Università di Genova | DIBRIS DIPARTIMENTO DI INFORMATICA, BIOINGEGNERIA, ROBOTICA E INGEGNERIA DEI SISTEMI

iit RBCS ISTITUTO ITALIANO DI TECNOLOGIA | iit CONTACT ISTITUTO ITALIANO DI TECNOLOGIA | WHISPER | MaLGa

Learning dictionaries of kinematic primitives for action classification

Alessia Vignolo[1], Nicoletta Noceti[2], Alessandra Sciutti[1], Francesca Odone[2], Giulio Sandini[3]

[1]CONTACT Unit - Isitituo Italiano di Tecnologia, Genova

[2]MaLGa – Machine Learning Genoa center, DIBRIS, Università di Genova

[3]RBCS Unit - Isitituo Italiano di Tecnologia, Genova

Motivations and objectives

- Among the earliest processing stages of human perception development is the ability to **precisely localize in space and time an action and its sub-parts**
- Our work focuses on this ability and explores the concept of visual motion primitives, i.e. a limited number of action sub-components used to reconstruct a wide range of different complex actions
- The goal of this research is to assess whether a simple representation based on kinematic motion primitives may be the backbone of a general approach to action understanding
- It has been shown that humans tend to segment hand/arm actions in points where the kinematic is subject to a change, i.e., change in direction, velocity and acceleration of the wrist.

Using sparse coding with a data-driven dictionary

- We approach dictionary learning as an unsupervised problem using K-Means

$$\min_{\mathbf{D},\mathbf{U}} \|\mathbf{X}-\mathbf{D}\mathbf{U}\|_F^2 \ . \ \text{s.t.} \quad \text{Card}(\mathbf{u}_i)=1, |\mathbf{u}_i|=1,$$
$$\mathbf{u}_i \geq 0, \forall i=1,\ldots,T$$

where $\mathbf{X}$ is the training set, $\mathbf{U}$ are the clusters membership codes, and $\mathbf{D}$ is the dictionary with K atoms

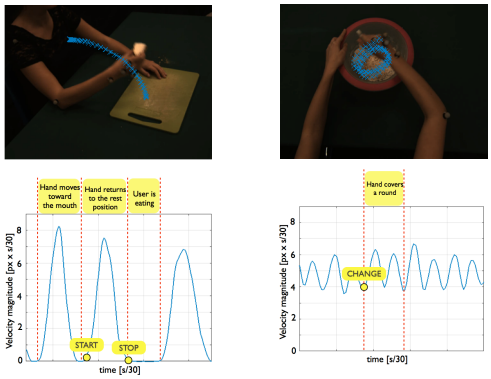
- We use Sparse Coding [1] to derive a sparse representation using the dictionary

$$\mathbf{u}^* = \arg\min_{\mathbf{u}} \|\mathbf{x}-\mathbf{D}\mathbf{u}\|^2 + \lambda\|\mathbf{u}\|_1$$

Our approach

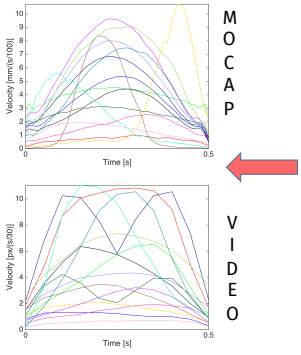
Segmentation in sub-movements

We represent a video as a sequence of velocities derived from **optical flow magnitudes** collapsed in a single point [2] and we segment the sequence detecting **dynamic instants** [3]



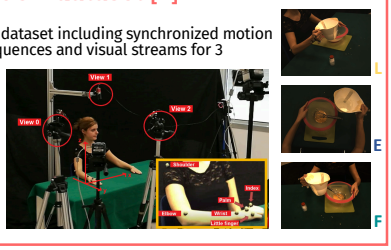
Action representation and classification using the dictionary

