

Josef Bigum

Fabrizio Smeraldi

Audio- and Biometric Person Authentication

**Third International Conference
Halmstad, Sweden, June 2009
Proceedings**

ISBN 978-91-85566-29-2

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

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag. Violations are liable for prosecution under the German Copyright Law.

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

<http://www.springer.de>

© Springer-Verlag Berlin Heidelberg 2001
Printed in Germany

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

Table of Contents

Face as Biometrics

Face Identification and Verification via ECOC	1
<i>Kittler J., Ghaderi R., Windeatt T., and Matas J.</i>	
Pose-Independent Face Identificataion from Video Sequences	14
<i>Lincoln M.C. and Clark A.F.</i>	
Face Recognition Using Independent Gabor Wavelet Features.....	20
<i>Liu C. and Wechsler H.</i>	
Face Recognition from 2D and 3D Images	26
<i>Wang Y., Chua C.-S., and Ho Y.-K.</i>	
Face Recognition Using Support Vector Machines with the Feature Set Extracted by Genetic Algorithms	32
<i>Lee K., Chung Y., and Byun H.</i>	
Comparative Performance Evaluation of Gray-Scale and Color Information for Face Recognition Tasks	38
<i>Gutta S., Huang J., Liu C., and Wechsler H.</i>	
Evidence on Skill Differences of Women and Men Concerning Face Recognition	44
<i>Bigun J., Choy K.-W., and Olsson H.</i>	
Face Recognition by Auto-associative Radial Basis Function Network	52
<i>Zhang B.L. and Guo Y.</i>	

Face Recognition Using Independent Component Analysis and Support Vector Machines	59
<i>Déniz O., Castrillón M., and Hernández M.</i>	

Face Image Processing

A Comparison of Face/Non-face Classifiers	65
<i>Hjelmås E. and Farup I.</i>	
Using Mixture Covariance Matrices to Improve Face and Facial Expression Recognitions	71
<i>Thomaz C.E., Gillies D.F., and Feitosa R.Q.</i>	
Real-Time Face Detection Using Edge-Orientation Matching	78
<i>Fröba B. and Küblbeck C.</i>	

Directional Properties of Colour Co-occurrence Features for Lip Location and Segmentation	84
<i>Chindaro C. and Deravi F.</i>	
Robust Face Detection Using the Hausdorff Distance	90
<i>Jesorsky O., Kirchberg K.J., and Frischholz R.W.</i>	
Multiple Landmark Feature Point Mapping for Robust Face Recognition ..	96
<i>Rajapakse M. and Guo Y.</i>	
Face Detection on Still Images Using HIT Maps	102
<i>García Mateos G. and Vicente Chicote C.</i>	
Lip Recognition Using Morphological Pattern Spectrum	108
<i>Omata M., Hamamoto T., and Hangai S.</i>	

A Face Location Algorithm Robust to Complex Lighting Conditions.....	115
<i>Mariani R.</i>	

Automatic Facial Feature Extraction and Facial Expression Recognition ..	121
<i>Dubuisson S., Davoine F., and Cocquerez J.P.</i>	

Speech as Biometrics and Speech Processing

Fusion of Audio-Visual Information for Integrated Speech Processing	127
<i>Nakamura S.</i>	

Revisiting Carl Bild's Impostor: Would a Speaker Verification System Foil Him?	144
<i>Sullivan K.P.H. and Pelecanos J.</i>	

Speaker Discriminative Weighting Method for VQ-Based Speaker Identification	150
<i>Kinnunen T. and Fränti P.</i>	

Visual Speech: A Physiological or Behavioural Biometric?	157
<i>Brand J.D., Mason J.S.D., and Colomb S.</i>	

An HMM-Based Subband Processing Approach to Speaker Identification ..	169
<i>Higgins J.E. and Damper R.I.</i>	

Affine-Invariant Visual Features Contain Supplementary Information to Enhance Speech Recognition	175
<i>Gurbuz S., Patterson E., Tufekci Z., and Gowdy J.N.</i>	

Fingerprints as Biometrics

Recent Advances in Fingerprint Verification (<i>Invited</i>)	182
<i>Jain A.K., Pankanti S., Prabhakar S., and Ross A.</i>	

Fast and Accurate Fingerprint Verification (Extended Abstract)	192
<i>Udupa U.R., Garg G., and Sharma P.K.</i>	

An Intrinsic Coordinate System for Fingerprint Matching	198
<i>Bazen A.M. and Gerez S.H.</i>	
A Triple Based Approach for Indexing of Fingerprint Database for Identification	205
<i>Bhanu B. and Tan X.</i>	
Twin Test: On Discriminability of Fingerprints	211
<i>Jain A.K., Prabhakar S., and Pankanti S.</i>	
An Improved Image Enhancement Scheme for Fingerprint Minutiae Extraction in Biometric Identification	217
<i>Simon-Zorita D., Ortega-Garcia J., Cruz-Llanas S., Sanchez-Bote J.L., and Glez-Rodriguez J.</i>	
An Analysis of Minutiae Matching Strength	223
<i>Ratha N.K., Connell J.H., and Bolle R.M.</i>	
Curvature-Based Singular Points Detection	229
<i>Koo W.M. and Kot A.</i>	
Algorithm for Detection and Elimination of False Minutiae in Fingerprint Images	235
<i>Kim S., Lee D., and Kim J.</i>	
Fingerprint Classification by Combination of Flat and Structural Approaches	241
<i>Marcialis G.L., Roli F., and Frasconi P.</i>	
Using Linear Symmetry Features as a Pre-processing Step for Fingerprint Images	247
<i>Nilsson K. and Bigun J.</i>	
Fingerprint Classification with Combinations of Support Vector Machines	253
<i>Yao Y., Frasconi P., and Pontil M.</i>	
Performance Evaluation of an Automatic Fingerprint Classification Algorithm Adapted to a Vucetich Based Classification System	259
<i>Bartesaghi A., Gómez A., and Fernández A.</i>	
Quality Measures of Fingerprint Images	266
<i>Shen L.L., Kot A., and Koo W.M.</i>	
Gait as Biometrics	
Automatic Gait Recognition by Symmetry Analysis	272
<i>Hayfron-Acquah J.B., Nixon M.S., and Carter J.N.</i>	
Extended Model-Based Automatic Gait Recognition of Walking and Running	278
<i>Yam C.-Y., Nixon M.S., and Carter J.N.</i>	

EigenGait: Motion-Based Recognition of People Using Image Self-Similarity	284
<i>BenAbdelkader C., Cutler R., Nanda H., and Davis L.</i>	

Visual Categorization of Children and Adult Walking Styles	295
<i>Davis, J.W.</i>	

A Multi-view Method for Gait Recognition Using Static Body Parameters	301
<i>Johnsson A.Y. and Bobick A.F.</i>	

New Area Based Metrics for Gait Recognition	312
<i>Foster J.P., Nixon M.S., and Prugel-Bennett A.</i>	

Hand, Signature, and Iris as Biometrics

On-Line Signature Verifier Incorporating Pen Position, Pen Pressure, and Pen Inclination Trajectories	318
<i>Morita H., Sakamoto T., Ohishi T., Komiya Y., and Matsumoto T.</i>	

Iris Recognition with Low Template Size	324
<i>Sanchez-Reillo R. and Sanchez-Avila C.</i>	

RBF Neural Networks for Hand-Based Biometric Recognition	330
<i>Sanchez-Reillo R. and Sanchez-Avila C.</i>	

Hand Recognition Using Implicit Polynomials and Geometric Features ..	336
<i>Öden C., Erçil A., Yıldız V.T., Kırmızıtاش H., and Büke B.</i>	

Multi-modal Analysis and System Integration

Including Biometric Authentication in a Smart Card Operating System ..	342
<i>Sanchez-Reillo R.</i>	

Hybrid Biometric Person Authentication Using Face and Voice Features ..	348
<i>Poh N. and Korczak J.</i>	

Information Fusion in Biometrics	354
<i>Ross A., Jain A.K, and Qian J.-Z.</i>	

PrimeEye: A Real-Time Face Detection and Recognition System Robust to Illumination Changes	360
<i>Choi J., Lee S., Lee C., and Yi J.</i>	

A Fast Anchor Person Searching Scheme in News Sequences	366
<i>Albiol A., Torres L., and Delp E.J.</i>	

