Gender and Identity Classification for a Naive and Evolving System

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, Spain
mcastrillon@iusiani.ulpgc.es

Abstract

This paper does not propose a new technique for face representation or classification. Instead the work described here investigates the evolution of an automatic system which, based on a currently common framework, and starting from an empty memory, modifies its classifiers according to experience. In the experiments we reproduce up to a certain extent the process of successive meetings. The results achieved, even when the number of different individuals is still reduced compared to off-line classifiers, are promising.

1 Introduction

Nowadays face analysis is a main topic of interest for computer scientists and psychologists. On the one side, different automatic facial analysis systems have been developed in recent years, particularly applied to face recognition. Most of them were designed for still images, or directly adapted to video streams [21], despite recent developments in face detection techniques.

On the other side, most face recognition theories consider that face processing improves linearly with age reaching a plateau of high performance in young adults [18]. Therefore, starting with the attraction of faces for newborns [24], an adult is able to process faces reliably after being exposed to thousands of facial stimuli.

That said, automatic face processing systems have rarely considered online learning, which occurs in humans, during system *life* based on its experience. Indeed they typically compute a classifier off-line that is later analyzed with different test sets, assuming that the performance can be extended to the whole face domain. For example, a known corpus used to evaluate recognition techniques is the FERET database [22] and more recently the Face Recognition Vendor Test or the Face Recognition Grand Challenge [21]. This database offers a large enough problem in terms of individuals and samples, but the video context is not considered. Verification approaches make use of the BANCA protocol [3] which tackles the video problem for 208 individuals. Therefore, hundreds of approaches [9, 26, 33] have been described and compared in restricted environments, but they are still not comparable to human performance in daily situations [1].

In this paper we focus on the idea of developing a system which evolves online based on its perceptual experience, i. e. its meetings or encounters with people. Making use of available tools for face detection, representation and classification, we reproduce up to a certain extent the process of successive meetings, typical in humans.

Section 2 describes some recent work about this topic. Section 3 presents the system adopted for face processing. Experimental results are analyzed in Section 4 comparing different online learned classifiers with classifiers computed off-line. Some conclusions are summarized in Section 5.

2 Previous Work on Face Processing

The literature referred to face processing in humans presents different models. A recent idea described in [10], 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 model would mean that some tasks related to facial analysis are performed after detection, but others not. 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.

Most current face recognition implementations are designed to work with a single high resolution image of a per-

son [21]. These systems are trying to recognize faces which are not familiar enough. Observing humans, we have an impressive ability to recognize familiar faces at low resolution [6], but we are not so reliable for this task with unfamiliar faces. This fact has been evidenced in experiments where the photo ID was not enough to avoid fraud with high levels of performance [14, 23, 5]. Thus, why are we developing automatic systems to recognize unfamiliar faces? Indeed, different experiments suggest that an object model requires a collection of images [31] or their combination [5], which are collected by humans along their life.

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 [16, 19, 27]. Video stream analysis presents a major difference in relation to still image processing: Individuals present variations along the image stream. The interaction with different individuals will provide the system with the source to build any particular model. Focusing on face analysis, it is not reliable 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 [2, 32]. Therefore the representation and/or classification of individuals should be evaluated in time rather than using an one-shot methodology.

The extraction of exemplars, to reduce redundancy is considered in [16]. That system selects the exemplars from a single gallery video 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, in authors language, is also considered in [32]. In that work, a tracking failure indicated that a new keyframe should be added to the representational database 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 [2] the authors implemented in a humanoid robot the ability to learn to recognize the people it interacts with. As a novelty, the system was launched 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 performances degradations of the eigenface approach for facial expressions, facial alignment and scale.

The authors of another system [11], 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.

In all these approaches the authors pointed out the absence of a large database of sequences in order to perform extensive experiments for this purpose.

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 [2], no face processing classifier is modified during system performance based on its perception.

3. System Setup

Everything will be done automatically by the system with already existing approaches: 1) face detection, 2) exemplar extraction, 3) face representation and 4) face learning. The following sections describe the approaches considered.

