Looking for John Doe

Conrado Santos Boeira, Soraia Raupp Musse
School of Technology
PUCRS
Porto Alegre, Rio Grande do Sul
Email: conrado.boeira@acad.pucrs.br;soraia.musse@pucrs.br

Adenir Brito da Silva, Moacir Almeida Simões Junior
Department of Public Safety of Rio Grande do Sul
SSP-RS
Porto Alegre, Rio Grande do Sul
Email: {brito;simoes}@ssp.rs.gov.br

Abstract—**The identification of a face in pictures is the focus of the area of computer vision known as face recognition. Face detection and recognition have become popular in recent years due to many applications and hardware development. There are a wide range of applications including: biometrics, content based image retrieval systems, video processing and entertainment. In this work we present an application of a facial recognition method for practical use in safety applications but it can be easily applied to provide users recognition in entertainment projects. It was result of a collaboration between the Department of Public Safety of Rio Grande do Sul (SSP-RS) and the Virtual Humans Lab from PUCRS University.**

***Keywords*-Face recognition; machine learning; security;**

I. Introduction

Nowadays, computing and visual detection technologies reached a situation in which economical, reliable, and accurate solutions are viable. Embedded applications integrated with cameras in smartphones and other portable devices allow to use face detection and recognition systems. One of the main applications of facial recognition system are bio-metric devices and many studies are under way.

Humans are able to recognize hundreds of faces learning throughout their entire life and identify and recognize easily familiar faces even after years. This skill is very important in humans to provide social, familiar and affecting events. Building a system similar to the human perception system is still an active research area [1].

Many of used approaches in face recognition consists of defining landmarks in the faces and comparing these measurements, as proposed by Wiscott et al. [2]. Recently, new data driven techniques lead to the emergence of neural networks as the main method for recognizing faces. Deep learning [3] are artificial neural networks architectures that learn high-level abstractions in the data, using several learning layers.

A robust and effective facial recognition method can immensely help a police department to identify and arrest fugitives using surveillance cameras in the city. The high level of criminality in Brazil is a major a concern for the Country, therefore the use of face recognition algorithms can be of great assistance. That is why the Department of Public Safety of Rio Grande do Sul (SSP-RS) contact the Virtual Humans Lab with the purpose of developing an application to recognize faces of people of interest in the cameras scattered through the city of Porto Alegre, Brazil. We are using the method proposed by [4] in order to develop an application that can be used with the safety purpose.

This paper is organized as follows: in Section II we present briefly some work in the area, while Section III describes our prototype and how it has been built. Section IV discusses some preliminary obtained results and Section V points out some final considerations and future ideas for development.

II. Related Work

There are many methods used in facial recognition, each one with different gains, characteristics and challenges. In Zafeiriou [5], authors propose to classify algorithms for face recognition into two major categories: *i)* rigid-templates techniques that include: variations of boosting represented by well-known ViolaJones face detection method [6] and algorithms based on Convolutional Neural Networks (CNNs) [7]; and *ii)* techniques that learn and apply a Deformable Parts-based Model (DPM) [8], [9].

Although there are many papers published in the area of facial recognition in photos and videos, we have selected one technology proposed by Geitgey [4] to be used in our prototype. He described a method to detect and compare faces based on the Machine Learning Toolkit DLIB [10]. The first part of the algorithm aims to detect the face in the still image. For that purpose, the HOG method [11] is used as it is effective and fast for this application.

After that, the face must be warped so that the main facial landmarks are located in close positions in all of the face photos. Geitgey [4] proposes the usage of 128 markers in the face. As mentioned by the author the specifics measurements chosen by humans do not have a good result as the machine do not understand human landmarks. The solution that was found was to train a deep convolutional neural network using a triplets training. This training method works by providing to the network 3 photos, two of the same individual and one of a different person. The network then learns to generate 128 measurements for each face that are roughly the same in all photos of the same person.

After the faces encoding is done with 128 features, the only thing left to do is to compare the encodings as the distance between the measurements mirror directly the similarity of the faces. In this case, we used a simple linear Support Vector Machine (SVM) classifier, as proposed by Geitgey [4]. Other algorithms can be used and this is part of the ongoing work we are currently developing.

