CS-E4850 Computer Vision

Exam 19th of February 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 concepts (e.g. what does the concept mean, what are its key properties, and how it is utilised in computer vision):

   (a) Separable filter (2 p)

   (b) Laplacian pyramid (2 p)

   (c) Hough transform (2 p)

   (d) Essential matrix (2 p)

   (e) Multi-view stereo (2 p)

   (f) Object detection by sliding windows and cascade classifiers (2 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. Large-scale object instance recognition

   (a) Describe the bag-of-visual-words image representation technique and its pros and cons for object instance recognition. (2 p)

   (b) Describe what is *inverted index* and how it can be used to improve efficiency of object instance recognition from large image databases? (1 p)

   (c) Explain the concept *term frequency - inverse document frequency* (tf-idf) weighting and its purpose. (1 p)

   (d) Explain what is the precision-recall curve (that is often used for evaluating image retrieval systems). Compute precision and recall in the following case: We search for car images from a database of 10000 images. It is known that there are 600 car images in the database. An automatic image retrieval system retrieves 200 car images and 40 other images from the database. (2 p)

4. 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)

5. Triangulation

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:

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

(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) A third camera $\mathbf{P}_3$ is added to the scene. Describe how the linear triangulation method above can be extended to use the information from all the three cameras. (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 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)

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) 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)

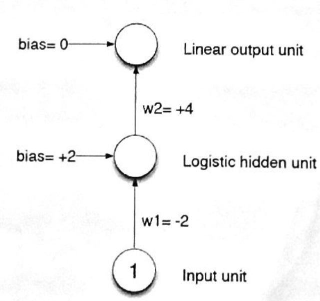

Figure 1: A small neural network with one hidden unit. The values for the weights and biases are given in the figure.