

CSE 252B: Computer Vision II, Winter 2017 – Assignment 5

Instructor: Ben Ochoa

Due: Tuesday, March 21, 2017, 11:59 PM

Instructions

- Review the academic integrity and collaboration policies on the course website.
- This assignment must be completed individually.
- This assignment contains both math and programming problems.
- Programming aspects of this assignment must be completed using MATLAB.
- You may not use MATLAB functions contained in the computer vision toolbox. Use the `MATLAB which` command to determine which toolbox a function is contained in. If you are unsure about using a specific function, then ask the instructor for clarification.
- You must prepare a report describing the problems, and your solutions and results. The report must contain enough information for a reader to understand the problems and replicate your work without also having the assignment (i.e., this document).
- Your report will be a pdf file named `CSE_252B_hw5_lastname_studentid.pdf`, where `lastname` is your last name and `studentid` is your student ID number. The report must be prepared using L^AT_EX.
- All of your MATLAB source code must be included in an appendix of your report. You may find the `listings` package useful for this.
- You must submit your report on Gradescope.
- Additionally, you must create a zip file named `CSE_252B_hw5_lastname_studentid.zip`, where `lastname` is your last name and `studentid` is your student ID number. This zip file will contain the pdf file and a directory named `code` that contains all of your MATLAB source code. Additionally, the `code` directory must contain a file named `README` containing details on how to execute your MATLAB source code.
- Submit your completed assignment by email to `bmankala@eng.ucsd.edu` and `atripath@eng.ucsd.edu`. The subject of the email message must be **CSE 252B Assignment 5**. Attach the zip file to the message.
- It is highly recommended that you begin working on this assignment early to ensure that you have sufficient time to correctly implement the algorithms and prepare a report.

Problems

1. Point on line closest to the origin (5 points)

Given a line $\mathbf{l} = (a, b, c)^\top$, show that the point on \mathbf{l} that is closest to the origin is the point $\mathbf{x} = (-ac, -bc, a^2 + b^2)^\top$ (hint: this calculation is needed in the two-view optimal triangulation method used below).

2. Programming: Automatic estimation of the fundamental matrix (150 points)

(a) Feature detection (20 points)

Download input data from the course website. The file `IMG_5030.JPG` contains image 1 and the file `IMG_5031.JPG` contains image 2. In your report, include a figure containing the pair of input images.

For each input image, calculate an image where each pixel value is the minor eigenvalue of the gradient matrix

$$\mathbf{N} = \begin{bmatrix} \sum_w I_x^2 & \sum_w I_x I_y \\ \sum_w I_x I_y & \sum_w I_y^2 \end{bmatrix}$$

where w is the window about the pixel, and I_x and I_y are the gradient images in the x and y direction, respectively. Calculate the gradient images using the five-point central difference operator. Set resulting minor eigenvalues that are below a specified threshold value to zero (hint: calculating the mean instead of the sum in \mathbf{N} allows for adjusting the size of the window without changing the threshold value). Apply an operation that suppresses (sets to 0) local (i.e., about a window) nonmaximum pixel values in the minor eigenvalue image. Vary these parameters such that around 1350–1400 features are detected in each image. For resulting nonzero pixel values, determine the subpixel feature coordinate using the Förstner corner point operator.

In your report, state the size of the feature detection window (i.e., the size of the window used to calculate the elements in the gradient matrix \mathbf{N} and corner points), the minor eigenvalue threshold value, the size of the local nonmaximum suppression window, and the resulting number of features detected in each image. Additionally, include a figure containing the pair of images, where the detected features (i.e., corner points) in each of the images are indicated by a square about the feature, where the size of the square is the size of the detection window.

(b) Feature matching (15 points)

Determine the set of one-to-one putative feature correspondences by performing a brute-force search for the greatest correlation coefficient value (in the range $[-1, 1]$) between windows centered about the detected features in image 1 and windows centered about the detected features in image 2. Only allow matches that are above a specified correlation coefficient threshold value (note that calculating the correlation coefficient allows for adjusting the size of the matching window without changing the threshold value). Further, only allow matches that are above

a specified distance ratio threshold value, where distance is measured to the next best match for a given feature. Vary these parameters such that around 300 putative feature correspondences are established. Optional: constrain the search to coordinates in image 2 that are within a proximity of the detected feature coordinates in image 1.

In your report, state the size of the proximity window (if used), the correlation coefficient threshold value, the distance ratio threshold value, and the resulting number of putative feature correspondences (i.e., matched features). Additionally, include a figure containing the pair of images, where the matched features in each of the images are indicated by a square about the feature, where the size of the square is the size of the matching window, and a line segment is drawn from the feature to the coordinates of the corresponding feature in the other image (see Fig. 11.4(e) in the Hartley & Zisserman book as an example).

