Face Exemplars Selection from Video Streams for Online Learning

M. Castrillón-Santana, O. Déniz-Suárez, and 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 tackles the problem of online acquisition of exemplars for dynamic updating of classifiers for facial analysis. Most facial analysis systems apply a previously computed classifier to a set of images, or recently to the output of real-time face detection systems. Here we describe an approach to select significant detected faces during interactive sessions in order to learn and modify online, with the initial help of an expert, a classifier for a given task. Preliminary experiments are reported related to gender recognition.

1. Introduction

In recent years different face detection approaches suitable for real-time applications have been described. However, most facial analysis systems are applied to still images, or they simply translate the approach designed for still images to video sequences. In those systems, the classifiers employed were typically computed off-line with a given training set, and later analyzed with different test sets. For example, a well known corpus used to evaluate recognition techniques is the FERET database [20], while the BANCA protocol has been designed for verification approaches [2].

In this work, we aim to move from the still image context, with the challenge of developing abilities for more natural and comfortable Human Computer Interaction (HCI) [19]. Any Vision Based Interface [24] must include face analysis in order to use computer vision technology to perceive the user in a HCI context. Therefore, it is assumed for these interfaces that a camera is continuously acquiring images, which can of course register individuals close to the system. In that context, where non invasive techniques are required for facial description, typical approaches are inappropriate as stated by different authors [14, 16, 23]. In that situation, the large number of faces collected by the face detection system must be processed considering temporal coherence, i.e. the representation and/or classification of individuals should be evaluated in time rather than using an one-shot methodology.

A problem rarely considered in the literature is the online learning or updating of any required classifier during system *life*, based on the system experiences. This ability is crucial for social beings. A successful system in this line would have great applications in the perceptual interfaces context. Can this 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 face this purpose.

1.1. Previous Work

Video stream analysis presents a major difference in relation to still image processing: Individuals present variations along the image stream. An object model seems to require a collection of images similarly to the way the human system does [28]. The source to set up these image collections are the interactive sessions that an automatic system has with the particular object. Focusing on face analysis, it is not reliable to use all the images present in a video stream to represent an specific facial class. It is obvious, that there is redundancy contained in them, and their use would produce massive computational and storage costs [1, 29].

The extraction of significative patterns, or exemplars, is tackled in [14]. 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 [29]. 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 [1] the authors implemented in an 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

Proceedings of the Second Canadian Conference on Computer and Robot Vision (CRV'05)
0-7695-2319-6/05 $ 20.00 **IEEE**

IEEE COMPUTER SOCIETY

made of the standard eigenface method [25], 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 justify this fact due to the known performances degradations of the eigenface approach for facial expressions, facial alignment and scale.

The authors of another system [9], 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.

2 Face Recognition

The literature referred to humans describes different models for the problem of face recognition. A recent focus described in [7] suggests a dual route model for face recognition, instead of the previous sequential or hierarchical models presented by other authors. Attending to the 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. As mentioned above, most automatic systems simplify this exposure to an off-line and fixed training stage. However, in this paper we are interested in the design of a system architecture able to accomplish online this extensive learning stage.

Taking into account these considerations, we tackle the problem considering two modules related with face detection and face identification. Initially, as we do not have an object recognition system available, we simplify the de Gelder model, designing the system to provide an identification suggestion only with the data provided by the face detection module.

As seen in Figure 1, the face detection module provides face detection data, which is later simplified by means of the exemplars selection. Once the exemplars have been selected, those contained in the temporal Window Of Attention (WOA) are used to suggest a classification. Later, classification errors are used by the system to recompute the classifier online. Using this system architecture, the face identification modules are dynamic while the face detection and representation modules are considered fixed.

In order to detect classification errors, we distinguish different epochs similarly to [1]. Both stages will allow the system to tune different classifiers by means of interactive sessions, i.e., during the system *life*.

