CS-E4850 Computer Vision

Exam 16th of December 2022, 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) Convolutional neural network (2 p)

   (b) Separable filter (2 p)

   (c) Hough transform (2 p)

   (d) Camera calibration (2 p)

   (e) Single shot multibox detector (SSD) (2 p)

   (f) Laplacian pyramid (2 p)

2. Local feature detection and description using SIFT

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

   (b) Is Harris corner detector rotation invariant? Could Harris corner detector and normalized cross-correlation based matching be used to match corner features in images related by a rotation? (2 p)

   (c) How do we compute a histogram of gradient orientations when generating a SIFT descriptor? (1 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 the pair of images? (1 p)

   (e) Describe RANSAC-based procedure for estimating the similarity transformation in a real world use case where there are both correct and incorrect correspondences among the SIFT region correspondences. (1 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) Describe what is the role of *spatial verification* in object instance recognition and how it is usually performed? (2 p)

4. Image formation

Assume that all coordinate frames are right handed.

Consider a camera with the following camera projection matrix:

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -10 \\ 0 & 1 & 0 & 10 \end{bmatrix}$$

(a) There is a triangle located in the world coordinate system with vertices: $(10, 0, 0)$, $(0, 10, 0)$, and $(0, 0, 10)$. Calculate the coordinates of the projection and draw the image of this triangle as seen by our camera. (2 p)

(b) What happens if you try to compute the projection of $(0, -10, -10)$, and why? (2 p)

Now we switch to a camera with the following intrinsic parameters:

- A 10 mm focal distance
- Rectangular pixels, 2.5 micron wide and 2 micron tall (one mm is 1000 micron)
- A camera sensor with 4000 pixels horizontally and 3000 vertically
- The principal point is at image point π =(2000,1500) pixels
- No distortion

The world reference system is the same as the camera's canonical reference system (camera is at the world origin and pointed towards the positive z-axis), except that the units for the axes are in centimeters. The cameras pixel image reference system measures image coordinates in pixels, and its origin is in the upper left corner of the image.

(c) Calculate the intrinsic, extrinsic and projection matrices. (1 p)

(d) A point $\mathbf{X}$ has coordinates (0, -10, 40) centimeters in the world reference system. Calculate the coordinates of the projection of X in the pixel image reference system. (1 p)

5. Epipolar geometry

We have a camera pair with projection matrices $P = [I \quad 0]$ and $P' = [R \quad \mathbf{t}]$ where R and $\mathbf{t}$ are such that the essential matrix E for the camera pair is the following:

$$E = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$$

(a) What is the epipolar constraint? What is the difference between fundamental matrix and essential matrix? (2 p)

(b) Give one possible value for the unit-norm vector $\mathbf{t}$ that points from the second camera center to the first camera center and for the rotation matrix R between the two cameras. Justify your answer briefly. (Hint: $E = [t_\times]R$.) (2 p)

(c) A point in the second image has coordinates $\mathbf{b} = \begin{bmatrix} 0.5 & 1 & 1 \end{bmatrix}^T$ in the canonical camera reference system (units are focal distances). Write the equation of the epipolar line of $\mathbf{b}$ in the canonical image reference system of the first camera. Show your derivation. Remember that the canonical (i.e. normalized) image coordinates of a point are the same as its first two canonical camera coordinates. (2 p)

6. Feature tracking

Let $I(\mathbf{x})$ and $J(\mathbf{x})$ be two grayscale images of the same scene taken from slightly different viewpoints and possibly slightly different orientations. We'd like to track a point $\mathbf{x}_I$ in image I to it's coordinate $\mathbf{x}_J$ in image J. That is we'd like to know the two dimensional displacement $\mathbf{d}^*$ of point $\mathbf{x}_I$ such that:

$$\mathbf{x}_J = \mathbf{x}_I + \mathbf{d}^*$$

To approximate $\mathbf{d}^*$ we look at a window (small square) $W(\mathbf{x}_I)$ of odd side-length 2h+1 pixels centered around the point $\mathbf{x}_I$ in image I and search for $\mathbf{d}$ that minimizes the dissimilarity between the windows in both images:

$$\mathbf{d}^* = \arg\min_{\mathbf{d}} \epsilon(\mathbf{d})$$

where the dissimilarity $\epsilon(\mathbf{d})$ is defined as a sum over the whole image $\mathbf{x} = (x_1, x_2)$:

$$\epsilon(\mathbf{d}) = \sum_{\mathbf{x}} [J(\mathbf{x} + \mathbf{d}) - I(\mathbf{x})]^2 w(\mathbf{x} - \mathbf{x}_I)$$

