Hosted by
BMVA

IAPR

ICPR2004

17th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
CAMBRIDGE, UNITED KINGDOM, 23-26 AUGUST 2004

IEEE COMPUTER SOCIETY

IEEE

ISBN: 0-7695-2128-2
IEEE Computer Society Order Number: P2128

Produced by InControl Productions, Inc.
80 Garden Court • Suite 260 • Monterey • CA • 93940
Phone: (831) 657-2424 • Fax: (831) 657-2428
www.incontrolproductions.com

ORIENTATION

Welcome to the 17th International Conference on Pattern Recognition, on CD-ROM. This material was published using Adobe® Acrobat technology. Included on the CD-ROM are versions of Acrobat® Reader 5.0 for Microsoft® Windows™ and Apple® Macintosh,™ and Acrobat Reader 4.0 for UNIX®.

This CD-ROM is a hybrid disc which allows PC, Macintosh, and UNIX users to share the same directory structure and access common files.

GETTING STARTED

The START.PDF file, located in the root directory of this CD-ROM, contains materials for the 17th International Conference on Pattern Recognition. Once you have installed Acrobat, open the START.PDF file to find the files that interest you.

INSTALLATION

To view the files on this CD-ROM, you must first install the Adobe Acrobat Reader version appropriate for your platform. Installation instructions can be found in the file called README.TXT in the root directory.

4P.We-ii Face and Gesture

4O.We-iii Multimodal Processing

Who Are You?

Modesto Castrillón-Santana
IUSIANI
Univ. de Las Palmas de G.C.
35017 Spain
mcastrillon@iusiani.ulpgc.es

Enrico Grosso
DEIR - Univ. of Sassari,
Via Sardegna, 58
07100 Sassari, Italy
grosso@uniss.it

Oscar Déniz-Suárez
IUSIANI
Univ. de Las Palmas de G. C.
35017 Spain
odeniz@dis.ulpgc.es

Abstract

Most automatic recognition systems are focused on recognizing, given a single mug-shot of an individual, any new image of that individual. Most verification systems are designed to authenticate an identity provided by the user. However, the previous work rarely focus on the problem of detecting when a new individual, i.e. an unknown one, is present. The goal of the work presented in this paper deals with the possibility of providing the system with basic tools to detect when a new individual starts an interactive session, in order to allow the system to add or improve an identity model in the database. Experiments carried out with a set of 36 different individuals show promising results.

1. Introduction

The face has been an object of analysis by humans for centuries. Facial expressions and facial details are the core of human-to-human communication [8]; they convey to humans a wealth of social signals, and humans are expert at reading them. Identity, gender, age, emotions are in our experience directly related to facial features. The extraordinary ability to decode these signals allows humans to react on the basis of the visual appearance of an individual and very often to derive information about character and personal attitudes.

Faces seem to be the best non invasive means for doing recognition/authentication as humans do everyday. This fact is not new to the scientific community. In a survey published at the beginning of the 90's [11] the face was selected as the main discriminant element for most automatic systems designed for recognition.

Automatic recognition systems based on visual information have been widely studied in the last decade [4, 11]. However, most systems described in the literature are focused on recognizing, given a single mug-shot of an individual, any new image of that individual provided to the system. The problem is so complex that restrictions are commonly applied to pose, illumination, etc. A well known corpus used to evaluate these techniques is the FERET database [10]. A large number of approaches, designed for database extraction, can obviously be tested against the FERET dataset. However, it is also clear that for Human Computer Interaction, where real-time non invasive recognition is required, these approaches could be unappropriate.

In fact, as recently pointed out by Torres and Vilá [12], in the context of video streaming, conventional recognition schemes are not well suited. In this context, a huge amount of images are provided to the system. These images must be processed considering temporal coherence and representation/classification of individuals should be evaluated in time rather than using a one-shot methodology.

This paper tackles the recognition problem in the context of video streaming with a basic goal which deals with the possibility of providing the system with basic tools in order to understand when a new individual starts an interactive session or if the appearance of an individual already contained in the training set is not properly modelled.

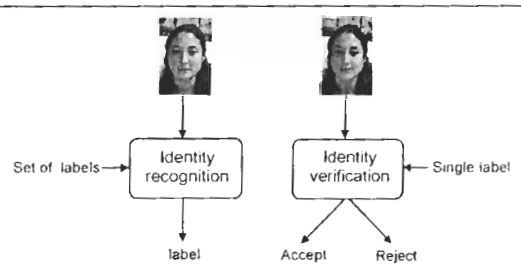

Figure 1. Recognition (left) and Verification Schemas (right).

0-7695-2128-2/04 $20.00 (C) 2004 IEEE

2. System Description

2.1. 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. The second problem is related to verification or authentication of an identity given by a subject, see Figure 1.

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 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 2-class 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 individual identity is not contained in the database., unless a rejection class.

