Hand Posture Recognition Using Convolutional Neural Networks

Dennis Núñez Fernández
National University of Engineering
Lima, Peru
dnunezf@uni.pe

Bogdan Kwolek
AGH University of Science and Technology
Krakow, Poland
bkw@agh.edu.pl

Hand gesture recognition is one evident way to build user-friendly interfaces between machines and users. In the near future, hand gesture recognition technology would allow for the operation of complex machines and smart devices through only hand postures, finger and hand movements, eliminating the necessity for physical contact between user and machine. However, gesture recognition on images from single camera is a difficult problem due to occlusions, variations of poses appearance, etc.

In recent years, Convolutional Neural Networks (CNNs) have become the state-of-the-art for object recognition in computer vision [1]. Despite high potential of CNNs in object detection [2] [3] and image segmentation [1], only few papers report promising results - a recent survey on hand gesture recognition [4] reports only one significant work [5]. Some obstacles to wider use of CNNs are high computational demands, lack of large datasets, and lack of hand detectors for CNN-based classifiers.

The proposed system operates on color RGB images. It has been designed to run in real-time with respect to low power usage, including devices without a GPU support. Given the extracted hand, a neural network based regressor is executed to estimate the wrist position, see Fig. 1. The wrist location is used to extract a sub-image with a hand, which is then fed to a CNN.

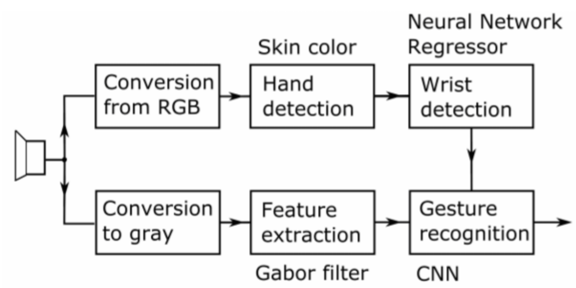

Figure 1: Block diagram of our method for hand posture recognition.

The CNN operates on gray images of size 38x38. We trained two models for gesture recognition: the first CNN was trained on gray images, whereas the second one on gray images filtered by a Gabor filter. We show that the system achieves far better classification performance on images pre-filtered by a Gabor filter. We show that a CNN operating over pre-filtered images gives good results on both images from the camera that has been used to acquire the dataset and a different camera. Training of the CNN was realized on a collected dataset of 6000 images representing ten hand postures. The dataset is available at: `http://home.agh.edu.pl/~bkw/code/ciarp2017`. The CNNs were trained using Caffe [6].

The first CNN was trained on raw gray images. Table 1 depicts the recognition performance for automatically and manually determined wrist position on test images acquired by different cameras.

Table 1: Performance measures using CNN on raw gray images

	Accuracy	Precision	Recall	F1 score
Kinect man.	0.950	0.945	0.960	0.949
Dif. cam. man.	0.928	0.931	0.937	0.934
Kinect aut.	0.905	0.906	0.932	0.906
Dif. cam. aut.	0.783	0.788	0.832	0.785

Table 2 depicts classification results that were obtained on images preprocessed by the Gabor filter. As we see, the results are far better than results in Table 1. Classification performance on images acquired by different camera is 87%. So, the improvement with respect to classification on raw images is 8.5%.

Table 2: Performance measures using CNN with Gabor-based preprocessing

	Accuracy	Precision	Recall	F1 score
Kinect man.	0.992	0.992	0.992	0.992
Dif. cam. man.	0.930	0.930	0.936	0.929
Kinect aut.	0.970	0.967	0.976	0.970
Dif. cam. aut.	0.868	0.868	0.882	0.866

Our system was designed to operate on humanoid Robot Nao as well as ARM processor-based mobile devices without GPU. In order to reduce the processing time of CNNs, we extract low-level features using techniques from biologically inspired computer vision, which are then further processed by moderate-size CNN. The recognition of hand posture on a single image is about 5 ms.

References

[1] Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: NIPS. (2012) 1097–1105

[2] Kwolek, B.: Face detection using convolutional neural networks and Gabor filters. In: Int. Conf. Artificial Neural Networks, LNCS, vol. 3696, Springer (2005) 551–556

[3] Arel, I., Rose, D., Karnowski, T.: Research frontier: Deep machine learning–a new frontier in artificial intelligence research. Comp. Intell. Mag. 5(4) (2010) 13–18

[4] Oyedotun, O., Khashman, A.: Deep learning in vision-based static hand gesture recognition. Neural Computing and Applications (2016) 1–11

[5] Tompson, J., Stein, M., Lecun, Y., Perlin, K.: Real-time continuous pose recovery of human hands using convolutional networks. ACM Trans. Graph. 33(5) (2014)

[6] Jia, Y., Shelhamer, E., Donahue, J., et al.: Caffe: Convolutional architecture for fast feature embedding. In: ACM Int. Conf. on Multimedia. (2014) 675–678