Learning to recognize faces by successive meetings

M. Castrillón-Santana, O. Déniz-Suárez,
J. Lorenzo-Navarro and M. Hernández-Tejera
IUSIANI
Edif. Ctral. del Parque Científico Tecnológico
Universidad de Las Palmas de Gran Canaria
Las Palmas de Gran Canaria, 35017
Spain
mcastrillon@iusiani.ulpgc.es

***Abstract*— In this paper we focus on the face recognition problem. However, instead of following the usual approach of manually gathering and registering face images to build a training set to compute a classifier off-line, the system will start with an empty training set, i.e. no experience, and it will build it autonomously by continuous on-line learning. In that way the classifier evolves with the perceptual experience of the system, similarly to the way humans do. Experiments have been performed with 310 sequences corresponding to 80 identities. Two different configurations have been analyzed depending on the ability to detect new, i.e. unknown, identities. The results achieved evidence that if a verification stage is included the system learns fast to detect new identities. For revisitors, the accumulated error rate decreases in both cases, reaching around 50% if no verification is included. These results seem to indicate that more interaction or meetings with the different individuals are needed to affirm that their identity is familiar enough to be recognized robustly.**

***Index Terms*— face recognition, face detection, exemplars selection, learning systems, online learning, support vector machines, incremental PCA**

I. INTRODUCTION

Face analysis is nowadays a main topic of interest for computer scientists and psychologists. On the one side, computer scientists have developed during recent years many approaches [1]–[4] particularly for the face recognition problem. Upon observing their descriptions, it is evidenced that they are typically based on classifiers which have been computed off-line. Their performance is later measured using different test sets. The average error rate is used to characterize the reliability of each system assuming that the performance can be extended to the whole face domain.

Most of these approaches have been designed for the still image context and rarely for continuous processing [5], despite recent developments in face detection techniques. Databases used for benchmark contain single images too. A well known corpus used to evaluate recognition techniques is the FERET database [6] and more recently the Face Recognition Vendor Test or the Face Recognition Grand Challenge [5]. This database offers a large enough set in terms of individuals and samples, but the video context is not considered. Verification approaches make use of the BANCA protocol [7] which tackles the video problem for 208 individuals.

However, it is not clear that the results achieved with those databases can be extended to the whole face domain. Learning algorithms use the available training samples to guide a search for a solution. A robust solution has to predict, i.e. produce the correct output for future, unseen samples. Training samples alone are not sufficient for this, for there is a large number of possible hypotheses that fit the training set. This is the essential idea behind Wolpert's No Free Lunch Theorem [8], which states that, on the criterion of prediction performance, there are no reasons to prefer the hypotheses selected by one learning algorithm over those of another. The perfect fit to a training set does not guarantee low error for future, unseen samples. This justification provides an answer to explain the fact that even the best automatic face recognizers, which report a low error rate for known databases, are still not comparable to human performance in most real situations [9], [10].

On the other side, recent theories about the psychology of face recognition consider that face processing improves linearly with age reaching a plateau of high performance in young adults [11]. A continuous learning process takes place starting with the initial attraction of face patterns for newborns [12], following with the exposition to a huge number of face patterns, and reaching after some years the reliability exhibited by normal adults, which are able to analyze faces successfully under many circumstances.

That said, automatic face processing systems have rarely considered continuous learning, which occurs in

This paper is based on "Face recognition from a *tabula rasa* perspective," M. Castrillón-Santana, O. Déniz-Suárez, J. Lorenzo-Navarro and M. Hernández-Tejera, which appeared in the Proceedings of the 7th International Conference Automatic Face and Gesture Recognition, FG2006, Southampton, UK, April 10-12 2006. © 2006 IEEE.

Work partially funded by research projects Univ. of Las Palmas de Gran Canaria UNI2005/18, Canary Islands Autonomous Government PI2003/160 and PI2003/165 and the Spanish Ministry of Education and Science and FEDER funds (TIN2004-07087).

humans based on their perceptual experience. Instead, as mentioned above, automatic systems are based on precomputed classifiers. In this paper our aim is to design a face recognition system which taking the considerations explained above, evolves continuously based on its perceptual experience, i. e. its meetings or encounters with people.

Our motivation comes from the challenge stated at the end of the 90s, which focused on outperforming human levels of performance in low-quality images where facial identities seem to be available for familiar faces [13]. Therefore, our system will be focused on developing a robust face recognizer for a bounded set of familiar individuals, i.e. those which are met multiple times. We consider that this ability is crucial for more natural and comfortable Human Computer Interaction (HCI) [14], which would integrate a Vision Based Interface [15] able to perceive the user using Computer Vision. Therefore, a camera will be continuously acquiring images, which can of course register individuals close to the system.

Can the learning process be done online with current technology? Can the system select from its interactive sessions the info needed to first create and later update the different classifiers according to its experience? In this paper, we describe an approach trying to tackle this purpose. This approach reproduces the process of successive meetings with people, typical in humans, making use of available tools for face detection, representation and classification.

