Numerisk analys och datalogi, KTH

Mårten Björkman

EXAM

Bildbehandling och datorseende 2D1421
Friday, March 18th 2005, 8.00-13.00

Allowed material: Calculator, mathematics handbook (e.g. Beta) and a hand-written (not copied) sheet of paper in A4 format with your own personal notes. These notes have to be handed in together with your answers and will be returned after answers have been corrected.
Language: Answers can be given in either English or Swedish.
General: The examination consists of **six** exercises that can give at most 50 credits. To pass the examination you need about half of all credits. The bonus credits (at most 5) will be added to the total sum of your credits, given that you passed the laboratory exercises on time during the course of this year. The results will be announced within three weeks.
Course evaluation: We would appreciate if you fill in the evaluation form available on the website.

The grades were set as follows: 3 (credits $\geq$ 28), 4 (credits $\geq$ 33) and 5 (credits $\geq$ 42).

Exercise 1 (5*2=10 credits)

Answer *five* out of the following *seven* short questions. If you respond to more than *five* questions, only the first *five* will be corrected and counted.

(a) What is the difference between “perspective” and (scaled) “orthographic” projection?

Perspective projections are necessary when the camera is located close to the object observed, whereas orthographic projection is feasible if objects are located far away and the field of view is small. Then the optical rays will be close to parallel and not converge into a single point. Furthermore, for orthogonal projections relative depth variations will go unnoticed and parallel lines in the world will be projected into parallel lines in the image.

(b) Explain the “cyclopean eye” and why we are usually interested in studying it.

A Cyclopean eye is an imaginary eye located inbetween the left and right cameras of the stereo set. Directions in the scene are usually related to this eye. It can further be used to represent world points the system is able to fuse and visualise cases of crossed or uncrossed diplopia.

(c) How is a signal in the Fourier domain affected by a translation and scaling in the spatial domain?

In the Fourier domain a signal is described by complex values. In cases of translation the phase will change, while the magnitude is kept constant. Furthermore, if the spatial domain representation is compressed, the corresponding Fourier domain representation will expand correspondingly.

(d) What are the three steps of the “Canny” edge detector?

Gradients are initially computed using derivate of Gaussian filters. Peaks in gradient magnitude in the gradient direction are then found using Non-Maximum suppression. Finally, Hysteris Thresholding is applied to extend edges from points of high gradient magnitude to nearby points of somewhat lower magnitude.

(e) Shortly describe how “K-means” works and what its purpose is.

The purpose of K-means is to automatically find clusters in some feature space, given a large number of feature points. First, a set of centres are randomly spread across the space. Each feature point is then associated to the closest one of these centres. Finally, the position is each centre is updated by computing the mean of all feature points associated to it. This process is repeated until convergence.

(f) What is a “Fundamental matrix” and what is it used for?

The Fundamental matrix is a 9×9 *matrix that describes the positional and orientational difference between two cameras in a stereo set. Unlike the Essential matrix it also includes the intrinsic camera parameters. From the fundamental matrix the epipolar line associated to a particular point can be found in the opposing camera image. Stereo matches can be found along this line, thus simplifying stereo matching to 1D searches.*

(g) Mention one common method for “lossy” compression and one for “loss-less” compression.

For example, quantization is a common method for lossy compression, whereas Huffman coding is used for loss-less compression.

Exercise 2 (1+1+2+2+3=9 credits)

Noise is a common problem is computer vision and careful steps have to be taken to reduce the influence of this noise. Typically, you blur images slightly before you do any further processing. Often we assume the noise to be Gaussian.

(a) For typical images, in which part of the frequency spectrum is image noise most severe? For which frequencies will on the other hand image data usually dominate the noise?

For high frequencies the noise usually dominates the signal and is thus most severe. The opposite is true for low frequencies. Here noise is typically not a serious problem.

(b) Blurring is not only used to reduce noise. What is the purpose of blurring in scale-space theory?