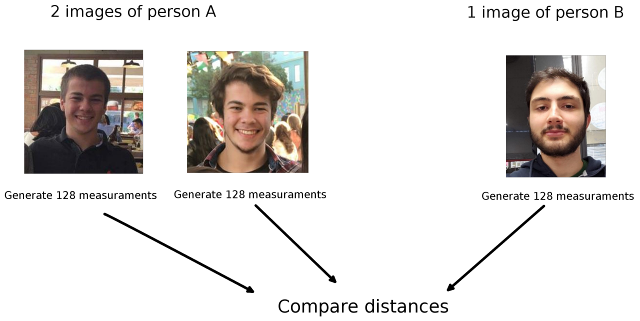
2 images of person A
1 image of person B
Generate 128 measurements
Generate 128 measurements
Generate 128 measurements
Compare distances

Figure 1. Example of the triplets training with 2 images of person A and 1 image of person B so that the network learns to create a similar embedding for images of the same person, as shown in [4].

III. Methodology

The arrangement with SSP (Department of Public Safety of Rio Grande do Sul) consists of exchanging data from public cameras for the development of methods and POCs for facial recognition in photos and videos.

A. Developed Prof-of-Concepts

The goal of the application is to help in the development of an application that would allow policeman to take a frontal photo of someone and compare with a database in order to confirm his/her identity. The model used was the one proposed by Geitgey as previously explained in Section II. We developed two proof-of-concepts (POCs) as following described.

Firstly a Verification POC was developed in order to provide a 1 to 1 comparison. It consists of generating the encoding of two arbitrary photos and returning the distance between them using SVM. Based on computed difference, we can answer if the person is the same or not, using predefined threshold values. It could be useful if a certain person presents an identification but the policeman wants to verify its reliability. In this case, having the person's document we search in the database for a picture of the individual and test the developed POC with the photo taken by the policeman to see if it is the same person or not. For this POC we used the threshold value of 0.5 as it showed 100% of correctness in our tests (as showed in Section IV.

The second POC aims to provide a comparison between 1 and N samples. In this case, we have a dataset containing images and the method generated the encoding for all of them. Then, our POC provides the encoding and comparison of one arbitrary photo (taken by the policeman, for instance) with all N pictures in the dataset, returning the M best matches, where M is a number defined by the user. In this case, the policeman has an application where he/she can look the M candidates (we have used $M = 3$ in our prototype). We also define a threshold value where distances > 0.6 are not presented, meaning that there is not the searched person in the dataset.

Indeed, generating encodings of many pictures can be a slow task. In our tests it takes 35.17 seconds to encode 262 photos (resolution 300 x 400 pixels) in a machine with a Intel Core i7-8700. Images with higher resolution 4200 x 3100 takes 5.43 seconds per picture. However, once encoding is done for a certain image, there is no need of repeating the process unless for new images included in the database. In addition, the time to compare data is not high. Our POC takes approximately 3 seconds to compare 1 x N pictures, where N=254.

B. Applications

Our POCs were included in an application developed by PROCERGS (www.procergs.com.br) in a client-server architecture. Pictures are taken by users and sent to the server which proceeds with the picture encoding, the classification algorithm and returns the computed measured distance. In the case of Verification (1 to 1 comparison), a document information is also sent, in order to find a second picture from the individual identity. In the second POC, the goal is to classify the similarity between the taken picture and the dataset (1 to N comparison). We do not have yet results with real dataset (from the Government) so we do not know the computation time of this part of the application.

IV. Results

This paper presents two developed proof-of-concepts applied to the problems of Face verification and recognition. We used the technology proposed by Geitgey [4]. As mentioned before, we integrated our algorithm with the application developed by PROCERGS. Therefore, tests with large dataset of population of Rio Grande do Sul have not yet been done. So, in this paper we included a preliminary test where we tested the two POCs with some people from our lab and some data data received from SSP.

Firstly, we have a dataset containing 238 photos of 238 different people, as received from SSP. We included 3 pictures from 8 researches from VHLab, resulting in more 24 images in the dataset. All images have been encoded to be prepared for POCs execution. Then we proceed to test the algorithm looking for the 3 best matches in the database to a set of 4th pictures of each one in the lab (total of 8 pictures to test).

The method returned exactly the 3 photos of each one of the 8 researchers as the best matches, from the dataset with a total of 262 images. Figure 2 illustrates the obtained results with 8 researchers from our lab and in all cases the extracted data was correct. This result was accordingly with obtained when tested in Labeled Faces in the Wild benchmark [12] (99.38%) performed by Geitgey [4]. This result is on par with other state-of-the-art recognition software that achieved over 95% accuracy on he LFW dataset [13] [14].

