Exercises (Short Answers)
Computer Vision 1, Master AI

1 Exercises (Lectures 2 and 3)

EXERCISE 1:
To calculate the color of light sources, the following intuitive color models are used: intensity I, chromaticity xy, hue H and saturation S. Let's assume, for simplicity reasons, that sunlight S is given by $X = Y = Z = 100$. Further, let $X = 100$, $Y = 100$ en $Z = 150$ be the values for a given artificial lamp A.

- (Q.a) Calculate the intensity I of the two light sources S and A.
- (A.a) $I_S = X + Y + Z = 300$, $I_A = 350$
- (Q.b) Calculate the chromaticy values $x = X/(X + Y + Z)$, $y = Y/(X + Y + Z)$ and plot these in the chromaticity diagram.
- (A.b) $x_S = \frac{1}{3}$, $y_S = \frac{1}{3}$, $x_A = 0.286$, $y_A = 0.286$.
- (Q.c) What the estimated hue of the object color B at $X = 120$, $Y = 100$ and $Z = 100$ with reference white light S.
- (A.c) NA.
- (Q.d) Rank the light sources with respect to their saturation S.
- (A.d) NA.
- (Q.e) Plot the region of colors which is produced through the mixture of S, A and B.
- (A.e) NA.

EXERCISE 2:
We consider the representation of colors in a color space. In Figure A.1 (see attachment), the color matching functions of the CIE X, Y and Z primary colors are given. Further, in table 1 (see attachment) their spectral values are given with 10 nm interval (e.g. the spectral color of 500 nm has the following tri-stimulus values $\bar{\mathrm{x}} = 0.0049$, $\bar{\mathrm{y}} = 0.323$ and $\bar{\mathrm{y}} = 0.2720$). Given a light source $K(\lambda)$ and an object $\rho(\lambda)$ with certain spectral distributions, then $X = \int_\lambda K(\lambda)\rho(\lambda)\bar{\mathrm{x}}(\lambda)d\lambda$, $Y = \int_\lambda K(\lambda)\rho(\lambda)\bar{\mathrm{y}}(\lambda)d\lambda$ and $Z = \int_\lambda K(\lambda)\rho(\lambda)\bar{\mathrm{z}}(\lambda)d\lambda$. It is assumed that $K(\lambda)$ is a white light source i.e. equal energy distribution over all wavelengths.

- (Q.a) Compute X, Y and Z for a given object color A of 500 nm i.e. $\rho(\lambda_{500}) = 1 and 0 otherwise$. Further, calculate the chromaticity coordinates $x = \frac{X}{X+Y+Z}$, $y = \frac{Y}{X+Y+Z}$ and $z = \frac{Z}{X+Y+X}$ of A.
- (A.a) $X_A = K0.0049, Y_A = K0.323, Z_A = K0.272$ and $x_A = 0.0081, y_A = 0.538, z_A = 0.453$.
- (Q.b) Plot color A as a small circle in the chromaticity diagram.
- (A.b) NA.
- (Q.c) Given an object color B of 580 nm (i.e. $\rho(\lambda_{580}) = 1$ and 0 otherwise), find X, Y and Z and the chromaticity coordinates x, y en z.
- (A.c) $X_B = K0.9163, Y_B = K0.8700, Z_B = K0.0017$ and $x_B = 0.512, y_B = 0.486, z_B = 0.0$.
- (Q.d) Plot color B as a small cross in the chromaticity diagram.
- (A.d) NA.
- (Q.e) Given a color C consisting of the colors A of 500 nm and B of 580 nm. Compute X, Y en Z and the chromaticity coordinates x, y en z.
- (A.e) $x_C = 0.385, y_C = 0.499, z_C = 0.115$.
- (Q.f) Plot the color as a small triangle in the chromaticity diagram.
- (A.f) NA.
- (Q.g) If the white light source $K(\lambda)$ varies (only) in intensity what would happen with the values X, Y, Z and x, y and z of the colors A, B and C? What will be the consequence?
- (A.g) X, Y, Z will change evenly. x, y and z are invariant to intensity changes.
- (Q.h) The tri-stimulus values of a given lamp L are as follows $X = 98.04$, $Y = 100.00$ and $Z = 118.12$. Compute the chromaticity coordinates x, y and z and plot color L with a small rectangle in the chromaticity diagram.
- (A.h) $x_L = 0.31, y_L = 0.316, z_L = 0.373$.
- (Q.i) Indicate, by three different lines, the colors which are generated by the mixture of L with A, B and C respectively.

