University of Burgundy

MsCV

VISUAL PERCEPTION

Lab 1

by

Gopikrishna Erabati

Supervisor: Dr. David Fofi

1. Pin Hole Camera Model

The most commonly used model is the so called pinhole camera. The model is inspired by the simplest cameras as shown in Fig. 1.

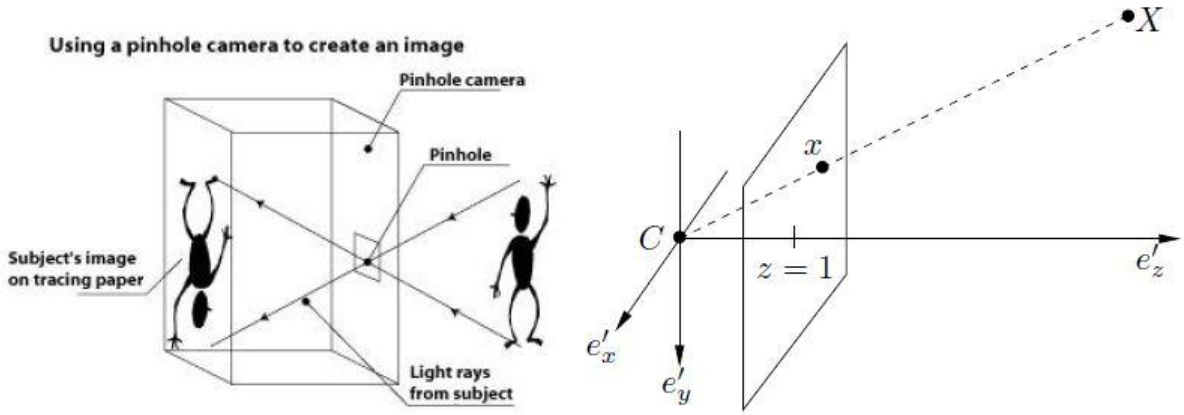

Figure 1 Pinhole camera model (left) Mathematical model (right)

To create a mathematical model we first select a coordinate system $[e'_x, e'_y, e'_z]$.We will refer to this system as the camera coordinate system. The origin C = (0, 0, 0) will represent the so called camera center (pinhole). To generate a projection $x = [x_1, x_{2,}, 1]$ of a scene point $X = [X'_1, X'_2, X'_3]$ we form the line between X and C and intersect it with the plane z = 1. We will refer to this plane as the image plane and the line as the viewing ray associated with x or X. The plane z = 1 has the normal e_z and lies at the distance 1 from the camera center. We will refer to ez as the viewing direction. Note that in contrast to a real pinhole camera we have placed the image plane in front of the camera center. This has the effect that the image will not appear upside down as in the real model.

1.1 Mappings

Here, we have 3 mappings from World coordinate system to Image plane. The four coordinate systems are as shown in Fig.2.

