RECENT ADVANCES IN BIOMETRIC PERSON AUTHENTICATION

J.-L. Dugelay[1] *J.-C. Junqua*[2] *C. Kotropoulos*[3*] *R. Kuhn*[2] *F. Perronnin*[1] *I. Pitas*[3]

[1] Institut EURECOM, Multimedia Communications Dept., 2229, Route des Cretes-B.P. 193, F-06904 Sophia Antipolis Cedex, France
{dugelay, perronni}@eurecom.fr

[2] Panasonic Speech Technology Lab, Suite #202, 3888 State Street, Santa Barbara, CA 93105, U.S.A.
{jcj, kuhn}@research.panasonic.com

[3] Dept. of Informatics, Aristotle Univ. of Thessaloniki, Box 451, GR-540 06 Thessaloniki, Greece
{costas, pitas}@zeus.csd.auth.gr

ABSTRACT

Biometrics is an emerging topic in the field of signal processing. While technologies (e.g. audio, video) for biometrics have mostly been studied separately, ultimately, biometric technologies could find their strongest role as interwined and complementary pieces of a multi-modal authentication system. In this paper, a short overview of voice, fingerprint, and face authentication algorithms is provided.

1. INTRODUCTION

Biometric person authentication deals with the following problem: given some physiological or behavioural characteristics of a subject, the so-called *biometrics*, and those of a reference person, whose identity is claimed by the subject, confirm or deny the claimed identity. Any human physiological or behavioural characteristic that is universal, unique, permanent, and collectable could be used as a biometric [1]. From a practical point of view, a biometric access control system should perform accurately, be acceptable by the society, be robust and should not be tampered.

Authentication (or verification) is closely related to recognition (or identification). However, the evaluation criteria for identity recognition are different from those used in authentication systems. The performance of identity recognition systems is quantified in terms of the *cumulative match score*, i.e., the percentage of correctly identified subjects within the N best matches versus N [2]. Recall-precision curves could also be used to evaluate identification algorithms. The performance of identity authentication systems is measured in terms of the *false rejection rate* (FRR) achieved at a fixed *false acceptance rate* (FAR) or vice versa. By varying FAR, the *receiver operating characteristic* (ROC) curve is obtained. A scalar figure of merit used to judge the performance of an authentication algorithm, is the so-called *equal error rate* (EER), corresponding to the ROC operating point having *FAR=FRR*.

A number of biometrics has been evaluated for identification and authentication applications. For example, voice, fingerprints, face, iris, infrared facial and hand vein thermograms, ear, retinal scans, hand and finger geometry are based on a physical characteristic, whereas signature and acoustic emissions emitted during a signature scribble, gait, keystroke dynamics are related to a behavioural characteristic [1]. In this paper we shall confine ourselves to the first three biometrics that appear to be the most popular ones in the scientific literature. The major strength of voice and face biometrics is their high acceptance by the society. Fingerprints are not as much socially acceptable as voice or face are, because they are related to forensic applications. However they offer a high performance level. The previously mentioned technologies are complementary in nature, a fact that has been partially exploited in multi-modal identification and authentication systems.

2. VOICE AUTHENTICATION

Voice authentication makes use of the unique characteristics of a user's speech to perform authentication. In the following subsections, we briefly summarize: 1) the phases of voice authentication, 2) the different degrees of text-dependence, which partition the technologies into different classes, 3) the desirable features of front-end parameters for voice authentication, 4) typical structures of voice authentication systems, 5) impostor models and score normalization methods, and 6) adaptation techniques which are essential for deploying the technology in the real world.

2.1. Phases of Voice Authentication

The various phases of voice authentication are as follows:

1. *Enrolment*: **P** (the rightful system user), speaks to it to train a voice model.
2. *Test*: **C** (the claimant) speaks to the system. The system accepts that **C** is **P**, or rejects the claim.