A dictionary of visual motion primitives is derived from the segmented sub-movements. They are represented using sparse coding, and the obtained codes are classified using a simple *Regularized Least Squares*, to focus on the **descriptive power of the representations**



The MoCA dataset [4]

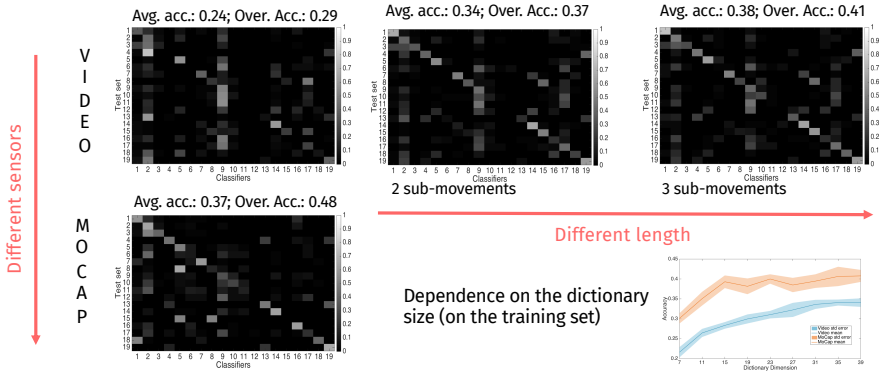
A bi-modal dataset including synchronized motion capture sequences and visual streams for 3 viewpoints



1-*Grating the carrot*, 2-*Cutting the bread*, 3- *Cleaning a dish*, 4-*Eating*, 5-*Beating eggs*, 6-*Squeezing the lemon*, 7-*Cutting with a mezzaluna*, 8-*Mixing*, 9-*Open the bottle*, 10-*Turning the omelette*, 11-*Pestling*, 12-*Pouring water*, 13-*Reaching an object*, 14-*Rolling the dough*, 15-*Washing the salad*, 16-*Salting*, 17-*Spreading cheese on a bread*, 18- *Cleaning the table*, 19-*Transporting an object*

Experimental analysis

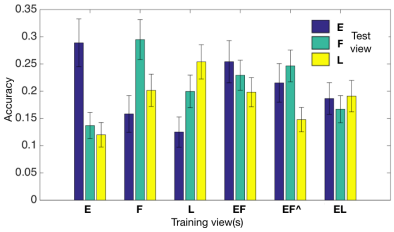
Confusion matrices on the test set



Different length

Dependence on the dictionary size (on the training set)

Cross-view recognition



We explore the generalization to new viewpoints, including a scenario (EF^) reminiscent of a situation in which a child learns by observing movements performed by himself and by a third person next to him.

REFERENCES

[1] J.Yang,K.Yu,Y.Gong,andT.Huang.Linearspatialpyramidmatching using sparse coding for image classification. In *CVPR*, 2009

[2] A. Vignolo, N. Noceti, F. Rea, A. Sciutti, F. Odone, and G. Sandini. Detecting biological motion for human-robot interaction: A link between perception and action. *Frontiers in Robotics and AI*, 2017

[3] F.Rea,A.Vignolo,A.Sciutti,andN.Noceti.Humanmotionunderstand- ing for selecting action timing in collaborative human-robot interaction. *Frontiers in Robotics and AI*, 2019

[4] E. Nicora, G. Goyal, N. Noceti, A. Vignolo, A. Sciutti, F. Odone, "The MoCA dataset, kinematic and multi-view visual streams of fine-grained cooking actions. *Scientific Data*, *to appear*

CONTACTS

- *Alessia Vignolo*, alessia.vignolo@iit.it
- *Nicoletta Noceti*, nicoletta.noceti@unige.it
- *Alessandra Sciutti*, alessandra.sciutti@iit.it
- *Francesca Odone*, francesca.odone@unige.it
- *Giulio sandini*, giulio.sandini@iit.it