Vision-based Hand Posture Classification using Deep Learning

Khadija Hesham
Electrical and computer engineering department
university of Ottawa
khesh072@uottawa.ca

Dina Abdelhady
Electrical and computer engineering department
university of Ottawa
dabde007@uottawa.ca

Mona Mansour
Electrical and computer engineering department
university of Ottawa
mmans104@uOttawa.ca

Gehad Abo Kamar
Electrical and computer engineering department
university of Ottawa
gabok033@uottawa.ca

Youssef Metwally
Electrical and computer engineering department
university of Ottawa
ymetw027@uottawa.ca

Abstract

Poorly designed user interfaces can result in less optimal or even a negative user experience although the system is functioning almost ideal. The degree to which a system is easy to learn and used is called "usability". There is a trade-off between the ease of use and the ease of learning, i.e., when a system is easier to use, it's often harder to learn.
Having a way to communicate with a software system without using keyboard, mouse, or buttons has a great advantage. Human-Computer Interaction (HCI) is a field of study concerned with the efficiency and effectiveness of user interfaces, human-oriented input and output technology, and psychological aspects of user interfaces. It aims to give the user a positive user experience. Instead of using a mouse or a keyboard to communicate with a computer, we can utilize the advances in the artificial intelligence field to make an easier solution for this.
Having such a solution has many advantages, especially for speech and hearing-impaired people. Speech and hearing impaired people require real-time sign language posture and gesture recognition to use daily live traditional applications. According to the World Health Organization (WHO), 5% of the world's total population (360 million people) has moderate or severe hearing loss. In addition, there are a large number of sign languages used around the world, 300 different sign languages approximately, which could create a new opportunity that we can make use of it to unify a sign language using machine learning techniques by classifying the different hand sign languages together and training the machine on the resulting data set, so it can classify new unseen hand postures. There are various domains in which we can utilize hand posture/gesture recognition such as Virtual Reality (VR), Robot Control, Natural User Interfaces, and many others. There are two types of hand gestures: a static hand gesture (posture) and a dynamic hand gesture. A static hand gesture, which is also called "hand posture" is defined as the position of hands and fingers in space without any movement concerning the time. We will adopt a vision-based approach to our solution in which we will need a camera only to capture the user's postures. This reduces the dependency of users on sensor devices which is more user-friendly. In this paper, experiments have been made on a self-collected dataset by using several methods and techniques to classify ten hand postures adopting a vision-based approach.
The remainder of this paper is organized as follows: in section 2, a brief literature review of related work done on hand posture/gesture recognition is presented. In section 3, the used methodology and detail of the self-collected dataset are presented. Section 4 presents the experimental data and analysis. Finally, the conclusion of this work is given in Section 5.

***Keywords*—hand posture recognition, Convolutional Neural Network, Sign language classification**

I. Introduction

Hand posture communication has the best ease of use required for human interaction or human-machine interaction across different countries without using any physical devices. Therefore, developing a means of recognizing different hand postures is very important from the user's point of view, especially for hearing and speech-impaired people. The main contribution of our work is to compare and experimentally investigate the results of three deep learning models using a self-collected data set of hand position images. The three deep learning models used are VGG-16 and VGG-19 using transfer learning, and a sequential convolutional neural network (CNN). Since there is a lack of research related to evaluating the efficiency of these deep learning models, we have incorporated some efficiency measures into our evaluation. The sequential convolutional neural network (CNN) outperforms the other two deep learning models with a test accuracy of %97.