$w(\mathbf{x})$ is the indicator function of a $W(\mathbf{x})$:

$$w(\mathbf{x}) = \begin{cases} 1 & \text{if } |x_1| \leq h \text{ and } |x_2| \leq h \\ 0 & \text{otherwise.} \end{cases}$$

We assume that the motion of the camera between the two images is so small that the magnitude of $\mathbf{d}^*$ is much smaller than the diameter of $W(\mathbf{x}_I)$ and use an iterative approach so that we can formulate the problem as follows: find a step displacement $\mathbf{s}_t$ that, when added to $\mathbf{d}_t$, yields a new displacement $\mathbf{d}_{t+1}$ at each iteration t such that $\epsilon(\mathbf{d}_t + \mathbf{s}_t)$ is minimized. We add $\mathbf{d}_t$ into $\mathbf{x}$ as follows $J_t(\mathbf{x}) = J(\mathbf{x} + \mathbf{d}_t)$ and approximate the image function $J_t(\mathbf{x} + \mathbf{s}_t)(= J(\mathbf{x} + \mathbf{d}_t + \mathbf{s}_t))$ with its first-order Taylor expansion:

$$J_t(\mathbf{x} + \mathbf{s}_t) \approx J_t(\mathbf{x}) + [\nabla J_t(\mathbf{x})]^T \mathbf{s}_t$$

Minimizing $\epsilon(\mathbf{d}_t + \mathbf{s}_t)$ leads to a linear system of equations $A\mathbf{s}_t = \mathbf{b}$ where

$$A = \sum_{\mathbf{x}} \nabla J_t(\mathbf{x}) [\nabla J_t(\mathbf{x})]^T w(\mathbf{x} - \mathbf{x}_I) \quad \text{and} \quad \mathbf{b} = \sum_{\mathbf{x}} \nabla J_t(\mathbf{x}) [I(\mathbf{x}) - J_t(\mathbf{x})] w(\mathbf{x} - \mathbf{x}_I)$$

The overall displacement is then the sum of all the steps:

$$\mathbf{d}^* = \sum_t \mathbf{s}_t$$

(a) Show that minimizing $\epsilon(\mathbf{d}_t + \mathbf{s}_t)$ leads to a linear system of equations $A\mathbf{s}_t = \mathbf{b}$ (1 p)

NOTE: the problems (b)-(f) below don't require that you have solved problem (a).

Assuming a window size of 3×3 ($h = 1$) and an initial guess of displacement $\mathbf{d}_0 = [0, 0]^T$. For a particular value of $\mathbf{x}_I$, the two components of $\nabla J_0(\mathbf{x})$ inside the window $W(\mathbf{x}_I)$ are:

$$\frac{\partial J_0}{\partial x_1} = \begin{bmatrix} 10 & 10 & 10 \\ 10 & 10 & 10 \\ 10 & 10 & 10 \end{bmatrix} \quad \text{and} \quad \frac{\partial J_0}{\partial x_2} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and the difference between the two images is:

$$I(\mathbf{x}) - J_0(\mathbf{x}) = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

(b) Compute A and $\mathbf{b}$. (1 p)

(c) Explain briefly what the so called aperture problem is. (1 p)

(d) Does the feature at $\mathbf{x}_I$ suffer from the aperture problem? Briefly justify your answer. (1 p)

(e) Give the minimum-norm solution $\mathbf{s}_0$ to the linear system $A\mathbf{s}_0 = \mathbf{b}$ (1 p)

(f) Assume that further iterations of the Lucas-Kanade algorithm do not change the solution $\mathbf{s}_0$ much. Does your answer to the previous question imply that the image motion between I and J at $\mathbf{x}_I$ is approximately horizontal? Briefly justify your answer. (1 p)