As seen in Figure 1, the face detection module provides face detection data, which are filtered by means of exemplar selection. These selected exemplars are used to suggest a classification for the meeting. Incorrectly classified exemplars are taken by the system to recompute the classifier online. Using this system architecture, the face classification module is modified online while the face detection and representation modules are fixed.

Similarly to [2], we distinguish two different epochs during system learning, but we tackle only the first one in the experiments presented below. 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 [4]. Later, we are guided or helped by our parents or other modalities (voice, context [25], etc.) to learn to distinguish different subclasses within face class

[10]: mom, dad, male/female, young/mature/old, familiar/unfamiliar, etc. This process requires an evolution till different robust classifiers are learned [18].

On the second epoch, once the system confidence is good enough, it will be able to autonomously select the misclassified patterns to be used to update the classifier.

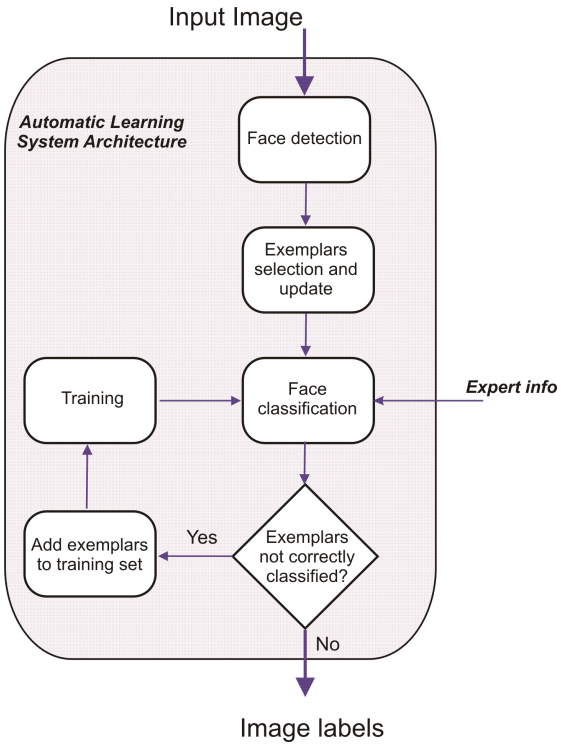

Figure 1. Graphical overview of the system.

3.1. Automatic Face Detection

The real-time face detector, see [7] for more details, combines different techniques providing robust performance in different conditions and environments. An initial detection is obtained by means of a window shift detector. Later, temporal coherence is used and each face is parameterized in terms of not only its position and size, but also its average color. Additionally, the skin color blob provides valuable information to detect eye positions for frontal faces.

In summary, each face detected in a frame can be characterized by different features $x_i = \langle pos, size, color, eyes_{pos}, eyes_{pattern}, face_{pattern} \rangle$. These features direct different cues in the next frames which are applied opportunistically, in an order based on computational cost and reliability:

- Eye tracking: A fast tracking algorithm [13] is applied in an area that surrounds previously detected eyes, if available.
- Face detector: The Viola-Jones face detector [30] 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 [17].
- 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 [13].

If the eyes are detected, the face is normalized to a 59×65 size. In absence of detections, the process will be started again using the Viola-Jones based detectors applied to the whole image.

3.2 Exemplars selection

During an interactive session, IS, i.e. during the video stream processing, the face detector gathers 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 information 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 to select an exemplar is based on tracking failures, as they show an evidence of substantial change in face appearance, which forces the tracker to lose the target. Under this circumstance, the system needs to use another cue to detect again first the face and later the eyes as explained above, or the detection thread will be considered lost. Once the eyes are detected again, that face is taken as a new exemplar. For each exemplar, its time life or persistence until the next tracking failure is stored. Therefore, an exemplar is described by the normalized detected face, x_j, its persistence, pe_j, and time-stamp, t_j, i.e. $e_j = \langle x_j, pe_j, t_j \rangle$.

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 by weighting

the binary classification for each exemplar according to its relative persistence. This is expressed as:

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

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).

3.3 Learning by Incorrectly Classified Patterns

As mentioned above, the initial stage of supervision is controlled by an expert. Therefore, the expert can correct the system after it has suggested a class for a detection thread. Any incorrectly labelled exemplar will be used to update and correct the classifier. Once the system is reliable, second epoch, it will request supervision only for doubtful situations.

