Detecting emotions with facial recognition

Effects of higher dimensional facial vectorization on the aspect of emotion detection

András Tóth   
Székesfehérvár, Hungary  
csonti04@gmail.com

Patrik Tóth  
Székesfehérvár, Hungary  
patrikthetoth@gmail.com

Dr. Halász József  
*Alba Regia Technical Faculty  
of Óbuda University*  
Székesfehérvár, Hungary  
halasz.jozsef@amk.uni-obuda.hu

As new technologies are introduced, their full potential might not be apparent at first. As facial recognition is used more and more in several different devices and services, one good idea might as well separate one product from the rest. As our team has demonstrated in our previous paper, introducing facial recognition into a smart mirror is not only feasible, it can also be practical. However, there are DIY solutions that provide the functionality mentioned above, and our team strives to offer something more.

Therefore, we investigated the possibility of detecting facial emotions via the same 128-dimensional facial recognition system and compared it one that utilizes 512 dimensional vectors to represent the human face.

Keywords— face recognition, neural network, emotion, Euclidean distance

# Introduction

Our team's smart mirror consist of three major components, one of these were the face recognition pipeline. Over the course of the project some results pointed out, that the current state of the pipeline is not sufficient, thus new technologies were adopted [ref here, mtcnn, 512d?]

Because of this change in the underlying technologies , we were able to achieve greater accuracy for our face recognition system. Using this newly gained accuracy a new goal was set, to detect the effects of different emotions on the face recognition pipeline accuracy, and using this data prepare our system to differenciate emotions on our users’ face.

# Methods

## Creating the face repository

For the creation of the face repository, a software called FaceGen Modeller (demo version) has been chosen. The 4 faces for this paper have been created with the software’s randomizer option. In this software it is possible to adjust the faces so called Action Units [Ref Here]. These action units are responsible to different movements in the face itself, in this way it is possible to create facial expression like happiness or anger. For every measuring point with given facial expression and intensity with the needed action units [Ref Here] an XML document has been created for the purpose of reusability, these were used for the creation of the faces presented in this dataset. For each model 2 different set of pictures were created with different intensity of anger or happiness applied.

There sets were created for each emotion: neutral withe the intensity between 0 and 10 %, low with the intensity between 20 and 30 % and high, with the intensity value between 50 and 60 % with 1 percent increments for all sets.

These were then captured with the help of a program called ShareX [Ref here]. For the proper file name format, that contained the user's name, the applied emotion, and its intensity the program Bulk Rename Utility was used.

## Facial Recognition pipeline

The pipeline consists of three layers, namely: Detection, Representation and Classification. The detection layer finds a face in the input image, and after cropping the image to reduce complexity, it is fed into the representation layer, which is tasked with vectorizing the cropped image, and as such, provides us with an n-dimension vector that is specific to the face in the input image. After this, we store these vectors, or embeddings and once another face is encountered, we can compare them, which ultimately is the responsibility of the classification layer.

## Measuring accuracy

Once we have these embeddings, we can easily compare two faces by way of calculating the Euclidian distance of the two vectors, thus giving us a scalar distance measurement that we can use to determine the accuracy of the recognition pipeline. For example, 0 means the two faces are identical, while a distance of 1.6 means they are nothing alike. In practice, we see a threshold of recognition at around 0.6, as in, if the distance is larger than 0.6, the two users are not the same.

## Evaluating the gathered data

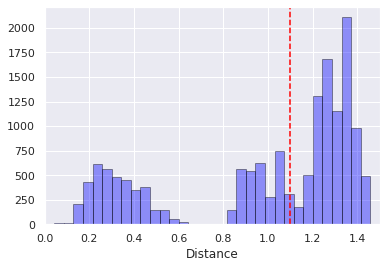
More than 30 thousand results were gathered during the test, these contained the data set user's data and evaluating user's data such as username applied emotion, the emotions intensity, and the Euclidean distance within the 2 users. This type of distance was used in the FaceNet paper [Ref here] to represent how closely two face represents the same person. The researchers in that paper used 1.1 as a segmentation threshold. Distances below 1.1 between two faces were considered to belong to the same user.

This project used Python 3 with Seaborn, Numpy and Pandas for data evaluation.

## Smart mirror system software architecture

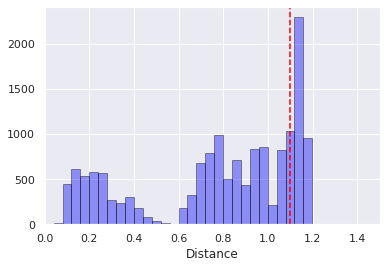
The given results were originally gathered to evaluate two different face recognition and identification systems. The chosen system was used as the basis for our software architectures face recognition system. This was used to access user related data from the central database to show relevant information to our users. These were transferred with simple HTTP Request and Responses, using a RESTful solution. The given software architectures design can be seen in [this] figure.

# Results



1. Figure Distances measured with 512D representational layer

The initial results showed that, the upgrade to the new 512-dimension representational layer was well worth it. We believe these additional dimensions in the representational layer help our system to represent the users with more granularity. As one can see in Figures 1 and two, the 512-dimensional system produces a bigger gap between those distance values that was measured with picture belonging to the same user, as opposed to those value that were the result of comparing two vectors that belonged to different users.

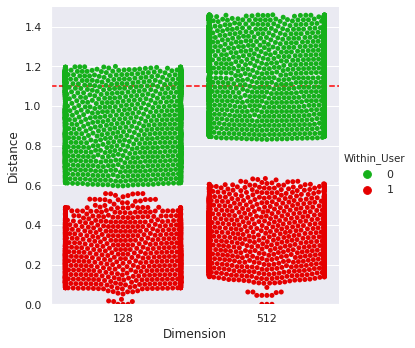


2. Figure Distances measured with 128D representational layer

All figures include a separator line at the distance value of 1.1, as recommended as a threshold value for separating users from each other [Facenet here].

On Figures 3 it is clearly seeable that the 512-dimension system produces higher distances in those situations when the vectors belonged to different users.

This means that with help of the new representational layer the system is more likely capable of evading user misclassification.



3. Figure Distances with given Dimension  
Within\_User indicates that the two vectors represented the same user

# Discussion

Text goes here brrrrr.

# Conclusion

Text goes here brrrrr.

##### Acknowledgment

The authors declare no conflict of interest. The authors would like to express their gratitude to those who have provided their help in the various experiments.

##### References

Text goes here brrrrr.

**IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.**