MALIS : Project proposal

Arnaud Minondo
Dorian Gailhard
Timothée Ducros

Recognition of the area of the screen looked at by the user

Problem tackled : Some people have motricity diseases and can't use a computer as it requires hands and fingers movements. Our goal is to apply machine learning techniques to know where they are looking on the screen, to build a basis on which we could improve existing techniques which allow them to use computers.

Method : We intend to feed a neural network images of faces looking at different parts of the screen - which will be segmented in different zones - and adapt its weights so that the most probable zone that is looked at matches the real zone.

The pictures will be taken by ourselves thanks to a program displaying dots in the different zones of the screen and taking pictures through a webcam, that we will program. The pictures will be very standardized, with the eyes always in the same place.

We will then try the classical architecture of a feedforward neural network and the stochastic gradient descent for the learning method, tinkering with the number of neurons and of layers. Then we will move on to the convolution neural network architecture and compare its efficiency with the previous one.

The initial results will most likely be awful so image processing techniques to enhance the contrast between the face and the eyes, or edge detection with the Fourier Transformation will be used to keep only the meaningful information.

The goal is then to :

- first, have good recognition for a screen that is divided into 4 parts
- then, to divide the screen in more parts and have good results, and so on until we reach the limits of our capacity
- finally, to have an understanding on why we can't push our neural network any further

Intended experiments :

The plan is to :

- first, program the tool that will take pictures of faces looking at the dots, and verifying manually that it works
- then, build a simple neural network that will recognise eyes looking at extreme directions - ie the pupil position is exaggerated - to begin with pictures that are really different

- then do the same but with pictures less distinct
- then do the same with more parts

The evaluation will be done through the results on a test database, comparing the different architectures and selecting the best one. These results will be expressed through a percentage of correct guesses.

Prior research : This kind of software already exists - see for example openEyes - https://people.cs.clemson.edu/~geist/public_html/seminar/openEyes.pdf - for Matlab, or PyGaze - http://www.pygaze.org/resources/downloads/Dalmaijer_Mathot_Stigchel_2013_PyGaze_manuscript.pdf - for Python, but is heavy on computing resources and you can't just detect a part of the screen that is looked at and you must have the exact location. Some existing softwares like openEyes even need their own equipment and can't work with just a webcam.

This will be what is interesting about our project : to have a good approximation of where the user is looking at without relying too heavily on resources.

We contacted Stéphane Lathuilière, teacher at Télécom Paris and he said that knowing the exact position is an extremely difficult problem, so this is why we stick with approximations.