# Försättsblad till skriftlig tentamen vid Linköpings universitet

| | |
|---|---|
| **Datum för tentamen** | 2019-05-29 |
| **Sal (1)** | TER3(26) |
| **Tid** | 14-18 |
| **Utb. kod** | TSBB15 |
| **Modul** | TEN1 |
| **Utb. kodnamn/benämning**<br>**Modulnamn/benämning** | Datorseende<br>Skriftlig tentamen |
| **Institution** | ISY |
| **Antal uppgifter som ingår i tentamen** | 12 |
| **Jour/Kursansvarig**<br>Ange vem som besöker salen | Per-Erik Forssén |
| **Telefon under skrivtiden** | 013-285654 |
| **Besöker salen ca klockan** | 15.00 |
| **Kursadministratör/kontaktperson**<br>(namn + tfnr + mailaddress) | Per-Erik Forssén, 013-285654<br>per-erik.forssen@liu.se |
| **Tillåtna hjälpmedel** | Dictionary Swedish-English-Swedish |
| **Övrigt** | See instructions on the next page Cross-ruled |
| **Antal exemplar i påsen** | |

# Instructions TEN1

The exam consists of 4 parts, each corresponding to one of the four lab exercises of TSBB15.

- Part 1 covers tracking
- Part 2 covers motion
- Part 3 covers denoising
- Part 4 covers multiple view geometry

Each part contains 3 tasks, two that require description of terms, phenomena, relations, etc. (type A) and one that goes more into detail and may require some calculations (type B).

Correct answers for type A give 2p and for type B give 4p, i.e., each part gives 8p, and a total of 32p for the whole exam.

In order to pass with grade 3, at least 15p are required.

In order to pass with grade 4, at least 22p are required.

In order to pass with grade 5, at least 27p are required.

Re-use of the midterm examination result: If you have written the midterm examination and obtained 8p or more in Parts 1 and 2, you may re-use this result, and not answer the corresponding parts in this exam.

All tasks should be answered on **separate sheets**. Do not write answers directly in the exam.

Write your AID-number and the date on all sheets that you hand in. In addition, the sheets should be numbered in consecutive order.

Good luck!

Per-Erik Forssén and Michael Felsberg

# PART 1: Tracking

**Task 1** (B, 4p) The *Expectation Maximization*(EM) algorithm can be used to fit a Gaussian mixture model to a set of samples $\{x_n\}_1^N$.
a) For the case of $K = 3$ mixture components, write down the expression for the mixture model, and describe what parameters are fitted (Hint: the samples are scalars). (2p)
b) In the E-step, *responsibilities* are computed. Explain what these are, and how they relate to the samples. (1p)
c) Describe what is computed in the M step, and how. (1p)

**Task 2** (A, 2p)
Tracking can be done by minimizing a cost function of the following form:

$$E(\mathbf{d}) = \sum_{\mathbf{x}\in\Omega} \|I(\mathbf{x}) + \mathbf{d}^T \nabla I(\mathbf{x}) - J(\mathbf{x})\|^2 . \quad (1)$$

Here $I$ and $J$ are two consecutive image frames, $\mathbf{d}$ is the sought displacement, and $\Omega$ is the region being tracked. Now consider the surface $E(\mathbf{d})$.
a) Explain when $E(\mathbf{d})$ is strictly convex in $\mathbf{d}$, and when it is not. (1p)
Hint: A function $f(\mathbf{x})$ is *strictly convex* in $\mathbf{x}$ if:

$$\lambda f(\mathbf{x}_1) + (1-\lambda) f(\mathbf{x}_2) > f(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2)\,,\ \lambda \in ]0,1[ \quad (2)$$

for all choices of $\mathbf{x}_1, \mathbf{x}_2$ (with $\mathbf{x}_1 \neq \mathbf{x}_2$).
b) Relate the convexity of $E(\mathbf{d})$ to concepts introduced in the course. (1p)

**Task 3** (A, 2p)
A classic problem in tracking by detection is the *assignment problem*: Given detected bounding boxes in one frame, which detected bounding boxes in the next frame correspond to these?

Consider the following *matching matrix* of Jaccard index values, where we denote the detections in frame 1 by $\{\mathbf{Ak}\}_{k=1}^3$ and the detections in frame 2 by $\{\mathbf{Bl}\}_{l=1}^4$:

| $J(\mathcal{A}, \mathcal{B})$ | **B1** | **B2** | **B3** | **B4** |
|---|---|---|---|---|
| **A1** | 0.7 | 0 | 0 | 0 |
| **A2** | 0 | 0.4 | 0 | 0.5 |
| **A3** | 0 | 0 | 0 | 0.4 |

The Jaccard index for two regions is defined as:

$$J(\mathcal{A}, \mathcal{B}) = \frac{|\mathcal{A} \cap \mathcal{B}|}{|\mathcal{A} \cup \mathcal{B}|}\,, \quad (3)$$

a) Denote the matching matrix elements above by $J(k, l)$. An *assignment* on this matrix is a set of index pairs $\{(k_a, l_a)\}_{a=1}^3$ such that each detection is used at most once, i.e. $k_a = k_b \Leftrightarrow a = b$, and $l_a = l_b \Leftrightarrow a = b$. Explain which assignment is the optimal one, and define a cost function that is maximal for this assignment. (1p)
b) A common assignment strategy is to select the maxima along rows or columns in the matching matrix, and then removing low-scoring pairs that use the same detection more than once. Will such a strategy maximise your cost function in (a)? Motivate! (1p)

