Försättsblad till skriftlig tentamen vid Linköpings universitet

Datum för tentamen	2018-05-30
Sal (1)	TER2(23)
Tid	14-18
Kurskod	TSBB15
Provkod	TEN1
Kursnamn/benämning **Provnamn/benämning**	Datorseende Skriftlig tentamen
Institution	ISY
Antal uppgifter som ingår i tentamen	12
Jour/Kursansvarig Ange vem som besöker salen	Per-Erik Forssén
Telefon under skrivtiden	013-285654
Besöker salen ca klockan	15.30
Kursadministratör/kontaktperson (namn + tfnr + mailaddress)	Per-Erik Forssén, 013-285654 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** that are to be attached to the exam.

Write your AID-number and the date on all paper sheets that you attach to the examination. In addition, these sheets should be numbered in consecutive order.

Good luck!

Per-Erik Forssén, Michael Felsberg, and Klas Nordberg

PART 1: Tracking

Task 1 (B, 4p) The *Expectation Maximization* 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) Describe and motivate a strategy to initialize each of the unknown parameters. (1p)
c) Explain why the choice of initialization matters. (1p)

Task 2 (A, 2p)
In tracking, a cost function of the following form is minimized:

$$E(\mathbf{d}) = \sum_{\mathbf{x}\in\Omega} \rho(\|I(\mathbf{x}+\mathbf{d}) - J(\mathbf{x})\|). \tag{1}$$

Here I and J are two consecutive image frames, $\rho(\epsilon)$ is an *error norm*, and Ω is the region being tracked. In the KLT algorithm, tracking is done using an iteration derived from a Taylor expansion, and setting the derivative to zero, for $\rho(\epsilon) = \epsilon^2$. For $\rho(\epsilon) = |\epsilon|$, we typically get better performance near motion discontinuities, i.e. when Ω contains more than one motion, but this choice has other disadvantages.
a) For $\rho(\epsilon) = |\epsilon|$, derivation of a KLT-style iteration is problematic. Explain why. (1p)
b) Explain how the cost function in (1) can be minimized without a KLT-style iteration. (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:

$$J(\mathcal{A},\mathcal{B}) = \frac{|\mathcal{A}\cap\mathcal{B}|}{|\mathcal{A}\cup\mathcal{B}|}, \tag{2}$$

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 (A, 2p)
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)

Task 5 (B, 4p)
a) Based on the principle of brightness constancy, formulate a differential equation that involves the image $I(\mathbf{x})$, and the displacement field $\mathbf{d}(\mathbf{x}) = (d_1(\mathbf{x}), d_2(\mathbf{x}))$. (1p)
b) Formulate a cost function $E(\mathbf{d}(\mathbf{x}))$ that replaces $E(\mathbf{d})$ in Eq. (1) for this case, such that E in general is minimized by a unique and well-defined $\mathbf{d}(\mathbf{x})$. How does Ω change? (1p)
c) Formulate a strategy which determines a displacement field $\mathbf{d}(\mathbf{x})$ that minimizes the new cost function $E(\mathbf{d}(\mathbf{x}))$. Motivate your formulation. (2p)

Task 6 (A, 2p)
a) Describe and explain the so-called aperture problem in motion estimation. (1p)
b) Let I be the intensity function in a region Ω (the aperture). In this region, I can be characterized in using (A) its intrinsic dimensionality, (B) its Fourier transform, and (C) its structure tensor. Describe these three properties of I for the case when the aperture problem occurs, and when it does not. (1p)

PART 3: Denoising

Task 7 (A, 2p)
The first version of diffusion based image enhancement was introduced by Perona & Malik. It uses a formulation of the blurring filter, at scale s, according to:

$$g_s(\mathbf{x}) = \frac{1}{2\pi\mu s}\, e^{-\frac{1}{2\mu s}\|\mathbf{x}\|^2} .$$

Here, μ is a position dependent parameter that controls the speed of the diffusion process at each image point.
a) Describe in text how μ depends on position. How is μ controlled by the content in a region around an image point? When should μ be large, and when should it be small? Explain, and also provide the expression proposed by Perona & Malik. (1p)
b) The Perona-Malik approach has a specific problem, which motivates alternative formulations of the filter g_s, e.g., as was done by Weickert. What problem is that? (1p)

Task 8 (A, 2p)
Image enhancement can be implemented by adaptive filtering where a filter is interpolated at each image position, by adapting it to the local region around the point. In the frequency domain, the angular component of the high-frequency part of the filter can, *in principle*, be written as $\mathbf{u}^\top \mathbf{T}\,\mathbf{u}$, where, $\mathbf{u}$ is the coordinate in the frequency domain, and $\mathbf{T}$ is the orientation tensor at the position where the filter is applied. Motivate why this formulation makes sense at a position centered on an i1D neighborhood, e.g., a line/edge that is perpendicular to a vector $\hat{\mathbf{n}}$.

