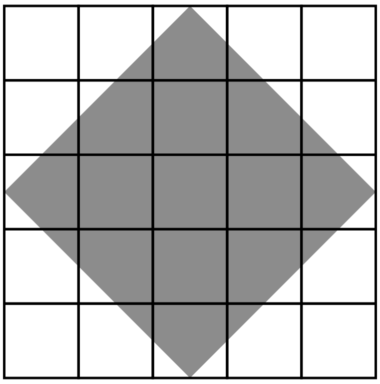
**Institut für Mess- und Regelungstechnik
mit Maschinenlaboratorium
Karlsruher Institut für Technologie
(KIT)
Prof. Dr.-Ing. C. Stiller**

Solutions for exam "Machine Vision" September 14, 2023

Question 1 (6 points)

The picture below shows a 5-by-5 pixel grid with a gray diamond shaped area. Design a filter mask that calculates the average gray value over the diamond shaped area.



Solution

We have to consider for each pixel which part of the pixel is covered by the diamond. Finally, we sum up the entries and use its reciprocal as normalization factor

$$\frac{2}{25} \cdot \begin{pmatrix} 0 & \frac{1}{8} & \frac{3}{4} & \frac{1}{8} & 0 \\ \frac{1}{8} & \frac{7}{8} & 1 & \frac{7}{8} & \frac{1}{8} \\ \frac{3}{4} & 1 & 1 & 1 & \frac{3}{4} \\ \frac{1}{8} & \frac{7}{8} & 1 & \frac{7}{8} & \frac{1}{8} \\ 0 & \frac{1}{8} & \frac{3}{4} & \frac{1}{8} & 0 \end{pmatrix}$$

Question 2 (4 points)

Design a potential function for a random field based segmentation that fosters that a pixel shares the same label as its neighbors if all eight neighbors share the same label. The potential function should not have any effect if the eight neighbors do not agree on the same label. Denote with l_0 the label of the pixel of concern and with $l_1, \ldots, l_8$ the label of the eight neighboring pixels. Furthermore, $c_0, \ldots, c_8$ denote the colors of the pixels.

Solution

$$\phi(l_0, l_1, \ldots, l_8) = \begin{cases} 1 & \text{if } l_1 = l_2 = \cdots = l_8 \text{ and } l_0 \neq l_1 \\ 0 & \text{otherwise} \end{cases}$$

Question 3 (4 points)

Assume that the loss term $g(w)$ of an artificial neural network has the form

$$g(w) = w^3 - 10w + 20$$

Perform two steps of gradient descent with learning rate $\varepsilon = \frac{1}{5}$ starting from the inital weight $w_0 = 0$. For each step provide the value of w and the value of $\frac{\partial g}{\partial w}$.

Solution

We calculate the derivative of g as $g'(w) := \frac{\partial g}{\partial w} = 3w^2 - 10$.

1st step:

$$g'(w_0) = -10 \quad \Rightarrow \quad w_1 = w_0 - \varepsilon g'(w_0) = 0 - \frac{1}{5} \cdot (-10) = 2$$

2nd step:

$$g'(w_1) = 2 \quad \Rightarrow \quad w_2 = w_1 - \varepsilon g'(w_1) = 2 - \frac{1}{5} \cdot 2 = 1.6$$

Question 4 (3+3 points)

Assume a single perceptron with ReLU activation and two inputs x_1 and x_2. We consider as input only values in the interval $-1 \leq x_1, x_2 \leq 1$. Other inputs are not considered. Which of the following functions can be implemented by the perceptron? Explain your answer briefly.

(a) $f(x_1, x_2) = x_1 + 2x_2 + 3$

(b) $g(x_1, x_2) = \begin{cases} x_1 + x_2 & \text{if } x_1 > 0 \text{ and } x_2 > 0 \\ 0 & \text{otherwise} \end{cases}$

Solution

(a) f can be implemented with weights $w_0 = 3$, $w_1 = 1$, $w_2 = 2$. Since $f(x_1,x_2) \geq 0$ for all relevant values of x_1 and x_2 the perceptron exactly implements f

(b) g cannot be implemented by a single perceptron. A single perceptron can only implement a single linear hyperplane that separates the input space into a half space for which the output is 0, and another half space in which the output depends linearly from the inputs. Two implement g we need to model two lines (i.e. $x_1 = 0$ and $x_2 = 0$) to separate the zero area from the non-zero area. Therefore, it can't be implemented by a single perceptron.

