Tutorial

LNCS 3161

Massimo Tistarelli
Josef Bigun
Enrico Grosso

Advanced Studies in Biometrics

Summer School on Biometrics
Alghero, Italy, June 2003
Revised Selected Lectures and Papers

Springer

Massimo Tistarelli Josef Bigun
Enrico Grosso

Advanced Studies in Biometrics

Summer School on Biometrics
Alghero, Italy, June 2-6, 2003
Revised Selected Lectures and Papers

Springer

Authors

Massimo Tistarelli
Enrico Grosso
University of Sassari, Computer Vision Laboratory
Palazzo del Pou Salit - Piazza Duomo 6, 07041 Alghero (SS), Italy
E-mail: {tista,grosso}@uniss.it

Josef Bigun
University of Halmstad, IDE
Box 823, 301 18 Halmstad, Sweden
E-mail: Josef.Bigun@ide.hh.se

Library of Congress Control Number: 2005926703

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

ISSN 0302-9743
ISBN-10 3-540-26204-0 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-26204-6 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springeronline.com

Typesetting: Camera-ready by author, data conversion by Boller Mediendesign
Printed on acid-free paper SPIN: 11493648 06/3142 5 4 3 2 1 0

Lecture Notes in Computer Science 3161

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Lecture Notes in Computer Science

For information about Vols. 1–3424

please contact your bookseller or Springer

Vol. 3535: M. Steffen, G. Zavattaro (Eds.), Formal Methods for Open Object-Based Distributed Systems. X, 323 pages. 2005.

Vol. 3532: A. Gómez-Pérez, J. Euzenat (Eds.), The Semantic Web: Research and Applications. XV, 728 pages. 2005.

Vol. 3526: S.B. Cooper, B. Löwe, L. Torenvliet (Eds.), New Computational Paradigms. XVII, 574 pages. 2005.

Vol. 3525: A.E. Abdallah, C.B. Jones, J.W. Sanders (Eds.), Communicating Sequential Processes. XIV, 321 pages. 2005.

Vol. 3524: R. Barták, M. Milano (Eds.), Integration of AI and OR Techniques in Constraint Programming for Combinatorial Problems. XI, 320 pages. 2005.

Vol. 3523: J.S. Marques, N.P. de la Blanca, P. Pina (Eds.), Pattern Recognition and Image Analysis, Part II. XXVI, 733 pages. 2005.

Vol. 3522: J.S. Marques, N.P. de la Blanca, P. Pina (Eds.), Pattern Recognition and Image Analysis, Part I. XXVI, 703 pages. 2005.

Vol. 3520: O. Pastor, J. Falcão e Cunha (Eds.), Advanced Information Systems Engineering. XVI, 584 pages. 2005.

Vol. 3518: T.B. Ho, D. Cheung, H. Li (Eds.), Advances in Knowledge Discovery and Data Mining. XXI, 864 pages. 2005. (Subseries LNAI).

Vol. 3517: H.S. Baird, D.P. Lopresti (Eds.), Human Interactive Proofs. IX, 143 pages. 2005.

Vol. 3516: V.S. Sunderam, G.D. van Albada, P.M.A. Sloot, J.J. Dongarra (Eds.), Computational Science – ICCS 2005, Part III. LXIII, 1143 pages. 2005.

Vol. 3515: V.S. Sunderam, G.D. van Albada, P.M.A. Sloot, J.J. Dongarra (Eds.), Computational Science – ICCS 2005, Part II. LXIII, 1101 pages. 2005.

Vol. 3514: V.S. Sunderam, G.D. van Albada, P.M.A. Sloot, J.J. Dongarra (Eds.), Computational Science – ICCS 2005, Part I. LXIII, 1089 pages. 2005.

Vol. 3513: A. Montoyo, R. Mu\~noz, E. Métais (Eds.), Natural Language Processing and Information Systems. XII, 408 pages. 2005.

Vol. 3510: T. Braun, G. Carle, Y. Koucheryavy, V. Tsaoussidis (Eds.), Wired/Wireless Internet Communications. XIV, 366 pages. 2005.

Vol. 3509: M. Jünger, V. Kaibel (Eds.), Integer Programming and Combinatorial Optimization. XI, 484 pages. 2005.

Vol. 3508: P. Bresciani, P. Giorgini, B. Henderson-Sellers, G. Low, M. Winikoff (Eds.), Agent-Oriented Information Systems II. X, 227 pages. 2005. (Subseries LNAI).

Vol. 3507: F. Crestani, I. Ruthven (Eds.), Information Context: Nature, Impact, and Role. XIII, 253 pages. 2005.

Vol. 3505: V. Gorodetsky, J. Liu, V. A. Skormin (Eds.), Autonomous Intelligent Systems: Agents and Data Mining. XIII, 303 pages. 2005. (Subseries LNAI).

Vol. 3503: S.E. Nikoletseas (Ed.), Experimental and Efficient Algorithms. XV, 624 pages. 2005.

Vol. 3502: F. Khendek, R. Dssouli (Eds.), Testing of Communicating Systems. X, 381 pages. 2005.

