

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

Andras Toth
Obuda University
Alba Regia Technical Faculty
sba.andras.toth@gmail.com

Patrik Toth
Obuda University
Alba Regia Technical Faculty
patrikthetoth@gmail.com

Szabolcs Meszaros
Obuda University
Alba Regia Technical Faculty
villam983@gmail.com

Jozsef Halasz
Obuda University,
Alba Regia Technical Faculty
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

I. INTRODUCTION

Smart devices are part of our everyday life now [1]. Most companies try to adapt their products to fit their customers as best as possible [2]. The team's vision is about a smart mirror, that utilizes face recognition to show meaningful information to its users while they perform menial tasks [3].

The team's smart mirror project 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 [4][5].

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.

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.

II. METHODS

A. Creating the face repository

For the creation of the face repository, a software called FaceGen Modeller (demo version) [6] has been chosen. The 4 basic face features (Figure 1) 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 [7]. 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 [7] 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. (Figure 2) All users were created with the male preset of FaceGen Modeller to avoid any potential biases in the face recognition pipeline.

These sets were created for each emotion: neutral with 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 [8]. For the proper file name format, that contained the user's name, the applied emotion, and its intensity the program Bulk Rename Utility [9] was used.

B. 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 (Figure 3).

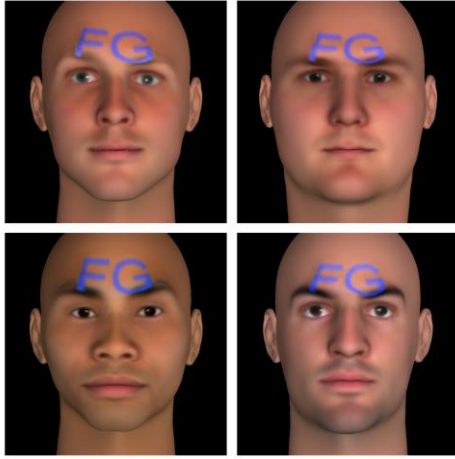


Figure 1. Four representative users in their neutral state
These were created using FaceGen Modeller [4].

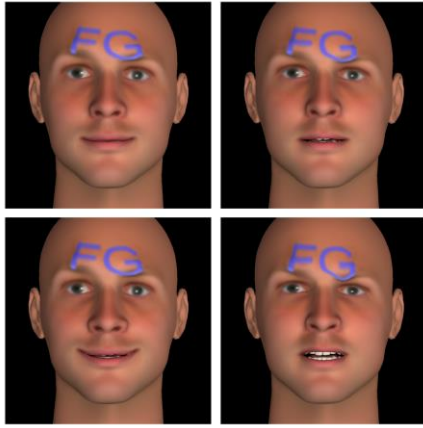


Figure 2. Representative emotions from a specified user
Top left low intensity happiness, lower left high intensity happiness
Top right low intensity anger, lower right high intensity anger
These were created using FaceGen Modeller [4].

C. 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.

D. Evaluating the gathered data

More than 30 thousand data points were gathered during the tests, 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 [10] 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. The project used Python 3 with Seaborn, Numpy and Pandas for data evaluation.