Thus, 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.

3.4 Representation-Classification Space

As we are tackling a face classification problem, we select first, similarly to [16], a well known face representation space in advance: the PCA space due to its economical advantages [15]. On this representation space, a Support Vector Machines (SVM) classifier [29] is trained using the training set for each considered problem. This combination $PCA+SVM$ has been chosen for being well known by the community and for the good performance results achieved.

4 Experiments and Results

We have considered two different face classification problems to solve: 1) gender classification, and 2) identity recognition. The problems have a different nature in the sense that for gender recognition the number of classes is known a priori, while it is not bounded for identity recognition. Indeed, for identity recognition it would also be desirable to have a system able to detect a new identity.

	Training set size		Test set size		Pattern
Descriptor	Female	Male	Female	Male	Size
Gender	1223	1523	835	2246	59×65

Table 1. Training and test sets.

4.1 Datasets

4.1.1 Static images.

This dataset contains 6000 face images taken randomly from Internet and selected samples from facial databases such as BIOID [12]. 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.

The first use of this dataset was to compute the PCA space employed for face representation using part of the dataset (4000 images, approximately 2000 male and 2000 female). The computation needed 12 hours in a PIV 2.2 Ghz, therefore its modification is still not affordable for real-time applications unless incremental PCA is used. Therefore, the PCA space used for projection is fixed and computed off-line. The second use of the dataset was to create an off-line classifier for gender recognition, and the corresponding test set, according to Table 1.

4.1.2 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 face databases contain still images but not video streams. Video streams facial databases are quite homogeneous and contain a reduced number of individuals. For that reason we have built up a database making use of broadcast television, but also friends and volunteers recorded with different webcams and cameras without controlled conditions. The database contains around 900 different video streams (320 by 240 pixels). The sequences correspond to approximately 725 individuals, i.e. some individuals were recorded more than once. Unfortunately we do not have permission of most of them to share the database.

4.2 Experiments related to Gender Classification

Gender classification is a problem which has been recently studied with good performance using off-line computed classifiers [20]. That said, before proceeding with the learning experiments, we decided to check the number of eigenfaces needed for reliable classification for this problem

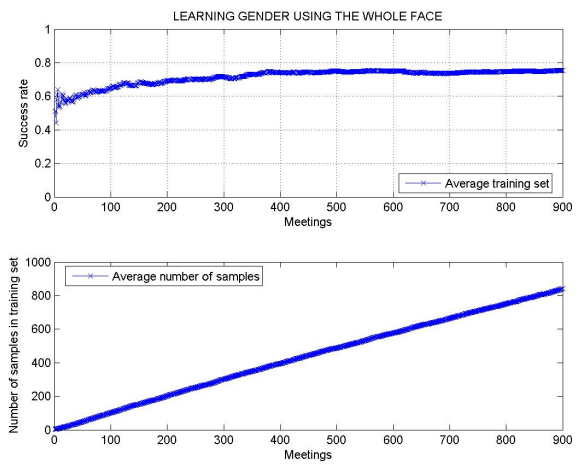

Figure 2. Average of accumulative successful rate evolution for the test set described in Table 1. The final performance is around 75%.

using an off-line computed classifier. Table 1 summarizes the composition of the test and training sets used.

The optimal value of eigenfeatures to be used in our configuration has been studied in [8]. Using the training set described in Table 1 around 40 components are required to perfectly classify that training set. For the test set, referred in Table 1, the classification performance with 70 components the rate is 80% and with 140 is hardly better than 83%. Therefore we decided to use the first 70 eigenfeatures for our learning experiments.

During the experiments, the system had 900 different meetings (500 with males and 400 with females) which correspond to approximately 730 individuals. The experiments were launched 10 times using a random order for each meeting agenda.

During a meeting, every exemplar selected is transformed to the face representation space described above and classified with the current SVM based classifier. Notice that the training set is initially empty and therefore the classifier is first computed when there is at least one sample per class.

