86735: Computer Vision
SIFT

Due date: Thu 23:59, 12.02.16

Prof. F. Odone, Prof. F. Solari, M. Chessa, N. Noceti

Rabbia Asghar, BEng, Ernest Skrzypczyk, BSc

Estimating the fundamental matrix

Stereopsis

Stereovision or Stereopsis refers to the aspect of extracting 3D information like perception of depth and 3D structure from two or more images taken from different viewpoints. This is similar to the functioning of two eyes in a human vision system.

The difference in retinal position of the eyes helps build the 3D perception of the world in human brain. Human eyes are positioned side-by-side and each eye takes view of the same area from slightly different angle. The two images arrive simultaneously in human brain and are then combined. The small differences in the two images result in building 3D stereo picture. The same principle is applied to camera images taken from two different viewpoints.

Epipolar geometry

Epipolar geometry is a very crucial part of stereo vision. It refers to the geometric relations between 3D points in the world frame and their projections onto 2D images in image planes of the camera. This acts as a constraint between the image points. Epipolar geometry is further explained with aid of figure (1).

The figure (1) shows two image planes of the acquisition system in blue for two respective cameras, left and right. The virtual image planes are placed in front of the focal plane of each camera. Points O_L and O_R represent the origins of symmetry of the two cameras lenses. X represents the point of interest in the 3D world, while x_R and x_L are the points of projection on right and left image plane respectively.

The terms specific to epipolar geometry are first defined below.

1. Epipole: Projection of optical centers of the cameras lenses O, into the other camera image plane:
   (a) Left epipole: Projection of O_R on the left image plane e_L.
   (b) Right epipole: Projection of O_L on the right image plane e_R.
2. Baseline: Line connecting O_L and O_R. The baseline intersects each image plane at the epipoles e_L and e_R.
3. Epipolar plane: Plane containing 3 points in space X, O_L and O_R.
4. Epipolar line: Intersection of the epipolar plane with the image plane.

The line $O_L - X$ is seen by left camera as a point, because it is directly in line with that cameras centre of projection. This means all the points on this line e.g. X, X_1, X_2, X_3 will be projected on x_L. However, the right camera sees this line as an actual line in its image plane. The projection of this line is in fact an epipolar line. In the same manner, line $O_R - X$ is projected on the epipolar line $x_L - e_L$ on the left image plane.

The epipolar plane, shown in green in the figure (1), contains the baseline $O_L - O_R$ and intersects the image planes at point e_R and e_L. For each point in an image plane, its corresponding point in the other image can be found by looking only along its epipolar line. This is called an **epipolar constraint**. The epipolar lines are an important constraint for relation between corresponding points of two stereo images.

Fundamental matrix F

On the basis of the above geometry, points in one image plane can be mapped on to the epipolar lines of the others. The basic relation of fundamental matrix F with the corresponding points is given by equation (1).

$$x_R^T F x_L = 0 \tag{1}$$

Following are a few properties of fundamental matrix which play important role in its estimation and applications:

1. The matrix encodes information on both the intrinsic and extrinsic parameters.
2. It is a 3x3 homogeneous matrix of rank 2, with 7 degrees of freedom.
3. If x_L and x_R are corresponding points, $x_R^T F x_L = 0$
4. For the same condition, if x_R and x_L are corresponding points, $x_L^T F^T x_R = 0$
5. Epipolar line $x_R - e_R$ corresponding to x_L is equal to $F x_L$.
6. Similarly, epipolar line $x_L - e_L$ corresponding to x_R is equal to $F^T x_R$.

8-point algorithm

Theory

The fundamental matrix can be simply estimated using point correspondences from two images and without any information on the intrinsic or extrinsic camera parameters.

Points p_l and p_r are the homogeneous representations of corresponding image coordinates x_L and x_R. As discussed before each correspondence leads to homogeneous equation of the form presented by equation (2):

$$p_r^T F p_l = 0 \tag{2}$$

where

$$p_r = \begin{bmatrix} x_r \\ y_r \\ 1 \end{bmatrix}, \quad p_l = \begin{bmatrix} x_l \\ y_l \\ 1 \end{bmatrix} \tag{3}$$

If fundamental matrix F is written as shown in (4)

$$F = \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}, \tag{4}$$

equation (2) can be rewritten to the from portrayed by equation (5)

$$x_l x_r f_{11} + x_l y_r f_{21} + x_l f_{31} + y_l x_r f_{12} + y_l y_r f_{22} + y_l f_{32} + x_r f_{13} + y_r f_{23} + f_{33} = 0 \tag{5}$$

The entries of the fundamental matrix F, can be determined by establishing eight or more correspondences. The equation (5) is rearranged to form a homogeneous system, as shown in (6).

