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

Solutions for exam "Machine Vision" August 9, 2019

Question 1 (6 points)

Assume a graylevel image g with gray values as illustrated in the figure below and a filter mask f as shown below. The origin of f is its central pixel. a, b, and c are unknown entries of the filter mask. The graylevel image h shown below was generated by convoluting g with f, i.e. $h = g * f$. Calculate the values of a, b, and c.

g:

1	0	2	5	3
0	3	2	0	4
4	1	2	0	0
2	0	0	1	6

f:

-1	0	a
b	c	b
a	0	-1

h:

	21	22	13	
	14	0	4	

Solution

By definition of the convolution operator, we obtain

$$
\begin{aligned}
21 &= 1\cdot(-1) + 0\cdot 0 + 2\cdot a + 0\cdot b + 3\cdot c + 2\cdot b + 4\cdot a + 1\cdot 0 + 2\cdot(-1)\\
22 &= 0\cdot(-1) + 2\cdot 0 + 5\cdot a + 3\cdot b + 2\cdot c + 0\cdot b + 1\cdot a + 2\cdot 0 + 0\cdot(-1)\\
13 &= 2\cdot(-1) + 5\cdot 0 + 3\cdot a + 2\cdot b + 0\cdot c + 4\cdot b + 2\cdot a + 0\cdot 0 + 0\cdot(-1)\\
14 &= 0\cdot(-1) + 3\cdot 0 + 2\cdot a + 4\cdot b + 1\cdot c + 2\cdot b + 2\cdot a + 0\cdot 0 + 0\cdot(-1)\\
0 &= 3\cdot(-1) + 2\cdot 0 + 0\cdot a + 1\cdot b + 2\cdot c + 0\cdot b + 0\cdot a + 0\cdot 0 + 1\cdot(-1)\\
4 &= 2\cdot(-1) + 0\cdot 0 + 4\cdot a + 2\cdot b + 0\cdot c + 0\cdot b + 0\cdot a + 1\cdot 0 + 6\cdot(-1)
\end{aligned}
$$

After simplification, we obtain

$$
\begin{aligned}
15 &= 5a + 6b\\
4 &= b + 2c\\
12 &= 4a + 2b
\end{aligned}
$$

Hence,

$$a = 3, b = 0, c = 2$$

Question 2 (6 points)

Let g be a graylevel image. Assume that the Harris corner detector classifies a certain pixel (u_1, v_1) in g as a corner pixel. We create a new image h from g by transposing g, i.e. $h(v,u) = g(u,v)$ for all pixels. Show that the Harris corner detector applied to image h classifies the pixel (v_1, u_1) of h as corner pixel.

Solution

Since h was generated from g by exchanging the u- and v-axis, it follows that $\frac{\partial h(v,u)}{\partial u} = \frac{\partial g(u,v)}{\partial v}$ and $\frac{\partial h(v,u)}{\partial v} = \frac{\partial g(u,v)}{\partial u}$. If we denote with $S_g = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$ the matrix that is used by the Harris corner detector to assess whether the pixel (u_1,v_1) is a corner point in g, then the respective matrix for for pixel (v_1,u_1) in image h is $S_h = \begin{pmatrix} c & b \\ b & a \end{pmatrix}$. Obviously, the Eigenvalues of S_g and S_h are the same. Hence, the point (u_1,v_1) is corner point in g if and only if point (v_1, u_1) is corner point in h.

