Research Journal of Applied Sciences 13 (1): 47-53, 2018
ISSN: 1815-932X

Face Recognition Access Control System using Convolutional Neural Networks

Paula Catalina Useche M., Javier O. Pinzo Arenas and
Robinson Jimeez Moreno
Faculty of Engineering, Nueva Granada Military University,
Bogota, Colombia

Abstract: The following study presents the development of a face access control system using convolutional neural networks where all the faces of a scene are recognized and classified to allow or deny an individual access according to their facial characteristics. The system allows the access of two specific people after individually recognizing them for 5 sec in a video sequence and prevents access when the presence of an outside person, located within the area of vision is detected. The program uses a face detection system by haar classifiers, a point tracking system by KLT Algorithm (Kanade-Lucas-Tomasi) and a classification technique by convolutional neural networks where accuracy percentages are reached above 96%.

Key words: Face recognition, computer vision, convolutional neural network, haar classifier, KLT Algorithm

INTRODUCTION

Developing a face detection and recognition system is not a simple task as this system must deal with multiple external factors that affect the recognition and correct classification of each individual such as variations of lighting, facial expressions, occlusions and facial poses as indicatedv by Ahmad *et al.* (2014). In addition, it demands a high level of accuracy, since, it is a biometric system whose applications are focused on access control systems, criminal identification or banking controls as explained by Dunstone and Yager (2008) and Arun *et al.* (2014).

Faced with the need to detect and identify people by intelligent computational means, numerous solutions have been proposed based on artificial intelligence techniques for the learning and classification of each face such as those presented.

By Guevara *et al.* (2008), cascade classifiers and haar base filters were used for the detection of faces, achieving a 100% success while by Ahmad *et al.* (2014) the detection was performed by means of Convolutional Neural Networks (CNN) with 99.5% accuracy using as an training database an AR base created by Aleix Martinez and Robert Benavente which contains 3276 color images of 126 subjects where 70 of them are men and each subject has 26 frontal images of the face with different facial expressions, illumination and occlusions. In the same way by Lawrence *et al.* (1997), Sharma *et al.* (2016) and Li *et al.* (2015) face recognition algorithms were developed by convolutional neural networks, achieving the identification of faces even with positions completely in profile or with changes of expression.

Within the area of the classification of faces, there are works such as those developed by Yin and Liu (2017) and Lyons *et al.* (1999) where the first estimates and classifies faces according to the illumination, facial expression and pose of each one while the second classifies them according to gender, race and facial expression.

In the field of biometrics, robust personal identification systems have been developed through automatic classifiers that combine voice, sound and face recognition information for the identification of people such as the one presented by Fox *et al.* (2007) where 75% accuracy is achieved with face recognition alone and 89.9% accuracy with the fusion of the three methods. In the case by Filkovic *et al.* (2016) and Konen *et al.* (1995) a system of re-identification of people was developed which consists of tracking a desired person through multiple cameras, keeping hidden their true identity where for the first case were used deep learning techniques such as the triplet network architecture while for the second case convolutional neural networks were used.

Finally by Konen *et al.* (1995), a face recognition access control system was performed in which the image acquisition, face location and user identification process

Corresponding Author: Paula Catalina Useche M., Faculty of Engineering, Nueva Granada Military University, Bogota, Colombia

were performed in approximately 3.5 sec with an accuracy of 96%, using a console of semi-automatic acquisition of images to make the face recognition.

The following study presents the development of a face recognition access control system where faces are detected and tracked in a video and the identification of users in real time using convolutional neural networks and achieving a percentage of accuracy >96%. The system allows to recognize more than one face at a time and to indicate for each user the level of access they have (allowed or denied), leading to the generation of a new technique of face recognition and identification of people in real time through neural convolutional networks, haar classifiers and KLT Algorithm where the system security level increases by avoiding access of any user when a person without access is within the program's area of vision.

MATERIALS AND METHODS

Access control system: The developed access control system is implemented for face recognition in video and face tracking in real time where all the faces recognized are classified in a scene by convolutional neural networks.