Figure 2 presents the accumulative performance results during the classifier evolution. The test set described in Table 1 is used to compute that rate. The final average performance after 10 random trials is 75% as described above. Compared with the classifier trained with the training set described in Table 1, the performance is five points lower. Different features must be noticed in the learned classifier:

- The number of individuals met is approximately one fourth (730/2748).
- The average number of samples contained in the training set is three times lower (700/2748) (see bottom graph in Figure 2).
- The eyes were located automatically and not manually (as done for the off-line classifier).
- Not all the individuals have any sample in the training set.
- The final number of male and female samples extracted for the training set is similar.
- The exemplars added to the training set are completely independent from the samples used to compute the PCA space (This is not so for for the off-line classifier).

Figure 2 indicates an improvement tendency which is smoothed after 400 meetings. However the number of samples being added to the training set keeps growing almost linearly. It seems that the learning framework requires more meetings to show if it can reach a performance similar to the off-line classifier.

In order to test with a larger test set, we analyzed the performance using the whole dataset of static images, which are completely independent (different source and individuals) from the video streams. The results reported in Figure 3 indicate a performance lower than in Figure 2, but it must be noticed that the number of images in the test set doubled the original.

4.3 Experiments related to Identity Recognition

With respect to face identification, there are two different problems that share similar techniques. The first one is associated to recognition from a database without a priori knowledge of the person's identity. The second problem is related to verification or authentication of an identity given by a subject.

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 learned 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 is not 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,

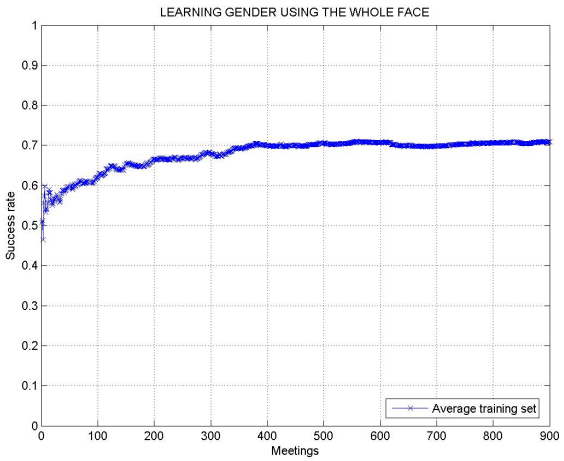

Figure 3. Average of performance evolution using as test set the fusion of training and test set described in Table 1. The final performance is around 71.7%. Notice that the test set is completely independent from the video streams data.

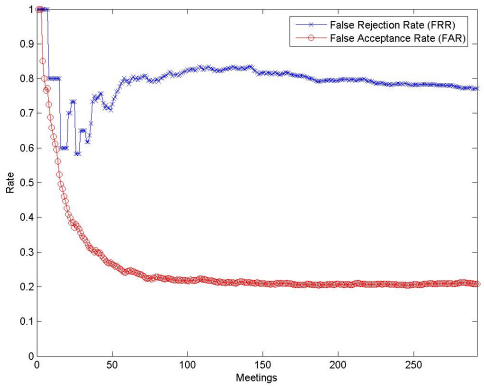

Figure 4. Results achieved for meetings with unknown individuals (FAR), 77% of the total number of meetings, and for already known individuals (FRR), 20% of the total number of meetings, along the system evolution. The final combined performance is successful for approximately the 80% of the meetings.

but do not guess if the individual's identity is not contained in the database.

To overcome the drawbacks of both systems, and to model the rejection class with available data, we decided to apply both approaches in cascade. 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. This approach forces the system to have a classifier for n classes for the first stage. Also n binary classifiers for verification are necessary; one of them is used for each processed face image.

For this experiment we selected only those identities for which more than one sequence was available. Therefore 300 sequences were employed.

For each meeting some exemplars are extracted and used first to suggest a classification based on the n-class classifier, and then verified using the binary classified of the class selected in the first stage. Incorrectly classified exemplars are added to each particular training set and used to retrain the classifiers.