Author Index	373
---------------------------	-----

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 $Z_{ij} = 0 \quad i = j$ and $Z_{ij} = 1 \quad i \neq j$. 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 (rota-

tion, 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 $x_i \in \mathbb{R}^D$, each belonging to one of c classes $\{C_1, C_2, \dots, C_c\}$, we compute the between-class scatter matrix, S_B ,

$$S_B = \sum_{i=1}^c (\theta_i \otimes \theta_i)(\theta_i \otimes \theta_i)^T \quad (1)$$

and within-class scatter matrix, S_W

$$S_W = \sum_{i=1}^c \sum_{x_k \in C_i} (x_k \otimes \theta_i)(x_k \otimes \theta_i)^T \quad (2)$$

where θ is the grand mean and θ_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 $\frac{\det(S_B W)}{\det(S_W W)}$. W_{opt} is known to be the solution of the following eigenvalue problem [3]:

$$S_B W \otimes S_W W^{-1} = 0 \quad (3)$$

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

$$(S_W^{\otimes 1} S_B) W = W \quad (4)$$

where λ is a diagonal matrix whose elements are the eigenvalues of matrix $S_W^{\otimes 1} S_B$. The column vectors w_i ($i = 1 \dots c \otimes 1$) 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 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 $(M \otimes c)$ 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:

$$W_{opt} = W_{lda} - W_{pca} \quad (5)$$

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

$$W_{lda} = \arg \max_W \frac{\mathbf{V}^T W_{pca}^T S_W W_{pca} W \mathbf{V}}{\mathbf{V}^T W_{pca}^T S_B W_{pca} W \mathbf{V}} \quad (6)$$

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

$$\underline{y} = [y_1 \ y_2 \ \dots \ y_b]^T \quad (7)$$

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

$$L_i = \sum_{j=1}^b Z_{i,j} \otimes y_j \quad (8)$$

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. Incidentally, 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.

$$Y_i = \{\underline{y}_i^l\}_{l=1}^N \quad (9)$$

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.

$$d_i(\underline{y}) = \frac{1}{N} \sum_{j=1}^N \|\underline{y}_j - \underline{y}\|_p \quad (10)$$

where y_j is the j th binary classifier output for the test pattern, and \underline{y}_j^l is the j th 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.

$$d_i(\underline{y}) \begin{cases} < t & \text{accept claim} \\ > t & \text{reject claim} \end{cases} \quad (11)$$

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.

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

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 σ . 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

$$k_i(\underline{y}) = w \exp \left(-\frac{d_i(\underline{y})}{\sigma^2} \right) \quad (13)$$

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

$$k_i(\underline{y}) \begin{cases} > 0.5 & \text{accept claim} \\ < 0.5 & \text{reject claim} \end{cases} \quad (14)$$

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 imposter set is partitioned into 25 evaluation and 70 test impostors. Within the

¹ <http://www.ee.surrey.ac.uk/Research/VSSP/xm2fdb.html>

protocol, the verification performance is measured using the false acceptance and the false rejection rates. The operating point where these two error rates equal each other is typically referred to as the equal error rate point. Details of the this protocol can be found in [13].²

5.2 System Description

All images are projected to a lower dimensional feature space as described in section 2, so that each pattern is represented by a vector with 199 elements. There are 200 clients, so from the identification viewpoint we are facing a 200 class problem. We use a BCH equi-distance code containing 200 codewords (compound labels) 511 bit long. The Hamming distance between any pair of labels is 256 bits. The choice of code and advantages of equi-distance code are discussed in [23].

For the verification task, the level-zero classifier is a Multi-Layer Perceptron (MLP) with one hidden layer containing 199 input nodes, 35 hidden nodes and two output nodes. The Back-propagation algorithm with fixed learning rate, momentum and number of epochs is used for training. The dual output is mapped to a value between “0” and “1” to give an estimation of probability of super-class membership. For the identification task, we used an MLP with three hidden nodes.

As explained in Section 3 the outputs of the MLPs define an ECOC feature vector, and from equation (10), $d_i(\underline{y})$ for the claimed identity i is calculated by averaging over respective class images. In identification, we simply assign to the test vector \underline{y} the class with minimum average distance $d_i(\underline{y})$. For verification we use the two different combining methods described in Section 4 and in both cases we attempt to minimise the error rates on the evaluation set of clients and impostors.

5.3 Identification

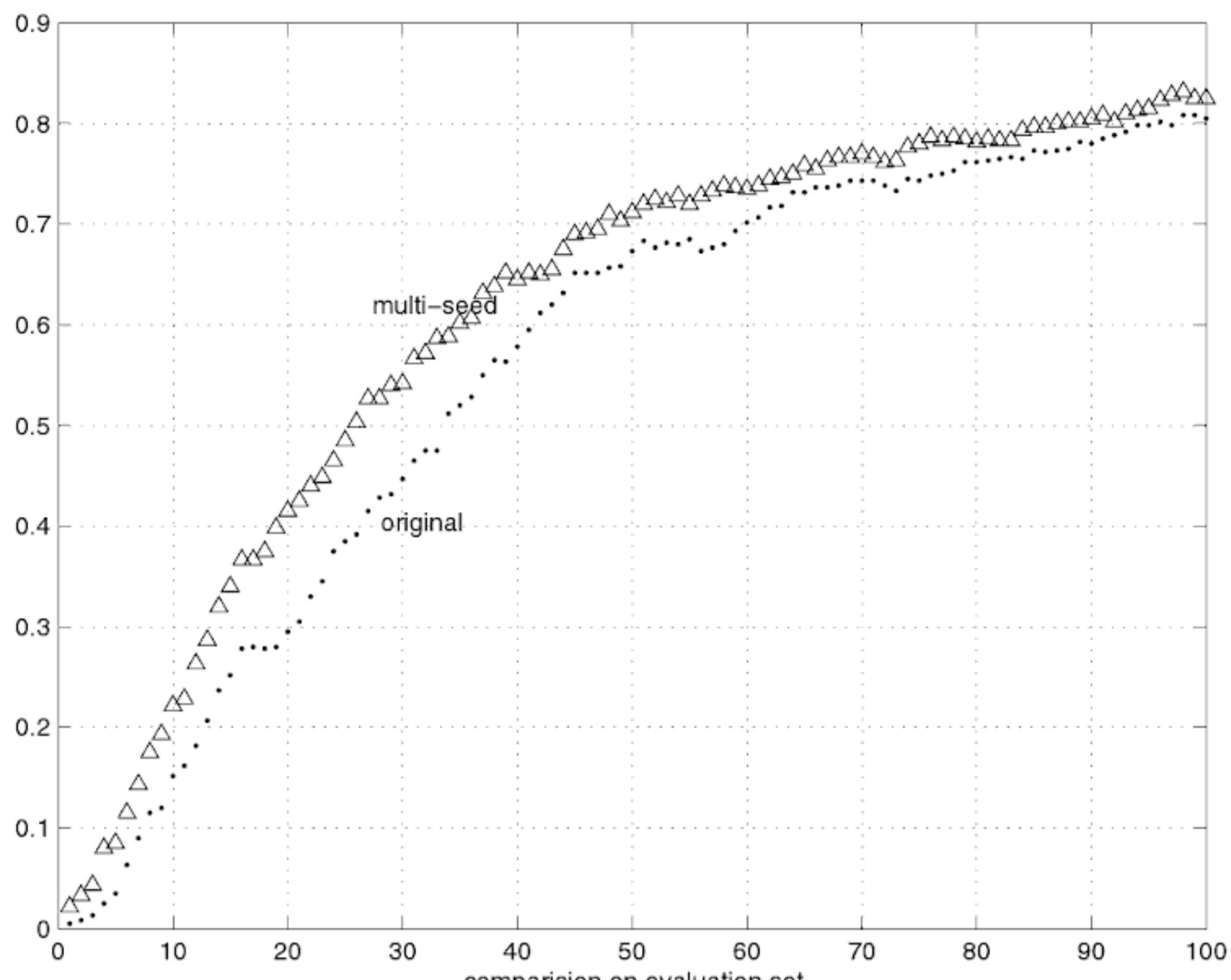
From the identification viewpoint, XM2VTS is a difficult database because the number of classes is high (200), the feature set has a high dimensionality (199 features) and the number of training samples is low (three samples for each subject).

