Access Control to Security Areas Based on Facial Classification

Aitor Moreno Fdz. de Leceta[1,⋆] and Mariano Rincón[2]

[1] Intelligent Systems of Control and Management Department, Ibermática
Parque Tecnológico de Álava, c/Leonardo Da Vinci, 9 2ª Planta - Edificio E-5
01510 Miñano, Álava, Spain
ai.moreno@ibermatica.com

[2] Departamento de Inteligencia Artificial, Escuela Técnica Superior de Ingeniería
Informática, Universidad Nacional de Educación a Distancia,
c/ Juan del Rosal 16, 28040 Madrid, Spain

Abstract. The methods of biometric access control are currently booming due to increased security checks at business and organizational areas. Belong to this area applications based on fingerprints and iris of the eye, among others. However, although there are many papers related to facial recognition, in fact it is difficult to apply to real-world applications because of variations in lighting, position and changing expressions and appearance. In addition, systems proposed in the laboratory do not usually contain a large volume of samples, or the test variations not may be used in applications in real environments. Works include the issue of recognition of the individual, but not the access control based only on facial detect, although there are applications that combine cards with facial recognition, working more on the verification that identification. This paper proposes a robust system of classification based on a multilayer neural network, whose input will be samples of facial photographs with different variations of lighting, position and even time, with a volume of samples that simulates a real environment. Output is not the recognition of the individual, but the class to which it belongs. Through the experiments, it is demonstrated that this relatively simple structure is enough to select the main characteristics of the individuals, and, in the same process, enable the network to correctly classify individuals before entering the restricted area.

1 Introduction

Face recognition is one of the problems that most challenges are proposing to technical computing nowadays, especially in security systems. Face is the most frequently used way to identify another individual. For this, the brain begins to establish the physical aspects of a face, and then determines whether these factions are known or not, and finally gives a name to what he sees [11]. This process seems so simple for us, but it can be very difficult for a machine. Therefore,

⋆ Corresponding author.

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 254–263, 2009.

before developing a biometric system, scientists have been dedicated to analyze the mental processes of facial recognition. So, they have found, for example, that there is a region in the back of the brain that responds preferentially when faces are detected in contrast with other parts of the anatomy or objects [6]. There is also evidence that the face gesture interpretation processes are independent of face identification [13], so a good system for facial recognition should be invariant to facial expression. A final challenge to overcome is the process speed: systems must operate in real time, with a very fast response time, and with the possibility of learning from failures.

This paper presents a part of a prototype for a logistics company access control, in which a winch or a door is connected to a camcorder, detecting a person who is going to enter into the lathe and approve or deny his access. In the case of denied access, an operator will record the identity of the individual, reclassified to next visit, if competent. The information is contained in the weights of a neural network. The use of the network can discriminate whether a record belongs to the set of authorized people or not, but can not retrieve the record. With this restriction, the network achieves a much higher ratio of capacity of discrimination compared to other models. The system requirements are an acceptable response time to a particular discrimination, easy deployment, and a robust and flexible learning process with unknown individuals and misclassification errors. The proposed solution offers some advantages over other methods of access control as a cheap solution, since it requires no expensive hardware and a non-intrusive architecture, with the advantage that the user should not do anything to access into the control area.

2 Prototype Description

The solution developed in this project is based on a grayscale image as input, linked to a classifier built on a multilayer neural network with backpropagation. The novel aspects incorporated are:

- **Use of neural networks for facial classification, not only as final classifier but also as feature detector:** As is clear from the state of the art [10], neural networks have been used in facial recognition systems to classify the characteristics of an individual. This characteristics or features are previously obtained by another processes and usually reduced with some feature reduction methods. This paper demonstrates that a single neural network is sufficient to make a correct classification of images of individuals without a prior extraction of key features. In other words, the network is capable of extracting intrinsic features before making the final classification.
- **Classification of individuals, without identifying them individually**: Another new aspect of this work is the classification of individuals into groups, forgetting the identification of each individual, looking for a technique that combines efficiency, adaptation, very short response times even with general purpose hardware (cameras and computers) and easy configuration.

For this, we need a classifier with the following characteristics:

- To be able to learn from their mistakes.
- Response times are acceptable to a particular post.
- Implementation must be simple, because simplicity implies robustness and flexibility in processing time.