Task 9 (B, 4p)

In natural 2D images $u_{i,j}$, the statistical distribution of image gradient magnitudes $m_{i,j} = |(\nabla u)_{i,j}|$ is long-tailed, i.e., most gradients are very small, but the distribution decays slower than e.g. a Gaussian distribution. Assume that $m_{i,j}$ is drawn from an exponential distribution

$$p(m_{i,j}) = \mu^{-1} \exp(-m_{i,j}/\mu)$$

where $\mu > 0$ is a contrast parameter (and the expectation of $m_{i,j}$). Assume further that the observed image value $g_{i,j}$ is measured with additive Gaussian noise with variance σ^2, i.e.

$$p(g_{i,j}|u_{i,j}) = (\sigma\sqrt{2\pi})^{-1} \exp(-(u_{i,j} - g_{i,j})^2/(2\sigma^2))$$

and that all image values are equally likely $p(g_{i,j}) = \text{const}$. The goal is now to find the image u with highest posterior probability, given the observation g

$$u^* = \arg\max_u p(u|g) \ . \tag{3}$$

a) Re-write $p(u_{i,j}|g_{i,j})$ using Bayes theorem[1] and the assumptions above (hints: collect unknown constants in one constant k; assume $p(u_{i,j}) = p(m_{i,j})$). (1p)
b) Re-write u^* using the result from a) under the assumption that all pixels are independent, i.e., that the joint distribution is the product of all distributions. (1p)
c) Re-formulate the maximum *a posteriori* problem in (3) for u^* as an energy minimization problem by taking the negative logarithm (hints: the energy should be a sum; re-collect constants for a meta-parameter λ). (1p)
d) Interpret the dependency of the meta parameter λ on σ and μ. What is the name of the energy model? (1p)

[1] $p(x|y) = p(y|x)p(x)/p(y)$

PART 4: Multiple View Geometry

Task 10 (A, 2p)

Assume that you have a set of corresponding image points $\{\mathbf{x}_k \leftrightarrow \mathbf{y}_k\}_1^K$ in two views, and a fundamental matrix $\mathbf{F}$ that relates them, such that $\mathbf{x}_k^T\mathbf{F}\mathbf{y}_k \approx 0\ \ \forall k \in [1, K]$.

a) Give the steps that compute the signed distances to the epipolar lines in the two images. (1p)

b) In the Gold Standard method to estimate $\mathbf{F}$, the *reprojection error* is used instead of the epipolar line distance. State what additional entities need to be calculated, and give an expression for the reprojection error given these additional entities. (1p)

Task 11 (A, 2p)

The BRISK feature uses a corner-based detector to find a canonical frame, and a binary descriptor to describe the region. The BRISK detector first finds corner *locations* in a scale pyramid and interpolates their response-magnitudes to find a corner *scale*. It then proceeds to find a dominant gradient *direction* that is associated with the direction $(1, 0)$ in the canonical frame of the patch.

a) Characterize the *geometric invariances* of this detector. (1p)

b) The BRISK descriptor forms a binary vector $b_1, \ldots, b_K$, from the sign of intensity differences for pixel pairs in a patch. This is done according to $b_k = (1 + \mathrm{sign}(I(\mathbf{p}_k) - I(\mathbf{p}_l)))/2$. Here $\mathbf{p}_k = (x_k, y_k, \sigma_k)$ are scale space coordinates, and the descriptor uses a pre-defined subset of pairs $(\mathbf{p}_k, \mathbf{p}_l)$ defined in the canonical frame. Characterize the *photometric invariances* of this descriptor. (1p)

Task 12 (B, 4p)

Structure from Motion (SfM) is typically based on bundle adjustment, and in this process it is critical that all measurements that participate are inliers.

a) Describe at least two distinct strategies to reduce the chances of an outlier correspondence being used. (2p)

b) The approach to SfM that is described in the course is *incremental*. Describe what this means in terms of how incremental SfM is implemented. Motivate why the incremental approach is useful in SfM. (2p)