School of Computer Science and Communication, KTH
Danica Kragic

EXAM SOLUTIONS

Image Processing and Computer Vision Course 2D1421
Monday, 13th of March 2006, 14.00–19.00

Grade table

0-25 U
26-35 3
36-45 4
46-50 5

Exercise 1 (5*2=10 credits)

(1) In what cases is spectral filtering more appropriate than spatial one? Give two examples.

1) If the noise is periodic, 2) If we want to filter the image using large kernels.

(2) What is an "ideal low-pass filter"? Is this filter suitable to use in terms of image processing? If yes, give an example of its application. If no, explain why.

The ideal low-pass filter is a mathematically idealized version of a smoothing filter. In a frequency-domain representation of the image, all frequencies above a threshold F are discarded, i.e. this filter passes low frequencies so image becomes blurred. This method of smoothing tends to create images with "ringing" at sharp boundaries in the picture which is its drawback. The observation that the application of the low-pass filter is equivalent to the convolution of the image with the sinc function provides an explanation for this phenomenon. This filter cannot be implemented in hardware. Ideal low pass filter in 2D:

$$\hat{h}(v) = rect(\frac{\|v\|}{2v_c}) \qquad \text{where} \qquad \|v\| = \sqrt{v_1^2 + v_2^2}$$

- v_c cut-off frequency
- Impulse response

$$h(x) = 2\pi v_c^2 \frac{J_1(2\pi v_c \|x\|)}{2\pi v_c \|x\|}$$

- J_1 - first order Bessel function

(3) Describe basics and draw figures of perspective and orthographic camera models. Given a set of parallel lines in 3D - explain what is the difference in their image projections for both models.

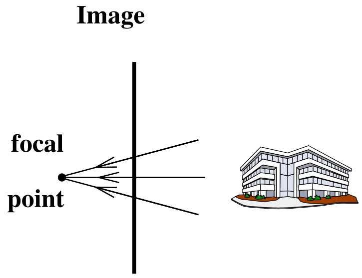
Perspective projection

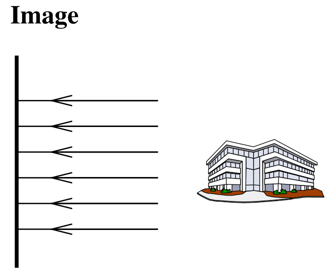
$$\frac{x}{f} = \frac{X}{Z}, \qquad \frac{y}{f} = \frac{Y}{Z}$$

Orthographic projection

$$x = X, \qquad y = Y$$

For the orthographic camera model, parallel lines remain parallel and for a perspective camera model the lines intersect in a vanishing point.





Perspective projection **Orthographic projection**

(4) What are the common image points distance measures? Give at least two examples.

Common distance measures:

– Euclidean distance

$$d(p,q) = \sqrt{(x-u)^2 + (y-v)^2}$$

– City block distance

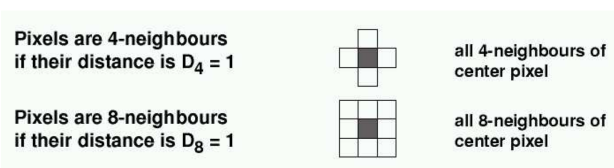
$$d(p,q) = |\, x-u \,| + |\, y-v \,|$$

– Chessboard distance

$$d(p,q) = \max(|\, x-u \,|, |\, y-v \,|)$$

(5) Explain the neighborhood concept is terms of images. What are the problems related to different concepts?

- 4-connected, 8-connected, hexagonal grid

- 

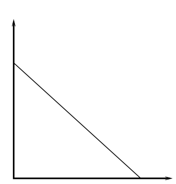
– 8-connectivity ⇒ outer boundary 4-connected
– 4-connectivity ⇒ outer boundary 8-connected
– if 4(8)-connectivity used for foreground ⇒ 8(4)-connectivity used for background
– Hexagonal grid - same connectivity concept for foreground and background

(6) Describe what scaling, rotation and translation of an image in the spatial domain result in in the spectral domain.

Scaling is inverse proportional. Fourie transform of an image rotates as the image itself. Translation affects only phase while power spectrum is translation invariant.

(7) Explain terms sampling and quantization.
Sampling: selection of a discrete grid to represent an image; Spatial deiscretization of an image
Quantization: Mapping of the brightness into a numerical value; Assigning a physical measurement to one of a discrete set of points in a range.