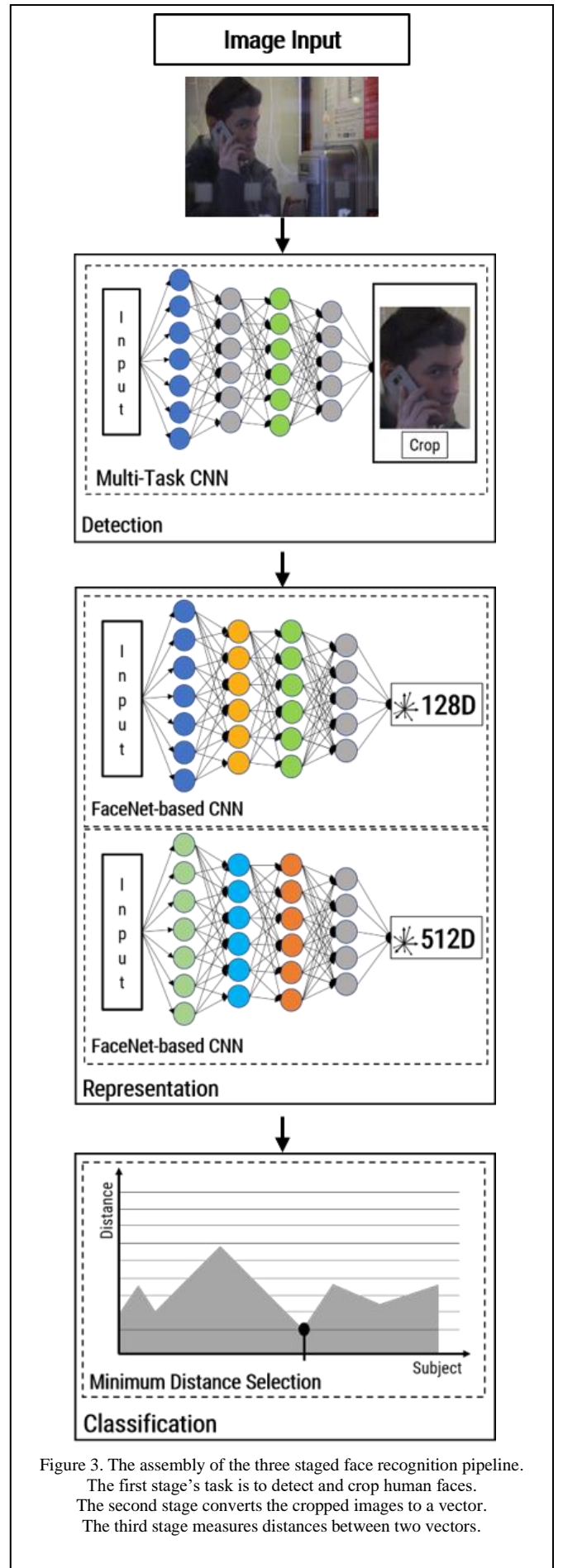


Figure 3. The assembly of the three staged face recognition pipeline.

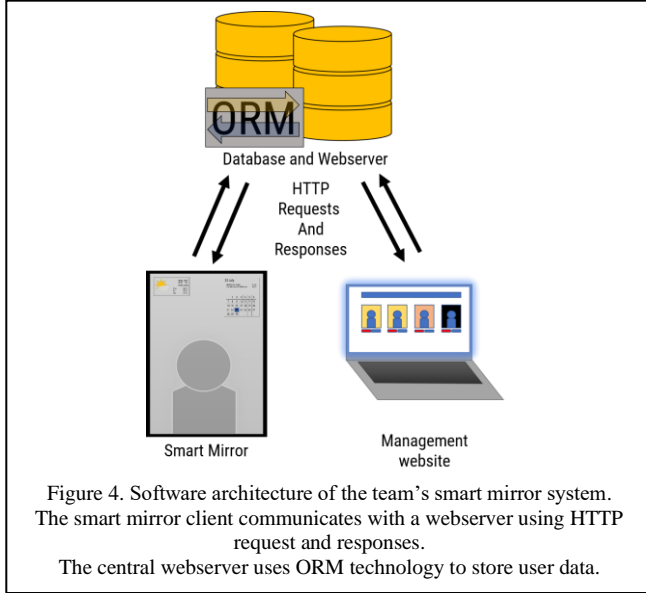
The first stage's task is to detect and crop human faces.

The second stage converts the cropped images to a vector.

The third stage measures distances between two vectors.

E. 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 [11] solution. The given software architectures design can be seen on Figure 4.



III. RESULTS

| Univariate Tests of Significance for Distance Sigma-restricted parameterization Effective hypothesis decomposition | | | | | |
|--|----------|------------------|----------|----------|----------|
| Effect | SS | Degr. of Freedom | MS | F | p |
| Intercept | 10828.53 | 1 | 10828.53 | 485749.1 | 0.000000 |
| (1)Emotion | 12.01 | 1 | 12.01 | 538.7 | 0.000000 |
| (2)Intensity | 22.82 | 1 | 22.82 | 1023.5 | 0.000000 |
| (3)Dimension | 167.37 | 1 | 167.37 | 7508.1 | 0.000000 |
| (4)Within_User | 3793.46 | 1 | 3793.46 | 170168.2 | 0.000000 |
| Emotion*Intensity | 4.00 | 1 | 4.00 | 179.3 | 0.000000 |
| Emotion*Dimension | 0.00 | 1 | 0.00 | 0.0 | 1.000000 |
| Intensity*Dimension | 0.88 | 1 | 0.88 | 39.6 | 0.000000 |
| Emotion*Within_User | 10.20 | 1 | 10.20 | 457.6 | 0.000000 |
| Intensity*Within_User | 18.78 | 1 | 18.78 | 842.5 | 0.000000 |
| Dimension*Within_User | 40.04 | 1 | 40.04 | 1796.0 | 0.000000 |
| Emotion*Intensity*Dimension | 0.32 | 1 | 0.32 | 14.2 | 0.000166 |
| Emotion*Intensity*Within_User | 2.46 | 1 | 2.46 | 110.5 | 0.000000 |
| Emotion*Dimension*Within_User | 0.35 | 1 | 0.35 | 15.6 | 0.000079 |
| Intensity*Dimension*Within_User | 0.75 | 1 | 0.75 | 33.5 | 0.000000 |
| 1*2*3*4 | 0.25 | 1 | 0.25 | 11.3 | 0.000776 |
| Error | 690.17 | 30960 | 0.02 | | |

Table 1. Distance statistics.

