Exam in Image analysis, SSY096, April 4, 2016

Allowed materials: Pencil, eraser.

The exam consists of six problems. Make sure that you have them all.

- Motivate all answers carefully.
- Use a new paper for each new numbered problem.
- Write your anonymous code on each page.
- Avoid using a red pen.

Grades

≥ 8 **points** Grade: 3

≥ 11 **points** Grade: 4

≥ 14 **points** Grade: 5

1 SIFT, 3 points

(a) A Sift-like descriptor (as the one in Lab 1) was computed for the image patch below. The regions used are indicated with grey dashed lines so these lines are not part of the actual image. Work out roughly how the final descriptor will look. You can either present it using vector bouquets or describe the actual elements in the descriptor vector. No exact numerical values (or vector lengths) are required, but try to get the relationships roughly correct. The structure of the gradient histograms is indicated below for reference.

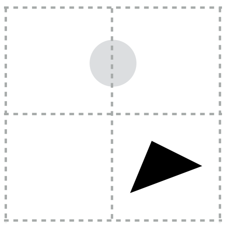
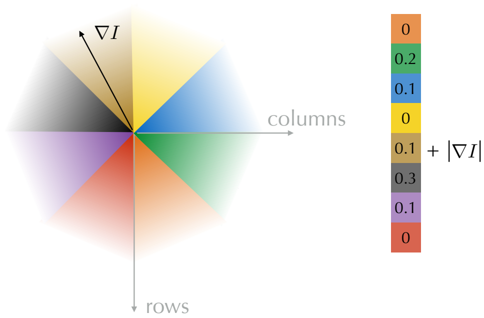
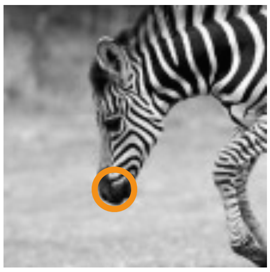


Image patch.



(b) The most common way to use Sift is to produce Sift features. Consider the two images below. A difference-of-gaussians detector has been used to detect interest points indicated with orange circles. Explain how Sift-like features handle scale differences such as this one. In other words how come these two points are assigned similar Sift-like descriptors?



2 Image registration, 3 points

Let's say we want to warp a source image I_s to a target image I_t.

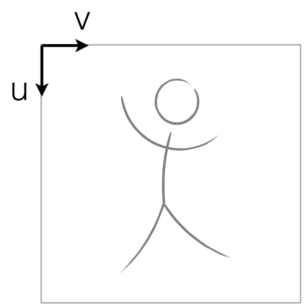
$$I_s = \begin{pmatrix} 18 & 6 & 29 & 11 & 10 \\ 12 & 1 & 23 & 30 & 3 \\ 26 & 15 & 9 & 4 & 24 \\ 14 & 15 & 22 & 8 & 19 \\ 6 & 28 & 21 & 27 & 20 \end{pmatrix} \quad I_t = \begin{pmatrix} 10 & 2 & 18 & 29 \\ 11 & 12 & 8 & 21 \\ 6 & 23 & 5 & 14 \\ 20 & 30 & 24 & 19 \\ 1 & 7 & 16 & 15 \end{pmatrix}. \tag{1}$$

We have used Ransac to estimate the appropriate transformation,

$$\begin{pmatrix} \bar{u} \\ \bar{v} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} + \begin{pmatrix} -1 \\ 1 \end{pmatrix}. \tag{2}$$

(a) Use this transformation to compute the warped image, I_w. Set undefined values to 0.

(b) Consider warping the source image below using the same transformation. How will this affect the stick figure? Make a sketch, or describe in words.



(c) Assume that we use the Sift features and Ransac to align two microscopy images. The images show the same tissue sample but at different magnifications. What is the appropriate transformation type? What are the disadvantages of using an affine transformation instead?

3 Triangulation, 3 points

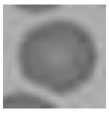
Given a set of Sift points with pixel coordinates (u_i, v_i) and corresponding camera matrices P_i we want to triangulate a 3D point X using Ransac.

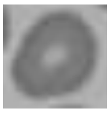
(a) Explain how to construct minimal solver for this problem. Show how to get a system $M\theta = b$ form, (where θ are all the unknowns) and explain how it can be solved in Matlab. Be sure to define all variables that you use in your explanation. (There are multiple correct ways to do this. Choose one.)

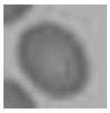
(b) An image is taken of a 50×50 cm painting. The image has 1000×1000 pixels. The camera has focal length, $f = 90\,\text{mm}$, pixel size, $\varrho = 0.01\,\text{mm}$ and principal point in the middle of the image. At what distance does the painting fit perfectly inside the image?

