Faculty of Electronics and Information Technology
Warsaw University of Technology
Syllabi of Electrical and Computer Engineering

Course Title	
	Image and Speech Recognition

Course Format	Lectures	Tutorials	Laboratory	Project
Hours/Week	2	1	-	1

Course Code	EIASR	Programme	M.Sc.
ECTS Points	6	Status	Elective
Placement (default)	Sem. 1	Specialization	CSN
Form of Passing	Pass	Course Class	ANGL, ATP

Required Prerequisites	
Suggested Prerequisites	
Similar Courses (incl. courses in Polish)	ROSM
ERASMUS Subject Code	11.3 Informatics, Computer Science 06.5 Electronic Engineering, Telecommunications

Course Description	
	Objective
	The goal is to learn about basic methods of pattern recognition and about algorithms applied for digital image- and speech-analysis. The students will be able to design image and speech analysis programs dealing with problems of pattern (image or speech) processing, pattern segmentation and object (or word) recognition.
	Summary
	In the first part elements of the pattern recognition theory are introduced. Among them are: basic feature space transformations PCA, LDA and ICA, pattern clustering approaches and various classifier types (linear discriminate-, a SVM-classifier, the Bayes, k-NN, and MLP-classifiers). The image analysis part covers low-level processing, segmentation and object recognition problems. The topics of low-level processing include: viewing geometry, camera calibration, color spaces and image compression, image binarization, normalization and filtering. Among image segmentation methods we introduce algorithms for edge following and line segment detection, Hough transforms, homogeneous region detection, texture- and shape description. Approaches to model-based sequence and object recognition are shown: dynamic programming search, heuristic matching strategies, graph search and MAP estimation. The area of speech recognition starts with basic signal processing in the time and frequency domain (speech source detection, basic frequency estimation, noise elimination, windowed Fourier transform, FFT). Then basic feature detection approaches are presented, leading to the MFCC and LPC-based features. We illustrate the phonetic model of speech by spectrograms for different phoneme types and we also introduce the tri-phone model. The spoken word recognition problem is solved by the use of Hidden Markow Models for word modeling with the Baum-Welch training and Viterbi search methods.
	Summary in Polish
	W pierwszej części przedstawione są podstawy teorii rozpoznawania wzorców. Wśród nich omawia się podstawowe przekształcenia przestrzeni cech: PCA, LDA i ICA, zagadnienia grupowania (klasteryzacji) cech i różne rodzaje klasyfikatorów: maszyna liniowa i SVM,

klasyfikator Bayesa i według najbliższych sąsiadów, oraz wielowarstwowy perceptron.
Druga część dotyczy analizy obrazu i obejmuje zagadnienia: reprezentacja i przetwarzanie niskiego poziomu, segmentacja obrazu i rozpoznawanie obiektów. Wśród problemów niskiego poziomu przetwarzania obrazu omawia się: geometrię odwzorowania sceny, kalibrację kamery, reprezentację barw i kompresję obrazu oraz binaryzację, normalizację i filtrację obrazu. Wprowadzane są algorytmy segmentacji obrazu, takie jak: śledzenie krawędzi, wyznaczanie linii, transformaty Hougha, wyznaczanie obszarów jednorodnych, tworzenie opisów tekstury obszaru i 2-wymiarowych kształtów. Pokazywane są podejścia do problemu rozpoznawania napisów i obiektów: przeszukiwanie metodą programowania dynamicznego, heurystyczne strategie dopasowania danych z modelem, przeszukiwanie grafów i estymacja MAP.
Problematyka rozpoznawania mowy rozpoczyna się od podstawowych sposobów analizy sygnału w dziedzinie czasu i częstotliwości (np. wykrycie sygnału mowy, oszacowanie podstawowej częstotliwości mowy, usuwanie szumu, okienkowa transformata Fouriera, szybka transformata Fouriera). Następnie przedstawia się podstawowe sposoby pozyskiwania cech, prowadzące do cech mel-cepstralnych MFCC i cech według liniowej predykcji LPC. Ilustrujemy fonetyczny model mowy spektrogramami tworzonymi dla dźwięków różnych typów a także przedstawiamy trzyfonowy model głoski. Dla rozwiązania problemu rozpoznawania słów stosujemy Ukryte Modele Markowa, uczenie według Bauma-Welcha i algorytm poszukiwania Viterbiego.

