PhD Program in Sciences & Technologies for Electronics & Telecommunication

YEAR 2019

Curriculum in

Computer Vision, Pattern Recognition and Machine Learning

Visione Computazionale, Riconoscimento e Apprendimento Automatico

ATTENTION

The PhD application also implies to submit a research proposal under one or more themes chosen among those below indicated.
To write a proper research proposal, please follow the instructions indicated in the following file:
https://pavisdata.iit.it/data/phd/ResearchProjectTemplate.pdf

More info about the ongoing lab research activities: https://pavis.iit.it

Research themes

Theme A
Computer vision for the prediction of human intentions, activities and behavior
Tutor: Vittorio Murino

Study and development of techniques and systems for the analysis of behaviours, actions, and intentions. We are also interesting in studying expressions/emotions, personality traits, and social signals in general, referred to both single persons and groups, preferably in the wild. In particular, we would like to go further in the analysis of human movements and explicit behaviours to design methods and algorithms for the prediction of human intentions, also exploiting hints and findings coming from social sciences. Cases of interest include, but are not limited to, human-object and human-human interactions in which the overarching goal is not declared in advance nor actually observed. The task is to anticipate the actual intention

subsuming that interaction using multimodal and multisensory (2D video sequences + 3D data) as input to develop the algorithms. Computer vision and machine learning constitute the focus of this research, and particular attention will be devoted to deep learning models.

Theme B
Human body modeling for Socially-Aware Computer Vision
Tutors: Cigdem Beyan, Vittorio Murino

To recognize and interpret human nonverbal behavior it is fundamental to identify the subjects involved, especially in the wild, that is, in real situations. To this end, part-based human body modeling is a mandatory task aimed at extracting from images the different components of the human body, such as head, torso, arms, legs, etc., to estimate posture, gesture, gaze, and other expressive behaviors widely known as useful hints to detect social groups, recognize their social relations and high level characteristics such as personality traits for instance. Further, real-time tracking of body parts is equally important to increase such recognition performances, possibly adding prediction functionalities to these algorithms. Computer vision and machine learning constitute the focus of this research, and particular attention will be specifically devoted to deep learning models. Applications in the contest of human-human and human-robot interactions, surveillance and security as well as in the biomedical field (e.g., monitoring of elderly people) are envisioned.

Theme C
3D scene understanding with geometrical and deep learning reasoning
Tutor: Alessio Del Bue

Classical multi-view geometry problems make use of geometric reasoning to infer the scene 3d structure and its dynamic. These approaches often neglects the semantic composition of the scene that instead provides important cues about objects motion and their current spatial configuration. Instead, such context and semantic information can be given by current deep learning architectures but very few works attempted to merge geometrical reasoning with such semantic information. This research theme will have the aim to bridge this gap and to provide solutions that can be applied on intelligent system for robotics and autonomous driving. This research will study different methods and tools involving scene graphs, object detection in multi-view images, deep learning methods for scene representation and large-scale 3D reconstruction in dynamic environments. The research activities will build over and expand the current expertise on these problems developed in the Visual Geometry and Modelling (VGM) lab at IIT (https://vgm.iit.it/).

Theme D
Multi-modal scene understanding
Tutor: Vittorio Murino

This topic focuses on the research and implementation of algorithms that leverage big data, often multimodal, coming from 24h/7d sensor streams, in order to monitor a certain environment and understand the scenario. In particular, we aim at exploiting optical (RGB), 3D

(e.g. Depth) and acoustic data, also exploiting our in-house opto-acoustical sensor – DualCam – to detect, track and classify persons and objects, events and behaviors in general. Due to the inherent ability of deep learning architectures to cope with multi-modal data, we first target such class of methods, Other machine learning approaches and fusion methods will in any case be investigated. For instance, additional information contained in different modalities can be exploited in the framework of "Learning with privileged information" or within a "Teacher-Student" paradigm.

Theme E
Computer vision into the wild
Tutor: Vittorio Murino

This research theme focuses on the development of machine learning models for computer vision that can be deployed into the wild. More specifically, one drawback of modern learning systems is that they strongly rely on the characteristics of the data they are trained with. This results in models that poorly generalize to context unexplored during training (for example, consider a home robot that is deployed in a new house). To overcome this liability, two main strategies are domain adaptation and domain generalization. In the former case, we can leverage non-annotated samples from a desired scenario during training, and design models that better *adapt* to that domain. In the latter, the goal is generalizing to domains that are utterly unseen during training. The design of new training procedures to solve these tasks and the identification of novel application settings represent the main directions of this research.