Vol. 3501: B. Kégl, G. Lapalme (Eds.), Advances in Artificial Intelligence. XV, 458 pages. 2005. (Subseries LNAI).

Vol. 3500: S. Miyano, J. Mesirov, S. Kasif, S. Istrail, P. Pevzner, M. Waterman (Eds.), Research in Computational Molecular Biology. XVII, 632 pages. 2005. (Subseries LNBI).

Vol. 3499: A. Pelc, M. Raynal (Eds.), Structural Information and Communication Complexity. X, 323 pages. 2005.

Vol. 3498: J. Wang, X. Liao, Z. Yi (Eds.), Advances in Neural Networks – ISNN 2005, Part III. L, 1077 pages. 2005.

Vol. 3497: J. Wang, X. Liao, Z. Yi (Eds.), Advances in Neural Networks – ISNN 2005, Part II. L, 947 pages. 2005.

Vol. 3496: J. Wang, X. Liao, Z. Yi (Eds.), Advances in Neural Networks – ISNN 2005, Part II. L, 1055 pages. 2005.

Vol. 3495: P. Kantor, G. Muresan, F. Roberts, D.D. Zeng, F.-Y. Wang, H. Chen, R.C. Merkle (Eds.), Intelligence and Security Informatics. XVIII, 674 pages. 2005.

Vol. 3494: R. Cramer (Ed.), Advances in Cryptology – EUROCRYPT 2005. XIV, 576 pages. 2005.

Vol. 3493: N. Fuhr, M. Lalmas, S. Malik, Z. Szlávik (Eds.), Advances in XML Information Retrieval. XI, 438 pages. 2005.

Vol. 3492: P. Blache, E. Stabler, J. Busquets, R. Moot (Eds.), Logical Aspects of Computational Linguistics. X, 363 pages. 2005. (Subseries LNAI).

Vol. 3489: G.T. Heineman, I. Crnkovic, H.W. Schmidt, J.A. Stafford, C. Szyperski, K. Wallnau (Eds.), Component-Based Software Engineering. XI, 358 pages. 2005.

Vol. 3488: M.-S. Hacid, N.V. Murray, Z.W. Raś, S. Tsumoto (Eds.), Foundations of Intelligent Systems. XIII, 700 pages. 2005. (Subseries LNAI).

Vol. 3486: T. Helleseth, D. Sarwate, H.-Y. Song, K. Yang (Eds.), Sequences and Their Applications - SETA 2004. XII, 451 pages. 2005.

Vol. 3483: O. Gervasi, M.L. Gavrilova, V. Kumar, A. Laganà, H.P. Lee, Y. Mun, D. Taniar, C.J.K. Tan (Eds.), Computational Science and Its Applications – ICCSA 2005, Part IV. XXVII, 1362 pages. 2005.

Vol. 3482: O. Gervasi, M.L. Gavrilova, V. Kumar, A. Laganà, H.P. Lee, Y. Mun, D. Taniar, C.J.K. Tan (Eds.), Computational Science and Its Applications – ICCSA 2005, Part III. LXVI, 1340 pages. 2005.

Vol. 3481: O. Gervasi, M.L. Gavrilova, V. Kumar, A. Laganà, H.P. Lee, Y. Mun, D. Taniar, C.J.K. Tan (Eds.), Computational Science and Its Applications – ICCSA 2005, Part II. LXIV, 1316 pages. 2005.

Vol. 3480: O. Gervasi, M.L. Gavrilova, V. Kumar, A. Laganà, H.P. Lee, Y. Mun, D. Taniar, C.J.K. Tan (Eds.), Computational Science and Its Applications – ICCSA 2005, Part I. LXV, 1234 pages. 2005.

Vol. 3479: T. Strang, C. Linnhoff-Popien (Eds.), Location- and Context-Awareness. XII, 378 pages. 2005.

Vol. 3478: C. Jermann, A. Neumaier, D. Sam (Eds.), Global Optimization and Constraint Satisfaction. XIII, 193 pages. 2005.

Vol. 3477: P. Herrmann, V. Issarny, S. Shiu (Eds.), Trust Management. XII, 426 pages. 2005.

Vol. 3475: N. Guelfi (Ed.), Rapid Integration of Software Engineering Techniques. X, 145 pages. 2005.

Vol. 3468: H.W. Gellersen, R. Want, A. Schmidt (Eds.), Pervasive Computing. XIII, 347 pages. 2005.

Vol. 3467: J. Giesl (Ed.), Term Rewriting and Applications. XIII, 517 pages. 2005.

Vol. 3465: M. Bernardo, A. Bogliolo (Eds.), Formal Methods for Mobile Computing. VII, 271 pages. 2005.

Vol. 3464: S.A. Brueckner, G.D.M. Serugendo, A. Karageorgos, R. Nagpal (Eds.), Engineering Self-Organising Systems. XIII, 299 pages. 2005. (Subseries LNAI).

Vol. 3463: M. Dal Cin, M. Kaâniche, A. Pataricza (Eds.), Dependable Computing - EDCC 2005. XVI, 472 pages. 2005.