When one of the recognized faces corresponds to any of the two persons who have allowed access to the system, a count of the person's permanency time in the video is started as shown in Fig. 1a and when the 5 sec of recognition are reached, access to the user is allowed as shown in Fig. 1b.

On the other hand, when the system recognizes a face that does not belong to any of the main users and another one that belongs, the count of the user with access is eliminated and is not executed until the person without access leaves the range of vision as illustrated in Fig. 2 where the main or authorized users (Javier and Paula) are shown at the top of the image.

The category labeling (Javier, Paula, others), recognition level (from 0-100%), time taken by user Javier or Paula in the video (time in seconds) and access status (allowed or denied) are indicated in the upper tab of each box. Following are the steps in which the operation of the access control program is divided and a brief explanation of each of them and their purpose in the system is given.

Step 1; face detection and tracking: The first step of the program is to detect and follow all the faces of a scene where tracking is done in each frame and the detection is generated every 10 frames in order to find new users in the video and ensure an appropriate follow-up of each person as discussed by Lisin (2016). When a face is detected, the program generates a box on it and crops the image in order to a square image is always fed into the network, focused only on the face of the person as shown at the top of Fig. 2.

For the face detection process, the Viola-Jones Algorithm presented by Viola and Jones (2004) is used

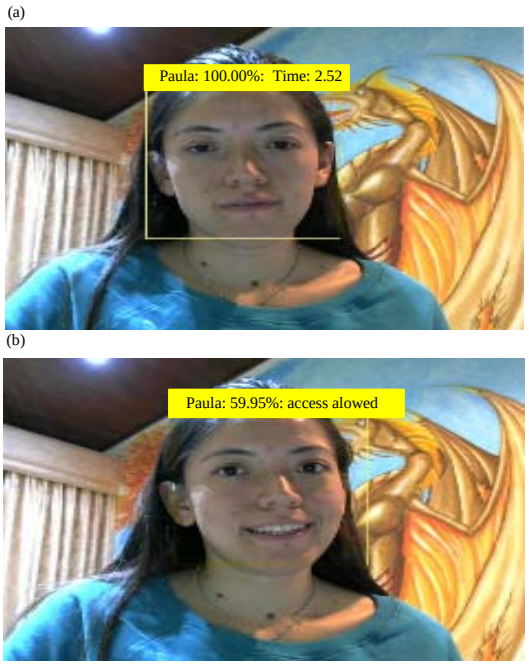

Fig. 1: Face recognition; a) during counting and b) with permitted access

Fig. 2: Users with and without access

Fig. 3: Tracking points

Fig. 4: Face positions trained from a frontal position to a rotation of approximately 45°C left and right

which consists of evaluating each image through its equivalent integral image where each pixel of the integral image is equal to the sum of all the pixels of the original image from the pixel (0, 0) to the location (x, y) of the pixel of the desired integral image as in Eq. 1 where ImI is the integral image and ImO the original image:

$$\mathrm{Iml}(x,y) = \sum_{x'\le x, y'\le y} \mathrm{Im0}(x',y') \qquad (1)$$

Then, it is used an AdaBoost learning algorithm, proposed by Freund and Schapire (1997) which uses the Littlestone and Warmuth weighting update rule, detailed by Littlestone and Warmuth (1994) to generate a learning reinforcement algorithm that does not require any previous knowledge of the classification algorithm. Finally, several weak classifiers in cascade are combined as discussed by Guevara *et al.* (2008) which allow to quickly discard the background to focus classification only the characteristics of the face.

For the tracking process, the Kanade-Lucas-Tomasi (KLT) Algorithm is used which calculates the displacement of the points of interest that have been previously placed on the face as shown in Fig. 3 using the Newton method to minimize the sum of square distances as explained by Sinha *et al.* (2011), Tomasi and Kanade (1991).

Step 2; face recognition: The face recognition of the access control system was divided into three categories of classification where two of them correspond to the users who have access to the system and the third includes all those that are not allowed access.