Section II describes recent work related to the topic covered in this paper. Section III presents the system used for continuous processing. Section IV describes the techniques used for face representation and classification. Experimental results are analyzed in Section V. Finally, some conclusions are summarized in Section VI.

II. Previous Work

Psychological theories of face analysis are not precise enough to be translated to a computer system, however they provide hints to understand the best known face processing system. A recent proposal described in [16], suggests a dual route model for face recognition, instead of the previous sequential or hierarchical models presented by other authors. On the basis of observations performed on prosopagnosic patients, which could not be explained by previous models, the authors have concluded that the process of face recognition is divided in two different processes located in different human brain areas. On one side, face detection, which would be a face-specific process. On the other, face identification, which would share part of the object recognition system.

This idea explains also the known ability of newborns to fixate on faces. According to some authors [17], that a priori or innate knowledge can be explained by evolution, i.e., babies who pay attention to their parents receive better care. Therefore, the face detection process has no sensitivity to face identity or any semantic aspect. Detection is fast, while identification depends on extensive exposure/learning from infancy through childhood. This fact seems to evidence the need of continuous learning along life as part of a practice and habituation process, until the moment that it becomes unconscious [18], [19].

The ability to recognize familiar faces at low resolution is impressive in humans [10], [13]. However, most computer based face recognition approaches tackle the problem using a single image per individual to recognize a large pool of identities [5]. These systems are trying to recognize faces which are not familiar enough, becoming less reliable with uncontrolled imagery. On the other side, humans are not so reliable for the task accomplished by computers, as evidenced in experiments where the photo ID was not enough to avoid fraud in high levels of performance [10], [13], [20], [21].

Different experiments suggest that an object model requires a collection of images [22] or their combination [10], which are actively collected by humans along their interaction sessions with a particular object or individual.

However, current automatic face recognizers insist on simplifying extraordinarily the learning process, neglecting the evidences exhibited by humans. Thus, why are we developing automatic systems to recognize unfamiliar faces? Why are we developing classifiers based on a single image extracted once?

Meetings, i.e. seeing other people's faces, along life are the obvious source to gather those collections of images used by continuous learning [10], [22]. Their analysis presents a major difference in relation to still image processing: Individuals present variations along the image stream. In this context it is hard to tackle the face recognition problem based on a single image per individual [22]. The crucial point here is that the ratio of intraclass, similarity between images of faces of the same individual), to interclass, similarity between images of faces of different individuals, variation is still very high in face recognition, even for a low number of individuals.

Different automatic systems provide nowadays visual information extracted from the face. In our context, a system performing live will have to manage video streams, and typical face processing approaches are inappropriate for the video stream context, as stated by different authors [23]–[25]. The interaction with different individuals will provide the system with the source to build any particular model. Focusing on face analysis, it is not adequate to use all the images present in a video stream to represent a specific facial class. It is obvious that there is redundancy in them, and their use would produce massive computational and storage costs [26], [27]. Therefore the representation and/or classification of individuals should be evaluated in time rather than using an one-shot methodology.

The extraction of significant patterns, or exemplars, is tackled in [23], where they are selected from a single video gallery of each individual. However, no further tuning is performed later during classification of new videos. That approach had the novelty of integrating temporal information in the classifier output, but did not alter the classifier by means of system experience.

The automatic selection of important patterns or *keyframes* is also considered in [27]. In that work, a tracking failure indicated that a new keyframe should be added to the keyframes database, each of them represented by a set of local features. Later, each new keyframe found during interaction would be compared with those already contained in an individual description and added if needed. This action required robust recognition.

In [26] the authors implemented in a humanoid robot the ability to learn to recognize the people it interacts with. As a novelty, the system started with an empty database, exactly the problem that we tackle, and developed a completely unsupervised face recognition system. The system used the standard eigenface method [28], distinguishing two stages: 1) an initial stage where the system must be able to cluster its visual stimuli, and 2) online training, which based the recognition of unknown individuals on a simple distance measure with already stored ones. The detection of an unknown individual allowed the system to create a new identity cluster. In a reduced set of 9 individuals, the system was unable to learn 5 of them using the unsupervised mechanism. The authors affirm that this fact is due to the known performance degradations of the eigenface approach for facial expressions, facial alignment and scale.

The authors of another system [29], made use of Modified Probabilistic Neural Networks being able to identify not only known, but also unknown subjects. Once the system detected an unknown subject, a fixed number of images in the buffer were selected to create new links in the Neural Network. These images were selected according to the difference with the average face computed during the interaction. Once a new model is learned, it will not be updated later. Some experiments were performed with a reduced number of subjects.

Exemplars are taken from single images and aligned by hand in [10]. They are later use to compute the average image per identity which is later used to learn. However, the process is only partially automatic, and errors are not used to retrain the system.

Summarizing, the approaches which tackle the face processing problem in video streams have been rarely designed for that context. As described above, just a few have focused on the automatic exemplar selection problem but, with the exception of the preliminary experiments described in [26], no face processing classifier is modified during system's life based on its perception.