$$Af = 0 \tag{6}$$

$$A = \begin{bmatrix} x_{l1}x_{r1} & x_{l1}y_{r1} & x_{l1} & y_{l1}x_{r1} & y_{l1}y_{r1} & y_{l1} & x_{r1} & y_{r1} & 1 \\ x_{l2}x_{r2} & x_{l2}y_{r2} & x_{l2} & y_{l2}x_{r2} & y_{l2}y_{r2} & y_{l2} & x_{r2} & y_{r2} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ x_{ln}x_{rn} & x_{ln}y_{rn} & x_{ln} & y_{ln}x_{rn} & y_{ln}y_{rn} & y_{ln} & x_{rn} & y_{rn} & 1 \end{bmatrix} \tag{7}$$

$$f = \begin{bmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{bmatrix} \tag{8}$$

Matrix A is nx9 matrix where n is the number of correspondences. However, the rank of matrix A is 8, which makes it rank deficient. This gives a unique solution up to a scale factor and the solution is proportional to the last column of V, where $A = UDV^T$.

The 8-point algorithm can be summarised in following steps.

1. Take n correspondences from 2 stereo pair images where n is greater than 8.
2. Construct homogeneous system $Ax = 0$, as described above, where A is an nx9 matrix.
3. Matrix F can be computed by singular value decomposition (SVD) of matrix A. The entries of F are proportional to the components of the last column of V.
4. Enforce the [singularity] constraint: $rank(F) = 2$ by the following steps:
   (a) Compute the SVD of F.
   (b) Set the smallest singular value equal to 0 and let D be the corrected matrix.
   (c) The corrected estimate of F is given by $F_e = UDV^T$.

Point normalization

In order to implement the 8-point algorithm and estimate the fundamental matrix F, corresponding points in the image are expressed in a vector form of 3 elements, as shown in equations (3) and (4). The 3^{rd} element of the vector is assigned the value of 1 and this is done to prepare a homogeneous vector. For most of the corresponding points, the first two elements are much larger than the 3^{rd} element. However, this results in vector pointing in more or less the same direction for all points. Similarly, the F estimated with this approach is not invariant to point transformations.

In order to make the algorithm numerically stable and the estimation more precise, it is more suitable to use normalized points. This is done by transforming the coordinates of each of the two images independently such that the *average* point is equal to $(1, 1, 1)^T$, which is achieved in 2 steps:

1. The origin of the new coordinate system is centered (has its origin) at the centroid (center of gravity) of the image points. This is accomplished by a translation of the original origin to the new one.

2. After the translation, the coordinates are uniformly scaled, so that their average distance from the origin equals $\sqrt{2}$.

This principle results in a distinct coordinate transformation for each of the two images. As a result, new homogeneous image coordinates p_l, p_r, are given by

$$\bar{p}_l = T_l p_l \tag{9}$$

$$\bar{p}_r = T_r p_r \tag{10}$$

where T_l, T_r are the normalized transformations, responsible for translation and scaling from the old to the new normalized image coordinates. This normalization is only dependent on the image points, which are used in a single image. Normalized transformation is given by equation (11)

$$T = \begin{bmatrix} s & 0 & -sc_x \\ 0 & s & -sc_y \\ 0 & 0 & 1 \end{bmatrix} \tag{11}$$

where c is the centroid of all points and s is the scale to make average distance equal to $\sqrt{2}$.

This normalization should be carried out before applying the 8-point algorithm. After the fundamental matrix $\bar{F}$ has been estimated, de-normalization is done to get the results in original coordinates system. $\bar{F}$ can be de-normalized to give F according to (12).

$$F = T_r^T \bar{F} T_l \tag{12}$$

In general, this estimate of the fundamental matrix is a more precise one, than one would have obtained by estimating from not normalized coordinates, even at the price of a few calculations more.

Implementation

The above script (**??**) residing in file `EightPointAlgorithm.m` represents the implementation of the 8-point algorithm described in depth in previous subsection *8-point algorithm - Theory.* It is also sufficiently commented in a self explanatory way and written according to requirements of the laboratory session and beyond in the script itself. Additional parameter `normalize` was added to the function `EightPointAlgorithm`, because of the normalization loop in the main script (**??**) described in next section.

