Information page for written examinations at Linköping University

Examination date	2013-05-29
Room	U11
Time	14-18
Course code	TSBB15
Exam code	TEN1
Course name **Exam name**	Computer Vision, written examination
Department	Dept of EE (ISY)
Number of questions in the examination	12
Teacher responsible/contact person during the exam time	Per-Erik Forssén
Contact number during the exam time	013-285654
Visit to the examination room approx.	15.30
Name and contact details of the course administrator	Klas Nordberg 281634 klas@isy.liu.se
Equipment permitted	Dictionary Swedish-English-Swedish
Other important information	See instructions on the next page
Which type of paper should be used, cross-ruled or lined	Cross-ruled
Number of exams in the bag	11

Instructions TEN1

The final 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 (A, 2p) A Gaussian mixture model has the following form:

$$p(I) = \sum_{k=1}^{K} p(I|\Gamma_k)P(\Gamma_k) \quad \text{where} \quad p(I|\Gamma_k) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{-0.5(I-\mu_k)^2/\sigma_k^2}$$

a) Describe what $P(\Gamma_k)$ is (what is it called?), and also list any constraints on $\{P(\Gamma_k)\}_1^K$. (1p)
b) Gaussian mixture models are *generative models*. Describe how you can use the model above to generate new data, e.g. when $K = 2$. All mixture parameters are known, and you have a access to two random number generators: $g()$ that generates univariate normally distributed samples in $\mathcal{N}(0, 1)$, and $r()$ that generates uniformly distributed samples in $[0, 1]$. (1p)

Task 2 (A, 2p) In KLT tracking, the local displacement $\mathbf{d}$ between two images is found as:

$$\mathbf{d} = \mathbf{M}^{-1}\mathbf{e}\,, \text{ where } \mathbf{M} = \sum_{\mathbf{x}\in\mathcal{N}} w(\mathbf{x})\nabla a(\mathbf{x})\nabla a(\mathbf{x})^T\,,\ \mathbf{e} = \sum_{\mathbf{x}\in\mathcal{N}} w(\mathbf{x})\nabla a(\mathbf{x})(b(\mathbf{x}) - a(\mathbf{x}))$$

a) Write down the cost function that the above expressions minimise. (1p)
b) Which quantities should be updated in the next iteration, and how? (1p)

Task 3 (B, 4p) In standard Kanade-Lucas tracking, not all regions are equally easy to track. In particular, an effect known as *the aperture problem* is well known to cause problems.
a) Describe what the aperture problem is, and also describe what will happen to the motion estimate in a region where it is present. (1p)
Two common measures for classifying the local image strucure are:

$$a = \frac{\mathrm{tr}^2\mathbf{T} - 4\mathrm{det}\mathbf{T}}{\mathrm{tr}^2\mathbf{T} - 2\mathrm{det}\mathbf{T}} \quad \text{and} \quad b = \frac{2\mathrm{det}\mathbf{T}}{\mathrm{tr}^2\mathbf{T} - 2\mathrm{det}\mathbf{T}}$$

b) For regions without the aperture problem, one of the two measures is greater than the other. Which one, and why? (2p)
c) A common way to select good regions to track is to use the Harris operator. Describe in what way the difference between the two measures above relates to the Harris operator. (1p)

PART 2: Motion

Task 4 (A, 2p)
a) Explain the concepts *motion field* and *optical flow*. (1p)
b) Describe similarities and differences between the two. (1p)

Task 5 (A, 2p) Global estimation of optical flow leads to the minimization of a cost function that has a BCCE term and a smoothness term. If we omit the smoothness term, we obtain a non-trivial solution with zero error.
a) What is this solution? (1p)
b) Why is this solution not a useful result? (1p)

