Image and speech recognition

Exercises

Włodzimierz Kasprzak

v.3, 2015

HUMAN CAPITAL
HUMAN – BEST INVESTMENT!

EUROPEAN UNION
EUROPEAN
SOCIAL FUND

Project is co-financed by European Union within European Social Fund

Exercises. Content

1. Pattern analysis (3)
2. Pattern transformation (4)
3. Pattern classification (7)

   3A. Pattern sequence recognition (3)
4. Iconic processing (7)
5. Boundary-based image segmentation (5)
6. Region-based image segmentation (3)
7. Object recognition (3)
8. Speech pre-processing (4)
9. Speech features (3)
10. Phonetic model (3)
11. Word and sentence recognition (2)

E1. Pattern analysis

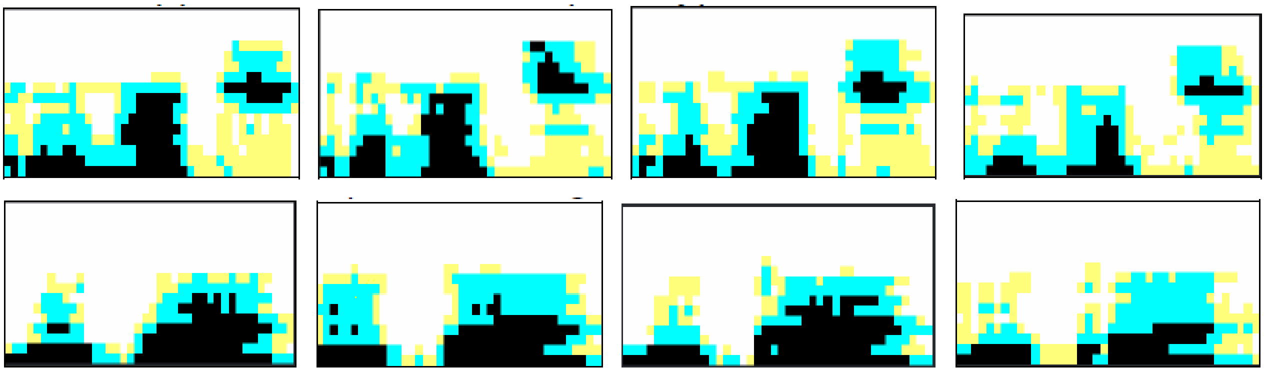
HUMAN CAPITAL
HUMAN – BEST INVESTMENT!

EUROPEAN UNION
EUROPEAN
SOCIAL FUND

Project is co-financed by European Union within European Social Fund

Task 1.1

4 spectrogram images (of resolution 40 frames × 16 frequencies) for spoken words "Koniec„ (top row) and "OK" (bottom row) are given:



A) Propose three types of feature vectors (of small size) for a direct classification of these images. B) For one feature type provide an approximation of class prototypes and run a classification process by a geometric minimum-distance classifier.

Task 1.2

The intensity distribution in the image is modelled as a stochastic variable with the pdf $p_X(A \le X \le B)$ and cumulated density $F_X(x) = p_X(X \le x)$. Let two images of resolution 4×4 are given:

a)

0	1	2	3
4	5	6	7
8	9	10	11
12	13	14	15

b)

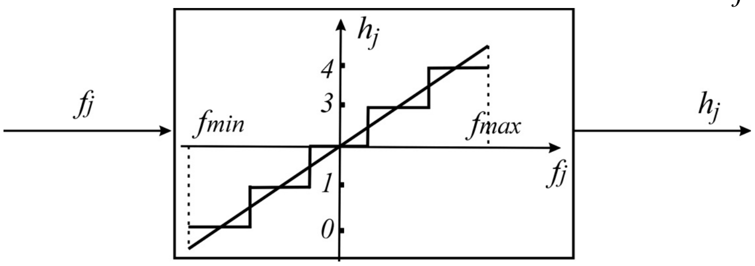
3	3	3	3
7	7	7	7
11	11	11	11
15	15	15	15

(1) Show the pdf-s of intensity variables Assuming, that two original analogue-valued variables are available, get their pdf-s.

(2) What are the means and variances of all these distributions?

Task 1.3

Signal digitalization problem: original analogue value f_j , discrete value: h_j.



Digitalization error: $n_j = f_j - h_j$. The signal-to-noise ratio (SNR) expresses the digitalization quality.

(1) Estimate the SNR value in case of signal digitalization with B bits.

(2) What is the change of the signal-to-noise ratio if the number of bits is increased by 1 ?

E2. Pattern transformation

HUMAN CAPITAL
HUMAN – BEST INVESTMENT!

EUROPEAN UNION
EUROPEAN
SOCIAL FUND

Project is co-financed by European Union within European Social Fund

Task 2.1

(PCA) A set of 2-D features is given below. Define the PCA problem and obtain the particular PCA-based transformation for given set of features.

	1	2	3	4	5	6
feature	(1,1)	(2,4)	(3,5)	(2,1)	(3,0)	(4,-1)

Task 2.2

(LDA) A set of 2-D features from 2 classes is given below. Define the LDA problem and obtain the particular LDA-based transformation for given set of features.

	1	2	3	4	5	6
feature	(1,1)	(2,4)	(3,5)	(2,1)	(3,0)	(4,-1)
class	1	1	2	2	2	2

Task 2.3

(ICA) The following 3 discrete-time source signals are available, S1 = "Triangle signal", S2 ="Step signal", S3="Noise" :

time source	1	2	3	4	5	6	7	8	9	10
S1	1	2	3	4	0	-1	-2	-3	-4	0
S2	1	0	0	-1	-1	0	1	1	0	-1
S3	2	-1	1	-2	2	-1	-1	-2	1	1

(1) Compute the **normalized kurtosis** of each source signal.

