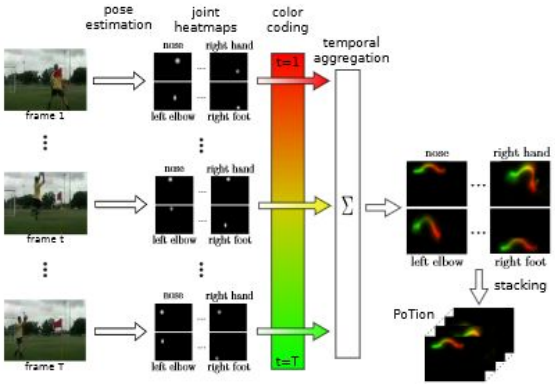
3D Pose Motion Representation for Action Recognition



Goal: Implement and evaluate an action recognition framework based on 3D human pose features

Description:

Action recognition is one of the most fundamentals problems of computer vision. Human pose features provide valuable cues for recognizing human actions. To this end, [1] recently proposed an efficient motion descriptor based on 2D pose features. Specifically, the authors first run a state- of-the-art human pose estimator and extract heatmaps for the human joints in each frame. Then, a motion descriptor is obtained by temporally aggregating these probability maps. The resulting motion descriptor is trained to recognize actions and is able provide the state-of-the-art performance even with shallow neural network architectures.

Two hand wave

Hand clap

While 2D pose features are helpful in estimating the human action, they lack depth information which is crucial for recognizing fine-grained actions. Therefore, we would like to account for the depth of human joints and extend this idea to the 3D setting, where 3D pose features are aggregated temporally within a volumetric representation. The resulting motion descriptor is then going to be trained to recognize human actions using different neural network architectures and compared against the state-of-the-art.

[1] “PoTion: Pose Motion Representation for Action Recognition”, Choutas et al. CVPR 2018

Requirements / Tools:

Python, PyTorch

Supervisor:

Bugra Tekin <bugra.tekin@microsoft.com>
Federica Bogo <febogo@microsoft.com>
Taein Kwon <taein.kwon@inf.ethz.ch>