ELSEVIER
Volume 27, issue 3, 2 February 2009 ISSN 0262-8856
Special Issue:
Multimodal Biometrics
image
AND
vision
COMPUTING
Available online at
ScienceDirect
www.sciencedirect.com

Available online at www.sciencedirect.com

ScienceDirect

Image and Vision Computing 27 (2009) 269–278

www.elsevier.com/locate/imavis

Handwriting verification – Comparison of a multi-algorithmic and a multi-semantic approach

Tobias Scheidat *, Claus Vielhauer, Jana Dittman

Department of Computer Science, Institute of Technical and Business Information Systems, Research Group Multimedia and Security, Otto-von-Guericke-University of Magdeburg, Universitätsplatz 2, 39106 Magdeburg, Germany

Received 3 May 2006; accepted 18 March 2007

Abstract

In this paper, a comparison of an existing multi-algorithmic and a new multi-semantic fusion approach for biometric online handwriting user verification is presented. First, in order to improve the authentication performance of a biometric online handwriting system four classification algorithms are combined using several weighting strategies for matching score level fusion. Second, based on the best two algorithms and the best weighting strategy found during the test of the multi-algorithmic approach, a new multi-semantic fusion approach using a pair wise combination of four semantics on matching score level is proposed. As semantics we understand alternative handwritten contents (e.g. symbols) in addition to signatures. We show that both fusion approaches, multi-algorithmic and multi-semantic, can lead to a fusion result which is better than the result of the best single algorithm or semantics involved. While the improvement for the multi-algorithmic system yields 19%, we observe more than 57% for the multi-semantic approach.

Keywords: Biometrics; Distance measures; Multi-algorithmic fusion; Multi-semantic fusion; Online handwriting

1. Introduction

Nowadays, authentication of persons and information has a great importance for data access and exchange of data. The traditional methods for user authentication are based on secret knowledge or personal possession. The disadvantages of these methods are that the authentication object can be stolen, lost or handed over, because the presented knowledge or object is authenticated, but it is not assured that the bearer is the real owner. On the other hand, a biometric system authenticates the user itself based on individual physiological and behavioural characteristics. Biometric traits can be divided into static or offline characteristics (e.g. fingerprint, iris) and dynamic or online characteristics (e.g. handwriting, speech). In general offline traits are based on physiological and online traits are based on behavioural characteristics of human beings.

By the fact that biometric data cannot be taken identically every time due to different sensors or circumstances, there is a random fuzziness between reference data and authentication data. A biometric system must be able to compensate these deviations, for example, by the use of thresholds, which fix the area of a successful authentication. This similarity of a single person's data is important for recognition, whereas the dissimilarity is also important for the biometric characteristics of different persons to keep the users apart.

In general a biometric system consists of four modules as shown in Fig. 1. The sensor module captures the physiological or behavioural characteristic of the user, and the feature extraction module determines the feature set from the captured data. This representation of the biometric characteristic is used by the matching module in order to calculate a score of similarity or dissimilarity between authentication and reference data. Using the matching

* Corresponding author. Tel.: +49 391 67 20123; fax: +49 391 67 18110.
E-mail addresses: tobias.scheidat@iti.cs.uni-magdeburg.de (T. Scheidat), claus.vielhauer@iti.cs.uni-magdeburg.de (C. Vielhauer), jana.dittmann@iti.cs.uni-magdeburg.de (J. Dittman).

0262-8856/$ - see front matter © 2008 Published by Elsevier B.V.
doi:10.1016/j.imavis.2007.03.006

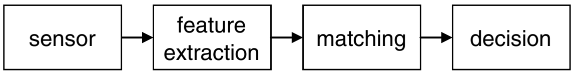

Fig. 1. User authentication based on Biometric Hash.

score, the decision module decides, whether a person is the one declared (verification) or who this person is (identification).

In order to solve the mentioned problems of biometric systems, some approaches attempt to reach a better performance, higher security level and user acceptance by combination of various biometric experts and/or modalities. This combination of different biological characteristics claims to adapt the human recognition system where multiple significant features, e.g. face, voice or typical behaviour, are used to recognize known persons. Biometric systems or algorithms can be combined with each other at different levels of the authentication process. Further four scenarios of the fusion using the sensors, involved units or algorithms can be differentiated. A description of scenarios and classification of fusion approaches can be found in Section 2.

