Dynamic 3D Hand Gesture Recognition with Learning Spatio-Temporal Aggregation from Different Representation

to be Submitted to IEEE Transactions on CSVT

Reza Azad
Computer Engineering Department
Sharif University of Technology, Tehran, Iran
razad@ce.sharif.edu

Maryam Asadi
Computer Engineering Department
Sharif University of Technology, Tehran, Iran
masadia@ce.sharif.edu

S. Kasaei
Computer Engineering Department
Sharif University of Technology, Tehran, Iran
skasaei@sharif.edu

***Abstract*—** Hand gesture recognition is one of the most applicable and hot research topics in computer vision community. The recent advances in imaging devices, like Microsoft Kinect, have received a great deal of attention from researchers to reconsider problems such as gesture recognition from depth information. Hand gesture recognition refers to the classification of dynamic hand movements in action videos. Generally, hand gesture recognition includes three main steps: hand detection, feature extraction and classification. The first step plays an important role in hand gesture recognition. The most challenging part of hand gesture recognition is the second step which is the process of extracting high level description from hand movements. Finally, machine learning tools are used to classify gestures as the last step.

This research aims at proposing an appropriate approach to recognize human hand gestures from depth videos recorded with mounted camera. In this thesis, a method has been presented for recognizing hand gesture, which is based on linearly aggregation of spatio-temporal information in different levels. The proposed method relies on the fact that there are considerably diversities of performing hand gestures per person. To deal with this challenge, a video summarization step for hierarchical representation of video has been proposed. Video summarization result in increasing intra-class similarity and in the meanwhile raises the inter-class dissimilarities. In addition, the research presents a new weighted motion mapping method to extract spatio-temporal information. This method has been evaluated on three public datasets, i.e., MSR Gesture 3D, SKIG and MSR Action 3D, which achieved state of the art results with 98.05, 97.31 and 95.24 accuracy rate, respectively. The experimental results have shown the robustness of the proposed method against occlusion, noise, diverse sizes of hand shape, variable length of videos, and its higher applicability in real world applications.

Keywords— **Hand gesture recognition; Binary weighted mapping; Visual words; spatio-temporal description; 4D videos; RGB-D camera**

Github: ***https://github.com/rezazad68/Dynamic-3D-Action-Recognition-on-RGB-D-Videos***