III. Continuous Learning

In human learning, it is known that only novel, unpredicted stimuli trigger the review of the existing working hypothesis [18]. Therefore, stimuli that are inconsistent with the current hypothesis will force its change to make it consistent with them. Following this consideration, our learning approach will follow the diagram depicted in Figure 1. During continuous processing held in an interaction meeting, faces are detected automatically. Among those detected faces some exemplars are used to label the individual identity. *Incorrectly classified exemplars are then used to update the existing classifier.*

In this section first the face detection approach is briefly outlined. Then the exemplar selection criteria and classification mechanism are described. The section is finished describing the learning procedure.

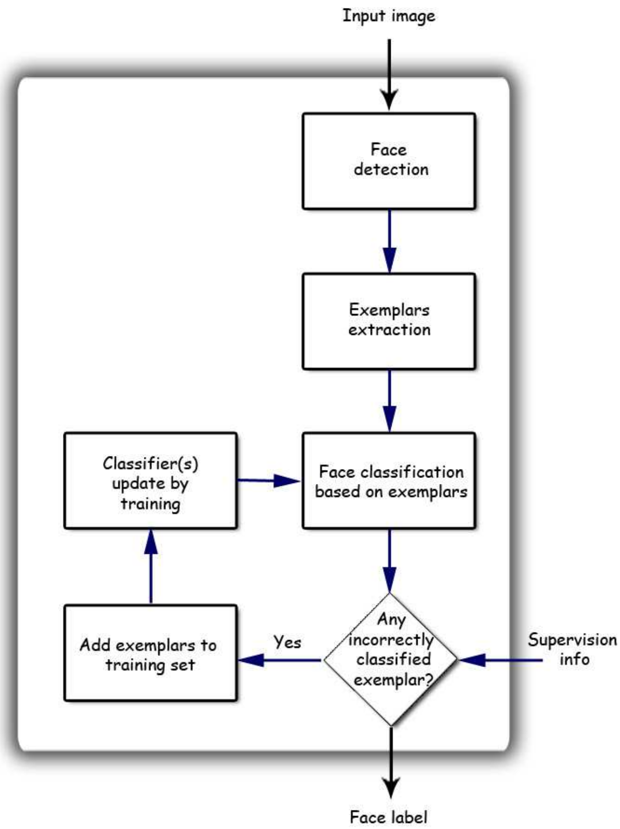

Fig. 1. Learning mechanism.

A. Automatic Face Detection

The real-time face detector, see [30] for more details, combines different techniques providing robust performance in different conditions and environments. An initial detection is obtained by means of window shift detectors [31], [32]. The face location is used, sometimes combined with skin color information, to detect eye, mouth and nose positions for frontal faces. At this point each detected face is parameterized in terms of its position, size, color and facial patterns: $x_i = \langle pos, size, color, eyes_{pos}, eyes_{pat}, mouth_{pos}, mouth_{pat}, nose_{pos}, nose_{pat}, face_{pat}\rangle$. Later, temporal coherence based on these features direct different cues in the next frames which are applied opportunistically in an order based on computational cost and reliability. For each previously detected face, if one of these cues reports a coherent detection, the process is finished. The different cues employed are roughly described as follows:

- Facial features tracking: A fast tracking algorithm [33] is applied in an area that surrounds previously detected eyes, mouth and nose, if available.
- Face detector: The Viola-Jones face detector [31] is applied in an area that covers the previous detection.
- Local context face detector: If previous techniques fail, it is applied in an area that includes the previous detection [32].
- Skin color: Skin color is searched in the window that contains the previous detection, and the new sizes and positions are coherently checked.
- Face tracking: If everything else fails, the prerecorded face pattern is searched in an area that covers previous detection [33].

If both eyes and another feature (nose or mouth) are detected, the face is normalized to a 59×65 size, and will be considered an exemplar candidate. In absence of detections, the process will be based on the standard window shift detectors [31], [32].

B. Exemplar Selection

The face detector processes the video stream and generates an Interactive Session, IS, gathering information about a set of detection threads, $IS = \{dt_1, dt_2, ..., dt_n\}$. A detection thread contains a set of continuous detections, i.e. detections which take place in different frames. These consecutive detections are related in terms of position, size and pattern matching techniques. Thus, for each detection thread, the face detector system provides a number of facial samples, $dt_p = \{x_1, ..., x_{m_p}\}$.

The huge amount of data extracted during an interactive session must be reduced in some way to avoid no relevant information and redundancy. From this collection of facial samples the exemplars $e_p = \{e_1, ..., e_{s_p}\}$ are extracted for each detection thread, dt_p.

The criterium used to select an exemplar is based on tracking failures, as they suggest an evidence of substantial change in face appearance, which forces the tracker to lose the target, see Figure 2. Under this circumstance, the system needs to use another cue to detect again first the face and later the eyes and another facial feature as explained above, or the detection thread will be considered lost. Once the eyes and mose/mouth are detected again, that face is taken as a new exemplar. For each exemplar, its life time or persistence until the next tracking failure is stored. Additionally its PCA reconstruction error (considering a PCA space) [34] is computed. Therefore, an exemplar is described by the normalized detected face, x_j, its persistence, pe_j, timestamp, t_j, and PCA reconstruction error, $PCAerror_j$, i.e. $e_j = \langle x_j, pe_j, t_j, PCAerror_j \rangle$.

