Robust Hand Detection and Classification in Vehicles and in the Wild

T. Hoang Ngan Le Kha Gia Quach Chenchen Zhu Chi Nhan Duong Khoa Luu
Marios Savvides
CyLab Biometrics Center, Carnegie Mellon University
{thihoanl, kquach, chenchez, chinhand, kluu, marioss }@andrew.cmu.edu

Abstract

Robust hand detection and classification is one of the most crucial pre-processing steps to support human computer interaction, driver behavior monitoring, virtual reality, etc. This problem, however, is very challenging due to numerous variations of hand images in real-world scenarios. This work presents a novel approach named Multiple Scale Region-based Fully Convolutional Networks (MS-RFCN) to robustly detect and classify human hand regions under various challenging conditions, e.g. occlusions, illumination, low-resolutions. In this approach, the whole image is passed through the proposed fully convolutional network to compute score maps. Those score maps with their position-sensitive properties can help to efficiently address a dilemma between translation-invariance in classification and detection. The method is evaluated on the challenging hand databases, i.e. the Vision for Intelligent Vehicles and Applications (VIVA) Challenge, Oxford hand dataset and compared against various recent hand detection methods. The experimental results show that our proposed MS-FRCN approach consistently achieves the state-of-the-art hand detection results, i.e. Average Precision (AP) / Average Recall (AR) of 95.1% / 94.5% at level 1 and 86.0% / 83.4% at level 2, on the VIVA challenge. In addition, the proposed method achieves the state-of-the-art results for left/right hand and driver/passenger classification tasks on the VIVA database with a significant improvement on AP/AR of 7% and 13% for both classification tasks, respectively. The hand detection performance of MS-RFCN reaches to 75.1% of AP and 77.8% of AR on Oxford database.

1. Introduction

The problems of hand detection and classification have been studied for years with the aim of ensuring the generalization of robust unconstrained hand detection algorithms to unseen images. However, the detection accuracy and classification score in recent hand detection systems [5, 3, 11, 10] are still far from achieving the same capabilities as a human due to a number of challenges in practice. For example, the hand variations, highly occlusions, low-resolution, strong lighting conditions, varied in shape and viewpoint as shown in Figure 1, are the important factors that need to be considered. Meanwhile, blurring of colors due to hand movement, skin tone variation in recorded videos due to camera quality are also the other difficulties in this problem.

(a) VIVA Hand database

(b) Oxford Hand database

Figure 1: Some examples of hand detection and classification (driver's left hand - RED, driver's right hand - GREEN, passenger's left hand - BLUE, passenger's right hand - YELLOW) results on VIVA database [2] (a); and hand detection Oxford database [14] (b) using our proposed MS-RFCN method (best view in color)

This paper presents an advanced Convolutional Neural Network (ConvNet) based approach, named Multiple Scale Region-based Fully Convolutional Networks (MS-RFCN), for hand detection and classification. In order to robustly deal with the challenging factors, we proposed to span the receptive fields in the ConvNet in multiple deep feature