(8) What is "contrast reversal" in terms of grey-level transformations? Draw the corresponding linear transformation.
Contrast reversal - basically inverting the grey-level values: $f(x) = g(1-x)$.



(9) What is “region growing”? When is it commonly used?

Region growing - homogeneous regions grows in size by including “similar” neighboring pixels, the final result does not necessarily need to cover the entire image
First, seed regions have to be extracted, and these seed regions are iteratively grown at their borders by accepting new pixels being consistent with the pixels already being contained in the region. After each iteration, the homogeneity value of the region has to be re-calculated using also the new pixels. The results of region growing heavily depend of a proper selection of seed points

Usage: divide image pixels into a set of classes based on local properties and features (classification), matching in stereo, split-and-merge transform etc.

(10) If a camera views a planar square, what characteristics in the image should be approximately true in order for an affine camera model to be appropriate to use?

Corresponding line pairs should remain paralllel in the image.

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

(a) Given the following perspective camera matrix:

$$P = \begin{bmatrix} 5 & -14 & 2 & 17 \\ -10 & -5 & -10 & 50 \\ 10 & 2 & -11 & 19 \end{bmatrix}$$

and a 3D point in homogeneous coordinates $X = [0\ 4\ 4\ 2]^T$

- What are the Cartesian coordinates of the point X in 3D ?

$$(0/2,\ 4/2,\ 4/2) = (0,\ 2,\ 2)$$

- What are the Cartesian image coordinates, (u,v) of the projection of X ?

$$P = \begin{bmatrix} 5 & -14 & 2 & 17 \\ -10 & -5 & -10 & 50 \\ 10 & 2 & -11 & 19 \end{bmatrix} [0\ 4\ 4\ 2]^T = [-14\ 40\ 2]^T$$

$$(u,v) = (-14/2,\ 40/2) = (-7,20)$$

(b) What is a vanishing point and under what conditions will a vanishing point for a perspectice camera model be at infinity ? Does a vanishing point exist for an affine camera model and why? Motivate your answer.

a) A point in the image that represents an intersection of parallel lines. The vanishing point will be at infinity in case when lines are in the plane parallel to the image plane. b) No. Affine camera preserves parallel lines.

(c) What is meant by “camera calibration”? Present it for a perspective camera model and define how many parameters have to be estimated. *Camera calibration is the process of estimating camera intrinsic and extrinsic parameters. For a perspective camera model with no distortion, intrinsic parameters are focal length, camera center and pixels sizes and teh angle between the image axis (5 parameters). Extrinsic camera parameters represent the pose (position and orientation) of the camera center and are modeled with 6 parameters (3 for the rotation and 3 for the translation). Total 11 parameters have to be estimated.*

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

(a) A discrete approximation of the second derivative $\frac{\partial^2 I}{\partial x^2}$ can be obtained by convolving an image $I(x,y)$ with the kernel

$$1\ \ -2\ \ 1$$

Use the given kernel to derive a 3×3 kernel that can be used to compute a discrete approximation to the 2D Laplacian. Apply the Laplacian kernel to a center pixel of the following image (Show all the convolution steps!):

$$\begin{array}{ccc} 3 & 2 & 1 \\ 6 & 5 & 4 \\ 9 & 8 & 7 \end{array}$$

Laplacian:

$$\begin{array}{ccc} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{array}$$

Convolution gives: $2+6+4+8-20=0$

(b) Why is it important to convolve an image with a Gaussian before convolving with a Laplacian? Motivate your answer by relating to how a Laplacian filter is defined.

Since Laplacian is a second order filter (second derivative) it amplifies noise. We first have to lowpass an image to attenuate some of the noise.

(c) Let us assume doing the following two operations: 1) I first convolve an image with a Gaussian and then take the Laplacian, $\nabla^2(G*I))$, and 2) I first apply the Laplacian to the Gaussian and then convolve the image, $(\nabla^2\ G)*I$. Will the results be the same? If yes - why? If no - why?

It is the same due to the linearity property.