Initially, 275 images were taken per category, each with faces in different positions as shown in Fig. 4 and all under different intensities of light, background and facial expressions. For the category with denied access were taken photos of several people with different skin tones,

Fig. 5: Data augmentation

face shapes, eye, nose and mouth proportions, amount of and others with similar characteristics to users with access. After obtaining an initial base of 275 images per category an augmentation of the database was generated by three changes: one in the light intensity in each image (brighter or darker), another in the value of each RGB component (increase of each component or two at the same time) and the last one in the addition of random noise type Speckle and Salt and Pepper, explained by MathWorks (2016), generating copies of the original image as shown in Fig. 5 and obtaining a final database of 5225 images per

category. The images in the first row, from left to right, correspond to the original image with brighter and darker light intensity, respectively. The intermediate images present changes in their RGB components and those in the last row correspond from left to right to Salt and Pepper noise and Speckle noise.

The architecture designed for a CNN focused on face recognition is presented in Fig. 6 where a 128×128 pixel color image is entered with the face found in the detection and tracking phase and the user's classification is obtained as a result along with a recognition percentage. The parameters of each layer of the network are specified in Table 1.

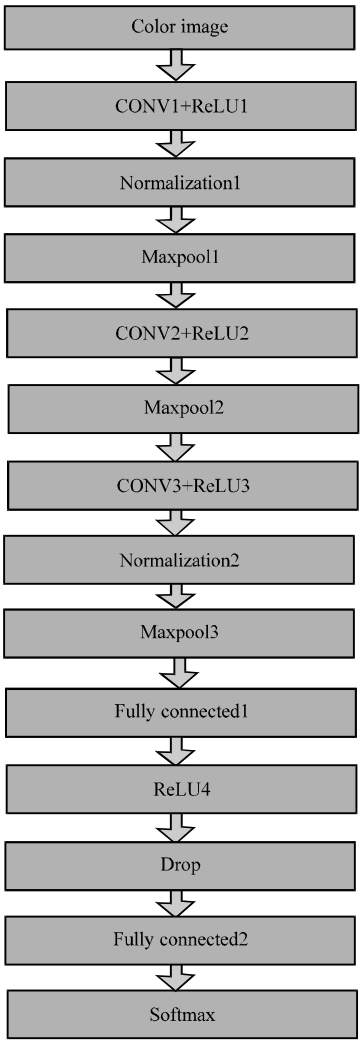

Fig. 6: Network architecture

The architecture employed achieved 99.3% accuracy over the original test images, i.e., without the changes shown in Fig. 5 with a confusion matrix as shown in Table 2, where each column of the confusion matrix represents the actual category to which each test image belongs and the rows represent the classification made by the network for each image which means that of 50 images of the authorized user "Javier", only one was classified as "Others" (Table 3).

In the same way that, the original database was augmented an increase of the test database was made, obtaining 520 images per category where a high percentage of accuracy of 97.9% was achieved where the results are presented in the confusion matrix in Table 2 which shows an erroneous classification of 19 "Javier" images as "Others" and 4 as "Paula", 5 images of "Others" as "Javier" and 4 images of "Paula" as "Others". The network used for the program with 97.9% accuracy was called "Final Network.

RESULTS AND DISCUSSION

CNN's first trainings were done with 275 images per category and a wide range of face positions as illustrated in Fig. 7 in order to recognize faces from any perspective. However, although, the program was able to correctly classify faces located in profile, it presented problems to differentiate users very similar to each other like those of Fig. 8.

