LNCS 2091

Josef Bigun
Fabrizio Smeraldi (Eds.)

# Audio- and Video-Based Biometric Person Authentication

Third International Conference, AVBPA 2001
Halmstad, Sweden, June 2001
Proceedings

AVBPA

IAPR

Springer

Josef Bigun   Fabrizio Smeraldi (Eds.)

# Audio- and Video-Based Biometric Person Authentication

Third International Conference, AVBPA 2001
Halmstad, Sweden, June 6-8, 2001
Proceedings

Springer

Series Editors

Gerhard Goos, Karlsruhe University, Germany
Juris Hartmanis, Cornell University, NY, USA
Jan van Leeuwen, Utrecht University, The Netherlands

Volume Editors

Josef Bigun
Fabrizio Smeraldi
Halmstad University
School of Information Science,
Computer and Electrical Engineering
P.O. Box 823, S-301 18 Halmstad, Sweden
E-mail: {josef.bigun/Fabrizio.smeraldi}@ide.hh.se

Cataloging-in-Publication Data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Audio- and video-based biometric person authentication : third international conference ; proceedings / AVBPA 2001, Halmstad, Sweden, June 6 - 8, 2001. Josef Bigun ; Fabrizio Smeraldi (ed.). - Berlin ; Heidelberg ; New York ; Barcelona ; Hong Kong ; London ; Milan ; Paris ; Singapore ; Tokyo : Springer, 2001
(Lecture notes in computer science ; Vol. 2091)
ISBN 3-540-42216-1

CR Subject Classification (1998): I.5, I.4, I.3, K.6.5, K.4.4, C.2.0

ISSN 0302-9743
ISBN 3-540-42216-1 Springer-Verlag Berlin Heidelberg New York



Springer-Verlag Berlin Heidelberg New York
a member of BertelsmannSpringer Science+Business Media GmbH

http://www.springer.de



Typesetting: Camera-ready by author
Printed on acid-free paper SPIN: 10839388 06/3142 5 4 3 2 1 0

# Preface

This book collects the research work presented at the Third International Conference on Audio- and Video- Based Biometric Person Authentication that took place in Halmstad, Sweden, in June 2001.

As in the preceding two cases, the conference announcement met with a consistent positive response both from industry and the research community. Since 1997, when the first conference took place, the field of Biometric Person Authentication has witnessed the development of commercial solutions that testify the practical relevance of the subject. On the other hand, the high quality of the research papers collected in this volume confirms the scientific importance and the challenging nature of the problems underlying this multi-disciplinary research field.

The volume represents a necessarily concise survey of state-of-the-art techniques in the field and addresses the topics:

- Face as biometrics
- Face image processing
- Speech as biometrics and speech processing
- Fingerprints as biometrics
- Gait as biometrics
- Hand, signature, and iris as biometrics
- Multi-modal analysis and system integration

Compared to the previous editions, fingerprints and gait have gained emphasis. The book also includes three invited contributions:

- Anil Jain (Michigan State University, USA),
- Josef Kittler (University of Surrey, UK), and
- Satoshi Nakamura (ATR, Japan).

We believe that a sizable contribution of the proceedings resides in its multi-disciplinary character. Growing demands for conjugating security with the mobility and flexibility required by emerging applications, e.g. mobile electronic commerce, can only be addressed through a close cooperation between the communication and the computer science communities. It is likely that multi-modality will play a key role in future authentication systems, which will afford a high degree of robustness to shifting usage conditions and adaptability to scalable security requirements.

We gratefully acknowledge the contributions of the Program Committee, the referees, as well as the sponsoring organizations.

April 2001

J. Bigun and F. Smeraldi

# Organization

The international conference AVBPA 2001 was organized by

- the **School of Information Science, Computer and Electrical Engineering**, of Halmstad Univeristy, Sweden,
- **TC-14** of IAPR (International Association for Pattern Recognition).

