Finding Fundamental Matrix for Stereo Vision
Afshin Sepehri

In this problem, we use singular value decomposition and rank enforcement tools to find fundamental matrix for stereo vision based upon the geometry of stereo vision called *epipolar geometry*.

In many applications in computer vision area, we need to use two cameras to be able to get depth information of the scene, in a similar way the human brain builds three dimensional information of the scenes from two images captured by the two eyes.

Terminology

Before starting the project, you need to get familiar with some terminology and jargon of computer vision. If you already know about these, you may skip this section:

- *Pixel*: Pixel stands for Picture Element. It is the smallest unit of every image. It is a sample of the image intensity quantized to an integer value. Image is a two-dimensional array of pixels.
- *Gray-Level Image*: Gray-level image is a type of image with pixel values in range $[0, 255]$, where 0 means black, 255 means white and other values in between represent different levels of gray color. Larger values are used for brighter colors.
- *Pinhole Camera Model*: The simplest model for the generation of an image is the so-called pinhole camera. In this camera model, light is let into a box through a pinhole and projected on a light sensitive screen on the back end of the box at a distance away from the pinhole.
- *Image Plane*: The screen mentioned in the above definition, where light rays are projected, is called image plane.
- *Center of Projection*: The point where all the light rays join is named center of projection. If we consider pinhole as the center of projection in the pinhole model, the constructed image will be upside down, what we usually avoid. So, we prefer to change our model such that the center of projection goes behind the image plane. On the other hand, center of projection is the mirror point of the pinhole with respect to the image plane. Figure 1 depicts this model for two cameras with O_l and O_r as center of projections. In this project, we will use this model.
- *Homogeneous coordinates*: Homogeneous coordinates utilize a mathematical trick to embed n dimensional coordinates and transformations into a $n+1$ dimensional matrix format. As a result, inversions or combinations of linear transformations are simplified to inversion or multiplication of the corresponding matrices. As an example of this coordinate system, instead of representing each point (x, y) in two-dimensional space with a single two-dimensional vector

$$[x \quad y]^T$$

homogeneous coordinates allow each point (x, y) to be represented by any of an infinite number of three dimensional vectors

$$[tx \quad ty \quad t]^T$$

where t is an arbitrary nonzero real value. The two-dimensional vector corresponding to any three-dimensional vector can be computed by dividing the first two elements by the third, and a three-dimensional vector corresponding to any two-dimensional vector can be created by simply adding a third element and setting it equal to one.

Epipolar Geometry and Fundamental Matrix

Using the camera model defined above, the image of a point in 3D is the intersection point of the light ray connecting the point and the center of projection, and the image plane. In figure 1, the left and right cameras have image planes π_l and π_r and center of projections O_l and O_r respectively. So, the image of point P in 3D in the two cameras would be points p_l and p_r. Given a stereo

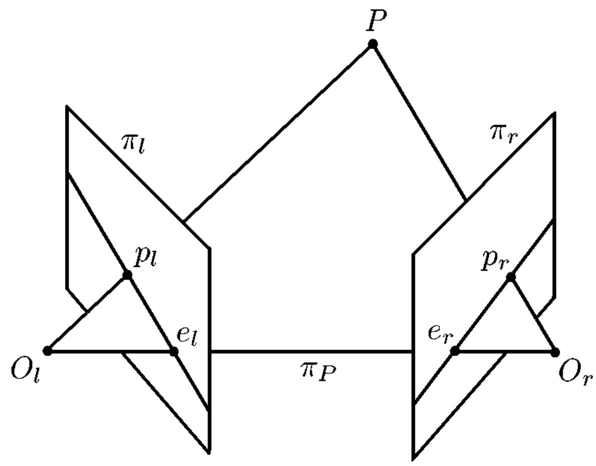

Figure 1: The Epipolar Geometry

pair of cameras, any point P in 3-D space, defines a plane π_P, going through P and the centers of projection of the two cameras. The plane π_P is called *epipolar plane*, and the lines where π_P intersects the image planes *conjugated epipolar lines*. The image in one camera of the center of projection of the other is called *epipole*. [1] In figure 1, epipoles are shown as e_l and e_r.

With the exception of the epipole, only one epipolar line goes through any image point. All the epipolar lines of one camera go through the camera's epipole.

According to *epipolar constraint*, corresponding points (e.g. p_l and p_r) must lie on conjugated epipolar lines. Meanwhile, the mapping between points and epipolar lines can be obtained from corresponding points only, with no prior information on the stereo system.

Without going through the details of the proof, the above theorem can lead us to the equation

$$p_r^T F p_l = 0 \tag{1}$$

where p_l and p_r, the image points in *homogeneous pixel coordinates* are defined as

$$p_l = [\,x_l \quad y_l \quad 1\,]^T$$
$$p_r = [\,x_r \quad y_r \quad 1\,]^T$$

and matrix F (3×3) is named *fundamental matrix.* It can be proven that this matrix has rank 2 and it is the same for all the corresponding points p_l and p_r in the two images.

