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

EXAM

Image Analysis and Computer Vision, DD2423

Wednesday, 12th of December 2012, 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 bonus credits from the labs will first be added to Part A until you reach 80% - the rest will then be used for Part B.

The results will be announced within three weeks.

Part A

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

1. Assume you have the choice between different cameras and camera settings. How can you reduce the noise per pixel, without applying spatial filtering?
2. If you are expressing perspective projections using homogeneous coordinates, why doesn't a scaling of the homogeneous coordinates matter?
3. What is the difference between the intrinsic and extrinsic camera parameters?
4. When you subsample an image, what is the reason for the aliasing artifacts that you might see in the subsampled image, if you think of the image in frequency domain?
5. What is a 'neighborhood system' and a 'connected component', and how are these two concepts related?
6. How can you measure the distance between two points in an image? Mention at least two possible distance measures.
7. What does a convolution in the spatial domain correspond to in the Fourier domain? Why is this relevant?
8. What does it mean that a 2D kernel is separable? Is a Fourier Transform separable?
9. Why are frequency domain representations of filters often easier to deal with than their corresponding spatial representations?
10. What is a 'difference of Gaussians' and why is it useful?
11. Mention a segmentation method that doesn't take spatial coherence into consideration. Why is it often a bad idea to ignore spatial coherence for segmentation?
12. Describe shortly how one can create a multi-scale (scale-space) representation of an image. Why are such representations useful in computer vision?
13. Assume you want to separate a bunch of nutts from screws based on shapes. Give an example of a one-dimensional shape description that could serve the purpose, and explain how you compute it.
14. Give two reasons why you are often interested in 'dimensionality reduction' in computer vision.
15. In what sense does nearest neighbour classification differ from Bayesian classification?

Part B

Exercise 1 (2+3+2=8 points)

As long as you know what model suits your camera, surprisingly much information can be derived about the world from simple measurements, in particular if the camera is in motion.

1. If we use a pin-hole camera model, what is the relation between the camera coordinates of a 3D point (X,Y,Z) and the coordinates of its projection in the image plane (x,y)? How does the relation look for an orthographic projection?
2. Assume that you are approaching a red door with a camera-equipped robot and that at time $t_0 = 0\ s$ the door covers 30'000 pixels of image area. Two seconds later, at time $t_2 = 2\ s$, the door covers an additional 10'000 pixels. When can you assume the robot (or at least the camera) to crash into the door, if you keep approaching with the same speed?
3. If you know that you were moving with a speed of about $0.25\ m/s$ and a door is about $2\ m$ in height and $0.75\ m$ in width, what is the approximate focal length f of the camera (measured in pixels)?

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

Most filters that we use for spatial filtering of images are linear. Unlike non-linear filter these are easier to understand, especially when studied in the frequency domain.

1. What are the three properties of linear shift-invariant (LSI) filters? Why do each of these properties make it easier to understand the behaviour of filters when you apply them to images?
2. What is the discrete Fourier Transform of the 4-pixel image $f = [2,3,1,4]$?
3. What is the difference in magnitude and phase between the Fourier Transform of $g = [0,x_1,x_2,x_3]$ and that of $h = [x_1,x_2,x_3,0]$? Why is this relevant when it comes to interpreting image data?
4. In the labs we have used the filter $\delta_x = [-\frac{1}{2},0,\frac{1}{2}]$ as an approximation of an x-wise derivative. For low frequencies, its continuous Fourier Transform $\hat{\delta}_x(\omega) = i\sin(\omega)$ is close to the transfer function of a true 1st order derivative, that is $D_x(\omega) = i\omega$. Instead of using a filter of length 3, find an even better approximation using a differential filter of length 5.

Exercise 3 (2+3+2=7 points)

Some mothers-in-law often take images that are not as good as you would have hoped. For some reason you always find a good reason to discard them, but before doing so you could try whatever method you knows to possibly improve the quality.

1. Draw a grey-level transformation that would enhance the contrasts when applied to most typical images. In which parts of the transformation do you get a stretching of grey-level values, and where do you get a compression?

2. For a 100 pixel image, you have a histogram $h_x(x) = [17, 28, 14, 20, 8, 5, 6, 2]$, where $x \in [0, 7]$ is a grey-level value. Unfortunately, the image is a quite dark with poor contrasts. Apply histogram equalization and find a grey-level transformation $y = T(x), y \in [0, 7]$, that would make a more balanced image. What is the transformation and what will the histogram $h_y(y)$ be after applying this transformation?

3. You are given an image that is somewhat blurred because it was taken slightly out of focus. Describe a method that could be used to sharpen the edges in this image, making the image more visibly pleasing. What linear filters do you suggest for the purpose?

Exercise 4 (2+3+2=8 points)

Feature (such as edges and corners) are important in computer vision, because the existence of a feature can tell us something about the world we are looking at.

