PART 1: Tracking

Task 1 (A, 2p) A stationary surveillance camera is imaging a section of a road in Cameroon. During the day, the road is periodically shadowed by a tree. This being in winter, there are hardly any clouds in the sky. Also note that as Cameroon is by the equator, nightfall and sunrise may be seen as instantaneous.

a) Write down a mixture distribution $p(I)$ for the intensity of a pixel on the road. Explain, and name all entities in your expression.

b) Lets say that the mixture components above are Gaussians. Interpret the mean and the standard deviation of each such Gaussian in terms of the input signal.

Task 2 (A, 2p) Consider KLT tracking of a 1D signal. The KLT tracker is based on the Taylor expansion of the shifted image $f(x+d)$. Now, instead make a Taylor expansion of the blurred image $(f*g)(x)$, where $g(x)$ is a Gaussian filter.

a) One of the terms in the expansion has the form $\frac{\partial}{\partial x}(f*g)(x)$. Derive the filter that can be used to compute these derivatives.

b) The solution for the displacement can be written $d = (I(x) - J(x))/G(x)$, where $G(x)$ is the output of the filter you just derived, but what are I and J?

Task 3 (B, 4p) The KLT tracker makes use of a continuous interpolation $c(x, y)$ of the discrete image $b_{m,n}$, and its x- and y- derivatives c_x and c_y to find displacements with sub-pixel accuracy. Usually, the interpolations are computed using 2D Gaussian kernels and derivatives, see Task 2. Instead of interpolating with a Gaussian kernel, c (also c_x and c_y) can be obtained by minimizing the functional

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

where $\delta(x-m, y-n)$ is the 2D delta function at (m, n). What is the Lagrangian $L(c, c_{x_w})$ of this functional (1p)? What is the Euler-Lagrange equation $L_c - \sum_w \partial_{x_w} L_{c_{x_w}} = 0$ for this functional (1p)? What is the effect of λ (1p)? What is the fix-point solution in the Fourier domain (hint: the continuous Fourier transform of $\delta(x-m, y-n)$ is a complex oscillation $\exp(-i2\pi(mu+nv))$) (1p)?

PART 2: Motion

Task 4 (A, 2p) Local displacements of regions in an image sequence can be estimated, for example, using block matching or based on the Lucas-Kanade equation over multiple scales, or a combination of the two methods. Generate a table with characteristics of these two approaches where they differ. Discuss why and how they can be combined in a suitable way.

Task 5 (A, 2p) The optical flow equation, or BCCE, is formulated as

$$\frac{\partial f}{\partial x_1} v_1 + \frac{\partial f}{\partial x_2} v_2 + \frac{\partial f}{\partial t} = 0$$

and can be seen as a constraint on the motion vector (v_1, v_1). This equation is, itself, based on a specific assumption (or constraint) on image intensities. Which assumption is this (1p)? Give two practical examples when this assumption is not valid (1p).

Task 6 (B, 4p) Consider the BCCE from Task 5 again and assume the situation where $f(x_1, x_2) = ax_1 + bx_2 + c$, with $a, b, c > 0$, and $\frac{\partial f}{\partial t} = d$. We shall now try to find out what BCCE tells us about $\mathbf{v} = (v_1, v_2)^T$!

a) Express BCCE in terms of a, b, c, d! If you plot all the valid solutions for $\mathbf{v}$ in the image plane, what do you get? (1.5p)
b) Which is the shortest velocity vector that satisfies BCCE? (This is called the normal velocity.) (1p)
c) In the general case, we locally integrate the square of the left hand side of the BCCE and minimise over (v_1, v_2). Which assumption about the motion vector do we make in this case? Give a practical example when this assumption is not valid. (1.5p)

PART 3: Denoising

Task 7 (A, 2p) The original idea by Perona & Malik on *inhomogeneous diffusion* implies that the image is filtered by a Gaussian filter of position dependent size, such that it locally depends on the presence of oriented structures in the image. Describe a typical situation when this method fails to effectively remove image noise and explain why.

Task 8 (A, 2p) In image enhancement, the local orientation tensor $\mathbf{T}(\mathbf{x})$ is mapped to a control tensor $\mathbf{C}(\mathbf{x})$ in such a way that if the norm of $\mathbf{T}(\mathbf{x})$ is very small then $\mathbf{C}(\mathbf{x})$ is approximately set to zero, i.e., in this case the adapted filter attenuates the high frequency content of image signal. Explain why.

Task 9 (B, 4p) The basic equation for non-linear diffusion is

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

where L is an image, s is the diffusion time, and μ usually depends on $|\mathrm{grad} L|$. In practice, L and s are discrete.

- Give an example for $\mu(|\mathrm{grad} L|)$.
- Expand the div-operator for your choice of μ above.
- What are the usually used discrete operators for the derivatives of L?
- How to discretize s, i.e., what does one iteration look like?

PART 4: Stereo

Task 10 (A, 2p) When the RANSAC algorithm is used for determining point correspondences between two images, it is often combined with a pre-processing procedure that removes hypothetical correspondences that have a low probability of being correct. Describe how to implement such a procedure and explain why it is critical to apply this produce before RANSAC is made.

Task 11 (A, 2p) Imagine that you have two image patches $I(\mathbf{x})$ and $J(\mathbf{x})$ extracted from two images. The patches depict the same 3D location, and are sampled in canonical frames given by the MSER detector. The two images were imaged using slightly different camera settings, and thus $J(\mathbf{x}) \approx k_1 I(\mathbf{x}) + k_2$ where k_1 and k_2 are unknown. Describe a way to convert I and J to normalised patches $\hat{I}$ and $\hat{J}$, that are approximately equal.

Task 12 (B, 4p) Given two images taken by a moving camera we can in some cases obtain the corresponding essential matrix $\mathbf{E}$. It contains information about the rotation and translation of the camera between the two images. This information, however, is not unique.

a) Describe which ambiguities there are for the rotation and translation of the camera, given that $\mathbf{E}$ is determined. (1p)
b) Describe how to resolve some of these ambiguities. (1p)
c) Describe what additional information about the camera is required in order to determine $\mathbf{E}$ from the images. (1p)
d) How do you obtain this additional information? (1p)