The use of grayscale images as input data has been referred to numerous times in machine vision [1], especially if the processing is done by means of neural networks [2]. The preprocessing phase in this case consists of the location of the face and the subsequent normalization of the image. In this paper, the normalization of the images consists of reducing the face region to 64x64 pixels with 256 levels.

Thereafter, a multilayer neural network with backpropagation is used. So we convert the photograph of the face into a matrix of 64x64 bytes, i.e. 4096 elements. It would be computationally very slow to train a network with one output for each individual differently. But despite that identification is not easy, the classification is cheaper, so that we can linearly discriminate individuals into two classes, we classify the sample "in" or "out" of a given set. The output is a layer built by two output neurons, indicating a degree of membership of each sample into the different classes. So, when the first output neuron is activated, the individual will belong to the set, and when the second is activated, the individual does not belong to this set. To develop this simple classification neural network with two output neurons the structure of the intermediate layer should be simpler than that proposed by the work of Cottrell and Fleming [3] (with 80 units in the hidden layer). So we took an intermediate layer of 10 neurons.

In summary, our first prototype is a network of three layers, the first of 4096 neurons, interconnected "completely" with a second layer of 10 neurons, which in turn is interconnected "completely" with third layer of 2 neurons in output.

3 Intermediate Tests and Results

For verification tests, two databases created for this purpose and documented in the literature have been used:

- Images in PGM format from the corpus *The UMIST Face Database* [14]
- Images in JPEG format in the *Feret* database [12].

3.1 Example 1: The UMIST Face Database

Images are taken from the corpus " The UMIST Face Database " [14] with the following characteristics:

- Background: The fund is not always the same tone
- Scale: there is variation.
- Point of view: Different angles of the same person.
- Position of the face in the picture: Different angles of the same person
- Light: Different Illuminations
- Expression: considerable variation

Training set	Pictures (P) / Individuals	Pictures (NP) / Individuals
A	26 / 5 Individuals	28 / 10 Individuals
B	26 / 5 Individuals	45 / 10 Individuals
C	26 / 5 Individuals	102 / 10 Individuals
D	93 / 5 Individuals	28 / 10 Individuals
E	93 / 5 Individuals	45 / 10Individuals
F	93 / 5 Individuals	102 / 10 Individuals

Fig. 1. Different combinations for each of the various trainings offered

Training type................	A (Test1)	B (Test2)	C (Test3)	D (Test4)	E (Test5)	F (Test6)
AUC	0,949	0,890	0,809	0,980	0,941	0,979

(a)

	D (Test 4)		F (Test 6)	
	Hits	%	Hits	%
P (20 samples)	19	95%	20	100%
D (20 samples)	18	90%	14	70%
NP (20 samples)	20	100%	20	100%
Total Hits %	57	95,00%	54	90,00%

(b)

Fig. 2. Results for networks trained with trainind sets A-F. a) AUC performance; b) Detail of the best two networks according to AUC performance.

In order to analyze the influence of the number of samples in the training set, different sets with an unbalanced number of pictures have been defined (see Figure 1). The validation set consists of 5 individuals of type P (Authorized) and 10 other samples of individuals of type NP (not authorized). Finally, the evaluation set consists of 20 samples from each of the three types of individuals: P, NP, and D (unknown).

It is presented the same testing set to each network, yielding the results summarized in Figure 2. To evaluate the classification success of each network, we use ROC curves, which are obtained by evaluating the value of the area under the curve (AUC performance).

Conclusions

– The best training is obtained with an average of 19 photographs of individuals (P) and 3 photographs of individuals (PN). That is, in our model, for a successful result with at least a 95% of successes, it is necessary an average of 19 photographs per individual in different positions, which is quite reasonable for training in a real environment.
– Training sample must be unbalanced, with a greater number of individuals (P) to train, with respect to individuals (NP). A ratio of 6 (P) per 1 (NP).

