Politecnico di Torino

Finanziato dall'Unione europea NextGenerationEU | Ministero dell'Università e della Ricerca | Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

ARTIFICIAL INTELLIGENCE

DM630/Makr Shakr - Tackling the challenges for fully autonomous manipulators in unstructured daily-living environments

Funded By	MAKR SHAKR S.R.L. [P.iva/CF:11162790015] MINISTERO DELL'UNIVERSITA' E DELLA RICERCA [P.iva/CF:97429780584] Politecnico di TORINO [P.iva/CF:00518460019]
Supervisor	DI CARLO STEFANO - stefano.dicarlo@polito.it
Contact	AVERTA GIUSEPPE BRUNO - giuseppe.averta@polito.it SAVINO ALESSANDRO - alessandro.savino@polito.it DI CARLO STEFANO - stefano.dicarlo@polito.it
Context of the research activity	This work aims to develop theories, methods, and algorithms for closed-loop control of manipulators working in complex and unpredictable environments. Indeed, while in many industrial tasks robots are tasked to use pre-defined tools to act in structured and known environments, deploying robots for food and beverage preparation, such as domestic helpers or bartenders, is far from being a solved task, with significant challenges from a perception (i.e. looking and understanding the scene around) and an action perspective (learning to execute complex and articulated (sequences of) actions). Progetto finanziato dal PNRR a valere sul DM 630/2024 - CUP: E14D24002330004
	Deploying autonomous intelligent manipulators in complex and unstructured environments is still a challenging task, with many open problems under both perception and action perspectives. Indeed, when the environment is not structured and well-known in advance, it is practically unfeasible to plan the set of actions the robot should take to complete a task. Such actions will indeed likely depend on the full state of the environment (where are placed objects of interest and obstacles), on its dynamics (i.e., when objects, humans, or other robots move around), and on the complexity of the task itself, for which actions taken may depend on what the robot did in the past. The scope of this research is to develop theory, methods, and algorithms to enable full autonomy of manipulators in challenging and unstructured scenarios, for example in food and beverage preparation. The student will deep dive into the challenge of developing computer vision methods for the

Objectives	understanding of unstructured environments, through the synthesis of dynamic and task-oriented scene graphs. Such representation strategy will be able to provide a compact yet accurate description of the surrounding scene, providing embedded knowledge about the position of objects, their physical characteristics (i.e. shapes, material, inertial properties), their functionality (i.e. objects and tools affordances) and mutual relationship (tools to act on other objects). Scene graphs will be used to feed robot learning methods that will be able to close the loop between perception and action. Human examples of task execution, such as meal preparation available in popular egocentric video datasets (like Ego4D [1]), visual instructions (GenHowTo [2]), and similar (eventually collected ad-hoc) visual-based 2D and 3D data to extract ground truth dynamic scene graphs for tasks execution, with specific attention to food and beverage preparation. Videos of such examples will be used as a source of information to imitate activities through autonomous manipulators, and ultimately close the loop between perception and action also through scene graph representations. Particular attention will be given to the robustness of the developed system under domain shifts, e.g. when the appearance of the environment changes (a different kitchen, novel brands of products, etc) and to the possibility of continuously adapting robot behavior when novel knowledge is available. The integrated perception-action system will be tested in collaboration with the industrial partner, which will provide support with data collection and testbed for end-user use cases. The candidate is expected to publish the outcome of this research in premium computer vision and robot learning conferences (CVPR, ECCV, ICCV, CoRL, IROS, ICRA) and journals (IEEE TrPami, IJCV, IEEE T-RO, IEEE RA-L) [1] Grauman, K., Westbury, A., Byrne, E., Chavis, Z., Furnari, A., Girdhar, R., Hamburger, J., Jiang, H., Liu, M., Liu, X. and Martin, M., 2022. Ego4d: Around the world in 3,000 hours of egocentric video. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 18995-19012). [2] Soucek, T., Damen, D., Wray, M., Laptev, I. and Sivic, J., 2024. Genhowto: Learning to generate actions and state transformations from instructional videos. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 6561-6571).

Skills and competencies for the development of the activity	Outstanding passion and motivation for research. Excellent programming skills (python and Pytorch) are required. Interest in deep learning for videos (egocentric or third-person video) is required. Experience with robot learning is not required, although preferred.