טכניון – מכון טכנולוגי לישראל
פקולטה למדעי המחשב

Computer vision 236873 Final Exam A

Guide lines

1. You are allowed to use two A4 pages of reference material.
2. You can use any non-graphical calculator (The use of other electronic devices such as lap tops, cell phones etc. is prohibited).
3. Exam duration 2 hours.
4. There are 4 questions worth 25 points, each with several sections.
5. There are 8 pages in this exam.
6. Write your answers in the allocated space for each question.
7. Please use clear hand writing and explain every phase of your answers. Only well explained (correct) answers will reward you full grade.

Good luck!

Question 1 –Camera Calibration: (25 points)

The relationship between a 3D point at world coordinates (X,Y,Z) and its corresponding 2D pixel at image coordinates (u,v) can be defined as a projective transformation using a 3 × 4 camera projection matrix P.

a. Give the main steps of an algorithm for computing the matrix P from a single image of a known 3D "calibration object." **Note:** Please write down the entire mathematical operation. (6 points)

Answer:

Let P_1,P_2,P_3 be the three rows of the projection matrix and let P_123 be a 12-vector of their concatenation. Every image coordinate, e.g. u, and the corresponding 3D coordinates X (4 vector) specify a constraint P_1^T X u = P_3^T X, which is linear in the components of P_123. The constraints for m points may be written as a matrix equation A P_123 = 0 where every line of the (2m x 12) matrix A contains the linear constraints coefficient associated with one image coordinate. With reasonable large m, the homogeneous system does not have a solution and a good approximation is the right singular vector with the lowest singular value (found using SVD).

For a more complete answer, refer to the camera calibration tutorial.

b. Consider the option of down-sampling each dimension of the image by 4.
   1. Does it affect the intrinsic camera matrix? If so explain how.
   2. Can this operation be replaced by a zoom operation? If so explain why. (6 points)

Answer:

1. It changes the intrinsic camera matrix by scaling Sx,Sy by 4 and also by changing the principal point (note that the sensor location relative to the camera does not change but the unit by which we measure the principal point, do.
2. Multiplying f by .25 compensates for the scaling but not for the principal point changes. Therefore, zoom operation cannot replace the scaling.

Note: There were many diverse answers to 2b, taking into account also factors such as aliasing.

c. Consider a set of parallel lines which are also parallel to the ZX plan (in the camera coordinate system, Z is the optical axis). What can you say about the intersection of these lines in the image? (6 points)
Answer:
A parametric expression for a 3D line is X_0+Dt (where X0,D are 3D vectors, and t is a scalar parameters. For lines parallel to the XZ plane the y component of D, denoted Dy, is 0. The vanishing point (in idea camera image) is (u,v)=(Dx/Dz,Dy/Dz)=(Dx/Dz,0). That is, all vanishing points have zero v coordinate and are of the line v=0. (The vanishing line).

d. Consider a point (u',v') in the image plane of a camera characterized by known intrinsic and extrinsic projection matrices. Give an expression (not necessarily explicit) for the set of 3D points projected to (u',v') in world coordinate system and in the camera coordinate system. (7 points)
Answer:
First, consider the case of the world coordinate system. Let K=M_int, M_ext be the intrinsic and extrinsic projection matrices respectively. The full projection matrix from world coordinates to image coordinates is P_w = K M_ext. Denote the rows of this projection matrix P1,P2,P3. All of them are 4-vectors. We know that u'=P1^TX/P3^TX, v'=P2^TX/P3^TX. This are two linear constraints on the homogeneous coordinates vector X, which (together) limit it to a line.
Consider now the case of camera coordinate system The extrinsic projection matrix in the camera coordinate system simply M_ext=[I | 0] . The rest is the same.

Question 2 –Edge Detection and Corner Detection: (25 points)

a. During the final project of the course, a student is required to find edges. Explain why it is a good idea to look for zero crossings in the image response to Difference of Gaussians. Explain what does the DoG approximates, and why it is commonly used. (5 points)
Answer:

The difference of Gaussians is an approximation to the Laplacian of Gaussian operator, but is more efficient computationally, since the Gaussian filter is linear, separable, and by the semigroup property, convolution with it may be calculated as several convolutions with smaller Gaussian, which is more efficient. Both the LoG and the DoG enable better localization of the edge, since it provides the (thin) location where the second derivative changes signs, meaning the location where the edge has the most contrast.

b. The student chose to apply Canny edge detector algorithm on the input images. Help him understand which claims regarding the algorithm are correct and which aren't. **Explain your answer with regard to each claim.** (10 points)
   1. Two threshold values are required to connect the edge points, since the threshold depends on the direction of advancement.
   2. Using two threshold values makes the non-maximal suppression step redundant.
   3. Using two threshold values prevents gaps that may be created in the final description of the edge.
   4. The oriented derivatives in x and y directions require different threshold values.
   5. Smoothing with a Gaussian with bigger variance will capture less fine edges.

   Answer:

   Claims 3 and 5 are the correct statements regarding Canny edge detector.

c. Harris corner detector algorithm is invariant to several transformations, meaning the same key-point would be detected after applying the transformation on the input image. For the following list of transformation, determine for which Harris corner detector is invariant and explain why. (4 points)
   1. Scale change.
   2. Translation.
   3. Rotation.
   4. Blur.