Main script

The script `L06.m` runs a main loop beginning at line 45 for all image sets and a secondary loop for normalization of points beginning at line 70. Each main loop consists of loading the point sets for both images (line $46 \div 61$), which are then converted to homogenous form by adding a third coordinate equal to 1 (lines $63 \div 67$). Next the normalize loop begins with line 69, where the 8-point algorithm is called to estimate the fundamental matrix F:

The additional argument holds the current state of variable `normalize`, which loops between 0 and 1, so that the loaded image sets can be compared using normalized and not normalized point sets. To confirm that the estimation of the fundamental matrix F was correct, the point sets for both images undergo an epipolar constraint check (lines $72 \div 92$) described by equation (14). Since calculations are made using numerical methods and there is no analytical analysis, the ideal case presented by equation (13) will never be fulfilled, which is the reason for introducing a very small number with variable ϵ. The variable `EPS` was set at 10^{-2}, which is high value for a tolerance. Below also that part of code is displayed. It should be noted that for practical reasons the arrays P_1 and P_2 stayed in column form, so the operation on them might be misleading in the code.

$$x_R^T F x_L = 0 \tag{13}$$

$$P_2^T F P_1 < \epsilon, \quad \epsilon = 10^{-2} \tag{14}$$

The results are printed on the screen with a summarizing success rate. Next step beginning with line 95 loads the stereo images, and visualizes them (lines $106 \div 110$) using the external function `VisualizeEpipolarLines` from file `VisualizeEpipolarLines.m`, which is a modified version of the provided original. The arguments of the `VisualizeEpipolarLines` function have been also expanded, so that the variable for normalizing the point set `normalize` and the figure reference variable `k` are passed on. The function itself has some additional features, but basically allows for an easier comparison of the results. Main addition is plotting of the centroid of the point set in each image and a surrounding box, so that the process of normalizing points can be presented more easily. This only applies if the point set is actually being normalized as shown below in the excerpt from file `VisualizeEpipolarLines.m`.

The difference between the files is included in the `VisualizeEpipolarLines.diff` file, which allows patching using the command:

```
$  patch < VisualizeEpipolarLines.diff
```

It is worth mentioning that certain operating systems do not respect small and capital letters in filenames. A very nice feature of the provided `VisualizeEpipolarLines` script enables the user the possibility to select custom points of interest, which will be used to estimate the fundamental matrix F. To activate this mode empty sets of points P_1 and P_2 need to be passed to the function.

Next the epipoles e_L and e_R are being calculated by using the property of being, respectively, the left and right null space of fundamental matrix F, and therefore can be obtained as last columns of vectors U and V by performing a singular value decomposition on the matrix F as per equation (15).

$$F = UWV^T \tag{15}$$

Additionally an epipolar check has been conducted, which should also result in 0 in the ideal case, as displayed by equations ($16 \div 19$). The algorithm of performing the test and printing the results (lines $125 \div 149$) is analogue to that of epipolar constraint check, with the difference that each point set is used with its corresponding epipole. Worth mentioning is the actual value of the tolerance used for this check. For the given image sets, ϵ can be as low as $\epsilon = 10^{-12}$.

$$e_L^T F P_1 < \epsilon \tag{16}$$
$$e_R^T F^T P_2 < \epsilon \tag{17}$$

$$e_L^T F x_L = 0 \tag{18}$$
$$e_R^T F^T x_R = 0 \tag{19}$$

At the end of the script the normalize and main loops are closed and the log file writing disabled, followed by a formatting script for the results, that is executed only on *nix systems.

Results

Epipolar constraint

8-point algorithm is implemented to a stereo set of images, which have characteristic points selected in them. Those point sets from both, left and right, images, are then homogenized and preferably also normalized. The theory behind normalization of the point sets is extensively described in section *Point normalization*. After the estimation of the fundamental matrix F all points from both images, should fulfil the epipolar constraint presented by equation (13).

The results, which excerpts for the first image set are presented above and the second image set below, show clearly that the success rate of satisfying the epipolar constraint is significantly higher, even perfect resulting in a 100% match, when point normalization is being utilized. This confirms previous assumption, that the estimated fundamental matrix F with point normalization is more precise at a cost of few additional calculations.

Epipoles

Similar to the epipolar constraint also a specific set of equations ($16 \div 19$), involving the epipoles, was tested. The following results, are also for both point sets, respectively normalized and not normalized.

There is a significant difference for the first image set in the epipoles coordinates. The default `eps` in `MATLAB` is set to 10^{-12}, which is the reason this tolerance was choosen. However even lower values of one order would still result in a near perfect success rate. A slight difference, approximately 10^{-3}, in the epipolar coordinates does results in a significant difference of the test result, which was tested manually. Below are the test results for the second image set. Also here a perfect match has been achieved, reassuring the correctness of the estimate of fundamental matrix F. The difference between the epipoles in case of not normalized and normalized point sets is very low compared to the first image set, however it still is a difference of at least a factor 10 up to 10^4. The analysis of the position of the epipoles is somewhat difficult at this point, but it should be possible to reconstruct figure (1) from section *Epipolar geometry* and show the the relations between elements of epipolar geometry on this particular acquisition equipment setup.

Running the script `L06.m` in `MATLAB` environment produces initially a slightly different log file `L06.sc`. It can be processed with the `BASH` script `cleanup.sh` for proper formatting, which is exactly what the main script does, if the host is a *nix derivative.