(2) Make 3 **instantaneous mixtures** of sources by using the matrix:

$$A_{3\times 3} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 2 & 2 & 1 \end{bmatrix}$$

(3) Compute the **correlation factor** of pairs of sources and pairs of mixtures.

Task 2.4* (1)

(ICA quality measures) Assume that we have tested a blind signal separation method and the following data are available.

The (usually unknown) 3 sources:

Time Source	1	2	3	4	5	6	7	8	9	10
S1	1	2	3	4	0	-1	-2	-3	-4	0
S2	1	0	0	-1	-1	0	1	1	0	-1
S3	2	-1	1	-2	2	-1	-1	-2	1	1

The 3 separated (output) signals:

Time Source	1	2	3	4	5	6	7	8	9	10
Y1	3.3	-0.7	2.2	-1.8	2.2	-1.6	-1.5	-3.1	0.1	0.9
Y2	1.4	2.1	3.4	4.1	0.1	-1.2	-2.2	-3.4	-4.3	0
Y3	1.2	0.2	0.3	-0.7	-1.1	-0.1	0.9	-0.8	-0.4	-1.1

Task 2.4* (2)

The (usually unknown) mixing matrix: $A_{3\times 3} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 2 & 2 & 1 \end{bmatrix}$

The final de-mixing matrix $\boldsymbol{W}$: $W_{3\times 3} = \begin{bmatrix} 2.2 & 0.1 & -1 \\ 0.1 & -1 & 1 \\ -1.1 & 1 & 0.1 \end{bmatrix}$

Express the **quality of separated results** if:

1. Only the outputs can be used.
2. If both the outputs and sources can be used.
3. If both the mixing and de-mixing matrix can be used.

E3. Pattern classification

HUMAN CAPITAL
HUMAN – BEST INVESTMENT!

EUROPEAN UNION
EUROPEAN
SOCIAL FUND

Project is co-financed by European Union within European Social Fund

Task 3.1

Define and solve the optimisation problem required for a linear potential functions classifier (define and solve it analytically).

Consider the case of 2-D feature vectors and 2 classes if following learning samples (feature vectors) are given:

i	1	2	3	4	5	6
$\mathbf{c}^i = (c^i_1, c^i_2)$	(1; 2)	(1; 3)	(1; 4)	(1; 1)	(2; 1)	(3; 1)
class	Ω_1	Ω_1	Ω_1	Ω_2	Ω_2	Ω_2

Task 3.2

Let 10 learning samples from 3 classes are given in the R^2 space (2-D feature vectors):

(x,y)	(1; 2,5)	(0.5; 3)	(0; 3,5)	(2; 3)	(4; 4)	(3; 5)	(1; 1)	(2; 0)	(3; 1)	(2;2)
Class	1	1	1	2	2	2	3	3	3	3

(A) Design the following classifiers:

1. the geometric minimum-distance classifier,
2. the Bayes classifier,
3. the k-NN classifier (k=4, the 4-NN classifier).

(B) Give classification results obtained by these three classifiers for a feature (2; 2,5) .

Task 3.3

Let the single observation of unknown state parameter x is

$z = x + w,$ where noise: $w \sim \mathrm{N}(0, \sigma^2)$.

A) What value takes the ML estimate ?

B) What value takes the MAP estimate ?

C) When is the MAP estimate equivalent to the ML one?

Task 3.4

Let a sequence of observations of unknown (constant) state parameter x is given, as: $z(t) = x + w(t), \quad t=1,...,N;$ where noise $w \sim N(0, \sigma^2)$.

A) What is the LS (least square) estimator?

B) What is the MMSE (minimum mean square error) estimator?

Task 3.5

Define and solve the optimisation problem for a linear SVM:

(A) Define the primary problem analytically and solve it by Matlab;

B) Define the dual problem and solve it like in case A);

C) Using a custom method (e.g. the “brute-force” method) check candidates for support vectors and define the separating plane;

D)Illustrate the support vectors and separating plane in graphical form.

Following learning samples (feature vectors) are given:

i	1	2	3	4	5	6
$\mathbf{c}^i = (c^i_1, c^i_2)$	(1; 2)	(1; 3)	(1; 4)	(1; 1)	(2; 1)	(3; 1)
class	Ω_1	Ω_1	Ω_1	Ω_2	Ω_2	Ω_2

Compare the result with the classifier designed in task 3.1.

Task 3.6

Simulate the first iteration of the backpropagation learning algorithm of a two-layer perceptron (one hidden layer), with 3 neurones in the hidden layer and two inputs and outputs, if:

- the learning rate is 0.1;
- sigmoid activation function: $z = \theta(y) = \dfrac{1}{1+\exp(-y)}$
- initial weight matrices $\mathbf{W}^{(1)}$ and $\mathbf{W}^{(2)}$: $\mathbf{W}^{(1)} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ -1 & 1 \end{bmatrix}$ $\mathbf{W}^{(2)} = \begin{bmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \end{bmatrix}$
- first learning sample: $\mathbf{x}=[1,2]$; required output: $f(\mathbf{x}) = [1, 0]$.

The activation values can be approximated from the table:

y	-3	-1	-0.9	-0.7	-0.5	-0.4	-0.3	-0.2	-0.1	-0.05	0.0
z=θ(y)	0.05	0.27	0.29	0.33	0.375	0.40	0.425	0.45	0.475	0.487	0.50

y	0.05	0.1	0.2	0.3	0.4	0.5	0.7	0.8	0.9	1.0	3.0
z=θ(y)	0.513	0.525	0.55	0.575	0.60	0.623	0.668	0.69	0.71	0.73	0.95

Task 3.7

Clustering. Let 15 training samples are available, as given in the table.