In many stereo applications, we require to find the corresponding image points of a 3D point P in the two images to retrieve the coordinates of P in 3D. For this purpose, we should be able to find the corresponding point of each point p_l of the left camera, in the image plane π_r and visa versa. By knowing fundamental matrix F, the search space for such a point is shrunk into a line and this makes the matching process faster and more accurate. Therefore, finding matrix F is very essential. Note that matrix F depends only on the relative position and orientation of the two cameras and does not depend on the position of scene points in 3D. So, finding the fundamental matrix is a single time process and the result can be used for any scene taken by the two cameras in the future, as long as the relative position and orientation of the cameras is preserved.

In this project, we try to find the fundamental matrix as accurate as possible based upon an algorithm called *eight-point algorithm.*

1 Constructing linear equations of coefficients

We can convert (1) to the form

$$a^T f = 0 \tag{2}$$

where f, the vector of the elements of the matrix F, is defined as

$$f = [\,F_{11} \quad F_{12} \quad F_{13} \quad F_{21} \quad F_{22} \quad F_{23} \quad F_{31} \quad F_{32} \quad F_{33}\,]^T \tag{3}$$

and a is the vector of the coefficients a_i $(i = 1, ..., 9)$, which depend on x_l, y_l, x_r and y_r, the coordinates of the two corresponding points.

Problem 1. (10%) Find the coefficient vector a as a function of values x_l, y_l, x_r and y_r.

Since all the coefficients of (2) are unique only up to a scaling factor, we assume one of the elements of vector f (e.g. F_{33}) to be 1 and decrease the number of unknowns to 8.

To solve the equation numerically, We use two corresponding images taken by the two cameras. We select n corresponding points ($n \geq 8$) manually from the images and extract their x and y coordinates in pixel with respect to the top left corner of the images. This will give us two sets of points, P_l and P_r. Now we can make a set of equations as

$$Af = 0 \tag{4}$$

where each row of A ($n \times 9$) has the same structure as a in (2) and is found using one of the pair of corresponding points.

Problem 2. (10%) Find the matrix A using the points provided in two matrices `Pl` and `Pr` in file `points.mat`. Coordinates of corresponding points can be found in the same rows of these two matrices, one point in a row. We refer to these two matrices as P_l and P_r.

2 Solving the equations and finding fundamental matrix F

If all the measurements were accurate, having 8 points would be enough to find the 8 unknown elements of vector f, but since there are different sources of error in these measurements, we usually need more than 8 points and should find the best solution using a numerical method. We use *singular value decomposition (SVD)* here.

Problem 3. (15%) Suppose we have SVD of matrix A, that is

$$A = UDV^T \tag{5}$$

Prove that the best approximation to f in (4), is the column of V corresponding to the least singular value of A.

Problem 4. (15%) Find SVD of matrix A as (5) and find the best approximation to f based upon the theorem of problem 3. Do not forget to scale the result vector f such that the last element gets value 1.

Build fundamental matrix F from the elements of vector f according to (3) and find the norm of the error vector containing $p_r^T F p_l$ for all corresponding points p_l and p_r in matrices P_l and P_r.

3 Rank enforcement of matrix F

As mentioned earlier, matrix F is inherently a rank 2 matrix, but finding from SVD of matrix A, we did not consider this constraint. Now to apply this, we force the desired rank in the following way:

Problem 5. (10%) Find SVD of matrix F, that is

$$F = UDV^T$$

Set the smallest singular value in the diagonal of D to 0, let D' be the corrected matrix.

Find the corrected estimate of F, F', as

$$F' = UD'V^T$$

Find the norm of the error vector containing $p_r^T F' p_l$ for all corresponding points p_l and p_r in matrices P_l and P_r and compare the result with result of problem 4. Verify that matrix F' has rank 2.

4 Data Normalization

In order to avoid numerical instabilities, the eight-point algorithm should be implemented with care. The most important action to take is to *normalize* the coordinates of the corresponding points so that the entries of A are of comparable size. Typically, the first two coordinates (in pixels) of an image point are referred to the top left corner of the image, and can vary between a few pixels to a few hundreds; The differences can make A seriously ill-conditioned. [1] To make things worse, the third (homogeneous) coordinate of image points is usually set to one. A simple procedure to avoid numerical instability is to translate the first two coordinates of each point to the centroid of each data set, and scale the norm of each point so that the average norm over each component of the data set is 1. This can be accomplished by multiplying each left (right) point by a suitable 3×3 matrix, H_l or H_r. To compute each one of these two matrices, we use the following steps:

Given a set of points $p_i = [x_i, y_i, 1]^T$ with $i = 1, ..., n$, we define

$$\overline{x} = \sum_i x_i / n$$

$$\overline{y} = \sum_i y_i / n$$

$$\overline{d} = \frac{\sum_i \sqrt{(x_i - \overline{x})^2 + (y_i - \overline{y})^2}}{n\sqrt{2}}$$

Now, matrix H (3×3) can be found such that

$$Hp_i = \widehat{p_i}$$

where

$$\widehat{p_i} = [(x_i - \overline{x})/\overline{d} \qquad (y_i - \overline{y})/\overline{d} \qquad 1]^T$$

Problem 6. (10%) Find matrices H_l and H_r using the points given in P_l and P_r and as described above. Verify that the average norm of each component of $\widehat{p_i}$ is almost 1.