Visualization of stereo pairs

To visually determine if the estimation of the fundamental matrix F has been done correctly, points of interest on one of the stereo images can be selected and then, depending on the direction of conversion, the corresponding points for the other image calculated using the appropriate formula.

Figures $(2 \div 5)$ represent both image sets for normalized and not normalized point sets. The figures (2) and (3) show a significant difference in both distribution of epipolar lines and the epipoles themselves, which has been shown in previous subsections of the *Results* section. Especially figure (2) clearly displays the importance of using normalization of point sets, since it is easily visible how much difference there is between the points and epipolar lines. The two points in the right down corner of the right image, are simply omitted by the lines, other points are almost not crossed. As expected the epipolar lines seem to be parallel, since the physical setup of the acquisition hardware is very compact, ergo both cameras for the left and right image are situated very near to each other.

One of the features of the modified `VisualizeEpipolarLines` function, is to show the centroid of the point set and a surrounding box. After normalizing the point sets, the epipoles are calculated to be at different location and the epipolar lines run without exception through the selected points. All points are crossed within the range determined by the shape of the rectangle symbol representing the points. Figure (3) also represent the centroid with a bounding box around all points of interest. The points themselves have been selected at very characteristic coordinates of the image, making it easier for humans to determine if the estimation was successful and precise enough. The centroids are near each other, however they do not lie on the same epipolar line. The bounding boxes are very similar in location and size, since the cameras are very near to each other. Another option of determining accuracy of the estimated fundamental matrix, could be achieved by compositing one of the images on top of the other, so that either the centroids would be the common point or the center masses of the bounding boxes.

The set of figures representing the left and right images of set 2, figure (4) and figure (5) additionaly use a 3D object, namely a Rubik's cube, in the scenery as a reference. Translation and rotation of points of interest takes place between the left and right image, which is especially visible, while looking at the selected points on the Rubik's cube. Similarly to the first image set, the image with not normalized point set has epipolar lines, that are less accurate. The point in the right top corner of the image, so representing the corner of the calibration board, is slightly missed by it's epipolar line in the right image. Other epipolar lines cross or at least touch the points.

The normalized point set has a visible difference in location of the centroid, where it almost covers the nearest to the image plane corner of the Rubik's cube in the left image of figure (5). It should be also easily recognizable, that the bounding box of the right point set is larger than that of the left image point set, both vertically, which should be more visible because of the images arrangement, and horizontally. This the strengthens previous observation of a significant transformation of the points referring to 3D scenery. The other visible difference between normalized and not normalized point sets is the change from almost perfectly parallel epipolar lines in the not normalized point set to clearly not parallel epipolar lines in the normalized one. This is the result of a difference in the epipole coordinates, which for the right epipole has only one significantly different coordinate, namely the second, with a value of $45 \cdot 10^{-3}$, and for the left epipole a small difference of $6 \cdot 10^{-3}$ also for the second coordinate. This is sufficient to create such an effect.

Conclusions

When image acquisition is done by a stereo system, the problem of finding the corresponding points in both images arises. In the presented results, two point sets of characteristic points in the 3D scenery were already provided. If there are several simplifying assumptions present, like that the distance between the cameras is small, so the acquired images are very similar, algorithms for template matching could be used to find the corresponding points in the other image. Normalized cross correlation would one candidate for this operation. However depending on the real world scenery, the acquired results might be inaccurate, even more so if the points were not preselected, but for example randomly generated or acquired by using a corner detection algorithm. To further increase the probability of finding the correct corresponding point in the other image, epipolar geometry, especially epipolar lines, could be used. As explained in the section *Epipolar geometry* the epipolar constraint allows for finding the corresponding point on the specific epipolar line, which greatly simplifies the search, therefore saves computing time, and increases the probability of finding the correct corresponding point. This however depends, as shown in the *Results - Epipolar constraint* section, on the fact if the point sets have been normalized or not. The point normalization significantly improves the estimation of fundamental matrix F, which in turn allows for more precise determination of corresponding epipolar lines.

The main benefit of estimating the fundamental matrix F is that very little information is needed. Here two sets of images have been presented, however script (**??**) allows for further expansion of image sets and pairs of images. Without intrinsic or extrinsic parameters, the estimation with normalized point set delivers very accurate results as shown in the section *Visualization of stereo pairs* and since one of the properties of fundamental matrix F is, that it includes both intrinsic and extrinsic parameters of the acquisition set, it is a very powerful tool for computer vision in general.

The process of point normalization is not very demanding on the computation resources, since there are only a few operations necessary. This makes the combination of estimating the fundamental matrix F using normalized point sets with little information in form of few image pairs, a very interesting approach for specific operations, like finding corresponding points and further image processing and analysis, for example creating maps of the environment in combination with odometry of a mobile robot or an unmanned air vehicle.