To overcome the drawbacks of both systems, and to model with available data the rejection class, we decided to apply both approaches in a cascade manner. 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 2.

This approach forces the system to have a classifier for n classes for the first stage. Also n 2-class classifiers for verification are necessary; one of them is used for each processed face image.

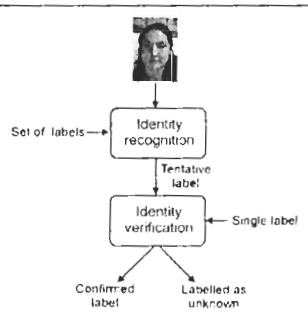

Figure 2. Identity recognition plus verification.

2.2. Representation Space and Classification

The face image to be analyzed has a high dimensionality, feature that makes the classification problem hardly tractable. In order to avoid this problem, Principal Components Analysis (PCA) decomposition [7] is applied to the training data provided. This action allows us to represent the appearance of the different individuals contained in the training set [13].

Using this representation space, different classifiers can be used to select a label for each face processed. The original implementation [13] makes use of Nearest Neighbor Classifier (NCC) for that purpose. However, different authors argued that the low reliability provided by this approach mainly if lighting conditions are not restricted. Under these circumstances PCA+NNC reduces its performance according to [2].

Recent developments use local representations such as Independent Components Analysis (ICA) [1] to get a better representation space. However, the work described in [6] proved that using any of both different representation spaces, PCA or ICA, and powerful classification criteria such as Support Vector Machines (SVMs) [14], which perform well with high dimensional data, instead of naive NNC, reported similar recognition rates. Thus, this work concluded that the classification criteria selection was more critical than the representation space used. According to these results, recognition experiments have been carried out using SVMs as classification criteria.

3. Experiments

3.1. Video Streams Data Set

In the video streams context, a main problem is the absence of standard video stream databases with the complexity typical of HCI environments. Most facial databases do not contain sequences offering the facial evolution of different individuals. The availability of an illumination controlled and background restricted database such as XM2VTS [9] is not well suited to verify the unrestricted problem tackled in this paper.

Due to that reason, the data set used to carry out the experiments presented in this document contains different video streams that have been acquired and recorded using different standard webcams. These sequences were taken on different days without special illumination restrictions. Therefore, some were taken with natural (therefore, variable) and others with artificial illumination. The sequences cover different gender, face sizes and hair styles. They were taken at 15 Hz during $15 - 30$ seconds, i.e., each sequence contains from 210 to 450 frames of 320×240 pixels. All the frames contain at least one individual in unrestricted pose,

i.e., there is a face in each frame but not always frontal. In the experiments considered, only the most salient frontal face detected using the real-time face detector described in [3] was analyzed.

Among the 36 sequences acquired, 23 of them have been selected to perform identity recognition experiments. This subset is used to model the identity of the 23 individuals contained in them. The second set, i.e. 13 sequences, contains individuals not included in the training set.

A single pattern is selected (the first one detected by the face detector in the sequence) from each of the 23 sequences to model the individual appearance for each individual. Those selected patterns are used to compute PCA. This PCA space defines the SVM based classifier used in this implementation. Any new face detected is projected to this PCA space and soon after classified.

3.2. Recognition Experiments

Figure 3 presents success and error rates for the 23 sequences which contain an individual included in the training set. As mentioned above these known users are modelled using a single pattern. The average success recognition rate was 0.8614, therefore the average error rate was 0.1384 processing around 4500 images with a training set of 23 patterns. It must be observed that the process is applied to video streams, therefore a simple temporal coherence criteria based on voting [5] would improve the system performance.

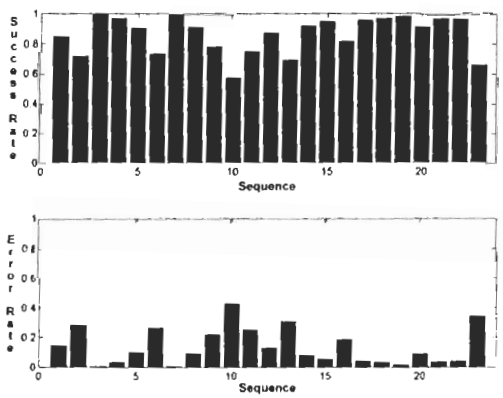

Figure 3. Success and error rate comparison.

3.3. Recognition plus Verification

The approach described in the previous section uses a single pattern to model each individual appearance. This training set defines a PCA representation space where a classifier based on SVMs provides an error rate of 0.1384. A main drawback of this approach is that if a new individual is processed by this classifier, the results are absolutely erroneous as the system is unable to recognize that situation. As mentioned previously, the integration of a second classifier using the verification approach is suitable to confirm an identity. It must be pointed out that the main training cost is in the PCA computation, which must be done also for the single multiclass classifier.

