School of Computer Science and Communication, KTH
Lecturer: Mårten Björkman

EXAM

Image Analysis and Computer Vision, DD2423

Tuesday, 14^{th} of January 2014, 14.00–19.00

Allowed helping material: Calculator, the mathematics handbook Beta (or similar).

Language: The answers can be given either in English or Swedish.

General: The examination consists of Part A and Part B. For the passing grade E, you have to answer correctly at least 80% of Part A. If your score is less than 80%, the rest of the exam will not be corrected. Part B of the exam consists of **six** exercises that can give at most 50 points.

The results will be announced within three weeks.

Part A

Provide short answers to the questions! Each answer is worth maximum one point.

1. What factors determine the intensity of a pixel when measured by a camera sensor?
2. What kind of errors affect the quality of pixels in the sampling process?
3. Why is a definition of neighbourhood system necessary for connected components?
4. Why is it enough to know the impulse response of a linear shift-invariant filter to know the effect it has on all signals (images)?
5. What is the benefit of Fourier transforms for image filtering?
6. What does it mean that a Fourier transform is conjugate symmetric?
7. In what sense is an ideal low-pass filter not really ideal?
8. Why is the notion of scale important in image analysis?
9. What characterizes an image feature that is good for stereo matching?
10. What is the similarity and difference between K-means and Mean-shift for image segmentation?
11. What do graph based segmentation methods normally try to optimize?
12. What invariances should an object recognition method typically be able to handle?
13. What components do most feature based recognition methods consist of?
14. Derive the relation between the disparity and the depth of a 3D point for parallel cameras.
15. Why can it be hard to separate rotations from translations when computing the motion of a camera using corner features tracked over time?

Part B

Exercise 1 (1+3+3=7 points)

1. Assuming that you have two perspective cameras with different focal lengths over-looking a scene, for which camera are parallel 3D lines closest to being parallel also in the 2D projection?

2. Compute the projections of the 3D points with homogeneous coordinates

$$\mathbf{X}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 1 \end{pmatrix}, \ \mathbf{X}_2 = \begin{pmatrix} 2 \\ 1 \\ 1 \\ 1 \end{pmatrix} \text{ and } \mathbf{X}_3 = \begin{pmatrix} 2 \\ 3 \\ -1 \\ 1 \end{pmatrix}$$

   in the camera with projection matrix

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}.$$

   What is the interpretation of the projection of $\mathbf{X}_3$?

3. Let $P_i = (X_i, Y_i, Z_i)^T$, $i = 1, \dots, N$, be a set of 3D points with image projections $p_i = (X_i/Z_i, Y_i/Z_i)^T$. Assuming a rigid camera motion, the transformed point coordinates are given by

$$P_i' = RP_i + T,$$

   where R is the rotation and T is the translation of the motion, with projections $p_i' = (X_i'/Z_i', Y_i'/Z_i')^T$. Show that in the case of pure rotation ($T = 0$), it is not possible to recover the depths Z_i given any number of matched pairs (p_i, p_i').

Exercise 2 (2+3+2=8 points)

1. What properties do Gaussian filters possess that make them suitable for scale-space representation?

2. Show through derivation that the Gaussian function

$$g(x,y;\, t) = \frac{1}{2\pi t} e^{-(x^2+y^2)/2t}$$

   satisfies the heat equation

$$L_t = \frac{1}{2}\nabla L, \quad \text{where } \nabla L = L_{xx} + L_{yy}.$$

3. From the above, show that the scale-space representation of an arbitrary function f, that is

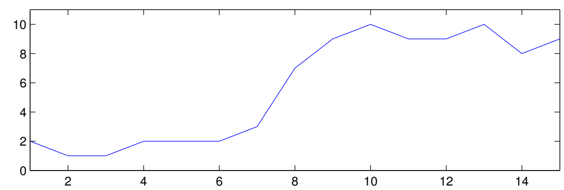
$$L(x,y;\, t) = g(x,y;\, t) * f(x,y)$$

   also satisfies the heat equation.

Exercise 3 (2+2+3+3=10 points)

1. To the 1D image $f = [\,2,1,1,2,2,2,3,7,9,10,9,9,10,8,9\,]$ (see below) apply the two filter kernels

$$g_1 = [\,1,2,1\,] \text{ and } g_2 = [\,1,0,-1\,].$$



2. The image obviously includes an edge at about $x = 8$. Propose a sharping method to enhance this edge and apply the method to the image.

3. Consider a 3×3 spatial filter mask

$$h = \frac{1}{8}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 4 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

that computes the weighted average of the center point and its four closest neighbours. Find the corresponding frequency space representation $H(u,v)$ and show that the filter is a lowpass filter.

4. Propose a suitable discrete filter that approximates a Laplacian and show using Taylor Series expansion how it compares to a true Laplacian with transfer function $G(u,v) = -(2\pi)^2(u^2+v^2)$.

Exercise 4 (3+1+2+3=9 points)

1. Assume you have the image below. If you where to perform histogram equalization on this image, how would the resulting image look like? Explain the individual steps of the computation.

0	0	0	0	0	0	0	0
0	1	1	1	1	1	1	1
0	1	2	2	2	2	2	2
0	1	2	3	3	3	3	3
0	1	2	3	4	4	4	4
0	1	2	3	4	5	5	5
0	1	2	3	4	5	6	6
0	1	2	3	4	5	6	7

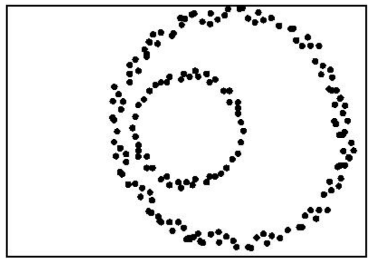
2. Why does discrete histogram equalization rarely lead to uniform histograms?

3. Suggest two 3×3 differential operators, one for x-wise derivatives and one for y-wise derivatives. Why are these just approximations of derivatives?

4. Apply the two operators to the image above and compute a second moment matrix for the whole image, using a uniform window function. What can the second moment matrix be used for?

Exercise 5 (2+3+3=8 points)

Consider an image that consists of two sets of points. For the first set, points are spread around a circle of radius R centered at point C_1, which is near the center of the image. The other set of points is spread on a circle of radius $2R$, which is centered at point C_2 located inside the other circle. Assume that the points in each set are distributed densely enough so that the distances between points on the same circle are smaller than the distances between points on different circles.



1. Assume that we want to divide the points into two clusters corresponding to the circles. Describe what result can be expected when using the K-means clustering algorithm on the point positions.
2. Explain how the circles could be found using RANSAC. What would the necessary steps be?
3. Assume that you instead try to use a Hough transform to find the circles. How could the accumulator space be set up and what steps would you need to perform?

Exercise 6 (1+3+3+1=8 points)

1. Why is the epipolar geometry constraint relevant to stereo matching?
2. Assume you have two cameras, c_1 and c_2, where c_2 is placed two units to the right of c_1 and one unit forward in the coordinate system of c_1. Also assume that c_2 is rotated 30° around the y-axis (see figure below). Compute the essential matrix that relates image points between the cameras.
3. Where are the epipoles in the images of c_1 and c_2? How are these related to the epipolar lines?
4. Assume that you don't know the position and orientation of c_2 with respect to c_1. How many image point matches between the cameras would you theoretically need to find the essential matrix?