# PART 2: Motion

**Task 4** (B, 4p)
a) Explain the two concepts *motion field* and *optical flow*. (1p)
b) Explain the difference between the two, e.g., by describing specific cases where they differ. (1p)
c) The local motion between two images $I$ and $J$ can be based on the *brightness constancy assumption*:

$$I(\mathbf{x} + \mathbf{d}(\mathbf{x})) = J(\mathbf{x}). \tag{4}$$

Here, $\mathbf{x}$ is an arbitrary point in image $I$ and $\mathbf{d}(\mathbf{x}) = (d_1(\mathbf{x}), d_2(\mathbf{x}))$ is the motion/displacement of this point in image $I$. An alternative formulation is the differential *brightness constancy constraint equation* (BCCE):

$$\frac{\partial I}{\partial x} \cdot d_1(\mathbf{x}) + \frac{\partial I}{\partial y} \cdot d_2(\mathbf{x}) = -\frac{\partial I}{\partial t}. \tag{5}$$

Describe how (5) is derived from the assumption of brightness constancy and what assumption is made. (2p)

**Task 5** (A, 2p)
If we estimate $\mathbf{d}$ from (5), we need to consider more than one point, e.g. a region $\Omega$. Why do we need more than one point and what phenomenon occurs if the region contains a simple signal? (2p)

**Task 6** (A, 2p)
Name one advantage of the *global method* introduced in the lecture compared to the region approach in Task 5. What is the name of the global method? (2p)

# PART 3: Denoising

**Task 7** (A, 2p)
Three different classes of diffusion filtering methods have been introduced in the lecture on image enhancement. To characterize the three different approaches, at least two properties are required.
a) Draw a table with the names of the three classes of diffusion filtering and at least two properties to distinguish those three classes. (1p)
b) In the table, indicate which of the approaches fulfill which of the properties. (1p)

**Task 8** (A, 2p)
In image enhancement, the local orientation tensor $\mathbf{T}(\mathbf{x})$ is mapped to a diffusion tensor $\mathbf{D}(\mathbf{x})$ in a non-linear way.
a) How do the eigenvalues change and why is this change necessary? (1p)
b) The mapping suggested in the lecture does not require an explicit eigenvalue extraction. Why is this the case? Hint: write $\mathbf{T}(\mathbf{x})$ in factorized form and look at a higher-order term in the expansion of the matrix-valued exponential function $\exp(\mathbf{A}) = \sum_{k=0}^{\infty} \frac{\mathbf{A}^k}{k!}$. (1p)

**Task 9** (B, 4p)
In linear image restoration, we minimize the functional

$$\varepsilon(u) = \int_{\Omega} L(u, u_x, u_y, x, y)\, dx\, dy$$

where $u_x$ and $u_y$ are the partial derivatives of the unknown image $u$, and $x$ and $y$ are the spatial coordinates. At the minimum, the Euler-Lagrange equation

$$L_u - \operatorname{div}(L_{u_x}, L_{u_y}) = u(x, y) - u_0(x, y) - \lambda \Delta u(x, y) = 0 \tag{6}$$

is fulfilled, where $u_0$ is the observed noisy image, $\lambda > 0$ is a constant, and $\Delta$ is the 2D Laplace operator, i.e., $\Delta u = u_{xx} + u_{yy}$.

a) Write down a suitable Lagrangian $L(u, u_x, u_y, x, y)$ for (6) and show that the left equality in (6) is fulfilled. (2p)

b) In the lecture, the Horn-Schunck iteration has been derived using the discrete Laplace operator $[1, -2, 1] = [1, 1, 1] - 3 \cdot [0, 1, 0]$ instead of $\Delta$. Apply the same trick to rewrite the right equality in (6) as the recursive formula

$$u^{(t+1)} = \bar{u}^{(t)} - \frac{\bar{u}^{(t)} - u_0}{1 + \lambda}$$

where $\bar{u}$ is the sum of horizontal mean and vertical mean of $u$ (over three samples each).(2p)

# PART 4: Multiple View Geometry

**Task 10** (A, 2p)

In calibrated stucture from motion, two of the components are often combined with RANSAC: 1) Essential matrix estimation, and 2) Perspective n-point estimation.

a) Write down the residuals used in the two components. (1p)

b) Explain which component is better at detecting outliers and why. (1p)

**Task 11** (A, 2p)

The MSER local feature uses thresholding to find regions that are then approximated as ellipses. For each region it also finds a reference direction, e.g. as the furthest distance from the centroid along the region boundary. As descriptor MSER originally used a subsampled RGB image patch sampled in a coordinate frame where the ellipse approximation of the region becomes a circle, and where the reference direction points to the right. The patch was then normalized in each colour channel by subtracting the mean, and then dividing by the standard deviation of the values in the patch.

a) Which *geometric invariances* does MSER have? (1p)

b) Which *photometric invariances* does MSER have? (1p)

**Task 12** (B, 4p)

An initial structure from motion solution can be refined by a technique called *bundle adjustment*.

a) Write down the bundle adjustment cost function and name all variables. Also state which variables in the cost function are known, and which are free parameters, to be adjusted in the optimization. (2p)

b) How can an initial solution of the free parameters be obtained? Describe one of the methods that has been presented in the course. (2p)