- (A.i) NA.
- (Q.j) What is the hue (dominant wavelength) of C with L as reference white?
- (A.j) Around 540 nm.
- (Q.k) Order the three colors A, B and C with respect to their saturation.
- (A.k) NA.
- (Q.l) What are the complementary colors A^c, B^c and C^c for A, B en C respectively with L as reference white? Are these complementary colors pure (wavelength) or a mixture of pure colors?
- (A.l) NA.
- (Q.m) Draw the region of colors which are generated by the mixture of A^c, B^c, C^c and L.
- (A.m) NA.
- (Q.n) Given is a color with a spectral power distribution given in Figure A.2 (see attachment). Estimate the hue (dominant wavelength) and describe the amount of the saturation and intensity. What should be the approximated position of this color in the chromaticity diagram?
- (A.n) NA.
- (Q.o) Given is a color with spectral power distribution given in Figure A.3. Estimate the hue (dominant wavelength) and describe the amount of the saturation and intensity. What should be the approximated position of this color in the chromaticity diagram?
- (A.o) NA.
- (Q.p) Given is a color with spectral power distribution given in Figure A.4. Estimate the hue (dominant wavelength) and describe the amount of the saturation and intensity. What should be the approximated position of this color in the chromaticity diagram?
- (A.p) NA.
- (Q.q) For which of the three spectra a human will perceive the highest intensity? Explain your answer.
- (A.q) Human perception of brightness is related to Y.

EXERCISE 3:

We consider the color of a matte, dull (not glossy) surface. The color at a specific location on the surface under white light illumination is given by the following simple reflection model $R = Ik_R \cos\theta$, $G = Ik_G \cos\theta$ and $B = Ik_B \cos\theta$, where I is the intensity of the white light source, k_R, k_G and k_B are the amount of red, green and blue reflected by the surface (i.e. color of the surface). Furthermore, $\cos\theta = \vec{n} \cdot \vec{l}$ is the dot product of the two-unit vectors $\vec{n}$ (i.e. surface normal) and $\vec{l}$ (i.e. direction of the light source), see Figure A.5.

- (Q.a) Assume that the surface is flat and homogeneously colored. Explain why the intensity is higher when the surface normal coincides with the direction of the light source than observed under an angle with respect to the direction of the light source.
- (A.a) Dependent on $\cos\theta$.
- (Q.b) Assume that the color of the surface is yellow i.e. $R = 100$, $G = 100$, and $B = 10$. Explain what will happen with the values R, G and B if (only) the intensity of the light source will diminish. Plot the positions of the colors in the RGB-color space.
- (A.b) R, G and B values change evenly.
- (Q.c) In case of a curved (not flat) surface, indicate where the colors will be positioned in the RGB-color space. Explain your answer.
- (A.c) Same as answer A.b.
- (Q.d) A simple color invariant is given by R/G. Proof that R/G is independent of the (intensity) light source I, object geometry and the direction of the light source.
- (A.d) $\frac{R}{G} = \frac{Ik_R \cos\theta}{Ik_G \cos\theta} = \frac{k_R}{k_G}$
- (Q.e) Which color models (R, G, B or R/G) will you choose for the recognition of objects under varying light intensity. Explain your answer.
- (A.e) R/G.
- (Q.f) Consider the same surface. Assume that the surface is glossy (instead of matte). The reflection model is now given by $R = Ik_R cos\theta + Ik_s cos^n\alpha$, $G = Ik_G cos\theta + Ik_s cos^n\alpha$ and $B = Ik_B cos\theta + Ik_s cos^n\alpha$. k_s is the specular reflection coefficient and cos^n depends on the glossiness and α depends on the viewing condition. Plot the colors of the homogeneously colored (shiny) surface in RGB- and rgb-color space.
- (A.f) NA.
- (Q.g) Proof that R/G is not a color invariant for shiny surfaces. Proof that $\frac{R-G}{R-B}$ is a color invariant for shiny surfaces.
- (A.g) $\frac{R-G}{R-B} = \frac{k_R-k_G}{k_R-k_B}$

2 Exercises (Lectures 3 and 4)

Exercise 1. Image Filtering

Below are four types of filters.