Vol. 3462: R. Boutaba, K.C. Almeroth, R. Puigjaner, S. Shen, J.P. Black (Eds.), NETWORKING 2005. XXX, 1483 pages. 2005.

Vol. 3461: P. Urzyczyn (Ed.), Typed Lambda Calculi and Applications. XI, 433 pages. 2005.

Vol. 3460: Ö. Babaoglu, M. Jelasity, A. Montresor, C. Fetzer, S. Leonardi, A. van Moorsel, M. van Steen (Eds.), Self-star Properties in Complex Information Systems. IX, 447 pages. 2005.

Vol. 3459: R. Kimmel, N.A. Sochen, J. Weickert (Eds.), Scale Space and PDE Methods in Computer Vision. XI, 634 pages. 2005.

Vol. 3458: P. Herrero, M.S. Pérez, V. Robles (Eds.), Scientific Applications of Grid Computing. X, 208 pages. 2005.

Vol. 3456: H. Rust, Operational Semantics for Timed Systems. XII, 223 pages. 2005.

Vol. 3455: H. Treharne, S. King, M. Henson, S. Schneider (Eds.), ZB 2005: Formal Specification and Development in Z and B. XV, 493 pages. 2005.

Vol. 3454: J.-M. Jacquet, G.P. Picco (Eds.), Coordination Models and Languages. X, 299 pages. 2005.

Vol. 3453: L. Zhou, B.C. Ooi, X. Meng (Eds.), Database Systems for Advanced Applications. XXVII, 929 pages. 2005.

Vol. 3452: F. Baader, A. Voronkov (Eds.), Logic for Programming, Artificial Intelligence, and Reasoning. XI, 562 pages. 2005. (Subseries LNAI).

Vol. 3450: D. Hutter, M. Ullmann (Eds.), Security in Pervasive Computing. XI, 239 pages. 2005.

Vol. 3449: F. Rothlauf, J. Branke, S. Cagnoni, D.W. Corne, R. Drechsler, Y. Jin, P. Machado, E. Marchiori, J. Romero, G.D. Smith, G. Squillero (Eds.), Applications of Evolutionary Computing. XX, 631 pages. 2005.

Vol. 3448: G.R. Raidl, J. Gottlieb (Eds.), Evolutionary Computation in Combinatorial Optimization. XI, 271 pages. 2005.

Vol. 3447: M. Keijzer, A. Tettamanzi, P. Collet, J.v. Hemert, M. Tomassini (Eds.), Genetic Programming. XIII, 382 pages. 2005.

Vol. 3444: M. Sagiv (Ed.), Programming Languages and Systems. XIII, 439 pages. 2005.

Vol. 3443: R. Bodik (Ed.), Compiler Construction. XI, 305 pages. 2005.

Vol. 3442: M. Cerioli (Ed.), Fundamental Approaches to Software Engineering. XIII, 373 pages. 2005.

Vol. 3441: V. Sassone (Ed.), Foundations of Software Science and Computational Structures. XVIII, 521 pages. 2005.

Vol. 3440: N. Halbwachs, L.D. Zuck (Eds.), Tools and Algorithms for the Construction and Analysis of Systems. XVII, 588 pages. 2005.

Vol. 3439: R.H. Deng, F. Bao, H. Pang, J. Zhou (Eds.), Information Security Practice and Experience. XII, 424 pages. 2005.

Vol. 3438: H. Christiansen, P.R. Skadhauge, J. Villadsen (Eds.), Constraint Solving and Language Processing. VIII, 205 pages. 2005. (Subseries LNAI).

Vol. 3437: T. Gschwind, C. Mascolo (Eds.), Software Engineering and Middleware. X, 245 pages. 2005.

Vol. 3436: B. Bouyssounouse, J. Sifakis (Eds.), Embedded Systems Design. XV, 492 pages. 2005.

Vol. 3434: L. Brun, M. Vento (Eds.), Graph-Based Representations in Pattern Recognition. XII, 384 pages. 2005.

Vol. 3433: S. Bhalla (Ed.), Databases in Networked Information Systems. VII, 319 pages. 2005.

Vol. 3432: M. Beigl, P. Lukowicz (Eds.), Systems Aspects in Organic and Pervasive Computing - ARCS 2005. X, 265 pages. 2005.

Vol. 3431: C. Dovrolis (Ed.), Passive and Active Network Measurement. XII, 374 pages. 2005.

Vol. 3430: S. Tsumoto, T. Yamaguchi, M. Numao, H. Motoda (Eds.), Active Mining. XII, 349 pages. 2005. (Subseries LNAI).

Vol. 3429: E. Andres, G. Damiand, P. Lienhardt (Eds.), Discrete Geometry for Computer Imagery. X, 428 pages. 2005.

Vol. 3428: Y.-J. Kwon, A. Bouju, C. Claramunt (Eds.), Web and Wireless Geographical Information Systems. XII, 255 pages. 2005.

Vol. 3427: G. Kotsis, O. Spaniol (Eds.), Wireless Systems and Mobility in Next Generation Internet. VIII, 249 pages. 2005.