4 Statistical learning, 3 points

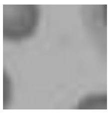
(a) Consider a cell classifier[1] producing a single output probability p. Below we see four positive cell examples and four negative examples together with the respective ouput probabilities. Use these values to write down a formula for the negative log-likelihood loss over this training set.


$p = 0.9$

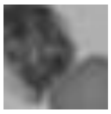

$p = 0.9$


$p = 0.8$


$p = 0.8$


$p = 0.1$

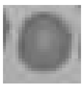

$p = 0.1$


$p = 0.3$


$p = 0.2$

(b) Consider comparing four convolutional neural networks for cell detection. The difference is the number of filters in the convolutional layers. With 5 filters we get a loss of 0.6 with 10 filters a loss of 0.4, with 30 filters a loss of 0.3 and with 50 filters a loss of 0.25. Why shouldn't we use these values to choose the best classifier? How would you normally make this choice?

(c) Explain what data augmentation is and how it can be used when training a cell classifier. How would this differ from data augmentation when training a classifier for handwritten digits?

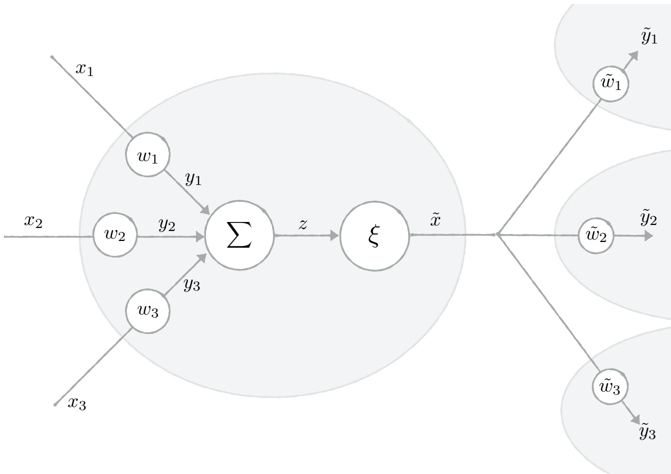


[1]as the one in Lab 4

5 Backpropagation, 3 points

The figure below shows a small part of a neural network. Here, ξ denotes the logistic sigmoid function

$$\xi(z) = \frac{e^z}{1 + e^z}. \tag{3}$$



Consider training the weights of this network using stochastic gradient descent on the loss function,

$$L(\theta) = \sum_{i=1}^{n} L_i(\theta), \tag{4}$$

where θ are all the weights. The current values are

$$w_1 = 1, \quad w_2 = -1, \quad w_3 = 2, \quad \tilde{w}_1 = 2, \quad \tilde{w}_2 = 4, \quad \tilde{w}_3 = 1. \tag{5}$$

When training a neural network backpropagation is used to find the gradient of the partial loss ∇L_i, but first a forward pass is performed to find all intermediate values in the network. This gave

$$x_1 = 2, \quad x_2 = 4, \quad x_3 = 1, \quad z = 0, \quad \tilde{x} = 0.5, \tag{6}$$

To find all the partial derivatives in ∇L_i we move backwards through the network. We have computed

$$\frac{\partial L_i}{\partial \tilde{y}_1} = 1, \quad \frac{\partial L_i}{\partial \tilde{y}_2} = 1, \quad \frac{\partial L_i}{\partial \tilde{y}_3} = 2. \tag{7}$$

(a) Derive a formula for

$$\frac{\partial L_i}{\partial w_1}. \tag{8}$$

Motivate your answer carefully.

(b) Compute a numerical value for

$$\frac{\partial L_i}{\partial w_1}. \tag{9}$$

(c) With a learning rate of 0.01, what would the next value for w_1 be?

6 Robust model estimation, 3 points*

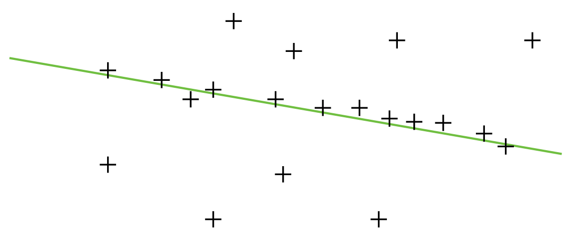
Consider line fitting from n noisy measurements $(u_1, v_1), \ldots, (u_n, v_n)$. A line can be parameterized with an angle α and a constant, b. The absolute residual

$$|r_i(\theta)| = |u_i \cos\alpha + v_i \sin\alpha + b| \tag{10}$$

is the perpendicular distance from the line to a point.[2] An outlier is a measurement such that

$$|r_i(\theta)| > \tau. \tag{11}$$

We want to find a line that minimizes the number of outliers.



(a) Show that there exists an optimal θ such that at least two absolute residuals are exactly equal to τ.

(b) Suggest an algorithm to find a line that minimizes the number of outliers.

[2]As usual θ are all the unknowns, in this case α and b.