University of Burgundy

MsCV

VISUAL PERCEPTION

Lab 2

by

Gopikrishna Erabati

Supervisor: Dr. David Fofi

1. Compute the Essential Matrix E, knowing R and T (Ground Truth)

In my simulation I had taken a point cloud as a3D scene. And my two camera parameters are as below

$$P_1 = \begin{bmatrix} 50 & 0 & 0 \\ 0 & 50 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{R} & \boldsymbol{t} \\ \boldsymbol{0} & 1 \end{bmatrix}$$

where, $\boldsymbol{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ and $\boldsymbol{t} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}'$

and,

$$P_2 = \begin{bmatrix} 50 & 0 & 0 \\ 0 & 50 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{R} & \boldsymbol{t} \\ \boldsymbol{0} & 1 \end{bmatrix}$$

where, $\boldsymbol{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ and $\boldsymbol{t} = \begin{bmatrix} -t_{xx} & 0 & 0 \end{bmatrix}'$

"here negative sign (-) because its translation of origin of world coordinate frame with respect to camera2"

To compute the essential matrix, we need to know the translation and rotation of camera1 with respect to camera2 [1],

$$E = [t]_x R$$

where, t = [-400 0 0]' , as I took a translation of 400 world units.

and R is an identity matrix as there is no rotation involved.

The result is as shown in Figure.1

```
***************************************************************
The computed Essential matrix is
     0       0       0
     0       0     400
     0    -400       0

***************************************************************
```

Figure 1 Computed Essential matrix (ground truth)

2. Extract Rotation(R) and Translation(T) matrices from Essential matrix(E) and
3. Cheirality constraint

R and T matrices are extracted using [2],

Let the singular value decomposition of the essential matrix be $E \sim U\, diag(1,1,0)V'$, *where U and V are choosen such that det(U) > 0 and det(V) > 0 then*

$$t \sim t_u \;\equiv\; [u_{13} \quad u_{23} \quad u_{33}]^T$$

and,

$$R_a = UDV^T, R_b = UD^TV^T$$

where $D = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Any combination of R and t according to the above prescription satisfies the epipolar constraint.

In order to solve the inherent ambiguities and to choose corresponds to the true configuration, the cheirality constraint is imposed [2].

The svd decomposition of E is [U S V] = svd(E).

In order to satisfy the condition of det(U) > 0 and det(V) > 0, we multiply the U with det(U) and V with det(V).

The result is as shown in Figure 2.

```
************************************************************
The estimated rotation is
     1     0     0
     0     1     0
     0     0     1

The estimated translation is
    -1
     0
     0

The estimated essential matrix is
     0     0     0
     0     0     1
     0    -1     0

************************************************************
```

Figure 2 R and T from E

As we can see from Fig. 2, the translation and essential matrices are scaled versions of original translation and essential matrices.

4. 3D reconstruction of scene using estimated rotation and translation

In this approach, we take camera1 at origin and camera2 with extrinsic parameters as estimated rotation and translation matrices and we triangulate the 2D points to get the 3D scene.

The result is as shown in Figure 3.