The goal of this experiment is to show that the ECOC technique can solve a complex problem using simple learning machines (a neural network with a few hidden nodes), and to compare the original and the multi-seed ECOC. In identification the evaluation set is not used. We report the results separately for the evaluation and test sets. The rate of correct classification is presented in table II. For shorter codes (0 to 100) in particular, the performance of multi-seed ECOC is better, as clearly shown in figure II

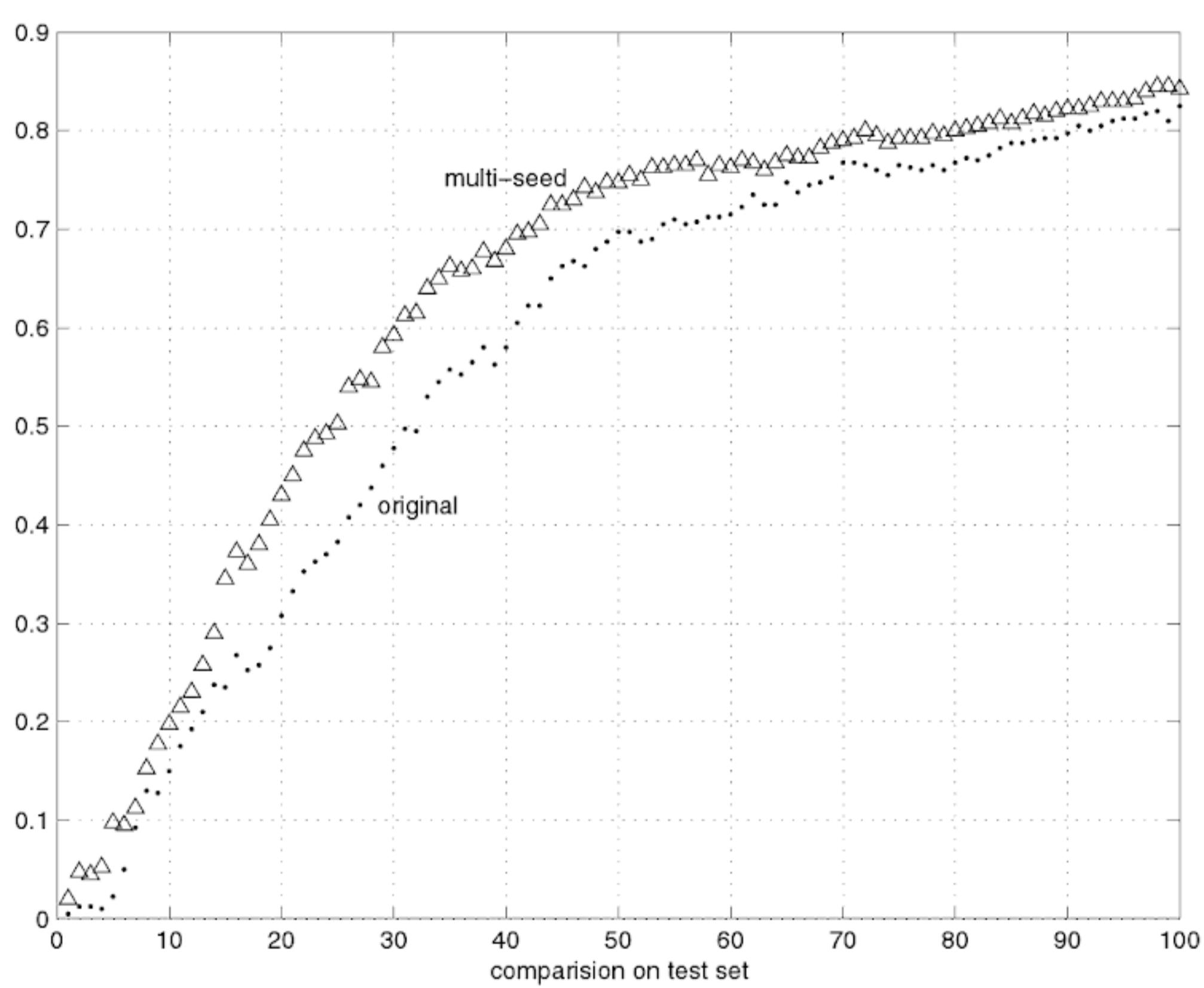
5.4 Verification

Both distance and similarity based rules for combining the outputs of the ECOC multiple classifiers have been investigated. Of the two decision functions, the distance based rule is the only one that depends on a parameter, the decision threshold, that has to be selected.

² http://www.idiap.ch/~m2vts/Experiments/xm2vtstdb_protocol_october.ps



(a)



(b)

Fig. 1. Comparing performance as a function of code length in original and multi-seed ECOC for face identification over a) Evaluation set. b) Test set.

Table 1. Comparison of original and multi-seed ECOC recognition rates (%) obtained in a subject identification experiment using an equi-distance code (200 – 511).

data set	evaluation	test
Original ECOC	91.83	92.50
Multi-seed ECOC	92.25	93.25

Distance Based Combination. Normally one would use the evaluation set data to compute the Receiver Operating Characteristics (ROC) curve which plots the relationship of false rejection rate and false acceptance rate as a function of threshold. A suitable threshold is then selected to achieve the required behaviour. For instance, one can specify the threshold that delivers equal false rejection and false acceptance rates. The threshold can be selected for each client separately, or globally by averaging the errors over all the clients.

One of the difficulties encountered with our ECOC based approach was that because the level-zero classifier was "too powerful", the FR and FA errors on the evaluation set were zero for a large range of thresholds. In such circumstances the ROC curve is not very useful in threshold setting. This problem was circumvented by the following procedure. Starting from $t = 0$ we successively increased the threshold in fixed steps to find the point where the total error (the sum of FR and FA errors) is minimum. If the total error was zero for several such increments the selected threshold would correspond to the point just before the total error would start rising.

The results obtained with the above threshold selection procedure using the evaluation set data are given in Table 2 as a function of step size. As different step sizes

Table 2. Result of verification when the clients in the evaluation set are used as seeds.

search step	FR(Ev)	FA (Ev)	FR(Ts)	FA(Ts)
.25	0	0	13.2500	0.1078
.2	0	0	10.5000	0.1422
.1	0	0	6.5000	0.2772
.05	0	0	5.2500	0.4130
.01	0	0	4.7500	0.6540
.005	0	0	4.7500	0.7111
.001	0	0	4.5000	0.7391

terminate the threshold selection procedure at different destinations from the impostors in the evaluation set the test set performance varies. In table 3 we report error rates when seeds from both the evaluation and training sets are used to set the thresholds. Even though generalisation has improved, it is not clear from the evaluation set performance how to select the best step size. One possibility is to combine the results from all step sizes, and the final row of table 3 shows the result of majority vote combination.

Table 3. Result of verification when the clients in the evaluation and training sets are used as seeds.

search step	FR(Ev)	FA(Ev)	FR(Ts)	FA(Ts)
.2	0	0.065	6.75	.1676
.1	0	0	4.50	.2174
.05	0	0	3.25	.3668
.01	0	0	1.25	.6495
.005	0	0	1.25	.7038
.001	0	0	1.25	.7482
combining	0	0	1.25	.6603

To demonstrate the effectiveness of ECOC we report in Table 4 the result of applying the exhaustive search method directly to the original 199 dimensional feature vectors. Comparing Tables 3 and 4, the benefits of mapping the input data onto the ECOC output vectors are clearly visible. Note also that in this case the evaluation set error rates are non zero, i.e. the population of clients and impostors are overlapping. In this particular case the ROC curve could have been computed but we did not pursue this particular scheme as it was clearly inferior to the ECOC based approach.

Table 4. Result of verification in the fisher face features space.

search step	FR(Ev)	FA (Ev)	FR(Ts)	FA(Ts)
.25	1.67	0.89	16.75	1.105
.2	0.83	1.07	15.25	1.144
.01	0.167	0.33	8.0	1.180
.005	0.167	0.31	8.0	1.239
.001	0.167	0.2925	8.0	1.310

Kernel Combination. Although the kernel combination method requires no thresholds, there are design parameters that can be varied to control the behaviour of the method. In particular, we can choose different ways to represent impostors. Each of the 25 evaluation impostors has 4 sets of 2 images as explained in Section 5.1. Therefore, as an alternative to 25 centres averaged over 4 sets we can choose 50 centres averaged over 2 sets or 100 centres averaged over 1 set. The error rates for 25, 50, 100 impostor centres, along with the results of combining by majority vote are shown in Table 5. In comparison with Table 3, there is a different trade-off between false acceptance and false rejection rates.

