ENGN1610 & 2605 Image Understanding
Lab07: Image Formation and Epipolar Constraint

The goal of this lab is to

- Learn how a point in world scene project to an image using camera intrinsic matrix.
- Learn the epipolar constraint arises from an essential matrix.

Problem 1. Image Formation Assume that the two cameras have the same resolution of 1200×1600 pixels (rows $\times$ columns) with the same calibration matrix capturing the same 3D scene, Figure 1. The 3D scene is represented by a dense cloud of 3D points storing $[X, Y, Z]$ coordinates together with their colors stored in (R, G, B), and the relative poses of the two cameras with respect to the world coordinate are represented by their rotation and translation matrices, *i.e.*, R_1, T_1, R_2, and T_2.

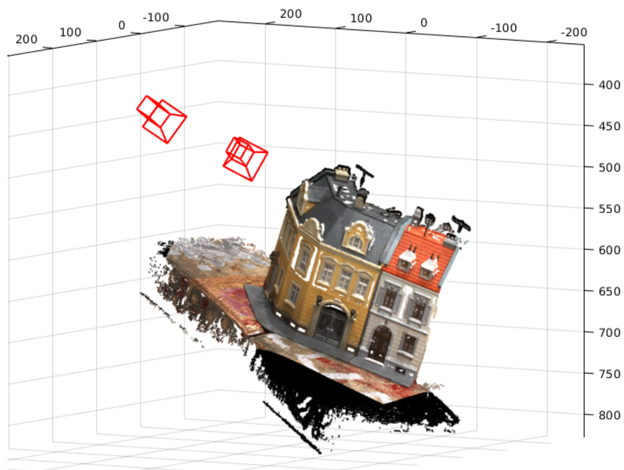


Figure 1: A 3D scene represented by a cloud of points is viewed by two cameras.

1. Show the expression that project a 3D point in world coordinates onto an image plane of a camera in pixels.
2. Use the provided scene points, camera intrinsic matrix, and relative pose encoded by `.mat` files, project the 3D points to the two image planes. Keep the following things in mind:
   (a) When projecting 3D points to a camera image plane, the 3D points have to be in the camera coordinate.
   (b) Projected points located outside the image boundaries should not be considered.
   (c) Direct projections on an empty image plane might give you an image shown in Figure 2(a). To construct a nicer and a rather compact image, Figure 2(b), you may need to make use of MATLAB's functions `meshgrid` and `griddata` to properly form the image.

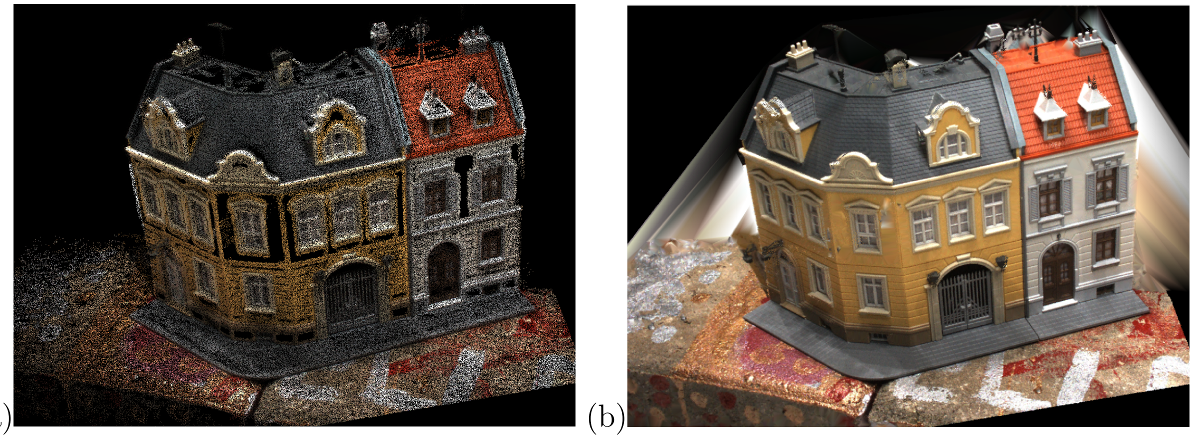
(a) (b)


Figure 2: Examples of projecting 3D points directly on an empty image plane (a) and an image with nicer and compact formation (b).

Problem 2. Essential Matrix Constraint Let γ from an image I and $\bar{\gamma}$ from another image $\bar{I}$ be a pair of feature correspondence in meters, both represented by homogeneous coordinates. In the lecture, we define the essential matrix as the relative pose between two cameras, and it satisfies $\bar{\gamma}^T E \gamma = 0$. The equation is often called the essential matrix constraint, or the *epipolar constraint*, meaning that given an essential matrix capturing the relative pose between two cameras, any pair of true correspondences in meters must follow that constraint.

