Date for the exam	2009-03-12
Room	TER1
Time	8:00–12:00
Course code	TSBB12
Exam code	TEN1
Course name	Computer Vision Theory
Department	ISY
Number of tasks	12
Number of pages	6
Examiner	Michael Felsberg
Watching teacher	Michael Felsberg
Telephone	013-282460
Visits exam	9:00
Allowed means	no
Misc	See instructions on the next page

Instructions TEN1

The exam consists of 4 parts, each corresponding to one of the four lab exercises of TSBB12.

- Part 1 covers tracking
- Part 2 covers motion
- Part 3 covers denoising
- Part 4 covers stereo

Each part contains 3 tasks, two that require description of terms, phenomenon, 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 16p are required.

In order to pass with grade 4, at least 23p are required.

In order to pass with grade 5, at least 28p are required.

Tasks of type A are to be answered directly on the exam sheets, tasks of type B are to be answered on a separate sheet that is to be attached.

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

Good luck! Michael Felsberg, Klas Nordberg, and Per-Erik Forssén

PART 1: Tracking

Task 1 (A, 2p) Local displacements between two image regions can be estimated, for example, using block matching or based on the Kanade-Lucas tracking method over multiple scales, or a combination of the two methods. Describe characteristics of these two approaches where they differ, as well as why and how they can be combined in a suitable way.

Task 2 (A, 2p) In standard Kanade-Lucas tracking, we minimize the error

$$\varepsilon = \|f_1(\mathbf{x}) - f_2(\mathbf{A}(\mathbf{x}))\|^2$$

where f_1 and f_2 are image regions in the image and in the template, and $\mathbf{A}$ is some transformation of the local coordinate system, in the simplest case only translation. The minimization is made over $\mathbf{A}$.

This formulation is not symmetric relative f_1 and f_2, which may result in inconsistent estimates of the transformation $\mathbf{A}$. How can we formulate a symmetric error function?

Task 3 (B, 4p) The KLT tracker makes use of a continuous interpolation $f(x, y)$ of the discrete image $b_{m,n}$, and its x- and y- derivatives f_x and f_y to find displacements with sub-pixel accuracy. For best results, the algorithm should use regularised derivatives. Explain why. Also derive the analytic expression for a regularised x-derivative when using a 2D Gaussian kernel.
Instead of Gaussian regularization, f (also f_x and f_y) can be obtained by minimizing the functional

$$\varepsilon(f) = \frac{1}{2}\int_\Omega \left(f(x,y) - \sum_{m,n} \delta(x-m, y-n) b_{m,n} \right)^2 + \lambda |\nabla f|^2 \, dx \, dy$$

where $\delta(x - m, y - n)$ is the 2D delta function at (m, n). What is the Euler-Lagrange equation $L_f - \sum_w \partial_{x_w} L_{f_{x_w}} = 0$ for this functional? What is the fix-point solution (hint: the Fourier transform of $\sum_{m,n} \delta(x - m, y - n)$ is again a sum of impulses $\sum_{k,l} \delta(u - k, v - l)$)? Please consider this last question as more difficult. The result may be presented neglecting constant factors.
WRITE YOUR ANSWERS ON A SEPARATE SHEET

PART 2: Motion

Task 4 (A, 2p) In some situations, moving objects can also be detected using background modelling. Describe when background modelling can be used to detect moving objects. Also describe the model used to represent the background, and how it is used to determine what is foreground.

Task 5 (A, 2p) Motion analysis in image sequences is sometimes based on the assumption that the 3D motion vectors of 3D points in the scene, when projected onto the image plane generates a *motion field* which can be determined. This assumption is not correct in general. Give two examples of why we cannot determine the motion field from measurements on the image sequence.

Task 6 (B, 4p) The brightness constancy constraint equation (BCCE) or optical flow equation is written as

$$\frac{\partial I}{\partial t} + v_x \frac{\partial I}{\partial x} + v_y \frac{\partial I}{\partial y} = 0$$

where I is the image intensity as a function of spatial coordinates (x, y) and time t, and $\mathbf{v} = (v_x, v_y)$ is the local optical flow vector.

- Describe a method that determines $\mathbf{v}$ based on local measurements on the image signal I. Give clear account on which computational step are needed to produce $\mathbf{v}$ as a function of (x, y, t).
- Describe how certainty measures of the estimate can be provided. Describe in detail at least two cases for the local signal which result in uncertainties in the estimate of $\mathbf{v}$ for your method.

WRITE YOUR ANSWERS ON A SEPARATE SHEET

PART 3: Denoising

Task 7 (A, 2p) Denoising based on adaptive filtering or on anistropic diffusion uses local information to control the image filtering. What type of local information is being used and in how does it control the filtering? Be specific in terms of how characteristic cases of the local image region correspond to particular types of filtering in that region.

Task 8 (A, 2p) Removal of noise can be done by means of low-pass filtering, which corresponds to computing local averages of the filtered signal. This can be done on the image signal or on features estimated from the signal, such as the structure tensor. Why is it not a good idea to do LP-filtering of the image and how can it be motivated that LP-filtering of the structure tensor is good idea?

Task 9 (B, 4p) The basic equation for adaptive diffusion is

$$\frac{\partial}{\partial s} L = \frac{1}{2} \mathrm{div}(\mathbf{D}\, \mathrm{grad} L)$$

- Explain what L and $\mathbf{D}$ are. Which primary variables do they map from and what are their ranges?
- What does $\frac{\partial}{\partial s} L$ mean in practice? Explain what we do with this information once it has been estimated in order to achieve anisotropic diffusion.
- How is $\mathbf{D}$ related to $\mathbf{T}$, the local orientation tensor? Be specific in terms of how the eigensystem of $\mathbf{D}$ is related to that of $\mathbf{T}$.
- How can the right hand side of the above equation be simplified (from a practical point of view) given that $\mathbf{T}$ varies slowly over the image?

WRITE YOUR ANSWERS ON A SEPARATE SHEET

PART 4: Stereo

Task 10 (A, 2p) The epipolar constraint describes a *necessary* but not *sufficient* condition for stereo image points to correspond to the same 3D point. Why is the constraint not sufficient? Draw a figure of a 3D scene with the two image planes and corresponding camera centers. Describe two *distinct* 3D points such that if one of them is projected into the first camera and the other in the second camera, the corresponding image points satisfy the epipolar constraint. Describe a general condition on the 3D points such that their image coordinates satisfy the epipolar constraint even if the 3D points are distinct.

Task 11 (A, 2p) In wide-baseline stereo, image patches sampled in canonical frames are used to find correspondences between two images. One particular way to find canonical frames is the SIFT detector. Describe the geometric invariances in the canonical frames used by SIFT. Also describe how these geometric invariances are found.

Task 12 (B, 4p) The fundamental matrix can be estimated together with a set of corresponding stereo image points based on the RANSAC algorithm. This algorithm is indeterministic and iterative, i.e., it produces the correct result only with a certain probability given that a certain number of iterations has been made.

We have a stereo image pair which is uncalibrated, i.e. the fundamental matrix is not known but we know that one exists. We also have a set of interest points that has been found in each image from which N hypothetical correspondences can be formed, each in the form of a point pair. Only M of these are correspondences that satisfy the epipolar constraint defined by the correct fundamental matrix. RANSAC is iterated P times and then stops and a well-known standard method is used to estimate the fundamental matrix in each iteration. What is the probability that a correct estimate of the fundamental matrix is found when the algorithm stops? Assume that N is very large, for example, relative to M. Use reasonable approximations.

WRITE YOUR ANSWERS ON A SEPARATE SHEET