Table 5. Result of verification using the kernel score with different numbers of centres for the impostors.

impostor centres	FR(Ev)	FA(Ev)	FR(Ts)	FA(Ts)
25	0	0	0.7500	0.8833
50	0	0	0.5000	0.8786
100	0	0	0.7500	1.2455
combining	0	0	0.7500	0.8596

5.5 Comparison with Other Methods

For comparison we are including the results obtained using three other methods on the same data set and with the same protocol. The methods use the same representation of image data in terms of 199 fisher face coefficients. They employ three different scores for decision making in this feature space. In particular, we use the Euclidean metric, s_E , Normalised correlation, s_N , and Gradient metric, s_O , as detailed in [9]. The results are summarised in Table 6.

Table 6. Performance of the three baseline matching scores on manually registered images.

Score	Evaluation set			Test set		
	FR	FA	TE	FR	FA	TE
s_E	7.83	7.83	15.66	5.50	7.35	12.85
s_N	2.50	2.50	5.00	2.25	2.56	4.81
s_O	1.74	1.74	3.48	1.75	1.70	3.45

The results show a number of interesting features. First of all, by comparing the Euclidean metric performance with the proposed distance $d_i(\underline{y})$ in Table 4 it would appear that the more robust metric used in $d_i(\underline{y})$ combined with the multi-seed representation of clients may be more effective than the Euclidean distance based score. Most importantly, all the ECOC based results are decisively superior to the decision making in the original Fisher face space. Finally, the combination of ECOC multiple classifier outputs by means of the relative similarity score in (14) appears to yield slightly better results than using the distance based score $d_i(\underline{y})$. The implication of this finding and of the work reported elsewhere is that the choice of decision (score) function plays an extremely important role in the design of verification systems and should receive more attention in the future.

6 Conclusion

We described 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 the ECOC feature space a novel distance measure and a kernel based similarity measure have been developed. The distance based score computes the average first order Minkowski distance between the probe and gallery images which is more effective than the Euclidean metric. The proposed method was shown to exhibit superior verification performance on the well known XM2VTS data set as compared with previously reported results.

Acknowledgements

The support received from OmniPerception Ltd, EPSRC Grant GR/M61320 and EU Framework V Project Banca is gratefully acknowledged.

References

1. P N Belhumeur, J P Hespanha, and D J Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. In *Proc. of ECCV'96*, pages 45–58, Cambridge, United Kingdom, 1996.
2. S Ben-Yacoub, J Luettin, K Jonsson, J Matas, and J Kittler. Audio-visual person verification. In *Computer Vision and Pattern Recognition*, pages 580–585, Los Alamitos, California, June 1999. IEEE Computer Society.
3. P.A. Devijver and J. Kittler. *Pattern Recognition: A Statistical Approach*. Prentice Hall, 1982.
4. T.G Dietterich and G. Bakiri. Error-correcting output codes: A general method for improving multiclass inductive learning programs. pages 572–577. Proceedings of the Ninth National Conference on Artificial Intelligence (AAAI-91), AAAI Pres, 1991.
5. T.G. Dietterich and G Bakiri. Solving multi-class learning problems via error-correcting output codes. *Journal of Artificial Intelligence Research*, 2:263–286, 1995.
6. R. Ghaderi and T. Windeatt. Circular ecoc, a theoretical and experimental analysis. pages 203–206, Barcelona, Spain, September 2000. International Conference of Pattern Recognition(ICPR2000).
7. G. James. *Majority Vote Classifiers: Theory and Applications*. PhD thesis, Dept. of Statistics, Univ. of Stanford, May 1998. <http://www-stat.stanford.edu/~gareth/>
8. J Kittler, Y P Li, and J Matas. Face verification using client specific fisher faces. In J T Kent and R G Aykroyd, editors, *The Statistics of Directions, Shapes and Images*, pages 63–66, 2000.
9. J Kittler, Y P Li, and J Matas. On matching scores for lda-based face verification. In M Mirmehdi and B Thomas, editors, *British Machine Vision Conference*, 2000.
10. J Kittler and F Roli. *Multiple Classifier Systems*. Springer-Verlag, Berlin, 2000.
11. E.B. Kong and T.G. Dietrich. Probability estimation via error-correcting output coding. Banff, Canada, 1997. Int. Conf. of Artificial Intelligence and soft computing. <http://www.cs.orst.edu/~tqd/cv/pubs.html>

12. Y.P. Li. *Linear Discriminant Analysis and its application to face Identification*. PhD thesis, School of Electronic Engineering, Information technology and Mathematics, University of Surrey, Guildford, Surrey, U.K. GU2 7X, September 2000.
13. J Luettin and G. Maitre. *Evaluation Protocol For The Extended M2VTS Database (XM2VTS)*. Dalle Molle Institute for Perceptual Artificial Intelligence, P.O. Box 592 Martigny, Valais, Switzerland, July 1998. IDIAP-Com 98-05.
14. J. Matas, M. Hamouz, M. Jonsson, J. Kittler, Y. Li, C. Kotropoulos, A. Tefas, I. Pitas, T. Tan, H. Yan, F. Smeraldi, J. Bigun, N. Capdevielle, w. Gerstner, S. Ben-Yacoub, Y. Abdullaoued, and Y. Majoraz. Comparison of face verification results on the xm2vts database. In A. A Sanfeliu, J.J Villanueva, M. Vanrell, R. Alqueraz, J. Crowley, and Y. Shirai, editors, *Proceedings of the 15th ICPR*, volume 4, pages 858–863, Los Alamitos, USA, September 2000. IEEE Computer Soc Press.
15. J Matas, K Jonsson, and J Kittler. Fast face localisation and verification. *IVC*, 17(8):578–581, June 1999.
16. K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre. XM2VTSDB: The extended M2VTS database. In *Proc. of AVBPA'99*, pages 72–77, 1999.
17. M. Nadler and E.P Smith. *Pattern Recognition Engineering*. John Weley and Sons INC., 1993.
18. W.W. Peterson and JR. Weldon. *Error-Correcting Codes*. MIT press, Cambridge,MA, 1972.
19. F.S. Samaria and A.C. Harter. Parameterisation of a stochastic model for human face identification. In *Proceeding of the 2nd IEEE Workshop on application of computer vision*, Sarasota,Florida, 1994. <http://mambo.ucsc.edu/psl/olivetti.html>
20. T.J. Senjnowski and C.R. Rosenberg. Parallel networks that learn to pronounce english text. *Complex systems*, 1(1):145–168, 1987.
21. L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization on human face. *Journal .Opt.Soc. Am. A*, 3(4):519–524, 1987.
22. D.B. Skalak. *Prototype Selection for Composite Nearest Neighbor Classifiers*. PhD thesis, Dept. of Computer Science, Univ. of Massachusetts Amherst, May 1997.
23. T. Windeatt and R. Ghaderi. Binary codes for multi-class decision combining. volume 4051, pages 23–34, Florida,USA, April 2000. 14th Annual International Conference of Society of Photo-Optical Instrumentation Engineers (SPIE).

Independent Face Identification from Video Sequences

Michael C. Lincoln and Adrian F. Clark

VASE Laboratory, University of Essex, Colchester CO4 3SQ, UK
{mclinc,alien}@essex.ac.uk

Abstract. A scheme for pose-independent face recognition is presented. An “unwrapped” texture map is constructed from a video sequence using a texture-from-motion approach, which is shown to be quite accurate. Recognition of single frames against calculated unwrapped textures is carried out using principal component analysis. The system is typically better than 90% correct in its identifications.

Introduction

Face identification is currently a particularly active area of computer vision. Although work on face analysis was performed as long ago as the 1970s [1], current interest was arguably inspired by the “eigenfaces” technique [2]. Subsequently, researchers have applied a wide variety of approaches, including various types of neural networks [3], hidden Markov models [4] and shape analysis [5]. The vast majority of face recognition techniques, including all those listed here, concentrate on full-face imagery. This is partly because such a constraint simplifies the problem and partly because typical current applications are for situations in which the subject is cooperative. There has been work on face recognition from profile imagery [6] but the more general problem in which the head orientation is unknown remains relatively unexplored. Full 3D face recognition is touched on in [7] and considered in more detail in [8]. The area of 3D face recognition technology arguably has the most potential but at present full-face imagery is rarely available. Hence, *pose-independent* schemes have practical value; indeed, the approach outlined herein is being developed in collaboration with Essex Police.