(1) Skip the class information and simulate the process of k-means clustering – show the code-book results after 2 first iterations if running the clustering for (a) 2 and (b) 4 classes.

(2) Compare the clustering results with the class information given in the last column of the table.

Index	0	1	2	*Class*
1	16	27	10	1
2	16	27	12	1
3	18	29	8	2
4	17	28	8	2
5	18	29	8	2
6	27	43	5	3
7	31	48	7	3
8	23	35	7	4
9	22	34	5	4
10	28	36	6	4
11	29	45	6	5
12	27	43	7	5
13	28	44	5	5
14	28	44	7	5
15	28	44	8	5

E3A. Pattern sequence recognition

HUMAN CAPITAL
HUMAN – BEST INVESTMENT!

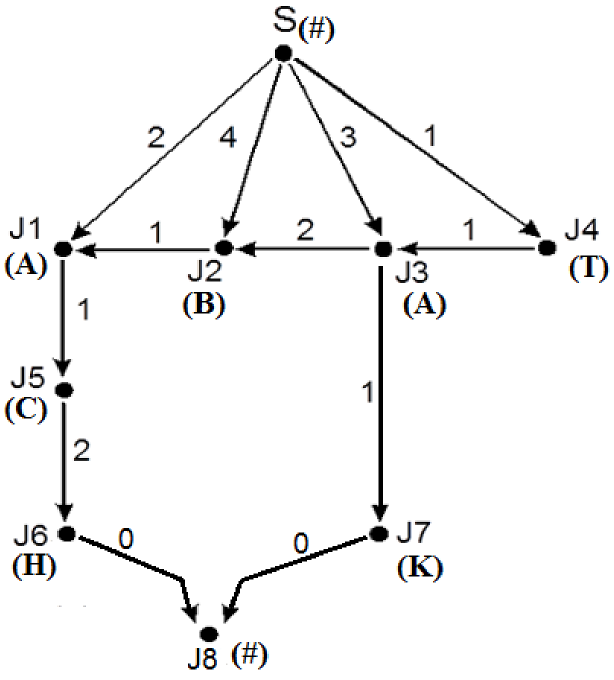
EUROPEAN UNION
EUROPEAN
SOCIAL FUND

Project is co-financed by European Union within European Social Fund

Task 3A.1

In the enclosed problem graph the decision costs are associated with links, while the letters accepted by nodes are given in brackets.

Simulate the prior path search process for this graph (from start node S to goal node J8) by the method of: **dynamic programming**.



Task 3A.2

Let an image-based text recognition system considers letters from a limited alphabet: { a, i, e, o, u, b, d, f, r}.

Design the problem graph (model) needed for dynamic programming-based recognition of words “road” and “four”, that tolerates selected letter errors:

- substitution errors of vowels (“e” instead of “o”, “u” instead of “a” and vice versa) and consonants (“f” instead of “r”, “b” instead of “d and vice versa”),
- deletion errors of inner-word letters and
- inner-word inclusion errors (“i” can be added).

For simplicity assume single-letter errors only, i.e. that an error cannot follow directly after another error.

Give the decision graph in dynamic programming for both models if the following observation sequence is given:

t	0	1	2	3	4	5	6
letter	#	r	o	i	u	r	Δ

Task 3A.3

Simplify the structure of **problem graph** defined in task 3A.2 by adding observation costs for every node *s*, i.e. turn the problem graph into a discrete Hidden Markov Model (with scalar costs of symbol emission per state).

Remark: skip the requirement of single-letter errors only.

E4. Iconic processing

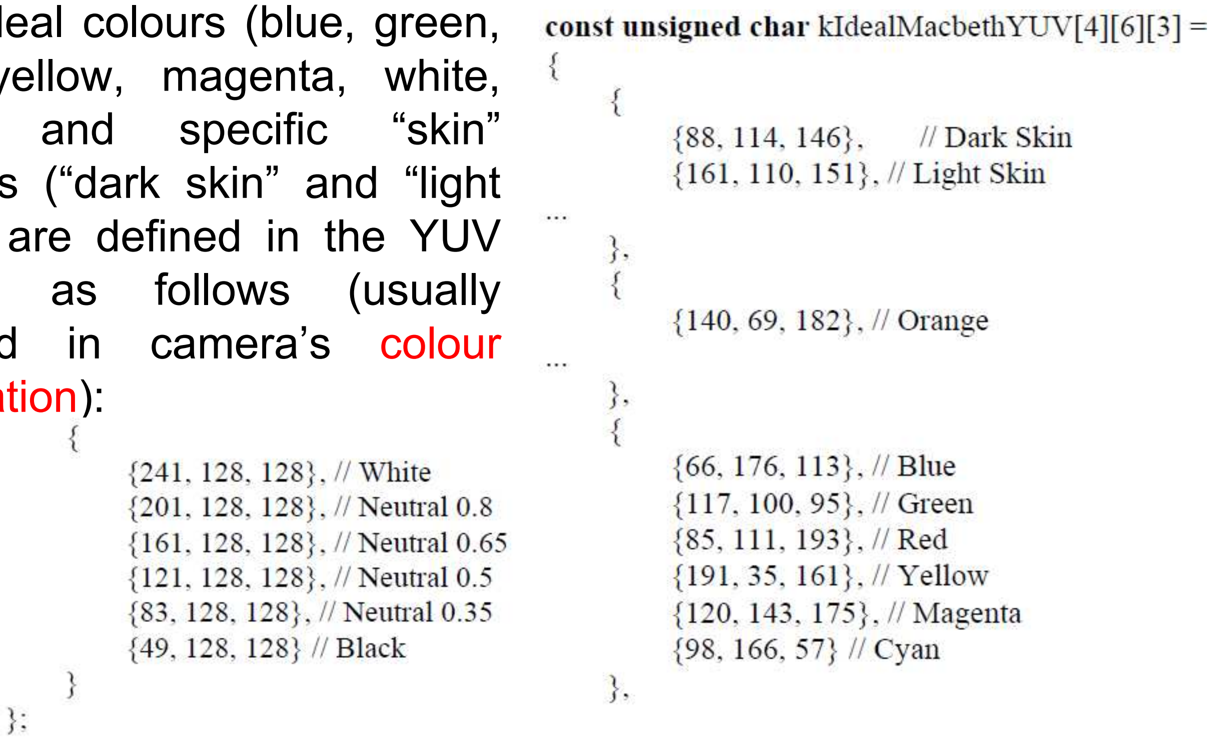
HUMAN CAPITAL
HUMAN – BEST INVESTMENT!