Question 3 (6 points)

Implement a MATLAB function *vertical_fit* that fits a vertical line to a set of 2-dimensional points. The function takes as input a $2 \times n$ matrix of 2-dimensional points where the x-coordinates of the points are provided in the first row and the y-coordinates in the second row. n is the number of points and it can be assumed to be greater than 2. The function should return as output two arguments, the unit normal vector of the line and the offset c. Use the provided header for implementation of the MATLAB function. Explain your approach briefly.

```
function [ normal, c ] = vertical_fit ( data )
% add your code below this line
```

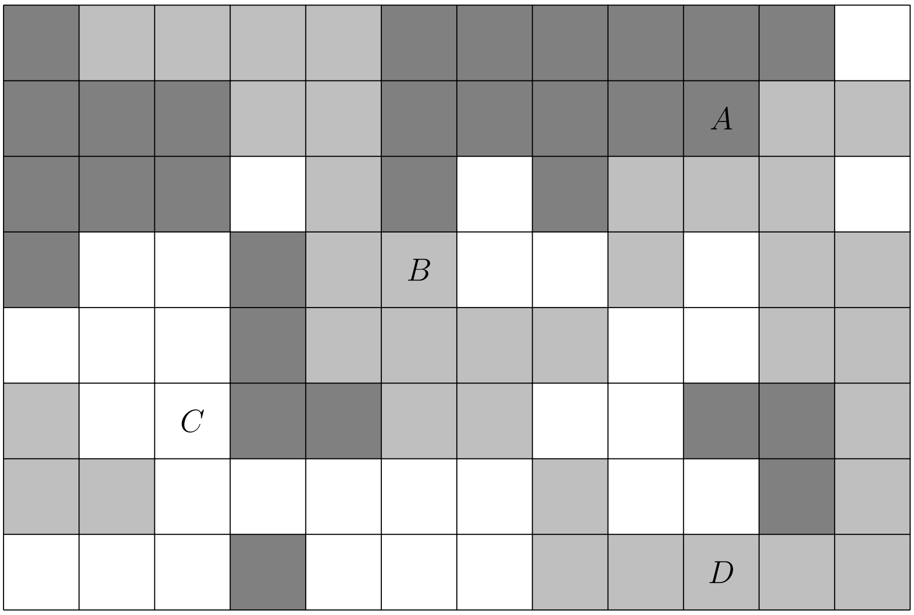
Solution

Since the line should be vertical, its normal vector is given by $(1,0)^T$ while its offset can be calculated with the formula from slide 4/27.

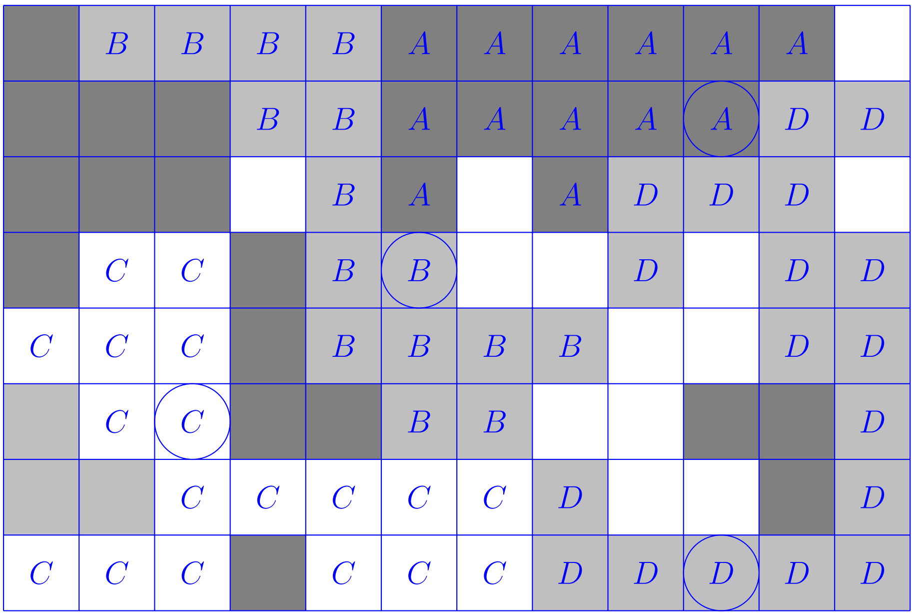
```
function [ normal, c ] = vertical_fit ( data )
normal = [1;0];
c = -sum(data(1,:))/length(data);
end
```

Question 4 (8 points)

A grayscale image with the size of 12×8 pixels should be segmented by region growing. For this purpose, four seed pixels A, B, C and D are already selected. Two pixels are assumed to be similar, if both of the gray values are equal. The similarity is determined in a 4-neighborhood of pixel, i.e. the neighborhood considers the upper, lower, left, and the right neighboring pixel. Execute the region growing algorithm on the image below by filling all segmented cells with the correct labels (A, B, C, or D).