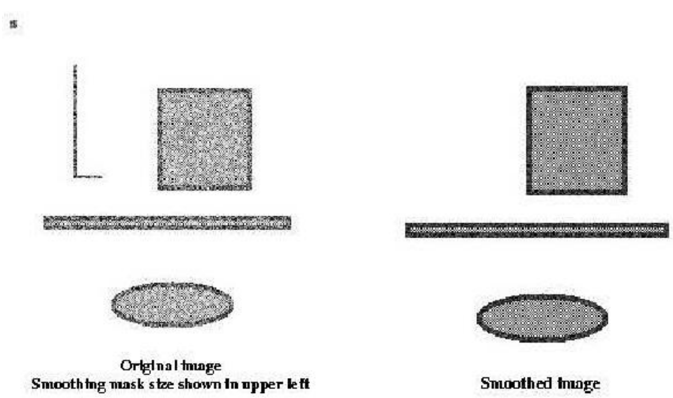
(d) Can a Laplacian of a Gaussian (LoG) operator be approximated in any other way than using a sum of second order derivatives? If yes, how?

Yes, as a difference of Gaussians (DoG filter).

(e) Given the image below before (left) and after (right) a smoothing filter was applied. The size of the filter is shown as a small square in the upper-left corner in the image (as you can see, its size is rather small compared to the image size). In your opinion, which one of the following filter types most likely produced the image on the right:

1) mean (unweighted averaging) filter,
2) median filter, or
3) Gaussian filter.

Motivate your answer.



Median filter since the line on the left of the image dissapperas - with any other filter it would remain and be extended.

(f) Give an example of a mean (unweighted averaging) filter. Is a mean filter in general separable? Why do we prefer separable filters?

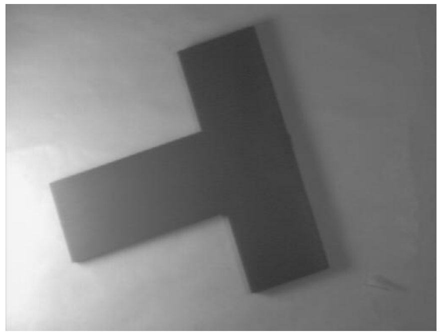
Let us take the simplest example:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

Since unweighted averaging is assumed, all elements in the matrix are equal. It is always separable since its rank is equal to one. Separable filters are prefered since they can be implemented as repeated convolutions with one-dimensional kernels.

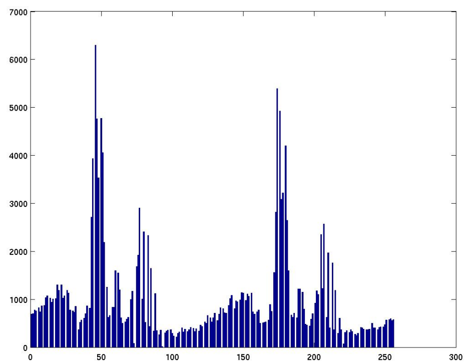
Exercise 4 (2+2+2=6 credits)

You are given an image of an object on a background containing a strong illumination gradient. You are supposed to segment the object from the background in order to estimate its moment descriptors.



(a) Sketch the histogram of the image and explain what are the problems related to choosing a suitable threshold.

The histogram is not clearly bimodal - difficult to obtain local minimum (suitable threshold).

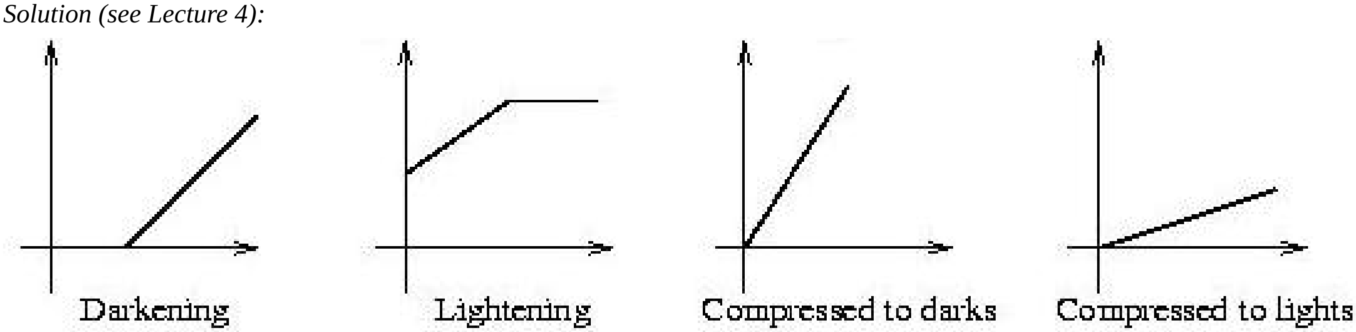


(b) Propose a suitable methodology that could be used to perform successful segmentation. *a) Local adaptive thresholding selects an individual threshold for each pixel based on the range of intensity values in its local neighbourhood. This allows for thresholding of an image whose global intensity histogram doesn't contain distinctive peaks. b) estimating the illumination gradient and substracting it from the image c) estimating gradient close to the object boundaries and using this to perform adaptive thresholding*

