \tag{2}$$

Then the rotation for the first frame was made equal to $I_{3\times 3}$.

For comparing estimated depth values z with the true depths z, we will use the mean square error between the normalized estimated depth and true depth. The normalization is done by subtracting the mean and dividing by the standard deviation.

Figure 1 shows the true face shape where the depth has been normalized. The projection for $F = 50$ frames were computed using perspective projection for $P = 100$ points.

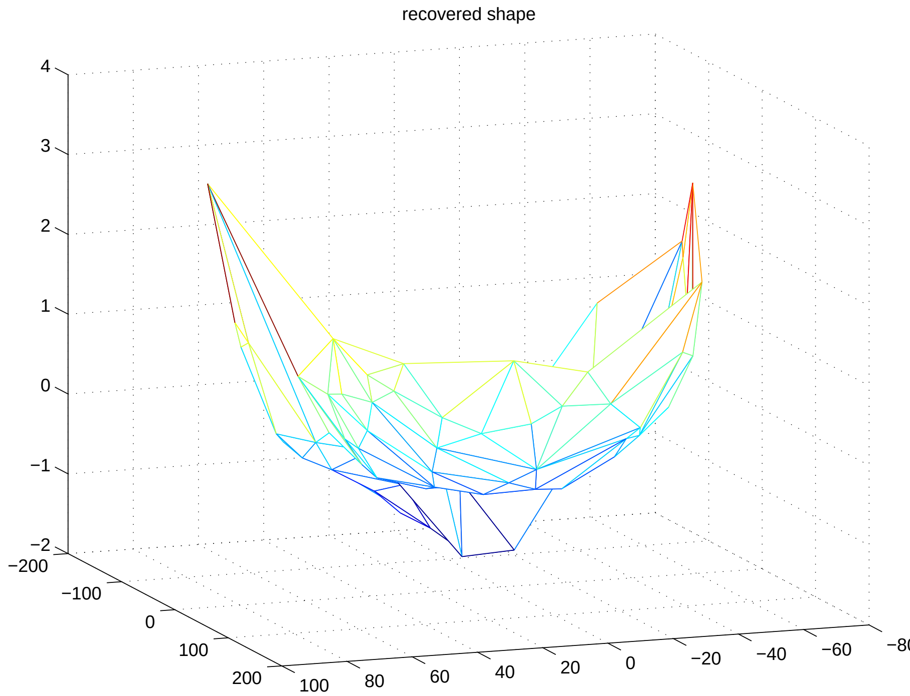

Figure 2: Estimated Shape: Para-perspective

3.1 Test 1

Here we have zero noise in the feature points. Note that the features has been generated using perspective projection, so even in zero noise case, there will be errors because the orthographic and para-perspective projection have been used as an approximation to the perspective projection. Figure 1 shows the true 3D shape of the face.

Figure 2, 3 4 shows the recovered shape using para-perspective projection which is close to the true one and the estimated translation and rotation between frames along with the true values. The errors in depth has been tabulated in Table 1.

Figure 5, 6 7 shows the recovered shape using orthographic projection and the estimated translation and rotation between frames along with the true values.

Figure 8, 9 10 shows the recovered shape using orthographic projection and the estimated translation and rotation between frames along with the true values.

We observe that in zero noise case, the translation and rotation estimates obtained from EKF are not as good as that from factorization theorems but the

