LUND UNIVERSITY
FACULTY OF ENGINEERING
MATHEMATICS

TAKE-HOME EXAM
COMPUTER VISION 2014

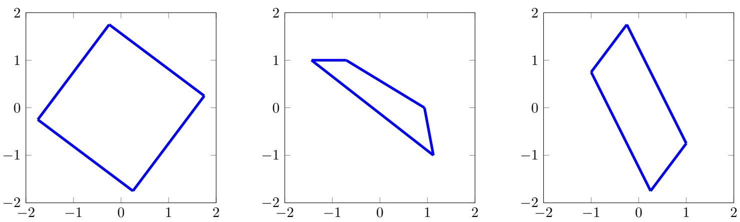
The solutions should be handed in no later than 48 hours after collecting the exam. You may use any books and computer programs (e.g., Matlab and Maple). Data for the exam can be downloaded from `www.maths.lth.se/matematiklth/vision/datorseende/exdata.zip`. *It is not permitted to get help from other persons. Credits can be given for partially solved problems. Write your solutions neatly and explain your calculations. All your programs should be submitted directly to* `calle@maths.lth.se`. *Both the content and the format of your solutions, and the difficulty of the problems solved, will affect your grade. For grade 4 or grade 5, it is not necessary to have solved four or five problems correctly, respectively, but it is sufficient.*

1. Compute the projection of the scene points $(1, 1, 0)$,$(0, 1, 2)$ and $(1, 1, 1)$ in the camera

$$P = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}. \tag{1}$$

Also compute the camera center of P in homogeneous coordinates. What is the geometric interpretation of the result?

2. The figures below show the output of applying 3 projective transformations to the unit square.



(a) (b) (c)

a) Characterize the transformations

$$H_1 = \begin{bmatrix} 3/4 & -1 & 0 \\ 1 & 3/4 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad H_2 = \begin{bmatrix} 3/4 & -1 & 0 \\ 1 & 3/4 & 0 \\ 0 & 0 & 5/4 \end{bmatrix}, \tag{2}$$

$$H_3 = \begin{bmatrix} 3/16 & -1 & -1/4 \\ 1/4 & 3/4 & 1/2 \\ 1/4 & 1/4 & 1 \end{bmatrix} \quad \text{and} \quad H_4 = \begin{bmatrix} 3/8 & -5/8 & 0 \\ 1/2 & 5/4 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{3}$$

as projective, affine, similarity or Euclidean. Determine which of the four transformations where used to generate the outputs in (a),(b) and (c).

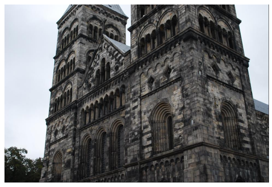
b) Show that a projective transformation H maps the conic

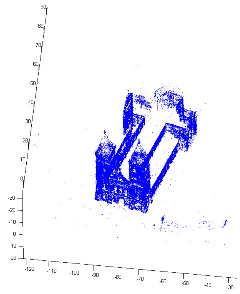
$$\{\mathbf{x} \in \mathbb{P}^2; \mathbf{x}^T C \mathbf{x} = 0\} \tag{4}$$

to

$$\{\mathbf{y} \in \mathbb{P}^2; \mathbf{y}^T H^{-T} C H^{-1} \mathbf{y} = 0\}. \tag{5}$$

3. Below are two images `cath1.jpg` and `cath2.jpg` depicting a famous cathedral. The file `ex3data.mat` contains a 3D model X (seen to the right below) and image points $x\{1\}$ and $x\{2\}$ for the two images that have been matched to the 3D model. (NaN in $x\{i\}$ means that the corresponding 3D point was not detected in image i.) The goal of this exercise is to determine the positions and orientations of the cameras that captured the images.





a) How many degrees of freedom does the uncalibrated pinhole camera have?
How many visible 3D points do you need to be able to compute the camera matrix?

b) Suppose that the number of mismatched points are roughly 20%. If you use a minimal set of correspondences, how many RANSAC iterations do you need to find an outlier-free set of point correspondences with 99.99% probability?

c) Write a function that computes a camera matrix from a minimal number of correspondences using DLT. Use RANSAC with this function to determine the two camera matrices. A point is considered to be an inlier if its projection is less than 5 pixels from the corresponding image point. Don't forget to make sure that your solutions have visible points in front of them. Plot the 3D model and the camera centers and principal axes in a 3D plot.

d) Compute the inner parameters of the cameras. Do they seems reasonable?