(c) For the images below, pair them with the correct expressions:

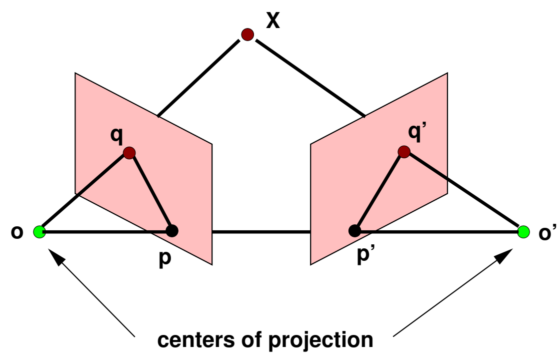
1. Darkening
2. Lightening
3. Compressed to darks
4. Compressed to lights

Solution (see Lecture 4):



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

(a) What is the epipolar constraint and how can it be used in stereo matching?
Represents geometry of two cameras, reduces a correspondance problem to 1D search along an epipolar line. A point in one view "generates" an epipolar line in the other view. The corresponding point lies on this line. Epipolar geometry is a result of coplanarity between camera centers and a world point - all of them lie in the epipolar plane.



(b) Assume a pair of parallel cameras with their optical axes perpendicular to the baseline. How do the epipolar lines look like? Where are the epipoles for this type of camera system?
The epipolar lines are parallel to each other (also, in the direction of th image y axis). If the optical centers of the camera are on the same height, the corresponding epipolar lines will be on the same height - one line for both images. The epipoles are in the infinity.

(c) 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 general

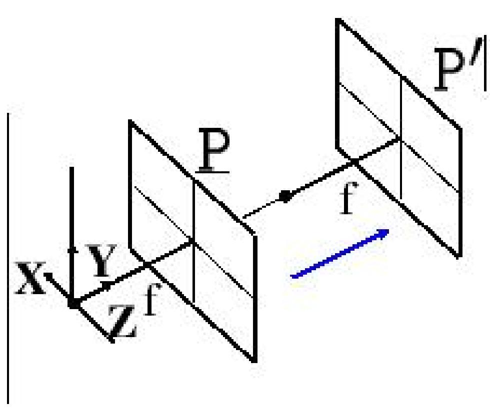
$$E = t_S R$$

where t_S is a skew-symmetric matrix related to translation vector. For a forward translating camera (see figure), we have

$$R = I, \text{ and } t_S = \begin{bmatrix} 0 & -t_Z & 0 \\ t_Z & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

therefore

$$E = \begin{bmatrix} 0 & -t_Z & 0 \\ t_Z & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$



From $l = E\ p$ the epipolar line for a point $p = [x\ \ y\ \ 1]$ is

$$l = \begin{bmatrix} 0 & -t_Z & 0 \\ t_Z & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} -t_Z\, y \\ t_Z\, x \\ 0 \end{bmatrix}$$

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

Given the two binary images below.

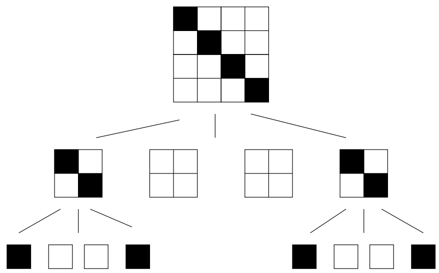
(a) Quad-tree representation (image on the left).

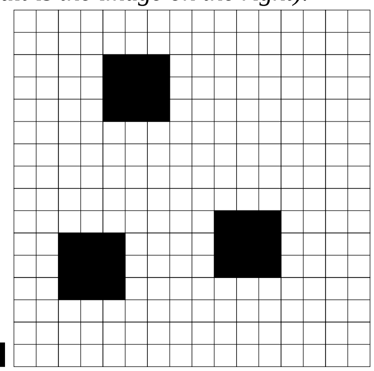
(b) Generate a one-dimensional run-length code with pixels ordered from top to bottom, left to right.

27,1,14,3,4,1,7,.....

