**ABSTRACT**

Human Activity Identification (HAI) in videos is one of the trendiest research fields in the computer visualization. Among various HAI techniques, Joints-pooled 3D-Deep convolutional Descriptors (JDD) have achieved effective performance by learning the body joint and capturing the spatiotemporal characteristics concurrently. However, the time consumption for estimating the locale of body joints by using large-scale dataset and computational cost of skeleton estimation algorithm were high. The recognition accuracy using traditional approaches need to be improved by considering both body joints and trajectory points together. Therefore, the key goal of this work is to improve the recognition accuracy using an optical flow integrated with a two-stream bilinear model, namely Joints and Trajectory-pooled 3D-Deep convolutional Descriptors (JTDD). In this model, an optical flow/trajectory point between video frames is also extracted at the body joint positions as input to the proposed JTDD. For this reason, two-streams of Convolutional 3D network (C3D) multiplied with the bilinear product is used for extracting the features, generating the joint descriptors for video sequences and capturing the spatiotemporal features. Then, the whole network is trained end-to-end based on the two-stream bilinear C3D model to obtain the video descriptors. Further, these video descriptors are classified by linear Support Vector Machine (SVM) to recognize human activities. Based on both body joints and trajectory points, action recognition is achieved efficiently. Finally, the recognition accuracy of the JTDD model and JDD model are compared