## Executive Committee

| | |
|---|---|
| Conference Chair: | Josef Bigun; Halmstad University |
| Program Chairs: | Josef Bigun and Fabrizio Smeraldi; Halmstad University |
| Local Organization: | Eva Nestius, Ulrika Hult, and Ulla Johansson; Halmstad University |

## Program Committee

| | |
|---|---|
| The *Program Chairs,* | |
| Frederic Bimbot | IRISA, France, |
| Mats Blomberg | Royal Institute of Technology, Sweden, |
| Jean-Francois Bonastre | Uni. d'Avignon e. d. Pays de Vaucluse, France, |
| Gunilla Borgefors | Swedish Univ. of Agricultural Sciences, Sweden, |
| Roberto Brunelli | ITC-irst, Italy, |
| J. M. Hans du Buf | University of Algarve, Portugal, |
| Horst Bunke | University of Bern, Switzerland, |
| Rama Chellapa | University of Maryland, USA, |
| Gerard Chollet | CNRS, France, |
| Robert Frischholz | Dialog Communication Systems AG, Germany, |
| Sadaoki Furui | Tokyo Inst. of Technology, Japan, |
| Dolores Garcia-Plaza Cuellar | Ibermatica, Spain, |
| Dominique Genoud | Nuance Communications, USA, |
| Bjorn Granstrom | Royal Institute of Technology, Sweden, |
| Kenneth Jonsson | Finger Prints AB, Sweden, |
| Jurgen Luettin | ASCOM, Switzerland, |
| John Mason | University of Swansea, UK, |
| George Matas | CVUT, Czech Republic, |
| Bruce Millar | Australian National University, Australia, |
| Jonathon Phillips | DARPA, USA, |
| Tomaso Pogio | MIT, USA, |
| Nalini Ratha | IBM, USA, |
| Gael Richard | Philips, France, |
| Massimo Tistarelli | University of Genova, Italy, |
| Harry Wechsler | George Mason University, USA, |
| Thomas Vetter | University of Freiburg, Germany. |

## Referees

The *Program Committee,*

| | |
|---|---|
| Roberto Cesar | University of Sao Paulo, Brazil, |
| Tony F. Ezzat | MIT, USA, |
| Cristina Fernandez Grande | Ibermatica, Spain, |
| Sami Romdhani | University of Freiburg, Germany, |
| Miguel Schneider-Fontan | Ibermatica, Spain. |

## Sponsoring Organizations

Halmstad University,
International Association for Pattern Recognition (IAPR),
VISIT program of the Swedish Foundation for Strategic Research,
Swedish Society for Automated Image Analysis, (SSAB).

# Table of Contents

## Face as Biometrics



## Face Image Processing

## Speech as Biometrics and Speech Processing

## Fingerprints as Biometrics

## Gait as Biometrics



## Hand, Signature and Iris as Biometrics



## Multi-modal Analysis and System Integration

# Face Identification and Verification via ECOC

J Kittler, R Ghaderi, T Windeatt and J Matas

Centre for Vision, Speech and Signal Processing,
University of Surrey, Guildford, Surrey GU2 7XH, UK
[J.Kittler, T.Windeatt]@eim.surrey.ac.uk

**Abstract.** We propose a novel approach to face identification and verification based on the Error Correcting Output Coding (ECOC) classifier design concept. In the training phase the client set is repeatedly divided into two ECOC specified sub-sets (super-classes) to train a set of binary classifiers. The output of the classifiers defines the ECOC feature space, in which it is easier to separate transformed patterns representing clients and impostors. As a matching score in this space we propose the average first order Minkowski distance between the probe and gallery images. The proposed method exhibits superior verification performance on the well known XM2VTS data set as compared with previously reported results.

## 1 Introduction

Automatic verification and authentication of personal identity based on biometric measurements has become popular in security applications. Existing commercial systems are exploiting a myriad of biometric modalities including voice characteristics, iris scan and finger print. However, as a source of biometric information, the human face plays a particularly important role as facial images (photographs) not only can easily be acquired but also they convey discriminatory features which are routinely used for recognition by humans without the need for specialist training. This opens the possibility for a close human - machine interaction and cooperation. Should the need arise, human operators may readily be called on to endorse machine decisions, as may be desirable, for instance, at border check points, or for access to high security sites. Furthermore, in comparison with other biometrics, face images can be collected in a natural way during the interaction of the subject with the verification system at the point of access. In contrast to other modalities face imaging also allows continuous verification during the client's access to services.

Unfortunately, the performance of automatic systems for face recognition or verification is often poor. Although a considerable progress has been made over recent years, face recognition and verification is still a challenging task. For this reason one of the recent paradigms has been to use multiple modalities to achieve robustness and improved performance. Typically, one would combine voice and face data [2] to achieve better verification rates (lower false rejection and false acceptance rates). However, the merits of the combination of other modalities including face profile, lip dynamics and 3D face information to name but a few have also been investigated. Although the multimodal approach has been demonstrated to achieve significant improvements, there is still the need to improve the performance of the constituent biometric subsystems to drive the error rates even lower. Some advances recently reported in this context include [9].