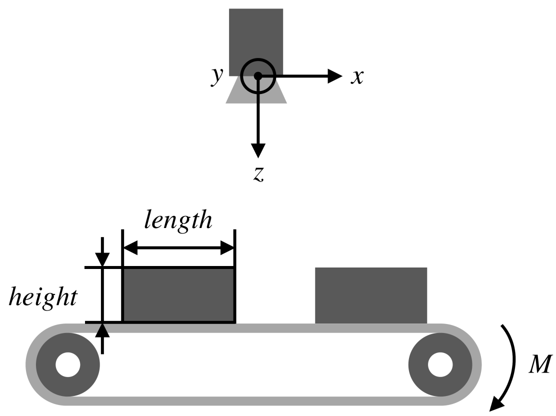
Solution



Question 5 (5+5 points)

Imagine you should design a camera mount in order to check the imprints on the top of boxes which are moving on a conveyer belt. Every box has the same measurements: length 400 mm, width 300 mm and height 200 mm. Your preselected camera provides a resolution of 2002×1001 pixels. The intrinsic parameters of the camera are given by the matrix A. For simplification, a pinhole camera model is assumed.

$$A = \begin{pmatrix} 800 & 0 & 1001 \\ 0 & 800 & 500 \\ 0 & 0 & 1 \end{pmatrix}$$



(a) You have decided to mount the camera straight above the conveyer belt. Calculate the maximum height of the camera above the conveyor belt that allows to perceive thin lines of 0.5 mm width on top of the boxes.

(b) In part (b) assume a camera height of 400 mm above the conveyer belt. Is it possible to see a whole box in a single camera image?

Remark: Part (b) can be solved independently of part (a)

Solution

(a) One pixel should display an area on top of the boxes that is at most 0.5 mm wide. With the use of the given camera matrix A, the height of the camera can be calculated.

$$z \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = A \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

$$z \cdot 1 < 800 \cdot 0.5mm = 400mm$$

z denotes the required height of the camera with regard to the top of the boxes. To calulate the height relative to the conveyor belt h, the height of a box must be added.

$$h = z + 200mm < 600mm$$

(b) Because a box with an aspect ratio of 4:3 should fit into an image with an aspect ratio of 2:1, only the y-dimension needs to be evaluated.

$$z \cdot v = 800 \cdot y + 500 \cdot z \Leftrightarrow y = \frac{(v - 500) \cdot z}{800}$$

The image resolution in v-direction reaches from $v_{min} = 0$ to $v_{max} = 1000$. The height of the camera with respect to the top of the boxes is calculated as $z = 200$ mm

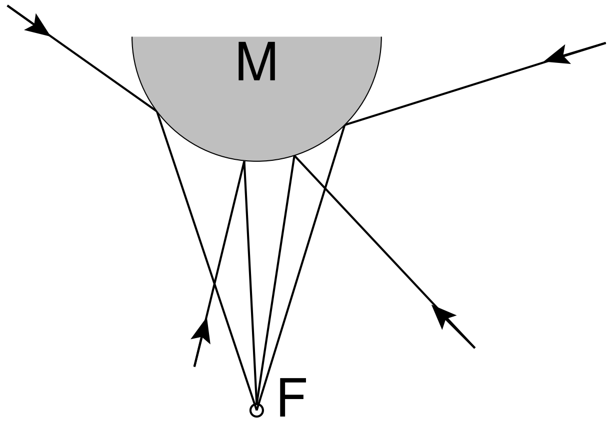
$$v \in [0, 1000], z = 200mm$$

$$y_{min} = \frac{(v_{min} - 500) \cdot z}{800} = \frac{(0 - 500) \cdot 200mm}{800} = -125mm$$
$$y_{max} = \frac{(v_{min} - 500) \cdot z}{800} = \frac{(1000 - 500) \cdot 200mm}{800} = 125mm$$
$$\Delta y = y_{max} - y_{min} = 125mm - (-125mm) = 250mm < 300mm$$

An entire box cannot be displayed in a single image because the Δy is smaller than the width of a box.

