Ph.D. in Information Technology
Thesis Defenses

July 27th, 2023
at 11:00
Aula PT1, Ed.20 and online by Webex

Matteo FROSI – XXXV Cycle

ART SLAM: A FRAMEWORK FOR REAL-TIME MULTI-SENSOR 3D SLAM

Supervisor: Prof. **Matteo Matteucci**

Abstract:

Autonomous robot navigation represents a core aspect in the field of robotics, due to the many applications (e.g., agriculture and farming and automated driving). Simultaneous Localization and Mapping (SLAM) methods address the problem of constructing a model of the environment surrounding the robot, i.e., the map, while simultaneously estimating its pose within it. In literature, a plethora of SLAM systems have been proposed in the last decade. The created methods, however, greatly differ one from the other, in terms of architecture, data, and adopted frameworks. For this reason, extending or just using a SLAM method that suits specific criteria has become increasingly difficult over the past few years. For this reason, in the thesis, we present a common framework to perform SLAM, by developing multiple novel methods, distinguishable one from the other by the main sensor used and the specific real-world problem tackled. These systems can be summed up as follows: SLAM using a LiDAR as the main sensor, SLAM using both cameras and LiDAR, SLAM exploiting radar information and SLAM aided with information coming from third-party mapping services. All presented methods contribute in multiple ways. First, they improve existing algorithms, to mainly increase accuracy and performance. Then, some of the systems deal with situations that are not fully explored in literature, such as for SLAM where the main sensor is a radar, localization in GNSS-denied environments, and SLAM aided with prior maps. Lastly, the development of a unique framework allows users to adopt a SLAM system suited to their needs. The proposed methods are tested on state-of-the-art datasets, including KITTI and MulRan.

Hao QUAN – XXXV Cycle

SEMANTIC HUMAN ACTIVITY RECOGNITION FROM MOBILE ROBOTS

Supervisor: Prof. **Andrea Bonarini**

Abstract:

Social robots have been deployed in different real application environments (In The Wild (ITW)) to provide services. A social mobile robot that is able to select targets to approach based on activities performed by people in public spaces can better optimize customer experience by providing guidelines, discount promotions, and emergency assistance. To select targets, it is important to understand what candidates are doing to approach persons who may accept the robot's offer properly. Activity recognition ITW is a challenging task and would need machine learning algorithms to generate models from datasets representative of the situations to be faced. The state-of-the-art datasets and deep learning models for the task of Human Activity Recognition (HAR) from mobile robots do not satisfy the requirements to produce reliable HAR for social robots in service tasks that could operate in public spaces. In this thesis, we address the issue from different perspectives. First, we suggest some characteristics of a reliable ITW dataset should have to support the development of activity recognition from mobile robots and motivate how no state-of-the-art dataset fully satisfies these criteria. Given the lack of datasets to support skeleton-based activity recognition from social robots in service tasks, we collected a large-scale video dataset for activity recognition from the viewpoints of mobile robots ITW: POLIMI-ITW-S. POLIMI-ITW-S dataset is composed of 22,161 RGB video clips (approximately 46 hours) on which we have identified 37 classes of activities performed by more than 50,000 people in real shopping centers. Secondly, we discuss labeling issues and propose a hierarchically detailed labeling approach needed in many applications. We developed a new Human Activity Video Pose Tracking Annotation Tool (HAVPTAT) that could support the semi-automatic labeling of such clips. We used HAVPTAT to label the new POLIMI-ITW-S dataset thoroughly. Based on the common practices, we exploited a novel data pre-processing framework to elaborate the collected raw video dataset to obtain the skeletal data ready-to-use for training and evaluating the popular skeleton-based HAR deep learning algorithms. Then, we evaluated some advanced skeleton-based HAR models on POLIMI-ITW- S dataset. The obtained results are much worse than the results obtained on the datasets collected from constrained environments. We deeply investigate the recognition results and summarize several potential reasons that prevent obtaining good results. Furthermore, we propose a new Shift Semantic Graph Convolutional Neural Network (SS-GCN) method, enhancing Human Activity Recognition (HAR) by understanding human-object interactions semantically. We give some suggestions to improve the performance. Finally, we propose a novel semantic hierarchical model that improves the performance of state-of-the-art skeleton-based activity recognition models on this dataset by about 30%. We depict several examples to visualize the decisions made by robots showing whether to approach people based on the recognition results from the proposed model. We released the complete resources (including the annotated data composed by people tracked bounding boxes, 2-D human body skeleton, and activity classes), HAVPTAT, and the proposed semantic hierarchical model for research use.

PhD Committee

Prof. **Luca Magri,** Politecnico di Milano

Prof**. Domenico Giorgio Sorrenti,** Università degli Studi di Milano-Bicocca

Prof. **Andrew David Bagdanov**, Università degli Studi di Firenze