Our approach is to study the recognition accuracy of a multi-algorithmic system and a multi-semantic approach. While the first is based on a method suggested in our earlier work [1], the multi-semantic approach is a new method introduced in this article. In a multi-algorithmic system two or more experts of one biometric modality are fused. The multi-semantic system fuses different semantics (alternative written contents, e.g. password) of the handwriting in order to reach an improvement of the verification performance. This suggestion unites two different versions of a single biometric modality, where these versions can be based on secret information.

This paper is structured as follows: Section 2 presents a summary of the state-of-the-art about multimodal and multi-algorithmic biometrics. This is a composition from the view of the authors without neglecting other not mentioned publications. In the third section, we discuss the fundamentals of the multi-algorithmic and multi-semantic biometric approaches we developed and evaluated. Also the basics of the used algorithm and the fusion strategies of our systems are presented in Section 3. Section 4 gives an overview of the test database and the test methodology. The test results and a discussion of their meaning are shown in Section 5. A short summary of this paper and an outlook of future work are given in Section 6.

2. State of the art

A multimodal system is based on one of three fusion levels [2] depending on the point of fusion: feature extraction level, matching score level or decision level.

In the *feature extraction level*, the information extracted from the different sensors and gained from the feature extractors are stored in separate feature vectors. These feature vectors are combined to a joint feature vector, which is the basis for the matching process. In some cases this results in a very high dimensional joint feature vector.

On *matching score level*, the fusion is based on the combination of matching scores after the comparison of reference data and test data. In this strategy an assessment of the systems involved by individual weights is possible. This weighting is generally simple since every system contributes one single value. The fusion results in a new matching score, which is the basis for decision.

With the fusion on the *decision level*, each biometric subsystem involved is completely processed. Afterwards, the individual decisions are combined to a final decision, e.g. by Boolean operations. The influence on the fusion is very small at the decision level, since the fusion is carried out at a very late time within the authentication process.

In the following, we discuss current publications based on the scheme of fusion scenarios proposed by Ross and Jain, and indicate the level of the fusion where possible. Ross and Jain differentiate in [2] between the following scenarios for automatic biometric fusion, based on the number of biometric traits, sensors, classifiers and units involved:

(1) *Single biometric trait, multiple sensors*
In this scenario multiple sensors capture one biometric trait.
Chang et al. employ 3D range data in combination with 2D image for face recognition [3]. They report a recognition rate of approximately 99% for the fusion of both approaches on matching score level based on 200 test subjects. The recognition rate is 94% for the 3D part and 98% for the 2D part of the system described. In [4] Kumar et al. describe the fusion on matching score level of palm print verification system and hand geometry system. Based on a test set of 100 individuals an improvement of the false acceptance rate is achieved from 4.49% for palm print and 5.29% for hand geometry to 0% for fusion. By the fusion the false rejection rate could be reduced from 2.04% for system 1 or 8.34% for system 2 to 1.41%.

(2) *Single biometric trait, multiple classifiers*
This category of systems is based on biometric characteristics of only one biometric modality whereby different (independent) experts are consulted for the authentication.
Czyz et al. use in [5] five face verification experts in several combinations. For the test they use a database of 295 individuals (200 clients, 95 impostors) and reached an improvement up to 45% for the best expert combination in comparison with the single experts. In [6] Ly Van et al. describe an online signature verification based on a fusion of HMM's Likelihood and Viterbi Path on matching score level. The fusion decreases the individual EERs from 6.45% (Likelihood), respectively, 4.07% (Viterbi Path) to a combined EER from 2.84%. The BIOMET database, they used for evaluation, contains 1266 genuine signatures by 87 persons. Scheidat et al. describe in [1] a

signature verification system based on four experts, fused on matching score level using a test set of 22 individuals with 1800 reference and 1100 verification samples. There the best fusion strategy results in a decrease of the EER of 12.1% in comparison to the best individual algorithm. We will confirm our earlier approach in this paper by the use of additional test data and we use another three alternative semantics in addition to the signature here. In [7] four systems based on online handwriting are fused on matching score level by Garcia-Salicetti et al., using a database of 330 users. The equal error rate of the fusion (EER = 0.0122) is better than the results of each single system. In Section 5.1 we will compare these results with those of our multi-algorithmic system more precisely.