   Answer:

1. Harris corner detector is not scale and blur invariant.
2. Harris corner detector is translation and rotation invariant.

The key-points detected in 2 images will be the same when both translation and rotation are applied, but not when scale and blur are applied.

d. The student used Harris corner detector output as an edge detector. Do you think it can work? If yes, explain why and if not, give an example of a situation where it will fail. (6 points)
Answer:

An edge detector should work for straight edges (and perhaps also for other edges). For straight edges the second largest eigenvalue is zero. Therefore, if we want to compare Shi-Tomassi version to a threshold, then this threshold must be zero for respond for straight edges implying that the detector would respond also to constant regions. In this conditions, the Harris-Stephens version responds negatively for edge and the situation is even worse. Note that:

a. For curves edges the ability to use the algorithm is improved.
b. Using the two eigenvectors in another way could work but this is no longer the Harris operator.

Question 3 –Epi-Polar Geometry: (25 points)

a. Suppose you want to 3D reconstruct the following scene using two cameras. You are given a choice of two stereo configurations: the first is with a horizontal (with regard to the staircase) baseline and with a vertical baseline. Which of the two configurations will provide a better reconstruction? Explain why. (8 points)

Answer:

A stereo pair with a vertical baseline is preferred since the correspondence along the vertical epi-polar lines would be more accurate.

b. Consider the following matrices:
   1. F – Fundamental matrix.
   2. E – Essential matrix
   3. H – Homography matrix

   Describe shortly in which conditions you would be able to use each one to describe the image correspondence. (3 points)

   Answer:

   The epi-polar constraint can always be used to describe correspondence between two images. If the intrinsic camera calibration matrixes are unknown then the Fundamental matrix should be used. If the intrinsic camera calibration matrixes are known, then the Essential matrix can be used. In case the images describe planar scene, then a Homography can be used to describe the transformation between the two images.

c. A photographer, using a single camera, took two successive images of the staircase, when he was walking straight towards the staircase, so that the center of projection stays on the optical axis. Draw the epi-polar lines in this case. Will he be able to 3D reconstruct the scene? (7 points)
   Answer:

   The Epi-polar lines would intersect at the optical centers of each image plane. Since there is a translation movement between the two optical centers, a 3D reconstruction would be possible, as there is a depth information.

d. Now the photographer took two successive images of the staircase, when he was standing in the same spot, but rotating the camera about the optical axis. Will he be able to 3D reconstruct the scene? (7 points)
Answer:

As there is no translation between the optical center, there is no additional depth information, and 3D reconstruction wouldn't be possible.

Question 4 –Recognition: (25 points)

A common approach to object detection consists of the following stages:

1. Calculate pixel-wise gradient.
2. Generate object proposals (sub-images). Each proposal is then divided into cells.
3. Calculate histogram of the gradient directions, weighted by the gradient magnitude in each cell and concatenating the histograms into feature vector X.
4. Classify the vector X using a classifier of choice.

Note: The proposal generation is an external process which is not part of the algorithms. Do not refer to it in the questions below.

a. What is the name if this process? (3 points)
Answer:
The algorithm considers several proposals, extract a vector of concatenated gradient direction histograms, and classify it. That is, in each proposal it runs the HOG algorithm.

b. Discuss briefly the different factors, which influence the input image and may interfere with the detection. Briefly explain each one, and how this algorithm reduces the interference. (7 points)

Answer:

a. The image of the same object changes with illumination – the algorithm is based on gradient direction which is less sensitive to illumination than the intensity.

b. The image of the same object changes with pose – the algorithm uses cell which contain parts, which change less. It uses histograms which are less sensitive to small pose changes. It uses a (learning based) classifier that can use examples of different poses.

c. The image changes with changes within the class - the algorithm uses a (learning based) classifier that use different examples from the class.

c. For an input image of size N pixels, there are M object proposals, K cells in a proposal and each histogram consists of B bins. Give an estimation of the detection algorithm computational complexity. You can assume that the classifier is linear. When calculating the complexity, do so for the worst case. (7 points)

Answer:

Gradient calculation O(N)

For each proposal: Histograms construction O(N) – (with worst case assumption that each of the proposals is nearly as large as the image.). Constructing X from histograms – O(KB). Classification with linear classifier O(KB). Overall O(N+KB).

Overall: O(N+M(N+KB))=O(M(N+KB)).

d. Can you suggest a method of lower computational complexity that implements the same detection process? Explain. **Note:** assume M is very large. (8 points)

Answer:

The trick is to use integral images:

a. First find gradients, and quantize them into B bins, and for each quantized value, separately, build an image containing in each pixel the gradient size value (if the gradient direction corresponds to this quantized value) or 0 (otherwise). For each such image build an integral image. This takes O(NB) time.

b. For each cell build gradient direction histogram from integral image in O(B) time. Build the vector for a proposal and classify in O(KB) time.

c. For all proposals, the overall complexity is O(NB+MKB). As M is large(st) and N is the 2nd largest, this is better than the previous algorithm's complexity.