3x3 uniform (box) filter:

$$T = \begin{array}{|c|c|c|} \hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline \end{array}$$

7x7 uniform (box) filter:

$U =$

1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1

3x3 edge filter

$V =$

-1	0	1
-1	0	1
-1	0	1

Laplacian using the following matrix values:

$W =$

1	-2	1
-2	4	-2
1	-2	1

- (Q.a) Which of the following would make an image blurrier, a 3x3 or a 7x7 uniform filter? Why?
- (A.a) A 7x7 uniform filter. Average over a larger area.
- (Q.b) What edges are highlighted with the 3x3 edge filter?
- (A.b) Vertical edges.
- (Q.c) You wish to transform an image by applying a 3x3 uniform filter followed by the 3x3 Laplacian filter. Show that this can be implemented in a single convolution using a 5x5 filter and calculate the elements of this filter.
- (A.c) $W =$

1	-1	0	-1	1
-1	1	0	1	-1
0	0	0	0	0
-1	1	0	1	-1
1	-1	0	-1	1

Exercise 2. Color Constancy

Color constancy is an important issue when recognizing an object independent of the color of the light source. Two simple color constancy algorithms are based on the white patch assumption and the grey-world hypothesis. The R, G and B channels of a small image are given in Table 1.

- (Q.a) Give an example of an image for which the white patch method will fail. Please explain.
- (A.a) An image containing a few colorful (object) colors. The assumption is that an achromatic patch is present in the image.

$R =$

120	120	120
120	120	120
180	180	180

$G =$

70	80	70
70	70	80
80	230	70

$B =$

100	50	30
90	220	20
150	120	80

Table 1: The R, G and B channels of a small image.

- (Q.b) Explain in words how these color constancy methods work.
- (A.b) Taking the maximum or average value of a region/image.
- (Q.c) Calculate the results of both algorithms for the image shown in Table 1.
- (A.c) White patch: $a_1 = \frac{255}{180}$, $a_2 = \frac{255}{230}$, $a_3 = \frac{255}{220}$. Grey-world: $a_1 = \frac{128}{140}$, $a_2 = \frac{128}{91}$, $a_3 = \frac{128}{95.4}$.

Exercise 3. Edge Classification

Edge classification is used to detect and classify transitions based on their physical nature. One possible transition type is a shadow one. An example is shown in Table 2 where the R, G and B values of a small image are given.

- (Q.a) Compute the derivative, of the image in Table 2, in the x-direction by the use of a simple differential filter [-1 1]. The origin is the left pixel.
- (A.a) $f'(x, y) =$

0	20	0	0
0	20	0	0
0	20	0	0
0	20	0	0

- (Q.b) Compute the normalized red (r) response of the image.
- (A.b) $r(x, y) =$

85	85	85	85
85	85	85	85
85	85	85	85
85	85	85	85

- (Q.c) Calculate the derivative in the x-direction for r.
- (A.c) $r'(x, y) =$

0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0

.

$$R = G = B = \begin{array}{|c|c|c|c|} \hline 20 & 20 & 40 & 40 \\ \hline 20 & 20 & 40 & 40 \\ \hline 20 & 20 & 40 & 40 \\ \hline 20 & 20 & 40 & 40 \\ \hline \end{array}$$

Table 2: The R, G and B values of a small image containing a shadow transition.