(c) Outlier rejection (20 points)

The resulting set of putative point correspondences should contain both inlier and outlier correspondences (i.e., false matches). Determine the set of inlier point correspondences using the M-estimator Sample Consensus (MSAC) algorithm, where the maximum number of attempts to find a consensus set is determined adaptively. For each trial, you must use the 7-point algorithm (as described in lecture) to estimate the fundamental matrix, resulting in 1 or 3 solutions. Calculate the (squared) Sampson error as a first order approximation to the geometric error.

In your report, describe any assumptions, including the probability p that at least one of the random samples does not contain any outliers (used to determine the number of attempts to find a consensus set), and the probability α that a given data point is an inlier and the variance σ^2 of the measurement error (both used to determine the distance threshold; hint: this problem has codimension 1). State the resulting number of inliers and the number of attempts to find the consensus set.

Additionally, include a figure containing the pair of images, where the inlier features in each of the images are indicated by a square about the feature, where the size of the square is the size of the matching window, and a line segment is drawn from the feature to the coordinates of the corresponding feature in the other image (see Fig. 11.4(g) in the Hartley & Zisserman book as an example).

(d) Linear estimation (15 points)

Estimate the fundamental matrix \mathbf{F}_{DLT} from the resulting set of inlier correspondences using the direct linear transformation (DLT) algorithm (with data normalization). Include the numerical values of the resulting \mathbf{F}_{DLT} , scaled such that $\|\mathbf{F}_{\text{DLT}}\|_{\text{Fro}} = 1$, in your report with sufficient precision such that it can be evaluated (hint: use `format longg` in MATLAB prior to displaying your results).

(e) Nonlinear estimation (70 points)

As described in lecture, parameterize the fundamental matrix as $(\boldsymbol{\omega}_u^\top, \boldsymbol{\omega}_v^\top, \boldsymbol{\sigma}^\top)^\top$, where $\|\boldsymbol{\sigma}\| = 1$, and calculate the camera projection matrices $\mathbf{P} = [\mathbf{I} \mid \mathbf{0}]$ and $\mathbf{P}' =$

$[\mathbf{M} | \mathbf{e}']$, where $\mathbf{M} = \mathbf{U}\mathbf{Z}\text{diag}(\sigma_1, \sigma_2, (\sigma_1 + \sigma_2)/2)\mathbf{V}^\top$, $\mathbf{U} = \exp([\boldsymbol{\omega}_{\mathbf{U}}]_{\times}) = [\mathbf{u}_1 | \mathbf{u}_2 | \mathbf{u}_3]$, $\boldsymbol{\sigma} = (\sigma_1, \sigma_2)^\top$, $\mathbf{V} = \exp([\boldsymbol{\omega}_{\mathbf{V}}]_{\times})$,

$$\mathbf{Z} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and $\mathbf{e}' = -\mathbf{u}_3$. Use \mathbf{F}_{DLT} and the triangulated 3D points as an initial estimate to an iterative estimation method, specifically the sparse Levenberg-Marquardt algorithm, to determine the Maximum Likelihood estimate of the fundamental matrix $\mathbf{F} = \exp([\boldsymbol{\omega}_{\mathbf{U}}]_{\times})\text{diag}(\boldsymbol{\sigma}^\top, 0)\exp([\boldsymbol{\omega}_{\mathbf{V}}]_{\times})^\top$ that minimizes the reprojection error. The initial estimate of the 3D points must be determined using the two-view optimal triangulation method described in lecture (algorithm 12.1 in the Hartley & Zisserman book, but use the ray-plane intersection method for the final step instead of the homogeneous method). Additionally, you must parameterize the homogeneous 3D scene points that are being adjusted using the parameterization of homogeneous vectors (see section A6.9.2 (page 624) of the textbook, and the corrections and errata).

In your report, show the initial cost (i.e., the cost at iteration 0) and the cost at the end of each successive iteration. Show the numerical values for the final estimate of the fundamental matrix \mathbf{F}_{LM} , scaled such that $\|\mathbf{F}_{\text{LM}}\|_{\text{Fro}} = 1$, in your report with sufficient precision such that it can be evaluated.

(f) **Point to line mapping (10 points)**

Qualitatively determine the accuracy of \mathbf{F}_{LM} by mapping points in image 1 to epipolar lines in image 2. Choose three distinct points $\mathbf{x}_{\{1,2,3\}}$ distributed in image 1 that are not in the set of inlier correspondences and map them to epipolar lines $\mathbf{l}'_{\{1,2,3\}} = \mathbf{F}_{\text{LM}}\mathbf{x}_{\{1,2,3\}}$ in the second image under the fundamental matrix \mathbf{F}_{LM} .

In your report, include a figure containing the pair of images, where the three points in image 1 are indicated by a square (or circle) about the feature and the corresponding epipolar lines are drawn in image 2. Comment on the qualitative accuracy of the mapping (hint: each line \mathbf{l}'_i should pass through the point \mathbf{x}'_i in image 2 that corresponds to the point \mathbf{x}_i in image 1).