In scale-space theory images are blurred such that high frequency components are successively suppressed. For example, for each new octave in a Gaussian pyramid the frequency contents is reduced by half. By detecting features at different scales, the projected size of features can be determined.

(c) Assume we reduce the influence of noise by applying a 2D separable binomial filter. What does it mean that a filter is separable and why is separability a preferable feature?

A $n \times n$*-point 2D separable filter can be separated into two n-point 1D filters, one filter for each dimension. Filtering will thus become more efficient.*

(d) Given the following portion of an image, apply the 1D binomial kernel $h = \frac{1}{4}[1, 2, 1]$. The results should be 7 pixels in size, i.e. we disregard the boundaries.

17	18	21	15	13	8	2	1	3

After low-pass filtering we get the following response: $\frac{1}{4}[74, 75, 64, 49, 31, 13, 7]$

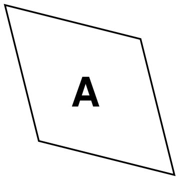
(e) After blurring with the binomial filter in (d) apply a centered differential kernel to the result. Also apply the same differentiation kernel to the original image portion without blurring. How many local maxima (in magnitude) will you get in each case? You only need to consider the center 5 pixels of the results.

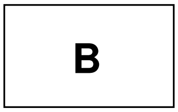
After blurring and differentiation we get $-\frac{1}{8}[10,26,33,36,24]$, *or* $-\frac{1}{2}[3,8,7,11,7]$ *without blurring. Thus we get two maxima if we do not blur before differentiation, instead of just one.*

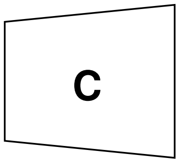
Exercise 3 (2+2+2+2=8 credits)

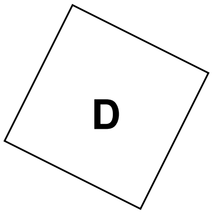
The projection of a two-dimensional surface can be described by a transformation, from the local coordinate system of the surface to the coordinate system of the image plane.

(a) Assume we have a square located somewhere in a 3D scene. Depending on the viewing conditions, the projection of this square might look like the examples below.









Which of these examples can **not** be described by an “affine transformation”? How can you directly see whether an affine transformation is possible?

The third example, **C**, *can not be described by an affine transform, since in this case parallel lines are not projected into parallel lines.*

(b) For at least two of the remaining three examples above, describe how the corresponding affine transformations look like, if we assume that the edges of the square are parallel to the axes of the surface coordinate system and the origin is in the centre. More explicitly, without determining any specific values, what are the relations between the parameters a_{ij} in the affine transformation

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

for each of these two cases?

A *has three degrees of freedom (scaling in a direction) and only* $a_{12} = a_{21}$ *has to be satisfied.* **B** *has two degrees of freedom (scaling), with the condition that* $a_{12} = a_{21} = 0$. *Finally,* **D** *has only one degree of freedom (rotation), satisfying* $a_{11}a_{22} - a_{12}a_{21} = 1$, $a_{11} = a_{22}$ *and* $a_{12} = -a_{21}$.

(c) Assume we have a moving camera and look at two images taken at different points in time. If we have a rigid scene, we can express the “optical flow” for each points in the scene, given that we have a set of motion parameters and the depth of each point. For which motions can we compute the optical flow without knowing these depths? What is the relation between the motion parameters and the optical flow for points located at infinity?

For a rigid scene, the optical flow depends on two components, translation and rotation. The translational component is inversely proportional to depth, while the rotational one is independent on depth. Thus for rotations we don't have to know the depth to compute the flow. For points at infinity, the translational component doesn't matter.

(d) The so called Optical Flow (or Brightness Constancy) constraint is given by $L_x u + L_y v + L_t = 0$. How can this constraint be derived? How does the "aperture problem" affect the application of the constraint?