Figure 4 shows the results achieved along the system evolution. These results have been averaged after 5 randomly ordered runs. For a specific meeting, the False Acceptance Rate (FAR) indicates the ratio, up to that moment, of the total number of meetings corresponding to unknown individuals which have been falsely accepted as known individuals. At the beginning the system seems not to have enough samples to model the unknown class. For that reason the error decreases notoriously until approximately 70 meetings, moment in which the error is lower than 20%. It must be noticed that for this particular experiments where each individual has at least two meetings, the likelihood of meeting new individuals is decreasing and therefore no improvement is possible after a certain point, as no new individuals are met.

On the other side, the False Rejection Rate (FRR) represents the ratio which corresponds to an already met identity which was falsely considered as unknown. These results are not good enough, approximately 77% of the identities are (incorrectly) not recognized, but it seems to have an improvement tendency. Observing in more detail the evolution for the individuals which had more encounters, see Figure 5, the improvement seems to be better than the average. These graphs indicate the accumulative performance for an identity, and show the evolution presented by the system as it is exposed repeatedly to an identity. Remember that these meetings are randomly held among those of other identities, and that an improvement for an identity is more valuable because the number of classes are also increasing with time. It is observed a clear tendency of improvement for two of the identities, while identities 1 and 60 have a more irregular behavior. It seems that they are not properly modelled, i. e. more meetings are needed to model

them. Observing the sequences, it is clear that they were recorded in particularly different conditions while the other two identities correspond to sequences extracted in different days, but keeping the illumination conditions constant (they correspond to news moderators).

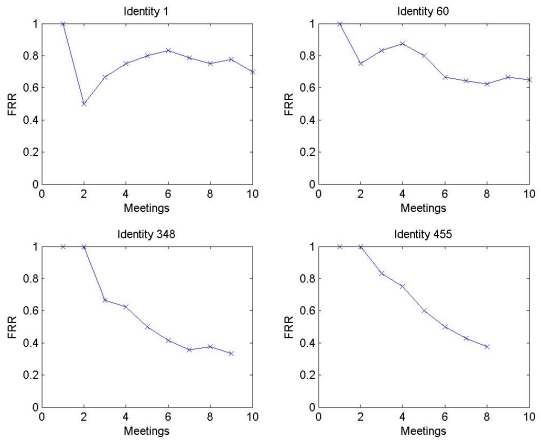

Figure 5. Accumulative FRR evolution for identities with more than seven meetings.

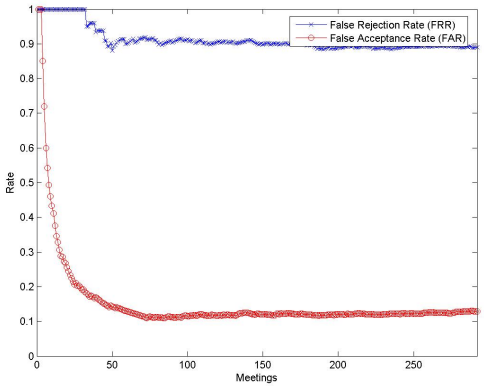

Figure 6. Results achieved for meetings with unknown individuals (FAR), 86% of the total number of meetings, and for already known individuals (FRR), 14% of the total number of meetings, along the system evolution. The final combined performance is successful for approximately the 80% of the meetings.

We also performed an experiment following the approach described in [5], where the exemplars average is used instead of them. The results presented in Figure 6 seems not to be coherent with those achieved in [5]. However it must be remembered that the authors in that paper aligned the face manually and not automatically, and therefore the average that we are using here could be misaligned, and therefore noisy for a reduced number of meetings.

These results are preliminary because the number of meetings is still reduced, but for most seen identities FRR improves and/or become stable, always better than the average shown in Figure 4. Thus, they suggest that successive meetings with an identity serve to improve the identity model.

5 Conclusions

Automatic face processing literature is vast in the still images context. However, the development in recent years of face detection approaches provides less restricted data to study the real problem of face processing. In this work our aim was to study the evolution of a system which is previously able to detect faces and to project them into a certain representation space.

We have presented a system which automatically detects faces in video streams, select some exemplars in real-time, and use them to suggest a label for gender and identity classification. Collaborating with an expert for those tasks, a human, the system evolves from an empty memory building training sets for each problem based on its experience. Incorrectly classified exemplars are used to iteratively tune the system classifiers.