Preface to the Lectures Book of the 1st Summer School for Advanced Studies on Biometrics 2003

Springer LNCS 3161

The ability to automatically recognize an individual has increasingly been acknowledged as a significant step in many application domains. In the last decade, several recognition and identification systems based on biometric measurements have been proposed. Many different biological signals have been utilized: fingerprints, face and facial features, retinal scans, iris patterns, hand geometry, DNA traces, and gait, and others. Not only have research tools been developed, but a notable number of new applications have been observed, making studies on biometrics a very stimulating but also a challenging arena.

All these issues pushed us to organize the 1st Summer School on Biometrics, which addressed the two facets of personal identity authentication: verification and identification. The school not only stressed the different techniques involved in the two processes, but also provided an in-depth roadmap on the algorithmic and technological issues involved in the development and integration of biometric systems.

This special LNCS volume offers the efforts and major achievements of both the school lecturers and some of the most outstanding students in the classes. The papers present different biometric authentication techniques in an attempt to provide a comprehensive selection of state-of-the-art methods used to address applications demanding robust solutions.

The volume is divided into two parts. The first part, composed of seven papers, covers a selection of the lectures given at the school classes, while the second part contains the four best contributions of the students.

In Part I, the first paper, by Bigun et al., covers a topic expected to alleviate concerns on performance and convenience, a combination of several sensing modalities or multimodal biometrics. The lecture discusses major issues involved in multi-biometrics to improve machine recognition performance while it exposes some recent findings on the human ability in person recognition. The second and third papers, by Boyd and Little, and Maltoni, respectively, address two specific biometric modalities: gait and fingerprint recognition. These papers describe two classical examples of behavioral (gait recognition) and physiological (fingerprint recognition) biometric modalities. The paper by Boyd and Little presents the psychophysics of gait recognition and different computational models to process image sequences to extract dynamic information for recognition. The paper by Maltoni is a comprehensive tutorial on fingerprint recognition, describing in detail all relevant issues in data acquisition and processing, including the latest advances in the state of the art. The fourth paper, by Tistarelli et al., analyzes the biological motivations for face-based authentication. The lecture, while exploring the psychophysics of human vision relevant to person authentication, highlights several biologically inspired processes to improve automatic face-based recognition. The application of statistical classifiers and the learning theory for robust biometric authentication are discussed in the fifth paper, by Verri et al. The application of support vector machines, firstly proposed by V. Vapnik, to biometric authentication and recognition is fully described. The sixth paper, by Yeshurun and Dganit, describes an exciting methodology and practice when using hand recognition. This contribution is well coupled with the last paper in this part, by Cipolla et al., which describes an interesting methodology to detect and track human gestures. A remarkable difference from other approaches is the use of 3D rather than 2D information for hand tracking and gesture recognition.

The presentations from the students, which we found to deserve further attention from the scientific community, were chosen to be included In Part II. The first student paper, by Castellani et al., introduces an interesting technique to exploit 3D stereoscopic data for face recognition. On a similar topic is the last paper in this section, by Conde et al.; in this case the influence of feature localization accuracy for classification is addressed. The second paper, by Gokberk et al., applies genetic algorithms to drive the feature extraction process. The proposed model is applied to a set of facial features extracted from Gabor filtered images. The paper by Campadelli and Lanzarotti, the third in this part, describes a novel method for face recognition based on elastic bunch graph matching. Differently from other approaches the set of features (jets vector) is extracted automatically from gray level and color images.

Last but not least, we wish to thank all lecturers and students and others who actively cooperated to make this event. We hope that the school contributed to the dissemination of state of the art in biometrics, as well as to advanced studies of it.

Massimo Tistarelli
Josef Bigun
Enrico Grosso

Table of Contents

Combining Biometric Evidence for Person Authentication

J. Bigun[1], J. Fierrez-Aguilar[2*], J. Ortega-Garcia[2], and J. Gonzalez-Rodriguez[2]

[1] Halmstad University, Sweden
josef.bigun@ide.hh.se
[2] Universidad Politecnica de Madrid, Spain
{jfierrez,jortega,jgonzalez}@diac.upm.es

Abstract. Humans are excellent experts in person recognition and yet they do not perform excessively well in recognizing others only based on one modality such as single facial image. Experimental evidence of this fact is reported concluding that even human authentication relies on multimodal signal analysis. The elements of automatic multimodal authentication along with system models are then presented. These include the machine experts as well as machine supervisors. In particular, fingerprint and speech based systems will serve as illustration. A signal adaptive supervisor based on the input biometric signal quality is evaluated. Experimental results on data collected from a mobile telephone prototype application are reported demonstrating the benefits of the reported scheme.

1 Introduction

Face recognition is an important element of person authentication in humans. Human face analysis engages special signal processing in visual cortex different than processing of other objects [2, 3]. It is reliably observed in a number of studies that negative bias in ability to recognize faces of another racial group versus own racial group exists [4, 5, 6]. It has been confirmed that the hair style and facial expressions are significant distraction factors for humans. It has recently been revealed [7] that the lack of caricature type information hampers the recognition more than the lack of silhouette and shading information and that there is a gender bias in women's and and men's abilities to recognize faces. In [7] it is shown that, depending on the gender to be recognized, humans were able to recognize unfamiliar faces from photographs at the success rate of 55-75%. This suggests that multimodal biometric information processing e.g. using signals from body motion including the head motion, speech, and lip movements, plays a significant role in human's efforts of authenticating other humans.