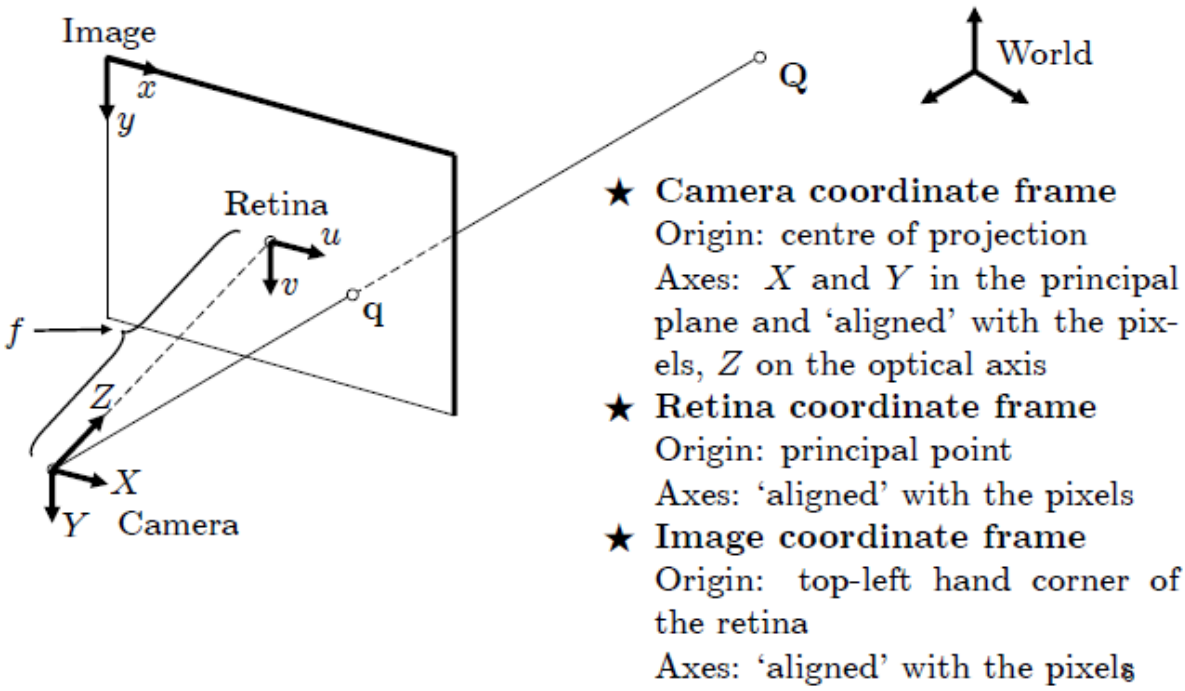

Figure 2 Coordinate frames in the model [1]

1.1.1 World to Camera coordinate system

This is a 3D displacement (Euclidean transformation).

$$Q_c = \begin{bmatrix} \boldsymbol{R} & \boldsymbol{t} \\ \boldsymbol{0} & 1 \end{bmatrix} Q$$

where, **R** is a Rotation matrix (3 x 3) with $R^T R = 1$ and det(R) = 1

t is a translational vector (3 x 1) "**translation of origin of world coordinate frame w.r.to camera coordinate frame**"

Q is world homogenous coordinate of 3D point

Q_c is camera coordinate of 3D point

These are extrinsic parameters.

1.1.2 Camera to Retina coordinate system

This is a projection.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix}$$

where, f is the focal length, i.e., the distance between center of projection to image plane.

1.1.3 Retinal to Image coordinate system

This is a 2D mapping

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} k_x & 0 & x_0 \\ 0 & k_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$$

where, k_x , k_y are density of pixels (pixels/mm)

u_0, v_0 is the translation from retinal to image coordinate system.

The complete mappings looks like:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \sim \begin{bmatrix} fk_x & 0 & x_0 \\ 0 & fk_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{R} & \boldsymbol{t} \\ \boldsymbol{0} & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

2. Exercise - Lab1

NOTE : A well commented code is attached with this report, there is a README.txt file in code folder which explains how to run the code and briefing about all other supporting functions.

In the lab tutorial, I had taken a 3D world scene as a point cloud. The focal length, I choose is 50 and I assume that my pixel density is same along X and y directions which is 1 pixel/mm and also my retinal and image plane are at same reference.

The camera 1 is at origin of world coordinate system and camera 1 is translated in X-direction by some value. So, here I define two camera matrices P_1 and P_2 for two cameras as :

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

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 $\boldsymbol{t} = [-t_{xx} \quad 0 \quad 0]'$

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

Task 1 : Projection of 3D points onto camera planes

We project the 3D points onto the camera planes by using the equations:

$$x = PX$$

where, X is 3D point, P is camera projection matrix and x is 2D point.

The results obtained are as shown in Fig.3.

