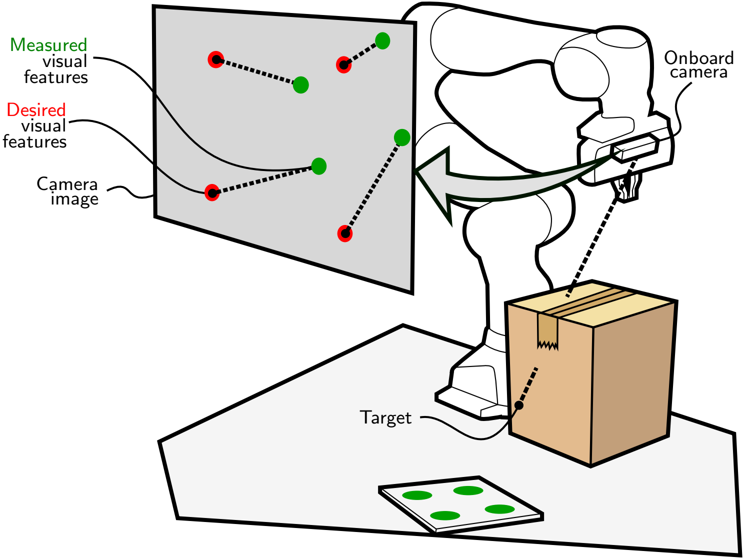
Imitation Learning-based Visual Servoing for Tracking Moving Objects

Rocco Felici[1], Matteo Saveriano[2], Loris Roveda[3], and Antonio Paolillo[3]

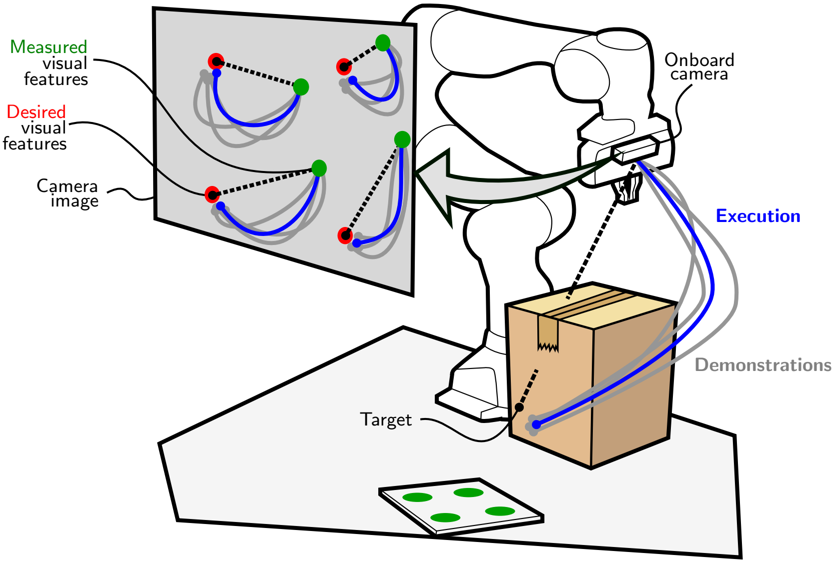
[1]USI, Lugano, Switzerland [2]DII, University of Trento, Italy [3]IDSIA, USI-SUPSI, Lugano, Switzerland

Munich, Germany
20-21 September 2023

Visual Servoing (VS)



Imitation Learning for Visual Servoing (ILVS)



ILVS formalism

- *Standard VS* task zeroes the error $\boldsymbol{e}$ between measured and desired features

$$\boldsymbol{v} = -\lambda \boldsymbol{L}^{+} \boldsymbol{e}$$

where $\boldsymbol{v}$ is the camera velocity, λ the control gain and $\boldsymbol{L}$ the interaction matrix

