ETH Zürich
Computer Science Department

Prof. Marc Pollefeys
Prof. Vittorio Ferrari
Prof. Luc Van Gool
HS 10

Computer Vision

Exercise 3

Hand-out: 14-10-2010
Hand-in: 21-10-2010 13:00

Objective:

The goal of this assignment is to estimate the epipolar geometry between two related views. First the fundamental matrix for two uncalibrated cameras is estimated, by implementing the eight-point algorithm. For the calibrated case you will estimate the essential matrix and extracted the relative camera poses using a sparse 3D reconstruction.

3.1 Image capturing (10%)

Capture some images from a static 3D scene from different viewpoints. The images should have some overlap such that point correspondences can be established. Un distort your images to make straight lines in the scene appear straight in the image. This can be achieved using the Bouget's toolbox used in the previous exercise.

3.2 Fundamental matrix, eight-point algorithm (30%)

The fundamental matrix is defined as follows

$$x'^T F x = 0, \tag{1}$$

for any pair of point correspondences $x'_i \leftrightarrow x_i$ in two images. One property of the fundamental matrix is that it is of rank 2 with 7 degrees of freedom, hence it can be recovered from 7 point correspondences. However, it is much simpler to compute it from 8 point correspondences, by solving the following homogeneous system of equations derived from eq. 1

$$Af = \begin{bmatrix} x'_1 x_1 & x'_1 y_1 & x'_1 & y'_1 x_1 & y'_1 y_1 & y'_1 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x'_n x_n & x'_n y_n & x'_n & y'_n x_n & y'_n y_n & y'_n & x_n & y_n & 1 \end{bmatrix} f = 0$$

- First, click at least eight point correspondences in the two captured images.
- As you did in the previous assignment, normalize your points such that their root-mean-square distance from the origin is $\sqrt{2}$. Use your own implementation or the one provided to compute the transformation matrix T_1 and T_2 such that $\hat{x}' = T_1 x'$ and $\hat{x} = T_2 x$.

- Each point correspondence gives one equation, stack all equations into A and find the right null vector of A, using the singular value decomposition $A = USV'$. The solution vector is the last column of V.

- Constraint enforcement: One important property of the fundamental matrix is that it is a singular matrix of rank 2. A convenient way to enforce the singularity constraint is by factorizing $F = USV^T$ using SVD, where S is a diagonal matrix S = diag(r, s,t) satisfying $r \geq s \geq t$. Then replace F by $\hat{F} = Udiag(r, s, 0)V^T$, which is the closest singular matrix to F under the Frobenius norm.

- Denormalize $\hat{F}$ to obtain the final fundamental matrix i.e., $\hat{F} = T_1'\hat{F}T_2$.

Draw the epipolar lines for both, the non-singular fundamental matrix F and the fundamental matrix $\hat{F}$ with enforced singularity constraint. Also show the epipole of F and $\hat{F}$ in both images, which are the right and left null-vectors of the fundamental matrix.

3.3 Essential matrix (30%)

The essential matrix is the special case of the fundamental matrix, to the case of the normalized image coordinates, where $\hat{x} = K^{-1}x$. Similar to the fundamental matrix it can be computed from 8 point correspondences. The method differs in the enforcement of the constraint, where the fundamental matrix satisfies $detF = 0$, the essential matrix satisfies the additional constrain that the two first singular values are equal.

a) Implement the eight-point algorithm to compute the essential matrix. First, transform the point correspondences into normalized image coordinates $\hat{x} = K^{-1}x$, using your camera calibration matrix K. Then, proceed as for the fundamental matrix. Use the constraint that the first two singular values must be equal and the third one is zero. Factorize your essential matrix $E = USV'$ using SVD, where S is a diagonal matrix S = diag(r, s, t) satisfying $r \geq s \geq t$. Replace E with $\hat{E} = Udiag((r+s)/2, (r+s)/2, 0)V^T$, which is the closest essential matrix to E under the Frobenius norm.

b) The essential matrix can be expressed as a rotation matrix R and a translation vector t,

$$E = [t]_\times R \tag{2}$$

- First, compute the translation vector $||t||$, which is the projection of the other camera center into the current image plane, which is the right null vector of E ($Et = 0$). You can normalize t to unit length, since the scale can not be determined.