Let $\gamma = [\xi \;\; \eta \;\; 1]^T$, $\bar{\gamma} = \left[\bar{\xi} \;\; \bar{\eta} \;\; 1\right]^T$, the epipolar constraint is,

$$\begin{bmatrix} \bar{\xi} & \bar{\eta} & 1 \end{bmatrix} \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix} \begin{bmatrix} \xi \\ \eta \\ 1 \end{bmatrix} = 0. \tag{1}$$

Expand the expression, we have, for each $(\bar{\xi}, \bar{\eta})$,

$$A\bar{\xi} + B\bar{\eta} + C = 0, \tag{2}$$

where the coefficients A, B, and C are,

$$\begin{aligned} A &= e_{11}\xi + e_{12}\eta + e_{13} \\ B &= e_{21}\xi + e_{22}\eta + e_{23} \\ C &= e_{31}\xi + e_{32}\eta + e_{33}. \end{aligned}$$

What kind of an equation is Equation 2? It is a line equation! This means that, given an essential matrix E, for each γ on image I, its correspondence $\bar{\gamma}$ must lie on a line on image $\bar{I}$. This line is often called an *epipolar line*. Figure 3 is an example, where the red squares from the left and the right images are a pair of SIFT feature correspondence. The red line on the right image is the epipolar line, passing through the SIFT feature on the right image. We will elaborate more on the epipolar constraint in the following lectures.

In this problem, you are asked to compute and draw an epipolar line as the one in Figure 3. The camera intrinsic matrix is also provided as a `.mat` file. The necessary steps to follow are:

1. **Feature Detection and Matching:** Use VLFeat to detect SIFT keypoints and extract SIFT descriptors from a pair of images in Figure 4.

2. **Find Reliable Feature Correspondences:** Use the function `Ransac4Homography.m` you implemented in the last lab to find the inliers of the feature correspondences. You may need to return the inlier indices if necessary.

3. **Compute Essential Matrices:** In the lecture, we say that to estimate an essential matrix, at least 5 pairs of correspondences are necessary. Randomly pick 5 pairs from the inliers given by the previous step and feed them into the provided function file `fivePointAlgorithmSelf.m`. Follow the instruction in that file to properly structure your input.

4. **Draw an Epipolar Line:** The return of the function file `fivePointAlgorithmSelf.m` is a cell array storing all possible, valid essential matrices. Pick one of the essential matrices, compute the epipolar line coefficients in Equation 2, and plot the line on the right image. An expected result should be the same as in Figure 3. Show at least three epipolar lines passing through different inlier features.

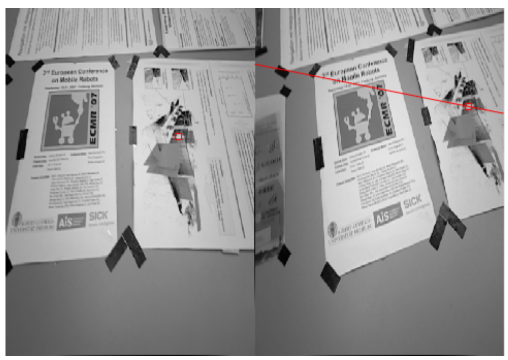


Figure 3: An example of an epipolar line.

(a)  (b) 

Figure 4

Problem 3. Answer the Following Questions (This part is mandatory for ENGN2605 students but optional for ENGN1610 students who will get 5 extra points if all the questions are answered correctly.)

Question 1: Show that $det(E) = 0$.

Question 2: In lab06, we implemented a RANSAC algorithm to compute a homography matrix that is supported by the maximal number of inlier features. Likewise, we can also use a RANSAC algorithm to compute an essential matrix. Recall that in estimating a homography matrix under a RANSAC scheme, a candidate correspondence is considered as an *inlier* if the Euclidean distance between the transformed coordinate and the matched coordinate is less than some small threshold, *e.g.*, 1 pixel. In estimating an essential matrix under a RANSAC scheme, what kind of measure do you think can be used to decide whether a pair of feature correspondences is an inlier or an outlier?

Extra Points (This part is optional for all students who will get 10 extra points if the question is answered correctly.)

Let R_1, T_1 and R_2, T_2 be the relative rotation matrices and translate vectors of camera 1 and camera 2 with respect to the world coordinate, respectively. In terms of R_1, R_2, T_1, and T_2, what is the relative pose of camera 2 with respect to camera 1?