By using points $\widehat{p_i}$ instead of p_i and applying eight-point algorithm, we can find a new matrix $\widehat{F}$. The desired fundamental matrix F for the original points can be found from matrix $\widehat{F}$ and the two normalization matrices H_l and H_r.

Problem 7. (10%) Prove that fundamental matrix F can be found as

$$F = H_r^T \widehat{F} H_l \tag{6}$$

Problem 8. (10%) Repeat the steps of problems 2, 4 and 5 using the normalized points $\widehat{P_l}$ and $\widehat{P_r}$. Find new fundamental matrix F using (6) and compare it with the one found in problem 5.

5 Finding epipolar lines

Finding the fundamental matrix F, we are able to find the corresponding epipolar line for each point p_l as

$$l_r = F p_l \tag{7}$$

where $l_r = [a, b, c]^T$ and the equation of the line in 2D is $ax + by + c = 0$.

According to the epipolar geometry, all the epipolar lines meet at the point epipole, which is the image of one center of projection in the other image plane. Note that based upon the relative orientation of the two image planes, this point may or may not be within the boundary of the image plane.

Problem 9. (10%) Find L_r, the matrix of epipolar lines where each row is the corresponding epipolar line for a point in P_l. Test if all these lines intersect in a single point and explain how you tested it.

6 Verify Fundamental Matrix on Real Images

In this section, you can verify the result of your program on a pair of real images. In fact, the provided points in file `points.mat` were selected manually from a pair of images (`left.bmp` and `right.bmp`)taken by two cameras of a stereo vision set. Some Matlab functions are provided for you to test the fundamental matrix you have found in earlier sections. Here is the name, parameters and usage of each one of these functions:

- `[h, w] = ViewImages(leftImage, rightImage)`: shows two gray-level images simultaneously on the screen. The input parameters are the name of the two gray-level image files. It returns the height and width of the left image, which are logically equivalent to the right image's ones.

- `[Pl, Pr] = SelectCorrespondingPoints()`: lets you select the corresponding points by marking them on the images via clicking the mouse buttons. Left and right buttons can select points on left and right images respectively. The output matrices have the same format as matrices P_l and P_r discussed earlier.

- `MarkPointsOnImages(Pl, Pr, r)`: marks the points given in the parameters `Pl` and `Pr` on the two images as cross ($\times$) with radius `r`.

- `MarkPointsAndLinesOnImages(Pl, Lr, h, w, r)`: marks a set of points `Pl` on the left image and a set of lines `Lr` on the right image. The height and width of the images should be given in the two scalar inputs `h` and `w`. `r` is the radius of the cross used to mark points. This function can be used to test visually how well the epipolar lines computed using the described method pass the corresponding points.

The above functions should be called in an appropriate sequence to fulfill the purpose. For instance, you can call the functions `ViewImages` and `MarkPointsOnImages` and send the point sets P_l and P_r to view the given corresponding points on the two images, or you can call `ViewImages` and `MarkPointsAndLinesOnImages` to see the points P_l and lines L_r found in previous sections. Also, by calling `ViewImages` and `SelectCorrespondingPoints`, you can select new points on one image and test the fundamental matrix F you have already found on them. You can check file `AllTogether.m` as a sample combination of the provided functions.

7 Complementary Comment

According to epipolar geometry, the fundamental matrix is a unidirectional quantity, which relates points in one image to lines in another image. In our problem, we found the fundamental matrix F_{lr}, which relates points of left image to the lines of right image. However, the other fundamental matrix, F_{rl}, which relates points of right image and lines of left image can be computed easily.

(Extra) Problem 10. You do not need to do this problem, and it will not give you extra credit. Prove that fundamental matrix F_{rl} can be found as

$$F_{rl} = F_{lr}^T$$

Tools

To study more about stereo vision, you can refer to [1] and [2]. Also, to see complementary discussions on eight-point algorithm used in this project, see [3].

In this problem, we use singular value decomposition (SVD). To study more about SVD, see [4].

1. E. Trucco, A. Verri, *"Introductory Techniques for 3-D Computer Vision"*, Prentice Hall 1998, Chapter 7.

2. R. Hartley and A. Zisserman, *"Multiple View Geometry in Computer Vision"*, Cambridge University Press, 2000.

3. R. Hartley, *"In defence of the 8-point Algorithm"*, Proc. 5th International Conference on Computer Vision, Cambridge (MA), pp. 1064-1070(1995).

4. Gene H. Golub and Charles F. Van Loan, *"Matrix Computations"* (2nd ed.), Johns Hopkins Press, 1989.