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 not able to describe them. Humans first recognize the face class, with the different considerations about the way this knowledge is achieved [3], and later we are in most cases educated to learn to distinguish different subclasses within face class [7]: males/females, young/mature/old, familiar/unfamiliar, etc. Some of these classifications present clearer borders than others, and for some of them we do not use only facial details but also the context [21].

On the second epoch, once the expert realizes that the system confidence is good enough, the system will be able to unsupervised select using temporal coherence in the WOA, the misclassified patterns.

3 The Approach

The problem that we are focusing in this paper is the online apprehension or acquisition of the significant data for any classifier useful for facial analysis. It is assumed that the only information used will be the normalized faces provided by a face detector system described below. The number of classifiers to be learned will be provided a priori, and for some of them also the number of classes. This is the case of gender classification, but not for identity for which a system would not have a predefined number of classes. Observing Figure 1, we will introduce briefly the face detection technique used, describe the techniques used for exemplars selection, and explain the way in which the system decides whether a classifier must be updated.

3.1. Real-Time Face Detection

It is assumed that the system would perform the learning online, therefore, a fast face detection approach is required to provide the image stream. For that purpose we employ

IEEE COMPUTER SOCIETY

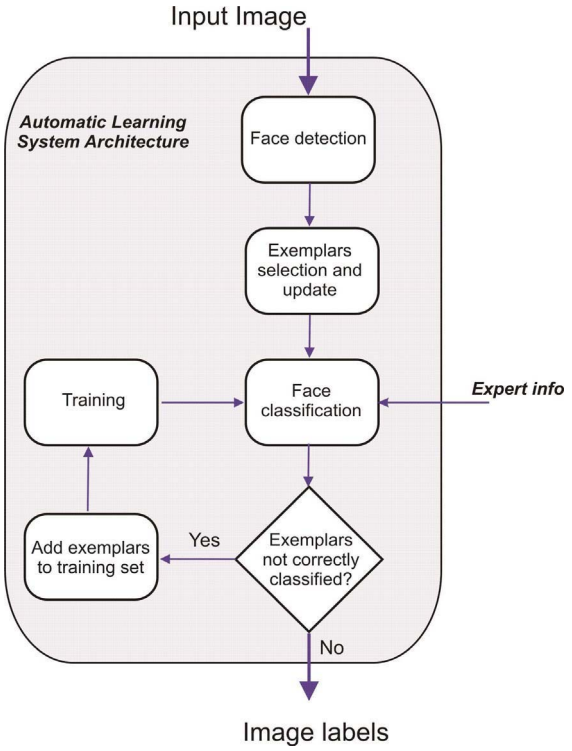

Figure 1: Graphical overview of the system.

a face detection system which provides real-time multiple face detection at different resolutions.

The detector, see [5] for more details, makes use for the first detection or after a failure, of two window shift detectors which provide acceptable processing rates. These brute force detectors are based on the general object detection framework by Viola and Jones [27]. They are the frontal face detector described in that paper, and the local context based face detector described in [15]. The last one achieves better recognition rates for low resolution images if the head and shoulders are visible.

A monolithic approach, considered as an approach that uses a single schema, presents some disadvantages for video stream processing, due to the fact that it is not making use of a main cue included in video streams, i.e. temporal coherence cue which can reduce processing cost and improve detection robustness. Indeed any face detected in a frame provides information useful to speed up the process in the next frames. Therefore, for each detected face, the system stores not only its position and size, but also its average color using red and green normalized color space [30].

Skin color based approaches for face detection have a lack of robustness for different conditions. A well known problem is the absence of a general skin color representation for any kind of light source and camera [22]. However, the skin color extracted from the face previously detected by the Viola-Jones framework can be used to estimate facial features position. The skin color blob provides valuable information to detect eye positions for frontal faces as described in [6].

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

- Eye tracking: A fast tracking algorithm [12] is applied in an area that surrounds previously detected eyes, if available. The tracker makes use of a fixed pattern size for both eyes, and searches the minimum difference in the search area as follows:

$$D(u,v) = \sum_{Area} |I(u+i, v+j) - P(i,j)| \qquad (1)$$

- Face detector: The Viola-Jones face detector [27] 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 [15].
- 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 [12].