* Contact author: C. Kotropoulos (costas@zeus.csd.auth.gr). C. Kotropoulos and I. Pitas have been supported by the European Union funded Research Training Network "Multi-modal Human-Computer Interaction" (HPRN-CT-2000-00111).

3. *Adaptation* (optional): when the system decides that **P** has spoken to it, it updates the model of **P**.

The performance of voice authentication systems is strongly influenced by the amount of data used during phase 1. It is worth noticing that the first two phases are met in any biometric authentication system.

2.2. Degrees of Text-Dependence

Depending on the generality of the text used during the test phase 2, we can classify voice authentication systems into 3 categories:

- *Fixed-phrase verification*: **P** trains the system on a phrase that will also be used for testing. This mode is technically easy, but insecure (someone could record and play back **P** saying the phrase).
- *Prompted-phrase verification*: at test time, the system prompts **C** to say a word sequence not used for enrolment or for previous tests. The system knows the test phoneme sequence, yet impostors cannot use recordings. This mode requires an interface for prompting **C**.
- *Text-independent verification*: **C** speaks freely. This mode is technically difficult; it is used for applications with little control over user input.

2.3. Features of Voice-Authentication

Features for voice authentication should be [3]:

- Practical (occur naturally and frequently in speech)
- Robust (not change over time, not be affected by channel or reasonable background noise)
- Secure (not be subject to mimicry).

Current speaker verification systems are mainly based on cepstral acoustic features derived from the speech spectrum (if done over telephone, only the 300-3300 Hz range used). To explore longer-term speech features, NIST has added an "extended data" speaker detection task to its evaluation. Recent work in this area [4] shows that speaker word choices (bigrams like "you bet", "for sure") are also very speaker correlated.

2.4. Structure of Voice-Authentication Systems

Current systems are based on Hidden Markov Models (HMMs). Depending on the degree of text-dependence, various structures have been considered [5]:

- for fixed-phrase, one HMM that models the phrase;
- for prompted-phrase, a set of HMMs modeling phonemes (as in speech recognition);
- for text-independent, single-state HMM with many Gaussians (Gaussian Mixture Model = GMM).

2.5. Impostor Models and Score Normalization Methods

Current voice authentication systems make use of a set of models **I** which represent the population around the space of the rightful speaker **P** (cohort models) or simply the general speaker population (Universal Background Model or UBM). Then a Log-likelihood test on speech ***S*** with threshold *T* for each model **I** is used:

- *log p(**S**|model of **P**) – log p(**S**|model of **I**) > T* =>**accept,** else **reject.**

State-of-the-art systems currently tend to use big UBMs (2000 or more Gaussians) which are adapted to train the model for **P** [6]. This supports fast scoring (the scoring function only needs to look at Gaussians that were adapted when the **P** model was trained). To deal with the substantial mismatch between the enrolment and the test conditions, score normalization methods have been proposed [6][7][8].

2.6. Adaptation of Speaker Models

Because the environment and even the speaker's voice characteristics may change over time, one can adapt the model for **P,** when one is sure that the current speaker is **P**. Maximum a posteriori probability (MAP) adaptation combined with confidence weighting improved authentication performance under channel mismatch conditions by 61%, despite impostor attacks [9]. To deal with a sparse amount of adaptation data, speaker-space methods can be used, similar to the ones used in speech recognition [10].

3. FINGERPRINT AUTHENTICATION

A fingerprint is the pattern of ridges and furrows on the surface of a fingertip. These patterns are *unique* and *permanent*. Identical twins have different fingerprints [11] and for the same person, fingerprints are different from hand to hand and finger to finger. Fingerprint recognition is one of the most mature biometrics and has been used since the beginning of the 20th century in forensics. Due to its criminal connotation, most users do not easily accept it. In the following subsections, we briefly summarize: 1) fingerprint acquisition, 2) classification, and 3) traditional matching techniques.