Task 6 (B, 4p) One way to determine optical flow is to formulate a cost function for a *local* region Ω as

$$\epsilon = \int_{\Omega} w(\mathbf{x}) \|\nabla_3 f \cdot \mathbf{v}\|^2 d\mathbf{x}, \qquad \text{where} \quad \nabla_3 f = \begin{pmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \frac{\partial f}{\partial t} \end{pmatrix}^T \qquad \text{and} \quad \mathbf{v} \sim \begin{pmatrix} v_1 & v_2 & 1 \end{pmatrix}^T,$$

and where $\sim$ denotes equality up to scale. We assume that $\nabla_3 f$ is known at all points. To estimate (v_1, v_2) we can minimize ϵ with an additional constraint on $\mathbf{v}$.
a) How is (v_1, v_2) determined if the third element of $\mathbf{v}$ is set equal to 1? (1p)
b) How is (v_1, v_2) determined if $\|\mathbf{v}\| = 1$? (1p)
c) How do the two approaches compare in terms of numerical stability for the case of moving edges/lines? (1p)
d) The second approach has the advantage of providing additional information about the solution. What information? (1p)

PART 3: Denoising

Task 7 (A, 2p) Three different diffusion filtering methods have been introduced in the lecture on image enhancement. To characterize three different methods, at least two properties are required. Draw a table with the names of the three methods (1p) and at least two properties and indicate which of the methods fulfill which of the properties, such that all three methods have different sets of properties (1p).

Task 8 (A, 2p) Inhomogeneous diffusion is applied to an image to reduce noise. This method computes a type of scale space $L(x, y, s)$ from the image, using the equation

$$\frac{\partial}{\partial s} L = \frac{\mu}{2} \nabla^2 L,$$

where the diffusion speed $\mu = \mu(x, y)$ is controlled by the local image content at point (x, y).

a) Give a general characterization of how μ depends on the local image structure at (x, t). Motivate your answer. (1p)

b) Perona and Malik have suggested a simple formulation of μ in terms of local image content, that is consistent with this characterization. Describe their choice of μ. (1p)

Task 9 (B, 4p) In natural 2D images, the magnitude of the gradient $m = |\nabla u|$ is long-tailed. Assume that the probability density of m is modeled according to

$$p(m) \propto \exp(-m/\mu)$$

where $\mu > 0$ is a contrast parameter (the expectation of m). Assume further that the observed image g is measured with additive Gaussian noise with variance σ^2 and that all image values are equally probable. The goal is now to find the image u that is globally most probable, given the observation g

$$u^* = \arg\max_u p(u|g) \ .$$

a) Re-write $p(u|g)$ using Bayes theorem[1] and the assumptions above (hints: collect unknown constants in one constant and assume all pixels being independent). (1p)

b) Re-formulate the resulting term as an energy term (hint: this should be a sum). (1p)

c) Re-formulate the maximum probability problem above as an energy minimization (hint: re-collect constants). (1p)

d) What is the name of the resulting energy model? (1p)

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

PART 4: Multiple View Geometry

Task 10 (A, 2p) You are using RANSAC for robust estimation of a homography between two images, using a correspondence set that contains outliers. For each iteration, RANSAC picks 4 putative pairs of corresponding points from the two images and estimates a homography from them.

Why is 4 pairs used here? Why not use more than 4 point pairs for the estimation of the homography? (2p)

Task 11 (A, 2p) Two images I and J depict exactly the same view of a scene, but under different illuminations. You have computed SIFT features at position $\mathbf{x}$ in both images.
a) The SIFT features will be identical if $I = 0.75J$, but which step in the algorithm makes this happen? (1p)
b) Will the feature descriptors also be the same if $I = 0.75J + 0.25$? Motivate! (1p)

Task 12 (B, 4p) You have determined the essential matrix $\mathbf{E}$ between two camera matrices that differ in pose by a rigid transformation given by a rotation $\mathbf{R}$ and a translation $\bar{\mathbf{t}}$. They are related as

$$\mathbf{E} \sim \mathbf{R}^T [\, \bar{\mathbf{t}} \,]_\times$$

a) How is $\bar{\mathbf{t}}$ related to $\mathbf{E}$, i.e., how can you determine $\bar{\mathbf{t}}$ from $\mathbf{E}$? (1p)

b) Given that $\mathbf{E}$ is known, you can determine two distinct rotations $\mathbf{R}_1$ and $\mathbf{R}_2$ and two directions of $\bar{\mathbf{t}}$, that are consistent with $\mathbf{E}$ in the above equation. Why is $\bar{\mathbf{t}}$ not fully determined from $\mathbf{E}$? (1p)

c) How can you determine which of the two rotations to use, and the direction of $\bar{\mathbf{t}}$? Motivate your answer (e.g. with a top-down illustration). (2p)