In order to perform pose-independent face recognition, one ideally would have images of subjects captured at all possible orientations. This is not a practical solution; but it is easy to consider an “image” that is a projection of the head shape onto a notional *cylinder* rather than onto a plane. We term this a *wrapped texture map*. Our scheme involves taking each image (planar projection) in a video sequence, tracking the head from frame to frame and deter-

gh the two were developed independently. Moreover,[9] did not at-tification.

cation schemes applied to conventional, planar images can be ex-unwrapped texture maps, though care is needed. For example, e distances between interior features can be used [1], as can eigen-ooches, as used here. Most importantly however, one can compare me of a person's head with a portion of a texture map to achieve on.

remainder of this paper is organized as follows. The construction of ed texture map, the most important component of the scheme, is n Sec 2. The use of these textures in an eigenfaces-like identification discussed in Sec 3. Conclusions are drawn in Sec 4.

Construction of an Unwrapped Texture Map

minaries

ce model of the head being tracked is required in order to eval-responding texture. An accurate model of the head is not required, or models are likely to affect the accuracy and stability of tracking. employs a tapered ellipsoid as a user-independent head model; this shape to control and, as it is convex, means that hidden surface re-be accomplished with back-face culling [10].

In computer graphics, a 2D texture is normally applied to a 3D model. As-
th each vertex in each facet of a 3D model is a 2D texture coordinate. ing process then determines the appearance of each screen pixel
acet by interpolating between the texture coordinates of the vertices. his technique requires the reverse operation: values are inserted *into*
map when the image positions of the projections of vertices of the l have been determined. Our implementation of this uses OpenGL,
ws this process to be carried out in hardware, even on PC-class sys-

ained above, not every pixel in the texture map will have the same ence, each pixel in the constructed texture map has a corresponding value (forming a *confidence map*). This is modeled as the ratio be-area of a pixel in texture space and the area in screen space that gave hese confidence values are central to the way in which image data into the unwrapped texture map.

a measure of the similarity between two textures is required. The sed herein is

$$\sqrt{\sum_{uv} C_{min}^2(u, v) d^2(u, v)} \quad (1)$$

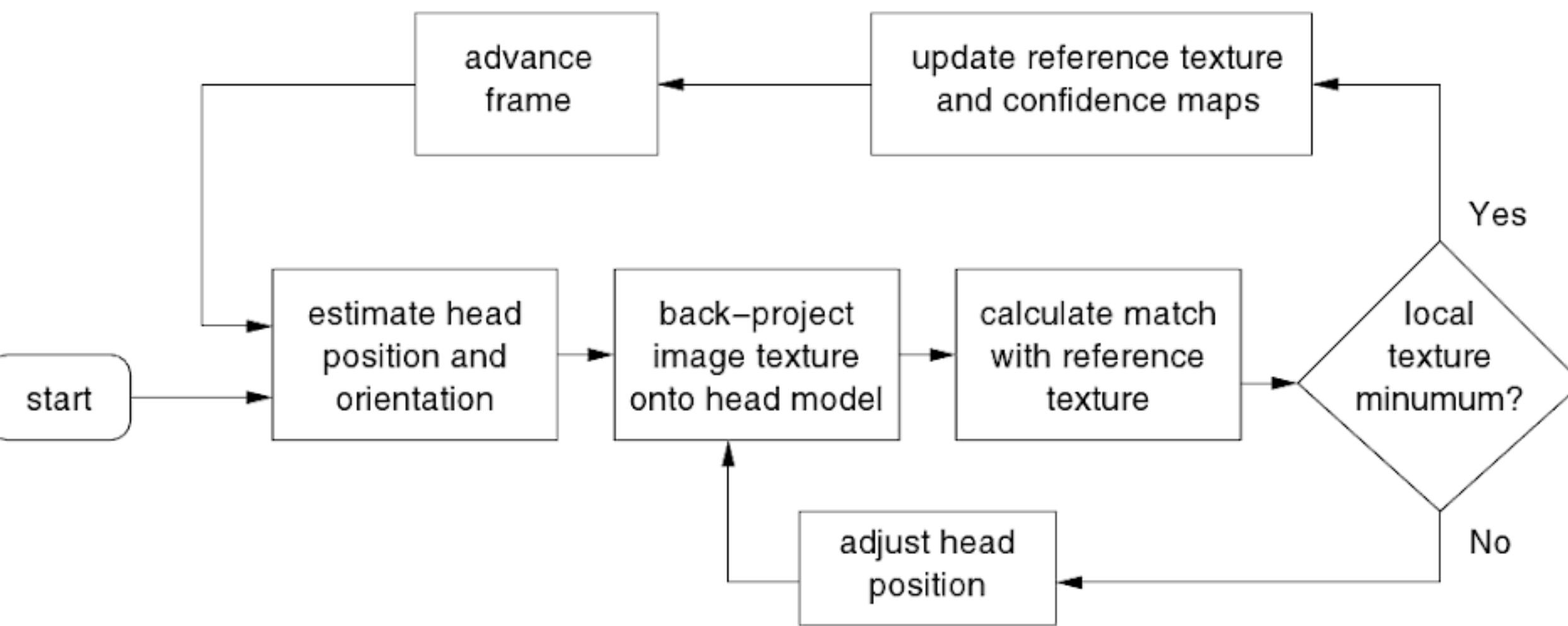


Figure 1. Procedure for constructing unwrapped texture map.

ing by Optimization

position and orientation of the head in the first frame of a sequence is currently manual, though this could be automated. As outlined above, “reference” unwrapped texture and confidence maps are initialized from the first frame. The procedure for accumulating the texture and confidence maps in subsequent frames is illustrated in Fig 1. An estimate for the head’s new position and orientation is made; this can be simply the same as in the previous frame, though some prediction scheme (e.g., Kalman filtering) is probably better. The head model is transformed to this new position and the image texture is *back-projected* onto it, facet by facet. A match with the reference head model is then performed. The six position and orientation parameters of the head model are adjusted using a simplex optimization scheme until the best match value is obtained. (Simplex seems to be as effective as the more complex scheme described in [9].

Once optimum parameters found, the back-projected texture for the current frame is merged into the reference texture map. A pixel in the texture map is updated only if it will result in a higher confidence value: if C_r is the confidence value for a pixel in the reference image and C_i the corresponding value for the current image, then

$$W_r = \frac{C_r}{C_r + C_i} \quad W_i = \frac{C_i}{C_r + C_i} \quad (2)$$

Assuming $C_i > C_r$, the texture map value V_r is updated to

$$V_r = V_r W_r + V_i W_i \quad (3)$$

the value of texture map for the current image.