3.1. Fingerprint Acquisition

A fingerprint can be either an *inked* or a *live-scan* fingerprint. In the first case, the finger is evenly coated with a thin layer of ink, then "rolled" or "dabbed" on a sheet of paper that can be scanned. Live-scan fingerprints are obtained without an intermediate medium like paper. Live-scan acquisition systems are optical, thermal, electromagnetic or ultrasound based. A detailed description and a comparison testing of these systems can be found in [12]. Quality fingerprint acquisition is extremely challenging due to elastic distortion of the finger on the acquisition surface, dry skin, worn-out ridges, or the presence of scars on the finger [13].

3.2. Fingerprint Classification

Global patterns of ridges and furrows form special configurations in the central region of fingerprints. These patterns can typically

be assigned to one of a small (usually six) pre-specified number of classes: *arch*, *tented arch*, *right loop*, *left loop*, *twin loop* and *whorl.* The class information is not sufficient to carry out recognition. However, it can be used for clustering: once a fingerprint is classified, it can be matched only with a subset of the database. An overview of fingerprint classification approaches can be found in [14].

3.3. Fingerprint Matching

The uniqueness of a fingerprint is determined by the *local* ridge characteristics called *minutiae*. Usually two types of minutiae are used for their robustness and stability: ridge *ending* and *bifurcation*. Most automatic fingerprint matching algorithms mimic the process used by forensic experts to perform recognition: minutiae are first extracted to form a template and then matched with another template. Minutiae templates offer a compact representation of the fingerprint. The steps of a typical minutiae extraction algorithm are: 1) orientation estimation, 2) segmentation, 3) ridge detection and thinning, 4) minutiae detection and 5) post-processing (discard spurious minutiae) [13],[15]. During the matching phase, the relative locations and orientations of the minutiae are compared with another template. In [13], this is performed via a string-matching algorithm. However, [11] exposed the shortcomings of the traditional minutiae representation and a very promising representation combining global and local information was explored in [16].

4. FACE AUTHENTICATION

Over the last twenty years, numerous algorithms have been proposed for face recognition [17]. The oldest ones were *geometric feature-based methods*. Despite their economical representation and their insensitivity to variations in illumination and viewpoint, such methods are very sensitive to the feature extraction process. Alternative to feature-based techniques are the *appearance-based methods*, such as the *Eigenfaces* [18][19], the *Fisherfaces* [20], etc. Eigenfaces rely on *the Karhunen-Loeve* (KL) transform or *Principal Component Analysis* (PCA) and produce the so-called *most expressive features* (MEFs) that are well-suited for an optimal low-dimensional face representation in the least-squares sense, i.e., for encoding or compression. For pattern classification, that is, face recognition or authentication tasks, we seek features that offer a clear separation between the pattern classes, the so-called *most discriminating features* (MDFs) that are provided by *Linear Discriminant Analysis* (LDA) or *Fisher Linear Discriminant* (FLD) [21]. Fisherfaces stem from the latter approach. There has been a tendency to prefer LDA over PCA because, as explained above the former deals directly with discrimination between classes, where as the latter aims at faithfully representing the data. It has been shown that LDA outperforms PCA only when large and representative training data sets are given [22]. A combined use of PCA and LDA like methods has frequently been proposed to cope with the curse of dimensionality problem [21]. However, one should bear in mind that LDA gives the optimal linear discriminant among faces when the distribution of each class is Gaussian. An inherent drawback of appearance-based methods is that the recognition of a face under a particular lighting and pose can be performed reliably when the face has been previously seen under similar circumstances. To alleviate this drawback, the use of generative models, that are able to synthesize novel images under changes in lighting and viewpoint based on a small number of training images, was proposed [23]. Another powerful face recognition technique, whose origin can be traced back in the neural network community, is the *dynamic link architecture* (DLA) [24]. A simplified implementation of dynamic link architecture, the so-called elastic graph matching (EGM), is often preferred for locating objects in a scene with a known reference [25].