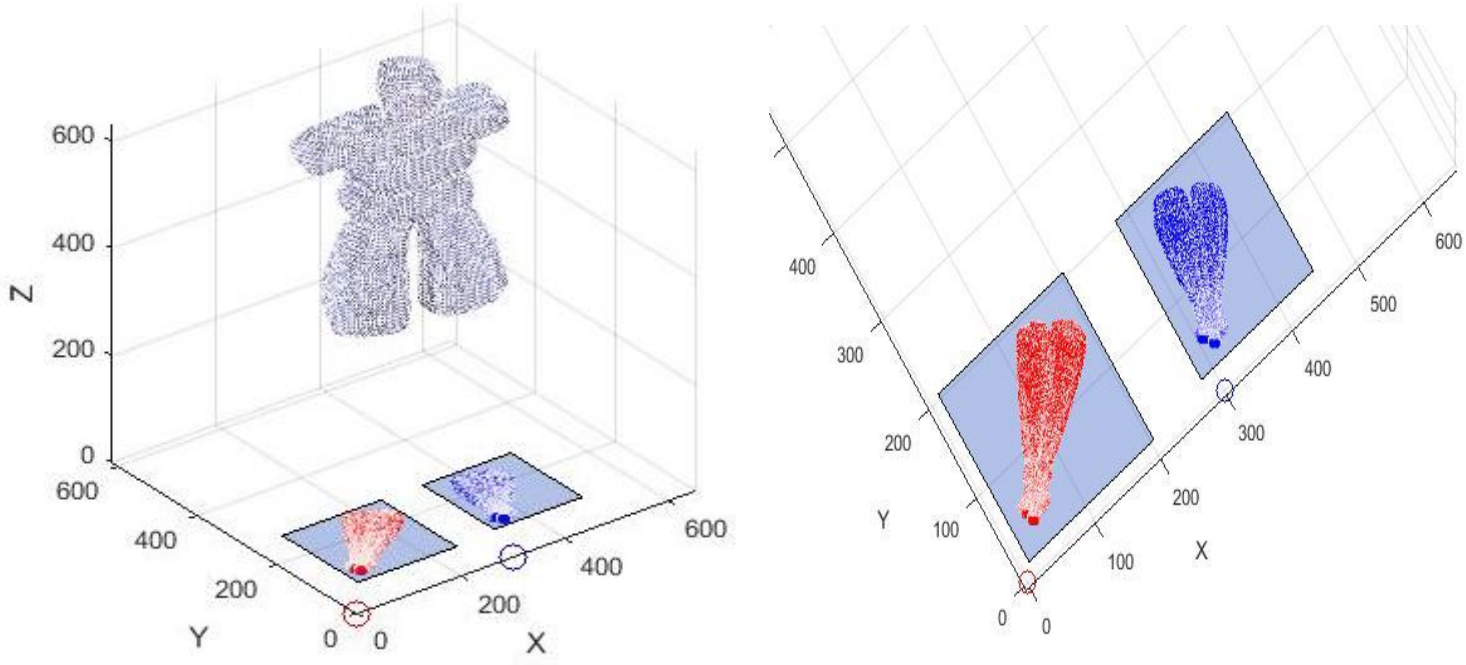

Figure 3 2D projections of 3D points for two cameras, located at origin and translated along X (300)

In the Fig.3, we can see the 3D scene as a point cloud and the two cameras located at origin and (300,0,0) as 'red' and 'blue' circles. The two patches have the projections of 3D points on them. The 'red' 2D points are projections with respect to camera1 and 'blue' 2D points are projections with respect to camera2.

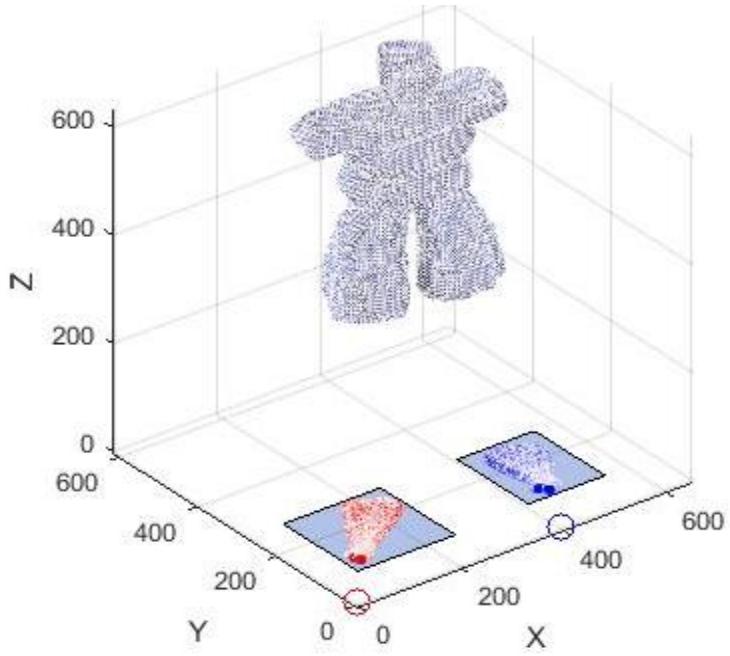

Figure 4 2D projections of 3D points for two cameras, located at origin and translated along X (400)

In the Fig.4, we can see the 3D scene as a point cloud and the two cameras located at origin and (400,0,0) as 'red' and 'blue' circles. The two patches have the projections of 3D points on them. The 'red' 2D points are projections with respect to camera1 and 'blue' 2D points are projections with respect to camera2.