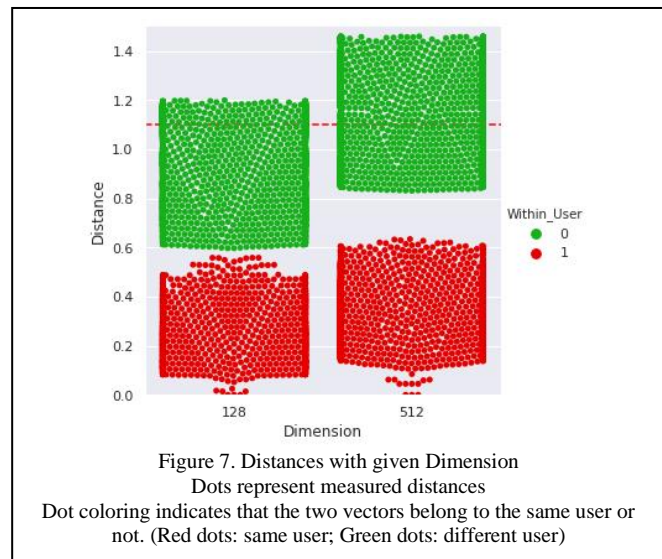
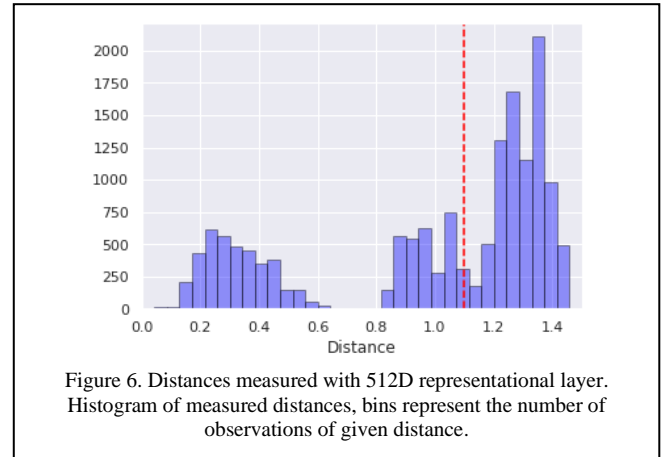
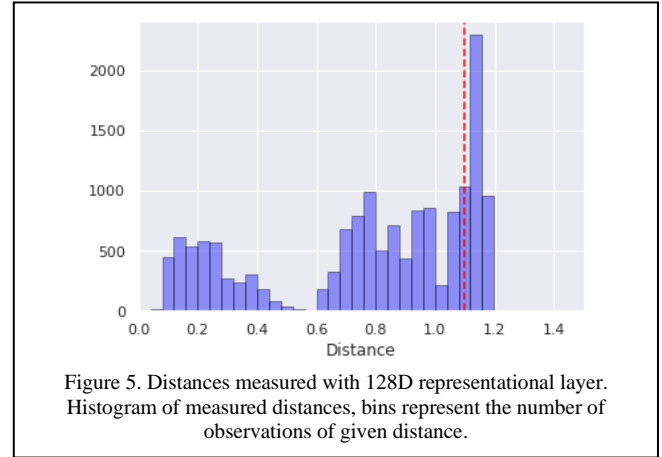
General Linear Model (STATISTICA 7.0) was used in the examined 4-factorial design. $p < 0.05$ represents statistically significant difference between groups. The results of the interactions are also presented.