- (Q.d) How can you classify the transition to be of a shadow type?
- (A.d) Edge responses present in RGB, but not present in normalized color rgb.
- (Q.e) With the same procedure, could you distinguish shadow edges from geometry edges? Please explain.
- (A.e) Not possible with color models. Smooth vs. abrupt transitions.
- (Q.f) With the same procedure, how can highlights be classified? Do you need more than two color features? Which ones?
- (A.f) H (no highlights) and rgb (highlights).

Exercise 4. Error propagation

Color invariants become unstable for certain imaging conditions. One way to handle instabilities is by error propagation. Consider a pixel having the following values $R = 20$, $G = 40$, $B = 60$ with $\sigma = 4$.

- (Q.a) Show that intensity $I = R+G+B$ is a color feature which is stable.
- (A.a) $\sigma_I = \sqrt{\sigma_R^2 + \sigma_G^2 + \sigma_B^2}$. Not dependent on R, G or B. Stable.
- (Q.b) Show that the color feature $1/R$ becomes unstable when the intensity is decreasing.
- (A.b) $\sigma_{\frac{1}{R}} = \frac{-\sigma_R}{R^2}$. Uncertainty depends on R and becomes unstable with decreasing intensity including R.
- (Q.c) Under which circumstances do you think that normalized color and hue will become unstable?
- (A.c) Normalized color = intensity. Hue = saturation.
- (Q.d) How can error propagation be used for histogram construction for image retrieval? Please explain.
- (A.d) Kernel density estimation, where the kernel width corresponds to the uncertainty q_σ of transformation q.

Exercise 5. Retrieval Effectiveness Measures

Given are two different image retrieval systems with the following characteristics. Firstly, the image database consists of 1000 images. The number of relevant images with respect to a given query is 10 composed of the following set $(A, B, C, D, E, F, G, H, I, J)$. The number of images shown to the user is 15 (Answer Set). Further, the order of the 15 highest ranked images of the two different image retrieval systems (for the same image query) is as follows:

S_1	S_2
1. A	1. K
2. L	2. A
3. B	3. M
4. N	4. N
5. O	5. O
6. P	6. B
7. Q	7. Q
8. C	8. R
9. S	9. S
10. T	10. T
11. D	11. C
12. V	12. V
13. W	13. W
14. X	14. X
15. Y	15. D

- (Q.a) Calculate the precision and recall.
- (A.a) Recall system 1: $R_{S_1} = \frac{4}{10}$, Precision system 1: $P_{S_1} = \frac{4}{15}$. Recall system 2: $R_{S_2} = \frac{4}{10}$, Precision system 2: $P_{S_2} = \frac{4}{15}$.
- (A.b) Compute the precision-recall graph for the two different image retrieval systems.
- (Q.b) System 1: (Recall, Precision) = (10%,100%), (20%,66.7%),(30%,37.5%), (40%,36.4%). System 2: (Recall, Precision) = (10%,50%), (20%,33.3%),(30%,27.3%), (40%,26.7%).
- (Q.c) Compute the R-Precision.
- (A.c) $R_1 = \frac{3}{10}$, $R_2 = \frac{2}{10}$

3 Exercises (Lectures 4 and 5)

Exercise 1. Edges and Corners

Consider the image patches A and B in Table 3.

- (Q.a) What are the interest points of image patch A?
- (A.a) Edges.
- (Q.b) Compute the gradient magnitude and the Harris corner response of image patch A (using a simple derivative filter e.g. [1-1]).

$$A = \begin{array}{|c|c|c|c|}\hline 1 & 1 & 1 & 0 \\ \hline 1 & 1 & 1 & 0 \\ \hline 1 & 1 & 1 & 0 \\ \hline 1 & 1 & 1 & 0 \\ \hline \end{array}$$

$$B = \begin{array}{|c|c|c|c|}\hline 0 & 0 & 0 & 0 \\ \hline 1 & 1 & 1 & 0 \\ \hline 1 & 1 & 1 & 0 \\ \hline 1 & 1 & 1 & 0 \\ \hline \end{array}$$

Table 3: Intensity values of image patches A and B.

