Audio processing, Video processing and Computer vision

Academic Year: (2022 / 2023) **Review date: 20-05-2022**

Department assigned to the subject: Department of Signal and Communications Theory

Coordinating teacher: GONZALEZ DIAZ, IVAN

Type: Compulsory ECTS Credits : 6.0

Year : 4 Semester : 1

REQUIREMENTS (SUBJECTS THAT ARE ASSUMED TO BE KNOWN)

Neural Networks
Signals and Systems
Machine Learning I and II

OBJECTIVES

Students must achieve the following objectives:

1) Know image, audio, sppech and video signals, as well as their main parameters and the digitization process.
2) Know the most important techniques of image, video and audio processing, as well as the main tasks in computer vision and audio.
3) Apply machine learning and deep learning techniques studied in previous subjects to the analysis of audiovisual content.
4) Develop intelligent applications that involve the automatic analysis of audiovisual content.

DESCRIPTION OF CONTENTS: PROGRAMME

The course is divided into two main blocks abased on the signal modalities: on the one hand, image and video and, on the other, voice and audio.
In both cases, signals and their main characteristics are presented first, including certain notions of the visual and auditory systems. Next, the most common techniques for each signal processing are studied, illustrating their use in selected applications. Finally, most modern approaches are introduced, based on the application of deep learning (e.g. CNNs and RNNs), which constitute nowadays the state of the art of technology.

The course program is organized as follows:

Block 1: Processing of visual signals: image and video
==========
Topic 1: Introduction to digital video and images
Topic 2: Fundamentals of image and video processing
Topic 3: Image Representation: low-level descriptors
Topic 4: Image Segmentation
Topic 5: Convolutional Neural Networks (CNNs) for image classification
Topic 6: Other applications of CNNs in visual analysis

Block 2: Processing of speech and audio signals
==========
Topic 7: Fundamentals of digital audio and speech: generation, perception and digitization
Topic 8: Time-located analysis for speec and audio signals
Topic 9: Low-level speech and audio descriptors
Topic 10: Recurrent Neural Networks (RNNs) for sequential data analysis

LEARNING ACTIVITIES AND METHODOLOGY

Two teaching activities are proposed: lectures and lab sessions.

LECTURES

The lecture sessions will be supported by slides or by any other means to illustrate the concepts

explained. In these classes the explanation will be completed with examples.

In these sessions the student will acquire the basic concepts of the course. It is important to highlight that these classes require the initiative and the personal and group involvement of the students (there will be concepts that the students themselves should develop).

LABORATORY SESSIONS

This is a course with a high practical component, and students will attend to laboratory sessions very often. In them, the concepts explained during the lectures will be put into practice using the programming language python, and software libraries for image analysis and computer vision (scikit-image, PIL, OpenCV), audio analysis (scikit-sound), machine learning (scikit-learn) and deep learning (pytorch). In the laboratory, machines equipped with high-performance GPUs are available and but students can also use free distributed computing systems such as Google Colab.

ASSESSMENT SYSTEM

SE1: FINAL EXAM. Knowledge, skills and abilities acquired throughout the course will be assessed globally.
SE2: CONTINUOUS ASSESSMENT. Works, presentations, debates, exercises, practices and work in the workshops throughout the course will be evaluated.

% end-of-term-examination:	40
% of continuous assessment (assigments, laboratory, practicals…):	60

BASIC BIBLIOGRAPHY

- Ian Goodfellow, Yoshoua Bengio, and Aaron Courville Deep Learning, The MIT Press, 2016
- Ken C. Pohlmann Principles of Digital Audio (5th Edition), McGraw-Hill/TAB Electronics, 2005
- Ken C. Pohlmann Principles of Digital Audio (5th Edition), McGraw-Hill/TAB Electronics, 2005
- N. Morgan and B. Gold Speech and Audio Signal Processing: Processing and Perception of Speech and Music, John Wiley & Sons, Inc. New York, NY, USA, 1999
- Rafael C. González and Richard E. Woods Digital Image Processing (4th Edition), Pearson, 2018

ADDITIONAL BIBLIOGRAPHY

- D. O'Shaughnessy Automatic speech recognition: History, methods and challenges, Pattern Recognition, 41 (10) pp. 2965-2979, 2008
- David A. Forsyth and Jean Ponce Computer Vision: A Modern Approach (2nd Edition), Pearson , 2012
- S. Huang, A. Acero, H.W. Hon Spoken Language Processing: A Guide to Theory, Algorithms and System Development, Prentice Hall, 2001
- Wilhelm Burger and Mark J. Burge Principles of Digital Image Processing: Fundamental Techniques, Springer-Verlag, 2009
- Wilhelm Burger and Mark J. Burge Principles of Digital Image Processing: Core Techniques, Springer-Verlag, 2009