Given an interactive session, IS, for any old enough detection thread (older than 20 frames), any facial classifier being considered by the system can compute the *a posteriori* probability for a class, C_k. This is done considering those exemplars whose PCA error is not notoriously bigger than the average, and weighting their binary classification according to its relative persistence:

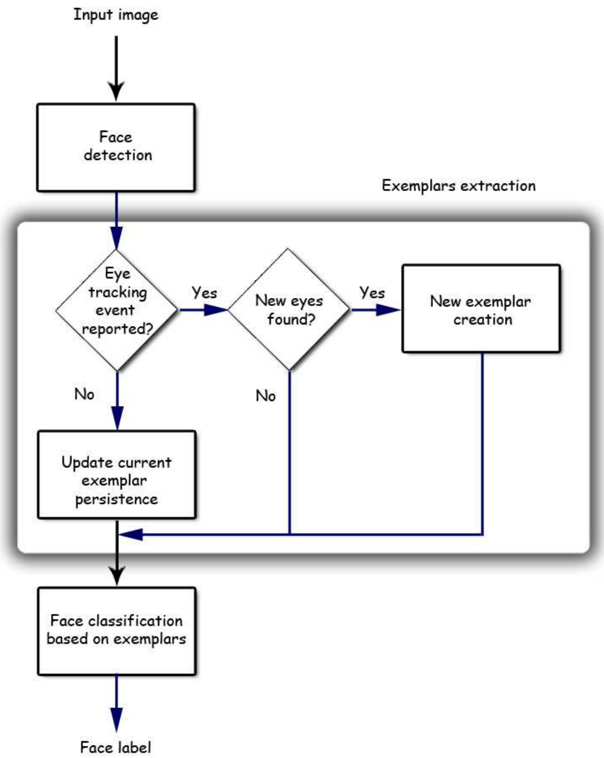

Fig. 2. Exemplar selection mechanism.

$$P(C_k|dt_p) = \frac{\sum_{j=1}^{s_p} P(C_k|e_j) * pe_j}{\sum_{n=1}^{s_p} pe_n} \qquad (1)$$

This value can be computed for the exemplars extracted during the interactive session, or only for those which have been selected within a recent Window Of Attention (WOA). In that case, in Equation 1 only those exemplars inside the WOA will be considered.

C. Learning by Incorrectly Classified Exemplars

The basic idea for updating the classifiers after a meeting is to make use of the incorrectly classified patterns. Similarly to [26], we distinguish two different epochs during system learning. However, the second epoch should be considered only when the system is reliable enough, which is not the case of the current implementation as shown in the experiments below:

1) First, we consider that at the initial stage, during the system *infancy*, the system must be necessarily supervised by a human expert. The system is able to detect faces but it is still not able to classify them reliably. Humans first recognize the face class, with the different considerations about the way this knowledge is achieved [17]. Later, we are guided or helped by our parents and/or through other modalities (voice, context [35], etc.) to learn to distinguish different subclasses within face class [16]: mom, dad, male/female, young/mature/old,

familiar/unfamiliar, etc. This process requires an evolution till different robust classifiers are learned [11].

2) On the second epoch, once the system confidence is good enough, it will first request supervision only for doubtful situations, and later autonomously select the misclassified patterns to be used to update the classifier.

During the first epoch an expert corrects the system, if necessary, after it has suggested a class for a detection thread. Any incorrectly labelled exemplar will be used to update and correct the classifier.

Thus, given an exemplar, e_j, if the system was corrected, and the correct class was C_c, all the incorrectly labelled exemplars, i. e. $P(C_c|e_j) = 0$, will be added to the training set. If the supervisor confirmed the class suggested by the system, C_k, similarly incorrectly assigned exemplars, $P(C_k|e_j) = 0$, will be added to the training set.

The result is that the samples added to the system during learning are given by incorrect classification during system *life*. A new interactive session with individuals of the same class (identity, gender, ...) will add exemplars to the training set if they were incorrectly classified. Therefore, the classifier evolves according to its perceptual experience, i. e. it is not previously fixed.

IV. Recognition vs. Verification

There are two different problems that share similar techniques in the face identification literature. The first one is associated to recognition from a database without a priori knowledge of the person's identity, see Figure 3-left. The second problem is related to verification or authentication of an identity given by a subject, see Figure 3-right.

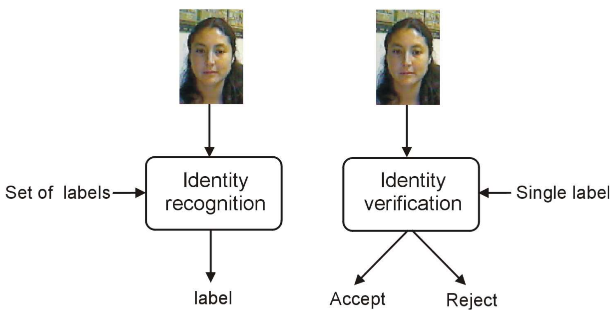

Fig. 3. Recognition (left) and Verification Schemas (right).

