Article

A Deep Learning-Based End-to-End Composite System for Hand Detection and Gesture Recognition

Adam Ahmed Qaid MOHAMMED[ORCID]**, Jiancheng Lv * and MD. Sajjatul Islam**

College of Computer Science, Sichuan University, Chengdu 610065, China;
adammohammed@stu.scu.edu.cn (A.A.Q.M.); sajjatulislam@stu.scu.edu.cn (M.S.I.)
* Correspondence: lvjiancheng@scu.edu.cn

check for updates

Abstract: Recent research on hand detection and gesture recognition has attracted increasing interest due to its broad range of potential applications, such as human-computer interaction, sign language recognition, hand action analysis, driver hand behavior monitoring, and virtual reality. In recent years, several approaches have been proposed with the aim of developing a robust algorithm which functions in complex and cluttered environments. Although several researchers have addressed this challenging problem, a robust system is still elusive. Therefore, we propose a deep learning-based architecture to jointly detect and classify hand gestures. In the proposed architecture, the whole image is passed through a one-stage dense object detector to extract hand regions, which, in turn, pass through a lightweight convolutional neural network (CNN) for hand gesture recognition. To evaluate our approach, we conducted extensive experiments on four publicly available datasets for hand detection, including the Oxford, 5-signers, EgoHands, and Indian classical dance (ICD) datasets, along with two hand gesture datasets with different gesture vocabularies for hand gesture recognition, namely, the LaRED and TinyHands datasets. Here, experimental results demonstrate that the proposed architecture is efficient and robust. In addition, it outperforms other approaches in both the hand detection and gesture classification tasks.

Keywords: hand detection; hand gesture recognition; human-computer interaction; deep learning

1. Introduction

Among the several human-computer research activities in computer vision and machine learning (e.g., human action recognition, pose estimation, and gesture recognition), hand gesture recognition is particularly important due to its various potential applications. Robust hand gesture detection and recognition in cluttered environments is a crucial task for many computer vision applications, such as human-computer interaction, sign language recognition, hand action analysis, driver hand behavior monitoring, and virtual reality. This task presents a challenging problem that has not yet been solved in computer vision and machine learning. Unlike many previous studies [1–7], which separately have tried to address hand detection or hand gesture recognition, our approach attempts to jointly solve the problem of hand localization and gesture recognition. This task, however, is very challenging, due to the significant variations of hand images in realistic scenarios.

The utilization of hand-crafted features has dominated early research in hand detection and gesture recognition. Most of these approaches have utilized hand skin color, texture, and appearance features for hand detection and gesture recognition [8–13]. However, their success is only found in certain well-prepared environments. The task remains challenging in real-life applications, due to problems posed by background complexity, occlusion, viewpoint, lighting changes, the deformable and articulated nature of hands, etc. Encouraged by recently emerged deep learning approaches for object detection and recognition [14–16], many researchers have proposed numerous approaches to