4. Consider the cameras $P_1 = \begin{bmatrix} I & 0 \end{bmatrix}$ and

$$P_2 = \begin{bmatrix} 1 & 0 & 0 & b \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

a) Show that the Essential matrix for this camera pair is

$$E = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -b \\ 0 & b & 0 \end{pmatrix}. \tag{6}$$

b) If $x = (x_1, x_2)$ and $y = (y_1, y_2)$ are projections (in regular Cartesian coordinates) of a scene point $\mathbf{X}$ in P_1 and P_2 respectively, show that

$$x_2 = y_2. \tag{7}$$

c) Suppose that $x = (x_1, x_2)$ and $y = (y_1, y_2)$ are projections of a scene point $\mathbf{X}$ in P_1 and P_2 respectively. If $\tilde{x} = (\tilde{x}_1, \tilde{x}_2)$ and $\tilde{y} = (\tilde{y}_1, \tilde{y}_2)$ are measurements of x and y corrupted by Gaussian Noise, that is,

$$\tilde{x} = x + \delta_1 \tag{8}$$
$$\tilde{y} = y + \delta_2 \tag{9}$$

where δ_i has the probability density function

$$p(\delta_i) = \frac{1}{2\pi} e^{-\|\delta_i\|^2/2}, \quad i = 1, 2, \tag{10}$$

show that the maximum likelihood estimation of x_1, x_2 and y_1 can be found by solving

$$\min_{x_1, x_2, y_1} (x_1 - \tilde{x}_1)^2 + (x_2 - \tilde{x}_2)^2 + (y_1 - \tilde{y}_1)^2 + (x_2 - \tilde{y}_2)^2. \tag{11}$$

What is y_2?

d) Show that any point on the line

$$\mathbf{X}(\rho) = \begin{pmatrix} x_1 \\ x_2 \\ 1 \\ \rho \end{pmatrix} \tag{12}$$

projects to the point $x = (x_1, x_2)$ in P_1 and that $\mathbf{X}\left(\frac{y_1 - x_1}{b}\right)$ also projects to $y = (y_1, y_2)$ in P_2 if $x_2 = y_2$.

e) The file `ex4.mat` contains the camera matrices `P1` and `P2` for the two images shown below. The variables `xtilde` and `ytilde` contain measured point projections in P1 and P2 respectively. Use rq-factorization to determine the calibration matrices. Normalize the image points and use the results from **c)** and **d)** to triangulate the 3D points.

5. In the 8-point algorithm the problem of estimating the camera motion and the scene structure is solved using 8 points. In this exercise we will see that 7 points is actually enough.

a) Using 7 point correspondences in two images one gets 7 linear constraints for the elements of the fundamental matrix F. Show that the set of all matrices that fulfills these 7 equations can be written

$$aF_0 + bF_1,$$

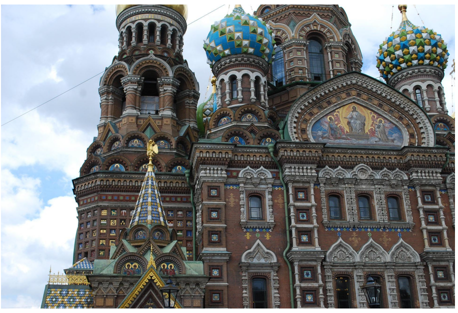
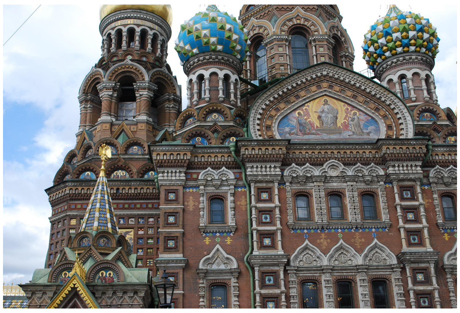
where F_0 and F_1 are 3×3 matrices, a and b are numbers.

b) Since the scale of a fundamental matrix is arbitrary we assume that $a = 1$ from now on. Use that any fundamental matrix F satisfies the constraint $\det(F) = 0$, to derive an equation for b. Show that this equation is a third degree polynomial in b.

c) Write a function that given 7 matches computes the solutions for F as outlined in **a)** and **b)**.