We can clearly see from the above Figs. 3 and 4, as the camera is translated the projections will change accordingly.

I also drawn lines of projection from 3D coordinates to center of projection to visually check the 2D projections on image plane. This is as shown in Fig. 5

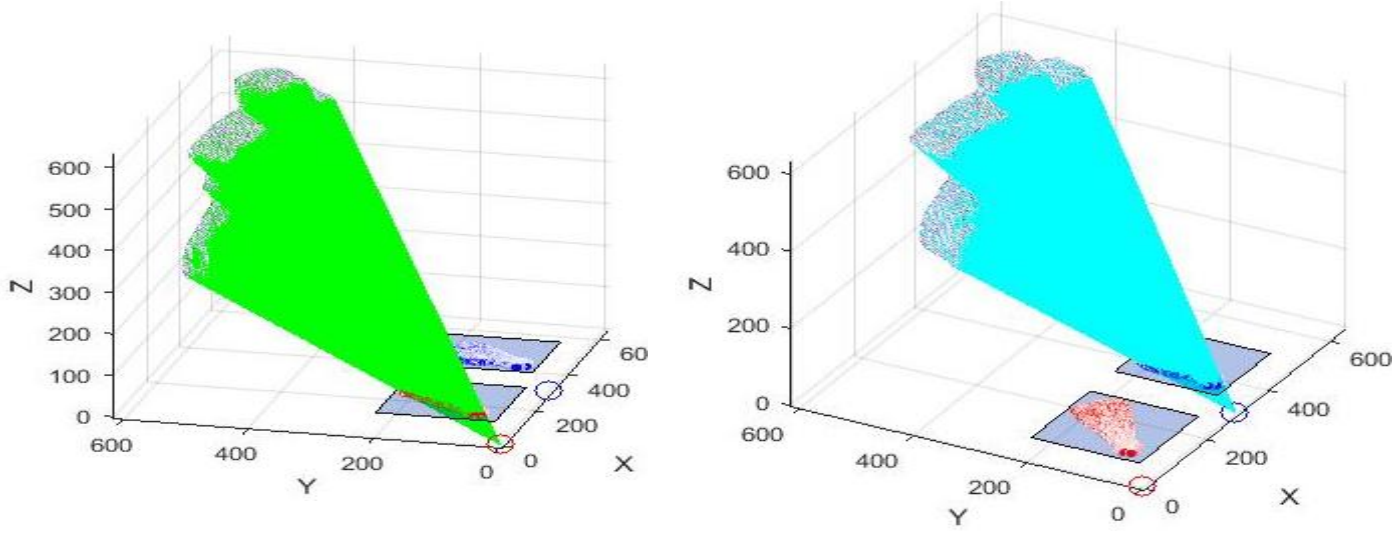

Fig. 5 Lines of projection from 3D coordinates to centers of projections

I tried to build images with the computed 2D points in a 512 x 512 image as shown in Fig. 6.

Fig. 6 Projected 2D points of 3D scene

Task2 : Compute Fundamental matrix using camera parameters

Epipolar geometry is the geometry of stereo vision[2]. When two cameras view a 3D scene from two distinct positions, there are a number of geometric relations between the 3D points and their projections onto the 2D images that lead to constraints between the image points.

The epipolar geometry is the intrinsic projective geometry between two views. It is independent of scene structure, and only depends on the cameras' internal parameters and relative pose.

The fundamental matrix F encapsulates this intrinsic geometry. It is a 3 × 3 matrix of rank 2. If a point in 3-space X is imaged as x in the first view, and x' in the second, then the image points satisfy the relation $x'Fx = 0$. The fundamental matrix is independent of scene structure. However, it can be computed from correspondences of imaged scene points alone, without requiring knowledge of the cameras' internal parameters or relative pose.

