Assignment 1: CS 763, Computer Vision

Due: 2nd February before 11:55 pm

Remember the honor code while submitting this (and every other) assignment. All members of the group should work on and understand all parts of the assignment. We will adopt a zero-tolerance policy against any violation.

Submission instructions: You should ideally type out all the answers in Word (with the equation editor) or using Latex. In either case, prepare a pdf file. Put the pdf file and the code for the programming parts all in one zip file. The pdf should contain the names and ID numbers of all students in the group within the header. The pdf file should also contain instructions for running your code. Name the zip file as follows: A1-IdNumberOfFirstStudent-IdNumberOfSecondStudent-IdNumberOfThirdStudent.zip. (If you are doing the assignment alone, the name of the zip file is A1-IdNumber.zip). Upload the file on moodle BEFORE 11:55 pm on 2nd February. Late assignments will be assessed a penalty of 50% per day late. Note that only one student per group should upload their work on moodle. Please preserve a copy of all your work until the end of the semester. If you have difficulties, please do not hesitate to seek help from me.

1. Suppose you have acquired the image of a cricket pitch at the time instant that a ball thrown by the bowler landed on the ground (somewhere on the pitch) at some point say Q. Given this image, your task is to determine the perpendicular distance from Q to the line containing the wickets on the bowler's side. Make use of the standard dimensions of a cricket pitch as seen on `https://en.wikipedia.org/wiki/Cricket_pitch#/media/File:Cricket_pitch.svg`. Assume that the ball and all the sides of the pitch were clearly visible in the image. Now prove analytically that you do not need a calibrated camera for this calculation (ignoring errors due to discretization of the spatial coordinates). [3+3 = 6 points]

2. Consider an image of a static scene acquired by a camera fixed on a tripod. Now the camera is rotated (but it remains fixed on the tripod without any translation) and another picture of the same scene is acquired. Let $\mathbf{p_1}$ and $\mathbf{p_2}$ be the pixel coordinates of the images of some physical point in the scene in the two images respectively. Note that $\mathbf{p_1}$ and $\mathbf{p_2}$ are in different coordinate systems. Derive a relation between $\mathbf{p_1}$ and $\mathbf{p_2}$ in terms of the matrix $\mathbf{R}$ which represents the rotational motion of the camera axes from the first position to the second, and the intrinsic parameter matrix $\mathbf{K}$ of the camera. Furthermore, if $\mathbf{K}$ is known, explain how will you determine $\mathbf{R}$. [6 points]

3. In class, we have seen image formation on a flat screen (i.e. image plane) with a pinhole camera. Now suppose the screen was wrapped on the inner surface of a hemisphere and hence, the 3D points were projected onto a concave hemispherical surface. Derive a relationship between the coordinates of a 3D point $P = (X, Y, Z)$ and its image on such a screen (both in camera coordinate system). If you had to calibrate this sort of a system, what are the additional intrinsic parameters of the camera as compared to the case of an image plane ? [6 points]

4. In this exercise, we will prove the orthocenter theorem pertaining to the vanishing points Q, R, S of three mutually perpendicular directions OQ, OR, OS, where O is the pinhole (origin of camera coordinate system). Let the image plane be $Z = f$ without any loss of generality. Recall that two directions v_1 and v_2 are orthogonal if $v_1^T v_2 = 0$. One can conclude that OS is orthogonal to $OR - OQ$ (why?). Also the optical axis Oo (where o is the optical center) is orthogonal to $OR - OQ$ (why?). Hence the plane formed by triangle OSo is orthogonal to $OR - OQ$ and hence line oS is perpendicular to $OR - OQ = QR$ (why?). Likewise oR and oQ are perpendicular to QS and RS. Hence we have proved that the altitudes of the triangle QRS are

concurrent at the point o. QED. Now, in this proof, I considered the three perpendicular lines to be passing through O. What do you think will happen if the three lines did not pass through O? [6 points]