This procedure is illustrated in Fig 2, which shows a single frame of a video

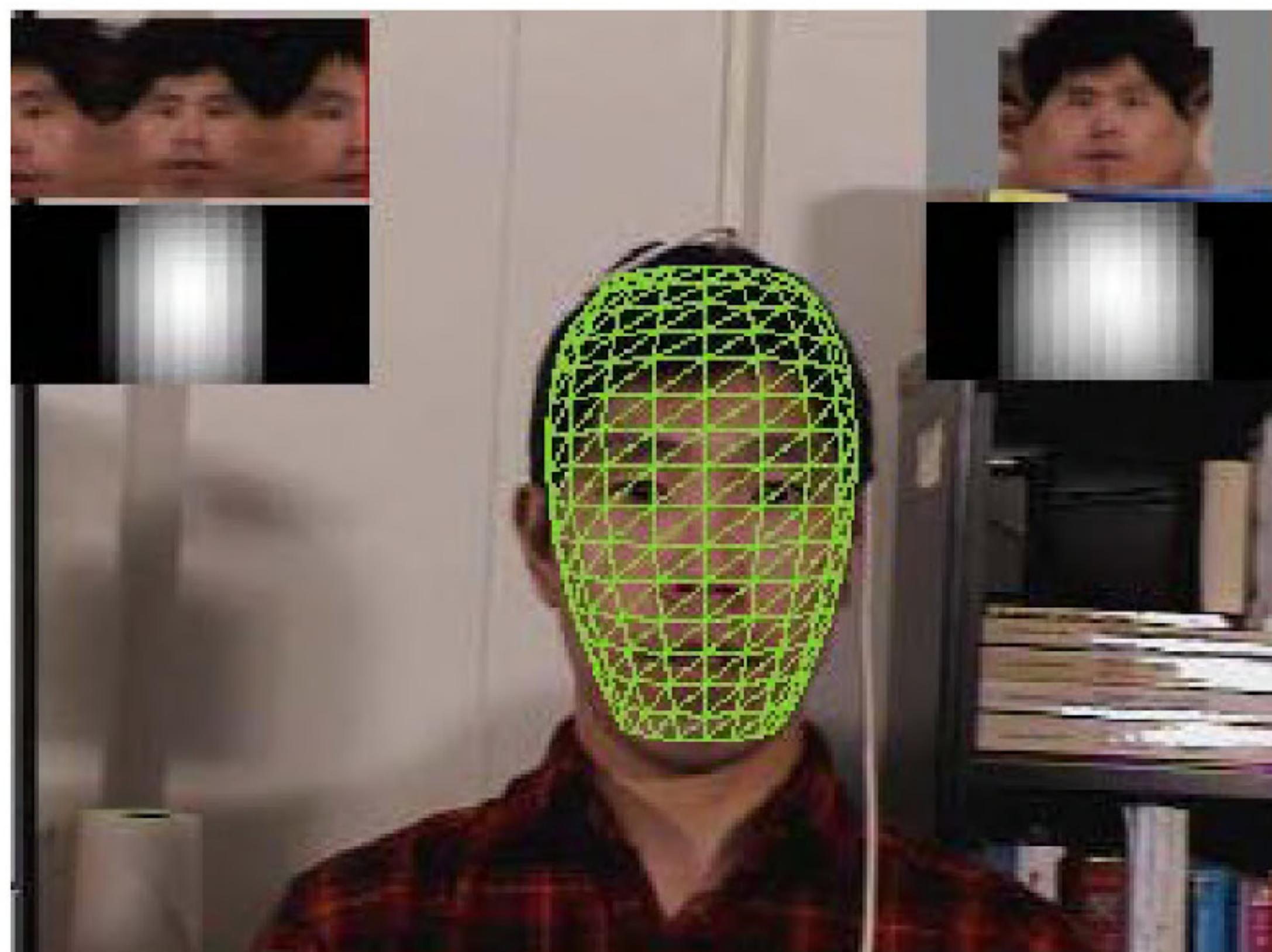


Figure 2. The construction of texture maps by tracking.

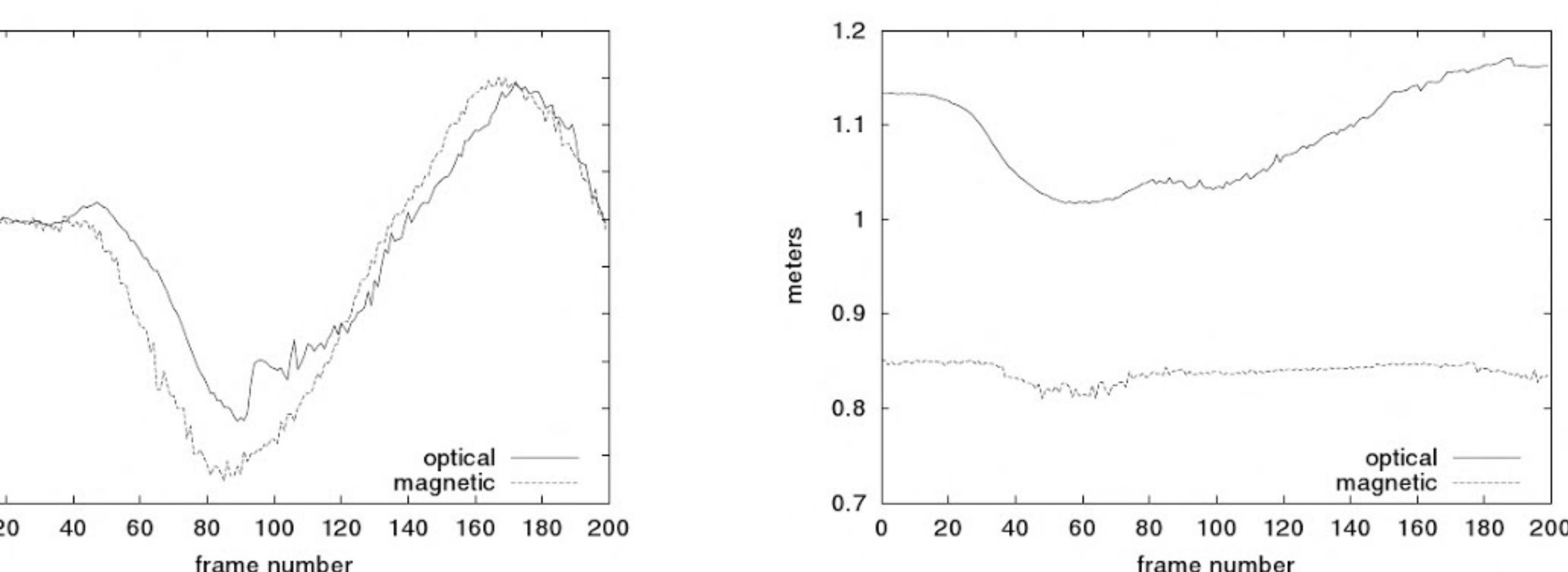
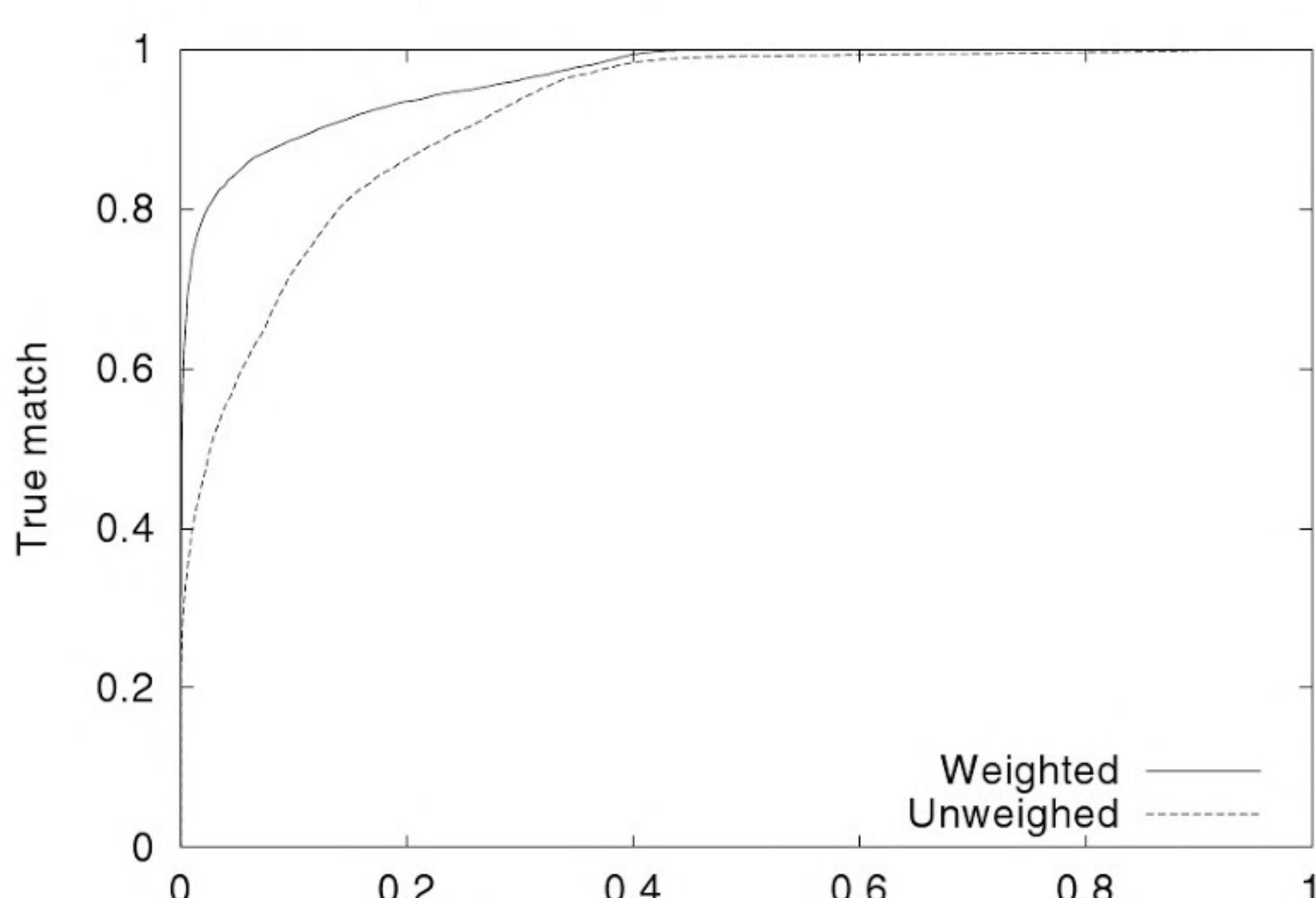


Figure 3. Head-tracking accuracy.