The main idea of epipolar geometry is as shown in Fig. 7.

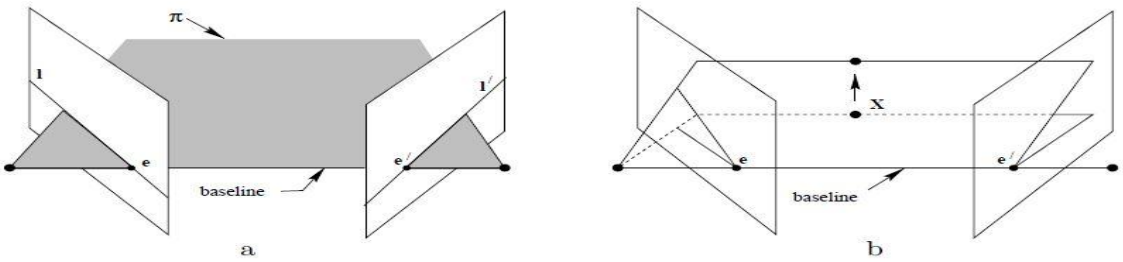

Fig. 8.2. **Epipolar geometry.** *(a) The camera baseline intersects each image plane at the epipoles* $\mathbf{e}$ *and* $\mathbf{e}'$*. Any plane* π *containing the baseline is an epipolar plane, and intersects the image planes in corresponding epipolar lines* $\mathbf{l}$ *and* $\mathbf{l}'$*. (b) As the position of the 3D point* $\mathbf{X}$ *varies, the epipolar planes "rotate" about the baseline. This family of planes is known as an epipolar pencil. All epipolar lines intersect at the epipole.*

Fig. 7 Epipolar Geometry [3]

The Fundamental Matrix, F

The fundamental matrix is the algebraic representation of epipolar geometry.

Given a pair of images, each point **x** in one image, there exists a corresponding epipolar line l' in the other image. Any point x' in the second image matching the point **x** must lie on the epipolar line l'. The epipolar line is the projection in the second image of the ray from the point **x** through the camera centre **C** of the first camera. Thus, there is a map from a point in one image to its corresponding epipolar line in the other image.

$$x \rightarrow l'$$

It will turn out that this mapping is a (singular) *correlation*, that is a projective mapping from points to lines, which is represented by a matrix F, the fundamental matrix.

The equation to find Fundamental matrix is:

$$F = K'^{-T}[t]_x R K^{-1}$$

where, K' is intrinsic parameter of camera2

K is intrinsic parameters of camera1

$[t]_x$ is skew symmetric matrix of translation of camera 2 with respect to camera1

R is roataion matrix of camera2 with respect to camera1.

With the selected stereo vision system as explained earlier, the result is as shown in Fig.8.

```
************************************************************
Computed Fundamental matrix from known camera parameters
     0     0     0
     0     0     6
     0    -6     0

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

Fig. 8 Computed Fundamental matrix from camera parameters (known) in MATLAB

Task 3 : Estimate the fundamental matrix from the two images

I used Salvi's [4] toolbox to estimate fundamental matrix from two images using RANSAC method. It takes the parameters as

1. 2D corresponding points of two images

2. Probability of F without outlier

3. Outlier ratio

The estimated fundamental matrix is as shown in Fig. 9

```
************************************************************
Method: RANSAC from Salvi Toolbox
Funamental Matrix
   -0.0000    0.0000   -0.0000
    0.0000   -0.0000    0.9999
    0.0000   -0.9999   -0.0001

Rank-2: 1
************************************************************
```

Fig. 9 Estimated Fundamental matrix using Salvi toolbox

As we can see from Fig. 8 and 9 that the fundamental matrix differs by a scale factor. And *Rank-2 shows '1' which means true*!

I also tried to estimate fundamental matrix using MATLAB's RANSAC method and the result is as shown in Fig. 10

```
************************************************************
Method: RANSAC from MATLAB Toolbox
Funamental Matrix
    0.0000    0.0000   -0.0000
   -0.0000   -0.0000    1.0000
    0.0000   -1.0000    0.0000