(3) *Single biometric trait, multiple units*
The user presents the system several different versions of the same biometric trait (e.g. prints of more than one finger).
Jain et al. show in [8] that by the combination of prints of two fingers or two versions of one finger improvements are possible. The evaluations are based on a test set of 160 persons. In [9] the authors describe a system which uses a combination of two fingerprints at the feature extraction level. During the tests an equal error rate of 1.9% is determined using a test database of 100 users, each finger captured two times.No methods were found by the authors, which use multiple units of the same single dynamic biometric modality (e.g. speech, online handwriting) for user authentication. Our multi-semantic approach starts here to use two different written contents of the online handwriting to the verification.

(4) *Multiple biometric traits*
Here, several biometric subsystems of different modalities decide on the authenticity of the user. Ly Van et al. [10] combine signature verification with text dependent and text independent speech verification, at a time. They report that fusion increases the performance by a factor 2 relatively to the best individual system. The test dataset contains five genuine bimodal values and 12 impostor bimodal values for each of 68 individuals. Jain and Ross present in [11] a biometric system that uses face, fingerprint and hand geometry of a user for authentication. The fusion on matching score level improves the three individual results considerably here. Vielhauer et al. present in [12] a multimodal system where a speech recognition system and a signature recognition system are fused on matching score level. In [13] an enhancement of the multimodal system by exchange of the single signature component by the multi-algorithmic handwriting subsystem proposed in [1] is suggested. This system is based on a combination of four different handwriting authentication experts. Finally, an improvement of approximately 15% for the overall system could be reached by the exchange described.

The online modality of handwriting is generally associated with signature verification in context of biometric user authentication. Contrary to other biometric modalities, the handwriting provides the possibility of an exchange of the written content, for example, in order to replace a compromised content or to use secret knowledge additionally to the writing behaviour. Schmidt described in [14] a database containing also writing samples of the same textual content (German word "Grünschnabel") in addition to the handwriting category of signatures. In [15] Kato et al. investigate the possibility to apply user-specific, more or less complex drawings as authentication information. In our approaches alternative handwritten contents in addition to signature are named as semantics. They are based on different properties and their combinations, such as secret knowledge, predefined textual contents, creativity and individuality during the writing process. In different publications (e.g. [16,17]) it has been shown that it is possible to use alternative contents for the authentication.

3. Multi-algorithmic and multi-semantic approaches

The idea of our multi-algorithmic fusion is the utilization of a given feature extraction scheme in combination with different distance measures. The matching module of the original Biometric Hash algorithm [16,18], which was conceived for the calculation of biometric hashes, determines the value of the dissimilarity using the Hamming Distance (HD). In order to create different algorithms, the matching module was extended by three additional distance functions: City Block Distance (CBD), Canberra Distance (CD) and Euclidian Distance (ED). In this section, the functionality of the algorithm and the distance measures used are described. The fusion of the algorithms shall be carried out using weighted scores on the matching level. For this, different strategies are introduced, based on the equal error rates (EER) of the individual algorithms to estimate the weights. The fusion itself represents a sum rule, where the matching scores multiplied by the weights are added. At the end of this section, a new method for the fusion of semantics of the biometric handwriting is introduced and the operation explained. To abstract their functionality within the fusion process, in the following we denote the algorithms or semantics also *as fusion components* or only *components*.

3.1. Matching algorithms

3.1.1. Biometric Hash algorithm

The verification algorithm for the modality of online handwriting is based on the Biometric Hash algorithm, as introduced in detail in [16,18]. In summary, this method calculates a statistical feature vector of $k = 69$ statistical parameters (online and offline features), which are transformed into the hash value space by an interval mapping function. This mapping, denoted as Key Generation,

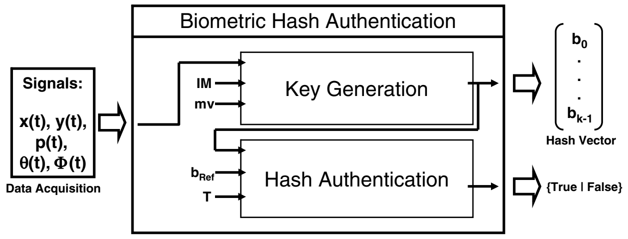

Fig. 2. User authentication based on Biometric Hash (original taken from [16]).

results in a feature vector representation $\vec{b} = (b_0, \ldots, b_{k-1})$ supported by a user-specific statistical model, consisting of an Interval Matrix (IM) and a Masking Vector (mv), which is obtained during enrolment process.