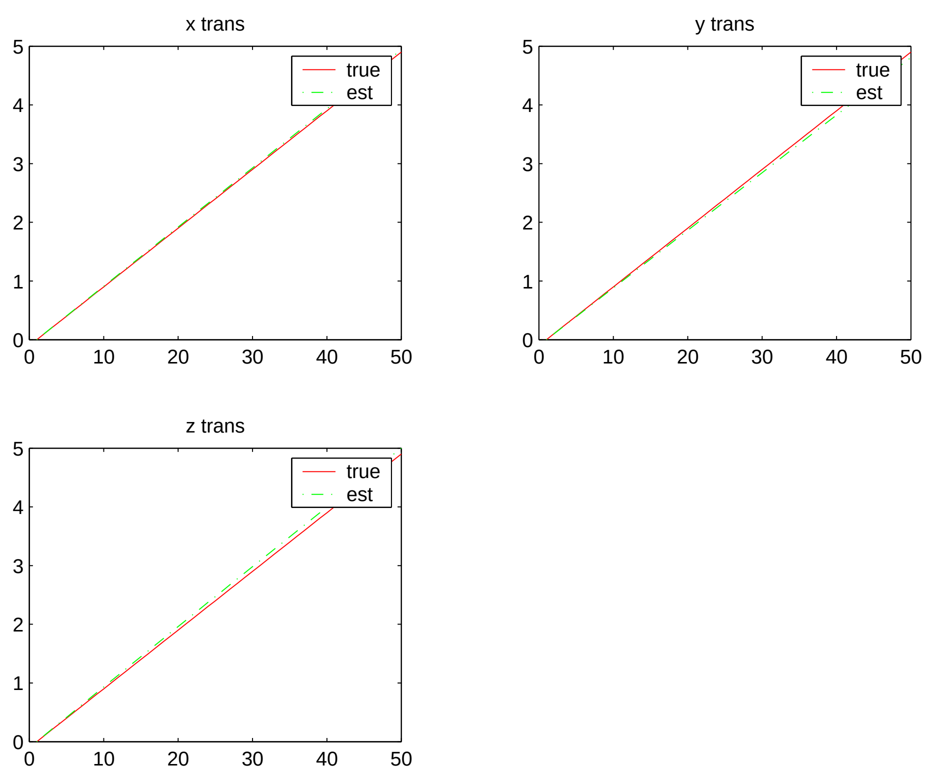

Figure 3: True and Estimated Translations: Para-perspective, zero noise

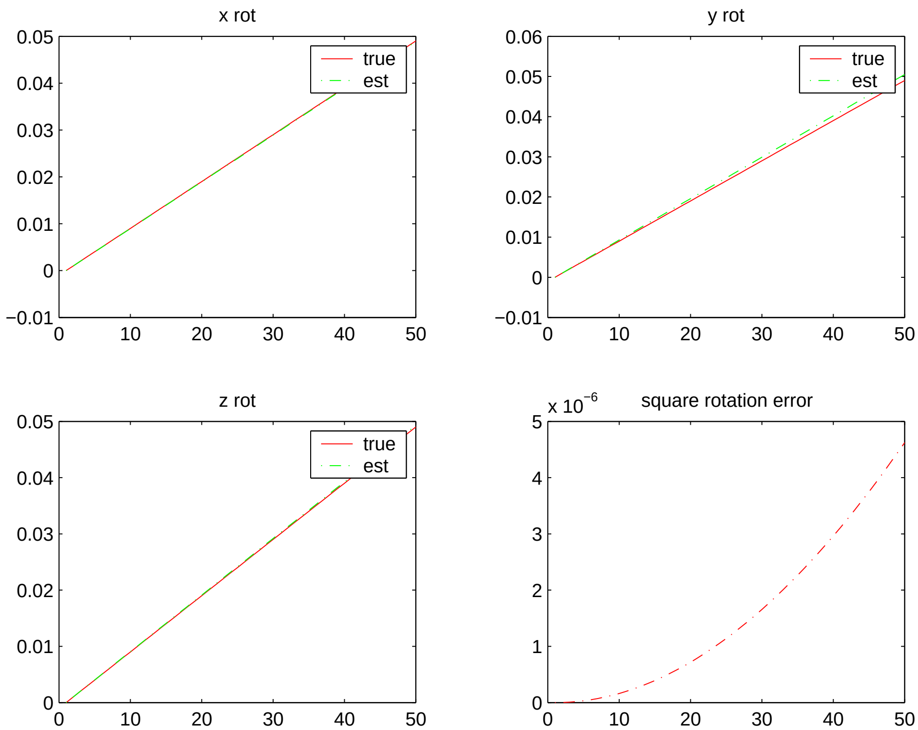

Figure 4: True and Estimated Rotations: Para-perspective, zero noise

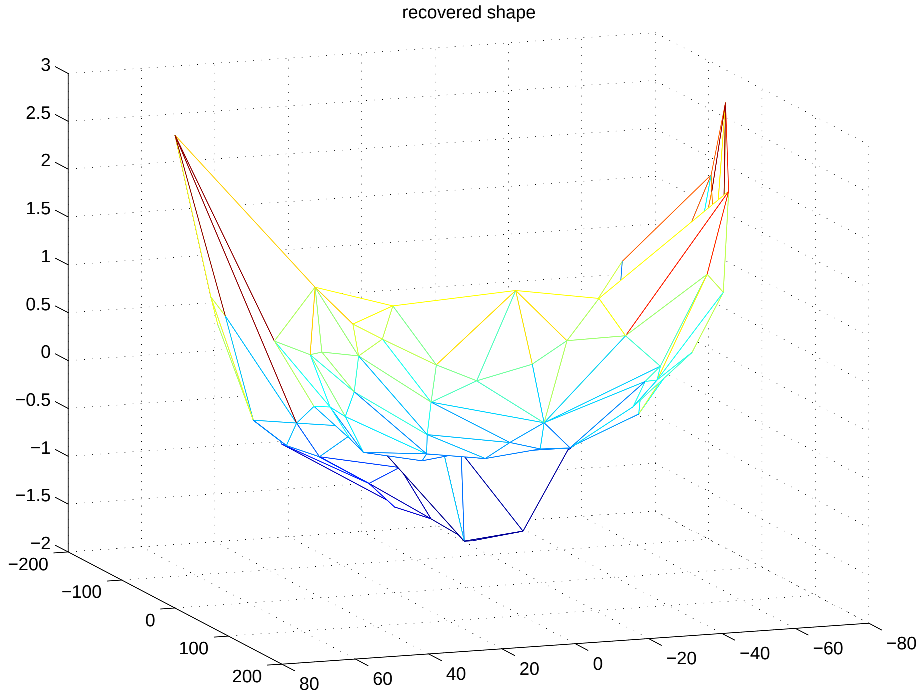

Figure 5: Estimated Shape: Orthographic

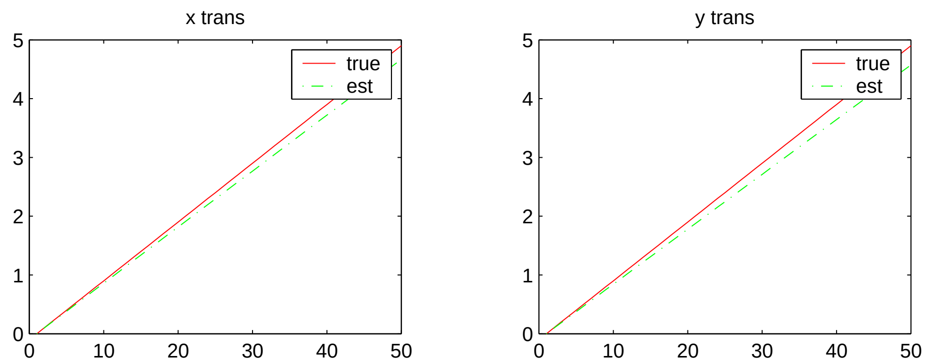

Figure 6: True and Estimated Translations: Orthographic, zero noise

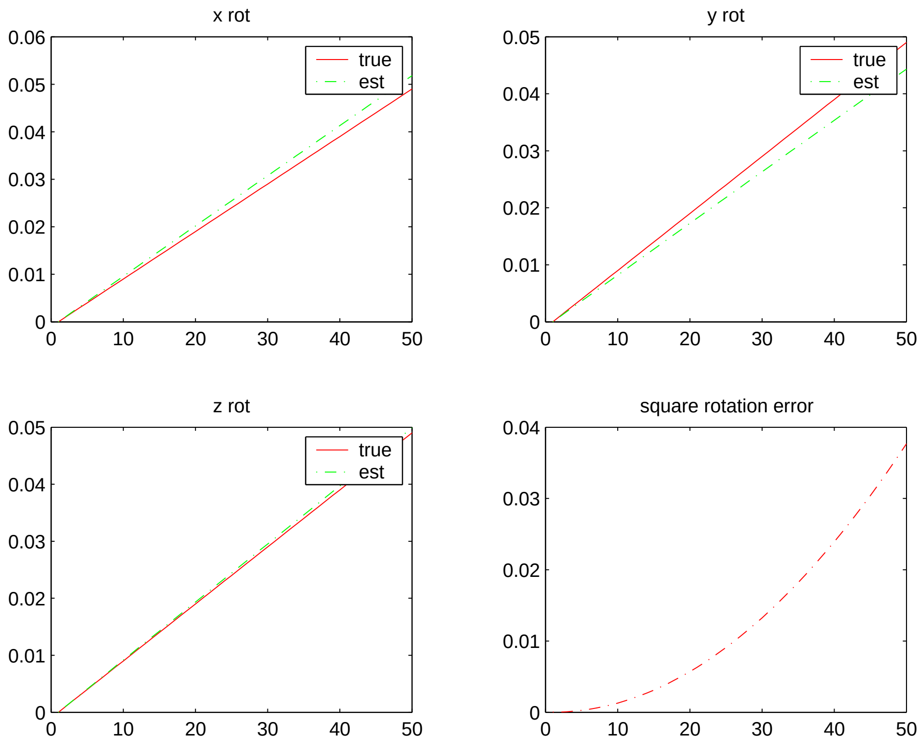

Figure 7: True and Estimated Rotations: Orthographic, zero noise