Theme F
Zero- and Few-Shot Learning
Tutor: Vittorio Murino

This topic focuses on the deployment of techniques allowing a recognition system to discriminate classes and categories which differ from the ones considered for training. We encompass cases where a few annotated instances of the novel classes are available (*few-shot* learning) or, when the novel classes and categories remain totally unseen, auxiliary textual descriptions/attributes are exploited to compensate the absence of visual data (*zero-shot* learning). Machine learning and computer vision constitute the focus of this research. Multimodal approaches to combine images/videos with attributes/text embeddings are of crucial interest and utility for this research theme. In this respect, a particular attention will be devoted to deep learning frameworks, capable of solving zero- and few-shot learning through an end-to-end pipeline.

Theme G
Crowd behavioral analysis and event recognition
Tutor: Vittorio Murino

Study and development of techniques and systems for the analysis of behaviours, events, social signals in general, referred to a large mass of people (crowd). More in detail, the analysis and modelling of behaviour of groups and crowd seen as single entities will be considered. There is

indeed evidence that large groups of people and crowds are characterised by a collective behaviour that may emerge in different situations and can lead to interesting outcome from the point of view of the surveillance applications, helping to detect and predict coming events. Machine learning as well as computer vision constitute the foundation of this research, starting from early work in human body modelling/tracking to novel social force models able to grasp the complex dynamics of the human flow. Particular attention will be specifically devoted to modern deep learning models.

Theme H
Re-identification
Tutor: Vittorio Murino

Study and development of biometric techniques for scene analysis and understanding. The research will mainly focus on person characterization, with possible focus on the usage of soft biometrics cues (3D, attributes) and in challenging conditions (e.g., crowd). The idea is to recover the identity of persons viewed in different times and places, also considering face/body attributes, the so-called re-identification problem. Not only optical cameras will be used, but also other information gathered from different sensors (e.g., depth). Moreover, the use of a pan-tilt-zoom (PTZ) camera able to identify specific features of a single person or groups, or addressing non-cooperative face recognition at distance, could be subjects of investigation. The robustness against environmental (real) conditions and the non-cooperation of the subjects are the main features that the developed techniques will have to cope with, in order to deploy this technology in real scenarios. Computer vision and machine learning constitute the focus of this research, with particular focus on deep learning models.

Theme I
Biomedical imaging & Neuroimaging
Tutors: Diego Sona, Vittorio Murino

The wide adoption of biomedical sensors (e.g., MRI, TAC, SPECT, MEG, EEG, Fluorescence Microscopy, etc.) in various medical and biological investigations is fostering an increasing interest in advanced tools supporting the expert in the analysis and interpretation of the produced 2D/3D/4D images, both in clinical and scientific applications.
In this perspective, this theme will address the research related to the development of computer aided diagnosis (CAD) systems, with tasks ranging from image processing, image segmentation and object detection, up to automatic determination of disease biomarkers and more advanced data analysis, with applications to connectomics and radiomics. Particular attention will be devoted to structural data, functional data and the investigation of relationships between function and the underlying structure. To this aim multi-modal data analysis and fusion will play a strategic role. The development of such CAD tools will also require the design of novel computer vision, pattern recognition and machine learning techniques for biomedical data. In this line, particular attention will be devoted to deep learning models. Research will address a range of possible applications, from biomedical image analysis (e.g., image segmentation of particular body organs, and classification of their conditions) to cell imaging (e.g., cells detections and segmentation, relationship between cell morphology and function), with particular attention to problems related to brain imaging such as, e.g., characterization of mental

or neuro-degenerative diseases, investigation of cognitive functions, functional and structural connectomics, etc.

Theme L

Movement and behavior analysis in biomedical applications

Tutors: Diego Sona, Vittorio Murino

Movement and behaviour analysis in neuroscience or clinical applications is a fundamental research topic studying the biological bases or the pathological elements of movements and behavior in humans or animals. In a neuroscientific perspective, this analysis allows for providing insights into the mechanisms of the nervous system producing abnormal behavior, and experimental subjects mostly involve animals. In a clinical perspective, this analysis allows for monitoring the health status of patients with neurological or physical inabilities. We are involved in a multidisciplinary research activity, which need the development of techniques and systems for the automatic analysis of movements, postures, actions and behaviour of mice in home cages or humans in monitored environments. In this framework, we aim at designing methods for tracking 24/7 from video recorded from multiple sensors and the subsequent multimodal high-level analysis of behaviors. Computer vision and machine learning constitute the focus of this research, with particular interest in methodologies exploiting the spatio-temporal information using computer vision and machine (deep) learning approaches.