These techniques are applied following this order, until one of them finds the searched face. Whenever a face is detected, the skin color cue is used for facial features detection. 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. Stable Patterns 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. The system relates these consecutive detections in terms of position, size and pattern matching techniques, allowing additionally brief gaps during the session. Thus, for each detection

IEEE COMPUTER SOCIETY

thread, the face detector system provides a number of facial samples, $dt_p = \{x_1, ..., x_{m_p}\}$. Ideally a detection thread contains samples detected from a single individual. However, different detection threads can correspond to the same individual, which is not checked by the current face detector.

Additionally, multiple detection threads can be active at a time, and the system stores those already lost. A detection thread is considered lost if it is not related with new detections after some frames. In the current implementation no identity recognition is integrated in order to fuse different detection threads which correspond to the same individual. The system is just able to guess continuity for a detection based on the recent position, size, color and eyes appearance. This fact can in some circumstances mix different individuals in a single detection thread, but this artifact is not covered in this paper.

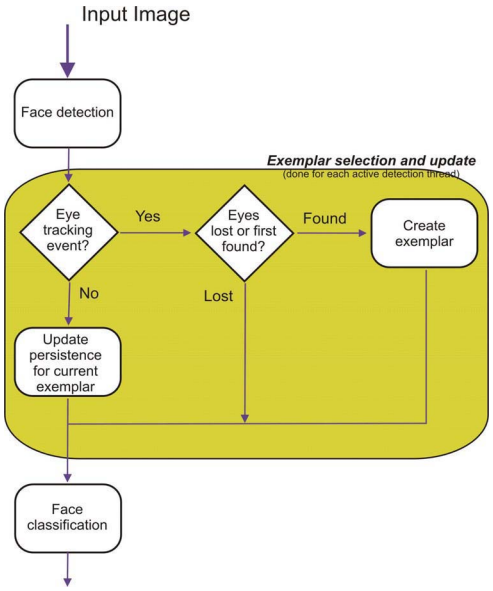

Figure 2: Detailed view of the *Update exemplars* task, see Figure 1

The huge amount of data extracted during an interactive session must be reduced in some way to avoid information redundancy. As explained above, the face detection system provides a set of detection threads, which consist in a collection of images associated to a detected individual. From this collection, some selected patterns, the exemplars $e_p = \{e_1, ..., e_{s_p}\}$, are extracted for each detection thread, dt_p. Later, the exemplars are employed for classification and classifiers tuning.

The criteria for selecting stable samples or exemplars, have been selected in order to be easily integrated in the detection process, similar to [29]. For that reason, it is based on the eye tracker events of a tracker which is integrated in the face detection system. A graphical description is presented in Figure 2.

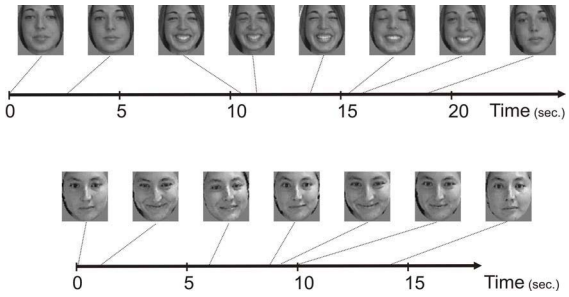

Figure 3: Stable patterns or exemplars extracted from two different detection threads. Dot lines indicates the moment in which they were extracted during interaction.

A tracking failure shows an evidence of a 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. After detecting the eyes by a cue different from the eye tracking, the first face detected in the next frames by the eye tracker 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 data provided by the normalized detected face, x_j, its persistence, pe_j, and its timestamp, t_j, i.e. $e_j = \langle x_j, pe_j, t_j \rangle$. In Figure 3, the exemplars extracted automatically for two individuals during sessions of more than 15 secs. are presented.

