SIGN LANGUAGE RECOGNITION AND TRANSLATION

Bassem Seddik

SAGE laboratory, National Engineering School of Sousse, University of Sousse, Tunisia
Email: bassem.seddik.tn@ieee.org

ABSTRACT

Human-action recognition is still an open research field in the computer vision community. The actual orientation leans towards the description from more than one modality combined with massive learning strategies [1] and the introduction of Microsoft's Kinect sensor accelerated this progress.

We aim in our thesis to recognize sign language and focus specially on the gestures performed by the deaf and dumb persons in a multi-modal context. For them who cannot speak, their unique communication solution is through sign language. The utility of our work is justified by the large number of the targeted population. According to the world federation of the deaf, there are over 70 million deaf people around the world. This increasing population uses 137 different sign languages.

For speaking and reading persons, we are able to use voice recognition services or text translation in order to allow the interaction with strangers or foreigners. But until now, there is no service capable of recognizing sign language actions. Such applications are still within science fiction.

In our works, we focus on overcoming two different challenges. The first one is related to the efficient recognition of actions presented to our software solution. The second concerns the ability to deal with continuous streams of confusing actions performed at different speeds as in real-life.

For this purpose, we first used the multi-modal information offered by the Kinect sensor. It is a condensation of technology that allows the simulation of the human vision system. It acquires colour images and offers an approximation of depth information as shown in Fig.1. In addition, it allowed us to estimate the orientations and speeds of the different body joints. We also focused on the face and hand regions as they convey more accurate details about the pronunciation and the signs gestures. This description allowed us to obtain high action-recognition rates.

For the second challenge related to the action temporal-segmentation, we proposed a solution that uses our rich description at different temporal scales [2]. As presented in Fig.2, by doing so, we have been able to deal with the non-interesting actions and to recognize the important ones as they are performed by the human actor.

Our experiments have been validated on the multi-modal dataset offered by the Chalearn 2014 gesture challenge [1].

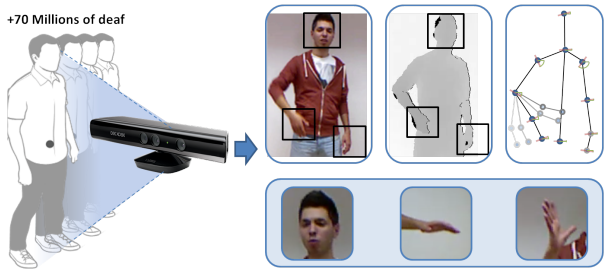

Fig. 1: From the kinect sensor, we are able to robustly describe the motion of the body, face and hands to recognize the action.

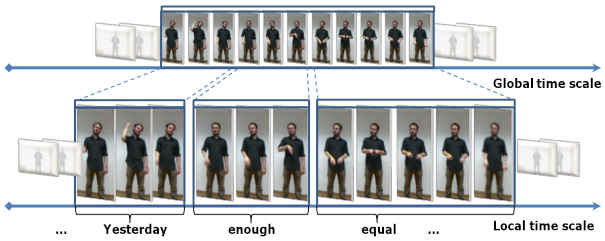

Fig. 2: Our solution analyses the actions, as they are performed, at different scales: wider global and closer local ones.

The obtained results proved the effectiveness of our solution within the literature and opened new perspectives for interaction with the deaf and dumb persons.

Index Terms— human action recognition, multi-modal data, multi-resolution analysis, temporal segmentation, Chalearn gesture challenge.

REFERENCES

[1] S. Escalera, X. Bar, J. Gonzlez, M.A. Bautista, M. Madadi, M. Reyes, V. Ponce-Lpez, H.J. Escalante, J. Shotton, and I. Guyon, "Chalearn looking at people challenge 2014: Dataset and results," in *ECCV Workshops*, Sept. 2014.

[2] B. Seddik, S. Gazzah, T. Chateau, and N. Essoukri Ben Amara, "Augmented skeletal joints for temporal segmentation of sign language actions," in *IEEE IPAS'14*, Hammamet, Tunisia, Nov. 2014.

978-0-9928626-3-3/15/$31.00 ©2015 IEEE