Automatic access of persons to services is becoming increasingly important in the information era. Although person authentication by machine has been a

* This study has been carried out while J. F.-A. and J. O.-G. were guest scientists at Halmstad University [1].

M. Tistarelli, J. Bigun, and E. Grosso (Eds.): Biometrics School 2003, LNCS 3161, pp. 1–18, 2005.

subject of study for more than thirty years [8, 9], it has not been until recently that the matter of combining a number of different traits for person verification has been considered [10, 11]. There are a number of benefits of doing so, just to name a few: false acceptance and false rejection error rates decrease, the authentication system becomes more robust against individual sensor or subsystem failures and the number of cases where the system is not able to give an answer (e.g. bad quality fingerprints due to manual work or larynx disorders) vanishes. The technological environment is also appropriate because of the widespread deployment of multimedia-enabled mobile devices (PDAs, 3G mobile phones, tablet PCs, laptops on wireless LANs, etc.). As a result, much research work is currently being done in order to fulfill the requirements of applications for masses.

Two early sound theoretical frameworks for combining different machine experts in a multimodal biometric system are described in [11] and [12], the former from a risk analysis perspective [13] and the later from a statistical pattern recognition point of view [14]. Both of them concluded (under some mild conditions which normally hold in practice) that the weighted average is a good way of conciliating the different experts. Soon after, multimodal fusion was studied as a two-class classification problem by using a number of machine learning paradigms [15, 16, 17], for example: neural networks, decision trees and support vector machines. They too confirmed the benefits of performance gains with trained classifiers, and favored support vector machines over neural networks and decision trees. The architecture of the system, ease of training, ease of implementation and generalization to mass use were however not considered in these studies. As happens in every pattern recognition problem which is application-oriented, these are important issues that influence the choice of a supervisor.

Interestingly enough, some recent works have nevertheless reported comparable performance between fixed and trained combining strategies [18, 19] and a debate has come out investigating the benefits of both approaches [20, 21]. As an example, and within this debate, some researches have shown how to learn user-specific parameters in a trained fusion scheme [22, 23]. As a result, they have showed that the overall verification performance can be improved significantly by considering user-dependent fusion schemes.

In this work we focus on some other benefits of a trained fusion strategy. In particular, an adaptive trained fusion scheme is introduced here. With adaptive fusion scheme, we mean that the supervisor readapts to each identity claim as a function of the quality of the input biometric signal, usually depending on external conditions such as light and background noise. Furthermore, experiments on real data from a prototype mobile authentication application combining fingerprint and speech data are reported.

This paper is structured as follows. In Section 2, we summarize the findings on mono-modal human recognition performance suggesting that individual modalities do not have to score high to yield robust multimodal systems [7]. Beginning in section 3 with some definitions, we discuss machine supervisors for multimodal authentication [1, 11] in the sequel. The elements of multimodal authentication along with major notations are introduced in section 4. In section 5, the statistical framework for conciliating the different expert opinions together with simplified and full supervisor algorithms are described. The components of our prototype mobile authentication application, namely fingerprint and speaker verification subsystems, are briefly described in section 6. Some experiments are reported in section 8 using the above mentioned multimodal authentication prototype and the performance evaluation methodology described in section 7. Conclusions will be finally given in section 9.

2 Human Face Recognition Performance

There is a general agreement on that, approximately at the age of 12 the performance of children in face recognition reaches adult levels, that there is already an impressive face recognition ability by the age of 5 and that measurable preferences for face stimuli exist in babies even younger than 10 minutes [24].

Our study [7], that aimed at quantifying the skills of humans in face recognition of unfamiliar faces, has been supported by more than 4000 volunteers[3]. We found that the lack of high spatial frequencies in visual stimuli, which result in blurred images as if face information were coming from an unfocused camera, hamper the recognition significantly more than the lack of low spatial frequencies, which result in stimuli similar to artist drawn faces, see Figure 1.

The face recognition questions. In all 8 questions (Q1-Q8) the task was to identify the picture of a stimulus person among a query set consisting of 10 pictures. The subjects were informed, before the start of the test, that the stimulus image and the image to be found in the query set were taken at two different occasions and that these two images could differ significantly in hair style, glasses, expression of the face, facial hair, clothing, etc. due to the natural changes in appearance that occur upon passage of time (a few months). In Q1-Q4 and Q8 the stimulus and the query set were shown simultaneously, in the same screen. Questions Q5-Q7 were similar to the other questions except that they included a memorization task in that the stimulus was shown in its own page without the query set. When the subject wished to continue, the stimulus was replaced by the query set, forcing the subjects to answer the question without a possibility to see the stimulus.

