CS-E4850 Computer Vision

Exam 19th of April 2023, 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 will need pen and paper, and also calculator is allowed but is not necessary.

1. Explain briefly the following terms and concepts:
   - (a) Precision and recall (2 p)
   - (b) Structure from motion (2 p)
   - (c) Essential matrix (2 p)
   - (d) Scale invariant feature transform (SIFT) (2 p)
   - (e) Multi-view stereo (2 p)
   - (f) Object detection by sliding windows (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. Image retrieval
   - (a) When matching features across two images, why does it make sense to use the ratio: (distance to best match) / (distance to second best match), as a way to judge if we have found a good match? (2 p)
   - (b) How do we use clustering to compute a bag-of-words image representation? Describe the process. (1 p)
   - (c) How can we find to which cluster we should assign a new feature, which was not part of the set of features used to compute the clustering? (1 p)
   - (d) When is it more efficient to create an inverted file index to match a query image to other images in the database, rather than comparing the query to all database images without an index? (1 p)
   - (e) Why do we need to measure both precision and recall in order to score the quality of retrieved results? (1 p)
4. Image filtering

(a) Filter image J with the gaussian filter G using zero padding. (1 p)

$$J = \begin{array}{|c|c|c|}\hline 0 & 1 & 0 \\ \hline 0 & 2 & 0 \\ \hline 0 & 3 & 0 \\ \hline \end{array} \qquad G = \frac{1}{16}\begin{array}{|c|c|c|}\hline 1 & 2 & 1 \\ \hline 2 & 4 & 2 \\ \hline 1 & 2 & 1 \\ \hline \end{array}$$

(b) Is it more efficient to filter an image with two 1D filters as opposed to a 2D filter? Why? How does the computational complexity relate to the size of the filter kernel (with K×K pixels) in both cases? (1 p)

(c) Is the following convolution kernel separable? If so, separate it. (1 p)

$$H = \begin{array}{|c|c|c|c|}\hline 1 & 2 & 3 & 4 \\ \hline 2 & 4 & 6 & 8 \\ \hline 3 & 6 & 9 & 12 \\ \hline 4 & 8 & 12 & 16 \\ \hline \end{array}$$

For the image I below apply the following filters to the pixel at the center (marked with a box).

$$I = \begin{array}{|c|c|c|c|c|}\hline 2 & 3 & 4 & 5 & 6 \\ \hline 3 & 4 & 5 & 6 & 8 \\ \hline 4 & 5 & \boxed{6} & 8 & 5 \\ \hline 5 & 7 & 8 & 9 & 3 \\ \hline 9 & 10 & 9 & 4 & 3 \\ \hline \end{array}$$

(d) 3×3 box filter (i.e. averaging in a 3×3 neighborhood). (1 p)

(e) 3×3 median filter. (1 p)

(f) Why is the Gaussian filter a better smoothing filter than a box filter? (1 p)

5. 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)

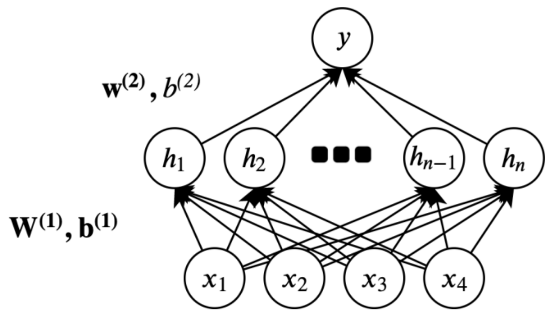
(b) Compute an expression for the flow vector $(u\ v)^\top$ by minimizing the sum of squared residuals. (Hint: Set the gradient of the cost function to zero.) (1 p)

(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)

6. Neural networks

In this problem, we'll find parameters for a multilayer perceptron to check whether four inputs x_1, x_2, x_3, x_4, where $x_i \in \mathbb{R}$, are unique. That is, our network should output 1 if no two input are equal with $x_i \neq x_j$ for $i \neq j$. We will use a two-layer neural network like this one:



All of the hidden units will use the impulse activation function, defined as follows:

$$\phi(x) = \begin{cases} 1, & \text{if } x = 0 \\ 0, & \text{otherwise} \end{cases}$$

(a) We will use n = 6 hidden units in our hand-coded neural network. Explain why we will not require more than 6 hidden units. (1 p)

(b) What are the shapes of each of the following quantities? (1 p)

- $\mathbf{W}^{(1)}$ - the weight matrix containing the weights of the first layer of the neural network
- $\mathbf{b}^{(1)}$ - the bias vector containing the baises to the hidden layers
- $\mathbf{W}^{(2)}$ - the weights containing the weights of the second layer of the neural network
- $\mathbf{b}^{(2)}$ - the bias containing the baises to the output layer

(c) Hand-pick a set of weights and biases so that the network correctly implements the desired functionality. Your answer should include the values of $\mathbf{W}^{(1)}$, $\mathbf{b}^{(1)}$, $\mathbf{W}^{(2)}$, $\mathbf{b}^{(2)}$. (1 p)

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

(e) 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)