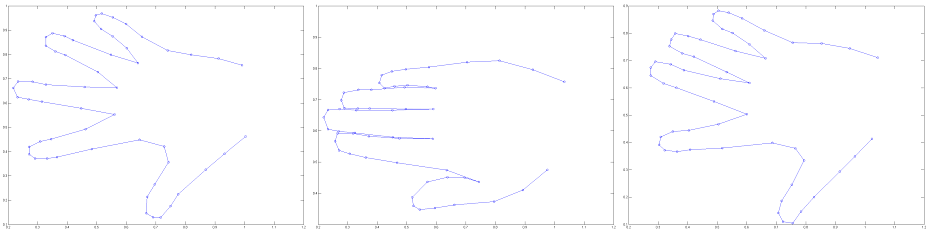
Hints: First form a matrix M such that $Mf = 0$, where f consists of the 9 entries of the fundamental matrix. Then compute F_0 and F_1 using svd and find the solutions for b that give determinant 0. To solve $\det(F_0 + bF_1) = 0$, without computing the polynomial by hand there are several options. If F_1 is invertible one can solve the eigenvalue problem $-F_1^{-1}F_0x = bx$. Another option is to evaluate $\det(F_0 + bF_1)$ for four different b's and and fit a polynomial $c_3b^3 + c_2b^2 + c_1b + c_0$ and then use the `roots` command.)

d) The cell `x` in the file `ex5.mat` contains corresponding points from the two images below. Use RANSAC with the 7 point solver from **c)** to find the fundamental matrix with the most inliers. A point is considered an inlier if the distance to its corresponding epipolar line is less than 5 pixels in both images. Note that the 7 point solver may give more than one solution. If this is the case then determine the number of inliers for all real-valued solutions. (Hint: Normalization should not be needed here.

e) Triangulate the points using DLT. Plot both the image points and projections in the images.

6. The file `ex6data.mat` contains points extracted from the silhouettes of 40 hands. The cell `a` contains the point coordinates of each image. You can plot the points and silhouette from image i using `plot(a{i}(1,:),a{i}(2,:),'o-');`.



a) Let the vector $x_{ij} \in \mathbb{R}^{2\times 1}$ ($i = 1, ..., m$, $j = 1, ..., n$) represent the noise free coordinates of the j'th point in image i. Further more let $a_{ij} \in \mathbb{R}^{2\times 1}$ be a noisy measurement of x_{ij},

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{12} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}. \tag{13}$$

Suppose that the coordinates fulfill

$$\begin{bmatrix} x_{i1} & x_{i2} & \dots & x_{in} \end{bmatrix} = \sum_{k=1}^{r} c_{ik} B_k, \tag{14}$$

where $c_{ik} \in \mathbb{R}^{2\times 1}$ and $B_k \in \mathbb{R}^{1\times n}$. Explain why the matrix X has rank r.

b) Form the matrix A from the measurements in the cell `a` and find the best rank 7 approximation X by solving

$$\min_X \quad ||A - X||_F^2 \tag{15}$$

$$\text{such that} \quad \text{rank}(X) = 7. \tag{16}$$

What is the resulting relative error $\frac{||A-X||_F}{||A||_F}$? Plot the original and the estimated point positions and silhouettes in the same image for frame 10. Do they look similar?

c) Determine a shape basis B_k, $k = 1, ..., 7$ such that (14) is fulfilled. Is the solution unique?

d) If the shape basis is known and we are given points from a previously unseen image we can find the coefficients that give the most similar shape in our model by solving

$$\min_{c_{ik}} \left\| \begin{bmatrix} a_{i1} & a_{i2} & \dots & a_{in} \end{bmatrix} - \sum_{k=1}^{r} c_{ik} B_k \right\|_F^2. \tag{17}$$

The cell `anoise` contains point coordinates (heavily corrupted by noise) from 3 previously unseen images. Solve (17) for the unseen images and plot the resulting silhouettes together with the noisy points. Does the result look reasonable?

e) Suppose that we are given a new image where only a few of the points have been detected. To determine the coefficients c_{ik} we can remove the entries of B_k from (17) and solve using the remaining residuals. How many points are needed to determine the coefficients c_{ik} if $r = 7$?

The cell `amissing` contains a subset of the point coordinates from 3 previously unseen images. (The missing entries are marked with NaN). Determine the coefficients c_{ik} by solving (17). Use the coefficients and the full B_k vectors to compute coordinates for all of the points. Plot the result together with the data from `amissing`. Does the it seem ok?

Good Luck!