CS-E4850 Computer Vision

Exam 13th of December 2019, Lecturer: Juho Kannala

There are plenty of questions, answer as many as you can in the available time. The number of points awarded from different parts is shown in parenthesis in the end of each question. The maximum score from the whole exam is 42 points.

You need pen and paper, also calculator is allowed but should not be necessary.

1. Explain briefly the following terms and concepts:
   (a) Separable filter (2 p)
   (b) RANSAC algorithm (2 p)
   (c) Precision and recall (2 p)
   (d) Inverted index (2 p)
   (e) Camera calibration (2 p)
   (f) Structure from motion (2 p)

2. Local feature detection and description using SIFT
   (a) Describe the detector part of the Scale Invariant Feature Transform (SIFT). In particular, explain the motivation and idea of the scale selection. (2 p)
   (b) Describe the descriptor part of SIFT. That is, describe how the pixel neighborhood around a detected keypoint is converted to a 128 dimensional feature vector. (2 p)
   (c) Mention at least two computer vision tasks or applications where SIFT is commonly used. Explain also what is the benefit of using SIFT in the applications (e.g. when compared to earlier methods which are not scale invariant). (2 p)

3. Lucas-Kanade optical flow

   The brightness constancy constraint that is utilized in optical flow computation can be written as follows

   $$(u\ v)^\top \cdot \nabla I + \frac{dI}{dt} = 0$$

   and it relates the flow to the spatial and temporal gradients of the image sequence.

   (a) Assuming that neighboring pixels have the same flow vector $(u\ v)^\top$, the brightness constancy constraint provides a set of linear equations for a given image patch in two consecutive frames of an image sequence (i.e. one equation per pixel). Write the system of linear equations in matrix form. (1 p)
   (b) Compute an expression for the flow vector $(u\ v)^\top$ by minimizing the sum of squared residuals. (Hint: Set the gradient of the cost function to zero.) (1 p)
   (c) When is the minimizing solution $(u\ v)^\top$ unique? How is the uniqueness of the solution related to the so called aperture problem? (2 p)
   (d) What are the pros and cons of Lucas-Kanade method when compared to template matching? (Template matching computes flow by comparing image patches explicitly using some similarity measure like normalised cross-correlation.) (2 p)

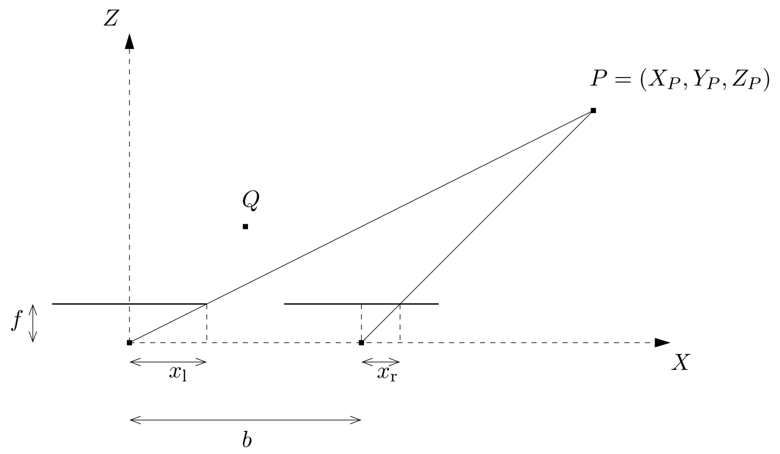

Figure 1: Top view of a stereo pair where two pinhole cameras are placed side by side.

4. Epipolar geometry and stereo

   (a) Figure 1 presents a stereo system with two parallel pinhole cameras separated by a baseline b so that the centers of the cameras are $\mathbf{c}_l = (0,0,0)$ and $\mathbf{c}_r = (b,0,0)$. Both cameras have the same focal length f. The point P is located in front of the cameras and its disparity d is the distance between corresponding image points, i.e., $d = |x_l - x_r|$. Assume that $d = 1\ cm$, $b = 6\ cm$ and $f = 1\ cm$. Compute Z_P. (2 p)

   (b) Let's denote the camera projection matrices of two cameras by $\mathbf{P} = \begin{bmatrix}\mathbf{I} & \mathbf{0}\end{bmatrix}$ and $\mathbf{P}' = \begin{bmatrix}\mathbf{R} & \mathbf{t}\end{bmatrix}$, where $\mathbf{R}$ is a rotation matrix and $\mathbf{t} = (t_1, t_2, t_3)^\top$ describes the translation between the cameras. Show that the epipolar constraint for corresponding image points $\mathbf{x}$ and $\mathbf{x}'$ can be written in the form $\mathbf{x}'^\top \mathbf{E}\mathbf{x} = 0$, where matrix $\mathbf{E}$ is the essential matrix $\mathbf{E} = [\mathbf{t}]_\times \mathbf{R}$. (2 p)

   (c) In the configuration illustrated in Figure 1 the camera matrices are $\mathbf{P} = \begin{bmatrix}\mathbf{I} & \mathbf{0}\end{bmatrix}$ and $\mathbf{P}' = \begin{bmatrix}\mathbf{I} & \mathbf{t}\end{bmatrix}$, where $\mathbf{I}$ is the identity matrix and $\mathbf{t} = (-6, 0, 0)^\top$. The point Q has coordinates $(3, 0, 3)$. Compute the image of Q on the image plane of the camera on the left and the corresponding epipolar line on the image plane of the camera on the right. (Hint: The epipolar line is computed using the essential matrix.) (2 p)

5. Geometric 2D transformations

   (a) Using homogeneous coordinates, write the matrix form of the following 2D transformations: translation, similarity (rotation+scaling+translation), affine and homography. How many degrees of freedom does each transformation have? How many point correspondences are needed to estimate each? (3 p)

   (b) A rectangle with corners $A = (-1, 1)$, $B = (1, 1)$, $C = (1, -1)$, $D = (-1, -1)$ is transformed by a transformation so that the new corners are $A' = (1, 3)$, $B' = (3, 3)$, $C' = (-2, 1)$, $D' = (-6, 1)$, respectively. An affine transformation does not explain the observations perfectly, but there is reason to believe that the transformation is affine and there is noise in the observations. Write down the equations to solve the transformation using the least squares method.
   Note: You don't actually have to solve the transformation. (3 p)

6. Neural networks

    (a) Explain how neural networks are typically used in image classification? What kind of neural networks are popular in this context and why? (2 p)

    (b) Explain the basic concepts of the backpropagation algorithm. (What it does? How it works? When it can be used? Why it may sometimes fail?) (2 p)

    (c) Explain the object detection problem and the basic principles of the single-shot multibox detector (SSD). In addition, describe the loss function and other key concepts of SSD training. (2 p)