Computer Science and Communication, KTH
Danica Kragic

Examination

Computer Vision course 2D1421

Wednesday, 07th of June 2006, 8.00–13.00

Allowed helping material: Calculator, mathematics handbook Beta and a hand-written (not copied) sheet of paper in A4 format with your personal notes. Your notes have to be handed in together with your answers and will be returned to you after the answers are corrected.
Language: The answers can be given either in English or Swedish.
General: The examination consists of **six** exercises that can give at most 50 credits. To pass the examination 25 credits is required. The bonus credits will be added to the total sum of your credits given that you passed the laboratory exercises on time during this year (these can give you at most 5 points). The results will be announced within three weeks.

Exercise 1 (5*2=10 credits)

Answer only 5 of the following questions. If you answer more than 5, only the first 5 answers will be corrected and the rest ignored.

(1) What is meant by "image warping"? Explain some of the problems related to warping.

(2) What is meant by a "pyramid representation". Give examples of different such representations.

(3) What is "fovea" and what is its function?

(4) Explain the terms "radiance" and "irradiance".

(5) What is the difference between "optical flow" and "motion field"?

(6) What are "epipolar lines" and why are these important in stereo matching?

(7) Mention at least two differences between "cones" and "rods".

Exercise 2 (3+4=7 credits)

During the course, it was mentioned that there are several camera models that can be used dependent on the scene structure, camera type, etc. We have also said that a perspective camera model is the most commonly used one and there are several other simpler models that are sometimes used.

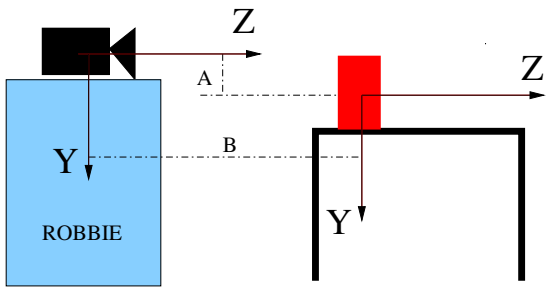
(a) What is the difference between an affine and a perspective camera model? Provide suitable equations in terms of intrinsic and extrinsic parameters.

(b) Assume a robot vehicle equipped with a camera on the top. The robot is placed in a living room in front of the dinner table on which there is a box of cereals. The camera image is a rectangle of height 500 pixels and width 600 pixels. The focal length of the camera is 690 pixels.

There is a coordinate system attached to the camera: Z and Y axes are pointing as shown and X axis is pointing so to make a right-hand coordinate system. There is also a coordinate system attached to the box and its coordinate axis are pointing in the same way as those of the camera's coordinate system. The coordinate system of the box in centered in the center of the mass of the box. The size of the box is the following: width=10cm, height=15cm and depth=5cm.

The distances between the camera and the object coordinate systems are A=15cm and B=1m. There is no relative rotation between the systems and their X axes are coincident.
Exercise: Determine the image plane positions (in pixels) for all 8 corners of the box.

Be very detailed with explaining all the steps!



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

Images are typically redundant and contain lots of irrelevant data. An important topic in computer vision is feature extraction. By extracting high information features, such as lines, blobs and edges, the total amount of data can be reduced and used more efficiently for scene understanding. One of the methods for feature extraction that you used in labs was Hough transform.

(a) Describe the main steps of line extraction using Hough Transform, beginning with the image itself and ending with a list of line segments. (I have never heard of Hough transform before so be thorough!) Point out the decisions you have to make that could potentially affect the quality of the results and the computational speed.

(b) How could you extend the method to detect not just single lines, but also parallel lines? What would be different if you search for circles instead of just lines?

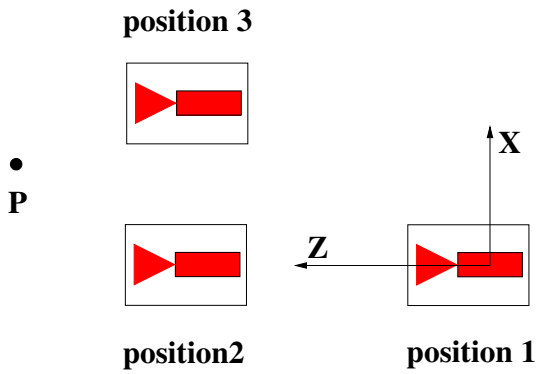
(c) Quite often you get multiple results for the same line in the image. In other cases lines are properly detected, but have a somewhat incorrect orientation. Mention at least two different reasons why these problems occur.

Exercise 4 (3+4=7 credits)

(a) What is meant by the epipolar constraint and essential matrix? Draw a picture and explain!

(b) Assume a robot with a camera on the top as shown in the below figure. There is a coordinate system attached to the camera with Z axis pointing in the direction of the camera axis, X axis as shown, and Y axis is pointing so to make a right-hand-rule coordinate system. The robot is first moving forward along the optical axis (to position 2) and then translating left without any rotation (to position 3). Estimate **the essential matrix** for the stereo systems given by positions 1 and 2, positions 2 and 3, and positions 1 and 3 respectively.

In addition, for the point P viewed by the camera in position 1, use the above essential matrices and estimate the equations of the corresponding **epipolar lines** in the next two camera positions.



Exercise 5 (2+2+3=7 credits)

Suppose there is a camera that undergoes a purely translational motion. The camera has a focal length of 100 and is moving with the velocity $(T_x, T_y, T_z) = (20, 30, 50)$. When we talked about optical flow and rigid body motion, we have derived the relationship between a moving camera and resulting motion in the image plane.

(1) Show how we can estimate the optical flow of a static point in the environment when the camera is moving.

(2) What is the aperture problem that affects the computation of the optical flow? How is it related to the linear constraint ("brightness constraint") determined by the differential technique for computing optical flow?

(3) Finally, what is the focus of expansion of a moving camera and how it can be estimated for the above stated camera velocity? What will the pattern of the flow vectors on the image plane be when the camera is translating this way.

Exercise 6 (2+2+2+2+2+2=12 credits)

(1) You are supposed to find positions of maximum spatial rate of change in an image. Give at least two convolution based methods that can be used for this purpose.

(2) If images are noisy and a) contain point like ('salt-and-pepper') noise or b) just random noise due to the sampling and quantization processes HOW would you process them prior to the above process? WHY do we have to do this (besides just improving the quality of the image)?

(3) Show how an approximation of the first derivative of an image can be obtained by convolving the image with a kernel $\delta_1 = [1, -1]$, where the image is defined as

[56 64 79 98 115 126 132 133]

Ignore computing a value for the first and last pixels, i.e. the output should be 6 values. In addition to showing the result of the convolution, indicate where edges would be detected.

(4) An alternative differentiation kernel is $\delta_2 = [1, 0, -1]$. In what sense is δ_1 preferable from δ_2?

(5) What is the purpose of (i) non-maximum suppression and of (ii) hysteresis thresholding that are done in the Canny edge detector?

(6) What is the purpose of scale-space representations? What is the most common way of their implementation?