The available results [7] reveal that in questions in which the face image to be recognized was not manipulated (e.g. the high frequencies were not depreciated), the recognition rate varied between 55-75 % in the average. A surprising result was that the females had in the average better success in all tasks than the males. A typical question in the test is illustrated by Figure 1.

The fact that the success rates are in the best cases (female subjects) around 80% suggests that not only mono-modal information but also multimodal biometric information processing e.g. using the signals from body motion including head motion, speech, and lip movements, plays a significant role when humans authenticate other humans.

[3] As of November 2003. The test is available at http://www.hh.se/facetest

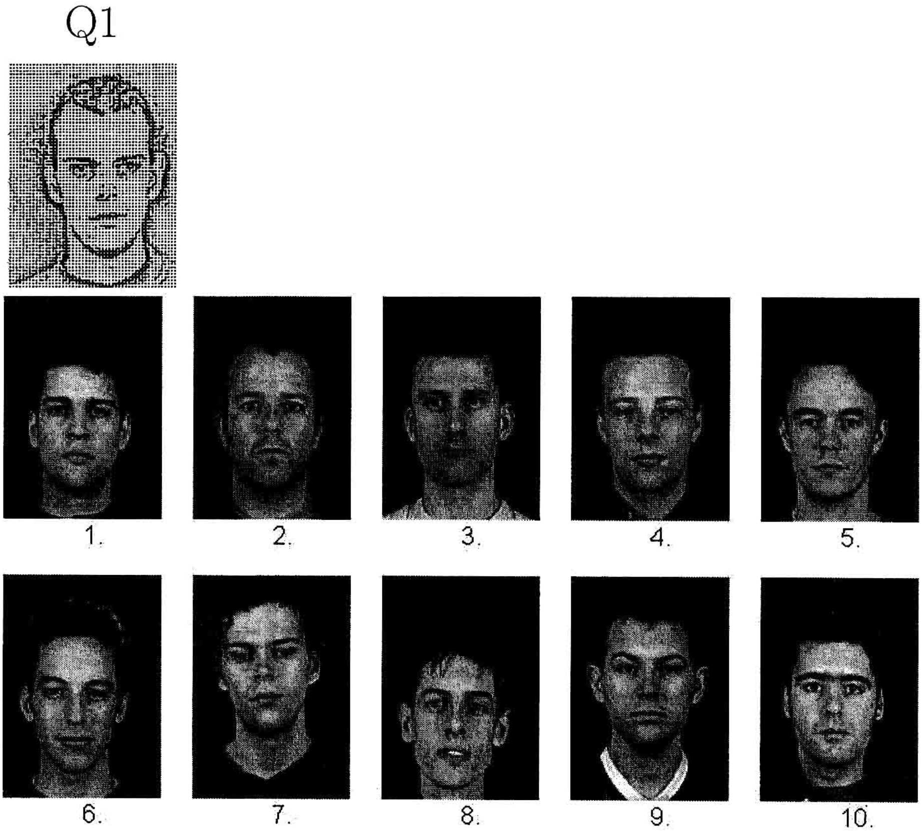

Fig. 1. A question illustrating the test. On the top the stimulus is given. The subject matched the stimulus with one of the 10 images below the stimulus. The low spatial frequencies of the stimulus were removed by signal processing.

3 Definitions

In *authentication* (also known as *verification*) applications, the users or *clients* are known to the system whereas the *impostors* can potentially be the world population. In such applications the users provide their claimed identities (either true or false) and a one-to-one matching is performed. If the result of the comparison (also *score* or *opinion*) is higher than a *verification threshold*, then the claim is accepted, otherwise the claim is rejected.

In *identification* applications, there is no identity claim and the candidate is compared to a database of client models, therefore a one-to-many matching is performed in this case. In the simplest form of identification, also known as *closed-set identification*, the best client model is selected. In *open-set identification*, the highest score is further compared to a verification threshold so as

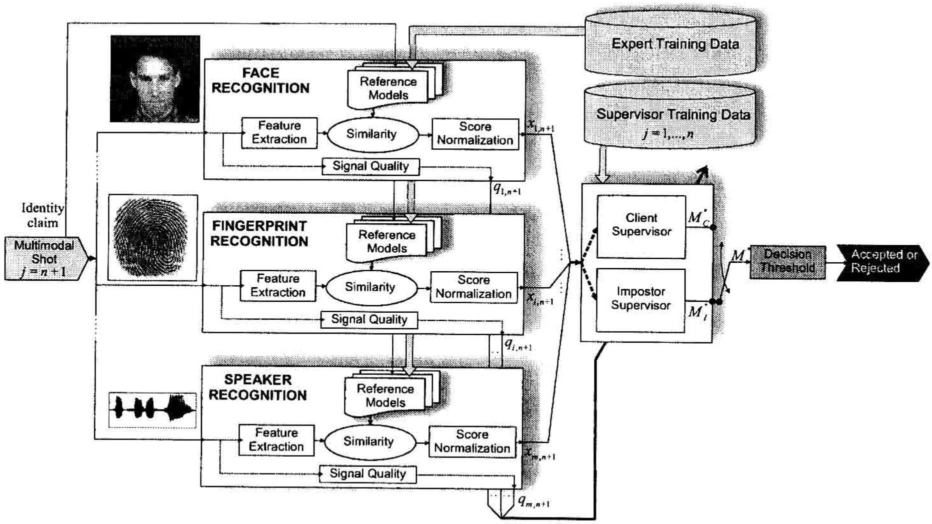

Fig. 2. The proposed system model of multi-modal person authentication.

to accept/reject this candidate as belonging or not to the database (an implicit authentication step).

