DATA.ML.300 Computer Vision

Final exam

The maximum number of points for each task is shown in parenthesis. The use of calculator is allowed, but it is not necessary.

1. Explain briefly the following terms and concepts:

   (a) Camera projection matrix (1 p)

   (b) Scale Invariant Feature Transform (SIFT) (1 p)

   (c) Camera calibration (1 p)

   (d) Structure from motion (1 p)

   (e) Essential matrix (1 p)

   (f) Hough transform (1 p)

2. Model fitting using RANSAC algorithm

   (a) Describe the main stages of the RANSAC algorithm in the general case. (2 p)

   (b) In this context, why it is usually beneficial to sample *minimal subsets* of data points instead of using more data points? (Minimal subsets have the minimal number of data points required for fitting.) (1 p)

   (c) Mention at least two examples of models that can be fitted using RANSAC. Describe how the models are used in computer vision and what is the size of the minimal subset of data points required for fitting in each case. (1 p)

   (d) Describe how RANSAC can be used for panoramic image stitching. Why is RANSAC needed and what is the model fitted in this case? (2 p)

3. Feature tracking

   (a) Describe the main elements of the Shi-Tomasi feature tracker (i.e. how the features are selected and tracked, and tracks terminated or added). (2 p)

   (b) Describe the Lucas-Kanade method for estimating the displacement of an image patch. What kind of equations need to be solved and what is the brightness constancy constraint in this context? (2 p)

   (c) What are the benefits of Lucas-Kanade method when compared to simple template matching? (2 p)

4. Neural networks

   (a) What is a Perceptron? Explain the construction and give a rough idea how it can be trained (hint: use picture). (2 p)

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

   (c) Explain how neural networks are typically used in image classification? What kind of neural networks are popular in this context and why? (1 p)

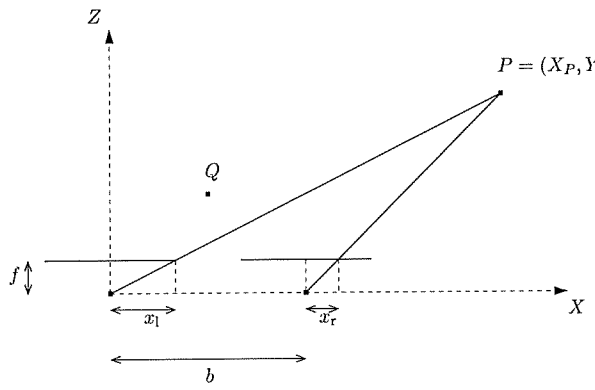

Figure 1: Top view of a stereo pair where two pinhole cameras are placed side by side.

(d) Explain the main differences between one and two stage object detection networks? Give an example of both types. (1 p)

5. Triangularisation 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) Two cameras are looking at the same scene. The projection matrices of the two cameras are $\mathbf{P}_1$ and $\mathbf{P}_2$. They see the same 3D point $\mathbf{X} = (X, Y, Z)^\top$. The observed coordinates for the projections of point $\mathbf{X}$ are $\mathbf{x}_1$ and $\mathbf{x}_2$ in the two images, respectively. The numerical values are as follows:

$$\mathbf{P}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{P}_2 = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & -1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}, \quad \mathbf{x}_1 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} \frac{3}{4} \\ 0 \end{bmatrix}.$$

Compute the 3D coordinates of the point $\mathbf{X}$. (Hint: Perhaps the simplest way in this case is to write the projection equations in homogeneous coordinates by explicitly writing out the unknown scale factors, and to solve X, Y, Z and the scale factors directly from those equations.) (1 p)

(c) Present a derivation for the linear triangulation method and explain how $\mathbf{X}$ can be solved using that approach in the general case (i.e. no need to compute with numbers in this subtask). (2 p)

(d) How does the triangulation approach differ from the bundle adjustment procedure which is commonly used in structure-from-motion problems (i.e. how is the bundle adjustment problem different)? (1 p)