The first problem is tackled by means of a single n-class classifier that assigns a label to any new image analyzed by the system. The classifier is learnt from a training set which contains samples of those n individuals. If a face image of an individual not contained in the training set is processed, the system will not be able to observe that circumstance, it will provide in any case one of those n labels. For the second problem, the literature offers the verification approach to confirm a given identity. Given n identities, the verification system needs n binary classifiers, i.e. a rejection class for each individual, in order to accept or reject the label provided by the user for the face image. These systems are mainly focused on confirming the label provided, but do not guess if the identity is not contained in the database.

Our system should be able to recognize new or unknown individuals [26], [29], i.e. individuals which are not already contained in the classifier, in order to create a new identity class. A simple way to implement that using recognition and/or verification techniques is to apply both approaches in a cascade configuration. The identity classifier has the drawback of not being able to verify if the user is contained in the training set. That can be achieved by a verification stage if a label is provided. Thus, the label provided by the identity classifier is used for the verification stage, see Figure 4.

This approach forces the system to have a classifier for n classes for the first stage, and n binary classifiers for verification in the second stage.

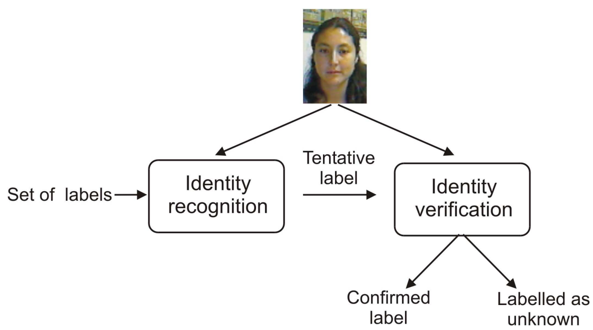

Fig. 4. Identity recognition plus verification.

The experiments presented below will consider a one stage Recognition approach, see Figure 3-left, but also the two stage approach depicted in Figure 4.

A. Representation Space and Classification

Face images are highly dimensional. In order to avoid the consequent processing overload problem, Principal Components Analysis (PCA) decomposition [36] is applied to the training data provided. This allows us to represent the appearance of the different individuals contained in the training set [28]. However, we have stated above that our system should evolve, and not necessarily an initial PCA space will contain all the information of the PCA space. Therefore, the system will also be tested considering that its representation space evolves. For that purpose we have followed the incremental PCA (IPCA) approach described in [37], [38].

Using the face representation space, different classifiers can be used to select a label for each face processed. The original implementation [28] makes use of Nearest Neighbor Classifier (NCC) for that purpose. However, different authors argued that this approach provides low reliability if lighting conditions are not restricted [39].

Recent developments use local representations such as Independent Components Analysis (ICA) [40] to get a better representation space. However, the work described in [41] showed that the selection of a powerful classification criteria was more critical than the representation space (PCA or ICA).

According to these results, recognition experiments have been carried out using Support Vector Machines (SVMs) [42], once that PCA-based dimension reduction has been performed. LIBSVM [43] has been the library employed for this purpose. Since SVM is essentially a biclass classifier, this library provides multi-class classification based on a voting approach computed with one-against-one biclass classifiers [44].

V. Experiments

A. Datasets

1) Static images: This dataset contains 7000 face images taken from Internet and selected samples from facial databases such as BIOID [45]. They have been annotated by hand to get their eye positions and labelled according to their gender. These images have been normalized according to eye positions obtaining 59×65 samples.

This dataset was used to compute off-line the initial PCA space employed for face representation. In the experiments, as commented above, we assume two different uses of this PCA space: 1) keeping it fixed without change along system performance, and 2) employing an incremental PCA (IPCA) approach [37], [38] to improve the representation space with new experience extracted from uncorrectly classified exemplars.

2) Video streams: Our aim is to produce successive meetings with different individuals. The search of video streams for that purpose is not an easy task. Most facial databases do not contain sequences offering the facial evolution of different individuals in different conditions and days. The availability of a controlled illumination and restricted background database such as XM2VTS [46] is not well suited to verify the unrestricted problem tackled in this paper. For that reason we have built up a database making use of broadcast television, and different people recorded with different webcams and cameras without controlled conditions.

The database contains around 310 different video streams (320 by 240 pixels or larger) corresponding to 80 identities. It must be remarked that the identities contained in this database are completely independent from those used to compute the initial PCA space. For each identity at least two sequences are available.

B. Results

As commented above, every valid exemplar extracted during detection is projected to the corresponding PCA space to get its representation which is later classified by the n-class classifier. The weighted combination of the classifications provided for a meeting reports a suggested class. In the experiments we have used that class to directly provide a label for the sequence, Figure 3 left, or it was further used to verify the identity assigned to the individual met, Figure 4.

