102470 - Computer Vision Course
Institut für Informatik
Universität Bern

MOCK EXAM

18/12/2018

- **You can use one A4 sized hand-written sheet of paper.**
- **No books, notes, computers, calculators and cellular phones are allowed.**

Optical Flow, Tracking, Registration, Fitting & Recognition [14 points total]

1. The task of optical flow is to find the motion field u and v by minimizing the functional

$$E[u,v] = |I_x^{t-1}u + I_y^{t-1}v + I^{t-1} - I^t|^2 + \lambda(|\nabla u| + |\nabla v|), \tag{1}$$

where I is the grayscale image, $|\nabla u|$ and $|\nabla v|$ are the total variation on u and v, and t is the index of the video frame. This is derived from the Taylor series expansion (up to the first order) of the brightness constancy equation

$$I(x - u(x,y), y - v(x,y), t-1) = I(x,y,t). \tag{2}$$

Give three scenarios (based on motion and brightness) where optical flow fails. Justify your answer by using the formulas above. **[6 points]**

Solution.
1. The motion is considerably larger than 1 pixel. In this case the Taylor approximation is not accurate. **2 p**
2. The motion is not uniform locally. TV prior favors locally uniform solutions. **2 p**
3. Brightness changes rapidly. The brightness constancy assumption (and the equation derived from it) does not hold. **2 p**

2. Consider the equation of a parabola $y = ax^2 + x$. Compute the parameter a that best fits the points (1, 3), (-1, 1) and (2, 0.5) with the least squares method. Write the least squares objective and show all your calculations. **[8 points]**

Solution.
The least squares objective is $L(a) = \sum_i (y_i - ax_i^2 - x_i)^2$ **2 p**. This is convex in a, the minimum can be found where $\frac{\partial L}{\partial a} = 0$ **2 p**. Because $\frac{\partial L}{\partial a} = -2\sum_i (y - ax_i^2 - x_i)x_i^2$, we get **2 p**

$$a = \frac{\sum_i (y_i x_i^2 - x_i^3)}{\sum_i x_i^4} = \frac{2+0-6}{1+1+16} = -\frac{4}{18}. \tag{3}$$

2 p for final answer.

Epipolar Geometry, Multiple Views & Motion [12 points total]

1. Why is image rectification useful in stereo matching? [**4 points**]

   Solution
   All epipolar lines are perfectly horizontal. [**2 points**] Image rectification makes the correspondence problem easier and reduces computation time. [**2 points**]

2. Epipolar geometry is the intrinsic projective geometry between two views I and I'. It depends only on the camera intrinsic parameters and their relative pose (rotation and translation between the camera centers). The Fundamental Matrix F is a 3×3 matrix.

   (a) How are the fundamental matrices F, going from I to I', and F' going from I' to I, related? [**2 points**]

   Solution
   $F^T = F'$. [**2 points**]

   (b) What is the geometric meaning of the epipoles **e** and **e**$'$? How are they (algebraically) related to the fundamental matrix? [**4 points**]

   Solution
   Geometrically the epipoles are the points of the intersection of the line joining the camera centers (the baseline) with the image planes. Equivalently, an epipole is the image in one view of the camera center of the other view. Algebraically, **e** is the right null space of **F**, i.e. **Fe** $= 0$. Similarly, **e'** is the left null space of **F**, i.e. **e'**^{T}**F** $= 0$. [**4 points**]

   (c) What is the effect of applying the fundamental matrix F to a point x? [**2 points**]

   Solution
   The fundamental matrix maps a point in one image to a line in the second image. $l' = Fx$ and $l = F^T x'$. [**2 points**]

Energy minimization & Bayesian estimation [23 points total]

1. Find the solution to the following energy minimization problem **[6 points]**

$$\arg\min_u |Au - f|^2 + \lambda |u - f|^2 \tag{4}$$

where $A \in \mathbf{R}^{n \times n}$ and $u, f \in \mathbf{R}^n$.

Solution To find the solution we need to compute the gradient and set it to zero **[3 points]**

$$\nabla E = 2A^\top (Au - f) + 2\lambda(u - f) = 0. \tag{5}$$

By rearranging the equation we obtain **[3 points]**

$$A^\top Au - A^\top f + \lambda(u - f) = 0 \tag{6}$$

$$\left(A^\top A + \lambda I\right) u = (A^\top + \lambda I) f \tag{7}$$

$$u = \left(A^\top A + \lambda I\right)^{-1} (A^\top + \lambda I) f. \tag{8}$$

2. Suppose you are given a collection of images $Y_i \in \mathbb{R}^m, i = 1, \dots n$, and you know Y_i are noisy measurements of an image $X \in \mathbb{R}^m$, such that

$$Y_i = X + \eta, \tag{9}$$

where the noise $\eta \sim \mathcal{N}(0, I)$ is assumed to be of zero mean and unit variance. Derive the maximum likelihood estimate of X.
Hint 1: $Y_i \sim \mathcal{N}(X, I)$.
Hint 2: The density of the multivariate normal distribution is

$$p(y; \mu, \Sigma) = \det(2\pi\Sigma)^{-\frac{1}{2}} e^{-\frac{1}{2}(y-\mu)^T \Sigma^{-1} (y-\mu)}. \tag{10}$$

[10 points]

Solution The likelihood of the data given the parameters is

$$p(Y; X) = \prod_{i=1}^{n} (2\pi)^{-\frac{m}{2}} e^{-\frac{1}{2}(Y_i - X)^T (Y_i - X)}. \tag{11}$$

Maximizing this is equivalent to maximizing the log-likelihood:

$$\sum_{i=1}^{n} -\frac{m}{2} \log(2\pi) - \frac{1}{2}(Y_i - X)^T (Y_i - X) \tag{12}$$

We take the derivative w.r.t. X and set it to zero:

$$\sum_{i=1}^{n} X - Y_i = 0. \tag{13}$$

By rearranging the terms, we obtain the estimate

$$X = \frac{1}{n} \sum_{i=1}^{n} Y_i. \tag{14}$$

3. Suppose we are given a task of fitting the parameters of a Gaussian Mixture Model (GMM) $p(x, z)$ to the data $\{x^{(1)}, \ldots, x^{(m)}\}$ consisting of m independent samples, where z denotes discrete latent variable. Each $z^{(i)}$ identifies the Gaussian from which the sample $x^{(i)}$ was generated.

   (a) Write the data log-likelihood under a Gaussian Mixture Model. [**3 points**]
   Solution The data likelihood can be expressed as

$$\begin{aligned}\ell(\theta) &= \sum_{i=1}^{m} \log p(x; \theta) \\ &= \sum_{i=1}^{m} \log \sum_{z} p(x, z; \theta).\end{aligned}$$

   (b) Why do we need the EM algorithm to fit the parameters of GMM? Why do we not simply maximize the likelihood by setting $\nabla_\theta \ell(\theta)$ to 0? [**4 points**]
   Solution The main difficulty comes from the fact that variable $z^{(i)}$ are not observed and the posterior does not factorize, making it much harder to compute. [**2 points**] This also complicates the computation of MAP and ML estimates of the parameters. [**2 points**]