Face Recognition Using Independent Gabor Wavelet Features

Chengjun Liu¹ and Harry Wechsler²

¹ Dept. of Math and Computer Science, University of Missouri, St. Louis, MO 63121
cliu@cs.umsl.edu, <http://www.cs.umsl.edu/~cliu/>

² Dept. of Computer Science, George Mason University, Fairfax, VA 22030
wechsler@cs.gmu.edu, <http://cs.gmu.edu/~wechsler/personal>

Abstract. We introduce in this paper a novel Independent Gabor wavelet Features (IGF) method for face recognition. The IGF method derives first an augmented Gabor feature vector based upon the Gabor wavelet transformation of face images and using different orientation and scale local features. Independent Component Analysis (ICA) operates then on the Gabor feature vector subject to sensitivity analysis for the ICA transformation. Finally, the IGF method applies the Probabilistic Reasoning Model for classification by exploiting the independence properties between the feature components derived by the ICA. The feasibility of the new IGF method has been successfully tested on face recognition using 600 FERET frontal face images corresponding to 200 subjects whose facial expressions and lighting conditions may vary.

1 Introduction

Face recognition has wide applications in security (biometrics and forensics), human-computer intelligent interaction, digital libraries and the web, and robotics [4], [18]. It usually employs various statistical techniques, such as PCA (principal component analysis) [20], [15], FLD (Fisher linear discriminant, a.k.a. LDA, linear discriminant analysis) [19], [2], [8], ICA (independent component analysis) [1], [7], [14], and Gabor and bunch graphs [21] to derive appearance-based models for classification.

Independent Component Analysis (ICA) has emerged recently as one powerful solution to the problem of blind source separation [5] while its possible use for face recognition has been shown in [1], [4] by using a neural network approximation. ICA searches for a linear transformation to express a set of random variables as linear combinations of statistically independent source variables [5]. The search criterion involves the minimization of the mutual information expressed as a function of high order cumulants. While PCA considers the 2nd order moments only and it uncorrelates the data, ICA would further reduce statistical dependencies and produce an independent code useful for subsequent pattern discrimination and associative recall [16]. ICA thus provides a more powerful data representation than PCA.

The Gabor wavelets, which capture the properties of spatial localization, orientation selectivity, spatial frequency selectivity, and quadrature phase relationship, seem to be a good approximation to the filter response profiles encountered experimentally in cortical neurons [6], [9]. The Gabor wavelets have been found to be particularly suitable

for image decomposition and representation when the goal is the derivation of local and discriminating features. Most recently, Donato et al [7] have experimentally shown that the Gabor filter representation is optimal for classifying facial actions.

This paper introduces a novel Independent Gabor wavelet Features (IGF) method for face recognition. The Gabor transformed face images exhibit strong characteristics of spatial locality, scale and orientation selectivity, similar to those displayed by the Gabor wavelets. Such characteristics produce salient local features, such as the eyes, nose and mouth, that are most suitable for face recognition. The feasibility of the new IGF method has been successfully tested on face recognition using 600 FERET frontal face images corresponding to 200 subjects whose facial expressions and lighting conditions may vary. The effectiveness of the IGF method is shown in terms of both absolute performance indices and comparative performance against some popular face recognition schemes such as the traditional Eigenfaces method and Gabor wavelet based classification method.

2 Gabor Feature Analysis

Gabor wavelets are used for image analysis because of their biological relevance and computational properties [6, 7]. The Gabor wavelets, whose kernels are similar to the 2D receptive field profiles of the mammalian cortical simple cells, exhibit strong characteristics of spatial locality and orientation selectivity, and are optimally localized in the space and frequency domains.

The Gabor wavelets (kernels, filters) can be defined as follows [11, 7]:

$$\theta^\circ(z) = \frac{k_{\theta^\circ}}{2} e^{\otimes \frac{k_{\theta^\circ}}{2} \frac{z}{2}^2} \left(e^{ik_{\theta^\circ} z} \otimes e^{\otimes \frac{z}{2}} \right) \quad (1)$$

where θ and σ define the orientation and scale of the Gabor kernels, $z = (x^\circ y)$, $\|\cdot\|$ denotes the norm operator, and the wave vector k_{θ° is defined as follows:

$$k_{\theta^\circ} = k e^{i \theta^\circ} \quad (2)$$

where $k = k_{max} \triangleq f$ and $\theta = \theta \triangleq 8$. f is the spacing factor between kernels in the frequency domain [11]. In most cases one would use Gabor wavelets at five different scales, $\|\cdot\| \leq 4$ and eight orientations, $\theta \in \{0^\circ, 22.5^\circ, 45^\circ, 67.5^\circ, 90^\circ, 112.5^\circ, 135^\circ, 157.5^\circ\}$.

The Gabor wavelet transformation of an image is the convolution of the image with a family of Gabor kernels as defined by Eq. 1. Let $O_{\theta^\circ}^{(\cdot)}$ denote a normalized convolution output (downsampled by σ and normalized to zero mean and unit variance), then the augmented feature vector $\mathcal{X}^{(\cdot)}$ is defined as follows:

$$\mathcal{X}^{(\cdot)} = \left\{ O_{0^\circ 0}^{(\cdot)^t}, O_{0^\circ 1}^{(\cdot)^t}, \dots, O_{4^\circ 7}^{(\cdot)^t} \right\}^t \quad (3)$$

where t is the transpose operator. The augmented feature vector thus encompasses all the outputs, $O_{\theta^\circ}^{(\cdot)}(z)$ ($\theta \in \{0^\circ, 22.5^\circ, 45^\circ, 67.5^\circ, 90^\circ, 112.5^\circ, 135^\circ, 157.5^\circ\}$, $\|\cdot\| \leq 4$), as important discriminating information.

3 Independent Component Analysis of the Gabor Features for Face Recognition

We now describe our novel Independent Gabor wavelet Features (IGF) method for face recognition. The augmented Gabor feature vector introduced in Sect. 2 resides in a space of very high dimensionality, and we first apply PCA for dimensionality reduction:

$$\mathcal{Y}^{()} = P^t \mathcal{X}^{()} \quad (4)$$

where $P = [P_1 P_2 \dots P_n]$ consists of the n eigenvectors corresponding to the leading eigenvalues of the covariance matrix of $\mathcal{X}^{()}$, $n < N$ and $P \in \mathbb{R}^{N \times n}$. The lower dimensional vector $\mathcal{Y}^{()} \in \mathbb{R}^n$ captures the most expressive features of the original data $\mathcal{X}^{()}$. The PCA output is then processed by the ICA method and its reduced dimension n is determined based upon the ICA sensitivity analysis.

ICA, which expands on PCA as it considers higher (> 2) order statistics, is used here to derive independent Gabor wavelet features for face recognition. ICA of a random vector seeks a linear transformation that minimizes the statistical dependence between its components [5]. In particular, let $\mathcal{Y} \in \mathbb{R}^n$ be a n dimensional random vector corresponding to the PCA output defined by Eq. 4. The ICA of the random vector \mathcal{Y} factorizes the covariance matrix Σ into the following form:

$$\Sigma = F \Sigma F^t \quad (5)$$

where $\Sigma \in \mathbb{R}^{m \times m}$ is diagonal real positive and $F \in \mathbb{R}^{n \times m}$, whose column vectors are orthogonal, transforms the original random vector $\mathcal{Y} \in \mathbb{R}^n$ to a new one $\mathcal{Z} \in \mathbb{R}^m$, where $\mathcal{Y} = F\mathcal{Z}$, such that the m components ($m \leq n$) of the new random vector \mathcal{Z} are independent or “the most independent possible” [5].

