
Séminaire Heudiasyc

Mardi 3 avril 2018 à 14 h en GI042 (Bât Blaise Pascal – UTC)

Models of Visual Attention for Understanding Humans

Christian Wolf, associate professor, INSA Lyon

Abstract:

We address human action recognition from RGB data and study the role of articulated pose and of visual attention mechanisms for this application. In particular, articulated pose is well established as an intermediate representation and capable of providing precise cues relevant to human motion and behavior. We describe two different methods which use pose in different ways, either during training and testing, or during training only.

The first method uses a trainable glimpse sensor to extract features on a set of predefined locations specified by the pose stream, namely the 4 hands of the two people involved in the activity. We show that it is of high interest to shift the attention to different hands at different time steps depending on the activity itself. The model not only learns to find choices relevant to the task, but also to draw away attention from joints which have been incorrectly located by the pose middleware.

A second method has been designed to explicitly remove the dependency on pose during training, making the method more broadly applicable in situations where pose is not available. Instead, a sparse representation of focus points is calculated by a dynamic visual attention model and passed to a set of distributed recurrent neural workers. State-of-the-art results are achieved on several datasets, among which is the largest dataset for human activity recognition, namely NTU-RGB+D.