Continuous Hand Gesture Recognition using Deep Coarse and Fine Hand Features

Hazem Wannous* and Jean-Philippe Vandeborre

IMT Nord Europe, Univ. Lille, CNRS, UMR 9189 - CRIStAL, F-59000 Lille, France

**Hazem.wannous@imt-nord-europe.fr*

Introduction

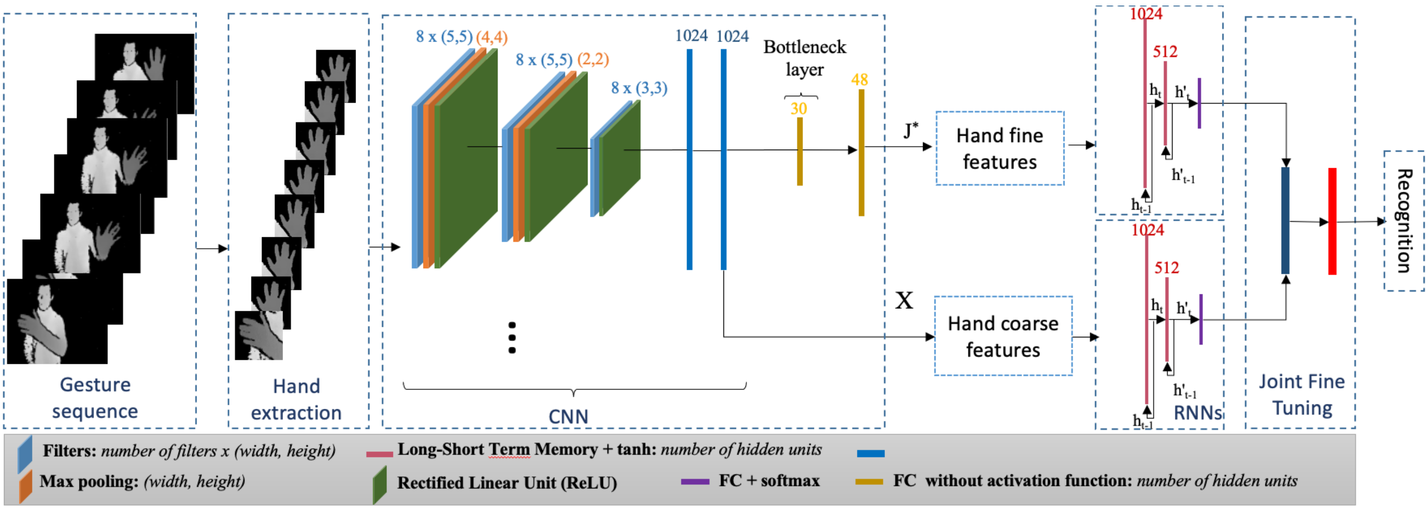
Context and Issues:

- Using hand gestures as a HCI modality introduces intuitive and easy-to-use interfaces for a wide range of applications.
- The hand is an object with a high number of degrees of freedom and with high similarities derived from the heterogeneities of possible gestures.
- Feature learning has to learn mutually spatial and temporal information, because gestures can be defined both by the shape variations and movements of the hand.
- The computation complexity has to be small enough so that the algorithm can predict an incoming gesture in real time

Contributions:

- We propose an end-to-end architecture based separately on learned temporal variation of coarse and fine features extracted from a CNN trained on depth sequences.
- Both features fed to two RNNs in order to model the temporal aspect of the hand poses and the shape variations over the time.
- We introduce of a new dataset of heterogeneous gestures recorded in an online scenario by a depth camera.
- We design a light efficient approach for online recognition of hand gestures. Simplicity and lightness is one of our goals for HCI applications.

Approach



Results

Datasets:

- **Online DHG**: 280 sequences of 10 continuous gestures,14 categories: fine and coarse gestures (14 and 28 classes). Pre-segmented sequences of this dataset constitutes the SHREK'17 dataset of 2800 gesture sequences [1]
- **NVIDIA** [2]: 1532 sequences of 25 gestures captured following a HCI based on hand gestures in a car scenario.

Metrics:

- Receiver Operating Characteristic (ROC) curve
- Normalized Time to Detect (NTtD) [3]
- Accuracy

Dataset	NTtD
Online DHG	0,2104
NVIDIA	0,2158

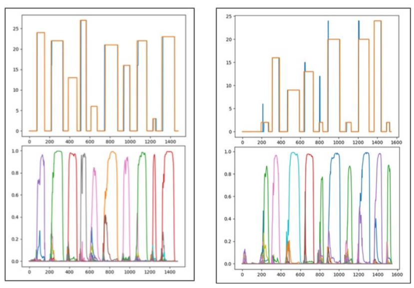
Method	14 ges.	28 ges.
Guerry *et al.*	82.9	71.9
Ohn-Bar *et al.*	83.8	76.5
Oreifej *et al.*	78.5	74.0
De Smedt *et al.*	88.2	81.9
Hou *et al.*	93.6	**90.7**
Chen *et al.*	**94.4**	**90.7**
Ours	94.2	90.5

Online DHG Dataset

Method	Features	Accuracy
Human		88.4%
HOG^2	HC	36.3%
SNV	HC	70.7%
C3D	DL	78.8%
R3DCNN	DL	80.3%
FOANET	DL	73.7%
Ours	DL	**81.3%**

NVIDIA Dataset

Obtained accuracies compared to some deep (DL) and hand-crafted (HC) state-of-the-art methods



Gesture detection performance on 10 continuous gestures (Online DHG). Obtained results (blue) versus the ground-truth (orange) where the x-axis is the time in number of frames and the y-axis represents the class outputs

Conclusion

- We proposed an online gesture recognition system as a whole pipeline from hand pose estimation to the classification step.
- We used a transfer learning strategy allowed to outperform state-of-the-art approaches using less than half of the number of parameters of the baseline model.
- The experiments showed that our framework is able to detect an occurring gesture after only about 21% and 22% of the nucleus phase, respectively for the Online DHG and NVIDIA datasets.

References

[1] Q. De Smedt, H. Wannous, J-P. Vandeborre, J.Guerry, B. Le Saux, and D. Filliat. 3D hand gesture recognition using a depth and skeletal dataset: Shrec'17 track. In Proceedings of the Workshop on 3D Object Retrieval, 3DOR '17, page 33–38, Goslar, DEU, 2017. Eurographics Association.

[2] P. Molchanov, X. Yang, S. Gupta, K. Kim, S. Tyree, and J. Kautz. Online detection and classification of dynamic hand gestures with recurrent 3d convolutional neural network. In Proceedings of the IEEE CVPR, pages 4207–4215, 2016

[3] M. Hoai and F. De la Torre. Max-margin early event detectors. International Journal of Computer Vision, 107(2): pages 191–202, 2014