Although the algorithms employed in both face recognition and face authentication are of common origin (for example, EGM), the evaluation methodologies (or experimental protocols) to assess their performance and the databases needed to conduct the experiments differ. Face recognition experiments are usually tested on the FERET database using the FERET evaluation methodology [2]. For face verification, two databases are more appropriate, namely the M2VTS database that contains 37 persons' video data which include speech consisting of uttering digits and color image sequences of rotated heads recorded in four sessions [26] and its extended XM2VTS version of 295 persons' video data [27]. A number of face authentication algorithms were developed and tested on the M2VTS and XM2VTS databases using the same experimental protocols such as gray level frontal face matching [28], EGM with local discriminants [29], optimized robust correlation [30], EGM that employs either multiscale dilation-erosion and combines linear projections at the graph nodes [31][32], or morphological signal decomposition [33] or weighting coefficients derived by reformulating Fisher's discriminant ratio to a quadratic optimization problem subject to a set of inequality constraints [34], and support vector machines [35].

5. FUSION TECHNIQUES

Multi-modal biometrics is a conventional decision fusion problem, where the evidence provided by each biometric is combined to improve the overall accuracy [1]. The geometric average and a HyperBF network were used to combine the normalized outputs of two different speech classifiers and three different face classifiers [36]. Commonly used classifier combination schemes such as the product rule, sum rule, min rule, max rule, median rule, and the majority rule were derived from a common theoretical framework under different assumptions by using different approximations [37]. Kittler *et al.* demonstrated that the sum rule outperforms the other classifier combination schemes when frontal face, face profile and voice biometrics are used. Bayes theory was used to estimate the biases of individual biometrics that were subsequently used to

calibrate and conciliate the decisions taken by the individual biometrics to a single decision [38]. Hard fusion schemes, such as the application of *AND/OR* operators on the receiver operating characteristics of individual biometrics as well as a linear combination of the individual scores were studied in [28]. A fusion scheme which integrates face, lip motion, and voice was proposed in [39]. Clustering algorithms such as the fuzzy *K*-means, fuzzy vector quantization algorithms, and a median radial basis function network were proposed for decision fusion in [40]. Support vector machines using polynomial kernels and Bayesian classifiers were shown to outperform Fisher's linear discriminant, C4.5 decision trees, and multilayer perceptrons in binary classification approaches applied to vectors comprising the decision scores provided by several face and voice verification modalities [41].

All the aforementioned decision fusion schemes aim at improving the verification accuracy in a multi-modal identification/verification system. However, for identification systems, there is a need to perform one-to-many comparisons to find a match. Consequently, one has to integrate biometrics that complement each other not only in identification accuracy, but in terms of identification speed as well. In [42], face recognition, a biometric technique not extremely reliable but suitable for database retrieval, is used to index the template database and fingerprint verification, which is reliable in deterring impostors, is used to ensure the overall identification accuracy.

6. CONCLUSIONS

Biometrics, an emerging field of signal processing, are exploited in complementary fashion to develop authentication technologies that are both accurate and not intrusive for the user. In the future, they will play a key role in access control enhancing security residing in smart cards and supporting personalized web e-commerce services. Personalization through person authentication is expected to be very appealing in the consumer product area as well.

7. REFERENCES

[1] A. Jain, R. Bolle, and S. Pankanti, Eds., *Biometrics. Personal Identification in Networked Society*. Boston, MA: Kluwer Academic, 1999.

[2] P.J. Phillips, H. Moon, S.A. Rizvi, and P.J. Rauss, "The FERET Evaluation Methodology for Face Recognition Algorithms", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090-1104, Oct. 2000.

[3] J. Wolf, "Efficient Acoustic Parameters for Speaker Recognition", *Journal of the Acoustical Society of America*, vol. 51, pp. 2044-2056, 1972.