(c) Perform a morphological opening using a kernel of size 3×3 with all elements equal to one (for the image on the right).

Opening = erosion + dilation (result is the image on the right).





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

(a) Consider two images taken by a camera that translates right along camera's x-axis direction while viewing a static scene. What does the motion field look like for this motion? Consider the left input image - where is the epipole located? Draw an image of the motion field. *If we consider x to be the horizontal axis pointing to the right, the motion field vector will be parallel to it and pointing left. The epipole is on the right, placed in the infinity.*

(b) What is an optical flow constraint equation? Write the equation and explain what each term represents. *Denote the intensity by $I(x,y,t)$, and assume that it is unchanging over time:*

$$I_x u + I_y v + I_t = 0$$

where the spatial and time derivatives of I are denoted by subscripts x, y, t, and u and v are the x and y components of the optical flow vector. The last equation is called the optical flow constraint equation.

(c) Imagine a robot moving in the environment with a constant velocity. We assume that the floor is perfectly planar and infinitely large. The robot has a camera mounted on the top of the base on height *h*, with the optical axis of the camera parallel to the floor. The robot can both rotate and translate on the floor.

Given that you can measure the robot's translational and rotational velocity, how would you estimate optical flow for points in the camera image? Express the optical flow with suitable equations. How do the equations change if the robot only translates?

Let us assume a world coordinate system with Z-axis horizontal pointing ahead, X-axis horizontal pointing on the right, and Y-axis pointing upward.

A robot that moves on a planar floor has 3 degrees of freedom, translational velocity U in X-direction, translational velocity V in Z-direction and rotational velocity ω around the Y-axis. Motion of the points in the world relative to the robot's coordinate system can be written:

$$\begin{pmatrix} \dot{X} \\ \dot{Y} \\ \dot{Z} \end{pmatrix} = - \begin{pmatrix} u \\ 0 \\ v \end{pmatrix} - \begin{pmatrix} 0 \\ \omega \\ 0 \end{pmatrix} \times \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = - \begin{pmatrix} U + \omega Z \\ 0 \\ V - \omega X \end{pmatrix}$$

If the camera is mounted so thet it optical axis is parallel to the Z-axis of the world coordinate system, its x-axel parallel to the X-axis and its y-axel parallel to the Y-axis, the equation for the perspective camera models can be written as:

$$\frac{x}{f} = \frac{X}{Z}, \qquad \frac{y}{f} = \frac{Y}{Z}$$

Taking the derivatives and combining it with the expressions for $\dot{X}$, $\dot{Y}$ and $\dot{Z}$ we have:

$$\frac{\dot{x}}{f} = \frac{Z\dot{X} - X\dot{Z}}{Z^2} = -\frac{Z(U + \omega Z) - X(V - \omega X)}{Z^2}$$
$$\frac{\dot{y}}{f} = \frac{Z\dot{Y} - Y\dot{Z}}{Z^2} = -\frac{0 - Y(V - \omega X)}{Z^2}$$

and together with the equations for the perspective camera it gives:

$$\frac{\dot{x}}{f} = -\frac{U}{Z} - \omega + \frac{x}{f}\frac{V}{Z} - \omega\frac{x^2}{f^2}$$
$$\frac{\dot{y}}{f} = \frac{y}{f}\frac{V}{Z} - \omega\frac{xy}{f^2}$$

This are general equations for the motion field of a robot moving on a planar surface. If tehre is a pure translation, we have:

$$\frac{\dot{x}}{f} = -\frac{U}{Z} + \frac{x}{f}\frac{V}{Z}$$
$$\frac{\dot{y}}{f} = \frac{y}{f}\frac{V}{Z}$$

where points in the image plane are represented by (x, y) and points in the world by $(X, -h, Z)$. From the later one and expression for the perspective model we have:

$$\frac{y}{f} = -\frac{h}{Z}$$

which gives the following relation beween y and Z (since $Z > 0$ is valid only when $y < 0$)

$$\frac{1}{Z} = -\frac{y}{fh}$$

After the combination with the above equations for motion, we get that the motion in the image $(\dot{x}, \dot{y})$ in points (x, y) with $y < 0$ is given by

$$\frac{\dot{x}}{f} = \frac{U}{h}\frac{y}{f} - \frac{V}{h}\frac{xy}{f^2}$$
$$\frac{\dot{y}}{f} = -\frac{V}{h}\frac{y^2}{f^2}$$

This means that motion field in the image is represented by a second order polinomial.