As another direction to gain performance improvements, attempts have been made to combine the outputs of several decision making systems. This approach draws on the results in multiple classifier fusion [10]. By combining several opinions one can reduce the error variance of the outputs of the individual experts and achieve better error rates. In [8] it was shown that by combining the scores of several diverse face verification systems the error rate of the best expert could be reduced by more than 42 %. However, such ad hoc designs of multiple expert systems may not necessarily produce the best solutions.

In this paper we propose a novel method for designing multiple expert face verification systems. It is based on the error correcting output codes (ECOC) approach developed for channel coding. The basic idea is to allocate additional bits over and above the bits required to code the source message in order to provide error correcting capability. In the context of pattern classification the idea implies that each class is represented by a more complex code than the conventional code $Z_{ij} = 0 \; \forall i \neq j$ and $Z_{ij} = 1 \;\; i = j$. The implementation of such error resilient code requires more than the usual number of classifiers.

The main difficulty in applying the ECOC classification method to the problem of face verification is that verification is a two class problem and ECOC is suited exclusively to multiclass problems. We overcome this difficulty by proposing a two stage solution to the verification problem. In the first stage we view the verification task as a recognition problem and develop an ECOC design to generate class specific discriminants. In fact we need only the discriminant for the class of the claimed identity. In the second stage we test the hypothesis that the generated discriminant is consistent with the distributions of responses for the particular client.

The proposed scheme leads to an effective design which exhibits the attractive properties of ECOC classifiers but at the same time it is applicable to the two class personal identity verification problem. The design approach has been tested on the XM2VTS face database using the Lausanne protocol. The false rejection and false acceptance rates achieved are superior to the best reported results on this database to date [14].

The paper is organised as follows. In Section 2 we describe how face images are represented. In Section 3 we outline the Error Correcting Output Code method and adapt it to the verification problem. In Section 4 we develop two hypothesis testing approaches which are the basis of the final stage of the verification process. The results of the proposed method obtained on the XM2VTS face database are reported in Section 5 which is followed by conclusions in Section 6.

## 2 Face Image Representation

Normalisation or standardisation is an important stage in face recognition or verification. Face images differ in both shape and intensity, so *shape alignment* (geometric normalisation) and *intensity correction* (photometric normalisation) can improve performance of the designed system. Our approach to geometric normalisation has been based on eye position. Four parameters are computed from the eye coordinates (rotation,scaling and translation in horizontal and vertical directions) to crop the face part from the original image and scale it to any desired resolution. Here we use "manually localised" eye coordinates to eliminate the dependency of the experiments on processes which may lack robustness. In this way, we can focus our investigation on how the performance is affected by the methodology of verification and in particular by the ECOC technique. For photometric normalisation we have used histogram equalisation as it has exhibited better performance in comparison with other existing methods[12].

Although it is possible to use gray levels directly, as demonstrated in earlier experiments[19,15], normally features are first extracted. There are many techniques in the pattern recognition literature for extracting and selecting effective features that provide maximal class separation in the feature space [3]. One popular approach is *Linear Discriminant Analysis (LDA)* which is used in our experiments. We briefly review the theory of LDA, and how it is applied to face recognition or verification. Further details may be found in [3] and [17].

Given a set of vectors $x_i, i = 1, \ldots, M$, $x_i \in R^D$, each belonging to one of c classes $\{C_1, C_2, \ldots, C_c\}$, we compute the between-class scatter matrix, $S_B$,

$$S_B = \sum_{i=1}^{c} (\mu_i - \mu)(\mu_i - \mu)^T \tag{1}$$

and within-class scatter matrix, $S_W$

$$S_W = \sum_{i=1}^{c} \sum_{x_k \in C_i} (x_k - \mu_i)(x_k - \mu_i)^T \tag{2}$$

where $\mu$ is the grand mean and $\mu_i$ is the mean of class $C_i$.