EUROPEAN UNION
EUROPEAN
SOCIAL FUND

Project is co-financed by European Union within European Social Fund

Task 4.1

The ideal colours (blue, green, red, yellow, magenta, white, dark) and specific “skin” colours (“dark skin” and “light skin”) are defined in the YUV space as follows (usually applied in camera’s colour calibration):



```
const unsigned char kIdealMacbethYUV[4][6][3] =
{
    {
        {88, 114, 146},     // Dark Skin
        {161, 110, 151}, // Light Skin
...
    },
    {
        {140, 69, 182}, // Orange
...
    },
    {
        {66, 176, 113}, // Blue
        {117, 100, 95}, // Green
        {85, 111, 193}, // Red
        {191, 35, 161}, // Yellow
        {120, 143, 175}, // Magenta
        {98, 166, 57} // Cyan
    },
    {
        {241, 128, 128}, // White
        {201, 128, 128}, // Neutral 0.8
        {161, 128, 128}, // Neutral 0.65
        {121, 128, 128}, // Neutral 0.5
        {83, 128, 128}, // Neutral 0.35
        {49, 128, 128} // Black
    }
};
```

Task 4.1 (1)

Assume that in a computer program we represent all the colour coefficients in terms of **8 bits** – using unsigned **integer** values from the interval [0, 255]:

1. Perform transformations between RGB and YUV colour spaces: for blue, green and red colours.
2. Find the linear Y-based normalization of the „skin" colour.

Solution.

1. Transform the coefficients Y(U-128)(V-128) into RGB.

Colors: → YUV → RGB

- Red → (85, 111, 193) →
- Green → (117, 100, 95) →
- Blue → (66, 176, 113) →

Task 4.2

Let an 8-level image of resolution 64 *x* 64 has the following histogram:

Pixel value *w*	Histogram value h(*w*)	$p_w(w)$?
0	90	
1	23	
2	1850	
3	656	
4	1029	
5	345	
6	22	
7	81	

Get the pdf of variable *w*. Perform *histogram stretching*. Assume an appropriate "cut-off" threshold. Give graphical illustration of both distributions – before and after the stretching operation.

Task 4.3

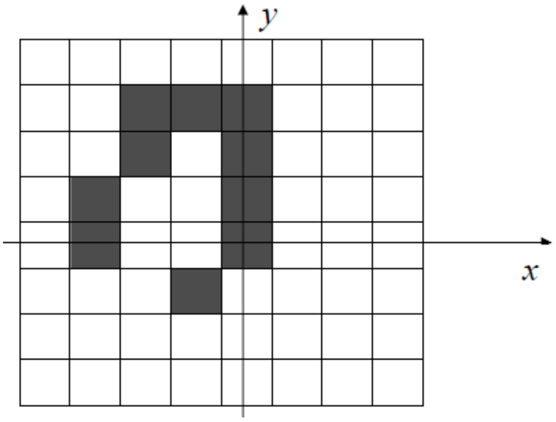
Let a 8-level image of resolution $64\ x\ 64$ has the following histogram:

Pixel value w	Histogram value h(w)	$p_w(w)$
0	790	0.19
1	1023	0.25
2	850	0.21
3	656	0.16
4	329	0.08
5	245	0.06
6	122	0.03
7	81	0.02

Perform *histogram equalization*. Give graphical illustration of both distributions – before and after the equalization operation. Determine the number of distinct levels in the output image.

Task 4.4

Let the binary image of some „character” pattern be given, as specified below. The image coordinate system is also shown.



Simulate the *moment-based normalization* procedure for this pattern. Determine intermediate transformations and show intermediate results.

Task 4.5

Let the following image be given.

3	2	8	2
5	1	7	1
9	1	6	1
8	0	5	2

1. Determine the threshold for image binarization using the Otsu method. Show intermediate results.
2. Perform image thresholding.

Task 4.6

Apply different *edge operators* (Roberts cross, *central differences, Sobel, Scharr*) to the two images given below. Obtain the images of edge strengths and orientations. Compare the results.

1)

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

2)

1	2	1	2
2	2	3	6
3	3	6	6
3	6	6	5
5	6	6	6

Task 4.7

Simulate 3 different *edge thinning* procedures as applied to the results of the Sobel operator for the first image in task 4.6.

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

E5. Boundary-based image segmentation

HUMAN CAPITAL
HUMAN – BEST INVESTMENT!