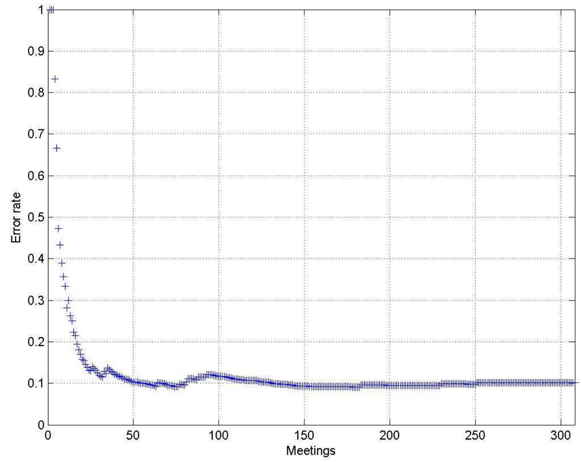

Fig. 5. FAR evolution achieved considering new identities.

Figure 5 presents results related to the False Acceptance Rate (FAR), the two stages approach, i.e. including classification and verification stages. These results are similar to those achieved in [47], in the sense that the system learns quickly to detect new or unknown identities. However, it must be noticed that the error does not decrease more due to the fact that in these experiments, the probability of meeting a new individual is very small once most of the individuals have been already met.

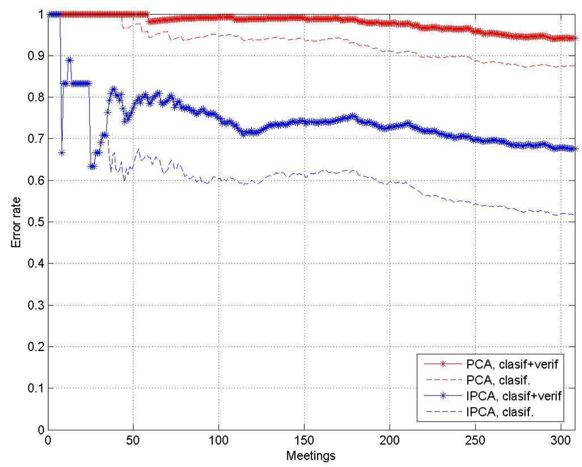

Fig. 6. Accumulated error rate evolution achieved considering only identities met more than once, but using the average image instead of a collection of exemplars to model an identity.

However, our main aim is to recognize *revisitors*, i.e. individuals who come back. That is related with the False Rejection Rate (FRR), that represents the error ratio which corresponds to an already met identity which was not properly recognized. When FRR is small it means that the system is good recognizing *revisitors*. The FRR rate evolution is presented in Figure 6. These results have been averaged after 3 randomly ordered runs, in order to avoid particularly good or bad performances. They are achieved starting from an empty training set, i.e. using a *tabula rasa* (blank slate) perspective. The figure presents four different error evolution curves, corresponding to different approaches as follows:

- Fixed PCA face representation space, single stage approach, i.e. no verification.
- Fixed PCA face representation space, two stages approach, i.e. recognition and verification are included.
- Incremental PCA face representation space, single stage approach, i.e. no verification.
- Incremental PCA face representation space, two stages approach, i.e. recognition and verification are included.

For a specific meeting, the FRR indicates the accumulated error, up to that moment, of identities previously seen, i.e. known individuals. The first look evidences that for all the approaches the recognition error decreases, which seems to indicate that the model for a particular identity is being improved. However, not all the approaches report a similar error. Particularly the integration of incremental PCA, which has real-time performance in these experiments, presents a notorious improvement.

On the other side, it is observed that the use of the two stages approach restricts the ability of the system to recognize already known individuals. The IPCA approach reports a final accumulated error lower than 70%, while using a single stage approach, it is slightly greater than 50%. These results suggest that the verification stage is at the moment quite restrictive. Some alternatives must be investigated to alleviate that effect, due to the fact that recognizing new individuals if of course a must.

Considering that the dataset contains from two sequences till nine per identity the results achieved are promising, and seems to indicate that the achievement of a reliable recognition system for a reduced set of individuals using real video without further restrictions, could be affordable. However, such a system would be available only after gathering enough information for each individual.

VI. Conclusions

Our main objective was to analyze if an individual can become familiar to a face recognition system after successive meetings. At the same time the system must be able to reject individuals not belonging to the familiar group. These abilities must be reached using an automatic face detector and using no more facial data than those extracted automatically from the system meetings.

In the experiments carried out, the system shows an improving performance in terms of rejecting unknown individuals. The performance achieved for revisiting people is still not reliable enough. However, the accumulated error rate keeps decreasing clearly. This fact is notoriously evidenced after the introduction of incremental PCA, which always outperforms the results achieved in previous experiments [47]. Can we assume that an individual becomes familiar when a collection of multi views is obtained? Future work must focus on providing more experience to the system, i.e. collecting more meetings for familiar individuals in order to verify this hypothesis. In that sense, we are also interested in investigating alternative face representation techniques debated in the psychology literature [10], where the average exemplars is used instead of all of them. The achievement of a similar performance, would be very interesting in the sense that the training stage would depend only in the number of classes, and not in the number of classes, and samples per class.