The objective of LDA is to find the transformation matrix, $W_{opt}$, that maximises the ratio of determinants $\frac{|W^T S_B W|}{|W^T S_W W|}$. $W_{opt}$ is known to be the solution of the following eigenvalue problem [3]:

$$S_B W - S_W W \Lambda = 0 \tag{3}$$

Premultiplying both sides by $S_W^{-1}$, (3) becomes:

$$(S_W^{-1} S_B) W = W \Lambda \tag{4}$$

where $\Lambda$ is a diagonal matrix whose elements are the eigenvalues of matrix $S_W^{-1} S_B$. The column vectors $\mathbf{w_i}$ ($i = 1, \ldots, c-1$) of matrix W are referred to as *fisherfaces* in [1].

In high dimensional problems (e.g. in the case where $x_i$ are images and $D$ is $\approx 10^5$) $S_W$ is almost always singular, since the number of training samples M is much smaller than D. Therefore, an initial dimensionality reduction must be applied before solving the eigenvalue problem in (3). Commonly, dimensionality reduction is achieved by Principal Component Analysis [21][1]; the first $(M - c)$ eigenprojections are used to represent vectors $x_i$. The dimensionality reduction also allows $S_W$ and $S_B$ to be efficiently calculated. The optimal linear feature extractor $W_{opt}$ is then defined as:

$$W_{opt} = W_{lda} * W_{pca} \tag{5}$$

where $W_{pca}$ is the PCA projection matrix and $W_{lda}$ is the optimal projection obtained by maximising

$$W_{lda} = arg\max_{W} \frac{|W^T W_{pca}^T S_W W_{pca} W|}{|W^T W_{pca}^T S_B W_{pca} W|} \tag{6}$$

## 3 ECOC Fundamentals

Error-Correcting Output Coding (ECOC) is an information theoretic concept which suggests that there may be advantages in employing ECOC codes to represent different signals which should be distinguished from each other after being corrupted while passing through a transmission channel. Dietterich and Bakiri [4] suggest that classification can be modelled as a transmission channel consisting of "input features", "training samples", and "learning paradigm". Classes are represented by *code words* with large Hamming distance between any pair. ECOC is believed to improve performance both by decomposing the multi-class problem as well as by correcting errors in the decision-making stage [5]. The binary values in the code word matrix are determined by the code generation procedure; it is possible to choose values that provide a meaningful decomposition [20], but usually there is no meaning attached [5,6,23,11]. There are a few methods to find a set of code words with a guaranteed minimum distance between any pair, the most popular being the BCH codes [5,18], which we use in our experiments.

To understand the ECOC algorithm, consider a $k \times b$ code word matrix $Z$ ($k$ is the number of classes) in which the $k$ rows represent code words (labels), one for each class. In the training phase, for each column, the patterns are re-labelled according to the binary values ("1s" and "0s"), thereby defining two *super classes.* A binary classifier is trained $b$ times, once for each column. Each pattern can now be transformed into ECOC feature space by the $b$ classifiers, giving a vector

$$\underline{y} = [y_1, y_2, \ldots y_b]^T \tag{7}$$

in which $y_j$ is the real-valued output of $j$th classifier. In the test phase, the distance between output vector and label for each class is determined by

$$L_i = \sum_{j=1}^{b} |Z_{i,j} - y_j| \tag{8}$$

and a pattern is assigned to the class corresponding to the code word having minimum distance to $\underline{y}$.

## 4 ECOC for Verification

In this section we discuss how the decision making strategy based on ECOC can be modified for the face verification task, which is characterised by a large number of two-class problems with a few training patterns for each client. As explained in Section 3, decision-making in the original ECOC multiple classifier is based on the distance, $L_i$ between the output of its constituent binary classifiers and the code words (compound labels), which act as representatives of the respective classes. The test pattern is then assigned to the class for which the distance $L_i$ is minimum.

In the case of verification, the task is somewhat different. We wish to ascertain whether the classifier outputs are jointly consistent with the claimed identity. This could be accomplished by setting a threshold on the distance of the outputs from the client code. However, the compound code represents an idealised target, rather than the real distribution of these outputs. Thus measuring the distance from the client code could be misleading, especially in spaces of high dimensionality.