Faced with this problem, tests were performed with changes in the network architecture, reduction in the range of face positions and training data augmentation to try to increase the percentage of accuracy of CNN where

Table 1: Network parameters

Symbols	Layers	Filter size (pixels)	No. of filters	Stride	Padding	Percentage
CONV1+ReLU1	Convolution+Rectified linear units	12×12	20	1	0	-
Maxpool1	Maxpooling	3×3	-	2	0	-
CONV2+ReLU2	Convolution+Rectified linear units	7×7	30	1	0	-
Maxpool2	Maxpooling	2×2	-	2	0	-
CONV3+ReLU3	Convolution+Rectified linear units	5×5	40	1	0	-
Maxpool3	Maxpooling	2×2	-	2	0	-
DROP	Dropout	-	-	-	-	50

Table 2: Confusion matrix for a test database with 50 images per category

Real class			
Classification	Javier	Others	Paula
Javier	49	0	0
Others	1	50	0
Paula	0	0	50

Accuracy; 99.3%

Table 3: Confusion matrix for a test database with 520 images per category

Real class			
Classification	Javier	Others	Paula
Javier	497	5	0
Others	19	515	40
Paula	40	0	516

Accuracy; 97.9%

the first and last network of Table 4 were trained only with frontal images like those of Fig. 4 while the others were trained with frontal and profile images as those of Fig. 7 and all were tested with the test images without augmentation with frontal and profile positions.

The first, second, third and last network of Table 4 have the same architecture as in Fig. 6, so that, it is possible to compare the behavior of the network by varying the number of filters per convolution and using different databases. On the other hand, the fourth network used an architecture similar to that of Fig. 6 but with an additional convolution before standardization 2 and the last network used three consecutive convolutions before standardization 1 and a single convolution between the following pooling layers.

Networks two to five showed a greater tendency to confuse users of the "Others" category with the "Javier" category in addition, they required more training time than the original network and presented a lower percentage of accuracy with respect to the first and last network, so, the initial architecture was maintained with only three layers of convolution and two fully connected with <40 filters per convolution and with the database of face positions with rotations <45° for the implementation of the control system.

It was established as a priority of the program that the largest number of users without access would be classified correctly as "Others", above the accuracy in the classification of the main users in order to avoid that people similar to the users with access can easily enter the system. The augmentation in the original database and the reduction of face positions led to an improvement in the classification of users as can be seen in the accuracy percentages in Table 4 and allowed to correctly classify two similar individuals but of different categories as shown in Fig. 8.

Fig. 7: Wide range of face positions from a front position to a rotation of approximately 90°C left and right

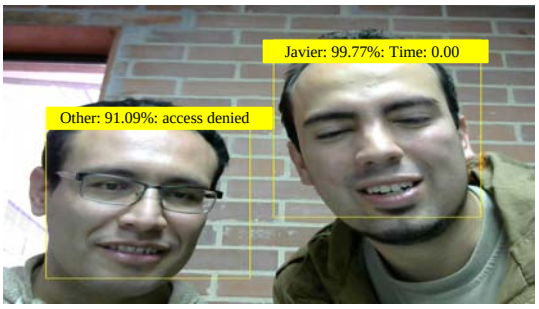

Fig. 8: Discrimination between similar faces

On the other hand in Fig. 9, it can be seen faces of different users that correspond to the category of "Others" where those in Fig. 9a belong to the training database while those in Fig. 9b are not found in the databases. Despite the difference, both were correctly classified as "Others" with recognition percentages above 80%. These tests demonstrate the ability of the final network to generalize the category of users without access and extend it to more people, overcoming external factors such as changes of light or shadows. On the other hand in Fig. 10, it is possible to observe the stronger activations of the CONV1+RELU1 layer and with them to visualize the most relevant characteristics that the network emphasizes of each user to generate the classification where the whitest areas are the ones that generate a greater activation.

As can be seen, the first layers of the network extract relevant characteristics of different parts of the face for

Table 4: Accuracy percentages based on changes in the network or database architecture

Changes	Architecture	Accuracy (%)	EPOCHS	Training time	Training database by categories
Original (Front)	3CONV/3Maxpool	92.7	100	11.33 min	275
Three layers of convolution	3CONV/3Maxpool	88.7	100	16.54 min	275
More convolution filters	3CONV/3Maxpool	88.7	100	17.37 min	
	Between 30 and 80 filters per convolution			275	
Four layers of convolution	4 CONV/2Maxpool	81.3	100	46.39 min	275
Five layers of convolution	5 CONV/3Maxpool	88.0	100	54.54 min	275
Final network (Front)	3 CONV/3Maxpool	99.3	100	4 h	5225