- (A.b) Gradient: $\bigtriangledown f = \sqrt{f_x^2 + f_y^2} = \begin{array}{|c|c|c|c|}\hline 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 1 & 0 \\ \hline \end{array}$. Harris corner response $R = \text{Det}M - k(\text{Trace}(M))^2 = -0.64$ where $M = \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix}$.
- (Q.c) What are the interest points of image patch B?
- (A.c) Edges and corners.
- (Q.d) Compute the gradient magnitude and the Harris corner response of image patch B (using a simple edge filter e.g. [1-1]).
- (A.d) $M = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$ and hence $R = 6.56$ which is > 0 and therefore a corner.
- (Q.e) Compute the eigenvalues of M for patch B where M is the 2x2 matrix computed from the image derivatives i.e. second moment matrix (autocorrelation matrix).
- (A.e) Eigenvalues are $\lambda_1 = 4$ and $\lambda_2 = 2$.

Exercise 2. Object Descriptors

Object recognition is important in computer vision. Objects can be recognized by considering their presence and locations in unknown images. The standard approach is to use image descriptors in a bag-of-features approach.

- (Q.a) What is an image descriptor?
- (A.a) Image representation describing the extracted (local) patch.
- (Q.b) What are the advantages of using histograms as image descriptors? What about quantization (number of bins)?
- (A.b) Pros: easy to compute. Fixed vector length. Cons: Spatial information is lost. Too many bins then noisy. Too few bins, then too coarse representation.

- (Q.c) What kind of image structures are descriptors made of?
- (A.c) Pixel intensities, color, texture, oriented gradients, etc.
- (Q.d) Compute the histogram of oriented gradients and pixel values for patch A given in Table 3?
- (A.d) $f_x = +3$ corresponds to gradients of 0 degrees bin and 3 accumulations, $f_y = +3$ corresponds to gradients of 90 degrees and 3 accumulations, and $f_{xy} = 1$ to diagonal gradients of 45 degrees and 1 accumulation.
- (Q.e) Compute the histogram of oriented gradients and pixel values for patch B given in Table 3. Which histogram is more discriminative?
- (A.e) Similar to Q.d. Patch B is more discriminative.
- (Q.f) What is the SIFT descriptor?
- (A.f) NA.
- (Q.g) Is the SIFT descriptor invariant under a change in (in-plane) rotation of the object? Please explain.
- (A.g) Normalized by max. gradient orientation.
- (Q.h) What is a color SIFT descriptor?
- (A.h) SIFT applied to different color models (RGB, Irg, HSV) where the final representation is the concatenation.

Exercise 3. Back-of-Features

The bag-of-features approach is an easy way to represent images for image classification.

- (Q.a) What is the difference between dense and point sampling?
- (A.a) Dense sampling: at every sth position where s is the stride. Point sampling: at salient points.
- (Q.b) What are the basic steps of the bag-of-feature approach?
- (A.b) (1) Feature extraction (e.g. SIFT), (2) Visual vocabulary (e.g. K-means of MoG), and (3) Classification (e.g. SVM).
- (Q.c) What are visual words and how is the visual vocabulary computed?
- (A.c) Visual words are prototypical image patches. Visual vocabulary is the dictionary.
- (Q.d) What are spatial pyramids and why are they useful?
- (A.d) To incorporate spatial information.
- (Q.e) Using the back-of-features approach with SVM for object recognition, do you expect that certain objects may be confused during recognition? Give examples.

- (A.e) Use your imagination.
- (Q.f) Do you think that context is important for object recognition? Can you give an example of certain objects?
- (A.f) Use your imagination.

Exercise 4. Object Tracking

After initialization (the initial image location and size of the object is provided) objects can be tracked over time by computing their locations in subsequent frames.

- (Q.a) What is template matching and how can this technique be used for tracking?
- (A.a) Tracking corresponds to searching the target object in a video frame by comparing/matching a model template image in a sliding window approach. Template is fixed and given in advance.
- (Q.b) Could you define a pixel-wise similarity measure for template matching?
- (A.b) SSD or cross-correlation.
- (Q.c) What are the possible image transformations between the template and possible candidates? What is the search area?
- (A.c) Translation, rotation, scale, affine and perspective projections.
- (Q.d) What are the pros and cons of template matching for object tracking?
- (A.d) Pros: robust and simple. Cons: time-consuming and fixed template.
- (Q.e) What is the difference between the similarity measure of template matching and mean-shift?
- (A.e) NA.