The performance evolution for gender and identity classification is of different nature. For gender we have compared the system with an off-line precomputed classifier achieving a not so lower performance with less face exposure. For identity the system seems to start to be able to distinguish unknown subjects, but it seems to require more meetings with individuals to get a better recognition performance, i.e. to make them familiar.

That said, we are optimistic about the future evolution of the system, but much more experience seems to be necessary to reach a final conclusion about the idea of creating a facial classifier which is tuned live. Therefore, future work will focus on gathering a larger video streams database with the aim to confirm that hypothesis in order to shift to the second epoch. As for incremental learning, currently the classifier retraining does not fit real-time restrictions for the largest numbers of meetings that we have managed. Indeed the gender classifier retraining requires 2 seconds and the identity classifiers take 600 msecs.

Acknowledgments

Work partially funded by research projects Univ. of Las Palmas de Gran Canaria UNI2003/06 and UNI2005/28, Canary Islands Autonomous Government PI2003/160 and PI2003/165 and the Spanish Ministry of Education and Science and FEDER funds (TIN2004-07087).

References

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

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

[3] 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 J. Kittler and M. Nixon, editors, *Proc. Audio- and Video-Based Biometric Person Authentication*, pages 625–638, Berlin, June 2003. Springer.

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

[5] A. M. Burton, R. Jenkins, P. J. Hancock, and D. White. Robust representations for face recognition: Te power of averages. *Cognitive Psychology*, 51(3):256–284, 2005.

[6] A. M. Burton, S. Wilson, M. Cowan, and V. Bruce. Face recognition in poor quality video: Evidence from security surveillance. *Psychological Science*, 10(3), 1999.

[7] M. Castrillón Santana, O. Déniz Suárez, M. Hernández Tejera, and C. Guerra Artal. Real-time detection of faces in video streams. In *2nd Workshop on Face Processing in Video*, Victoria, Canada, May 2005.

[8] M. Castrillón Santana, J. Lorenzo Navarro, D. Hernández Sosa, and Y. Rodríguez-Domínguez. An analysis of facial description in static images and video streams. In *2nd Iberian Conference on Pattern Recognition and Image Analysis*, Estoril, Portugal, June 2005.

[9] R. Chellappa, C. Wilson, and S. Sirohey. Human and machine recognition of faces: A survey. *Proceedings IEEE*, 83(5):705–740, 1995.

[10] B. de Gelder and R. Rouw. Beyond localisation: a dynamical dual route account of face recognition. *Acta Psychologica*, (107):183–207, 2001.

[11] 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*, pages 104–110, Singapore, 2004.

[12] R. W. Frischholz and U. Dieckmann. Bioid: A multimodal biometric identification system. *IEEE Computer*, 33(2), February 2000.

[13] C. Guerra Artal. *Contribuciones al seguimiento visual precategórico*. PhD thesis, Universidad de Las Palmas de Gran Canaria, Octubre 2002.

[14] R. Kemp, N. Towell, and P. G. When seeing should not be believing: Photographs, credit cards and fraud. *Applied Cognitive Psychology*, 11(3):211–222, 1997.

[15] Y. Kirby and L. Sirovich. Application of the karhunen-love procedure for the characterization of human faces. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(1), July 1990.

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

[17] 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)*, pages 157–164, October 2003.

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

[19] 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.

[20] B. Moghaddam and M.-H. Yang. Learning gender with support faces. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(5):707–711, 2002.

[21] 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.

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

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

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

[25] P. Sinha and T. P. Torralba. I think i know that face... *Nature*, 384(6608):384–404, 1996.

[26] A. Tolba, A. El-Baz, and A. El-Hardy. Face recognition: A literature survey. *International Journal of Signal Processing*, 2(1):88–103, August 2005.

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

[28] M. Turk and A. Pentland. Eigenfaces for recognition. *J. Cognitive Neuroscience*, 3(1):71–86, 1991.

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

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

[31] G. Wallis and H. Buelthoff. Learning to recognize objects. *Trends in Cognitive Sciences*, 3(1):22–31, January 2000.

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

[33] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM*, 35(4):399–458, 2003.