Question 6 (4 points)

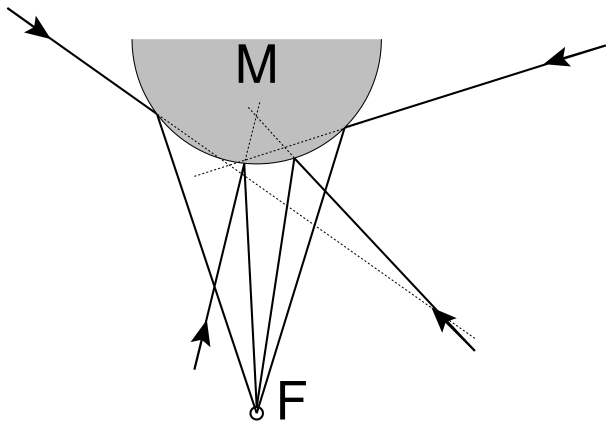
The figure below illustrates a catadioptric camerasystem with a mirror M and a pinhole camera with focal point F. Four lines of sight are shown in the figure. Explain whether or not this catadioptric camera has a single viewpoint.



Remark: You may use and modify the figure above to explain your solution.

Solution

A catadioptric camera has a single viewpoint if all lines of sight intersect in a single point. The figure belows shows the extensions of the lines of sight as dashed lines. Obviously, they are intersecting in various different points. Hence, the camera system does not have a single viewpoint.



Question 7 (6 points)

Assume a classification problem in which the training examples are described by a single feature $x \in \mathbb{R}$. The training set consists out of the following weighted examples

x	1	3	5	7	9	11	13	15
weight	5	3	1	4	5	2	2	2
class label	$+1$	$+1$	-1	-1	-1	-1	$+1$	$+1$

Calculate the best threshold classifier that minimizes the sum of weighted errors on the training set. Provide the threshold $\theta \in \mathbb{R}$ and the orientation paramter $z \in \{-1, +1\}$. The threshold classifier returns the value of $sign(z \cdot (x - \theta))$ for a given pattern x .

Solution

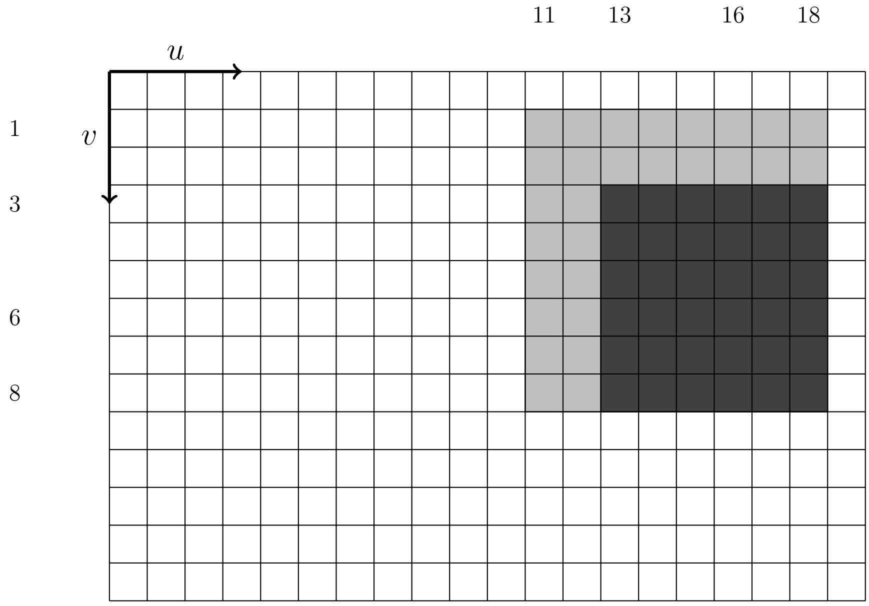
We try different possibilities for θ and z