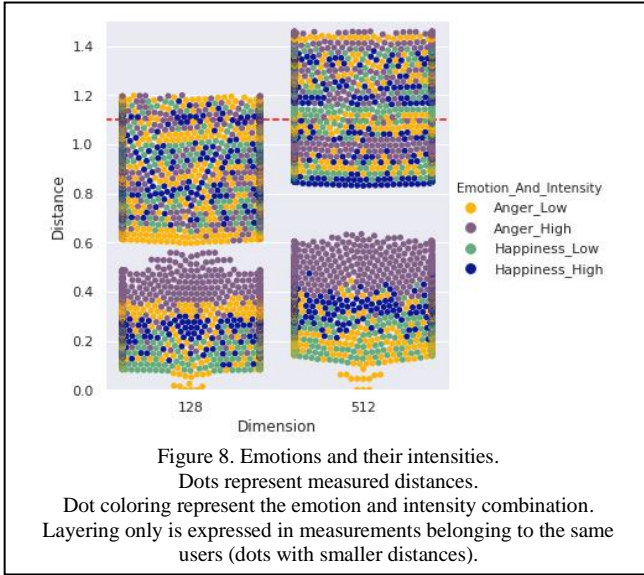
STATISTICA 7.0 was used to establish General Linear Model with 4 factors. Factor 1 was Emotion (Anger versus Happiness), Factor 2 was Intensity (Low versus High), Factor 3 was Dimension (128 versus 512) and finally, Factor 4 was Within User (Within User versus Between User). Interactions between factors were also analyzed. $p < 0.05$ were considered statistically significant (Table 1).

The initial results showed quantifiable difference with the new, 512-dimension representational layer compared to the former 128-dimensional one (Figures 5 and 6).

As one can see in Figures 5 and 6, the 512-dimensional system produces a bigger gap between those distance values that were measured with pictures 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 [10]. On Figure 7, it is clearly visible that the 512-dimension system produces higher distances, this effect is more pronounced in those situations when the compared





vectors belonged to different users. 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. 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 can be seen on Figure 8.

IV. DISCUSSION

The main findings of present study were the following:

A. Dimension findings

We believe the additional dimensions in the representational layer helped our system to represent the users with more granularity. With help of this new representational layer the system is more likely capable of evading user misclassification.

B. Within and between users findings

The two measured systems performed well enough, given that all the within user measurements were much lower than the 1.1 value provided by the FaceNet paper[10]. The 512-dimensional system provided a much bigger gap between the pools of measurements, thus lowering the possibility of misclassification.

C. Emotion findings

As one can see, the measured distances for the emotion anger produced higher distances between the low intensity and high intensity datapoints. This effect is the result of more active Action Units, [7] i.e. more facial features are active at the same time. Generally, the human face uses more muscle movement to represent anger, these muscle groups are called Action Units, these were introduced in the system Facial Action Coding System, by Carl-Herman Hjortsjö [12].

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.

D. Intensity findings

As the provided measurements shows, the data points with higher intensities provided higher distances between the two data points.

V. CONCLUSIONS

The project was successfully finished, it resulted in a working system with a central database and webserver, a client program, and an administrator web interface.

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.

It is self-evident, that those measurements that were performed between two different users resulted in higher distances, in this case emotion and their intensities did not resulted in a marked layering. Within the same users, different emotions resulted in a significant increase of distances, e.g. in high intensity anger, the distances doubled, compared to the neutral faces in 512-dimensional measurements.

The presented study proved, that the emotional qualities have great impact on the measured distances. In addition, it has been proven, that the 512-dimensional system handles these emotional qualities with greater finesse.

ACKNOWLEDGMENT

The research has been supported by the NTP-HHTDK-19 grant. 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

- [1] M. R. Alam, M. B. I. Reaz, and M. A. M. Ali, "A review of smart homes - Past, present, and future," *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 42, no. 6, pp. 1190–1203, 2012, doi: 10.1109/TSMCC.2012.2189204.
- [2] G. Brussenskiy, C. Chiarella, and V. Nagda, "Smart Mirror Sponsored by Central Florida Inpatient Medicine Group 10 Senior Design II-Project Documentation 4-28-2014," 2014.
- [3] P. S. Tóth, A. T. Tóth, and S. Mészáros, "Concept and implementation of a smart mirror," in *14 th International Symposium on Applied Informatics and Related Areas organized in the frame of Hungarian Science Festival 2019 by Óbuda University*, 2019.
- [4] 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.
- [5] A. Paszke *et al.*, "PyTorch: An Imperative Style, High-Performance Deep Learning Library," in *Advances in Neural Information Processing Systems 32 (NIPS 2019)*, 2019.
- [6] Singular Inversions Inc., "FaceGen Modeller." Singular Inversions Inc., 2020, [Online]. Available:

www.FaceGen.com.

- [7] 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.
- [8] ShareX Team, "ShareX." 2020, [Online]. Available: <https://getsharex.com/>.
- [9] J. Willsher, "Bulk Rename Utility." Acebrook Pty Ltd, 2019, [Online]. Available: <https://www.bulkrenameutility.co.uk/>.
- [10] 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.
- [11] R. T. Fielding, "Architectural Styles and the Design of Network-based Software Architectures," 2000.
- [12] C.-H. Hjortsjö, *Man's face and mimic language*. Studentlitteratur, 1970.