Lectures

1. Introduction to pattern recognition. (2h)
    - 1.1 Basic terms and pattern recognition approaches
    - 1.2 Structure and tasks of image recognition systems
    - 1.3 Structure and tasks of speech recognition systems
    - 1.4 Sampling and digitalization
2. Pattern transformation. (2h)
    - 2.1 Data-dependent feature space transformations
    - 2.2 Principal Component Analysis
    - 2.3 Linear Discriminate Analysis
    - 2.4 Independent Component Analysis
3. Pattern classification. (3 h)
    - 3.1 Pattern classification problem
    - 3.2 Linear discriminate classifier
    - 3.3 Stochastic classifiers (Bayes, ML)
    - 3.4 Geometric classifier
    - 3.5 k-nearest neighbor classifier
    - 3.6 Support Vector Machine
    - 3.7 Multilayer perceptron
4. Pattern clustering. (1 h)
    - 4.1 k-means and unsupervised EM
    - 4.2 PCM and vector quantization
    - 4.3 Competitive learning
5. Digital image representation. (2 h)
    - 5.1 Scene viewing geometry
    - 5.2 Camera model and camera calibration
    - 5.3 Internal image representation (color spaces)
    - 5.4 External image representation
    - 5.5 JPEG and MPEG compression
6. Image processing. (2 h)
    - 6.1 Image binarization
    - 6.2 Object normalization
    - 6.3 Basic image filters
    - 6.4 Edge detection
    - 6.5 Edge thinning

	7. Line-based image segmentation. (2 h) 7.1 Edge following 7.2 Line segment detection 7.3 Hough transforms 8. Region-based image segmentation. (2 h) 8.1 Homogeneous region detection 8.2 Texture descriptors 8.3 Shape descriptors 8.4 Active contours 9. Object recognition. (2 h) 9.1 Dynamic programming search 9.2 Object model-to-image matching – heuristic strategies 9.3 Generic model-to-image matching by search 9.4 Object recognition by state estimation 10. Speech signal representation. (2 h) 10.1 Digital audio signal representation 10.2 The Fourier transform 10.3 The Fast Fourier Transform 10.5 Speech signal pre-preprocessing 11. Speech feature detection. (3 h) 11.1 Mel-cepstrum features 11.2 LPC-based features 11.3 Signal window classification 12. Acoustic-phonetic speech models. (1 h) 12.1 Phonetic categories 12.2 Typical spectrograms 12.3 Context-dependent sub-sound model 13. Word recognition. (2 h) 13.1 Hidden Markow Models for speech 13.2 Viterbi search 13.3 Baum-Welch training 14. Spoken sentence recognition. (2 h) 14.1 Natural language processing (NLP) 14.2 N-gram models 14.3 Token passing search Midterm test (1 h) Final test (1 h) Remark. The lecture chapters are given in following order:1 (Introduction), 5, 10, 6, 2 (low-level processing), 7, 8, 11, 12 (segmentation), 3, 4 (classification and clustering), 9, 13, 14 (object and word recognition).
	Tutorials
	During exercises most of the algorithms for image and speech analysis introduced in the lecture are practically explained. Every lecture section is accomplished by several exercise examples. E1. Image representation and processing (2 h) E2. Speech signal transformation (2 h) E3. Pattern transformation (2 h) E4. Image segmentation (3 h) E5. Speech segmentation (2 h) E6. Pattern classification and clustering (2 h) E7. Word and object recognition (2 h)
	Laboratory
	...
	Project

	The project introduces two open source program libraries (e.g. OpenCV and Marf) that contain implementations of basic tasks in image and speech analysis as well as pattern classification. The students are selecting and running a project work in which they solve a pattern classification task and secondly an image or speech analysis task. The program implementation can be made in C++, C#, Java, Pascal/Delphi or Matlab.
	Assessment Method
	There is a continuous assessment method applied in this course. The points are collected during the semester time, a they can come from two tests (2 x 30 pts.), covering the lecture and exercise material, and from a project work (0 - 40 pts.). The final result is based on the following pattern: • A: 91-100 points • B: 81-90 points • C: 71-80 points • D: 61-70 points • E: 51-60 points • F/FX: 0 – 50 points
	References
	1. W. Kasprzak: *Image and Speech Recognition*. (In polish: *Rozpoznawanie obrazów i sygnałów mowy*. WUT publishing house, Warszawa, 2009. 2. R. Duda, P. Hart, D. Stork: *Pattern Classification.* 2nd edition, John Wiley & Sons, New York, 2001. 3. I. Pitas: *Digital Image Processing Algorithms and Applications*, John Wiley, New York etc. 2000. 4. L. Rabiner, B.-H. Juang: *Fundamentals of speech recognition.* Prentice Hall, New York, 1993. 5. J. Benesty, M.M. Sondhi, Y. Huang (eds): *Handbook of Speech Processing*, Springer, Berlin, 2008.

Responsible Person	prof. nzw. dr hab. inż. Włodzimierz Kasprzak
Date of Last Revision	20th May 2009

Explanations:

Programme	**B.Sc. | M.Sc.**
Status	**Compulsory | Elective**
Specialization	**Common | CSN | TCM**
Form of Passing	**Examination | Pass**
ERASMUS Subject Code	**11.3 Informatics, Computer Science |** **06.5 Electronic Engineering, Telecommunications |** **08.0 Humanities**