EUROPEAN UNION
EUROPEAN
SOCIAL FUND

Project is co-financed by European Union within European Social Fund

Task 5.1

Simulate the process of chain detection by the method of "*hysteresis thresholds*" for the edge image below. Assume the following thresholds: lower T = 20, upper T_0= 32, T2 = 30°. The edge orientation number O_{NUM} = 32.

y \ x	1: (**s** , r)	2	3	4
1	21, 8	22, 8	11, 10	6, 12
2	30, 8	34, 8	35, 9	20, 10
3	6, 11	37, 8	36, 9	33, 10
4	13, 14	12, 15	20, 8	30, 10
5	12, 13	12, 10	11, 7	12, 8

Output data:

1) “Chain image” with segment indices starting from 1.

2) Every segment has a **start edge** and an **end edge.**

3) Every edge element has a **successor** element (or zero).

Task 5.2

Let 8 following edge elements [$\mathbf{e}_i=(x_i, y_i)$, r(e_i)] be given (where the number of discrete directions is set to: O_{NUM} = 8), where r( ) – orientation index:

i	x_i	y_i	r(e_i)
1	1.0	1.0	1
2	1.0	2.0	2
3	3.0	3.0	3
4	2.0	2.0	1
5	0.0	2.0	1
6	4.0	0.0	1
7	3.0	2.0	2
8	-1.0	-1.0	3

Obtain points in **Hough** space (designed for **straight line** detection) that correspond to given edge elements.

What lines can been detected in this particular Hough space. Give them in parametric line form: $y = ax + b$.

Task 5.3

Fill the **Hough** accumulator (designed for **circle centre** detection) if following edge elements are given:

i	x_i	y_i	∠ (r(**e**))
1	2.0	4.0	270 °
2	0.0	2.0	0 °
3	2.0	0.0	90 °
4	4.0	2.0	180 °
5	0.0	0.0	45 °
6	4.0	4.0	225 °

Find the candidate for circle's centre point (x_c, y_c).

Task 5.4

Let the following points in the Hough accumulator, created for circle centre detection, are given:

i	a_i	b_i
1	-2	0.5
2	0	1
3	1	1.5
4	2	2
5	4	3

Get the line equation (in Hough space) best approximated (by the LSE approach) by above observations.

To which circle centre point in image space this line corresponds?

Task 5.5

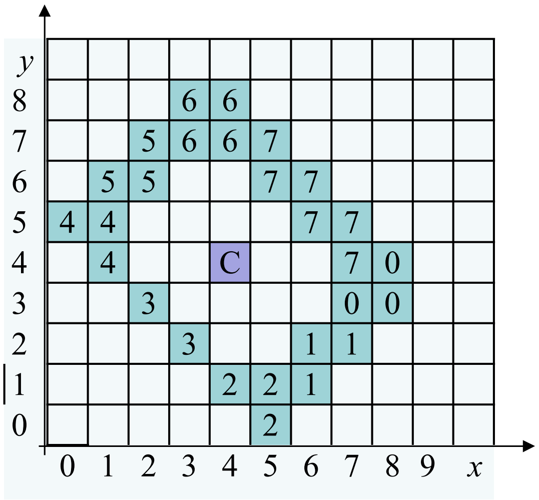
Simulate the contour recognition process by a generalized Hough Transform, given a rectangle-like contour model with smooth vertices and with basic size 2 × 3. The particular edge image is shown below – 8 orientations are distinguished ($O_{NUM}=8$) (orientations are given only).

Define the contour model in terms of 8 border points, alternatively for two scales ($s=1,2$) and two rotations ($\varphi = 0^o, -45^o$).

Solution:

Centre point $C = (4, 4)$.

$s=2$, $\varphi = -45^o$.



y											
8				6	6						
7			5	6	6	7					
6		5	5			7	7				
5	4	4					7	7			
4		4			C			7	0		
3			3					0	0		
2				3			1	1			
1					2	2	1				
0						2					
	0	1	2	3	4	5	6	7	8	9	x

E6. Region-based image segmentation

HUMAN CAPITAL
HUMAN – BEST INVESTMENT!
EUROPEAN UNION
EUROPEAN SOCIAL FUND
Project is co-financed by European Union within European Social Fund

Task 6.1

Let the following image of size 4×4 image is given.

8	3	21	33
3	4	20	32
6	6	19	31
12	7	18	30

Simulate step-by-step the combined region detection procedure (i.e. with SPLIT-and-MERGE steps and with region merging steps) applied to this image with the thresholds:

- for split-and-merge $\theta = 4$,
- for region merging $\theta_A = 3$, and
- for region merging with adaptive threshold $\theta(F) = 3 - F/16 * (3 - 11)$.

Task 6.2

Let a texture f is given (of size 5×5 and 4 intensity levels).

Define the intensity co-occurrence matrices (ICM) $G(1,0)$ and $G(1,1)$ (i.e. distance one for horizontal and vertical direction) for the texture below.

Calculate the contrast features for such ICM.

Calculate the combined ICM $G(1)$ from $G(1,0)$ and $G(1,1)$.

Shift the texture cyclically by one column ($f_{shifted}$) or produce a mirror texture (f_{mirror}) w.r.t. the Y axis.

Are there any differences of the ICM's for the shifted or mirrored texture?

$$f = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 2 & 2 & 2 & 0 \\ 2 & 2 & 3 & 3 & 0 \\ 2 & 2 & 2 & 2 & 2 \end{bmatrix}$$

$$f_{shifted} = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 2 & 2 & 2 \\ 0 & 2 & 2 & 3 & 3 \\ 2 & 2 & 2 & 2 & 2 \end{bmatrix}$$

$$f_{mirror} = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 2 & 2 & 2 & 0 \\ 0 & 3 & 3 & 2 & 2 \\ 2 & 2 & 2 & 2 & 2 \end{bmatrix}$$

Task 6.3

Let a texture f is given (of size 5×5 and 4 intensity levels).

