Short Story Abstract

A Survey on 3D Skeleton-Based Action Recognition Using Learning Method

The paper describes how human posture can help detect the actions with the help of projecting human skeleton system. The earlier versions of action recognition systems analysed video or RGB data dominant methods. It was noticed that thought this field had a lot of scope with perspective of deep learning architectures, it didn't have quite a lot of research happening in this area. Computer vision is a broad domain with a very wide scope of applications in the area.

One innovative way to recognize gestures in all directions irrespective of the angle of the video recorded for the person is to use the RGB image sequences to recognise the body structure of the human based on his skeleton. The former method of breaking down into representation for joints and bones, required more computational consumption, less robustness with complex background imagery as well has changes in the motion speed and body scale.

The developments and improving performances in Deep Learning methods using Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), and Graph Convolutional Neural Network (GCN) have opened pool of possibilities in the field of body gesture recognition. For RNN method, skeleton sequences are made up of time series sequences of coordinate positions of joints.

The three kinds of methods had exemplary performance but most review work only contained RGB and RGB-D image data method. The paper focusses on the three deep learning methods, general pipeline in 3D skeleton-based action recognition, spatial-temporal modelling and challenging datasets like NTU-RGB+D accounting in 3D skeleton-based action recognition.

There are available datasets which enable us to train and test the model, such as the video samples collected by Microsoft Kinect v2. This is one of the largest skeleton-based action recognition databases. It contains samples for two views: cross-view and cross-subject.

At last, compares performances of skeleton based techniques from which it is concluded that the conventional methods have better performances in original dataset. The accuracy on NTU-RGB+D dataset is quite high so it's hard to difficult to improve it further. In long term, real time monitoring will still be an open-ended problem and will might see leads on its research in future.