1. In the labs we have used a number of differential operators for edge detection.

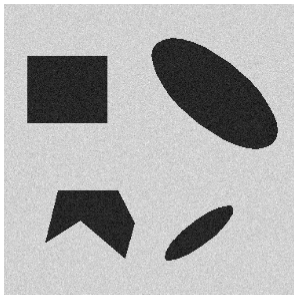
   From a given image f that looks like the one to the right, the following differential quantities are computed:

   (a) L_x (x-wise derivative)

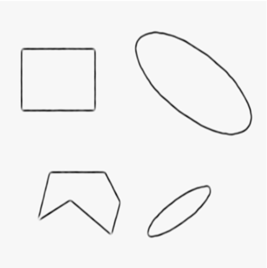
   (b) L_y (y-wise derivative)

   (c) $\nabla^2 L = L_{xx} + L_{yy}$

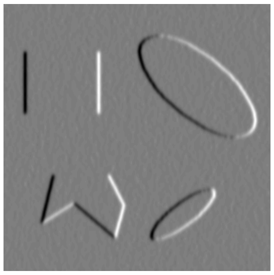
   (d) $\tilde{L}_{vvv} = L_x^3 L_{xxx} + 3L_x^2 L_y L_{xxy} + 3L_x L_y^2 L_{xyy} + L_y^3 L_{yyy}$



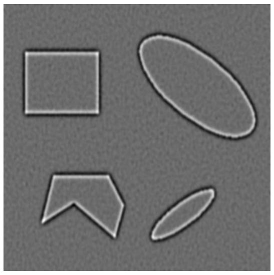
   Here $L = g * f$ where g is a small Gaussian kernel. Below you see these four quantities as grey-level images. Pair the differential quantities above to the correct images below. Motivate the matches!



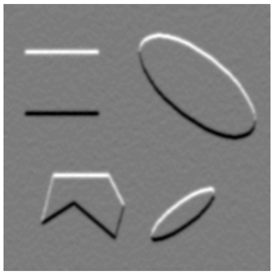
A



B



C



D

2. For Harris corner detector you first blur the image (to get rid of noise) and then compute the x-wise and y-wise derivatives. Describe (in some detail) the remaining steps of the method. How can you tell that you have found a corner and not an edge pixel?

3. Assume you match corner features $p_i = (x_i, y_i)$ in one image I to corner features $p_i' = (x_i', y_i')$ in another image I', and that all these features come from the same plane in 3D space. How does the relation between the coordinates of the matching features look like? What is the minimum number of matching features you need in order to determine the parameters of the relation?

Exercise 5 (5+2+4=11 points)

1. You have a camera image of a blue table with a collection of fruit in-front of you. Let us assume that you want to classify the individual pixels into either **fruit** (class A) or **table** (class B) pixels, given 2D color measurements $z = (x,y)^T \in [-1,1]$. Here component x represents red-green colours, while y represents blue-yellow colors. Since you want to make the classification invariant to white illumination, the (grey-scale) luminance component is ignored.

   The prior probabilities of the two classes are assumed to be $p(k = A) = \frac{1}{4}$ and $p(k = B) = \frac{3}{4}$, that is most pixels can be expected to belong to the table. Furthermore, the fruit and table pixels are assumed to be distributed according to

   $$p(z|k) = \frac{1}{2\pi|det\Sigma_k|^{1/2}} e^{-(z-m_k)^T\Sigma_k^{-1}(z-m_k)/2}$$

   with

   $$\Sigma_A = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{4} \end{pmatrix}, \Sigma_B = \begin{pmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{4} \end{pmatrix}$$

   and mean colours $m_A = (0,0)^T$ and $m_B = (0,1)^T$. Given a particular color measure z, what is the most probable classification of the corresponding pixel? Draw an illustration of the distributions, derive the decision function and show the decision boundaries in the illustration.

2. You are given a set of seven points placed on a 2D plane

   $$z_1 = (8,6),\ z_2 = (1,5),\ z_3 = (5,5),\ z_4 = (0,4),\ z_5 = (7,3),\ z_6 = (5,3) \text{ and } z_7 = (2,2).$$

   Apply K-means clustering with $c_1 = (0,0)$ and $c_2 = (7,7)$ as the two initial cluster centers, and show how the division of points will be.

3. Assume that the seven points z_i mentioned in question 2 actually come from a one-dimensional distribution, i.e. without noise they would be placed along a line. Find the line that best fits the seven points, if you apply Principal Component Analysis (PCA).

Exercise 6 (1+2+3=6 points)

As long as we can match points between the two images of a stereo pair, the distances to points that we observe in the world can be determined. However, due to repetitive structures in the scene, matching can be quite difficult.

1. What is the epipolar constraint and how can it be used in stereo matching? Draw a picture!

2. Assume a pair of parallel cameras with their optical axes perpendicular to the baseline. What is the relation between the disparities and the depth? How do the epipolar lines look like? Where are the epipoles for this type of camera system?

3. Estimate the essential matrix between two consecutive images for a forward translating camera. What is the equation of the epipolar line for the point p=[x y 1] in this case?