Future work should also consider the coordination with other modalities which could supervise the system in case of doubt. Note that, even if we try to recognize only a low number of individuals, face recognition may still fail because there is always the possibility that unseen facial images confuse the system (i.e. an individual in certain pose or under certain illumination is misrecognized). The experiments in the final stages required less than a second for updating the classifiers. Incremental approaches would be of great interest for large scale experiments.

Short Biographies

Modesto Castrillón-Santana is an Assistant Professor at the Department of Computer Science and a research member of the Institute of Intelligent Systems and Numerical Applications in Engineering at the ULPGC (University of Las Palmas de Gran Canaria), Spain. His research interests include facial detection and recognition, and computer vision for human computer interaction. He holds since 2003 a PhD from the ULPGC, and is member of AEPIA, AERFAI and the IEEE.

Oscar Déniz-Suárez holds a PhD from the ULPGC, Spain. Currently, he teaches Computing Fundamentals at the same university. His main interest is in computer vision for human-robot interaction. He is member of the University Institute for Intelligent Systems and Numerical Applications in Engineering.

Javier Lorenzo-Navarro is Associate Professor at the Department of Computer Science and a research member of the Institute of Intelligent Systems and Numerical Applications in Engineering at the ULPGC. He received his PhD in Computer Science in 2001. He has participated as research member and project coordinator in several research projects. His current research areas are computer vision applied to human computer interaction, homeostasis control in vision and data mining.

Mario Hernández-Tejera is Full Professor in Computer Science and Artificial Intelligence at the ULPGC and member of the Institute for Intelligent Systems and Numerical Applications in Engineering at the same University. His research interests includes artificial intelligence, computer vision and pattern recognition and its applications to perceptual user interfaces. He has directed several research projects in computer vision and pattern recognition and is co-author of more than 50 papers.

REFERENCES

[1] A. Samal and P. A. Iyengar, "Automatic recognition and analysis of human faces and facial expressions: A survey," *Pattern Recognition*, vol. 25, no. 1, 1992.

[2] R. Chellappa, C. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey," *Proceedings IEEE*, vol. 83, no. 5, pp. 705–740, 1995.

[3] A. Tolba, A. El-Baz, and A. El-Hardy, "Face recognition: A literature survey," *International Journal of Signal Processing*, vol. 2, no. 1, pp. 88–103, August 2005.

[4] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *Association for Computing Machinery*, vol. 35, no. 4, pp. 399–458, 2003.

[5] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *IEEE Conference on Computer Vision and Pattern Recognition 2005*, 2005.

[6] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face recognition algorithms," NISTIR, TR 6264, January 1999.

[7] E. Bailly-Bailliere, S. Bengio, F. Bimbot, M. Hamouz, J. Kittler, J. Mariethoz, J. Matas, K. Messer, V. Popovici, F. Poree, B. Ruiz, and J.-P. Thiran, "The BANCA database and evaluation protocol," in *Proc. Audio- and Video-Based Biometric Person Authentication*, J. Kittler and M. Nixon, Eds. Berlin: Springer, June 2003, pp. 625–638.

[8] D. Wolpert, "The existence of a priori distinctions between learning algorithms," *Neural Computation*, vol. 8, no. 7, p. 13911420, 1996.

[9] A. Adler and J. Maclean, "Performance comparison of human and automatic face recognition," in *Biometrics Symposium*, 2004.

[10] A. M. Burton, R. Jenkins, P. J. Hancock, and D. White, "Robust representations for face recognition: The power of averages," *Cognitive Psychology*, vol. 51, no. 3, pp. 256–284, 2005.

[11] H. Leder, G. Schwarzer, and S. Langton, *Development of Face Processing in Early Adolescence*. Hogrefe & Huber, 2003, ch. 5, pp. 69–80.

[12] G. Schwarzer, N. Zauner, and M. Korell, *Face Processing during the first Decade of Life*. Hogrefe & Huber, 2003, ch. 4, pp. 55–68.

[13] A. M. Burton, S. Wilson, M. Cowan, and V. Bruce, "Face recognition in poor quality video: Evidence from security surveillance," *Psychological Science*, vol. 10, no. 3, 1999.

[14] A. Pentland, "Looking at people: Sensing for ubiquitous and wearable computing," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, pp. 107–119, January 2000.

[15] M. Turk, "Computer vision in the interface," *Communications of the Association for Computing Machinery*, vol. 47, no. 1, pp. 61–67, January 2004.

[16] B. de Gelder and R. Rouw, "Beyond localisation: a dynamical dual route account of face recognition," *Acta Psychologica*, no. 107, pp. 183–207, 2001.

[17] V. Bruce and A. Young, *The eye of the beholder*. Oxford University Press, 1998.

