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. If you answer the whole exam, the score will be converted to a final score according to: `p_final = (p_all_parts-24)*2+16`

In order to pass with grade 3, at least 12fp are required.

In order to pass with grade 4, at least 19fp are required.

In order to pass with grade 5, at least 23fp are required.

Re-use of the midterm examination result: If you have written the midterm examination and obtained bonus points, these may be used by not handing in answers to parts 1 and 2. Your final score is then computed according to: `p_final = p_bonus + (p_part3+p_part4-12)*2+8`.

Be brief and to the point when answering. You are allowed to hand in **at most one page per question**. However, more brief answers are recommended. If you include excessive amounts of irrelevant information in your answer, this indicates lack of understanding, and leads to deduction of points.

Write your name, and personal number (or LiU-ID) on the cover sheet, and attach it before the other sheets that you hand in. In case you do not have access to a printer, you may write the same information on a sheet of paper instead.

Good luck!

Per-Erik Forssén, Mårten Wadenbäck and Michael Felsberg

Questions about the tasks in the exam are answered in the following Zoom room:
`https://liu-se.zoom.us/j/65168610393`

Questions are answered 15.00-15.30 and 16.30-17.00 on the exam day.

TSBB15 Computer Vision
Final Exam

August 18, 2020. 14:00 – 18.00
electronic submission through Lisam

Print out this cover page. Then fill in your name, and personal number. Indicate solved tasks with a "×" in the "Task solved" row below. When handing in the exam, attach this page in front of the others. In case you do not have access to a printer, you may write the same information on a sheet of paper instead.

Name:

Personal Number/LiU-ID:

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Task solved:												
Score:												

Total score:

Grade:

PART 1: Tracking

Task 1 (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, $\mathbf{d}$ is the sought displacement, Ω is the region being tracked, and $\rho(\cdot)$ is the used error norm. Compare the cases $\rho_{\mathrm{LK}}(\mathbf{x}) = \|\mathbf{x}\|^2$ (Lucas-Kanade), and $\rho_{\mathrm{MA}}(\mathbf{x}) = \|\mathbf{x}\|$.
a) What improvements can be expected by switching from ρ_{LK} to ρ_{MA}, and why? (1p)
b) The LK update does not work for ρ_{MA}. Suggest a way to minimize (1) for the ρ_{MA} case. (1p)

Task 2 (A, 2p)
When evaluating tracking performance, a predicted bounding box $\mathcal{Q}$ is compared with the ground truth bounding box $\mathcal{G}$. This is normally done using the Jaccard index:

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

Here $\cap$ is the set intersection operator, $\cup$ is set union, and $|\mathcal{A}|$ is the area of set $\mathcal{A}$. The result of (2) is a value $J \in [0,1]$. Two other options are bounding box recall, R, and bounding box precision, P:

$$R(\mathcal{Q},\mathcal{G}) = \frac{|\mathcal{Q}\cap\mathcal{G}|}{|\mathcal{G}|}\,,\ \text{and}\ P(\mathcal{Q},\mathcal{G}) = \frac{|\mathcal{Q}\cap\mathcal{G}|}{|\mathcal{Q}|}\,. \tag{3}$$

a) If either of these measures is used instead of J it is possible for the tracker to cheat. Explain how. (1p)
In project 1, true positives(TP) false positives(FP) and false negatives(FN) were defined by counting bounding boxes that had large and small overlaps. It is also possible to define these counts on a pixel level, and this has been used to define the measures in (3).
b) Define TP, FP and FN as areas using the bounding boxes $\mathcal{Q}$,$\mathcal{G}$, arithmetic operators, and set operators. (1p)