In addition, considering the 1x1 comparison, we observed that acceptable matches are included in the threshold interval of $(0.3, 0.5)$. Anyway for the POCs, the result presented to the user is the best match image (and not a text result). So the user can visually inspect the best match image and decide if it is the found person or not. The usability takes into account the hypothesis that the method finds the most similar person, and quantitatively the distance can indicate if it is the same person or not.

Table I presents some quantitative data obtained in our analyses. In the first column there are the IDs of tested people. In second, third and fourth columns we present the distances from the image used as query and the 3 best matches. In addition, in remaining columns we present the fourth and the last match distances. In all 8 tested cases, the fourth match represents a picture of a different person once we included 3 pictures of each researcher in the dataset [1].

Table I

Table showing the distance measured between the encoded features of each tested picture. The first 3 matches (which are photos of the same person), the fourth match (the first photo of a different person) and the photo with the largest measured distance.

Person	First Match	Second Match	Third Match	Fourth Match	Last Match
1	0.39	0.401	0.442	0.591	0.99
2	0.345	0.515	0.525	0.604	0.919
3	0.345	0.413	0.47	0.514	0.947
4	0.313	0.351	0.436	0.576	0.934
5	0.434	0.453	0.543	0.593	0.932
6	0.482	0.551	0.575	0.6	0.997
7	0.447	0.448	0.497	0.574	0.985
8	0.308	0.427	0.485	0.557	0.895

Although the distances present varied values, we can see that the smaller differences always correspond to the pictures from the same person. As said before, we included 3 pictures in order to show a large variety of data.

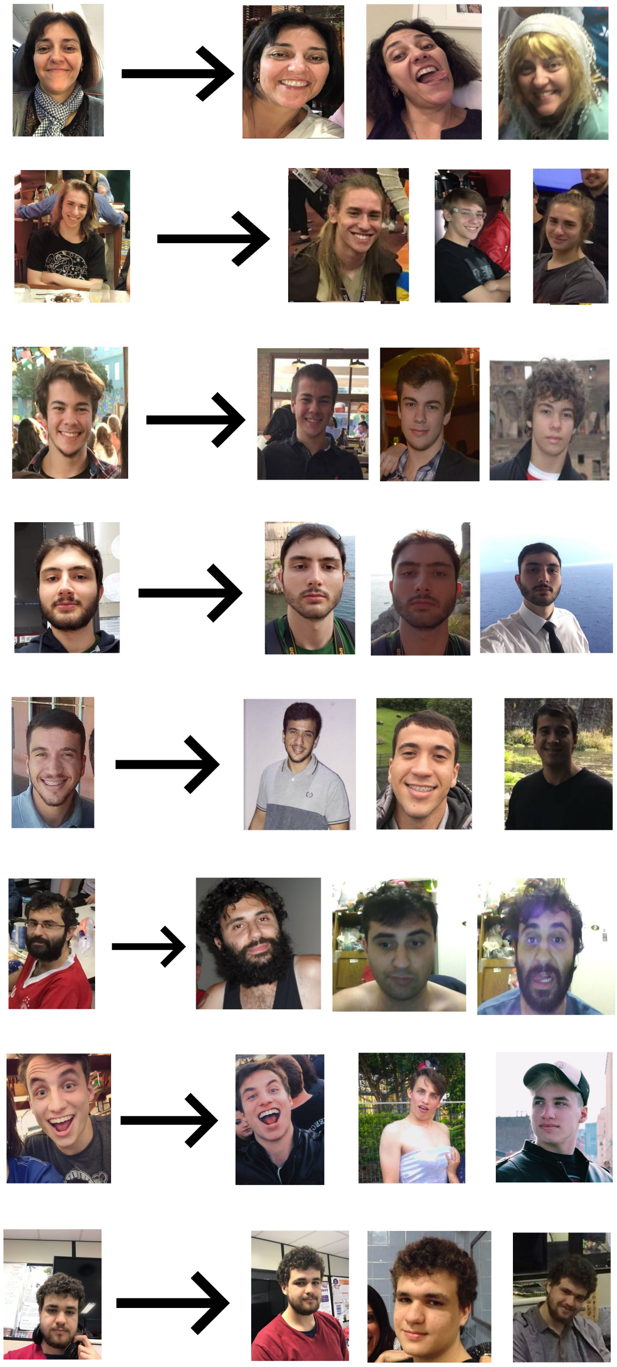

Figure 2. Example of photos used to test the method (before the arrow) and the obtained results (3 best pictures).