In a multimodal authentication framework, various subsystems (also denoted as *experts*) are present, each one of them specialized on a different trait. Each expert delivers its opinion on a "package" of data containing an identity claim (e.g. face images, fingerprint images, speech data, etc.) that will be referred to as a *shot*. This paper is focused on combining the experts opinions (also known as *soft decisions*). It will be shown that a careful design of the *supervisor* (also known as *fusion* strategy) yields a combined opinion which is more reliable than the best expert opinion.

4 System Model

Below is a list of the major notations we use throughout the paper, see also Figure 2.

i Index of the experts, $i \in 1 \ldots m$
j Index of the shots, $j \in 1 \ldots n, n+1$
x_{ij} Authenticity score delivered by expert i on shot j
s_{ij} Variance of x_{ij} as estimated by expert i
y_j The true authenticity score of shot j
z_{ij} The error score of an expert $z_{ij} = y_j - x_{ij}$

Note that the experts are allowed to provide a quality of the score which is modelled to be inversely proportional to s_{ij}. This strategy is novel with respect to the implemented supervisors reported so far in that it is the expert who is providing a variance on every authenticity score it delivers, not the supervisor. It is also worth pointing out that y_j can take only two numerical values corresponding to "False" and "True". If x_{ij} is between 0 and 1 then these values are chosen to be 0 and 1 respectively. We assume that the experts have been trained on other shots apart from $j \in 1 \ldots n, n+1$. The supervisor is trained on shots $j \in 1 \ldots n$ (i.e. x_{ij} and y_j are known for $j \in 1 \ldots n$) and we consider shot $n+1$ as a test shot on the working multimodal system (i.e. $x_{i,n+1}$ is known, but y_{n+1} is not known and the supervisor task is to estimate it).

5 Statistical Model

The model for combining the different experts is based on Bayesian statistics and the assumption of normal distributed expert errors, i.e. z_{ij} is considered to be a sample of the random variable $Z_{ij} \sim N(b_i, \sigma_{ij}^2)$. It has been shown experimentally [11] that this assumption does not strictly hold for common audio- and video-based biometric machine experts, but it is shown that it holds reasonably well when client and impostor distributions are considered separately. Taking this result into account, two different supervisors are constructed, one of them based on expert opinions where $y_j = 1$

$$\mathcal{C} = \{x_{ij}, s_{ij} | y_j = 1 \text{ and } 1 \leq j \leq n\} \tag{1}$$

while the other is based on expert opinions where $y_j = 0$

$$\mathcal{I} = \{x_{ij}, s_{ij} | y_j = 0 \text{ and } 1 \leq j \leq n\} \tag{2}$$

The two supervisors will be referred to as *client supervisor* and *impostor supervisor*, respectively (see Figure 2).

The client supervisor estimates the expected true authenticity score of an input claim based on its expertise on recognizing client data. More formally, it computes $M_{\mathcal{C}}'' = E[Y_{n+1}|\mathcal{C}, x_{i,n+1}]$ (the prime notation will become apparent later on). In case of impostor supervisor, $M_{\mathcal{I}}'' = E[Y_{n+1}|\mathcal{I}, x_{i,n+1}]$ is computed. The conciliated overall score M'' takes into account the different expertise of the two supervisors and chooses the one which came closest to its goal, i.e. 0 for the impostor supervisor and 1 for the client supervisor:

$$M'' = \begin{cases} M_{\mathcal{C}}'' \text{ if } |1 - M_{\mathcal{C}}''| - |0 - M_{\mathcal{I}}''| < 0 \\ M_{\mathcal{I}}'' \text{ otherwise} \end{cases} \tag{3}$$

Based on the normality assumption of the errors, the supervisor algorithm described in [11] is obtained (see [13] for further background and details). In the following, we summarize this algorithm in the two cases where it can be applied.

5.1 Simplified Supervisor Algorithm

When no quality information is available, the following simplified supervisor algorithm is obtained by using $s_{ij} = 1$:

1. (Supervisor Training) Estimate the bias parameters of each expert. In case of the client supervisor the bias parameters are

$$M_{\mathcal{C}i} = \frac{1}{n_{\mathcal{C}}} \sum_j z_{ij} \quad \text{and} \quad V_{\mathcal{C}i} = \frac{\alpha_{\mathcal{C}i}}{n_{\mathcal{C}}} \tag{4}$$

where j indexes the training set $\mathcal{C}$, $n_{\mathcal{C}}$ is the number of shots in $\mathcal{C}$ and

$$\alpha_{\mathcal{C}i} = \frac{1}{n_{\mathcal{C}} - 3} \left(\sum_j z_{ij}^2 - \frac{1}{n_{\mathcal{C}}} \left(\sum_j z_{ij} \right)^2 \right) \tag{5}$$

