FONDAZIONE BRUNO KESSLER
MAGAZINE

MaGe, the robotic arm that combines vision and language

October 9, 2025

The TeV-Technologies of Vision project to study the integration between natural language, vision and robotics

A robotic system capable of performing manipulation operations based on instructions expressed in natural language: this is the goal of MaGe **– Make Grasping Easy**, a research project conducted by the Technologies of Vision (TeV) Unit at Fondazione Bruno Kessler.

The project focuses on the interaction between **three-dimensional machine vision, precision robotics** and **large language models,** with the aim of improving the autonomy and flexibility of grasping systems in unstructured environments.

Starting from a **textual description and a visual representation of the scene**, the system is able to identify relevant objects, estimate their spatial position and plan the movements necessary to interact with them safely. The **robotic arm**, guided by this data, can thus perform manipulation actions based on commands expressed in natural language—both written and spoken—autonomously determining the most suitable *grasping* strategy.

"Our goal is to study how language can become an effective means of controlling physical systems, even in complex contexts," explained **Fabio Poiesi** with TeV. *"The integration between semantic understanding and spatial perception represents an open challenge in robotics."*

Among the difficulties faced, there is the limited ability of current language models to manage the spatial dimension. For this reason, the team is working on the development of **new datasets and training strategies**, with the aim of improving the understanding of the geometric relationships between objects in a three-dimensional space.

The visual component of the system is based on a 3D camera capable of generating a point cloud, from which depth and color information are derived. To ensure precision in manipulation, a calibration process has been implemented between the reference frames of the camera and the robotic arm, also using visual markers.

The system's output is a **target pose**: a combination of optimal position and orientation for grasping. The robotic arm autonomously calculates the trajectories necessary to reach the target, avoiding collisions with the environment or with its own structure, thanks to the use of reverse kinematics algorithms.

The project uses technical equipment that includes a high-precision industrial camera, capable of detecting details with an accuracy of up to 0.2 mm, and a robotic arm with a load capacity of up to 5 kg and a maximum extension of 85 cm.

Thanks to the support of the Fondazione Caritro's VRT program, which made the acquisition of the robotic arm possible, a **multidisciplinary research team was formed consisting of** Runyu **Jiao, Alice Fasoli, Francesco Giuliari, Matteo Bortolon, Sergio Povoli, Yiming Wang** and FabioPoiesi, head of the Technologies of Vision (TeV) unit. This represented a fundamental step for the evolution of the MaGe project, allowing in-depth experiments that led to the drafting of a scientific paper to be presented at the **IROS international conference** (IEEE/RSJ International Conference on Intelligent Robots and Systems) this October in Hangzhou, China. A milestone that testifies to the quality of the work carried out and the recognition obtained internationally.

The project is a concrete example of **applied research at the intersection of language, vision and robotics**, an area still little explored in Italy. The activity of the TeV unit is at the forefront of the development of robotic cognitive systems, with the aim of contributing to the definition of new solutions for the interaction between humans and machines, in both industrial and

experimental contexts.

The quality of the research developed by the TeV unit is also confirmed by the results obtained at the international level. In November 2024, it took **first place at the BOP Benchmark for 6D Object Pose Estimation Challenge 2024**, outperforming over 50 global competitors, including teams from giants like NVIDIA, Meta, and Naver Labs. The winning method, *FreeZe v2*, was able to handle complex scenarios without the need for a training phase, demonstrating the effectiveness of the FBK approach in combining vision, language and robotics.

A result that confirms the ability to compete and succeed even starting from non-metropolitan contexts, thanks to a combination of talent, strategic vision and targeted investments. The challenge now is to turn these results into concrete applications, in advanced manufacturing, space robotics, predictive maintenance and many other fields.

PERMALINK

https://magazine.fbk.eu/en/news/mage-the-robotic-arm-that-combines-vision-and-language/

TAGS

- #3d
- #augmentedintelligence
- #dataset
- #digitalindustry
- #dvl
- #grasping
- #llm
- #MaGe
- #robotic arm
- #robotica
- #robotics
- #tev

AUTHORS

- Michela Antino