If we have a moving point of constant brightness in the scene, we have the following equation: $L(x(t), y(t), t) = L_0$. If we take the t-derivative of this equation we get: $L_x u + L_y v + L_t = 0$, where $u = x'(t)$ and $v = y'(t)$ represent the optical flow. To determine the optical flow, we thus need to find two parameters, but we only have one equation. To solve this "aperture problem" we have to integrate over some window around the image point.

Exercise 4 (3+3+3=9 credits)

Histogram equalization can be used to transform the histogram of an image to the histogram of another. This is beneficial for operations like stereo matching in which image data is matched between two different cameras, that might have very different characteristics.

(a) Assume that the histogram in an image is given by $p(z) = 5z^2 - 3z + 5/6,\ z \in [0,1]$. Determine a transformation $z' = T(z)$, such that the histogram in the new image is $p'(z') = 1,\ z' \in [0,1]$. For which values of z does the transformation result in stretching?

Since the number of samples from an interval Δz to the corresponding interval $\Delta z'$ must be equal, $p(z)dz = p'(z')dz' \Rightarrow T'(z) = dz'/dz = p(z)/p'(z') = 5z^2 - 3z + 5/6 \Rightarrow T(z) = (10z^3 - 9z^2 + 5z)/6$. Stretching occurs for those z that satisfy $T'(z) > 1 \Rightarrow 5z^2 - 3z + 5/6 > 1 \Rightarrow 5z^2 - 3z - 1/6 > 0$. Let us solve the equation $z^2 - 3/5z - 1/30 = 0 \Rightarrow z_1 = 3/10 + (9/100 + 1/30)^{\frac{1}{2}} \approx 0.6512$. In conclusion, stretching occurs if $z > 0.6512$.

(b) Further, assume we wish to transform an image with the histogram $p(z) = (5 - 2z)/4,\ z \in [0,1]$ into a new histogram $p'(z') = 2z',\ z' \in [0,1]$. What is the transformation in this case?

Similar to previous exercise $p(z)dz = p'(z')dz' \Rightarrow T'(z) = dz'/dz = p(z)/p'(z') = (5-2z)/4/2z' \Rightarrow 2z'T'(z) = (5-2z)/4 = 2T(z)T'(z) = \delta T^2(z)/\delta z \Rightarrow T^2(z) = 5z/4 - z^2/4 \Rightarrow T(z) = (5z - z^2)^{\frac{1}{2}}/2$.

(c) Let's say we want to use thresholding to segment a dark cup from a somewhat lighter table. How should the threshold be chosen? Unfortunately, no single threshold results in a satisfactory segmentation. Could you mention (and explain) any other method that could lead to better results?

We could create a histogram and search for a local minimum between two distinct peaks, and use this minimum as a threshold. A segmentation of the cup could then hopefully be found as the largest connected component of points below the threshold. If this doesn't work we could instead use a region growing method that starts in a local minimum in the image and gradually adds neighbouring points that have similar luminance. If the change in luminance is gradual (and we are lucky) we will end up with a connected component representing the cup.

Exercise 5 (2+3+1=6 credits)

In object recognition and classification we often like to go from a high-dimensional image space to a lower-dimensional feature space, in which comparisons between new images and previously stored models can be made.

(a) In order for comparisons to be successful, they have to be made invariant to a number of changes that frequently appear in real images. Mention at least three such invariances.

Comparisons have to be invariant to differences in lighting conditions, scale, pose, deformations, clutter, cameras, etc.

(b) One way of reducing the dimensionality is by applying Principle Component Analysis (PCA). From a set of training images $\{\mathbf{X_i}\}$ we compute a number of eigenimages (eigenspaces, eigenfaces) $\{\mathbf{U_k}\}$, that represent a basis with which every (training and test) image can be described.

- Given the eigenimages, each new image $\mathbf{X}$ can be represented by a set of coefficients $\{c_k\}$. How do you compute these coefficients?
- What will the number of coefficients you need depend on?
- Once you have the coefficients for a particular image, how do you compare these to those of other images (that are usually stored in a database)?