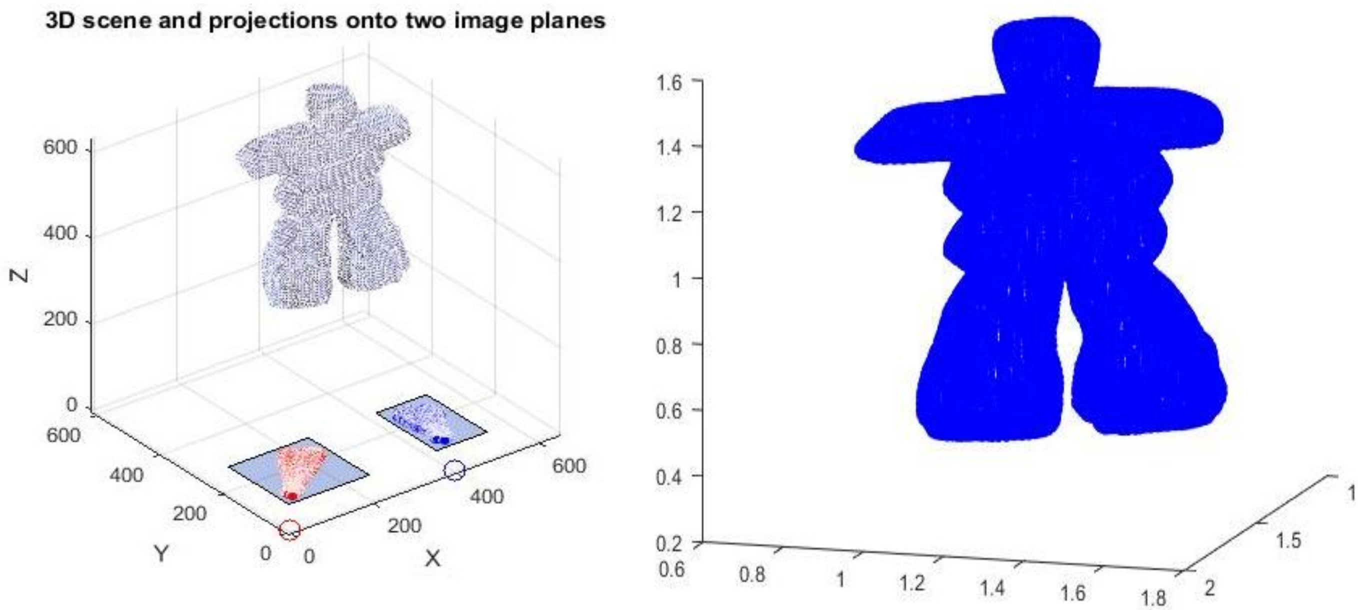

Fig. 3Original 3D scene and projections (left), Reconstructed 3D scene from estimated R and T(right)

As we can see in Fig. 3, the scale of 3D reconstruction is not preserved as we lost the scale during the estimate of R and T from E.

I also tried to view the reconstructions from all four R and T possibilities as shown in fig. 4.

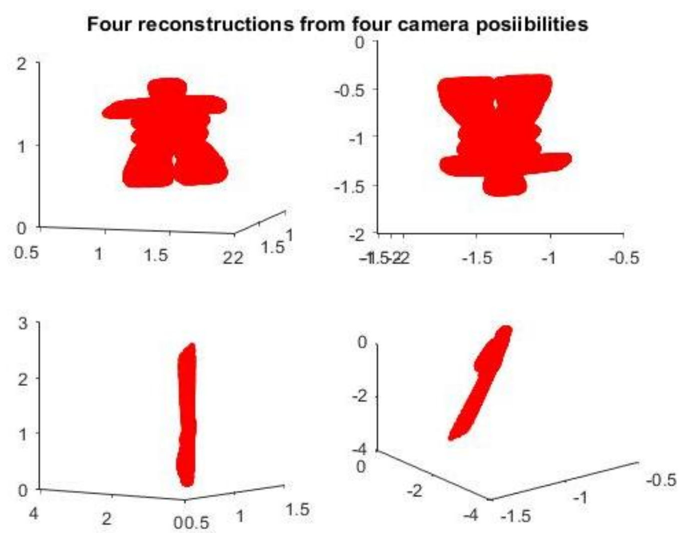

Fig 4 Reconstruction of 3D scene from four R and T possibilities

From the fig.4 we can see that the reconstruction is same (except the negative coordinate frame) because here we are just translating the camera to a opposite direction in the same collinear line. So, when we translate a camera , the 3D points appears to slide along parallel rails, so we get affine reconstruction.

5. Estimate Essential matrix from 2D set correspondences

The essential matrix can be estimated from 2D set correspondences using an Eight-point-algorithm[3],

$$p_2^T E p_1 = 0$$

where, p_1 and p_2 are normalised pixel cordinates

$$vec(p_2^T E p_1) = 0$$

$$(p_2^T \otimes p_1^T)\, vec(E) = 0$$

By using the above equations and using SVD we can solve for 'E'.