Rank-2: 1
************************************************************
```

Fig. 10 Estimated Fundamental matrix using MATLAB's RANSAC

To **check the accuracy of estimated fundamental matrix** I tried to draw the epipolar lines on right image and check whether the 2D points correspondences on right image of that of left image are lying on epipolar lines or not.

Some words about Fundamental matrix by pure translation (as in this case):

In considering pure translations of the camera, one may consider the equivalent situation in which the camera is stationary, and the world undergoes a translation $-\mathbf{t}$. In this situation points in 3-space move on straight lines parallel to **t**, and the imaged intersection of these parallel lines is the vanishing point **v** in the direction of **t**. This is shown in Fig. 11.

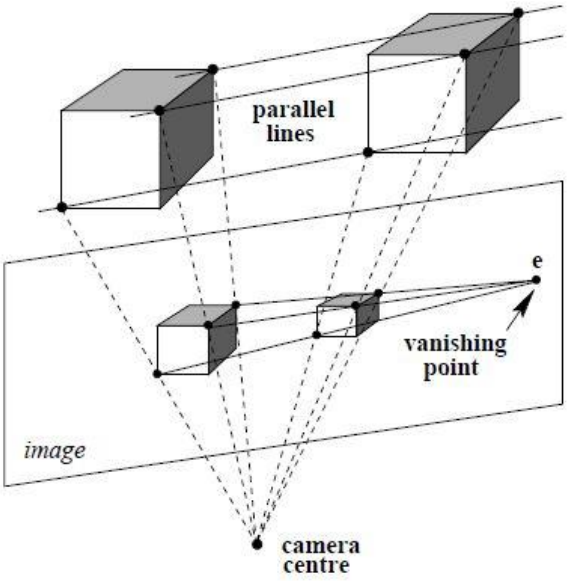

Fig. 9.7. *Under a pure translational camera motion, 3D points appear to slide along parallel rails. The images of these parallel lines intersect in a vanishing point corresponding to the translation direction. The epipole* e *is the vanishing point.*

Fig. 11 Pure translation motion [3]

As in our case, the internal parameters of two cameras are same and no rotation between cameras but only pure translations. So,

$P_1 = K[I \quad 0]$ and $P_2 = K\,[I \quad t]$

then $F = [e']_x$.

If translation is parallel to X-axis then $e' = [t_x\ 0\ 0]$, so

$$F = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -t_x \\ 0 & t_x & 0 \end{bmatrix}$$

It can be inferred from above that, F in case of pure translation is skew symmetric matrix and has 2 degrees of freedom. The epipolar line of x is $l' = Fx = [e']_x x$. and x lies on this line as $x^T[e']_x x = 0$. So, x and x' are collinear so as e and e'.

This collinearity property is termed *auto-epipolar*.

The epipolar lines drawn on second image in pure translational motion of camera2 is as shown in Fig. 12.

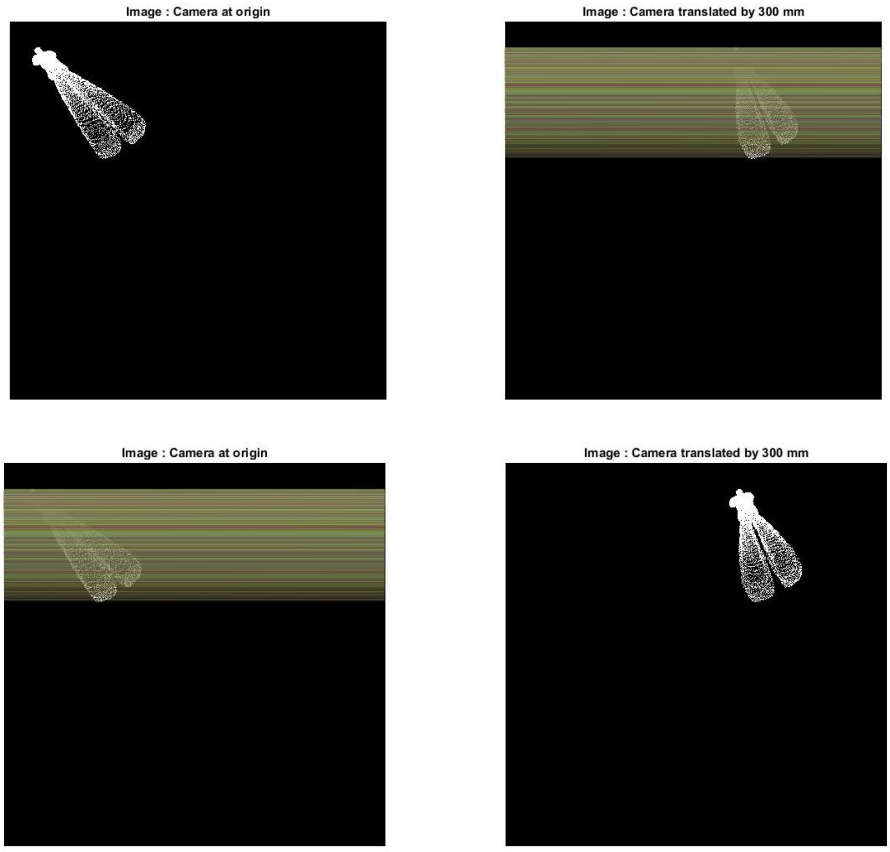

Fig. 12 Epipolar lines in pure translational motion of camera2

I tried to calculate the error of the correspondence points in image2 (of image1) lying on epipolar lines and the result is as shown in Fig. 13.