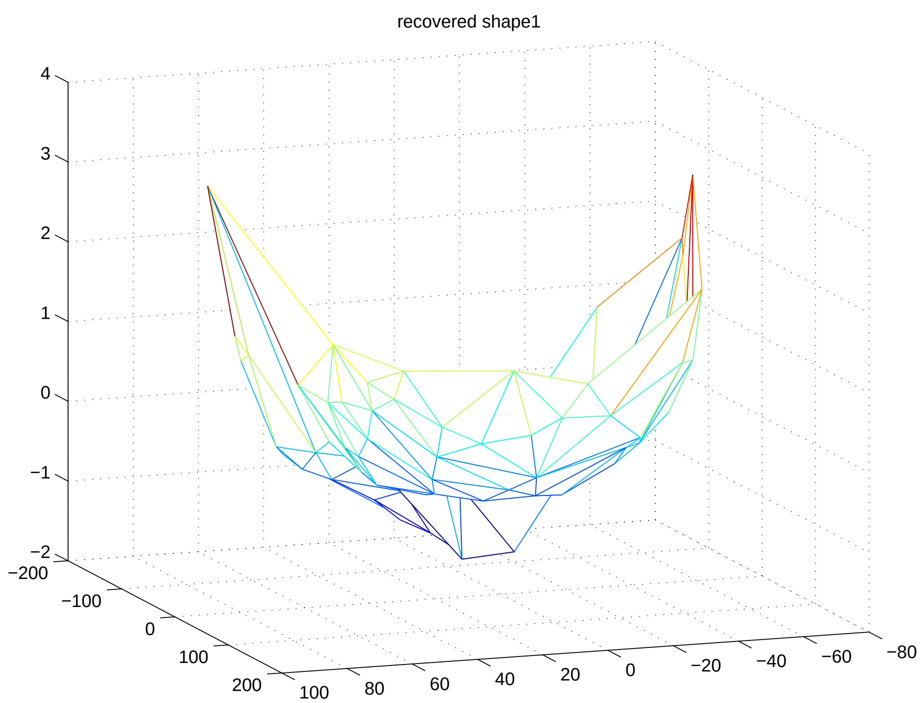

Figure 8: Estimated Shape: EKF

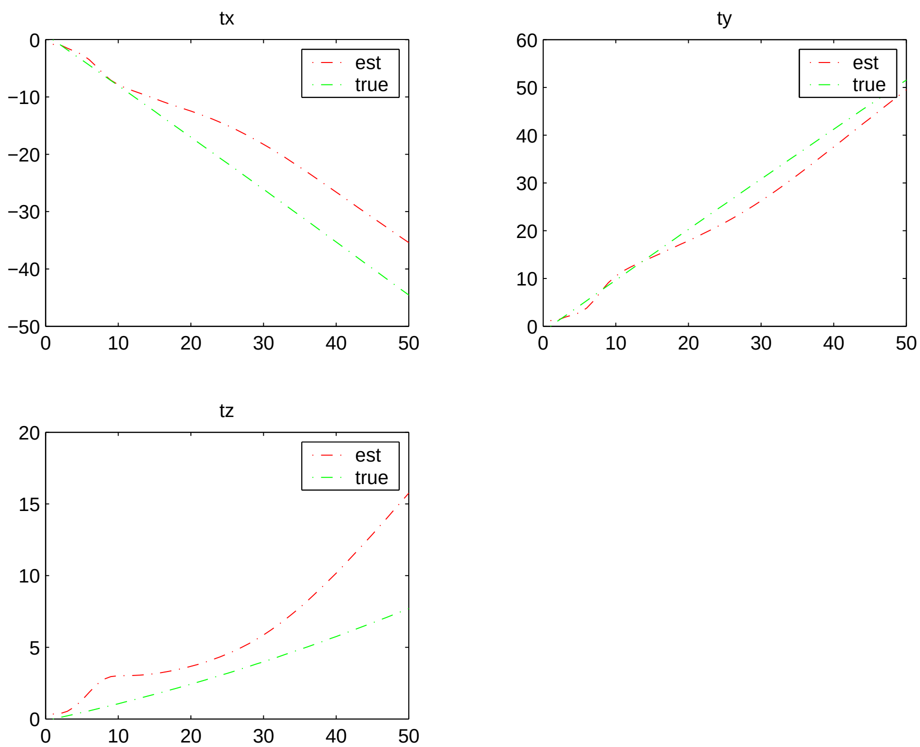

Figure 9: True and Estimated Translations: EKF, zero noise

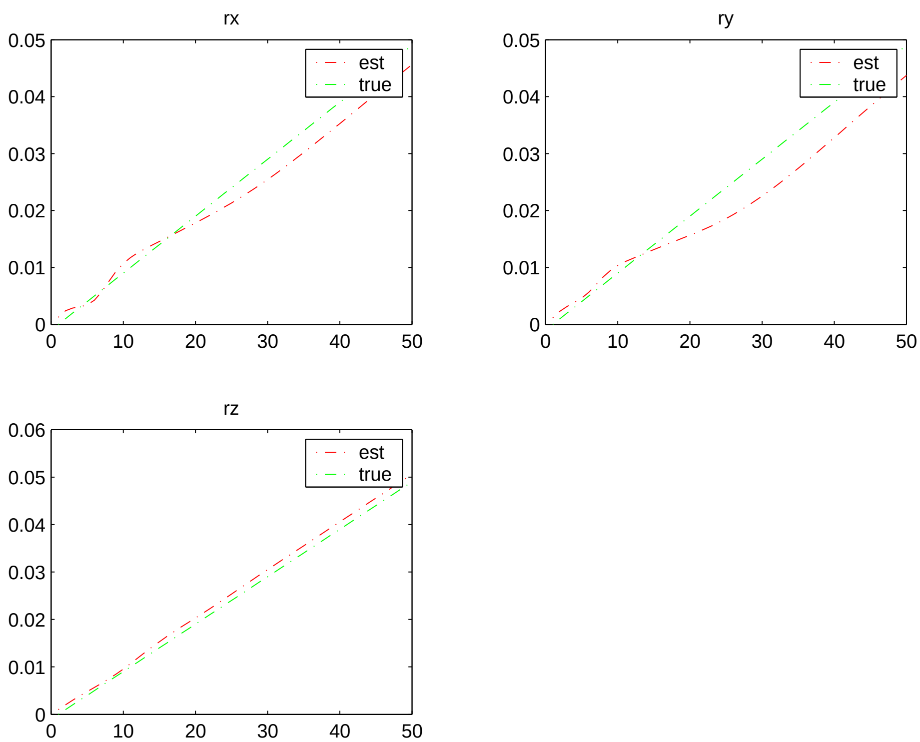

Figure 10: True and Estimated Rotations: EKF, zero noise

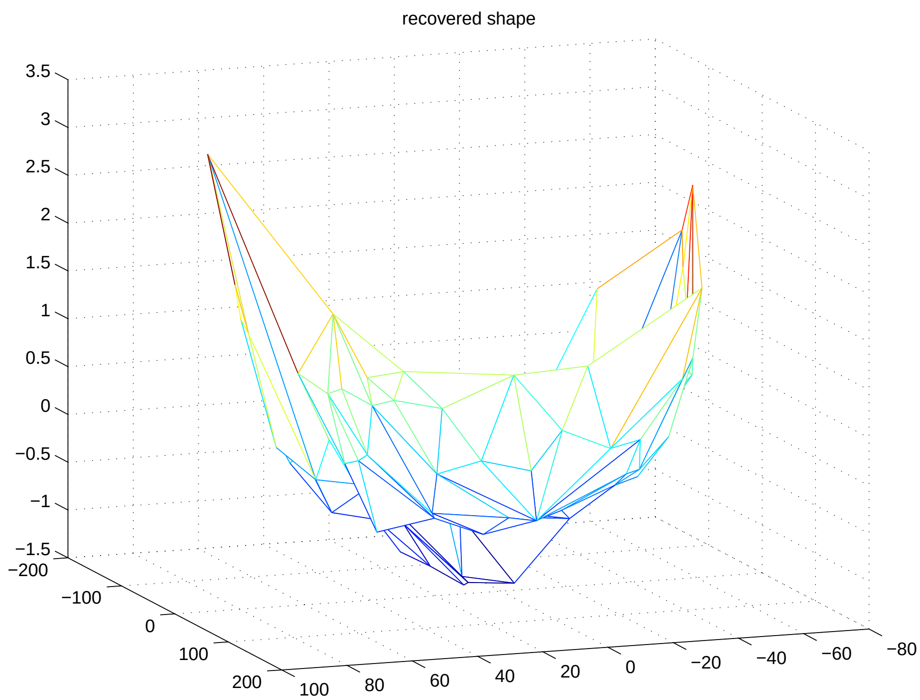

Figure 11: Estimated Shape: Para-perspective

depth estimate from EKF is better than that obtained from factorization. The translation and rotation estimates obtained from para-perspective projection are closer to the true values as compared to orthographic projection.

3.2 Test 2

Here we add 2 pixel random noise to the feature points and estimate the motion and shape.

Figure 11, 12 13 shows the recovered shape using para-perspective projection and the estimated translation and rotation between frames along with the true values. The errors in depth has been tabulated in Table 1.

Figure 14, 15 16 shows the recovered shape using orthographic projection and the estimated translation and rotation between frames along with the true values for 2 pixel noise.

Figure 17, 18 19 shows the recovered shape and the estimated translation and rotation between frames along with the true values using EKF.

Table 1 shows the mean square error between the normalized estimated depth values and the normalized true depth values for different approaches and noise values. We observe that the fusion using EKF approach gives the best result