Given an interactive session, IS, for any old enough detection thread (older than 20 frames), dt_p, 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 the relative persistence in relation to the total persistence of the detection thread. This is expressed as:

$$P(C_k|dt_p) = \frac{\sum_{j=1}^{s_p} P(C_k|e_j) * pe_j}{\sum_{n=1}^{s_p} pe_n} \quad (2)$$

Therefore, the system is able during an interactive session, or summarize at the end, to suggest a likely class for each detection thread. This value can be computed for the exemplars extracted during the whole interactive session, or for those which have been selected within a recent Window Of Attention (WOA). In that case, in Equation 2 only those exemplars inside the WOA will be considered.

IEEE COMPUTER SOCIETY

3.3 Recognizing classification errors

The initial stage of supervision is provided by an expert, which has the faculty of correcting the system after it has suggested a class for a detection thread, see 3.1. This would be similar to how humans learn, for gender recognition we are said at the beginning who is male and female, and later we are able to tune our final classifier from successive experiences. Not correctly labelled samples will be logically used to update and correct the up to date classifier. Once the system is reliable, it can request for supervision only for doubtful situations.

But there is another mechanism which will allow the system to learn during its *life* once it is not supervised. For example, for identity recognition it is crucial to detect unknown individuals [1, 4, 9], i.e. individuals which are not already contained in the classifier. The ability of distinguishing an unknown individual allows the system to ask the identity to the user, and therefore to create a new identity class. During our babyhood we perceive different facial patterns which are fixed by reinforcement. How many times a baby hears "Who is he/she?". However, children easily forget a previously known model if there is no reinforcement during a long period, so many faces fall easily in the unknown class. Later we are able to retain, normally, our learned models longer, and we easily realize when a new face is present, e.g. an unknown individual. In that sense the human system is also self supervised once it is reliable; we are introduced in some way, refreshed when we have forgotten someone, or we inquire a name.

Figure 4: Normalized face, and eyes area used for experiments in section 4.1.

In any of those configurations, wrongly classified exemplars are used to retrain the classifier. For example, when the classifier is still in its learning stage, an expert will supervise the class assigned to the detection thread. If the system were corrected, and the correct class were C_c, all the incorrectly labelled exemplars, i. e. $P(C_c|e_j) = 0$, will be added to the classifier iteratively whenever the recomputed classifier keeps classifying them wrongly. If the supervisor confirmed the class suggested, C_k, similarly incorrectly assigned exemplars, $P(C_k|e_j) = 0$, will be added iteratively to the classifier.

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 provide additional exemplars or the training set if they were incorrectly classified. Therefore, the classifier evolves according to its perceptual experience, i. e. it is not previously fixed. This focus is well suited for contexts like identity where the individual facial appearance changes in time, a fact which could not be completely tackled by a fixed training set.

4. The application: Gender Recognition

The previous section has described the system architecture, giving emphasis in the automatic selection of exemplars from video streams in order to suggest classification results which can later be used to update the classifier.

In order to perform experiments, we have found two different problems: 1) Gathering sequences of several individuals with different appearances is not a simple and fast task, and 2) sequences are typically short. If we had the intention to experiment with individual identities we would require interactive sessions which take place in different days for each individual, and readers would not accept a reduced number of individuals. Thus, the experimental setup would be particularly expensive in time.

These restrictions have forced us to perform a first test on a prototype problem for which we have the best balanced training and test set, with a sufficient number of samples per class. For those reasons we have selected a semantic classification problem with a static number of classes, such as gender recognition.

However, we must point out that with this learning schema in mind, whatever classification technique which could potentially provide good performance can be used in order to tune it based on the system experience, i.e., its interactive sessions.

Gender recognition is a task which has been recently studied with good performance using precomputed classifiers [18]. The image areas analyzed in the experiments are two different, see Figure 4: 1) the whole normalized image, and 2) a reduced area centered on the eyes, attending to psychophysical evidences for gender recognition in humans [11].