```
************************************************************
Error in corresponding points of image 1 lying on epipolar lines generated by points of image 2
6.6105e-05
************************************************************
Drawing Epipolar Lines...
Error in corresponding points of image 2 lying on epipolar lines generated by points of image 1
6.615e-05
************************************************************
```

Fig. 13 The error of points lying on epipolar lines

The error is very small as seen in Fig. 13, so the estimated fundamental matrix is accurate enough.

Task 4 : Compute the 3D scene from canonical representation

The fundamental matrix may be used to determine the camera matrices of the two views.

$l' = Fx$ and the correspondance equation $x'^T Fx = 0$ are *projective* relationships: the derivations have involved only projective geometric relationships, such as the intersection of lines and planes. Consequently, the relationships depend only on projective coordinates in the image, and not, for example on Euclidean measurements such as the angle between rays.

Similarly, F only depends on projective properties of the cameras P, P'. The camera matrix relates 3-space measurements to image measurements and so depends on both the image coordinate frame and the choice of world coordinate frame. F does not depend on the choice of world frame, for example a rotation of world coordinates changes P, P', but not F. In fact, the fundamental matrix is unchanged by a projective transformation of 3-space.

A given fundamental matrix determines the pair of camera matrices up to right multiplication by a projective transformation. Thus, the fundamental matrix captures the projective relationship of the two cameras.

So, what we compute the 3D scene from fundamental matrix and 2D points is only the projective reconstruction of original scene.

The camera matrices corresponding to fundamental matrix F can be choosen as

$$P = [I \mid 0] \; and \; P' = [[e']_x F \mid e']$$

Now, we can triangulate the 2D points in two images using the above camera matrices.

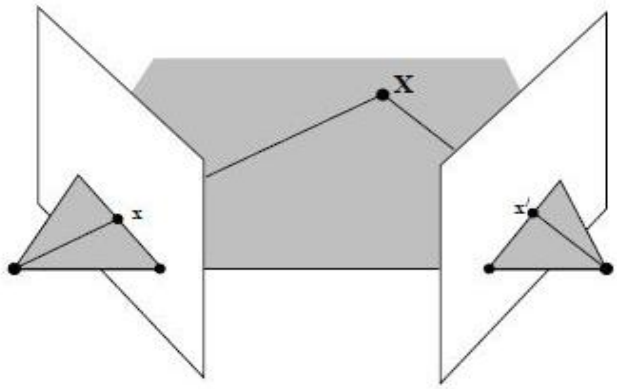

Fig. 10.1. Triangulation. *The image points* $\mathbf{x}$ *and* $\mathbf{x}'$ *back project to rays. If the epipolar constraint* $\mathbf{x}'^T \mathbf{F}\mathbf{x} = 0$ *is satisfied, then these two rays lie in a plane, and so intersect in a point* $\mathbf{X}$ *in 3-space.*