5. Consider two sets of corresponding points $\{\mathbf{p}_{1i} = (x_{1i}, y_{1i})\}_{i=1}^{n}$ and $\{\mathbf{p}_{2i} = (x_{2i}, y_{2i})\}_{i=1}^{n}$. Assume that each pair of corresponding points is related as follows: $\mathbf{p}_{2i} = \alpha \mathbf{R} \mathbf{p}_{1i} + \mathbf{t} + \eta_i$ where $\mathbf{R}$ is an unknown rotation matrix, $\mathbf{t}$ is an unknown translation vector, α is an unknown scalar factor and η_i is a vector (unknown) representing noise. Explain how you will extend the method we studied in class for estimation of $\mathbf{R}$ to estimate α and $\mathbf{t}$ as well. Derive all necessary equations (do not merely guess the answers). [6 points]

6. You are given two datasets in the folder `http://www.cse.iitb.ac.in/~ajitvr/CS763_Spring2016/HW1/Calib_data`. The file names are Features2D_dataset1.mat, Features3D_dataset1.mat, Features2D_dataset2.mat and
Features3D_dataset2.mat. Each dataset contains (1) the XYZ coordinates of N points marked out on a calibration object, and (2) the XY coordinates of their corresponding projections onto an image plane. Your job is to write a MATLAB program which will determine the 3×4 projection matrix $\mathbf{M}$ such that $\mathbf{P_1} = \mathbf{MP}$ where $\mathbf{P}$ is a $4 \times N$ matrix containing the 3D object points (in homogeneous coordinates) and $\mathbf{P_1}$ is a $3 \times N$ matrix containing the image points (in homogeneous coordinates). Use the SVD method and print out the matrix $\mathbf{M}$ on screen (include it in your pdf file as well). Write a piece of code to verify that your computed $\mathbf{M}$ is correct. For any one dataset, repeat the computation of the matrix $\mathbf{M}$ after adding zero mean i.i.d. Gaussian noise of standard deviation $\sigma = 0.05 \times max_c$ (where max_c is the maximum absolute value of the X,Y,Z coordinate across all points) to every coordinate of $\mathbf{P}$ and $\mathbf{P_1}$ (leave the homogeneous coordinates unchanged). Comment on your results. Include these comments in your pdf file that you will submit. **Tips:** A mat file can be loaded into MATLAB memory using the 'load' command. To add Gaussian noise, use the command 'randn'. [10 points]

7. In this exercise, you will estimate the homography between a pair of images using the method we studied in class. You should use the well-known SIFT algorithm to (1) detect salient feature points in both the images, and (2) determine pairs of matching points given the two point sets ('matching point pair' refers to points in the two images representing the same physical entity). The code for performing both these tasks is available at `http://www.cs.ubc.ca/~lowe/keypoints/`. We may study the internal details of how SIFT works in a separate set of lectures in class, but for this exercise, just assume this package is a magic blackbox. Now, given this set of matching pairs of points produced by the SIFT package, your job is to estimate the homography between the point sets. Write a routine of the form H = homography(im1,im2) where H is the homography matrix that will transform the first image. You will use data from the folder `http://www.cse.iitb.ac.in/~ajitvr/CS763_Spring2016/HW1/Homography/`. Do as follows:

   (a) Apply the homography transformation in the file 'Hmodel.mat' to the image 'goi1_downsampled.jpg' using reverse warping to generate a warped image. Now estimate the homography that transforms the first image into its warped version. Apply the estimated transformation to the first image (using reverse warping) and display all three images side by side in your report. Also print the model and estimated homography matrices (make sure you normalize both so that $H(3,3) = 1$ in both cases).

   (b) Determine the homography that transforms the image 'goi1_downsampled.jpg' to the second image 'goi2_downsampled.jpg'. Warp the first image (using reverse warping) and compare it to the second. Display all three images side by side in your report. Also print the estimated homography matrix normalized so that $H(3,3) = 1$.

   Note: You may not get perfect answers for the motion estimate due to errors in SIFT, but you should get a reasonable alignment. While warping, crop off the portions of the image that do not fit into the original size. You may use the nearest neighbor method for interpolation during warping. I encourage you to try out this experiment on images of planar surfaces from different viewpoints that you should take with a real camera. You will notice that the warp estimate will often be very wrong due to several incorrect matches (called as 'outliers'). In a subsequent assignment, we will implement a method that will be reasonably immune to these outliers. At that point, we will attempt to mosaic together two or more pictures as well. [10 points]