Exam – CS-E4850 Computer Vision

Date: December 12, 2024 **Lecturer: Juho Kannala**
Time: 3 hours

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 for the whole exam is 42 points.

You will need pen and paper. Calculator is allowed but is not necessary.

1. **General concepts** **(12 p)**

   Explain briefly the following terms and concepts (e.g. what does the concept mean, what are its key properties, and how it is utilised in computer vision):

   (a) Precision and recall (2 p)

   (b) Laplacian pyramid (2 p)

   (c) Camera calibration (2 p)

   (d) Structure from motion (2 p)

   (e) Object detection by the single shot multibox detector (SSD) (2 p)

   (f) Multi-view stereo (2 p)

2. **Local feature detection and description using SIFT(Scale-Invariant Feature Transform)** **(6 p)**

   (a) Explain the difference between feature detector and descriptor. (1 p)

   (b) What are the main stages of the SIFT algorithm, and how does each stage contribute to the feature detection and description process? (2 p)

   (c) How does SIFT handle changes in scale, rotation, and illumination, and why is this important for computer vision tasks? (2 p)

   Let's assume that we detected SIFT regions from two images (i.e. circular regions with assigned orientations) of the same textured plane.

   (d) What is the minimum number of SIFT region correspondence pairs needed for computing a similarity transformation between two images, and why? (Briefly explain your reasoning.) (1 p)

3. **Model fitting using RANSAC algorithm** **(6 p)**

   (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. (2 p)

(d) What are the limitations of RANSAC, and how can they be mitigated in practice? (1 p)

4. Triangulation **(6 p)**

Two cameras are observing 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:

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

(a) 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)

(b) 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)

(c) Explain how adding more cameras (e.g., a third camera) can improve the accuracy of the triangulation process. (1 p)

(d) If there is noise (i.e. measurement errors) in the observed image coordinates of point $\mathbf{X}$, the linear triangulation method above is not the optimal choice but a nonlinear approach can be used instead. What error(loss) function is typically minimized in the nonlinear approach? (1 p)

(e) How does the nonlinear 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)

5. Lucas-Kanade optical flow **(6 p)**

The brightness constancy constraint used in optical flow can be written as:

$$uI_x + vI_y + I_t = 0,$$

which relates the flow vector (u, v) to the spatial (I_x, I_y) and temporal (I_t) gradients of the image sequence.

(a) Assuming that neighboring pixels in an image patch have the same flow vector (u, v), write the brightness constancy constraint for all pixels in the patch as a system of linear equations in matrix form. (1 p)

(b) Derive an expression for the flow vector (u, v) by minimizing the sum of squared errors of the brightness constancy constraint. (Hint: Use the gradient of the cost function.) (1 p)

(c) Under what condition is the flow vector (u, v) unique? How is this condition related to the aperture problem? (2 p)

(d) How does the choice of patch size affect the accuracy and robustness of the Lucas-Kanade method? Discuss the trade-offs between small and large patch sizes, and explain how the selection of patch size impacts performance in the presence of noise, motion variations, and computational cost. (2 p)

6. Neural networks (6 p)

(a) Explain how neural networks are typically used in image classification. What kind of neural networks are common in this context and why? What kind of loss function is typically used in image classification? (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 to produce a good solution?) (2 p)

(c) In Figure 1 below you see a very small neural network, which has one input unit, one hidden unit (logistic), and one output unit (linear). The nonlinear function σ in the logistic unit is defined by the formula $\sigma(z) = 1/(1 + e^{-z})$. Let's consider one training case. For that training case, the input value is 1 (as shown in the figure) and the target output value t is 1. We are using the standard squared loss function: $E = (t - y)^2/2$, where y is the output of the network. The values of the weights and biases are shown in the figure and they have been constructed in such a way that you don't need a calculator.

Answer the following questions:

i. What is the output of the hidden unit and the output unit, for this training case? (0.5 p)

ii. What is the loss, for this training case? (0.5 p)

iii. What is the derivative of the loss with respect to w2, for this training case? (0.5 p)

iv. What is the derivative of the loss with respect to w1, for this training case? (0.5 p)

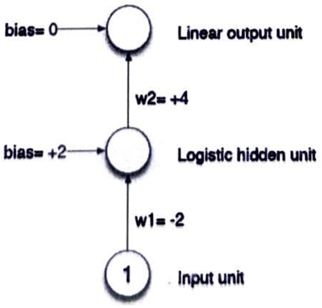

Figure 1: A small neural network with one hidden unit. The values for the weights and biases are given in the figure.