Analysis of False Positives. As seen in the previous section, Test D is able to classify 95% of allowed individuals (P) and unauthorized (NP), but has problems in rejecting some of the unknowns (D). To determine which are the

Test Individual	Individual to refuse	Trained "confused"	Percentage recognition
#9.1			0.94877
#10.1			0.96323

Fig. 3. Comparison of samples of wrong type (D) with the training "confused" samples

Network with two intermediate layers of 10 neurons		
Types of Samples	Hits	%
P (20 samples***)***	**19**	95%
D (20 samples***)***	**20**	100%
NP (20 samples***)***	**20**	100%

Fig. 4. Results with a network with two intermediate layers of 10 neurons

individuals (P) with which the system " confuses " the unknown individuals (D), it is created a new neural network, with the same structure as the original, but with 6 output neurons, 5 outlets that rank the entries from each of the 5 individuals of all training, plus an entry for all the unknown (a network of 64x64 neurons in the entry, 10 in the hidden layer, and 6 outputs, the individual 1-5, and sixth out for unknown). The result is shown in Figure 3. (Attached is a column with the picture of the individual that the system is confusing with the unknown).

It is noted that the net recognizes by mistake two samples of an individual never seen before (D) but quite similar to another already known. Indeed, the human mind also produces such errors.

The advantage of the network is that if we retrain the network, indicating that the unknown patterns are NP, the network correctly classified with 100% success. After several studies of other network configurations, if we apply the same sets of training and testing a network composed of 2 layers of intermediate neurons 10 each, we obtain the following result (see Figure 4):

3.2 Example 2: The Feret Face Database

To verify that our model can work fine in real environments, we need to increase the number of samples of individuals (P) and individuals (NP). For this, we take as reference images in the database Feret [12], which contains an extensive database with the following characteristics:

- Background: Very variable.
- Scale: much variation.
- Point of view: Different angles of the same person.
- Position of the face in the picture: Different angles of the same person.

Sample 3	***Type***	***Samples***	***Individuals***
Training	P	309	**46**
	NP	91	16
Test	P	81	41
	NP	40	**20**
	D	33	**17**
Total		**554**	**83**

Fig. 5. Set Distributions of all Training Example 3

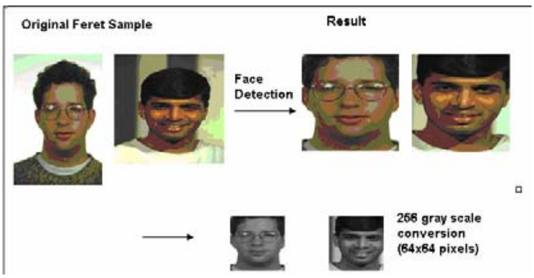

Fig. 6. Normalization of Feret image database

- Light: Different Illuminations
- Expression: quite variations

In addition, photographs are taken in different years, so there is enough variation regarding age and physiognomy of the same people. Assuming than over the 100 potential employees, 46 individuals have access to a restricted area (type P). We get, as a set of training, different distributions of photos for a total of 83 individuals, 554 photos in total, according to the distribution shown in Figure 5. (Individuals (P) and (NP) in the two subsets are the same).

The Feret database contains images taken at different points in time (along years) and in different places. Thus, for the network to function properly, we must normalize the images (see Figure 6) so that the faces are located on the same coordinates regardless of their position on the original images. We must center the faces as closely as possible in the normalizing matrix (in our case of 64x64 pixels). To center the face in the image, the Fdlib library was used [5].

As is clear from the results shown in Figure 7, the percentage of success obtained is not as the good as expected, considering the ratios of the previous tests. We study the distribution of samples in the training phase, concluding that the number of samples by individuals does not affect the quality of the final classification.

In this set of experiments, we have measured performance of two different NN architectures: the first one with one hidden layer (10 units), and the second one with two hidden layers (10 units per layer). Second architecture needs 4000 cycles to classify all the examples in order to get the same hits as the first architecture

