Ph.D. in Information Technology: Final Dissertations

DEIB Seminar Room "Alessandra Alario"

February 22nd, 2017

10.30 am

First Ph.D. presentation and discussion:

Alessio Mauro FRANCHI – XXIX Cycle

"Embodiment in robotics: the sensorimotor loop toward learning and cognition"

Advisor: Prof. **Giuseppina Gini**

Abstract:

Robots nowadays are not widespread among human beings but are still confined into plants and research laboratories; this is mainly because they lack autonomous behaviors and they are not able to learn and adapt to the real environment; shortly, they still miss what is called cognitive development. A new perspective in robotics and AI states that cognitive processes are strongly tied to the physical structure of the body: the body shapes the mind and vice versa. This view is called embodiment and tries to highlight the importance of the interplay between brain, body and world for learning in cognitive systems; it affects different aspects of these processes, firstly motor and language learning.

This research explores cognition in artificial agents from the point of view of embodied robotics. It wants to propose a new starting point for building a robotic architecture able to show simple child-like learning capabilities based on the sensorimotor circuit.

It proposes a model for motor learning based on the motor primitives paradigm; it transposes this concepts into language acquisition with the introduction of the linguistic primitives; it defines a model for the working memory, able to perform context-sensitive selection and retention of both information and actions. An intentional architecture for the autonomous generation of new motivation integrates and connect the aforementioned models.

The proposed model have been implemented in Matlab and thoroughly tested; even though some important aspects of cognition are still to be integrated, first results are encouraging and positively compare with scientific evidences about children.

Second Ph.D. presentation and discussion:

Matteo LUPERTO – XXIX Cycle

“Semantic Modeling of the Global Structure of Buildings”

Advisor: Prof. **Francesco Amigoni**

Abstract:

Autonomous mobile robots can perform many different tasks to help humans during their activities or to replace them in hazardous environments and in simple routine operations. When we consider indoor tasks, robots have to interact with environments that are specifically designed for human activities and for interaction between humans, buildings. Buildings are strongly structured environments that are organized in regular patterns. For instance, rooms typically have a geometrical structure that is characterized by features, such as walls perpendicular to the floor and to the ceiling, and by a layout that can be, in most cases, approximated by a box-like model. In order to increase their ability to autonomously operate in indoor environments, robots must have a good understanding of buildings, similarly to that human beings exploit during their everyday activities. If we consider how people and robots interact with indoor environments, it can be said that people naturally understand and "read" buildings as human-made environments (and act in them accordingly), and that this is hardly the case for autonomous mobile robots. One of the most important tools that researchers have developed to address a robot's needs for interacting with an indoor environment are semantic maps. Semantic maps are abstract representations which aim to represent the meaning of parts of the perceived environment in order to provide robots a human-like understanding of their surroundings. Semantic maps can be used for describing heterogeneous concepts that can be useful for robots, such as objects and rooms. In this dissertation, we focus on a particular type of semantic maps, which identifies rooms, represents I how rooms are connected, and assigns to each room a semantic label indicating its function, such as 'corridor', 'classroom', 'office', or 'bathroom'. Semantic maps are usually build on metric maps, that represent the space occupation and are particularly useful for tasks such as path planning and localization. Typically, the interaction between a robot and its environment is heavily based on data acquired with perception. Mapping methods usually provide reliable knowledge only on parts of the environment that have been already visited. This approach often implies that what has not been seen by the robot does not exists, adopting, in a sense, a closed world assumption on the environment. This statement is true considering both semantic and metric maps. This form of interaction with the environment is radically different from that of humans, who can easily navigate and comprehend the structure of buildings even without having seen them before. The contribution of this thesis moves from the consideration that the global structure of buildings, which is often neglected when building semantic maps, could be exploited to increase the autonomous abilities of robots when operating in indoor environments. Our proposed framework aims at identifying and at overcoming the limitations in standard semantic mapping methods by

starting from two insights on indoor environments. In first place, we consider an entire floor of a building as a single object, by identifying relations between different (and potentially unconnected) parts of the building. This can be done both considering the metric map of the environment, for example by identifying that rooms in different parts of the building share one or more walls, and by considering the topology of the environment, namely how rooms are connected with each other, and for example observing that parts of the building with a similar function have a similar structure. Moreover, we consider each building in relation with other buildings with the same function. The function of a building is represented by the main activity that each building is designed for and is captured by the concept of building type. Examples of building types are schools, offices, hospitals, university, shopping malls, and others. The function of a building imposes its structure, its floor plan, and the structure of its rooms. Each building, having a precise function, shares some structural features with all other buildings with the same purpose. Exploiting these two observations, we provide an analytical model of the structure of a building that considers altogether all the relations between its single parts and that considers the features shared by the set of buildings belonging to the same type. We start from reconstructing the layout of an II indoor environment from its metric map. The layout can then be used for obtaining a graph representation the building. In our approach, data from the metric maps are used in combination with data representing floor plans of buildings belonging to the same type. Using graph kernels and Monte Carlo Markov Chains, we provide a method able to generate new instances of building structures from a set of examples. We outline some possible applications of our approach, involving reasoning on unknown parts of buildings and labelling entire floors of buildings accordingly to their function.