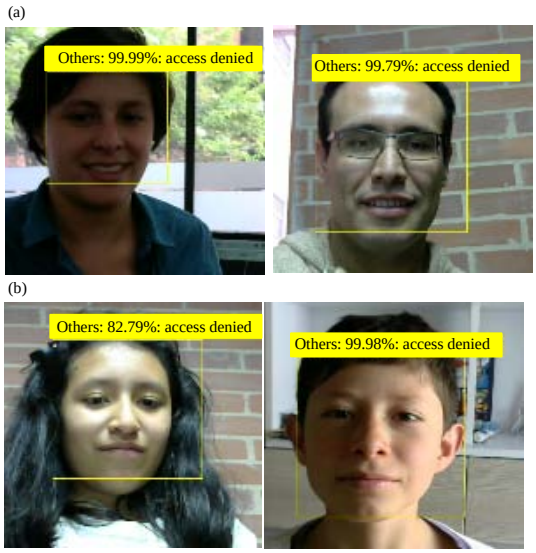

Fig. 9: Classification of users without access; a) In the training database and b) Outside the training database

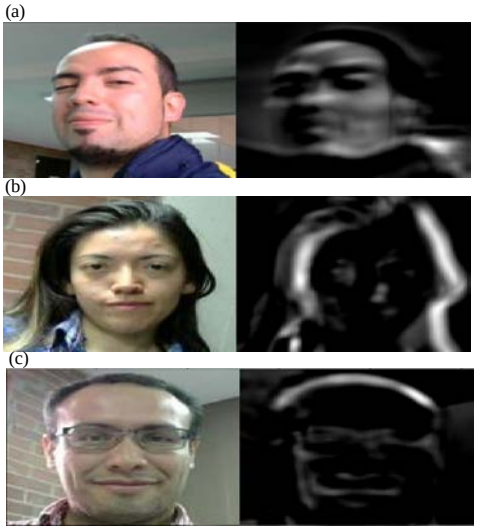

Fig. 10: Activations of the network for a) Javier; b) Paula and c) An user without access

each user where for the case of Fig. 10a the forehead, nose and upper brows are highlighted, Fig. 10b it highlights the hair and part of the nose and eyes whereas for Fig. 10c, the most active areas are the top of the head, the line of the mouth and the glasses.

When the network is able to recognize features such as those in Fig. 10, it means that the convolution filters were properly trained and that the CNN is focused on easy recognition and not in the background or other non-user aspects.

CONCLUSION

The architecture of the convolutional neural network has the ability to facilitate or hinder the learning of training images as was observed in Table 4, however, the size of the training database also generates a strong influence on the recognition capacity of the database, leading to the need to use large databases to achieve better results.

CNN has the ability to recognize faces from any perspective, however by requiring a more detailed classification of users, it is necessary to restrict face positions in order to root the training of the network to the details of each face and thus make it possible to differentiate each category more easily.

An augmentation of the database as presented in Fig. 5 increases the robustness of the network by simulating different lighting conditions for the same image which prevents the network from differentiating the categories according to light intensity, backlighting situations or very marked shadows.

Adding convolution layers to the architecture of a CNN does not imply in all cases, an improvement to the accuracy of the network, since, an excess of filters can lead to an unnecessary increase of the training time without representing a greater facility for the learning of each category as it was observed when training a network with five layers of convolution where the accuracy lowered near 1% with respect to the network with 3 layers of convolution and the training time increased little more than three times.

Increasing the number of convolution filters also does not equate to an improvement in the accuracy of the network as was observed when comparing networks two and three of Table 4 but it does influence the training time, showing an increase of about 1 min of the third network with respect to the second one.

The access control system developed is able to detect and deny the entry of users without access to the system even though they have not been previously trained in the network, however, it may fail to recognize faces with great similarity to each other, so that, the system could be supplemented with other personal information such as other biometric access systems or a password to increase the level of security.

