KUWAIT COLLEGE OF SCIENCE & TECHNOLOGY
كليـــة الكـــويت للعلـــوم والتكـنولـوجيا
Private University
جامعـــة خـاصـة

Vision Based Activity Monitoring for Human Behavior Modelling

(*PR18-18QS-01*)

Camera based surveillance systems capture subtle details of human movements which can provide important behavioral patterns which play important role in many real-life applications. For intelligent processing of such video streams, we need robust and accurate human detection and tracking algorithms which can build the foundation for higher level analysis of tracked patterns such as recognition of activities, actions, emotions and interactions. In this project, we are focusing on following two problems from the computer vision perspective: Human tracking, and Action recognition. The solution for these problems provides the basic inputs required for vision based human behavior modeling. The objective is to apply the latest developments in neural networks based machine learning models - Convolutional Neural networks (CNNs) with Recurrent Neural networks (RNNs) to design novel solutions for the problems under study. We follow complete data driven approach for content modeling which can be easily retrained on other datasets.

In the proposed human tracking algorithm, we apply tracking by detection approach. We use a CNN detector to recognizes the human objects in video frames as bounding boxes, and the association between the detections in current frame with active tracks is resolved by a Long Short Term Memory (LSTM) based RNN. We use the *maskRCNN* [16] detector for human recognition task which generates instance level segmentation in addition with the human region detection. We use the *appearance* and *motion* features of the detected targets and model them using LSTM based recurrent neural network for solving the association problem. The approach has been validated on MIT challenge datasets [1,14,17] where we achieve competitive performance in comparison with latest state-of-the-arts..

We also apply CNN and LSTM combination for action recognition task. Our work uses *appearance* features based low level representation for video frames which use pre-trained CNN for feature extraction in hierarchical LSTM framework for learning different action categories. The proposed method capitalizes on this memory attribute for fusing input streams in high-dimensional space by exploiting the spatial and temporal correlation. In this method, we define the temporal stream input on the LSTM learned spatio-temporal summary of the video frame sequence. Action recognition from videos focuses on temporal evolution in low-level features captured in video data. We demonstrate that the proposed two-stream approach provides and efficient and robust solution or action recognition problem. The experimental validation of our solution is presented on UCF-101 dataset (93.1%) and HMDB-51 dataset (71.3%).