Question 5 (8 points)

Implement a Python function *make_binary_histogram* that calculates the binary histogram for local binary patterns from an image. The function should take as input an eight-bit encoded graylevel image and should yield a 256-dimensional vector that implements the histogram. It should consider all pixels in the image except of the boundary pixels.

You might make use of a function *make_decimal* that takes as input a binary vector (i.e. a vector composed out of entries 0 and 1) that is interpreted as binary number. *make_decimal* returns this number. Example: `make_decimal([ 0,0,1,1,0,0,1,1 ])` returns 51.

```
import numpy as np

def make_decimal(pattern:  np.ndarray) -> int:
  v = 2**np.array([7,6,5,4,3,2,1,0])
  return np.sum(pattern * v)

# @student:  implement function make_binary_histogram
```

Solution

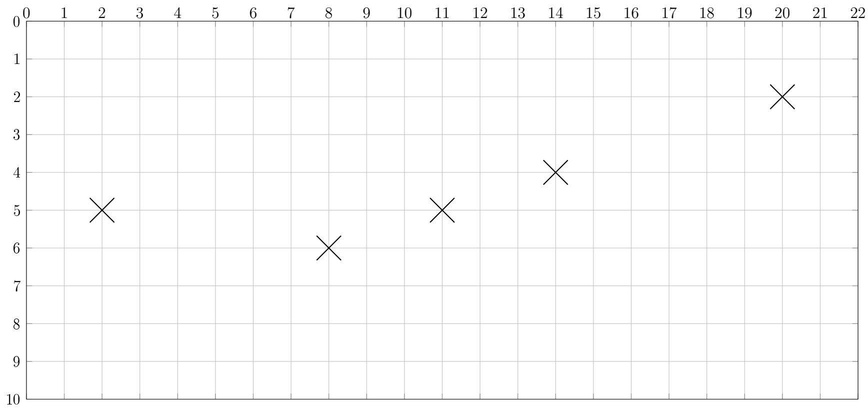
```
def make_binary_histogram(image):
  hist = np.zeros(shape=(256))
  height = image.shape[0]
  width = image.shape[1]

  for v in range(1, height):
    for u in range(1, width):
      pixel = image[v,u]
      pattern = np.zeros(shape=(8))
      pattern[0] = pixel < image[v-1, u-1]
      pattern[1] = pixel < image[v-1, u]
      pattern[2] = pixel < image[v-1, u+1]
      pattern[3] = pixel < image[v, u+1]
      pattern[4] = pixel < image[v+1, u+1]
      pattern[5] = pixel < image[v+1, u]
      pattern[6] = pixel < image[v+1, u-1]
      pattern[7] = pixel < image[v, u-1]
      hist[make_decimal(pattern)] += 1

  return hist
```

Question 6 (5+3 points)

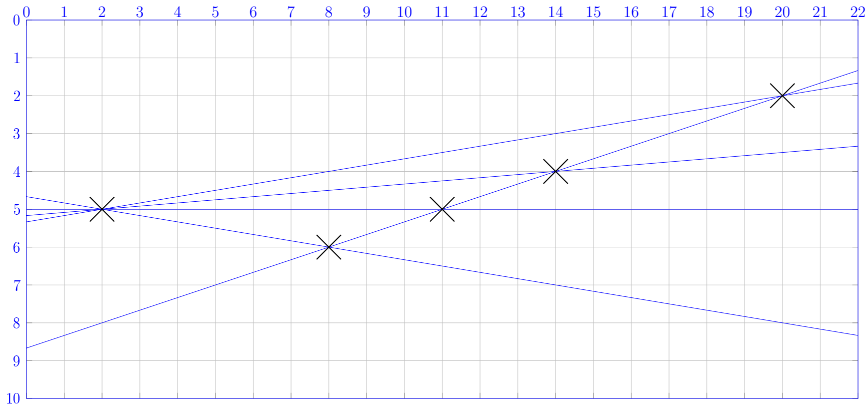
A linefitting must be applied to five sample points which are displayed below. To reduce the impact of outliers the RANSAC algorithm is used. All possible combinations of sample points are used to check the number of outliers in the tolerance band. The threshold θ can be either set to 0.5 or to 2 pixels.