The results of the application of this cascade classifier to the video streams containing the individuals are presented in Figure 4. The first classifier applied is a n-class classifier that provides a label. This label is used to select one of the bi-class classifiers which will allow the label verification. The results after the application of two classifiers in a cascade fashion decreases both success and error rates. On the one hand, the success rates are slightly reduced (as a new filter is applied), but on the other hand the error rates are almost eliminated. Using this approach the average success rate is 0.7213 and the average error rate is 0.0154. The rest are considered as unknown by the system.

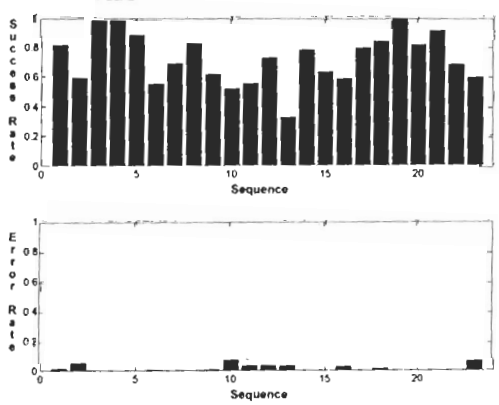

Figure 4. Success and error rate comparison for sequences that contain an individual included in training set.

This approach also offers the possibility of being able to refuse unknown individuals. Figure 5 presents the results achieved processing 13 different sequences of individuals, aprox. 2500 images, not included in the training set. The success rate reflects for these sequences those faces considered correctly unknown by the verification classifier. The average success rate is 0.9715 and the average error rate is 0.0285.

Again, as the procedure is applied to a sequence, these

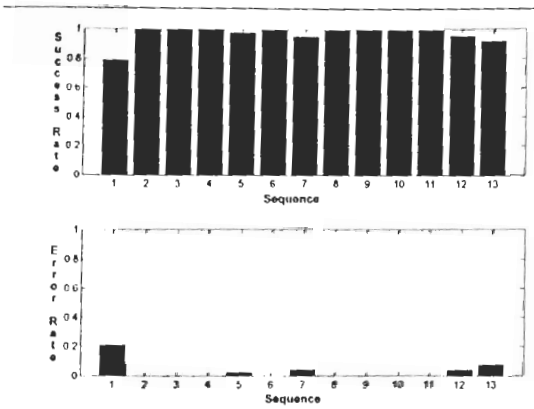

Figure 5. Success and error rate comparison for sequences that do not include any individual contained in the training set.

results can be easily analyzed in order to interpret the data reported by the system. Therefore, the system can assume some rules to make the system decide if the current subject is a known or unknown person.

Given a number m of detected faces during an interaction session:

1. If the system has considered that more than 50% of the images belongs to an individual and not more than 10% are assigned to other identities, then the system accepts the identity.
2. In any other case, the system considers that two situations are possible: 1) A new user is present, thus his appearance must be modelled, and 2) an already contained user is not well modelled in the database, so the system needs to update his model. In both cases the system should ask *who are you?*.

These simple rules work with any of the sequences used in these experiments.

4. Conclusions and Future Work

We have developed a basic ability to allow a system to autonomously suggest if an individual is not contained or not properly modelled in the database. According to the results achieved, this ability needs under some circumstances a supervised activation similar to humans, but under some circumstances allows the system to inquire the identity of a new person.

Our next objectives focus on the selection of multiple significant patterns of individuals and the addition of more individuals to the system by means of allowing it to work continuously instead of analyzing preregistered sequences.

Acknowledgments

Work partially funded by research projects Univ. of Las Palmas de Gran Canaria UNI2003/06 and Canary Islands Autonomous Government PI2003/165.

References

[1] 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.

[2] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. on PAMI*, 19(7):711–720, 1997.

[3] M. Castrillón Santana. *On Real-Time Face Detection in Video Streams. An Opportunistic Approach.* PhD thesis, Universidad de Las Palmas de Gran Canaria, March 2003.

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

[5] O. Déniz, J. Lorenzo, M. Castrillón, and M. Hernández. Estudio experimental sobre la combinación temporal de resultados en el reconocimiento de caras con secuencias de video. In *IX Conferencia de la Asociación Espaola para la Inteligencia Artificial, Gijón*, pages 59–64, November 2001.

[6] O. Déniz Suárez, M. Castrillón Santana, and H. Tejera. Face recognition using independent component analysis and support vector machines. *Pattern Recognition Letters*, 24(13):2153–2157, September 2003.

[7] 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.

[8] C. L. Lisetti and D. J. Schiano. Automatic facial expression interpretation: Where human-computer interaction, artificial intelligence and cognitive science intersect. *Pragmatics and Cognition (Special Issue on Facial Information Processing: A Multidisciplinary Perspective*, 8(1):185–235, 2000.

[9] 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.

[10] 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.

[11] A. Samal and P. A. Iyengar. Automatic recognition and analysis of human faces and facial expressions: A survey. *Pattern Recognition*, 25(1), 1992.

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

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

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