Similarly $M_{\mathcal{I}i}$ and $V_{\mathcal{I}i}$ are obtained for the impostor supervisor.
2. (Authentication Phase) At this step, both supervisors are operational, so that the time instant is always $n+1$ and the supervisors have access to expert opinions $x_{i,n+1}$ but not access to the true authenticity score y_{n+1}. Client and impostor supervisors calibrate the experts according to their past performance, yielding (for the client supervisor)

$$M'_{\mathcal{C}i} = x_{i,n+1} + M_{\mathcal{C}i} \quad \text{and} \quad V'_{\mathcal{C}i} = (n_{\mathcal{C}} + 1) V_{\mathcal{C}i} \tag{6}$$

and then the different calibrated experts are combined according to

$$M''_{\mathcal{C}} = \frac{\sum_{i=1}^{m} \frac{M'_{\mathcal{C}i}}{V'_{\mathcal{C}i}}}{\sum_{i=1}^{m} \frac{1}{V'_{\mathcal{C}i}}} \tag{7}$$

Similarly, $M'_{\mathcal{I}i}$, $V'_{\mathcal{I}i}$ and $M''_{\mathcal{I}}$ are obtained. The final supervisor opinion is obtained according to (3).

The algorithm described above has been successfully applied in [25] in a multimodal authentication system combining face and speech data. Verification performance improvements of almost an order magnitude were reported as compared to the best modality.

5.2 Full Supervisor Algorithm

When not only the experts scores but also the quality of the scores are available, the following algorithm is obtained:

1. (Supervisor Training) Estimate the bias parameters. For the client supervisor

$$M_{\mathcal{C}i} = \frac{\sum_j \frac{z_{ij}}{\sigma_{ij}^2}}{\sum_j \frac{1}{\sigma_{ij}^2}} \quad \text{and} \quad V_{\mathcal{C}i} = \frac{1}{\sum_j \frac{1}{\sigma_{ij}^2}} \tag{8}$$

where the training set $\mathcal{C}$ is used. The variances σ_{ij}^2 are estimated through $\bar{\sigma}_{ij}^2 = s_{ij} \cdot \alpha_{\mathcal{C}i}$, where

$$\alpha_{\mathcal{C}i} = \frac{1}{n_{\mathcal{C}} - 3} \left(\sum_j \frac{z_{ij}^2}{s_{ij}} - \left(\sum_j \frac{z_{ij}}{s_{ij}} \right)^2 \left(\sum_j \frac{1}{s_{ij}} \right)^{-1} \right) \tag{9}$$

Similarly $M_{\mathcal{I}i}$ and $V_{\mathcal{I}i}$ are obtained for the impostor supervisor.

2. (Authentication Phase) First the supervisors calibrate the experts according to their past performance, for the client supervisor

$$M'_{\mathcal{C}i} = x_{i,n+1} + M_{\mathcal{C}i} \text{ and } V'_{\mathcal{C}i} = s_{i,n+1}\alpha_{\mathcal{C}i} + V_{\mathcal{C}i} \tag{10}$$

and then the different calibrated experts are combined according to (7). Similarly, $M'_{\mathcal{I}i}$, $V'_{\mathcal{I}i}$ and $M''_{\mathcal{I}}$ are obtained. The final supervisor opinion is obtained according to (3).

The algorithm described above has been successfully applied in [13] combining human expert opinions but not in a multimodal authentication application.

5.3 Adaptive Strategy

The variance s_{ij} of the score x_{ij} is provided by the expert and concerns a particular authentication assessment. It is not a general reliability measure for the expert itself, but a certainty measure based on qualitative knowledge of the expert and the data the expert assesses. Typically the variance of the score is chosen as the width of the range in which one can place the score. Because such intervals can be conveniently provided by a human expert, the algorithm in section 5.2 constitutes a systematic way of combining human and machine expertise in an authentication application. An example of such an application is forensics, where machine expert approaches have been proposed [26] and human opinions must be taken into consideration.

In this work, we propose to calculate s_{ij} for a machine expert by using a quality measure of the input biometric signal (see Figure 2). This implies taking into account equation (10) right, that the trained supervisor adapts the weights of the experts using the input signal quality. First we define the quality q_{ij} of the score x_{ij} according to

$$q_{ij} = \sqrt{Q_{ij} \cdot Q_{i,claim}} \tag{11}$$

where Q_{ij} and $Q_{i,claim}$ are respectively the quality label of the biometric sample used by expert i in shot j and the average quality of the biometric samples used by expert i for modelling the claimed identity. The two quality labels Q_{ij} and $Q_{i,claim}$ are supposed to be in the range $[0, q_{max}]$ with $q_{max} > 1$ where 0 corresponds to the poorest quality, 1 corresponds to normal quality and q_{max} corresponds to the highest quality. Finally, the variance parameter is calculated according to

$$s_{ij} = \frac{1}{q_{ij}^2} \tag{12}$$