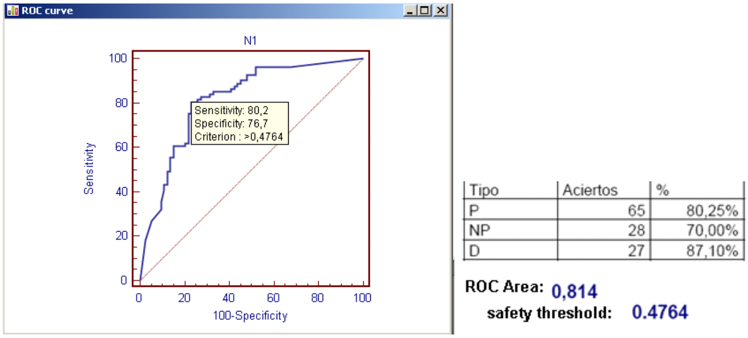

Fig. 7. Results of the evaluation database Feret evaluation

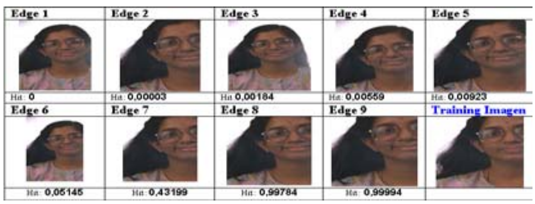

Fig. 8. Behavior of the network according to the center of the taken samples

with only 1000 cycles. So, by using the first architecture, we can achieve better performance.

A review of wrong samples on testing, compared with samples taken for the same individuals in training, indicates that the face is not in the same position. In order to determine if this is the origin of the increase in error rates, we take one wrong pictures at random to compare them with their respective images in the training set.

As it is clear from the study (see Figure 8), more left displaced images (as Edge2 or Edge7) are not giving the better results, but those which the square trimming is closest to the training images (Edge8 and Edge9). In conclusion, we can admit that in real environments, the sets of images must be framed in the area closer to the training examples, so we should use robust face tracking and focusing algorithms for better results.

4 Evaluation

It is difficult to find papers related to the classification for the authorization or denial of individuals belonging to classes with facial recognition. Usually the term face recognition is used to refer to two different applications: identification and verification. We will discuss identification in the event that the identity of the subject is only inferred from their facial features. Verification systems are those in which, besides the image, it is indicated who is the person that is claiming to belong to a specific set. Most jobs are focused on verifying the identity of an individual, i.e., compare the image features with those assotiated to the individual who claims to be. Moreover, usually it is used a neural network only as the end classifier, obtaining the values that identify the individual (main features) through other algorithms [8][9][7]. However, we can compare our method with the methods seen in the state of art (see the comparison chart of Figure 9).

In addition, there are few platforms for evaluating algorithms, such as "The CSU Face Identification Evaluation System [4], which based on the database Feret, has already published a series of statistics on the performance of these algorithms. The best yields can be seen in Figure 10.

Paper	Method	Success Percentage	Database size (Individuals/ images)
Brunelli y Poggio [Bru93]	Templates system	100%	47
Cristina Conde Vilda. [Cri06]	Hybrid 2D-3D	99%-97%-95%	105
Cottrell y Fleming [Cot90]	Neural networks - face detection, not verification.	97%	11 / 64
M. Tistarelli, E. Grosso [Tis00]	Polar coordinates.	97%	75 / 488
S. Lawrence, C. L. Giles [Law97]	Self-organizing map neural network.	96,20%	40 / 400
Moghaddam y Pentland [Mog97]	PCA (Principal Component Analysis)	96%	150 / 150
Turk y Pentland [Tur91]	PCA (Principal Component Analysis)	96% - 85%	16 / 2500
I. J. Cox [Cox96]	Geometric Features	95%	95 / 95
Howell, A.J. and Buxton, H. [How96]	RBF networks	95%	10
Brunelli y Poggio [Bru93]	Geometric Features	90%	47
O'Toole [Oto91]	Eigenvectors	88,60%	
F. S. Samaria [Sam94]	Hidden Markov Models	87%	40 / 240
M. J. Escobar, J. Ruiz-del-Solar. [Esc02]	EBGM and images processed into log-polar space	83.1% - 88.93%	15 / 165

(a)

Neural Network Characteristics	**Sample Database**	**Database size** (invidivuals/samples)	**Success** (%)
One intermediate layer: 10 units. Training: 1500 cycles.	UMIST Face Database	25/201	95%
Two intermediate layers: 10 units each Training: 500 cycles.	UMIST Face Database	25/201	98.33%
One intermediate layer: 10 units Training: 1000 cycles.	Feret Database	83/554	79.39%

(b)