The result is as shown in fig. 5.

```
***********************************************************
The estimated essential matrix from point correspondances (Eight Point Algo) is
   -0.0000    0.0000   -0.0000
   -0.0000    0.0000    0.7071
    0.0000   -0.7071   -0.0000

***********************************************************
```

fig. 5 Estimated essential matrix from 2D point correspondences (eight point algorithm)

We can see that the estimated essential matrix also differs by scale factor.

I provide this estimated essential matrix from point correspondences to again estimate R and T and the results are as shown in fig.6

```
*********************************************************
The estimated essential matrix from point correspondances (Eight Point Algo) is
   -0.0000    0.0000   -0.0000
   -0.0000    0.0000    0.7071
    0.0000   -0.7071   -0.0000

*********************************************************
*********************************************************
The estimated rotation is
    1.0000    0.0000   -0.0000
   -0.0000    1.0000    0.0000
    0.0000   -0.0000    1.0000

The estimated translation is
   -1.0000
   -0.0000
   -0.0000

The estimated essential matrix is
   -0.0000    0.0000   -0.0000
   -0.0000   -0.0000    1.0000
    0.0000   -1.0000   -0.0000

*********************************************************
```

Fig. 6 Results of estimating R and T from essential matrix (estimated from point correspondences)

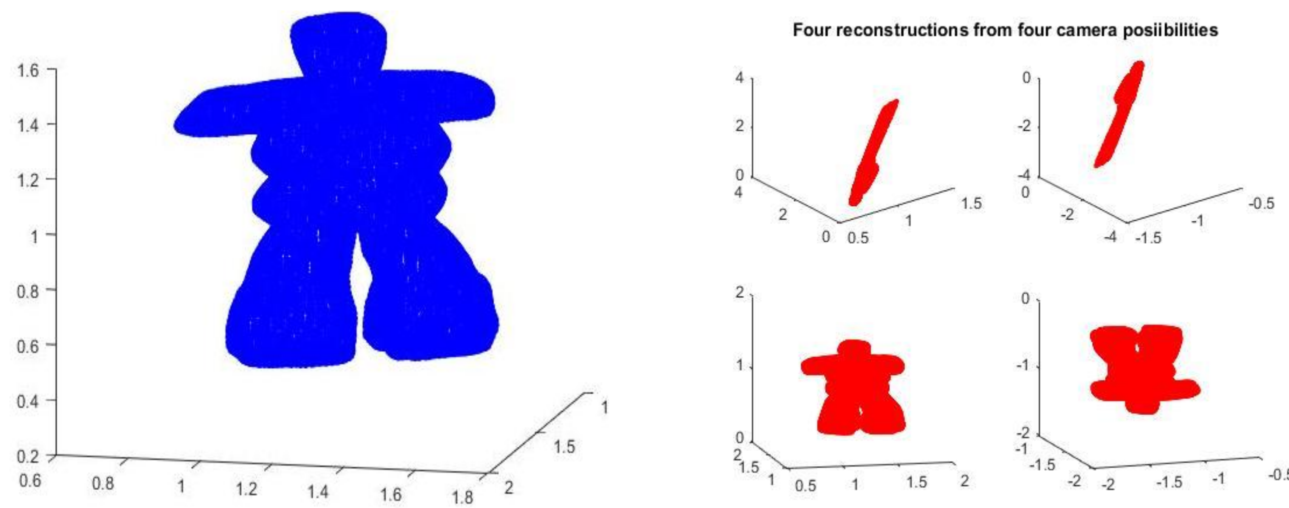

Fig. 7 Reconstruction using estimated R and T from essential matrix (estimated from point correspondances) (left), four possible reconstruction (right)

From above results we can say that, the estimation of R and T using estimated essential matrix (from point correspondences) also gives similar results as ground truth essential matrix, which also differs by scale.

References

[1] Lecture slides of Dr. David Fofi

[2] David Nister, "An Efficient Solution to the Five-Point Relative Pose Problem", IEEE Trans. in Pattern Analysis and Machine Intelligence, 2004.

[3] Eight point algorithm, Wikipedia