University of Bucharest
Faculty of Mathematics and Computer Science
Department of Computer Science

Alexandra Diaconu
Bogdan Alexe
MSC in Computer Science

Computer Vision - Laboratory class 6

Camera Calibration and Fundamental Matrix Estimation with RANSAC

Objective

The goal of this lab is to introduce you to camera and scene geometry. We will use images (Figure 1) taken from different viewpoints of the same scene to understand these concepts.

Structure of the lab

This laboratory consists of three parts:

(1) estimating the projection matrix;

(2) estimating the fundamental matrix;

(3) estimating the fundamental matrix with unreliable SIFT matches using RANSAC.

We will estimate the camera projection matrix, which maps 3D world coordinates to image coordinates, as well as the fundamental matrix, which relates points in one scene to epipolar lines in another. The camera projection matrix and the fundamental matrix can each be estimated using point correspondences. To estimate the projection matrix (camera calibration), the input is corresponding 3D and 2D points. To estimate the fundamental

Figure 1: *Left and right view images of the same scene. It is easy to find point correspondences between the two images.*

matrix the input is corresponding 2D points across two images.

We start by estimating the projection matrix and the fundamental matrix for a scene with ground truth correspondences. Then we'll move on to estimating the fundamental matrix using point correspondences from SIFT. Apart from Lecture 11 from class you can also consult a tutorial on epipolar geometry here here.

Data

We provide 2D and 3D ground truth point correspondences for the base image pair (*pic_a.jpg* and *pic_b.jpg*, see Figure 1), as well as other images for which we do not provide any ground truth annotation.

Part I: Camera Projection Matrix

The goal is to compute the projection matrix P that goes from world 3D coordinates to 2D image coordinates. Recall that using homogeneous coordinates the equation for moving from 3D world coordinates (a point has coordinates (X,Y,Z)) to 2D camera coordinates (a point has coordinates (u,v)) is:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \cong \begin{pmatrix} u * s \\ v * s \\ s \end{pmatrix} = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = P \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

Another way of writing this equation is:

$$s = p_{31}X + p_{32}Y + p_{33}Z + p_{34}$$

$$p_{11}X + p_{12}Y + p_{13}Z + p_{14} - u * s = 0$$
$$p_{21}X + p_{22}Y + p_{23}Z + p_{24} - v * s = 0$$

We obtain that for each pair of corespondences between a 3D point in world coordinates (X_i, Y_i, Z_i) and a 2D point in camera coordinates (u_i, v_i) we have two equations:

$$p_{11}X_i + p_{12}Y_i + p_{13}Z_i + p_{14} - u_i p_{31}X_i - u_i p_{32}Y_i - u_i p_{33}Z_i - u_i p_{34} = 0$$

$$p_{21}X_i + p_{22}Y_i + p_{23}Z_i + p_{24} - v_i p_{31}X_i - v_i p_{32}Y_i - v_i p_{33}Z_i - v_i p_{34} = 0$$

We obtain a homogenous linear system of $2{\cdot}n$ equations (n is the number of points correspondences) that can be written as:

$$\begin{pmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -u_1X_1 & -u_1Y_1 \\ -u_1Z_1 & -u_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -v_1X_1 & -v_1Y_1 \\ -v_1Z_1 & -v_1 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots \\ X_n & Y_n & Z_n & 1 & 0 & 0 & 0 & 0 & -u_nX_n & -u_nY_n \\ -u_nZ_n & -u_n \\ 0 & 0 & 0 & 0 & X_n & Y_n & Z_n & 1 & -v_nX_n & -v_nY_n \\ -v_nZ_n & -v_n \end{pmatrix} \begin{pmatrix} p_{11} \\ p_{12} \\ p_{13} \\ p_{14} \\ p_{21} \\ p_{22} \\ p_{23} \\ p_{24} \\ p_{31} \\ p_{32} \\ p_{33} \\ p_{34} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

The homogenous linear system can be written more condensed as:

$$\mathbf{Ap} = \mathbf{0}$$

In the above equation, the projection matrix $\mathbf{P}$ of dimensions 3×4 is written as the vector $\mathbf{p}$ of dimension 12×1. Reshaping the vector $\mathbf{p}$ will give you the projection matrix $\mathbf{P}$.