[4] G. Doddington, "Speaker Recognition Based on Idiolectal Differences between Speakers", in Proc. *Eurospeech 2001*, vol. 4, pp. 2521-2524, Aalborg, Denmark, Sept. 3-7, 2001.

[5] D. Reynolds and L. Heck, "Speaker Verification: From Research to Reality", ICASSP Tutorial, *ICASSP-2001*, Salt Lake City, Utah, May 7, 2001.

[6] D. Reynolds, "Comparison of Background Normalization Methods for Text-Independent Speaker Verification", in Proc. *Eurospeech-97*, vol. 2, pp. 963-966, Rhodes, Greece, Sept. 1997.

[7] J. Koolwaij, "Fundamentals of HMM Based Speaker Verification", http://www.ispeak.nl/start.html

[8] R. Auckenthaler, M. Carey, and H. Lloyd-Thomas, "Score Normalization for Text-Independent Speaker Verification Systems", *Digital Signal Processing*, vol. 10, pp. 42-54, 2000.

[9] L. Heck and N. Mirghafori, "On-Line Unsupervised Adaptation in Speaker Verification", in Proc. *ICSLP-2000*, vol. 2, pp. 454-457, Beijing, China, Oct. 2000.

[10] O. Thyes, R. Kuhn, P. Nguyen and J-C. Junqua, "Speaker Identification and Verification Using Eigenvoices", in Proc. *ICSLP-2000*, vol. 2, pp. 242-245, Beijing China, Oct. 2000.

[11] S. Prabhakar, "Fingerprint Classification and Matching Using a Filterbank", PhD thesis, Michigan State University, 1998.

[12] R. Adhami and P. Meenen, "Fingerprinting for Security", *IEEE Potentials*, vol. 20, no. 3, pp. 33-38, Aug.-Sept. 2001.

[13] A. Jain and S. Pankanti, "Automated Fingerprint Identification and Imaging Systems", *Advances in Fingerprint Technology*, 2nd Ed., Elsevier Science, New York, 2001.

[14] L. Chung Ern and G. Sulong, "Fingerprint Classification Approaches", in Proc. *ISSPA*, vol. 1, pp. 347-350, Kuala Lumpur, Malaysia, 13-16 Aug. 2001.

[15] A. Jain and S. Pankanti, "Fingerprint Classification and Recognition", in *The Image and Video Processing Handbook* (A. Bovik, Ed.), Academic Press, April 2000.

[16] A. Jain and S. Prabhakar, "Fingerprint Matching Using Minutiae and Texture Features", in Proc. *ICIP-2001*, pp. 282-285, Greece, Oct 7-10 2001.

[17] R. Chellappa, C.L. Wilson, and S. Sirohey, "Human and Machine Recognition of Faces: A Survey", *Proceedings of the IEEE*, vol. 83, no. 5, pp. 705-740, May 1995.

[18] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve Procedure for the Characterization of Faces", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 103-108, Jan. 1990.

[19] M. Turk and A. Pentland, "Eigenfaces for Recognition", *J. Cognitive Neuroscience*, vol. 3, no. 1, pp. 71-86, 1991.

[20] P.N. Belhumer, J.P. Hespanha, and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, July 1997.

[21] D.L. Swets and J. Weng, "Using Discriminant Eigenfeatures for Image Retrieval", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 831-836, August 1996.

[22] A.M. Martinez and A.C. Kak, "PCA versus LDA", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228-233, Feb. 2001.

[23] A. Georghiades, P.N. Belhumeur, and D.J. Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643-660, June 2001.

[24] M. Lades, J.C. Vorbruggen, J. Buhmann, J. Lange, C.v.d. Malsburg, R.P. Wurtz, and W. Konen, "Distortion Invariant Object Recognition in the Dynamic Link Architecture", *IEEE Trans. on Computers*, vol. 42, no. 3, pp. 300-311, Mar. 1993.