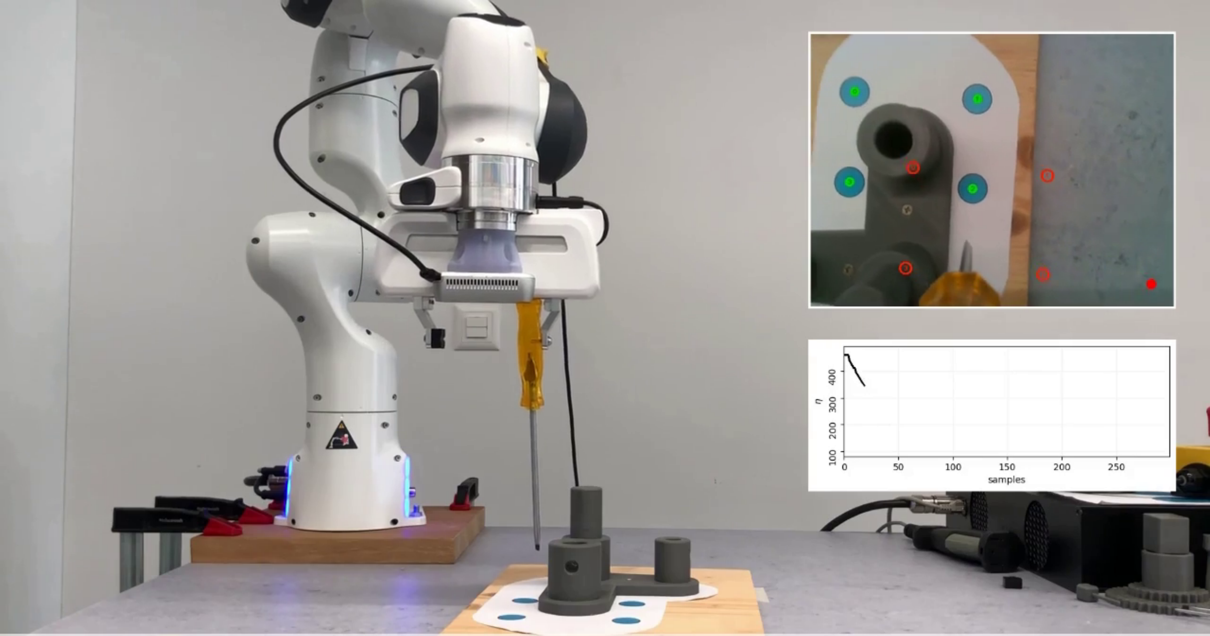
- An *additional task* can be added using the *dynamical system* formalism [1]

$$\boldsymbol{v} = -\lambda \boldsymbol{L}^{+} \boldsymbol{e} + h\boldsymbol{\rho}$$

where $\boldsymbol{\rho}$ is camera velocity augmenting the skill of the standard law

- We have used this control structure to *imitate complex demonstrations* and *keep the stability* of the original controller [2] [3]

[1] M. Saveriano and D. Lee, "Incremental skill learning of stable dynamical systems, " IROS 18.
[2] A. Paolillo, M. Saveriano, "Learning Stable Dynamical Systems for Visual Servoing," ICRA 22.
[3] A. Paolillo, P. R. Giordano, M. Saveriano, "Dynamical System-based Imitation Learning for Visual Servoing using the Large Projection Formulation," ICRA 23.



the classic approach leads to a collision

ILVS for Visual Tracking

- *Standard VS tracking* has this shape

$$\boldsymbol{v} = -\lambda \boldsymbol{L}^{+} \boldsymbol{e} - \lambda \boldsymbol{L}^{+} \frac{\partial \boldsymbol{e}}{\partial t}$$

where $\partial \boldsymbol{e} / \partial t$ is an estimate of the error variation due to the motion of the object

- We propose to *imitate the feedforward term*, as inspired by the dynamical system:

$$\boldsymbol{v} = -\lambda \boldsymbol{L}^{+} \boldsymbol{e} + \boldsymbol{\rho}(\boldsymbol{e}), \quad \boldsymbol{\rho} = -\lambda \boldsymbol{L}^{+} \frac{\partial \boldsymbol{e}}{\partial t}$$

where $\boldsymbol{\rho}$ is *learnt from demonstrations*

Learning the compensation term

- We consider data from D demonstrations of VS tasks composed of N samples:

$$\mathcal{D} = \left\{ \boldsymbol{e}_n^d,\ \boldsymbol{v}_n^d \right\}_{n=1,d=1}^{N,D} \quad \Longrightarrow \quad \mathcal{T} = \left\{ \boldsymbol{e}_n^d,\ \boldsymbol{\rho}_n^d \right\}_{n=1,d=1}^{N,D}$$

