Mihalea Matei
AI

Challenge 0 – The researcher's challenge

1.Introduction

Controlling machines using hand gestures used to be a topic present in movies in the past that would let us wonder if it will ever be put into practice, but here we are today when more and more people are starting to show interest to the subject because of the large number of possible applications and domains it covers, but also because it can make the user's life much easier.

2.Roadmap

In my opinion, the plan for developing a project on this matter should be composed of the following steps, no matter what kind of applications we choose to develop: firstly, we decide on the type of software that we wish to build, and we choose to do that because this decision will affect a lot of the further development of the project(and after that we can create a states-actions diagram of the application, which will define what events each action will trigger for all states), we also need to have a part of research in order to estimate the necessary resources we will need to finalize the app. Moving forward is the development phase, in which we think of algorithms and technologies we are going to implement and use, and also think about alternative solutions, then we write the code. In the final stages of the project we can do small adjustments if necessary.

3.Development and difficulties

Depending on the type of application, the project could take slightly different directions, but even so it will be centered around the hand gesture recognition. Knowing that we will use an RGB camera, the project will most likely also contain some computer vision components. Another big part of programming this application is defining actions for each gesture, keeping in mind the current state. This may lead to some difficulties in implementation depending on application, because while we will have to define action for each gesture in the database, the same gesture could have different interpretations in different circumstances. For example, in an application that controls your PC using hand gestures, swiping left or right while browsing the internet could result in switching between tabs, while performing the same gesture could result in lowering or increasing the volume while listening to music or watching a movie. With all this in consideration, we will have to think of all the states that the application could transition to, and define sets of actions for each of these states.

Coming to the algorithms that we will employ, my first and best idea would be using Convolutional Neural Networks for hand gesture recognition, combined with some image filtering to make sure the image is smooth and the gestures are recognized by the algorithm in all circumstances. With the nature of the hand gesture recognition problem being that of a classification problem, it is obvious the we will need to use an already existing dataset, and although there are online examples of such databases, these could be limited and therefore could constrain the application that we wish to develop. Consequently, the first challenge in developing the application would be to find an adequate dataset for our problem, meaning it should contain a large enough number of gestures such that we can define all the actions that can happen for all scenarios of utilization. There is also the issue of the gestures themselves, and keeping in mind that an intuitive application is desirable, so the gestures should have a direct correlation to the action that they trigger, so even though there are enough gestures in the database, we need to make sure that they are suitable for the kind of software that we are producing.

Another challenge in developing the application could be delimiting between gestures and not trying to classify for the period of time inbetween consecutive gestures. This part can prove to be pretty difficult in my opinion because of the unpredictability of human movements, thing that can make it hard for the program to detect if a hand movement is a gesture and should be recognized or if it is a transition between gestures and shouldn't.

4. Implementation and difficulties

Next, we will take a closer look at the algorithms and methods we decided to utilize, and we will start with the processing of the recording, and more specifically the filtering of the frames, component which might be overlooked at the beginning of the project but in fact constitutes a very important part in assuring a smooth and precise functioning of the application as a whole. To ensure the highest precision and the best results, we will use: background subtraction, to ensure that the classification is not affected by the environment, color filtering, which make sure that the color of the hand doesn't interfere with the classification process, hand region extraction and image resizing, to make sure all the dataset images and the recorded ones have the same size and format.

Regarding the process of classification of the gestures, I chose to utilize a Convolutional Neural Network because it can produce good results without using a very complicated network, and also doesn't depend on the feature learning process, unlike other methods like Support Vector Machines or Hidden Markov Models, reason why Convolutional Neural Networks are becoming more and more popular, especially in the domain of image classification.

Below, there are presented the main difficulties or challenges that might arise during the implementation of the application. Firstly, in the case of the neural network, a complicated design can cause a slow functioning of the application overall, but on the other side, a network that is not that well structured or too simple can cause a deficit in the precision of the classification .Therefore, we would ideally want a trade off between speed and precision, and not tipping the scale in one direction.

Another instance in which problems might appear is the the case of pre-processing and, not unlike the choosing of the network design, this presents concerns regarding the execution time . This is very important in the context of this application because we would like it to run in real-time, meaning that the processing or filtering each frame can cause delay in classification and therefore delay in the application responding to gestures if not carefully done.

As I mentioned in the beginning, hand gesture recognition can have applications in various domains, but a common denominator for most of them is the precision of actions resulted from the recognition process . A good example of such application is controlling the PC using hand gestures, in this regard controlling the cursor being the action which would require a fine tuning.

As a short conclusion, I would say this project may present some obstacles, but overcoming them and building a strong application could be a big achievement in a domain on the rise that still has room for innovations and developments.