Obtain the histograms of sums and differences for the displacement vector $v=(0,1)^T$.

Shift the texture cyclically by one column and produce a mirror texture along the Y axis.

$$f=\begin{bmatrix}0&0&1&1&0\\0&0&1&1&0\\0&2&2&2&0\\2&2&3&3&0\\2&2&2&2&2\end{bmatrix}$$

$$f_{shifted}=\begin{bmatrix}0&0&0&1&1\\0&0&0&1&1\\0&0&2&2&2\\0&2&2&3&3\\2&2&2&2&2\end{bmatrix}\qquad f_{mirror}=\begin{bmatrix}0&1&1&0&0\\0&1&1&0&0\\0&2&2&2&0\\0&3&3&2&2\\2&2&2&2&2\end{bmatrix}$$

Are there any differences of the sum and difference histograms if compared to the previous histograms?

E7. Object recognition

HUMAN CAPITAL
HUMAN – BEST INVESTMENT!

EUROPEAN UNION
EUROPEAN
SOCIAL FUND

Project is co-financed by European Union within European Social Fund

Task 7.1

(Calibration of the camera) Let 3 points in 3D space are specified in global coordinates (X_i, Y_i, Z_i):

Point	X_i	Y_i	Z_i
P1	-2	7	1
P2	-1	6	1
P3	-0	5	2

Let the corresponding image points (x_i, y_i) are:

Image point	x_i	y_i
p1	-1	4
p2	0	3
p3	1	1

Assuming, there is no rotation between global coordinates and camera coordinates and an unary scaling coefficient:

Task 7.1 (2)

1. Show, that the 3 pairs of points are sufficient for camera calibration;

2. Check, if it leads to a system of linear equations that have an unique solution (i.e. perform LSE minimization and check the system’s determinant)

3. How many points are required if there is a non-zero rotation around the Z axis?

Task 7.2

Assume, a single observation vector is given, $z(1) = [4, 2]^T$, of a hidden constant state vector, $s = [x_1, x_2]^T$, for a system with the following observation model:

$$z(t) = A\, s + w(t), \quad t=1,...,N;$$

where $A = \begin{bmatrix} 1 & 0.1 \\ 0.1 & 2 \end{bmatrix}$

Gaussian distributions of $p(\mathbf{s}) \sim N([1, 1]^T, \boldsymbol{P})$ and $p(\boldsymbol{w}) \sim N(\mathbf{0}, \boldsymbol{R})$, with covariance matrices:

$$\boldsymbol{P} = \begin{bmatrix} 1 & 0.2 \\ 0.3 & 2 \end{bmatrix} \qquad \boldsymbol{R} = \begin{bmatrix} 1 & 0.2 \\ 0.3 & 2 \end{bmatrix}$$

A) Derive the analytic solution of iterative MAP estimation.

B) Use the above data to compute the MAP estimate $s^{MAP}(1)$.

Task 7.3 (1)

Propose a general object recognition strategy for combined top-down and bottom-up matching of model with image data, and express it in terms of A* search.

Solution

A) Define a model (problem graph) that contains two hierarchies:

1. a vertical (is-part-of) hierarchy
2. a horizontal (is-specialization-of) hierarchy.

The IMAGE DATA correspond to leafs of such problem graph.

B) Define several INFERENCE steps (i.e. ways of matching DATA to MODEL NODES) to create modified concepts Q(A) and INSTANCES I(A) of model concepts A

Task 7.3 (2)

Solution (cont.) C) For a GOAL concept define costs for A* search:

- the costs of already matched parts and
- the expected costs of matching yet unmatched parts.

For an instance of some intermediate GOAL define expected remaining costs of reaching an instance of the ULTIMATE GOAL concept.

D) Consider the following iterative strategy:

```
1. Select an intermediate goal node A – create Q(A).
2. Make top-down model expansion for Q(A)
3. Make bottom-up model instantiation based on DATA.
4. If I(A) is ultimate goal then RETURN(I(A))
   else infer next goal $A^1$ from I(A):
        create Q($A^1$),    set A=$A^1$ and    repeat from step 2.
```

E8. Speech pre-processing

HUMAN CAPITAL
HUMAN – BEST INVESTMENT!
EUROPEAN UNION
EUROPEAN SOCIAL FUND

Project is co-financed by European Union within European Social Fund

Task 8.1

The Euler formula. Using the Euler formula compute the frequency-based decomposition of the following function:

$$f(t) = \begin{cases} 1, & if 0 \le t < \pi/2 \\ -1, & if \ \pi/2 < t < 3\pi/2 \\ 1, & if \ 3\pi/2 < t \le 2\pi \end{cases}$$

Solution. We assume that the function $f(t)$ is a 2π–periodic function. Then the Fourier coefficients can be computed using the following formulas:

$$a_k = \frac{1}{\pi}\int_0^{2\pi} f(t)\cos(2\pi f_0 kt)dt \qquad b_k = \frac{1}{\pi}\int_0^{2\pi} f(t)\sin(2\pi f_0 kt)dt$$

$$a_k = \frac{1}{\pi}\int_0^{2\pi} f(t)\cos(2\pi f_0 kt)dt = \frac{1}{\pi}[\int_0^{\pi/2} 1\cdot\cos(2\pi f_0 kt)dt + \int_{\pi/2}^{3\pi/2} -1\cdot\cos(2\pi f_0 kt)dt + \int_{3\pi/2}^{2\pi} 1\cdot\cos(2\pi f_0 kt)dt]$$

