Tratamiento de audio, video y visión artificial

Curso Académico: (2022 / 2023) **Fecha de revisión: 20/05/2022 12:20:24**

Departamento asignado a la asignatura: Departamento de Teoría de la Señal y Comunicaciones

Coordinador/a: GONZALEZ DIAZ, IVAN

Tipo: Obligatoria Créditos ECTS : 6.0

Curso : 4 Cuatrimestre : 1

REQUISITOS (ASIGNATURAS O MATERIAS CUYO CONOCIMIENTO SE PRESUPONE)

Redes Neuronales
Señales y Sistemas
Aprendizaje Automático I y II

OBJETIVOS

Los estudiantes deberán alcanzar los siguientes objetivos:

1) Conocer las señales digitales de imagen, audio, voz y vídeo, así como sus principales parámetros y el proceso de digitalización.
2) Conocer las técnicas más importantes del procesado de imagen, vídeo y audio, así como las tareas habituales de la visión y audio por ordenador .
3) Aplicar técnicas de aprendizaje automático y aprendizaje profundo vistas en asignaturas previas de la titulación al análisis de contenidos audiovisuales.
4) Desarrollar aplicaciones inteligentes que impliquen el análisis automático de contenidos audiovisuales.

DESCRIPCIÓN DE CONTENIDOS: PROGRAMA

La asignatura se divide en dos grandes bloques: por una parte, imagen y vídeo y, por otra, voz y audio.
En cualquiera de los casos, se presentan en primer lugar las señales y sus características, incluyendo ciertas nociones de los sistemas visual y auditivo. Seguidamente, se presentan las técnicas más habituales de tratamiento de señales específicas para cada caso, ilustrándose su uso en aplicaciones seleccionadas. Por último, se estudian las aproximaciones más modernas, basadas en soluciones de aprendizaje profundo (CNNs y RNNs), que hoy en día constituyen el estado del arte de la tecnología.

El programa de la asignatura se organiza como sigue:

Bloque 1: Tratamiento de señales visuales: imagen y vídeo
==========
Tema 1: Introducción a las imágenes y vídeo digital
Tema 2: Fundamentos del procesado de imagen y vídeo
Tema 3: Representación de imágenes: descriptores de bajo nivel
Tema 4: Segmentación de imágenes
Tema 5: Redes Neuronales Convolucionales (CNNs) para clasificación de imágenes
Tema 6: Otras aplicaciones de las CNNs en el análisis visual

Bloque 2: Tratamiento de la señal de voz y el audio
==========
Tema 7: Fundamentos del audio y la voz digital: generación, percepción y digitalización
Tema 8: Análisis localizado en el tiempo para señales de voz y audio
Tema 9: Descriptores de bajo nivel de voz y audio
Tema 10: Redes Neuronales Recurrentes (RNNs) para el análisis de datos secuenciales

ACTIVIDADES FORMATIVAS, METODOLOGÍA A UTILIZAR Y RÉGIMEN DE TUTORÍAS

Se proponen dos tipos de actividades formativas: clases de teoría, y prácticas de laboratorio.

CLASES DE TEORÍA

Las clases de teoría serán lecciones magistrales con uso de transparencias u otros medios audiovisuales para ilustrar determinados conceptos.

Mediante estas sesiones el alumno adquirirá los contenidos básicos de la asignatura. Es importante destacar que estas clases requerirán iniciativa y trabajo personal y en grupo por parte del alumno (habrá conceptos que deberán estudiar personalmente a partir de algunas indicaciones, casos particulares de tendrán que desarrollar, etc.)

PRÁCTICAS

Es una asignatura de elevado componente práctico, en lo que los alumnos asistirán regularmente a sesiones de laboratorio. En ellas, los conceptos adquiridos en las clases de teoría se pondrán en práctica utilizando el lenguaje de programación python, y bibliotecas software para el análisis de imagen y la visión artificial (scikit-image, PIL, OpenCV), el procesado de audio (scikit-sound), el aprendizaje automático (scikit-learn) y el aprendizaje profundo (pytorch). En los laboratorios se dispone de máquinas equipadas con GPUs de altas prestaciones y se utilizarán también sistemas de computación distribuidos y gratuitos como Google Colab.

SISTEMA DE EVALUACIÓN

Peso porcentual del Examen/Prueba Final:	40
Peso porcentual del resto de la evaluación:	60

SE1: EXAMEN FINAL. En el que se valorarán de forma global los conocimientos, destrezas y capacidades adquiridas a lo largo del curso.
SE2: EVALUACIÓN CONTINUA. En ella se valorarán los trabajos, presentaciones, actuación en debates, exposiciones en clase, ejercicios, prácticas y trabajo en los talleres a lo largo del curso.

BIBLIOGRAFÍA BÁSICA

- 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

BIBLIOGRAFÍA COMPLEMENTARIA

- 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