ACKNOWLEDGEMENTS

The researchers are grateful to the Nueva Granada Military University which through its Vice Chancellor for Research, Finances the present Project with code IMP-ING-2290 and titled "Prototype of Robot Assistance for Surgery" from which the present study is derived.

REFERENCES

Ahmad R.S., K.H. Mohamad, S.S. Liew and R. Bakhteri, 2014. Convolutional neural network for face recognition with pose and illumination variation. Intl. J. Eng. Technol., 6: 44-57.

Arun, S., V. Joshua, A. Pillai and A. Nageswaran, 2014. Fingerprints age determination in different methods using image processing. Intl. J. Soft Comput., 9: 85-87.

Dunstone, T. and N. Yager, 2008. Biometric System and Data Analysis: Design, Evaluation and Data Mining. Springer,Berlin,Germany,ISBN-13:978-387-77625-5, Pages: 229.

Filkovic, I., Z. Kalafatic and T. Hrkac, 2016. Deep metric learning for person Re-identification and De-identification. Proceedings of the 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO'16), May 30-June 3, 2016, IEEE, Opatija, Croatia, ISBN:978-1-5090-2543-5, pp: 1360-1364.

Fox, N.A., R. Gross, J.F. Cohn and R.B. Reilly, 2007. Robust biometric person identification using automatic classifier fusion of speech, mouth and face experts. IEEE. Trans. Multimedia, 9: 701-714.

Freund, Y. and R.E. Schapire, 1997. A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci., 55: 119-139.

Guevara, M.L., J.D. Echeverry and W.A. Uruena, 2008. [Detection of faces in digital images using cascade classifiers (In Spanish)]. Sci. Tech., 38: 1-6.

Konen, W. and E. Schulze-Kruger, 1995. ZN-Face: A system for access control using automated face recognition. Proceedings of the 1995 International Workshop on Automated Face and Gesture-Recognition, June 26-28, 1995, University of Zurich, Zurich, Switzerland, pp: 18-23.

Lawrence, S., C.L. Giles, A.C. Tsoi and A.D. Back, 1997. Face recognition: A convolutional neural-network approach. IEEE. Trans. Neural Networks, 8: 98-113.

Li, H., Z. Lin, X. Shen, J. Brandt and G. Hua, 2015. A convolutional neural network cascade for face detection. IEEE. Conf. Comput. Vision Pattern Recognit., 2015: 5325-5334.

Lisin, D., 2016. Detect and track multiple faces. MathWorks, Natick, Massachusetts, USA. https://www.mathworks.com/matlabcentral/fileexchange/47105-detect-and-track-multiple-faces

Littlestone, N. and M.K. Warmuth, 1994. The weighted majority algorithm. Inf. Comput., 108: 212-261.

Lyons, M.J., J. Budynek and S. Akamatsu, 1999. Automatic classification of single facial images. IEEE Trans. Pattern Anal. Mach. Intell., 21: 1357-1362.

MathWorks, 2016. Detect and track multiple faces. MathWorks, Natick, Massachusetts, USA.

Sharma, S., K. Shanmugasundaram and S.K. Ramasamy, 2016. FAREC-CNN based efficient face recognition technique using Dlib. Proceedings of the 2016 International Conference on Advanced Communication Control and Computing Technologies (ICACCCT'16), May 25-27, 2016, IEEE, Ramanathapuram,India,ISBN:978-1-4673-9546-5,pp: 192-195.

Sinha, S.N., J.M. Frahm, M. Pollefeys and Y. Genc, 2011. Feature tracking and matching in video using programmable graphics hardware. Mach. Vision Appl., 22: 207-217.

Tomasi, C. and T. Kanade, 1991. Detection and tracking of point features. Technical Report CMU-CS-91-132, Carnegie Mellon University School of Computer Science Press, Pittsburgh, PA., USA.

Viola, P. and M.J. Jones, 2004. Robust real-time face detection. Int. J. Comput. Vision, 57: 137-154.

Yin, X. and X. Liu, 2017. Multi-task convolutional neural network for face recognition. Comput. Vision Pattern Recogni., 1: 1-12.