AI-based Hand Gesture Recognition System for Improved Human Machine Interaction in Cockpits

H. Sigurthorsdottir, S. A. Bigdeli, S. Saeedi, E. Türetken, L. A. Dunbar

Interest in contact-less Human Computer Interaction (HCI) systems is growing at an accelerated pace due to the recent COVID-19 outbreak. At CSEM, we are developing robust vision-based HCI systems integrated in airplane cockpits and cars. In this study, we present a hand gesture recognition system that achieves 98% accuracy over a diverse set of user participants. The system is comprised of a neural network with convolutional and fully-connected layers that can be run in real-time on low-cost processors with limited computational resources.

With the COVID-19 pandemic, interest in touch-free Human Computer Interaction (HCI), including vision-based hand gesture recognition, has skyrocketed. The worldwide market is expected to grow by a projected US$ 22.8 billion by 2027 [1] and vision based HCI systems are of particular interest at the moment.

As with many other vision applications, increased accuracy brought by deep learning techniques has recently been a major catalyst for the resurgence of HCI applications. In recent years, interest in HCI has been driven mainly by virtual reality systems, video games, control of robotic systems, and sign language interpretation. These applications are not mission-critical and therefore typically do not require an extremely high-level of accuracy.

At CSEM, we are developing innovative smart vision systems for cockpits in the framework of the CleanSky-sponsored project PEGGASUS **. The goal of which is to create a natural HCI system for pilots to alleviate some of the psychological pressure due to the complexity of modern flight management procedures, tools, and data streams.

When it comes to hand gesture recognition in the cockpit, error tolerance is very low, and security is the priority. Target recognition errors are therefore set below 1%, which is a challenging task for still images that lack temporal information [2].

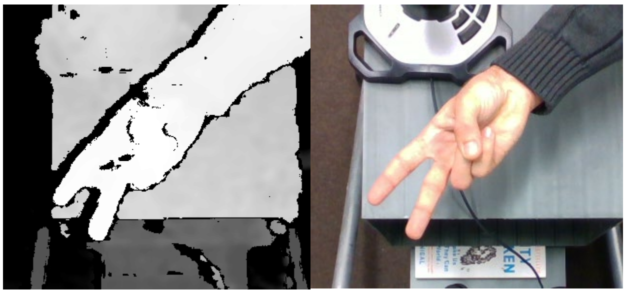

Figure 1: Depth map (left) and RGB image (right) from the PT1 dataset.

To this end, a hand gesture dataset specifically designed to emulate typical use-cases in the cockpit was collected by the PEGGASUS consortium (hereinafter PT1 dataset). The dataset includes more than 9K sample gestures from 30 subjects, captured by an RGB-D camera (Figure 1). It features five positive static gesture classes, which are selected to be easily distinguishable as well as a negative class composed of other hand poses usually seen in the cockpit.

We design a convolutional neural network (CNN) that maps its inputs to 6 output classes (5 PT1 gestures and a negative class for irrelevant gestures). The proposed network architecture is illustrated in Figure 2. The network is trained on depth maps only, as RGB images are not reliable in illumination-wise uncontrolled environments such as airplane cockpits.

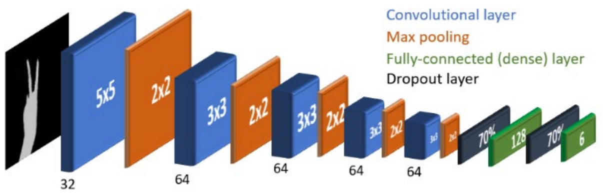

Figure 2: CNN network structure consisting of 5 convolutional blocks with max-pooling layers, and two fully connected layers at the end.

Training the neural network on PT1 led to 65% accuracy due to the relatively small size of this dataset. We therefore looked to transfer-learning to increase this accuracy. To do this first we trained our model on the public EgoGesture dataset [3]. We also resampled the gestures in this dataset to better represent the PT1 distribution. The CNN was benchmarked by hyperparameter-tuning (including the network architecture) on EgoGesture. Data augmentation techniques were used during the training to improve the generalization performance of the model on the PT1 dataset.

The model preformed with 92% accuracy on the EgoGesture validation set and was further fine-tuned on the PT1 dataset to compensate for the differences between the two datasets (camera calibration, capture angle, depth range, etc.). This transfer-learning strategy led to a significant reduction of the errors, bringing the accuracy to 98%.

The trained model will be tested further for robustness on-site in real cockpits. Future directions include incorporating temporal information, embedding the model on edge devices for real-time operation on a low power and computational budget, and using other modalities of sensing such as short-range radars.

[1] R. and M. (n.d.), "Gesture Recognition—Global Market Trajectory & Analytics." Retrieved June 11, 2020, from researchandmarkets.com

[2] B. Liao, *et al*., "Hand Gesture Recognition with Generalized Hough Transform and DC-CNN Using Realsense", ICIST (2018).

[3] Y. Zhang, *et al*., "EgoGesture: A New Dataset and Benchmark for Egocentric Hand Gesture Recognition", IEEE TOM (2018).

** This project has received funding from the Clean Sky 2 Joint Undertaking under the European Union's Horizon 2020 research and innovation program under grant agreement No. 821461.

Clean Sky