At this point, we are almost able to set up the linear regression to find the elements of the matrix $\mathbf{P}$. There's only one problem, the matrix $\mathbf{P}$ is only defined up to a scale (you can always multiple $\mathbf{P}$ by a constant k and you will obtain the same result as all the components will be multiplied by k and then we divide by the third component s - see the above equation). So these equations have many different possible solutions, in particular $\mathbf{P}$ = all zeros is a solution which is not very helpful in our context. The way around this is to first fix a scale and then do the regression. There are several options for doing this:

(1) we can use the singular value decomposition to directly solve the constrained optimization problem:

$$\min \|Ap\|$$
$$s.t\ \|p\| = 1$$

In this case the solution is given by the eigenvector of the matrix A corresponding to the smallest eigenvalue.

(2) we can fix the last element (p_34) to 1 and then find the remaining coefficients

Once we have an accurate projection matrix $\mathbf{P}$, it is possible to tease it apart into the more familiar and more useful matrix $\mathbf{K}$ of intrinsic parameters and matrix $[\mathbf{R}|\mathbf{T}]$ of extrinsic parameters. We do know that:

$$\mathbf{P} = \mathbf{K}[\mathbf{R}|\mathbf{t}]$$

For this lab exercise we will estimate one particular extrinsic parameter: the camera center in world coordinates. Let us define **P** as being made up of a 3×3 matrix that we will call Q and a 4th column will call p_4 :

$$\mathbf{P} = [\mathbf{Q}|\mathbf{p_4}]$$

The center of the camera **C** can be found by:

$$\mathbf{C} = -\mathbf{Q^{-1}p_4}$$

Part II: Fundamental Matrix Estimation

In this part we estimate the mapping of points in one image to lines in another by means of the fundamental matrix. This requires using similar methods to those in part 1. We will make use of the corresponding point locations listed in *pts2d-pic_a.txt* and *pts2d-pic_b.txt*. Recall that the definition of the fundamental matrix $F = (f_{ij})$ is:

$$\begin{pmatrix} u' & v' & 1 \end{pmatrix} \begin{pmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{pmatrix} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = 0.$$

where (u,v) and (u′, v′) are pairs of correspondence points in two images of the same scene.

Note: the fundamental matrix is sometimes defined as the transpose of the above matrix with the left and right image points swapped. Both are valid fundamental matrices, but the visualization functions should be change in this case.

Another way of writing this matrix equations is:

$$\begin{pmatrix} u' & v' & 1 \end{pmatrix} \begin{pmatrix} f_{11}u + f_{12}v + f_{13} \\ f_{21}u + f_{22}v + f_{23} \\ f_{31}u + f_{32}v + f_{33} \end{pmatrix} = 0$$

Which is the same as:

$$f_{11}uu' + f_{12}vu' + f_{13}u' + f_{21}uv' + f_{22}vv' + f_{23}v' + f_{31}u + f_{32}v + f_{33} = 0$$

This equation resembles very much with the equations from part I. Given corresponding points you get one equation per point pair. With 8 or more points you can solve this (we cannot derive the scale parameter). As in part I there's an issue here where the matrix is only defined up to scale and the degenerate zero solution solves these equations. So we need to solve using the same method we used in part I of first fixing the scale and then solving the regression.

The least squares estimate of F is full rank; however, the fundamental matrix is a rank 2 matrix. As such we must reduce its rank. In order to do this we can decompose F using singular value decomposition into the matrices $U\Sigma V^T = F$. We can then estimate a rank 2 matrix by setting the smallest singular value in Σ to zero thus generating Σ_2 . The fundamental matrix is then easily calculated as $F = U\Sigma_2 V^T$. We can check the fundamental matrix estimation by plotting the epipolar lines using the plotting function.

Part III: Fundamental Matrix with RANSAC

For two images of a scene it's unlikely that you'd have perfect point correspondences for running the regression for the fundamental matrix. So, next we are going to compute the fundamental matrix with unreliable point correspondences computed with SIFT. Least squares regression is not appropriate in this scenario due to the presence of multiple outliers. In order to estimate the fundamental matrix from this noisy data we will need to use RANSAC in conjunction with the fundamental matrix estimation.

This is calculated from matching points from both the images. A minimum of 8 such points are required to find the fundamental matrix (while using 8-point algorithm). More points are preferred and use RANSAC to get a more robust result.

So first we need to find as many possible matches between two images to find the fundamental matrix. For this, we use SIFT descriptors with FLANN based matcher and ratio test.

We will first obtain SIFT correspondences (might not be reliable) for any image pair. We will use these initial point correspondences and RANSAC to find the "best" fundamental matrix. We will iteratively choose some number of point correspondences (8, 9, or some small number), solve for the fundamental matrix and then count the number of inliers. Inliers in this context will be point correspondences that "agree" with the estimated fundamental matrix.

Our code will return the fundamental matrix with the most number of inliers.