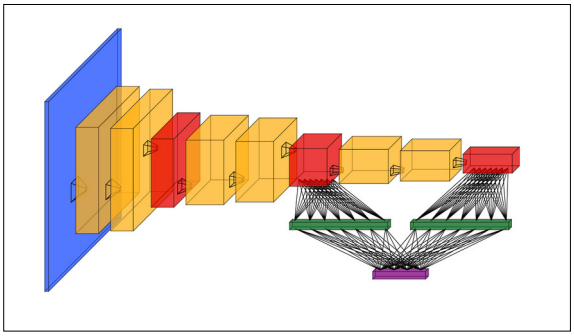
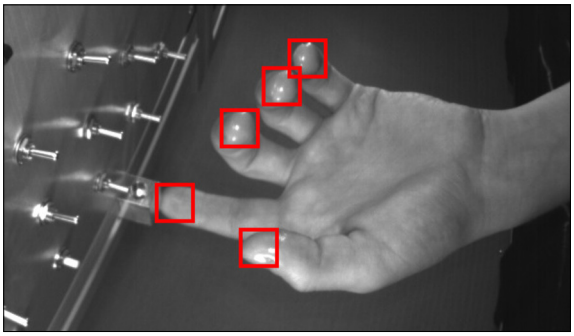
Student	Jonas Schmid
Examiner	Prof. Dr. Guido Schuster
Subject Area	Sensor, Actuator and Communication Systems

Hand Pose Estimation

Hand and Fingertip Detection using Deep Convolutional Neural Networks



The architecture of the second stage CNN.



The second stage CNN estimates the position and distinguishes all fingertips.

Introduction: Accurate hand pose estimation is an important requirement for augmented, virtual reality and other human computer interactions. The ICOM at HSR develops a simulation platform, focusing on a flight simulation application with a moving cockpit. In order to provide a realistic training environment for the pilot, the hands must be displayed in virtual reality. This enables the pilot to operate accurate on the control panel. To achieve this, the position and orientation of the hand in particular of the fingertips needs to be known. Therefore, a hand pose estimator is necessary. A setup with four monochrome cameras and ultraviolet illumination can be used to track marked fingertips. In this work the position of the fingertips is estimated directly from natural images without the case of any marks.

Solution: In the last years, deep learning, more precisely deep convolutional neural networks (CNNs), has outperformed and replaced several classical image processing algorithms. This work aims to show, that a CNN is a feasible approach to detect fingertips. Two available CNN architectures are adjusted and tested for our purpose. A first stage CNN estimates the position, size and confidence value of two bounding boxes for the left and right hand. A second CNN recognizes and distinguishes all fingertips of a hand, considering only the area within the bounding box. With the above mentioned setup with four cameras, we are able to record the required training data.

Result: Two CNNs are implemented, tested and evaluated using Googles machine learning library TensorFlow. A data set containing more than 50'000 images has been recorded and processed to train these CNN. The first localizes the hand with an average accuracy of 4.3% with respect to the input image. The second locates all fingertips with an average error of 9.2% with respect to the cropped image. A simple demonstration software was developed to detect and visualize the hands and fingertips. This allows a more subjective evaluation.

The first stage CNN estimates the location and size of the hand.