4.1. Representation and Classification

4.1.1 Representation space

According to this schema, we select first, similarly to [14], a well known face representation space in advance: the PCA space due to its economical advantages [13]. On this representation space, a Support Vector Machines (SVM) classifier [26] will be modified during system *life* tuning its output for each classifier considered. This combination

IEEE COMPUTER SOCIETY

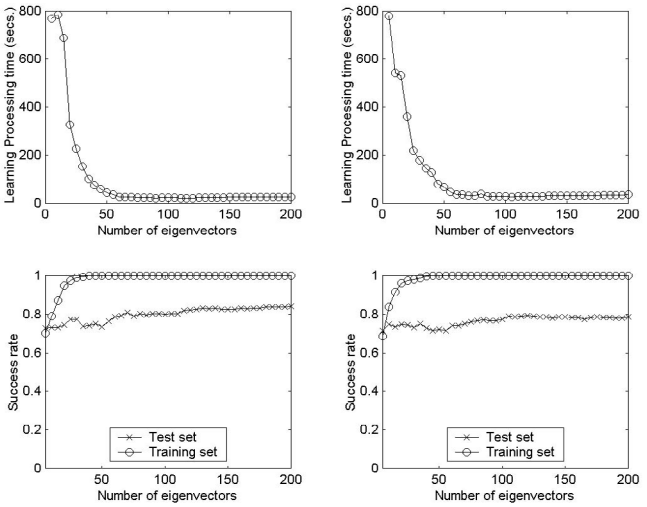

Figure 5: Gender recognition results computing a SVM based classifier considering a different number of eigenvectors (from 5 eigenvectors, up to 200). Top) Model training time, down) success rates. Left column corresponds to results using the whole normalized image, right column reflects the results using a reduced area around the eyes (see Fig. 4).

$PCA+SVM$ has been chosen for being well known by the community and for the good performance results achieved.

Before proceeding with the learning, we decided to check the number of eigenfaces needed for reliable classification for this problem. To define the PCA space, we have annotated the eye positions for 6000 (4000 males and 2000 females) faces taken from internet and selected samples from facial databases such as the BIOID [10]. This images have been normalized according to eye position obtaining 59×65 samples. To have a balanced (male/female) PCA space approximately 4000 of them were used to compute the PCA space. This task needed 12 hours in a PIV 2.2 Ghz, therefore its modification is still not affordable for real-time applications. Therefore, the PCA space used for projection is fixed and computed offline.

4.1.2 The optimal number of eigenvalues

Once the PCA space was computed, a training and a test set were created. The training set is composed by randomly selected 1523 male and 1223 female samples, and has been used to compute different SVM classifiers with a minimum number of 5 eigenvectors, up to 200. The test set contains 835 female and 2246 male samples.

The top plots of Figure 5 reflect the time needed for training the different classifiers, resulting similar for both images areas used, i.e. the whole image and the eyes area. It must be mentioned that classification meets real-time requirements (1 msec. with 50 eigenvalues and 5 with 200). The fastest learning times are achieved using more than 70 eigenvalues, needing 20 seconds (reasonable for online performance). The high cost of the classifiers that use a low number of eigenvalues seems to be justified as those eigenvalues provide not enough information to separate the two classes considered. The bottom graphs reflect the performance for the training and test set with a minimum number of 5 eigenvectors, up to 200 eigenvectors. It can be seen that the training set needs around 40 components to be perfectly recognized, however the test set reduces its progressive improvement in the range from 70 to 120 components. With 70 components the rate is 80% and with 140 is hardly better than 83%. The results are a little bit lower using only the eyes area, reaching 75% with 70 eigenvalues. This fact reflects the big amount of information contained in that area, or the redundancy contained in the whole face image for this problem.

These results prove that the training set used does not contain enough information. One option would be to increase the number of training samples, but it is likely that some samples already included are redundant. We can previously consider to reduce its size by removing redundant samples by means of prototype selection [8], or try to learn the classifier progressively, as we propose in this paper.