4 Exercises (Lecture 6)

Exercise 1. Deep Learning

Deep learning and ConvNets are very useful for object recognition and detection.

- (Q.a) Consider the following linear perceptron classifier:

$$h(\vec{x}, \theta) = \begin{cases} \text{-1, if } 3x_1 + 4x_2 - 24 \\ +1\text{, if } 3x_1 + 4x_2 - 24 \end{cases} . \tag{1}$$

  Depict the neuron for this perceptron. What is the logit and which non-linear function f is used?

- (A.a) Logit is equal to $z = 3x_1 + 4x_2 - 24$ and $f(z) = \begin{cases} \text{-1, if } f(z) > 0 \\ \text{+1, if } f(z) \leq 0 \end{cases}$.

- (Q.b) Consider an image with pixel values $\vec{x} = [10, 11, 1, 2]^T$, weight matrix $\vec{M} = \begin{bmatrix} 1 & 0.5 & -0.2 & 0.3 \\ -0.3 & -0.1 & 0.7 & 0.3 \\ 0.5 & 0.2 & 1.0 & -0.1 \\ -0.1 & 0.3 & 0.2 & 0.3 \end{bmatrix}$, and bias $\vec{b} = [1, 0.5, -0.1, 0.4]^T$.
  Compute the logit and output y_i using a softmax layer i.e. $y_i = \frac{e^{z_i}}{\sum_j e^{z_j}}$. What can you conclude about the prediction?

- (A.b) $\vec{z} = [21856305, 0.1, 7.9, -5.5]^T$ and therefore $\vec{y} = [0.99, 0, 0.0001, 0]^T$. Strong predictor: single entry in the softmax layer is close to 1.

- (Q.c) Consider a full-color 250x250 pixel image. What would be the number of weights for the input layer?

- (A.c) 250x250x3=187,500.

- (Q.d) You have a 32x32x5 image and k filters with an extent of 5 and consider the way most convolutional neural networks are implemented. What is the depth of the k filters? If you use no padding and stride $s = 1$, what will be the output size of the activation map?

- (A.d) Depth of filter is 5. Output size is equal to 28x28xk.

- (Q.e) Consider a feature map with values:

20	23	12	1
12	10	2	5
200	190	6	7
10	20	1	2

  . Calculate the max pooling layer with an extent e and stride s of 2. When are max pooling layers useful?

- (A.e)

23	12
200	7

  Useful: Local invariance and dimensionality reduction of feature maps.

- (Q.f) What are the different layers in a standard CNN? What are the layers of VGG net?

- (A.f) Input layer, convolutional layer, pooling layer, batch normalization layer, fully connected layer, softmax layer.

- (Q.g) What is transfer learning and when is it useful?

- (A.g) Using an existing trained network to improve learning. Suited for small and medium size datasets.

Exercise 2. Sliding Window Approach

Sliding window approach is a popular way to detect objects in images.

- (Q.a) What is the basic pipeline for window-based object detection?

- (A.a) (1) Build object model; learn classifier, (2) generate candidates in new image, (3) score the candidates.
- (Q.b) What are the advantages of a sliding-window approach?
- (A.b) Simple detection protocol.
- (Q.c) Given an image of 256x256, how many windows are required to detect objects for 8 different orientations and 6 scales?
- (A.c) 256x256x8x6 = 3,145,728.
- (Q.d) Assume that for a strong classifier (e.g. non-linear SVM), the time for window classification is about 0.01 seconds. How many hours does it take to detect objects in 10,000 images?
- (A.d) 87381 hours.
- (Q.e) How can one reduce the number of bounding boxes for detection? Now, how long does it take to detect objects in 10,000 images?
- (A.e) Jump positions (strides) or segmentation. Approximately 42 hours.