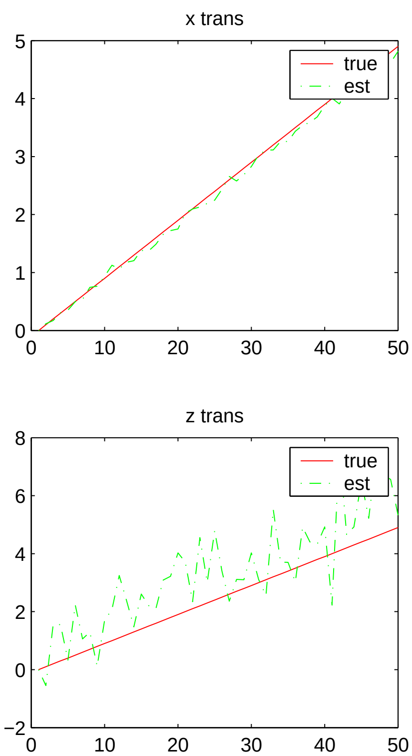

Figure 12: True and Estimated Translations: Para-perspective, 2 pixel noise

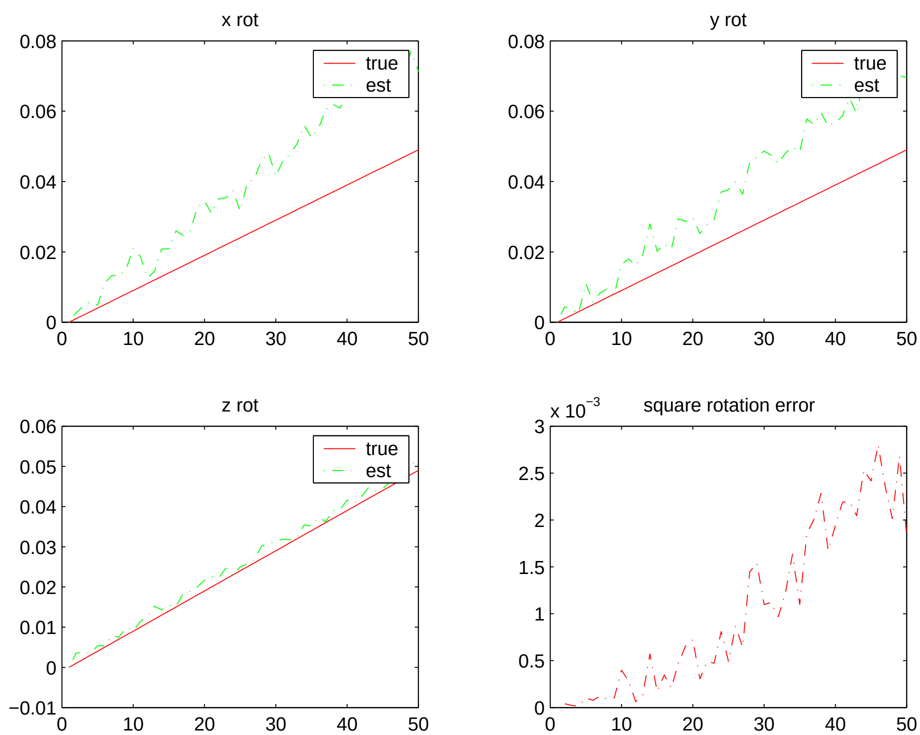

Figure 13: True and Estimated Rotations: Para-perspective, 2 pixel noise

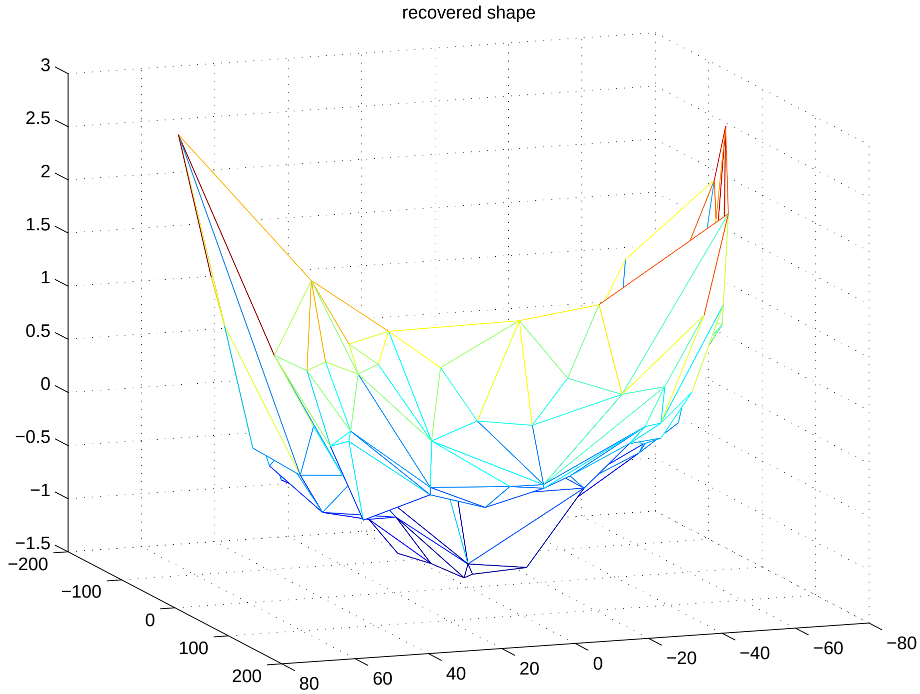

Figure 14: Estimated Shape: Orthographic

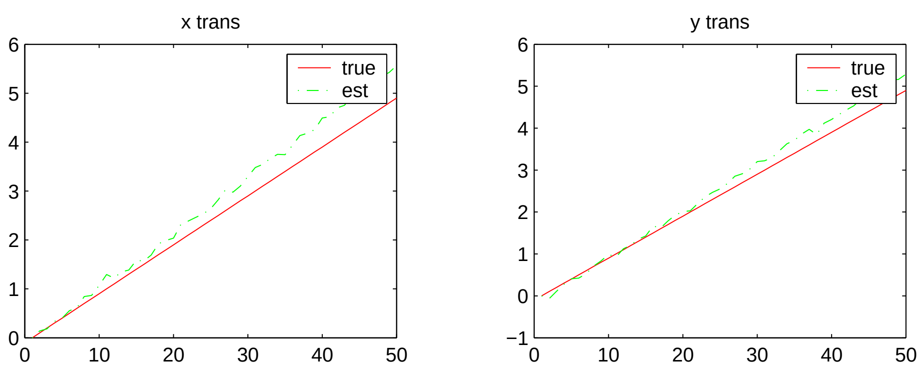

Figure 15: True and Estimated Translations: Orthographic, 2 pixel noise

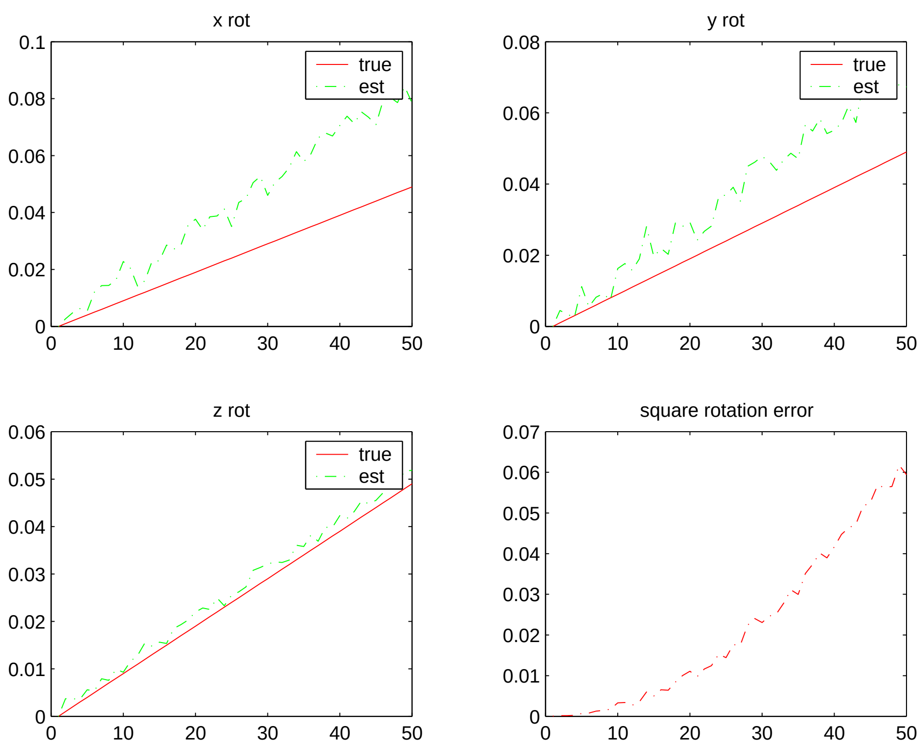

Figure 16: True and Estimated Rotations: Orthographic, 2 pixel noise

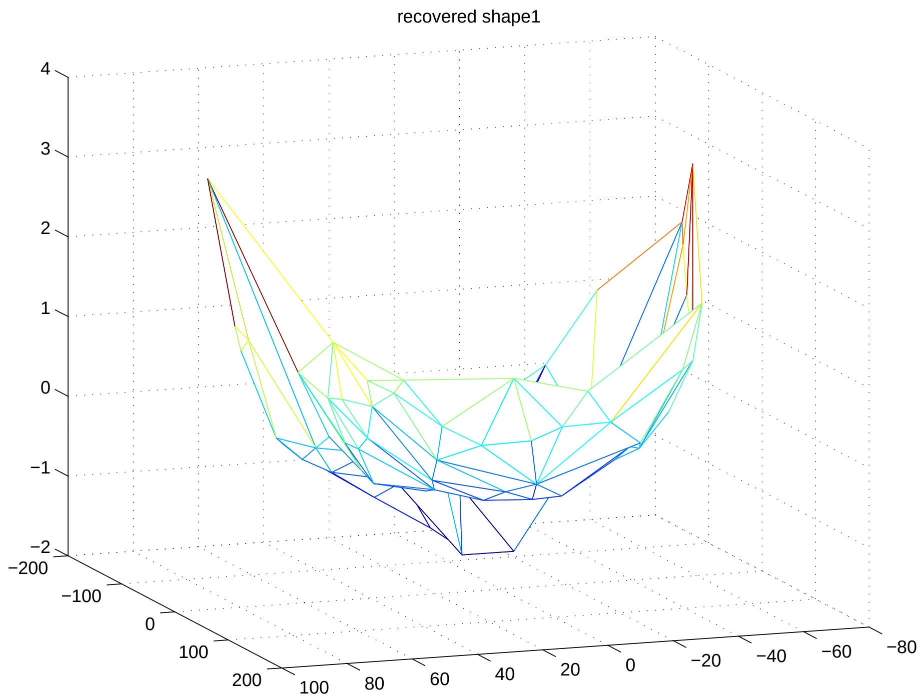

Figure 17: Estimated Shape: EKF

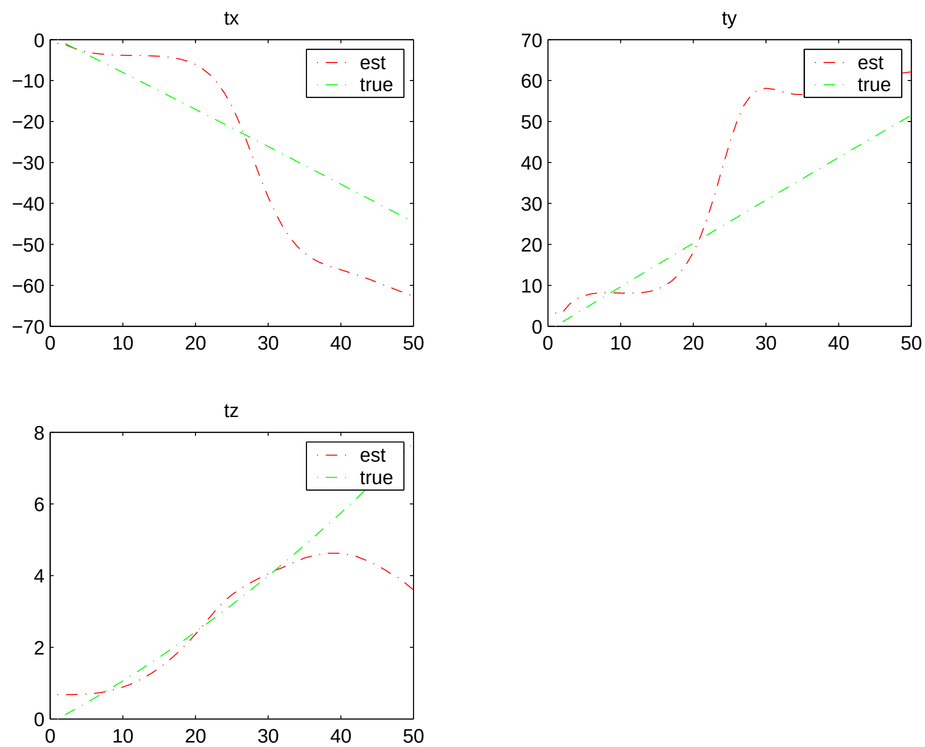