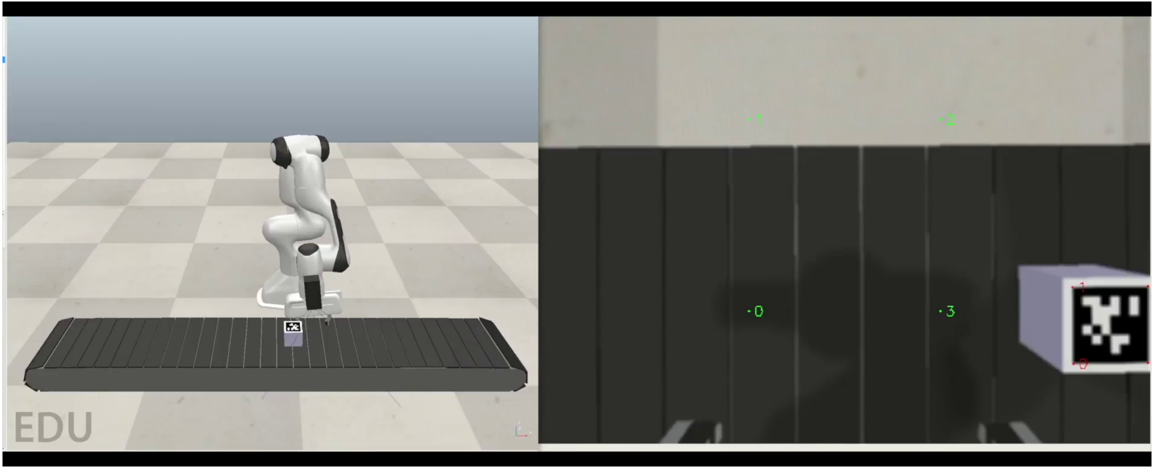
where $\boldsymbol{\rho}_n^d = \boldsymbol{v}_n^d + \lambda \boldsymbol{L}^+ \boldsymbol{e}$ (in the demonstration we assume to know the control structure)

- The compensation term can be inferred from $\mathcal{T}$ using the current value of the visual error

$$\boldsymbol{\rho} = \boldsymbol{r}\left(\boldsymbol{e} \,|\, \mathcal{T}\right)$$

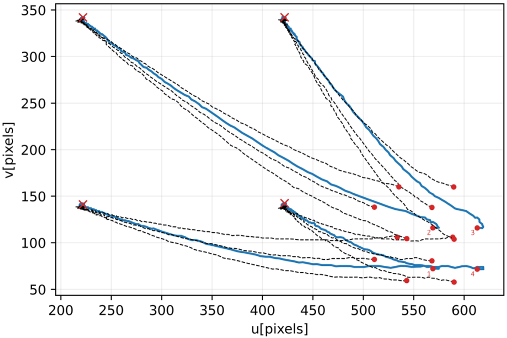
- We train GMR on $\mathcal{T}$ and use GMR to infer $\boldsymbol{\rho}$ from data

Results

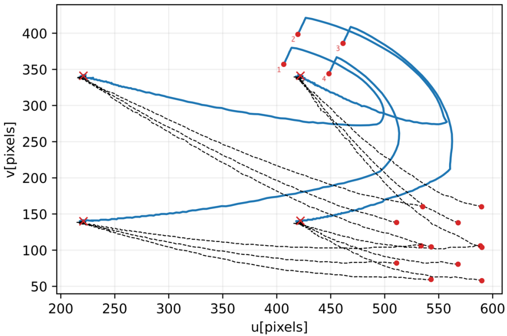


Results

- Visual features trajectories as executed by ILVS vs the demonstration.



with similar initial conditions of the demos



with unseen initial conditions

Conclusions

- We have exploit the *ILVS* paradigm to realize easy visual object tracking
- The approach is based on the *DS* rationale
- The approach realizes trackers *with no specific implementation* of estimators or observers
- Future work will
  - Further experimentation to test the *generalization* of the method
  - Formally investigate the *stability* of the controller
  - Real test within *industrial domain*

Thank you for your attention!