- To compute the rotation matrix R, factorize E such that $E = USV^T$ using the SVD. There are two different solutions for R

$$R_1 = UWV^T,\ R_2 = UW^TV^T \ \text{ with } \ W = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

Once the rotation matirx is extracted, make sure that the determinant of R is positive, if $det(R) = -1$ your rotation matrix is represented in a left hand coordinate system and needs to be multiplied by -1 to be transformed into a right hand coordinate system. With this rotation and translation, the final four possible projection matrices for the second camera are

$$P_1 = R_1[I_{3\times3}|t],\ P_2 = R_1[I_{3\times3}|-t],\ P_3 = R_2[I_{3\times3}|t],\ P_4 = R_2[I_{3\times3}|-t]. \tag{4}$$

The ambiguity of the solution is illustrated in fig. 1. Assume the first camera is at the origin of your camera coordinate system ($P = [I|0]$) and the second camera P' is rotated and translated according to P_i. Then the correct solution for P' is the transformation matrix P_i which triangulates a point correspondence in front of both cameras. Use the triangulation function provided in the framework. You can verify that a point lays in front of a camera, by transforming the 3D point into the camera coordinate system and looking at the Z component of the point.

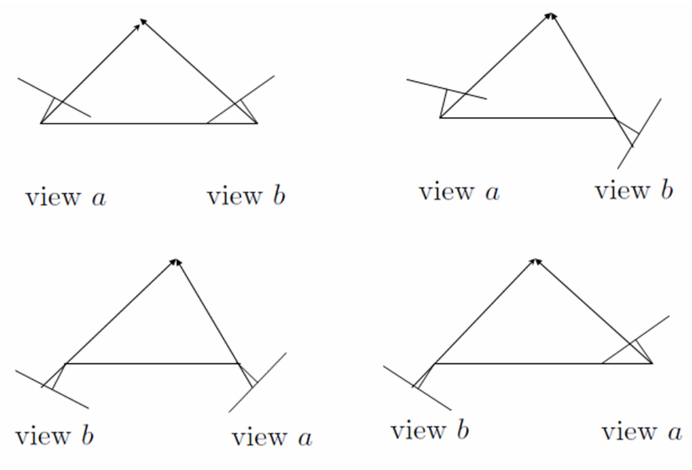

Figure 1: Rotation and translation ambiguity.

3.4 Non-linear optimization (30%)

To improve the accuracy of the essential matirx apply a non-linear least square optimization over P' and the triangulated 3D points.

- Choose camera matrices $P = [I|\mathbf{0}]$ and $P' = [R|t]$, where R and t are obtained from E as described in task 3.3.
- Triangulate all point correspondences $x'_i \leftrightarrow x_i$ using P and P'. A triangulation function is provided in the framework.
- Minimize the cost function

$$\sum_i d(x_i, PX_i)^2 + d(x'_i, P'X_i)^2 \tag{5}$$

where $d(.,.)$ is the Euclidean distance between the true coordinate x_i and the estimated coodinates PX. Use fminsearch to minimize over $3n + 12$ variables: $3n$ for the 3D points X_i, and 12 for the camera matrix $P' = [R|t]$.

Report the decrease of your reprojection error after the optimization. Draw the epipolar lines of the optimized essential matrix $E = [t]_\times R$.

3.5 Bonus

a) 7-point algorithm (20%)
The fundamental matrix has seven degrees of freedom and therefore seven point correspondences suffice to solve the non-linear system of equations. Implement the seven point algorithm, more information about the algorithm is provided in the course notes [2], section 4.2.2. To find the roots of a polynomial you can use the Matlab function *roots*.

3.6 Hand in:

Write up a short report explaining the main steps of your implementation and discussing the results of the methods. The report should include

- The intrinsics of your camera, used to solve the assignment, hard code it in the main.m file.
- The two undistorted images with the clicked point correspondences, also store the points as a .mat file and include it in your zip file.
- The two images showing the epipolar lines and the epipole for both fundamental matrices F and $\hat{F}$.
- The cameras, together with the triangulated points.

Send the report together with your source code to either saurero@inf.ethz.ch or tpetri@inf.ethz.ch.

3.7 References:

[1] `http://www.vision.caltech.edu/bouguetj/calib_doc/index.html`

[2] `http://www.cvg.ethz.ch/teaching/2010fall/compvis/tutorial.pdf`