Task 3 (B, 4p) The *Expectation Maximization* (EM) algorithm can be used to fit a Gaussian mixture model (GMM) to data. In the multivariate case a GMM is written as:

$$p(\mathbf{x}|\boldsymbol{\Theta}) = \sum_{k=1}^{K} w_k (2\pi)^{-D/2} \det(\boldsymbol{\Sigma}_k^{-\frac{1}{2}}) e^{-0.5(\mathbf{x}-\boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}(\mathbf{x}-\boldsymbol{\mu}_k)} \tag{4}$$

where $\boldsymbol{\Theta} = (w_1, \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \ldots, w_K, \boldsymbol{\mu}_K, \boldsymbol{\Sigma}_K)$ are the mixture parameters, K is the number of mixtures, and D is the dimensionality of the data, i.e. $\mathbf{x} \in \mathbb{R}^D$. EM can be seen as a generalization of the K-means algorithm. For each type of parameter $(w, \boldsymbol{\mu}, \boldsymbol{\Sigma})$:
a) Describe what the parameter corresponds to in the K-means algorithm. (2p)
b) Describe in what way the parameter generalizes EM compared to K-means. (2p)

PART 2: Motion

Task 4 (A, 2p)
We have seen in the lectures that the rank of the 3D *spatio-temporal structure tensor* $\mathbf{T}_{3\mathrm{D}}$ reveals important information about the local signal.

Explain why the case $\operatorname{rank} \mathbf{T}_{3\mathrm{D}} = 3$ cannot be consistent with the BCCE[1]. (2p)

Task 5 (A, 2p)
The *aperture problem* occurs when the local signal is intrinsically 1D (e.g. a line or an edge), and leads to the Lucas–Kanade equation not having a unique solution. In such cases, we have suggested computing the *normal velocity*, which gives one particular solution among infinitely many.

a) Apart from yielding a unique and well-defined solution, name another desirable property that the normal velocity possesses. (1p)

b) Compute the normal velocity $\mathbf{d} = (d_1, d_2)$ at $\mathbf{x}_0$, if the Lucas–Kanade equation at $\mathbf{x}_0$ is

$$\underbrace{\begin{pmatrix} 16 & 8 \\ 8 & 4 \end{pmatrix}}_{\mathbf{T}(\mathbf{x}_0)} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = \underbrace{\begin{pmatrix} 2 \\ 1 \end{pmatrix}}_{\mathbf{s}(\mathbf{x}_0)} . \tag{5}$$

HINT: Note that $\mathbf{T}(\mathbf{x}_0) = \begin{pmatrix} 4 \\ 2 \end{pmatrix} \begin{pmatrix} 4 & 2 \end{pmatrix}$. (1p)

Task 6 (B, 4p)
The continuous time version of the BCCE is

$$\frac{\partial I}{\partial t} + \frac{\partial I}{\partial u} v_1 + \frac{\partial I}{\partial v} v_2 = 0, \tag{6}$$

where $\mathbf{v} = (v_1, v_2)$ is the velocity vector at (u, v, t). This can be used to derive a continuous time version of the Lucas–Kanade equation, which can be written as $\mathbf{T}(\mathbf{x})\,\mathbf{v} = \mathbf{s}(\mathbf{x})$, where

$$\mathbf{T}(\mathbf{x}) = \int_{\Omega_0} w(\mathbf{y}) \nabla I(\mathbf{x} + \mathbf{y}) \nabla^\top I(\mathbf{x} + \mathbf{y})\, d\mathbf{y} \tag{7}$$

and

$$\mathbf{s}(\mathbf{x}) = -\int_{\Omega_0} w(\mathbf{y}) \frac{\partial I}{\partial t} \nabla I(\mathbf{x} + \mathbf{y})\, d\mathbf{y}. \tag{8}$$

a) What cost function leads to this formulation? (1p)

b) Show how the expressions for $\mathbf{T}(\mathbf{x})$ and $\mathbf{s}(\mathbf{x})$ are derived from the cost function in a). (3p)

[1] BCCE = Brightness Constancy Constraint Equation

PART 3: Denoising and Multi-View Stereo

Task 7 (A, 2p)
When building 3D models from images it is common to first solve for sparse 3D points, and then do densification to obtain a point cloud with more 3D points. Many densification methods use two views, but some use more.
a) Explain one advantage and one distadvantage with using only two views. (1p)
b) After densification, meshing is usually performed. Explain two advantages with a mesh compared to a dense point cloud. (1p)

