

Action Recognition Scheme based on Skeleton Representation with DS-LSTM Network

Xinghao Jiang, Ke Xu, Tanfeng Sun

Abstract— Skeleton-based human action recognition has been a popular research field during the past few years. With the help of cameras equipping deep sensors, such as the Kinect, a human action can be represented by a sequence of human skeleton data. Inspired by the skeleton descriptors based on Lie group, a Spatial Temporal Skeleton Transformation Descriptor (ST-STD) is proposed in this paper. The ST-STD describes the relative transformations of skeletons, including the rotation and translation during movement. It gives a comprehensive view of the skeleton in both spatial and temporal domain for each frame. To capture the temporal connections in the skeleton sequence, a Denoising Sparse Long Short Term Memory (DS-LSTM) network is proposed in this paper. The DS-LSTM is designed to deal with two problems in action recognition. Firstly, to decrease the intra-class diversity, the Spatial Temporal Auto-Encoder (STAE) is proposed in this paper to generate representations with higher abstractness. The denoising constraint and the sparsity constraint are applied on both spatial and temporal domain to enhance robustness and to reduce action misalignment. Secondly, to model the action sequence, a three-layer LSTM structure is trained with STAE representations for temporal modelling and classification. The experiments are carried out on four popular datasets. The results show that our approach performs better than several existing skeleton-based action recognition methods, which prove the effectiveness of our method.

For the published version of record document, go to:

<http://dx.doi.org/10.1109/TCSVT.2019.2914137>