Fig. 14 Concept of triangulation [3]

The goal of triangulation is to recover the coordinates of 3D point,

we have,

$$x_1 \mathbb{x} P_1 X_1 = 0$$

$$x_2 \mathbb{x} P_2 X_2 = 0$$

The projective reconstruction theorem[3]:

If a set of point correspondences in two views determine the fundamental matrix uniquely, then the scene and cameras may be reconstructed from these correspondences alone, and any two such reconstructions from these correspondences are protectively equivalent.

Fig. 15 Reconstruction using canonical form (projective)

As in case of pure translational motion of camera2, with no rotation and no change in the internal parameters. for an affine reconstruction we may choose the two cameras as:

$$P = [I \mid 0] \; and \; P' = [I \mid e']$$

By taking the above camera parameters the result is as shown in Fig. 16.

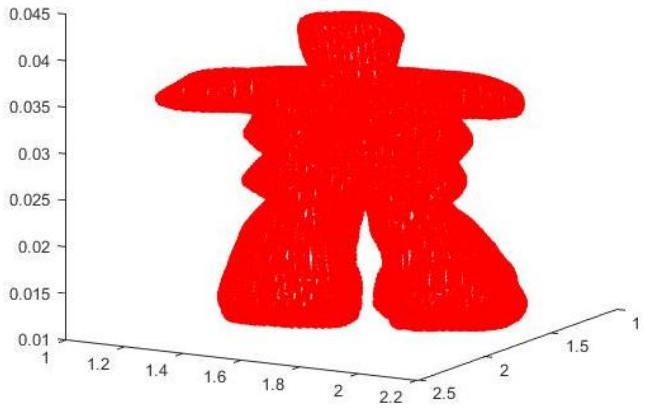

Fig. 16 Affine reconstruction for translation only motion of camera2

Task 5 : Compute the residual error

After we get the projective reconstruction from camera parameters in canonical form, we can project the 3D scene reconstructed using the camera parameters to form 2D points.

The result is as shown in Fig. 17

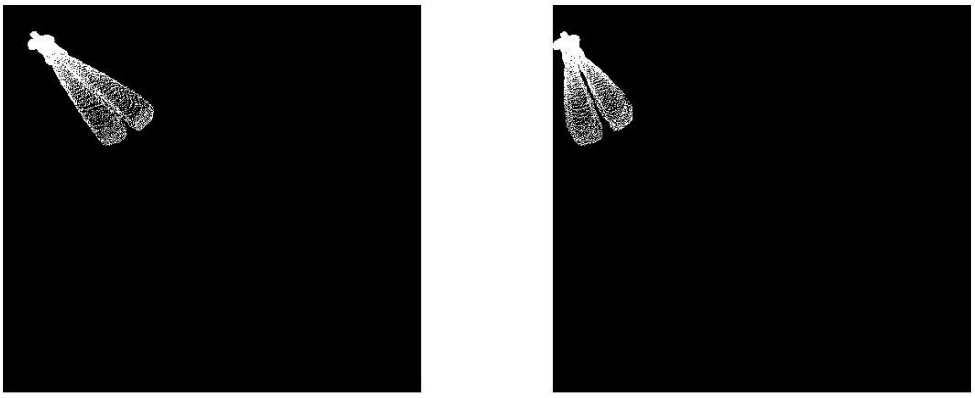

Fig. 17 Projections of 3D reconstruction using canonical camera parameters. (left) camera1 (right) camera2

I calculated the residual error (2D error) between the projected 3D scene with known camera parameters and projected 3D scene with canonical camera parameters. The result is as shown in Fig. 18

```
************************************************************
The Residual error (2D error) for image 1 is
0.00046744
************************************************************
The Residual error (2D error) for image 2 is
0.00035261
************************************************************
```

Fig. 18 The residual error results

3. Conclusion

This lab provides deep insights into the epipolar geometry and stereo vision concepts. It also provide good leanings into camera parameters and how to choose them accordingly. It tells us the significant role of fundamental and essential matrices in epipolar geometry.

References

[1] Lecture slides of Dr. David Fofi

[2] Epipolar geometry, Wikipedia

[3] Multi View Geometry in Computer Vision by Richard Hartley and Andrew Zisserman

[4] Dr. Salvi Toolbox, Univeristy of Girona, Spain