Gesture Recognition Research Proposal

Background

Hand gesture control is enabled by hand gesture recognition - interpreting human gestures and translating them into machine-readable entities. This is achieved by employing several techniques from fields like image processing, computer vision, and more recently, machine learning.

Objective

We aim to devise a hand gesture recognition system, which will take its input from an RGB camera.

Challenges

An important issue is the fact that restricting the input retrieval to an RGB camera will constrain our algorithms to infer meaning from 2D frames. Where other hand gesture recognition approaches make use of additional IR sensors and the like to gather information about depth and add another dimension to the input, we will have make up for it in other ways. On the flip side, one less input dimension gives us more freedom with algorithm complexity.

One of the challenges that could arise along the way is how lighting will impact the video. Telling the hand apart from the rest of the image can be quite difficult if, for example, the hand is not well-lit and there is a dark object right behind it.
A second challenge might come from the quality of the video feed. Some algortihms are sensitive to noise, so we will have to handle this by trying to reduce it to a minimum in the pre-processing phase.

As for the implementation part, a first issue we will have to take into consideration is how we acquire our data, and as a consequence our training data. If the hand gesture feature will be used in certain environments, we could take the liberty to choose similar environments to acquire data for training our models, leading to better accuracy in our predictions. If it is meant to be used anywhere and in any condition, that needs to reflect in our training data. We also need to make sure we cover a large enough range of motions, as different people can have quite different ways of motioning the same gestures.
A second problem we need to address is the frame rate of the video feed. While less FPS would make processing easier, we have to ensure the gestures will still be recognisable.
At the opposite end of the spectrum, if we aim for a high frame rate for the video feed, the processing power could become an issue. As video processing is inherently demanding from a

computational point of view, an embedded device might face problems in this regard at higher frame rates.

Roadmap

A rough description of the steps we will take along the way is as follows:

1. Getting training data

After we decide which type of data we need as per the issue described earlier, we will have to analyse the trade-off between crafting and curating a training dataset ourselves and searching for a premade one.

2. Preprocessing

For the preprocessing of data, we will look into two approaches for identifying the hand in a frame: (1) edge detection with the Canny edge detector and (2) background substraction. While likely more accurate, the second approach would have to use the assumption that the hand is in the foreground of the image.

3. Building models

We will design a Convolutional Neural Network in order to classify the hand gestures.

4. Testing models

5. Fine-tuning approach

In case it turns out the accuracy is not as good as we would like it to be when testing our models, we should make sure that we are not overfitting. In order to avoid that, we could train several models on different samples of the dataset, by bagging or boosting. The predictions would then be made by running data through all of the models and having them vote on the final gesture class.