The whitening transformation of the ICA derivation can be rearranged as follows [12]:

$$\mathcal{U} = \Sigma^{-1/2} F^t \mathcal{Y} \quad (6)$$

The above equation shows that during whitening the eigenvalues appear in the denominator. The trailing eigenvalues, which tend to capture noise as their values are fairly small, cause the whitening step to fit for misleading variations and make the overall method generalize poorly when it is presented with new data. If the whitening step, however, is preceded by a dimensionality reduction procedure (see Eq. 4 and a proper dimensionality is chosen, ICA performance would be enhanced and the computational complexity reduced [12].

3.1 The Probabilistic Reasoning Model for the Independent Gabor Features Method

The novel IGF method applies the independent component analysis on the (lower dimensional) augmented Gabor feature vector. In particular, the augmented Gabor feature vector $\mathcal{X}^{()}$ of an image is first calculated as detailed in Sect. 2. The IGF method determines, then, the dimensionality of the lower dimensional feature space n according to the sensitivity analysis of ICA (Sect. 3) and derives the lower dimensional feature, $\mathcal{Y}^{()}$

(Eq. 4). Finally, the IGF method derives the overall (the combination of the whitening, rotation, and normalization transformations) ICA transformation matrix, F , as defined by Eq. 5. The new feature vector, $\mathcal{Z}^{()}$, of the image is thus defined as follows:

$$\mathcal{Y}^{()} = F\mathcal{Z}^{()} \quad (7)$$

After the extraction of an appropriate set of features, the IGF method applies the Probabilistic Reasoning Model (PRM) [13] for classification. In particular, Let \mathcal{M}_k^0 , $k = 1, 2, \dots, L$, be the mean of the training samples for class k after the ICA transformation. The IGF method exploits, then, the following MAP classification rule of the PRM method [13]:

$$\min_j \frac{\sum_{i=1}^m (z_i \otimes m_{j,i})^2}{\sum_{i=1}^m z_i^2} \otimes \mathcal{Z}^{()} \quad k \quad (8)$$

where z_i and $m_{j,i}$, $i = 1, \dots, m$, are the components of $\mathcal{Z}^{()}$ and \mathcal{M}_k^0 , respectively, and z_i^2 is estimated by sample variance in the one dimensional ICA space:

$$z_i^2 = \frac{1}{L} \sum_{k=1}^L \frac{1}{N_k \otimes 1} \sum_{j=1}^{N_k} \left\{ y_{j,i}^{(k)} \otimes m_{j,i} \right\}^2 \quad (9)$$

where $y_{j,i}^{(k)}$ is the i-th element of the ICA feature $Y_j^{(k)}$ of the training image that belongs to class k , and N_k is the number of training images available for class k . The MAP classification rule of Eq. 8 thus classifies the image feature vector, $\mathcal{Z}^{()}$, as belonging to the class k .

4 Experiments

We assess the feasibility and performance of our novel Independent Gabor Features (IGF) method on the face recognition task, using 600 face images corresponding to 200 subjects from the FERET standard facial database [17]. The effectiveness of the IGF method is shown in terms of both absolute performance indices and comparative performance against some popular face recognition schemes such as the traditional Eigenfaces (PCA) method and Gabor wavelet based classification methods.

For comparison purpose, we use the following nearest neighbor (to the mean) classification rule:

$$(\mathcal{X} \cdot \mathcal{M}_k^0) = \min_j (\mathcal{X} \cdot \mathcal{M}_j^0) \otimes \mathcal{X} \quad k \quad (10)$$

The image feature vector, \mathcal{X} , is classified as belonging to the class of the closest mean, \mathcal{M}_k^0 , using the similarity measure . The similarity measures used in our experiments to evaluate the efficiency of different representation and recognition methods include L_1 distance measure, L_1 , L_2 distance measure, L_2 , Mahalanobis distance measure, M_d , and cosine similarity measure, \cos .

We first implemented the Eigenfaces method [20] on the original images, using the four different similarity measures: L_1 distance measure, L_1 , L_2 distance measure, L_2 ,

Table 1. Face recognition performance on the Gabor convolution outputs, using the three different similarity measures.

measure \ representation	$O_{\mu,\nu}^{(16)}$	$\mathcal{X}^{(4)}$	$\mathcal{X}^{(16)}$	$\mathcal{X}^{(64)}$
L_1	76%	76.5%	76.5%	76.5%
L_2	73.5%	72%	72%	72%
cosine	72%	70.5%	70.5%	70%

Mahalanobis distance measure, M_d , and cosine similarity measure, cos . The Mahalanobis distance measure performs better than the L_1 distance measure, followed in order by the L_2 distance measure and the cosine similarity measure. In particular, when 180 features (the specific number of features chosen here facilitates later comparisons with other methods) are used, the recognition rates are 76%, 70.5%, 42.5%, and 38%, accordingly. The reason that the Mahalanobis distance measure performs better than the other similarity measures is that the Mahalanobis distance measure counteracts the fact that L_1 and L_2 distance measures in the PCA space weight preferentially for low frequencies. As the L_2 measure weights more the low frequencies than L_1 does, the L_1 distance measure should perform better than the L_2 distance measure, a conjecture validated by our experiments.

The next series of experiments used the Gabor convolution outputs, $O_{\theta^*}(z)$, derived in Sect. 2 with the L_1 , L_2 and cosine similarity measures, respectively. For the first set of experiments, we downsampled the Gabor convolution outputs by a factor 16 to reduce the dimensionality and normalized them to unit length, as suggested by Donato et al. [7]. The classification performance using such Gabor outputs is shown in Table II. The best performance is achieved using the L_1 similarity measure. We have also experimented on the augmented Gabor feature vector $\mathcal{X}^{()}$ as defined by Eq. 3 with three different downsampling factors: = 4, 16, and 64, respectively. From the classification performance shown in Table II we found that (i) the augmented Gabor feature vector $\mathcal{X}^{()}$ carries quite similar discriminant information to the one used by Donato et al. [7]; and (ii) the performance differences using the three different downsampling factors are not significant. As a result, we choose the downsampling factor 64 for our novel IGF method, since it reduces to a larger extent the dimensionality of the vector space than the other two factors do. (We experimented with other downsampling factors as well. When the downsampling factors are 256 and 1024, the performance is marginally less effective; when the factor is 4096, however, the recognition rate drops drastically.)

The last experiment, performed using the novel Independent Gabor Features (IGF) method described in this paper, shows that the IGF derives independent Gabor features with low dimensionality and enhanced discrimination power. In particular, when 180 features are used by the IGF method, the correct recognition rate is 98.5%.

*image
not
available*

*image
not
available*

*image
not
available*

- Mariani R. 115
Mason J.S.D. 157
Matas J. [1](#)
Matsumoto T. 318
Morita [H.](#) 318
- Nakamura S. 127
Nanda [H.](#) 284
Nilsson K. 247
Nixon M.S. 272, 278, 312
- Öden C. 336
Ohishi T. 318
Olsson [H.](#) 44
Omata M. 108
Ortega-Garcia J. 217
- Pankanti S. 182, 211
Patterson E. 175
Pelecanos J. 144
Poh N. 348
Pontil M. 253
Prabhakar S. 182, 211
Prugel-Bennett A. 312
- Qian J.-Z. 354
Rajapakse M. 96
Ratha N.K. 223
Roli F. 241
- Ross A. 182, 354
Sakamoto T. 318
Sanchez-Avila C. 324, 330
Sanchez-Bote J.L. 217
Sanchez-Reillo R. 324, 330, 342
Sharma P. K. 192
Shen L.L. 266
Simon-Zorita D. 217
Sullivan [K.P.H.](#) 144
- Tan X. 205
Thomaz C.E. 71
Torres L. 366
Tufekci Z. 175
- Udupa U.R. 192
Vicente Chicote C. 102
- Wang Y. [26](#)
Wechsler [H.](#) [20](#), 38
Windeatt T. [1](#)
- Yam C.-Y. 278
Yao Y. 253
Yi J. 360
Yıldız V.T. 336
- Zhang B.L. 52