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

EXAM

Image Analysis and Computer Vision, DD2423

Friday, 15^{th} of January 2015, 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. Using a drawing describe how the world gets projected to an image with a pin-hole camera model. Why isn't it possible to manufacture a real pin-hole cameras?
2. Why is it valuable to detect image discontinuities (edges), if you want to understand the 3D world through the study of images?
3. What benefits to you get by representing image and world points using homogeneous coordinates, instead of Cartesian coordinates?
4. What kind of information exists in the phase and magnitude of the Fourier transform of an image?
5. If you want to reduce the dimensionality of data using PCA, how can you tell how many dimensions you should use, for the errors introduced not to be too large?
6. What kind of imperfections does the intrinsic camera parameters try to compensate for?
7. Why do you usually use bilinear interpolation for image transformations?
8. A frequency space decomposition of a signal assumes that the signal is periodic, but that is not true for most images. Why can you still use Fourier transforms for images?
9. According to the Sampling Theorem, when can you assume to get aliasing in image sampling?
10. What properties do Gaussian filters possess that make them suitable for scale-space representation? Mention at least two relevant such properties.
11. How do you compute a second moment matrix that is used for corner detection?
12. What kind of information does the SIFT descriptor contain? Think of how it is computed.
13. Explain briefly how RANSAC works, if you for example like to detect a line.
14. Why are template based methods rarely used for object recognition?
15. In what sense is it usually harder to compute optical flow than binocular disparities?

Part B

Exercise 1 (4+3=7 points)

1. Assume you have a 2D square in 3D Euclidean space, with corners $x_1 = (0,0,0)^\top$, $x_2 = (1,0,0)^\top$, $x_3 = (1,1,0)^\top$ and $x_4 = (0,1,0)^\top$. This square is projected onto two different cameras, camera A and camera B, that have the projection matrices

$$P_A = \begin{pmatrix} 1 & 2 & 1 & 0 \\ 3 & 1 & 2 & 3 \\ 1 & 0 & 2 & 1 \end{pmatrix}, \; P_B = \begin{pmatrix} 1 & 2 & 1 & 0 \\ 3 & 1 & 2 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

   Apply the two projections and draw the square in the 2D image plane for each of the two cameras. How can you tell from the projections of the square, which camera is affine and which is perspective? How can you draw the same conclusion by directly studying the projection matrices?

2. In an image you observe a polygon with corners at $x_1 = (0,1)^\top$,$x_2 = (4,0)^\top$,$x_3 = (4,2)^\top$ and $x_4 = (0,3)^\top$. You believe the polygon is actually a square in 3D and want to transform the polygon into a square. Find an image transformation T that transforms the corners into a new set of corners with coordinates $y_1 = (0,0)^\top$,$y_2 = (1,0)^\top$,$y_3 = (1,1)^\top$ and $y_4 = (0,1)^\top$.

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

1. Computing the average of neighbouring points in an image can be seen as applying a box filter to the image. What is the Fourier Transform of the box filter $f(x)$ defined below?

$$f(x) = \begin{cases} 1, & -\frac{1}{2} \leq x \leq \frac{1}{2} \\ 0, & otherwise \end{cases}$$

   Why does averaging often lead to undesirable effects in the resulting images, unlike for example what happens when you apply Gaussian filters?

2. How do you compute a convolution in the general case? What is the convolution of two box filters, that is $f(x) * f(x)$ given the definition of $f(x)$ above?

3. A box filter has the benefit of being easy to compute, since it's just a summation of neighbouring pixels, but there are other filters that are almost as simple. Compute the Fourier Transform of the triangular shaped filter $g(x)$ defined below.

$$g(x) = \begin{cases} 1 - |x|, & -1 \leq x \leq 1 \\ 0, & otherwise \end{cases}$$

   If a box filter has undesirable properties, does a triangular shaped filter also have problems? If any, what would those problems be?

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

1. Assume the histogram of a grey-level image is given by $p(z) = 2z^4 - 3z^2 + 8/5,\ z \in [0,1]$. Determine a transformation $z' = T(z)$, such that the histogram of the resulting image is $p(z') = 1,\ z' \in [0,1]$. For which grey-level values z does the transformation increase the contrasts and for which do they decrease?

2. Propose an isotropic sharpening filter based on differential operators that leaves uniform image areas unchanged. What does isotropic mean in this case? In what sense is sharping different from contrast enhancement, such as the histogram equalization above?

3. On the 1D image

$$f(x) = [\ 2\ 5\ 4\ 3\ 6\ 10\ 12\ 11\ 14\ 10\ ]$$

apply three different filters of size 3; a) a mean filter, b) a binomial filter and c) a median filter. Assume that pixels outside the left and right borders are equal to zero. Apply the filters individually, not in sequence.

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

1. Assume you have detected an image region, as indicated by the 1s in the figure below. Compute the zeroth and first order image moments of the region, given a suitable choice of coordinate axes.

0	0	0	1	1
0	0	1	1	1
0	0	1	1	0
1	1	1	0	0
0	1	1	0	0

2. Obviously, the region is close to elliptic in shape and can be represented by a covariance matrix. Using image moments of higher order and the choice of coordinate axes you defined before, compute the covariance matrix. Also explain how you can use the matrix to compute the dominating orientation of the ellipse (without actually doing it on the matrix you just computed).

3. Using the morphological structural element

0	1	0
1	1	1
0	1	0

compute an *opening* operation on the image region above. Assume that pixels outside the window are always set to 0, during all steps of the operation.

Exercise 5 (2+3+4=9 points)

1. What are the similarities and differences between K-means clustering and fitting of Gaussian Mixture models with points representing pixels in an image?

2. Methods using active contours (or snakes) are based on energy minimization. How do such methods represent a contour and what terms does the energy formulation typically consist of? Shortly describe at least two such energy terms. How do you go about to minimize the energy in practice, in order to get the final contour?

3. With the density used for Mean Shift segmentation defined as

$$f(x) = \frac{1}{N}\sum_{i=1}^{N} k(|x - x_i|^2),$$

where $k(|x|^2)$ is some continuous density function, derive the update function that is used to maximize the density. Also explain how the result of using this update function can be used to get a segmentation of an image.

Exercise 6 (3+3+3=9 points)

1. Assume you have a translating (but not rotating) robot moving around in a world and that you track two points, A and B, with a camera that has the focal length $f = 1$. At time t the points are located at image points $\mathbf{p}_A^t = (x_A^t, y_A^t, f)^\top$ and $\mathbf{p}_B^t = (x_B^t, y_B^t, f)^\top$ in homogeneous coordinates respectively. Show that using image positions from two points in time, $t = 1$ and $t = 2$, you can compute the translational direction $\mathbf{t} = (T_x, T_y, T_z)^\top$ in 3D space.

2. Using a stereo system consisting of two cameras, with focal lengths $f = 1000$ pixels separated by a baseline $b = 10$ *cm*, you measure the binocular disparity d of an observed 3D point to determine its depth Z. If you want to have an accuracy in depth of 1 *mm* on the depth $Z = 1$ *m*, how accurately do you need to be able to measure the disparity?

3. For two parallel cameras with a horizontal baseline and vertical y-axes the essential matrix has a particularly simple form. What is the relative rotation and translation between the cameras and what is the corresponding essential matrix?