Task 8.1 (2)

Solution (cont.) Observe - the base frequency is: $f_0 = \frac{1}{2\pi}$

$$a_k = \frac{1}{\pi}(\frac{\sin(kt)}{k}\Big|_0^{\pi/2} - \frac{\sin(kt)}{k}\Big|_{\pi/2}^{3\pi/2} + \frac{\sin(kt)}{k}\Big|_{3\pi/2}^{2\pi}) = \frac{1}{k\pi}[2\sin(k\pi/2) - 2\sin(3k\pi/2)]$$

$$a_k = \begin{cases} 0, & if \ k \ is \ even \\ \frac{4}{k\pi}\sin(\frac{k\pi}{2}), & otherwise \end{cases}$$

$$b_k = \frac{1}{\pi}(-\frac{\cos(kt)}{k}\Big|_0^{\pi/2} + \frac{\cos(kt)}{k}\Big|_{\pi/2}^{3\pi/2} - \frac{\cos(kt)}{k}\Big|_{3\pi/2}^{2\pi}) = \frac{1}{k\pi}[2\cos(3k\pi/2) - 2\cos(k\pi/2)] = \frac{1}{k\pi}[0-0] = 0$$

Task 8.2

Prove the following statements:

1. "The DFT transformation matrix $\boldsymbol{D}_M$ is a symmetric and unitary matrix (i.e. $\boldsymbol{D}_M \boldsymbol{D}_M^* = M\,\mathbf{I}$)."

2. "If the input signal $\boldsymbol{x}(t)$ has a real part only, then the vector $\mathbf{F}$ of Fourier coefficients is a symmetric conjugate around $M/2$, i.e.

$$\mathrm{F}_{M/2-k} = \mathrm{F}^*_{M/2+k}\text{ , for } k = 0, ..., M/2 -1. \text{"}$$

3. "Furthermore for a real-valued $\boldsymbol{x}(t)$ if M is an even number then F_0 and $\mathrm{F}_{M/2}$ are real numbers too."

Task 8.3

Simulate the steps of FFT computation (with decimation in frequency domain) for $M = 8$ and the following signal $x[n]$:

n	0	1	2	3	4	5	6	7
x[n]	20	10	5	5	5	0	-10	-10

Task 8.4

The following signal $x[n]$ is given:

n signal	1	2	3	4	5	6	7	8	9	10
$x[n]$	1	2	3	4	0	-1	-2	-3	-4	0

n signal	11	12	13	14	15	16	17	18	19	20
$x[n]$	1	0	0	-1	-1	0	1	1	0	-1

1) Compute this signal's auto-correlation factors r_k , k=1,…,10.

2) Estimate the fundamental frequency of this signal, assuming a sampling frequency of 1 sample / ms.

E9. Speech features

HUMAN CAPITAL
HUMAN – BEST INVESTMENT!

EUROPEAN UNION
EUROPEAN
SOCIAL FUND

Project is co-financed by European Union within European Social Fund

Task 9.1

A) Get the Fourier coefficients for the signal $x[n]$, given below:

n	0	1	2	3	4	5	6	7
$x[n]$	20	10	5	5	5	0	-10	-10

B) Get the mean value of $x[n]$ and generate the zero-mean signal: $x_0[n] = x[n] - \text{mean}(x[n])$. Compute the $\text{DFT}(\text{x}_0)$. Compare results with A).

C) By padding with 8 zeros the above signal is extended to 16 samples: $x1[n] = [x[n] \mid 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ ]$, i.e.

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$x1[n]$	20	10	5	5	5	0	-10	-10	0	0	0	0	0	0	0	0

Compute the $\text{DFT}(x1)$. Compare results with A).

D) Obtain the mean value of $x1[n]$ and generate the zero-mean signal: $x1_0[n] = x1[n] - \text{mean}(x1[n])$. Compute the $\text{DFT}(x1_0)$. Compare results with A).

Task 9.2

Compute MFC features for the following signal frame:

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$x[n]$	20	10	5	5	5	0	-10	-10	0	0	0	0	0	0	0	0

Assume that the sampling rate is 8 kHz. Use 3 triangle filters uniformly located according to the Mel-scale.

Task 9.3

Compute the set of 4 LPC features for the following signal frame:

n	0	1	2	3	4	5	6	7
$x[n]$	20	10	5	5	5	0	-10	-10

Define and solve the linear system given by a Toeplitz matrix:

$$\begin{pmatrix} r_0 & r_1 & r_2 & \cdots & r_{m-1} \\ r_1 & r_0 & r_1 & \cdots & r_{m-2} \\ \vdots & & & & \vdots \\ r_{m-1} & r_{m-2} & r_{m-3} & \cdots & r_0 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} = - \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{pmatrix}$$

E10. Phonetic speech model

HUMAN CAPITAL
HUMAN – BEST INVESTMENT!

EUROPEAN UNION
EUROPEAN
SOCIAL FUND

Project is co-financed by European Union within European Social Fund

Task 10.1

Using the 4 features: F1, F2, tongue location and mouth opening, find the expected locations of cluster centers of vowel samples.

Task 10.2

Give (A) the phonetic description (pronunciation) in terms of phonemes and (B) the expected spectrograms, of following spoken words: (1) "ok.", (2) "three", (3) "end".