- if all examples are classified as positive or as negative we get an error of $1+4+5+2 = 12$ or $5+3+2+2 = 12$, resp.
- for $\theta = 4$ and $z = +1$ we get an error of $5+3+1+4+5+2 = 20$
- for $\theta = 4$ and $z = -1$ we get an error of $2+2 = 4$
- for $\theta = 12$ and $z = +1$ we get an error of $5+3 = 8$
- for $\theta = 12$ and $z = -1$ we get an error of $1+4+5+2+2+2 = 16$

Other choices of θ are suboptimal for obvious reasons. Hence, the best solution is $\theta = 4$ and $z = -1$.

Question 8 (7 points)

The figure below shows an image as grid of pixels and a Haar feature illustrated by light gray and dark gray areas. The Haar feature is calculated by subtracting the average gray value of the dark gray area from the average gray value of the light gray area. Provide a formula to calculate the value of the Haar feature based on the integral image $I(u, v)$. Do not add individual gray values in your formula. The numbers in the figure provide row and column coordinates. You may introduce additional variables to represent intermediate results and build your solution on those variables.



Solution

There are several possible solutions. The example solution is based on the idea to partition the light gray area into two rectangles, one that contains the columns 11 and 12, and the other one that contains the rows 1 and 2 and columns 13 to 18.

We calculate the sums over the three areas:

$$A = I(12,8) + I(10,0) - I(12,0) - I(10,8)$$
$$B = I(18,2) + I(12,0) - I(18,0) - I(12,2)$$
$$C = I(18,8) + I(12,2) - I(18,2) - I(12,8)$$

Hence, we obtain as average for the light gray area (positive) and dark gray area (negative):

$$P = \frac{A+B}{28}$$
$$N = \frac{C}{36}$$

Hence, the value of the Haar feature is $P - N = \frac{1}{28}(A+B) - \frac{1}{36}C$

Question 9 (2+3+2 points)

(a) What are the two main differences between a fully connected layer and a convolutional layer in a multi-layer-perceptron?

(b) Assume a multi-layer-perceptron with two single channel convolutional layers. Let us denote the output of the perceptrons in the first layer by $x_{u,v}$ and the output of the perceptrons in the second layer by $y_{u,v}$. Provide a formula that describes the output of an arbitrary perceptron in the second layer y_{u_0,v_0} for a non-boundary pixel (u_0, v_0) as a function of the outputs of the perceptrons in the first layer $x_{u,v}$ and the weights w_0, w_1, w_2, and w_3. The perceptron should use the ReLU activation function and it should implement a convolution with a covolution kernel f of the following form

f :

0	w_1	0
0	0	w_2
w_3	0	0

Do not forget the bias weight w_0.

(c) Apply max-pooling with a 3-by-3 filter mask to the single-channel layer that is shown in the figue below. The result is a layer with 2-by-2 entries.

y :

4	0	1	3
6	2	0	2
9	4	1	0
10	4	1	2

Solution

(a) local receptive fields, weight sharing

(b)

$$y_{u_0,v_0} = max\{0, w_0 + w_1 \cdot x_{u_0,v_0+1} + w_2 \cdot x_{u_0-1,v_0} + w_3 \cdot x_{u_0+1,v_0-1}\}$$

(c)

4	0	1	3
6	2	0	2
9	4	1	0
10	4	1	2

$\xrightarrow[\text{3 × 3 filtermask}]{\text{max pooling with a}}$

9	4
10	4

Gesamtpunkte: 60