[18] B. J. Baars, *A cognitive theory of consciousness*. Cambridge University Press, NY, 1988.

[19] G. Mandler, *Mind and body: Psychology of emotion and stress*. W W Norton & Co Inc, 1984.

[20] R. Kemp, N. Towell, and P. G., "When seeing should not be believing: Photographs, credit cards and fraud," *Applied Cognitive Psychology*, vol. 11, no. 3, pp. 211–222, 1997.

[21] G. Pike, R. Kemp, and N. Brace, "The psychology of human face recognition," in *IEEE Colloquium on Visual Biometrics*, 2000.

[22] G. Wallis and H. Buelthoff, "Learning to recognize objects," *Trends in Cognitive Sciences*, vol. 3, no. 1, pp. 22–31, January 2000.

[23] V. Krueger and S. Zhou, "Exemplar-based face recognition from video," in *European Conference on Computer Vision (ECCV), Kbenhavn, Denmark*, May 2002.

[24] L. Lorente and L. Torres, "Face recognition of video sequences in a mpeg-7 context using a global eigen approach," in *International Conference on Image Processing'99, Kobe, Japan*, 1999.

[25] L. Torres and J. Vilá, "Automatic face recognition for video indexing appplications," *Pattern Recognition*, vol. 35, pp. 615–625, March 2002.

[26] L. Aryananda, "Recognizing and remembering individuals: Online and unsupervised face recognition for humanoid robot," in *Proc. of IROS*, 2002.

[27] C. Wallraven and H. Buelthoff, "Automatic acquisition of exemplar-based representations for recognition from images sequences," in *Computer Vision and Pattern Recognition*, 2001.

[28] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[29] J. Fan, N. Dimitrova, and V. Philomin, "Online face recognition system for videos based on the modified probabilistic neural networks," in *Proceeding of IEEE ICIP 2004*, Singapore, 2004, pp. 104–110.

[30] M. Castrillón Santana, O. Déniz Suárez, M. Hernández Tejera, and C. Guerra Artal, "Encara2: Real-time detection of multiple faces at different resolutions in video streams," *Journal of Visual Communication and Image Representation*, 2007, (in press).

[31] P. Viola and M. J. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition*, 2001, pp. 511–518.

[32] H. Kruppa, M. Castrillón Santana, and B. Schiele, "Fast and robust face finding via local context," in *Joint IEEE Internacional Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS)*, October 2003, pp. 157–164.

[33] C. Guerra, M. Hernández, A. Domínguez, and D. Hernández, "A new approach to the template update problem," in *2nd Iberian Conference on Pattern Recognition and Image Analysis*, Estoril, Portugal, June 2005.

[34] E. Hjelmas and I. Farup, "Experimental comparison of face/non-face classifiers," in *Procs. of the Third International Conference on Audio- and Video-Based Person Authentication. Lecture Notes in Computer Science 2091*, 2001, pp. 65–70.

[35] P. Sinha and T. Poggio, "I think i know that face ..." *Nature*, vol. 384, no. 6608, pp. 384–404, 1996.

[36] Y. Kirby and L. Sirovich, "Application of the Karhunen-Loéve procedure for the characterization of human faces," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, July 1990.

[37] P. Hall, D. Marshall, and R. Martin, "Incremental eigenanalysis for classification," in *British Machine Vision Conference*, vol. 1, September 1998, pp. 286–295.

[38] M. Artac, M. Jogan, and A. Leonardis, "Incremental pca for on-line visual learning and recognition," in *Proceedings 16th International Conference on Pattern Recognition*, 2002, pp. 781–784.

[39] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. on PAMI*, vol. 19, no. 7, pp. 711–720, 1997.

[40] M. Bartlett and T. Sejnowski, "Independent component of face images: a representation for face recognition," in *Procs. of the Annual Joint Symposium on Neural Computation,Pasadena, CA*, May 1997.

[41] O. Déniz Suárez, M. Castrillón Santana, and F. M. Hernández Tejera, "Face recognition using independent component analysis and support vector machines," *Pattern Recognition Letters*, vol. 24, no. 13, pp. 2153–2157, September 2003.

[42] V. Vapnik, *The nature of statistical learning theory*. New York: Springer, 1995.

[43] C.-C. Chang and C.-J. Lin, *LIBSVM: a library for support vector machines*, 2001, software available at `http://www.csie.ntu.edu.tw/~cjlin/ libsvm`.

[44] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multi-class support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, pp. 415–425, 2002.

[45] R. W. Frischholz and U. Dieckmann, "Bioid: A multimodal biometric identification system," *IEEE Computer*, vol. 33, no. 2, February 2000.

[46] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, "Xm2vtsdb: The extended m2vts database," in *Second International Conference on Audio and Video-based Biometric Person Authentication*, March 1999.

[47] M. Castrillón Santana, O. Déniz Suárez, J. Lorenzo Navarro, and M. Hernńdez Tejera, "Face recognition from a *tabula rasa* perspective," in *Proceedings of the International Conference on Automatic Face and Gesture Recognition*, Southsampton, UK, April 2006.