Audio and Visual Analytics

Academic Year: (2020 / 2021)   **Review date: 09/07/2020 17:01:56**

Department assigned to the subject: Signal and Communications Theory Department

Coordinating teacher: DIAZ DE MARIA, FERNANDO

Type: Electives ECTS Credits : 6.0

Year : 4 Semester :

OBJECTIVES

The main goal of this course is to provide the students with the theoretical and methodological knowledge about algorithms and methods for multimedia information indexing and retrieval.

At the end of the course the students are expected to have acquired (or progress in the acquisition -for transversal competences-) the following competences:

1. TRANSVERSAL/GENERAL COMPETENCES:
1.1 Personal work abilities.
1.2 Analysis and Synthesis abilities.
1.3 Abilities for applying theoretical concepts to practical uses.
1.4 Abilities related to team work and collaboration.
1.5 Abilities related to oral and written presentations.

2. SPECIFIC COMPETENCES:
2.1 To understand the fundamentals of audio-visual data analytics and its applications.
2.2 To understand the basics of speech, audio, image and video description and representation.
2.3 To understand the methods and technologies used for classification, detection or recognition of voice, audio, image or video.
2.4 Ability to design and implement the above methods and technologies in practical problems of automatic analysis of voice, audio, image and video.

CB1, CB2
CG3, CG11
ETEGITT9, ETEGITT3

DESCRIPTION OF CONTENTS: PROGRAMME

The modern information overload problem caused by the availability of enormous amounts of information through internet makes it necessary to design systems that allow us to find the information we search and filter or personalize the information according to our needs.

The objective of this course is to introduce the basic techniques of voice, audio, image and video processing, with a notable practical orientation. This will be supported by a project-based learning methodology. In particular, the methods necessary to carry out two projects will be presented both in the classroom and in the laboratory. One project in the field of image processing and the other in the field of audio processing, namely:

- Image: face recognition, construction of panoramic images, vehicle detection, etc.
- Audio: clustering systems or classification of audio by genres, classification of electrocardiograms, classification of emotions, etc.

Both projects will be presented on the Kaggle platform as "challenges", so that students can compete

with each other.

The course will be closed with an introductory lesson to neural networks and their applications in speech, audio, image and video processing, which will be continued in two optional courses of the second term:
- Deep learning for image analysis
- Natural Language Processing

Program of the course

1. Introduction to audio & visual analytics
2. Audiovisual data: digital representation
3. Digital image and video processing
3.1. Point-to-point operations and filters
3.2. Image segmentation and morphological processing
3.3. Image representation for classification and detection
3.4. Integrating project (e.g. face recognition, construction of panoramic images, vehicle detection, etc.)
4. Voice and audio processing
4.1. Speech production and audio perception
4.2. Localized time analysis.
4.3. How does Shazam work?
4.4. Speech and audio representation for classification, detection and recognition
4.4. Integrating project: (e.g. audio clustering or classification, electrocardiogram classification, emotion classification, etc.)
5. Introduction to Neural Networks for voice, audio, image and video analysis

LEARNING ACTIVITIES AND METHODOLOGY

Several types of learning activities are proposed: theoretical and practical lessons, lab assignments and final project.

Several methodologies will be adopted: theoretical lessons and problem-based learning (with different levels of supervision and guidance).

THEORETICAL LESSONS (2.5 ECTS)
Theoretical lessons provide an overview of the main theoretical and mathematical concepts together with explanations about the analytical tools employed for analysis of audio, imagen and video.

GUIDED LAB ASSIGNMENTS (1.75 ECTS)
Several guided lab assignments have been designed with the purpose of allowing the students to put into practice the mathematical tools explained in the theoretical lessons. The students will learn to use different audio and image analysis methods, such as audio clustering, face recognition and textual indexing, and learn to make sense of the results obtained.

FINAL PROJECTS (1.75 ECTS)
The students will develop two simple projects, one related to image analysis and the other related to audio analysis.

ASSESSMENT SYSTEM

% end-of-term-examination/test:	0
% of continuous assessment (assigments, laboratory, practicals…):	100

In line with the project-based learning approach the course, it will be evaluated exclusively on the basis of the two projects carried out:

50% Project 1: report, results and presentation.
50% Project 2: report, results and presentation.

BASIC BIBLIOGRAPHY

 - C. D. Manning, P. Raghavan and H. Schultze Introduction to Information Retrieval, MIT press, 2008

- 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, Fourth Edition, Pearson, 2018

ADDITIONAL BIBLIOGRAPHY

- Ricardo Baeza-Yates, Berthier Ribeiro-Neto Modern Information Retrieval: the concepts and technology behind search, 2nd Edition, Pearson, 2011

- S. Theodoridis and K. Koutroumbas Pattern Recognition, 4th ed., Academic Press, 2008

- Wilhelm Burger and Mark J. Burge Principles of Digital Image Processing: Fundamental Techniques, Springer-Verlag, 2009