Modeling visual features to recognize biological motion: a developmental approach

Giulio Sandini¹, Nicoletta Noceti², Alessia Vignolo¹,², Alessandra Scuitti¹, Francesco Rea¹, Alessandro Verri², Francesca Odone²

¹Robotics, Brain and Cognitive Sciences Department, Istituto Italiano di Tecnologia, Italy
²Department of Computer Science, Bioengineering, Robotics and System Engineering, University of Genova, Italy

This research has been conducted in the framework of the European Project CODEFROR (FP7-PIRSES-2013-612555)

The framework of this research is the European Project CODEFROR (FP7-PIRSES-2013-612555).

**Experimental setting**

We acquired indoor videos of two subjects, observed from two slightly different viewpoints, while performing repetitions of a given action. We considered a repertoire of dynamic movements typical of an interaction setting: Lifting an object from a height to a different one (see Fig.1); Throwing and object away; Pointing a finger towards a 3D location; Transporting an object from left to right, or from and to random positions, on a table; Waving the hand; Gesticulating while talking. For each subject, we acquired 2 video sequences for each viewpoint. Each sequence depicts 20 repetitions of a given action. As for the non-biological counterpart, we consider videos of bouncing and rolling balls, a toy car, a pendulum and a lever. The whole data set has been split in training set - used only for model estimation - and test set - used only for performance evaluation.

**Computation model**

We initially relied on the optical flow, to obtain a first low-level motion representation. Then, inspired by the Two-Thirds Power Law, we defined on it a set of dynamic features, computational counterpart of the elements having a role in the law, namely tangential velocity, curvature, radius of curvature and angular velocity (Fig. 2). At time t, we segmented in the image a (moving) region of interest R(t), and associated each point in it, p(t) ∈ R(t), with a feature vector F(p(t)) = [v(t), c(t), r(t), a(t)] where each entry is computed as in Tab. 1.

On a first experiment, we measured such features in correspondence to the centroid of the moving region; then, we also considered a histogram-based representation. In both cases, we designed appropriate Multi-Cue Kernels as weighted sum of similarities between homogeneous features. We then applied the kernels within SVMs binary classifiers to the biological vs non-biological motion classification problem.

**Results and conclusion**

- The effect of the Two-Thirds Power Law can be appreciated also on a video analysis setting (in Fig. 3 samples from lifting actions).
- Although the coarse motion representation, the classification accuracy is around 89%, reminiscent of newborns vision abilities;

* The model showed tolerance to changes in the viewpoint. Nevertheless, it does not require a priori knowledge of the scene, nor on the human body kinematics.

This research has been conducted in the framework of the European Project CODEFROR (FP7-PIRSES-2013-612555)