Figure 18: True and Estimated Translations: EKF, zero noise

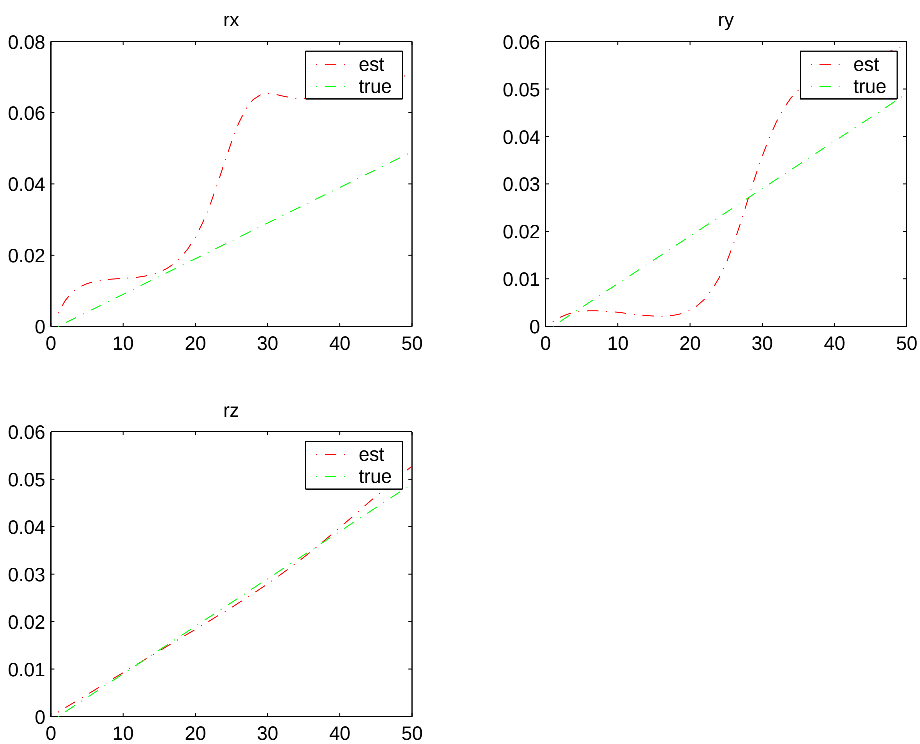

Figure 19: True and Estimated Rotations: EKF, zero noise

Table 1: Mean square error between normalized estimated depth and normalized true depth

Method	Noise=0	Noise = 2 pix
Ortho	0.02880	0.033
Para	0.01040	0.040
EKF	0.00284	0.006

for depth estimates as it optimally combines the measurements to estimate depth. However, the translation and rotation estimates are less accurate as compared to factorization method. Also, the estimates obtained using para-perspective model are better than those obtained using orthographic model as para-perspective is a better approximation to the perspective projection.

References

[1] Azarbayejani A, Pentland AP, "Recursive Estimation of Motion, Structure and Focal Length", IEEE Trans. on PAMI, Vol. 17, NO. 6, June 1995

[2] Broida TJ, Chandrashekhar S, Chellappa R, "Recursive 3-D Motion Estimation from a Monocular Image Sequence", Signal and Image Processing Institute, USC, CA

[3] Tomasi C, Kanade T, "Shape and Motion from Image Streams: A factorization Method", Cornell TR 92-1270 and CMU-CS-92-104

[4] Poelman CJ, Kanade T, "A Paraperspective Factorization Method for Shape and Motion Recovery", Dec 1993, CMU-CS-93-219