Third Ph.D. presentation and discussion:

Andrea ROMANONI – XXVIII Cycle

“Incremental Large-Scale Visual 3D Mesh Reconstruction”

Advisor: Prof. **Matteo Matteucci**

Abstract:

In the last decade, the growing interest about autonomous driving brought many computer vision and robotics researchers to focus on the vehicles understanding of the surrounding environment through a map of it. A map is needed to plan a path to reach a specific destination, or to estimate the current position by comparing the current perception of the environment against a reference. In robotics, a suitable mapping, or reconstruction, algorithm needs to be scalable, incremental and to provide a dense map. Scalability is needed especially in large-scale environments; an incremental algorithm allows map update as new data are

acquired; density enables a consistent and coherent navigability. Researchers, in computer vision, focused their reconstruction algorithms on accurate and dense results, disregarding any incremental processing, and only few works show large-scale capabilities. Instead, in robotics the focus is mainly on incremental algorithms but the output maps are usually point clouds; only a very limited amount of works estimate dense and continuous surfaces, but they are limited to small scale scenes. As the main contribution of this thesis, we propose a novel incremental, automatic and scalable reconstruction pipeline to estimate continuous dense manifold meshes; we especially focused on keeping the manifold property valid, to enable a coherent mesh refinement based on image appearance. Our contribution first improves the accuracy of the state-of-the-art incremental reconstruction algorithm both in case of video sequences of urban landscape, and in case of unordered set of images. Then, to embed and refine automatically and incrementally new part of the scene in a reference model, we proposed a novel mesh merging algorithm that preserves the manifold property. Finally, we extended our work to jointly deal with laser range finders and images, exploiting the accuracy of the laser range measurement and the appearance provided by the images. We tested our proposals against publicly available KITTI, Middlebury and EPFL datasets, which proi vide different scenarios in order to stress the flexibility of our approach.

Fourth Ph.D. presentation and discussion:

Francesco VISIN – XXVIII Cycle

"Deep Recurrent Neural Networks for Visual Scene Understanding"

Advisor: Prof. **Matteo Matteucci**

Abstract:

Machine Learning (ML) is a fascinating field of research. In the era of knowledge, being able to find the right information in enormous amounts of data (e.g., the internet) and summarize it in a form that is compact and yet retains all the content one is interested in, is a key factor of success or failure in many fields. I am particularly interested in applying ML to vision problems because we, as humans, rely heavily on vision for our daily operations. Improvements in the technology at our disposal to interpret visual data can have a direct and remarkably rapid impact on many practical applications such as assist or automate driving, analyze medical images, aid surgeons in the operating room or improve the quality of life for visually impaired people. This manuscript presents my work on Recurrent Neural Networks (RNNs) and RNN-based models applied to visual data, describing three models I proposed, namely ReNet, ReSeg and DEConvLSTM. The first is an RNN-based alternative to Convolutional Neural Networks (CNNs) for object classification. The carefully designed interaction between the RNNs in the architecture allows the model to capture the full context of the image

in just two levels of hierarchy as opposed to the many layers typically required by CNN-based models. The evolution of this model for semantic segmentation, called ReSeg, takes advantage of a similar inner structure as ReNet, further improved by the adoption of pretrained CNNs as well as the addition of transposed convolutional layers. Finally, the DEConvLSTM architecture addresses the much harder task of semantic segmentation in videos. To address this task I proposed a model that merges direct convolutions, transposed convolutions and RNNs in a unique coherent structure. The DEConvLSTM model exploits the speed of CNNs to process spatial information and the ability of RNNs to retain information through several steps of computation, and proved to be a valid architecture for video semantic segmentation. For each model, the architecture is first presented in detail, followed by a description of the experimental settings and of the datasets used for its evaluation. Results on publicly available dataset are compared to the state-of-the-art and discussed thoroughly.

PhD Committee:

Prof. **Matteo Matteucci**, DEIB – Politecnico di Milano

Prof. **Stefano Caselli**, Universita' di Parma

Prof. **Luca Iocchi**, Universita' di Roma "La Sapienza