The coefficients are computed by projection onto each $\mathbf{U_k}$*, that is* $c_k =< \mathbf{U_k}, \mathbf{X} >$*. The number of coefficients one needs depends on the number of classes and the relative spread of these classes. Comparisons can be made by simply calculating the distance (using the 2-norm) to each training example and assigning the class to the one of the closest example.*

(c) If you compare PCA representations the way you described in (b), will the invariances you mentioned in (a) be satisfied?

Usually no.

Exercise 6 (3+2+3=8 credits)

Using Taylor expansion

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2!} f''(x) + \frac{h^3}{3!} f'''(x) + O(h^4),$$

where h is the distance between two neighbouring pixels, it can be shown that a result of a differentiation kernel $d_x = \frac{1}{2}[-1, 0, 1]$ applied to an image $f(x)$ is

$$d_x * f(x) = \frac{1}{2}(f(x+h) - f(x-h)) = hf'(x) + \frac{h^3}{6} f'''(x) + O(h^4)$$

Thus, a derivative approximation based on d_x will result in a dominating error proportional to h^3 (if we assume that $h \ll 1$). Unfortunately, we cannot find a better approximation of a first order derivative than d_x, if the kernel is limited to only 3 pixels in width.

(a) Using the same method based on Taylor expansion, determine how well $d_{xx} = [1, -2, 1]$ approximates a second order derivative.

Here we need to consider an additional term of the Taylor expansion, which then leads to $d_{xx} * f(x) = f(x+h) - 2f(x) + f(x-h) = (f(x) + hf'(x) + \frac{h^2}{2!}f''(x) + \frac{h^3}{3!}f'''(x) + \frac{h^4}{4!}f^{(4)}(x)) - 2f(x) + (f(x) - hf'(x) + \frac{h^2}{2!}f''(x) - \frac{h^3}{3!}f'''(x) + \frac{h^4}{4!}f^{(4)}(x)) + O(h^5) = h^2 f''(x) + \frac{h^4}{12}f^{(4)}(x)) + O(h^5)$. The dominating error term is then $\frac{h^4}{12}f^{(4)}(x)$, or $\frac{h^2}{12}f^{(4)}(x)$ if we divide by h^2.

(b) One might wonder whether a better approximation of a first order derivative is possible, if we increase the length of the kernel. Without knowing the best kernel, can you tell whether it ought to be symmetric or anti-symmetric? How many parameters do we have to determine if we search for a kernel of length 5?

If a signal $f(x)$ is symmetric around $x = 0$, the derivative $f'(x)$ will be anti-symmetric around $x = 0$. Thus the differentiation kernel ought to be anti-symmetric and the filter will be of the type $g = [+a, +b, 0, -b, -a]$, i.e. there are two unknown parameters, a and b.

(c) Find a differentiation kernel of length 5, such that also the error term proportional to h^3 disappears. Which error term will now dominate?

First we conclude that

$$[1,0,-1] * f(x) = f(x+h) - f(x-h) = 2hf'(x) + \frac{h^3}{3}f'''(x) + \frac{h^5}{60}f^{(5)}(x) + O(h^7) \text{ and}$$

$$[1,0,0,0,-1] * f(x) = f(x+2h) - f(x-2h) = 4hf'(x) + \frac{2^3h^3}{3}f'''(x) + \frac{2^5h^5}{60}f^{(5)}(x) + O(h^7).$$

We see that we get rid of the h^3 term if we select $b = -8a$. For the filter $[-1, 8, 0, -8, 1]$ we get

$$[-1,8,0,-8,1] * f(x) = 12hf'(x) - \frac{2h^5}{5}f^{(5)}(x) + O(h^7).$$

We end up with a filter $\frac{1}{12}[-1, 8, 0, -8, 1]$ and a dominating error term $\frac{h^5}{30}f^{(5)}(x)$.

Good luck!