Solution. A)

Word	Pronunciation
ok	/oU/ /k^c/ /k^h/ /eI/
three	/T/ /9r/ /i:/
end	/E/ /n/ /d/

Task 10.3

Create a ten-digit word recogniser – specify the following:

1. the pronunciations of considered words,
2. the classification of phonemes into 8 context categories,
3. the number of sub-phonemes for each considered phoneme,
4. all sub-phonemes needed for signal frame classification.

E11. Word and sentence recognition

HUMAN CAPITAL
HUMAN – BEST INVESTMENT!

EUROPEAN UNION
EUROPEAN
SOCIAL FUND

Project is co-financed by European Union within European Social Fund

Task 11.1 (1)

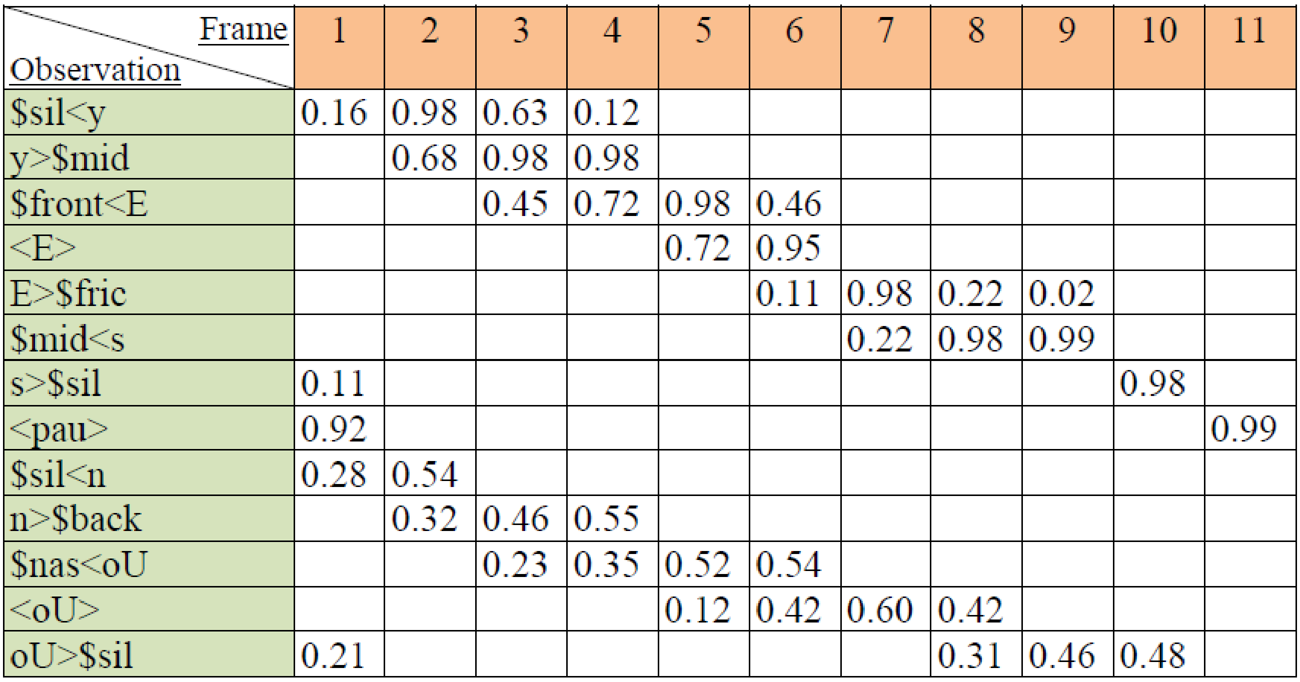
Define the **HMM model** for phonetic representation of **spoken words** "yes" and "no" in terms of sub-phonemes given below. Assume a **discrete and scalar** form of the observation model with single class-per-state only.

The observations are summarized in a sub-phoneme probability matrix, given on next page.

By the Viterbi search method find the best path in this observation matrix, w.r.t. the "yes"- and "no"- models.

Specify step-by-step actions and results of the Viterbi search-process.

Task 11.1 (2)



Frame Observation	1	2	3	4	5	6	7	8	9	10	11
$sil<y	0.16	0.98	0.63	0.12							
y>$mid		0.68	0.98	0.98							
$front<E			0.45	0.72	0.98	0.46					
<E>					0.72	0.95					
E>$fric						0.11	0.98	0.22	0.02		
$mid<s							0.22	0.98	0.99		
s>$sil	0.11									0.98	
<pau>	0.92										0.99
$sil<n	0.28	0.54									
n>$back		0.32	0.46	0.55							
$nas<oU			0.23	0.35	0.52	0.54					
<oU>					0.12	0.42	0.60	0.42			
oU>$sil	0.21							0.31	0.46	0.48	

Task 11.2

A) Design a Markow Model (HMM), required for the recognition of the **written** string "**wtorek**" in **images**. Assume a **discrete vector** form of the emission (output) model.

B) Describe and compute a particular decision space that is built by a Viterbi search for given word model, if (during the segmentation process) five letters (segments) have been detected with following probabilities resulting from single-letter classification:

letter \ segment	1	2	3	4	5
w	0.8	0.0	0.0	0.0	0.0
o	0.1	0.8	0.0	0.1	0.0
r	0.0	0.0	0.8	0.0	0.0
a	0.1	0.1	0.1	0.8	0.0
t	0.0	0.0	0.1	0.0	0.9
k	0.0	0.0	0.0	0.0	0.1
e	0.0	0.1	0.0	0.1	0.0

Thank you

HUMAN CAPITAL
HUMAN – BEST INVESTMENT!

EUROPEAN UNION
EUROPEAN
SOCIAL FUND