[25] J. Zhang, Y. Yan, and M. Lades, "Face Recognition: Eigenface, Elastic Matching and Neural Nets", *Proceedings of the IEEE*, vol. 85, no. 9, pp. 1423-1435, Sep. 1997.

[26] S. Pigeon and L.Vandendorpe, "The M2VTS multimodal Face Database", *Lecture Notes in Computer Science: Audio- and Video-Based Biometric Person Authentication* (J. Bigun, G. Chollet, and G. Borgefors Eds.), vol. 1206, pp. 403-409, 1997.

[27] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, "XM2VTSDB: The Extended M2VTS Database", in *Proc. 2nd Int. Conf. Audio- and Video-Based Biometric Person Authentication* (R. Chellappa, Ed.), pp. 72-77, Mar. 1999.

[28] S. Pigeon and L. Vandendorpe, "Image-based Multi-Modal Face Authentication", *Signal Processing*, vol. 69, pp. 59-79, August 1998.

[29] B. Duc, S. Fischer, and J.Bigun, "Face Authentication with Gabor Information on Deformable Graphs", *IEEE Trans. Image Processing*, vol. 8, no. 4, pp. 504-516, Apr. 1999.

[30] J. Matas, K. Jonsson, and J. Kittler, "Fast Face Localization and Verification", *Image and Vision Computing*, vol. 17, pp. 575-581, 1999.

[31] C. Kotropoulos, A. Tefas, and I. Pitas, "Frontal Face Authentication Using Discriminating Grids with Morphological Feature Vectors", *IEEE Trans. Multimedia*, vol. 2, no. 1, pp. 14-26, Mar. 2000.

[32] C. Kotropoulos, A. Tefas, and I. Pitas, "Morphological Elastic Graph Matching Applied to Frontal Face Authentication Under Well-Controlled and Real Conditions", *Pattern Recognition*, vol. 33, no. 12, pp. 1935-1947, Dec. 2000.

[33] A. Tefas, C. Kotropoulos, and I. Pitas, "Face Verification Using Elastic Graph Matching Based on Morphological Signal Decomposition", *Signal Processing*, 2002, to appear.

[34] A. Tefas, C. Kotropoulos, and I. Pitas, "Using Support Vector Machines to Enhance the Performance of Elastic Graph Matching for Face Authentication", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 7, pp. 735-746, July 2001.

[35] J. Matas, M. Hamouz, K. Jonsson, J. Kittler, Y. Li, et al. "Comparison of face verification results on the XM2VTS database", in Proc. *15th Int. Conf. Pattern Recognition*, pp. 858-863, 2000.

[36] R. Brunelli and D. Falavigna, "Person Identification Using Multiple Cues", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, no. 10, pp. 955-966, October 1995.

[37] J. Kittler, M. Hatef, R.P.W. Duin, and J. Matas, "On Combining Classifiers", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226-239, Mar. 1998.

[38] B. Duc, E.S. Bigun, J. Bigun, G. Maitre, and S. Fischer, "Fusion of Audio and Video Information for Multi-Modal Person Authentication", *Pattern Recognition Letters*, vol. 18, pp. 835-843, 1997.

[39] R.W. Frischholz and U. Dieckmann, "BioID: A Multimodal Biometric Identification System", *IEEE Computer*, vol. 33, no. 2, pp. 64-68, February 2000.

[40] V. Chatzis, A.G. Bors, and I. Pitas, "Multimodal Decision-Level Fusion for Person Authentication", *IEEE Trans. Systems, Man and Cybernetics, Part A*, vol. 29, pp. 674-680, Nov. 1999.

[41] S. Ben-Yacoub, Y. Abdeljaoued, and E. Mayoraz, "Fusion of Face and Speech Data for Person Identity Verification", *IEEE Trans. Neural Networks*, vol. 10, no. 5, pp. 1065-1074, Sep. 1999.

[42] L. Hong and A. Jain, "Integrating Faces and Fingerprints for Personal Identification", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 12, pp. 1295-1307, Dec. 1998.