One alternative would be to adopt the *centroid* of the joint classifier outputs to characterise each client and to measure the consistency of a new client claim from this representation. Incidently, the use of centroid in the context of ECOC classifiers is also advocated in [7]. However, as we have only a very small number of training samples, the estimated centroid would be very unreliable. We propose to represent each client $i$ by a set $Y_i$ of $N$ ECOC classifier output vectors, i.e.

$$Y_i = \{\underline{y}_i^l | l = 1, 2, ..., N\} \tag{9}$$

where $N$ is the number of $i-th$ client patterns available for training. In order to test the hypothesis that the client claim is authentic we adopt as a test statistic the average distance between vector $\underline{y}$ and the elements of set $Y_i$. The distance is measured using first order Minkowski metric, i.e.

$$d_i(\underline{y}) = \frac{1}{N} \sum_{l=1}^{N} \sum_{j=1}^{b} |y_j^l - y_j| \tag{10}$$

where $y_j$ is the $jth$ binary classifier output for the test pattern, and $y_j^l$ is the $jth$ classifier output for the $l-th$ member of class $i$. The distance is checked against a decision threshold, $t$. If the distance is below the threshold, client's claim is accepted, otherwise it is rejected, i.e.

$$d_i(\underline{y})\{ \begin{array}{ll} \leq t & accept\ claim \\ > t & reject\ claim \end{array} \tag{11}$$

It should be noted that the measure in (10) can also be used for identification by finding the argument $i$ for which the the distance $d_i(\underline{y})$ is minimum, i.e.

$$assign\ \underline{y}\ to\ class\ i\ if \qquad d_i(\underline{y}) = min_j d_j(\underline{y}) \tag{12}$$

Regardless of whether it is used in the identification or verification mode, we shall refer to the ECOC algorithm deploying measure (10) as multi-seed ECOC.

It is also interesting to note that ECOC can be interpreted as a version of *stacked generaliser* in which level zero multiple classifiers are binary and at level one we have an appropriate classifier for the ultimate task - verification or identification [22]. Although nearest neighbour classifiers advocated for level one by Skalak [22] have exhibited good performance in many applications, they do not perform well when the number of patterns is too low. Our approach is to use the decision rules in (11) and (12) that are based on average distance instead. The motivation for using first order Minkowski metric as in (8) rather than second order (Euclidean metric) is the greater robustness of the former to outliers (highly erroneous outputs of the level zero binary classifiers).

Note that instead of measuring the distance between points, we could measure a between point similarity which can be expressed by a kernel function that assumes a maximum when the distance is zero and monotonically decreases as the distance increases. The design of the decision function cannot involve any training as the number of points available is extremely small. We simply use exponential kernels with fixed width $\sigma$. The centres do not need to be explicitly determined because we use $d_i(\underline{y})$ in the exponent of the kernel to measure similarity of $\underline{y}$ to class $i$. We allocate one kernel per client and a number of kernels for each imposter. We measure the relative similarities of a test vector to the claimed identity and to the impostors as

$$k_i(\underline{y}) = \sum_{\alpha} w_\alpha \exp\{-\frac{d_\alpha(\underline{y})}{\sigma}\} \tag{13}$$

where index $\alpha$ runs over all imposter kernel placements and client $i$, the weights $w_\alpha$ are estimated and $\sigma^2$ defines the width of the kernel. The client claim test is carried out as follows:

$$k_i(\underline{y})\{ \begin{array}{ll} \geq 0.5 & accept\ \ claim \\ < 0.5 & reject\ \ claim \end{array} \tag{14}$$

## 5 Experiments on XM2VTS Data Base

The aim of the experiments reported in this section is to evaluate the proposed approach to personal identity verification and to compare the results with other verification methods. We use the XM2VTS face database for this purpose as it is known to be challenging and several results of experiments, carried out according to an internationally agreed protocol using other verification methods, are readily available in the literature.

### 5.1 Database and experimental protocol

The extended M2VTS (XM2VTS) database contains 295 subjects. The subjects were recorded in four separate sessions uniformly distributed over a period of 5 months, and within each session a number of shots were taken including both frontal-view and rotation sequences. In the frontal-view sequences the subjects read a specific text (providing synchronised image and speech data), and in the rotation sequences the head was moved vertically and horizontally (providing information useful for 3D surface modelling of the head). Further details of this database can be found in [16]. [1]

The experimental protocol (known as Lausanne evaluation protocol) provides a framework within which the performance of vision-based (and speech-based) person authentication systems running on the extended M2VTS database can be measured. The protocol assigns 200 clients and 95 impostors. Two shots of each session for each subject's frontal or near frontal images are selected to compose two configurations. We used the first configuration which is more difficult as the reported results show [14]. In this configuration, for each client there are 3 training, 3 evaluation and 2 test images. The impostor set is partitioned into 25 evaluation and 70 test impostors. Within the

[1] *http://www.ee.surrey.ac.uk/Research/VSSP/xm2fdb.html*