V. Conclusion

In this paper, we presented a face recognition application to help users to identify people of their interest. The scope is frontal (or almost frontal) poses. The application is motivated by a partnership between our University and the Police Department in our State. The model uses a technology proposed by Geitgey [4] and involves a pre-trained neural network to find features information in faces. We are currently testing the technology with more than 100.000 pictures and we want to improve the classification method in order to optimize the process of images comparison.

Currently we are working on two aspects: firstly, we want to recognize a person in a video sequence taken in a urban

[1] We are not allowed to show the images included in the SSP dataset, so we just included the images from researchers in our lab.

context. The approach chosen is to train a neural network to identify a specific person face in videos. The network used is the YOLO (You Only Look Once), proposed by Redmon and Farhadi [15]. The second aspect is to use the facial recognition framework to serve as interface to entertainment applications. Our idea is to recognize the user and in a future we want to detect facial emotion expression in order to change the game/application narrative. This could be implemented with the Microsoft Kinect sensors which have already been used for successful recognition of hand gestures for example [16].

Acknowledgment

The authors would like to thank the Department of Public Safety of Rio Grande do Sul (SSP-RS) and PROCERGS for the collaboration in this work. We would also like to thank all members of VHLab and especially those who allowed us to use their photos on performed tests. In addition, we want to thank CNPq and CAPES for partially funding this work.

References

[1] A. zdil and M. M. zbilen, "A survey on comparison of face recognition algorithms," in *2014 IEEE 8th International Conference on Application of Information and Communication Technologies (AICT)*, Oct 2014, pp. 1–3.

[2] L. Wiskott, J.-M. Fellous, N. Krüger, and C. Von Der Malsburg, "Face recognition by elastic bunch graph matching," in *International Conference on Computer Analysis of Images and Patterns*. Springer, 1997, pp. 456–463.

[3] J. Schmidhuber, "Deep learning in neural networks," *Neural Netw.*, vol. 61, no. C, pp. 85–117, Jan. 2015. [Online]. Available: http://dx.doi.org/10.1016/j.neunet.2014.09.003

[4] A. Geitgey, "Machine learning is fun! part 4: Modern face recognition with deep learning," *Medium. https://medium. com/@ ageitgey/machine-learning-is-fun-part-4-modern-face-recognition-with-deep-learning-c3cffc121d78*, 2016.

[5] S. Zafeiriou, C. Zhang, and Z. Zhang, "A survey on face detection in the wild: Past, present and future," *Computer Vision and Image Understanding*, vol. 138, pp. 1 – 24, 2015.

[6] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," 2001, pp. 511–518.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, ser. NIPS'12. USA: Curran Associates Inc., 2012, pp. 1097–1105. [Online]. Available: http://dl.acm.org/citation.cfm?id=2999134.2999257

[8] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1627–1645, Sep. 2010. [Online]. Available: http://dx.doi.org/10.1109/TPAMI.2009.167

[9] P. F. Felzenszwalb and D. P. Huttenlocher, "Pictorial structures for object recognition," *International Journal of Computer Vision*, vol. 61, no. 1, pp. 55–79, Jan 2005. [Online]. Available: https://doi.org/10.1023/B:VISI.0000042934.15159.49

[10] D. E. King, "Dlib-ml: A machine learning toolkit," *Journal of Machine Learning Research*, vol. 10, pp. 1755–1758, 2009.

[11] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1. IEEE, 2005, pp. 886–893.

[12] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," University of Massachusetts, Amherst, Tech. Rep. 07-49, October 2007.

[13] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[14] Y. Sun, D. Liang, X. Wang, and X. Tang, "Deepid3: Face recognition with very deep neural networks," *arXiv preprint arXiv:1502.00873*, 2015.

[15] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv*, 2018.

[16] Z. Ren, J. Yuan, J. Meng, and Z. Zhang, "Robust part-based hand gesture recognition using kinect sensor," *IEEE transactions on multimedia*, vol. 15, no. 5, pp. 1110–1120, 2013.