

# The MoCA dataset project

---

## Description and motivations

The Cooking Actions Dataset is a multimodal dataset in which we collect MoCap data and video sequences acquired from multiple views of upper body actions in a cooking scenario. It has been collected with the specific purpose of investigating view-invariant action properties in both biological and artificial systems, and in this sense it may be of interest for multiple research communities in the cognitive and computational domains. Beside addressing classical action recognition tasks, the dataset enables research on different nuances of action understanding, from the segmentation of action primitives robust across different sensors and viewpoints, to the detection of actions categories depending on their dynamic evolution or the goal.

The dataset includes 20 cooking actions involving one or two arms of a volunteer, some of them including tools which may require different forces. Three different view-points have been considered for the acquisitions, i.e. lateral, egocentric, and frontal. For each action a training and a test sequence is available, each containing, on average, 25 repetitions of the action. Furthermore, acquisitions of more structured activities (we called scenes) are included, in which the actions are performed in sequence for a final, more complex goal. Specifically, scenes are composed as following:

An annotation is available, which includes the segmentation of single action instances in terms of time instants in the MOCAP reference frame. A function then allows to map the time instants on the corresponding frame in the video sequences. In addition, functionalities to load, segment, and visualize the data are also provided, as described in the following.

## Technical information

MATLAB structures containing the MoCap streams are composed by the following fields:

- Shoulder, elbow, wrist, palm, index finger and little finger complete streams, without any filtering;
- index, array containing segmentation indices, this is the information used in `segmentAction` to separate the streams;
- labels (present only in the scenes structures), array containing the labels of the actions that succeed one another in the scene (in temporal order). True labels include also `'pause'`, a moment in the sequence in which the actor does not move;

The point of view of the video recordings is specified by the number at the end of each filename:

- "\_0.avi" lateral PoV;
- "\_1.avi" egocentric PoV;
- "\_2.avi" frontal PoV;

### **List of the actions included**

1. Shredding a carrot
2. Cutting the bread
3. Cleaning a dish
4. Eating
5. Beating eggs
6. Squeezing a lemon
7. Mincing with a mezzaluna
8. Mixing in a bowl
9. Open a bottle
10. Turn the frittata in a pan
11. Pestling
12. Pouring water in multiple containers
13. Pouring water in a mug
14. Reaching an object
15. Rolling the dough
16. Washing the salad
17. Salting
18. Spreading cheese on a slice of bread
19. Cleaning the table
20. Transporting an object

### **Scenes description**

- Scene #1 The actor mixes ingredients in a bowl, then adds salt and pours some water. Finally the ingredients are mixed again.
- Scene #2 The actor reaches a slice of cheese, grabs it and shreds it. Then the actor moves the cheese back to the original position.
- Scene #3 The actor reaches a bottle, moves it and removes the cap. Some water is poured in a bowl, then the bottle is put in the previous position. The actor then mixes the ingredients in the bowl.
- Scene #4 The actor cuts a slice of bread and spreads some nuts cream on it, then the actor eats it.
- Scene #5 The actor reaches a lemon and squeezes it. Then all the objects are moved away and the actor cleans the table.

### **Available functions**

Loading and visualisation functions (in MATLAB) allow users to access RGB and Kinematic streams.

Functions "loadDataset" and "loadAction" allow the user to load and save the Cooking Actions Dataset in an easy-to-use data structure. "loadAction" gives the user the possibility of loading only part of the Cooking Actions Dataset, as for instance an action, a marker, or an instance.

Syntax:

- loadDataset(folder, 'training') and loadDataset(folder, 'test') load MoCap data of all actions for the specified set.
- loadAction(folder, action) and loadAction(folder, action, 'ALL') return a struct containing data of all markers for the specified action. Notice that "action" is a string that can be derived from the names of the data files. Example: carrot\_tr.mat -> carrot
- loadAction(folder, action, marker) returns a struct containing data related to the specified action and limited to the specified marker
- loadAction(folder, action, 'ALL', instance) returns a struct containing data of all markers at the specified instance of the action.
- loadAction(folder, action, marker, instance) returns a struct containing data of a single marker at the specified instance of the action.

The function "segmentAction" extracts the instances of actions from the MoCap streams. Same function can be used to segment the scenes in different actions.

Three types of visualisation functions are available:

- "visualiseAction", for 3D plot of each marker's trajectory
- "visualiseSkeleton", for a simulation of the arm executing the complete action using MoCap data
- "initSynch" and "synchronizedView" for a joint view of RGB and Kinematic data.

All the functions provided can be used also on the test scenes in /data/mocap/scenes.

Type help "name of the function" in the Command Window for more information on syntax and on how to use the functions.