Fig. 9. Comparison of state-of-the-art methods with the current work

Current work	*Méthod*	**Success Percentage**	**Database size** (indiv./samples)
Sample 1. The UMIST Face Database	Neural network with an intermediate layer of 10 units. 1500 cycles of training.	95%	25 / 201
Sample 1. The UMIST Face Database	Neural network with 2 intermediate layers of 10 units each other. (500 cycles of training).	98,33%	25 / 201
Sample 2. The Feret Face Database	Neural network with an intermediate layer of 10 units. 1000 cycles of training.	79,39%	83 / 554

Algoritmo	**Upper%**
Bayesian_MAP	77.5%
Bayesian_ML	77.5%
EBGM_Standard	70.6%
LDA_Euclidean	69.4%
LDA_ldaSoft	69.4%
PCA_Euclidean	68.1%
PCA_MahCosine	77.5%

Fig. 10. Results on the platform "The CSU Face Identification Evaluation System" and comparison with the current work

5 Conclusions

The proposed neural network is an initial solution to build a good model for access control in a visual way into an environment of around 80 people, and those with authority (P) to access into the restricted area can be about a total of 46 individuals.

• The number of samples selected in the training set does not directly influence the outcome of the classification

• The quality of the samples in training and the testing of all to classify is the basis on which we must base the model. Both sets must have a similarity position with respect to the area and the central positions.

• The response time, both the training and the results are very short, thus allowing its implementation with accessible hardware and not dedicated.

• The results of this study are consistent with the revised state of the art, with a very significant improvement in performance in the training time, added benefit that involves to deploy a system with a single simple to implement (neural network for extracting features and classification), with results within a high threshold of confidence, agile performance on the classification, and learning capability incorporated.

Acknowledgements

The authors would like to thank the CYCIT for financial support via project TIN-2007-67586-C02-01. We also thank all Ibermática staff for their dedication and professionalism.

References

1. Ballard, D.H., Brown, C.M.: Computer Vision. Prentice Hall, Englewood Cliffs (1982)
2. Bischof, H., Pinz, A.: Neural Networks in Image Pyramids. In: International Joint Conference on Neural Networks (IJCNN 1992), vol. 4, pp. 374–379 (1992)

3. Cottrell, G.W., Fleming, M.K.: Face recognition using unsupervised feature extraction. In: Proc. Int. Conf. Neural Network, pp. 322–325 (1990)
4. Beveridge, R., Bolme, D., Teixeira, M., Draper, B.: The CSU Face Identification Evaluation System User's Guide: Version 5.0. Computer Science Department Colorado State University (2003)
5. Kienzle, W., Bakir, G., Franz, M., Scholkopf, B.: Face Detection - Efficient and Rank Deficient. In: Advances in Neural Information Processing Systems, vol. 17, pp. 673–680 (2005)
6. Ministerio de Educación, Política Social y Deporte Instituto Superior de Formación y Recursos en Red para el Profesorado. Organización para la Cooperación y el Desarrollo Económico, http://w3.cnice.mec.es/oecd/dataoecd/8/7/glosario.htm
7. Howell, A.J., Buxton, H.: Invariance in Radial Basis Function Neural Networks in Human Face Classification. Neural Processing Letters 2(3), 26–30 (1996)
8. Guerrero, J.A.G.: Utilización del análisis de componentes principales para la clasificación de blancos de radar mediante resonancias naturales. Dpto. Fiísica Aplicada, Grupo de sistemas, señales y ondas Universidad de Granada (2006)
9. Lawrence, S., Giles, C.L., Tsoi, A.C., Back, A.D.: Face recognition: a convolutional neural-network approach. IEEE Transactions on Neural Networks 8(1), 98–113 (1997)
10. Li, Jain: Handbook Of Face Recognition. Springer, Heidelberg (2005)
11. Kilner, J.M., Vargas, C., Duval, S., Blakemore, S.-J., Sirigu, A.: Motor activation prior to observation of a predicted movement. Nature Neuroscience 7, 1299–1301 (2004)
12. Phillips, P.J., Moon, H., Rauss, P.J., Rizvi, S.: The FERET evaluation methodology for face recognition algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(10), 1090–1104 (2000)
13. Scruton, R.: La experiencia esttica. México, F.C.E, pp. 117–119 (1987)
14. UMIST Face Database, http://images.ee.umist.ac.uk/danny/database.html