Task 8 (A, 2p)
In image enhancement, the locally adapted filter $g_{\mathbf{x}}$ at position $\mathbf{x}$ can be written as a linear combination of a low-pass filter g_{LP} and a set of fixed high-pass filters $g_{\mathrm{HP},k}$

$$g_{\mathbf{x}} = g_{\mathrm{LP}} + \sum_{k=1}^{N} g_{\mathrm{HP},k} \cdot c_{\mathbf{x},k} \tag{9}$$

where the coefficients $c_{\mathbf{x},k}$ depend on the image position $\mathbf{x}$. This formulation of $g_{\mathbf{x}}$ makes it possible to implement the image enhancement operation in terms of a number of convolutions between the image and some filters followed by a linear combination of the convolution results.
a) Why is $g_{\mathbf{x}}$ implemented this way instead of shift-variant filter? (1p)
b) Why is only the high-pass part used in the linear combination? (1p)

Task 9 (B, 4p)

Many optimization problems are formulated using an energy functional of the form:

$$\varepsilon(I) = \langle \texttt{data\_term} \rangle + \lambda \, \langle \texttt{smoothness\_term} \rangle \tag{10}$$

The absolute error is less sensitive to outliers than the quadratic error, which is why regularization (smoothness) using the L_1-norm is often considered to be preferable to a quadratic smoothness term $f_x^2 + f_y^2$.
a) Write down the L_1-regularization term as it is commonly used in variational methods for image denoising. (1p)
b) What is the name of the variational denoising model consisting of the term from a) and a quadratic data term (fidelity term)? (1p)
c) What is the effect of changing from the quadratic smoothness to L_1-smoothness at edges? (1p)
d) What happens with the result if the meta-parameter λ is very large? (1p)

PART 4: Multiple View Geometry

Task 10 (A, 2p)
There are several different algorithms that can estimate a fundamental matrix $\mathbf{F}$ from a set of input correspondences $\{\mathbf{x}_k \leftrightarrow \mathbf{y}_k\}_1^K$. We will now consider two such methods:
Method A: Define equations $\mathbf{x}_k\mathbf{F}\mathbf{y}_k = 0$ for each correspondence, and solve the resulting equation system to find $\mathbf{F}$.
Method B: Define a cost function

$$\epsilon(\{\mathbf{z}_k\}_1^K, \mathbf{C}_2) = \sum_{k=1}^{K} d_{\mathrm{PP}}(\mathbf{x}_k, \mathbf{C}_1\mathbf{z}_k)^2 + d_{\mathrm{PP}}(\mathbf{y}_k, \mathbf{C}_2\mathbf{z}_k)^2 \tag{11}$$

where $\mathbf{C}_1 = [\mathbf{I}|\mathbf{0}]$ is constant, and $\mathbf{C}_2$ and $\{\mathbf{z}_k\}_1^K$ are unknowns that have been properly initialized. Minimize (11) with respect to the unknowns, and then compute $\mathbf{F} = [\mathbf{e}_1]_\times \mathbf{C}_1\mathbf{C}_2^\dagger$, where $\mathbf{e}_1$ is the epipole in camera 1.
a) Explain which of the two methods is better to use inside a RANSAC loop. (1p)
b) What principle is used to derive method B? (1p)

Task 11 (A, 2p)
In a *linear response camera*, an exposure change corresponds to a scaling of the intensity. However, most cameras also have a so called *gamma correction* step, that maps the intensity as $I_\gamma(\mathbf{x}) = I(\mathbf{x})^{1/\gamma}$.
a) What will happen to MSER detections if gamma is changed from $\gamma = 1$ to $\gamma = 2$? (1p)
b) Is the magnitude of the image gradient invariant to exposure change if $\gamma = 1$? Explain! (1p)

Task 12 (B, 4p)
An initial structure from motion solution can be refined by a technique called *bundle adjustment* (BA).
a) Write down the BA 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) The Jacobian matrix used in BA iterations is normally very sparse. Sparsity of a matrix can be measured as the ratio of non-zero elements divided by the total number of elements. Write down an expression of the sparisty of the Jacobian for a BA problem with N 3D points, K cameras, and R residuals. Also motivate any constants in the expression. (2p)