Detecting the effects of emotions and higher dimensional facial vectorization on facial recognition in a smart mirror system

**Andras Toth, Patrik Toth, Szabolcs Meszaros, Jozsef Halasz**

Obuda University, Alba Regia Technical Faculty  
Institute of Natural Sciences and Software Technologies

[sba.andras.toth@gmail.com](mailto:sba.andras.toth@gmail.com)

[patrikthetoth@gmail.com](mailto:patrikthetoth@gmail.com)

[villam983@gmail.com](mailto:villam983@gmail.com)

[halasz.jozsef@amk.uni-obuda.hu](mailto:halasz.jozsef@amk.uni-obuda.hu)

Abstract—

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 it has been demonstrated in a previous paper, introducing facial recognition into a smart mirror is not only feasible, it can also be practical. There are DIY solutions that provide the functionality mentioned above, the present concept strives to offer something more.

The aim of this paper was to investigate the possibility of detecting the effect of facial emotions via the same 128-dimensional facial recognition system and compared it one that utilizes 512-dimensional vectors to represent the human face.

The project utilized a face recognition pipeline used Euclidian distances to classify the users. These users were artificially created, with neutral, angry and happy emotions were applied to their faces. All together more than 30000 distances were measured, these were the basis for this paper. General linear model was used to analyze these distances.

The results showed that the solution with 512-dimensional vectors revealed significantly higher distances between different users. Within the same users, the emotional content was able to increase distances, and this effect was more prominent with 512-dimensional vectors compared to 128-dimensional ones.

In conclusion, our result indicate that the 512-dimensional solution had higher sensitivity and the effect of emotional content on facial detection must be considered in later studies.

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

# 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 [1][2].

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 differentiate emotions on our users’ face.

# Methods

## Creating the face repository

For the creation of the face repository, a software called FaceGen Modeller (demo version) [3] 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 [4]. 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 [4] 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. All 4 users were created with the male preset of FaceGen Modeller (demo version) to avoid any potential biases in the face recognition pipeline.

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 [5]. 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.

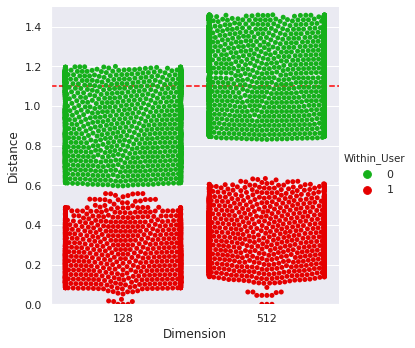
## Evaluating the gathered data

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 [6] 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.

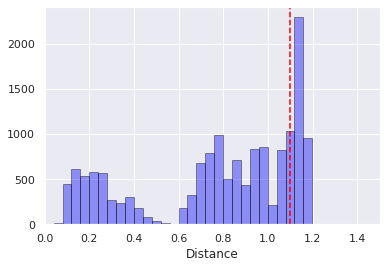
This project used Python 3 with Seaborn, Numpy and Pandas for data evaluation.

## 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.



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



1. Figure Distances measured with 128D representational layer

# Results

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.

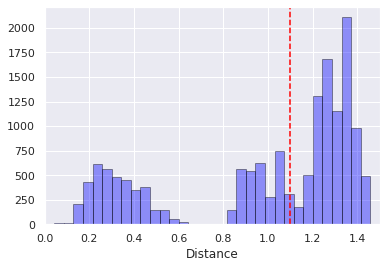
All figures include a separator line at the distance value of 1.1, as recommended as a threshold value for separating users from each other [6].

On Figure 3 it is clearly visible that the 512-dimension system produces higher distances, this effect is more pronounced in those situations when the vectors compared belonged to different users. This means that with help of the new representational layer the system is more likely capable of evading user misclassification.

This effect can be the result of more active Action Units, [4] i.e. more facial features are active at the same time.

With the gathered data it was possible to deduce which emotion had bigger effect on the measured distances, and how much the measured distance and its intensity correlate.

In measurements, when the two compared vectors belonged to the same user the lowest possible distance values were generated by those vectors that represented a user with low intensity anger, in turn the maximum possible values were generated by faces that represented high intensity anger.



2. Figure Distances measured with 512D representational layer

On the 512-dimensional system between these datapoints were the ones generated by vectors that represented faces with happiness. In general vectors containing low intensity happiness had lower distances than those with high intensity happiness.

These effects are shown on Figure 4 and 5. The curiosity on Figure 4 is that the higher percentage of low intensity anger are closer to the lower precent of high intensity anger. These effects are absent on Figure 4, due to the aforementioned formation of measured distances, this again implicates that, the 512-dimensional system is better at representing complex facial features, such as the emotion anger.

# Discussion

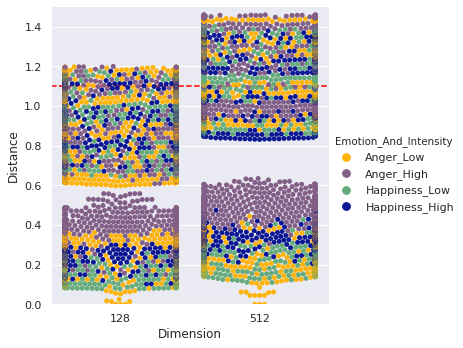
Text goes here brrrrr.

# 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.



4. Figure Emotions and their intensities

##### References

[1] K. Zhang, Z. Zhang, Z. Li, S. Member, Y. Qiao, and S. Member, “(MTCNN) Multi-task Cascaded Convolutional Networks,” *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1499–1503, 2016, doi: 10.1109/LSP.2016.2603342.

[2] A. Paszke *et al.*, “PyTorch: An Imperative Style, High-Performance Deep Learning Library,” 2019.

[3] Singular Inversions Inc., “FaceGen Modeller.” Singular Inversions Inc., 2020, [Online]. Available: www.FaceGen.com.

[4] E. B. Roesch, L. Tamarit, L. Reveret, D. Grandjean, D. Sander, and K. R. Scherer, “FACSGen: A Tool to Synthesize Emotional Facial Expressions Through Systematic Manipulation of Facial Action Units,” *J. Nonverbal Behav.*, vol. 35, no. 1, pp. 1–16, 2011, doi: 10.1007/s10919-010-0095-9.

[5] ShareX Team, “ShareX.” 2020, [Online]. Available: https://getsharex.com/.

[6] F. Schroff, D. Kalenichenko, and J. Philbin, “FaceNet: A unified embedding for face recognition and clustering,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, pp. 815–823, 2015, doi: 10.1109/CVPR.2015.7298682.

**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.**