(a) Which combination of two points and a threshold θ returns the best result?

(b) Assume that all sample point that are not outliers are perceived with very low noise. What threshold θ must be chosen and why? Which sample points provide the best result?



Solution



(a) Point 1 and 4, $\theta = 2$

(b) Point 2, 3, 4, or 5, , $\theta = 0.5$

Question 7 (3+2+3 points)

A scene is illuminated by bluish light. To achieve color constancy, white balance is applied. A white reference surface is used to calculate average values for RGB.

(a) Are the correction factors c_R and c_B less than, greater than or equal to 1?

(b) How does the white balance effect the image brightness.

(c) What happens if the lighting changes to greenish light and white balance is performed again?

Justify your answers briefly.

Solution

(a) $c_R = 1$
$c_B < 1$

(b) The image is darker than before white balance.

(c) Both correction factors c_R and c_R are greater than 1. The image is brighter than before the white balance.

Question 8 (3+4+1 points)

Assume a camera with thin lens. The focal lens of the lens is $f_{lens} = 10$. The focal length of the camera is $f_{camera} = 11$. The diameter of the aperture is $D = 20$.

(a) Calculate the optimal object distance z_0 at which points create a perfectly sharp image in the camera

(b) Based on your result from part (a) calculate how the diameter of the unsharpness circle in the image plane ε depends on the object distance z for distances $z \geq z_0$. The unsharpness circle is the area on the image plane which is illuminated by a point light source at distance z.

(c) Based on your result from part (b) how large can ε become at most for $z \geq z_0$?

Solution

(a) We use the lens equation from slide 07/19 for this calculation and get

$$\frac{1}{f_{lens}} = \frac{1}{f_{camera}} + \frac{1}{z_0} \quad \Rightarrow \quad \frac{1}{10} = \frac{1}{11} + \frac{1}{z_0} \quad \Rightarrow \quad z_0 = \frac{11 \cdot 10}{11 - 10} = 110$$

(b) To determine ε we use the upper equation from slide 07/21 which yields

$$\varepsilon = D \cdot \frac{f_{lens} \cdot (z - z_0)}{z \cdot (z_0 - f_{lens})} = 20 \cdot \frac{10 \cdot (z - 110)}{z \cdot (110 - 10)} = 2 \cdot \frac{z - 110}{z} = 2 - \frac{220}{z}$$

(c)

$$\sup_{z \geq z_0} \varepsilon(z) = 2$$

Question 9 (4+4 points)

Assume a binary classification task and an ensemble classifier with 20 ensemble members that was trained for that classification task. We apply the ensemble to a test set of 12 examples (named as A-L) and obtain the classification results as follows

test pattern	true class label	number of positive votes	number of negative votes
A	positive	17	3
B	positive	15	5
C	positive	9	11
D	positive	14	6
E	positive	20	0
F	positive	7	13
G	negative	15	5
H	negative	7	13
I	negative	2	18
J	negative	0	20
K	negative	1	19
L	negative	8	12

The ensemble classifier is comparing the number of positive votes minus the number of negative votes against a threshold δ.

(a) Calculate the precision and recall of the ensemble classifier for threshold $\delta = 0$

(b) Which values of δ maximize the recall? What is the highest precision that we can achieve when we maximize the recall?

Solution

(a) For $\delta = 0$ we obtain two false negatives (C and F), one false positive (G), and 4 true positives (A,B,D,E). Hence

$$precision = \frac{4}{4+1} = \frac{4}{5}$$

$$recall = \frac{4}{4+2} = \frac{4}{6}$$

(b) To maximize the recall we have to decrease δ such that all positive examples become true positives. This happens if $\delta < -6$. If we choose δ a little bit lower than -6 (e.g. -6.1) we obtain 6 true positives (A-F) and 3 false positives (G,H,L). Hence

$$precision = \frac{6}{6+3} = \frac{6}{9}$$

Gesamtpunkte: 60