4.2. Learning online. Experimental results

The experiments here described were performed with a system which was initially only able to detect faces, to represent them in a face space (PCA), and had an empty training set. Observing the experiments commented in the previous section, 70 eigenvalues were taken as enough to create a decent gender recognition classifier. As the system is still naive or infant, a human expert is used to correct classifications in this bootstrap, considering the two classes in this problem: male/female.

The system was successively introduced to 50 different individuals acquired by means of a standard webcam without any controlled condition. The order has been randomly selected, but we have tried to present alternatively a male and a female to allow the classifier to grow balanced in these early stages. Due to the short duration of these sequences, the WOA considered the whole sequence for classification in these experiments.

The initial empty training set is later filled by the exemplars extracted during interaction. The test dataset used to check the dynamic classifier contains the whole set of 6000 faces (i.e. both datasets, training and test, described in Section 4.1.2). Figure 6 shows the progressive performance of the system applied to the problem of gender recognition.

IEEE COMPUTER SOCIETY

On one side, the final performance for the dynamic classifier is above 68% using the whole face, and 62% using the eyes area (for the whole combination of test and training sets used in Section 4.1, i.e. aprox. 6000 images). On the other side, different batch computed classifiers achieved an average performance of 62%, i. e. aprox. 6 points lower than the learned classifier.

The results do not present a clear improvement after $10 - 14$ meetings. Indeed the learned classifier seems to be sensitive to the addition of new individuals due to the still reduced number of individuals met by the system, and therefore presents some oscillations, but these are reduced in comparison with the performance after the first meetings. This can also produce that the tendency of improvement seems to be still not evident (whole face) or slight (eyes area). The resulting training set is perhaps not optimal, but it is based on the sequential system experience. The performance is relatively far from the rate achieved by the classifier computed in the previous section. However, it must be observed that the classifier computed using 2746 samples contains 34 times more samples (54 times more individuals). It must also be noticed that there is not perfect alignment, the test set was annotated by hand while training set was built up with the more imprecise face detector. In any case, we consider that it is necessary to introduce much more individuals to the system in order to build more powerful system.

Providing some details, using the whole normalized face, the final automatic learned classifier gathers just 37 female samples and 43 male ones. These images do not correspond to all the individuals introduced, indeed they only represent 16 females and 19 males. This is due to the fact that during the learning process, some detection threads were correctly classified and therefore they did not alter the training set. Others needed more than one sample to be added into the training set. During these interactive sessions the computation of the SVM classifier needed around $10 - 20$ msecs., therefore the process is suitable for real-time. This cost will be increased adding more samples (individuals) to the training set, however we know that for 3000 (the batch training set) the time needed was 20 secs.

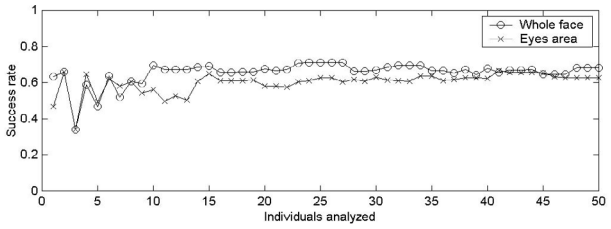

Figure 6: Performance during learning.

5. Conclusions

We have presented an approach which is suitable for online and real-time acquisition of facial data based on system experience. The approach integrates in a real-time face detector system the ability to select exemplars based on the eye tracker enclosed in the system. These exemplars are first projected on a fixed PCA space and later used to learn from scratch a facial based classifier.

Due to the lack of a large database of facial stimuli similar in size to the one that a human can perceive during his childhood. A prototype problem, gender recognition, has been selected to test the possibility of learning a classifier during system *life*. The system met with 50 different individuals for less than a minute. The online facial information extracted during those meetings was used to dynamically update the training set used for the gender classifier.

