Hand gesture control - Prototype research

Introduction

Interacting with computers through hand gestures has become a research topic of more and more interest in the last decades, as an alternative to the standard input methods like the keyboard and mouse. It has many applications such as virtual/augmented reality, sign language recognition and interpreting for disabled people, robot control, etc. However, detecting gestures quickly and reliably still remains a challenge, many researchers striving to improve hand gesture recognition technology.

This project is especially challenging due to the fact that the software has to detect hand gestures based only on an RGB camera. At present, many solutions for hand gesture recognition rely on additional information, such as depth information from an RGB-D camera, or solutions that require the user to wear a special glove with sensors attached to it, so that the movements of the hand and fingers can be more reliably and accurately detected.

Proposed solution

Since the reliability of the gesture detection heavily influences the user experience, this should be the top-most priority. The proposed solution is based on a research paper titled [Real-Time Hand Gesture Recognition Using Finger Segmentation]

First, the software will attempt to segment the current frame, and extract all the pixels that belong to the hands. Second, after the image has been segmented, various features can be extracted based on the silhouette of the hand, such as the palm point, the position and shape of each finger, all which can be used to infer the skeleton of the hand. Finally, a rule classifier will be applied in order to predict the hand gesture with accuracy.

In the paper, the first step (image segmentation) is performed using background subtraction in the case of static cameras, or detecting skin color using the HSV model. However, both methods can be unreliable under certain conditions, therefore I propose using a Convolutional Neural Network in order to extract the hands. The EgoHands dataset can be used in order to train the Neural Network. It contains 48 Google Glass videos of complex, first-person hand gestures, with a significant degree of difficulty.

Due to the above decision of using neural networks in order to segment the image, the software will have to be developed only with PC or laptops in mind. Due the computational costs of running a neural network, most mobile and embedded devices would not be able to perform the hand gesture detection in real-time, which means increased lag and less reliability for the user.

At first, the prototype will be built in a high level language, like Python in order to make sure that the proposed solution is at least viable. After that, the project will transition to be developed in C++ for higher computational efficiency.

Challenges

The first and foremost challenge will be the image segmentation. The neural network must have a high enough accuracy in order to further be able to detect the hand gestures. According to this paper, [Analysis of Hand Segmentation in the Wild], the chosen data set is hard, with the best method at the time obtaining around 75% accuracy. But luckily, the situations present in the data set are much more complex than in our software's case.

Even after segmentation, the method is very dependent on being able to detect the skeleton of the hand properly. Without depth or other information, the software will have to rely on the user to properly make his gestures as clear as possible to the camera.

One final challenge that the proposed algorithm faces is that it can only detect static gestures. This challenge is made even more difficult due to the lack of an RGB-D camera. Therefore, the dynamic hand gestures, if any, will have to be limited to very simple motions.

The issue of image segmentation also presents a development challenge. In the case that the segmentation using neural networks does not work, we will have to ultimately use more primitive methods such as background subtraction or skin color extraction based on HSV.

One other development issue will be to give the user the impression that the software is fast , reliable and that it behaves the way the user expects it to behave. For instance, there must always be enough feedback given to the user, so that it is clear what the software recognizes and doesn't recognize as a gesture.

Finally, transitioning from the prototype to the development of the final product will undoubtedly pose a challenge, much of the code for the prototype having to be rewritten in C++.

Bibliography

Real-Time Hand Gesture Recognition Using Finger Segmentation: Zhi-hua Chen, Jung-Tae Kim, Jianning Liang, Jing Zhang, and Yu-Bo Yuan, Real-Time Hand Gesture Recognition Using Finger Segmentation,, 2014
Analysis of Hand Segmentation in the Wild: Khan, Aisha Urooj, and Ali Borji, Analysis of Hand Segmentation in the Wild, 2018