The progressively updated gender classifier was tested with 6000 independent images of different individuals achieving after the last meeting a success rate of 68%. This rate is not great, but could a human be able to classify better under those conditions? That is something that we do not know. At least these preliminary experiments report that the system performance seems to be slowly increasing after the initial unstable phase. Certainly the prototype results indicate clearly that the system is still in its *infancy*. In other words, it still requires human supervision before becoming reliable as the system performance is still 15 points lower than the baseline considered, i.e. an off-line classifier computed with randomly selected single images of 3000 different individuals.

In order to perform further experiments we have to gather a larger video stream database with also more individuals and maybe multiple sessions per individual. The comparison with other exemplar selection approaches or classifiers is possible but seems to be reduced if previously the system experience, i.e. the video streams database, is not enlarged.

Future work must also focus on the application of the approach to other semantic classification problems. The application of the approach to other problems seems to be straightforward, except by the fact that for a main problem such as identity recognition, the system must be able to detect autonomously unknown individuals, or what related with other descriptors would mean that the system must be able to integrate the ability to detect novelties, which would be translated to a new class (a new identity in a identity recognizer) or a new facial descriptor (a new feature not previously considered by the existing classifiers, e.g. red beards).

In this sense, we also consider that the system must integrate in short time and incremental PCA representation space in order to be able to represent those novelties [17].

IEEE COMPUTER SOCIETY

Acknowledgments

Work partially funded by research projects Univ. of Las Palmas de Gran Canaria UNI2003/06, UNI2004/10 and UNI2004/25, Canary Islands Autonomous Government PI2003/160 and PI2003/165 and the Spanish Ministry of Education and Science and FEDER funds (TIN2004-07087).

References

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

[2] 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.

[3] Vicki Bruce and Andy Young. *The eye of the beholder*. Oxford University Press, 1998.

[4] M. Castrillón, E. Grosso, and O. Déniz. Who are you? In *Proceedings of the 17th International Conference on Pattern Recognition*, Cambridge, UK, August 2004.

[5] 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.

[6] M. Castrillón Santana, F.M. Hernández Tejera, and J. Cabrera Gámez. Encara: real-time detection of frontal faces. In *International Conference on Image Processing*, Barcelona, Spain, September 2003.

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

[8] P. A. Devijver and J. Kittler. *Pattern Recognition: A Statistical Approach*. Prentice-Hall, Englewood Cliffs, New Jersey, 1982.

[9] Jun Fan, Nevenka Dimitrova, and Vasanth 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.

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

[11] F. Gosselin and P. G. Schyns. Bubbles: a technique to reveal the use of information in recognition tasks. *Vision Research*, pages 2261–2271, 2001.

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

[13] 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.

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

[15] Hannes Kruppa, Modesto Castrillón Santana, and Bernt 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.

[16] 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.

[17] Javier Melenchón, Lourdes Meler, and Ignasi Iriondo. On-the-fly training. In *Articulated Motion and Deformable Objects: Third International Workshop, AMDO 2004*, pages 146–154, September 2004.

[18] Baback Moghaddam and Ming-Hsuan Yang. Learning gender with support faces. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(5):707–711, 2002.

[19] Alex Pentland. Looking at people: Sensing for ubiquitous and wearable computing. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, January 2000.

[20] P. Jonathon Phillips, Hyeonjoon Moon, Syed A. Rizvi, and Patrick J. Rauss. The feret evaluation methodology for face recognition algorithms. TR 6264, NISTIR, January 1999.

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

[22] Moritz Storring, Hans J. Andersen, and Erik Granum. Physics-based modelling of human skin colour under mixed illuminants. *Robotics and Autonomous Systems*, 2001.

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

[24] M. Turk. Computer vision in the interface. *Communications of the ACM*, 47(1):61–67, January 2004.

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

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

[27] Paul Viola and Michael J. Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition*, 2001.

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

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

[30] Christopher Wren, Ali Azarrbayejani, Trevor Darrell, and Alex Pentland. Pfinder: Real-time tracking of the human body. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(7), July 1997.