During verification, five discrete signals based on measurements of horizontal and vertical pen position $x(t)$ and $y(t)$, pen tip pressure $p(t)$ and pen azimuth and altitude $\Theta(t)$ and $\Phi(t)$, respectively, are taken from the digitizer tablet as shown in the left part of Fig. 2.

Based on these five signals, the Key Generation module will calculate an actual feature vector $\vec{b}$, which is compared to a stored reference vector $\vec{b}_{\mathrm{Ref}}$ against some decision threshold value T in the Hash Authentication Module. In the initial version of the algorithm, this authentication is performed by calculation of the Hamming Distance between the two vectors. Finally, this verification method results in a binary True/False decision with respect to the actual biometric data and the given threshold.

3.1.2. Distance measures

Amongst the numerous feature distance measures, we have chosen four selected reference functions for our first evaluation: Canberra, City Block (or Manhattan), Euclidian and Hamming Distance. The mathematical functions are described briefly in this subsection. For the descriptions, we instantiate two Biometric Hash vectors $x = (x_0, \ldots, x_{k-1})$ and $y = (y_0, \ldots, y_{k-1})$, each of integer value and dimensionality k ($k = 69$ in our test scenarios). Smaller distance between any two vectors x and y denotes greater similarity than larger.

Canberra Distance

The Canberra Distance calculates the sum of a set of ratios between appropriate values. It considers the distance between two points but also their relation to the origin

$$\mathrm{cd}(x,y) = \sum_{i=o}^{k-1} \frac{|x_i - y_i|}{|x_i| + |y_i|}$$

The result is in the interval $[0, k-1]$, in our system k is equal to 69 for 69 statistical features of one handwriting sample.

City Block Distance

The City Block Distance is the sum of the single distances along each dimension. It is based on the idea of a city walk, at which only right-angled direction changes are possible to reach the end-point

$$\mathrm{cbd}(x,y) = \sum_{i=0}^{k-1} |x_i - y_i|$$

The range, in that the value lies, cannot be predicted. Therefore, it must be normalized on the desired interval.

Euclidian Distance

The Euclidian Distance is general the shortest connection between two points

$$\mathrm{ed}(x,y) = \sqrt{\sum_{i=o}^{k-1} (x_i - y_i)^2}$$

The maximum size of the distance value cannot be indicated before and normalization is necessary too.

Hamming Distance

With the Hamming Distance, each of the components of the two Biometric Hash vectors, having the same index, are compared with each other. If they are identical, the result of the comparison is *0*, in the other case *1* and the distance is the sum of the single results. For this reason the distance is at least *0* and at the most $k-1$.

3.1.3. Distance normalization

Since the scales of the City Block Distance and Euclidian Distance functions are in a non-predictable interval, normalization is necessary. Because our initial reference distance functions, Canberra Distance and Hamming Distance, result in integer values within the interval $[0, k-1]$, we normalize all other distance values p_t to this interval according to the following transformation function T:

$$T(p_i) = \alpha p_i + \beta$$

The parameters α and β can be determined using the following system of equations:

$$T(p_{\min}) = \alpha p_{\min} + \beta = I_{\min} \quad (1)$$

$$T(p_{\max}) = \alpha p_{\max} + \beta = I_{\max} \quad (2)$$

where $p_{\min}$ and $p_{\max}$ are the borders of the original interval and $I_{\min}$ and $I_{\max}$ the borders of the targeted interval, i.e. *0* and $k-1$, respectively.

3.2. Fusion strategies

The multi-algorithmic fusion is carried out on the matching score level and is based on the combination of the four different Biometric Hash algorithms by means of different weighting strategies [1]. The multi-semantic fusion is based on one algorithm each and two different semantics classes.

3.2.1. Weighting parameter estimation

After creating the individual experts, we developed several weighting strategies for combining their results. Four strategies for weighting the match scores were developed, which are based on the EERs of the tests of the individual fusion components. For each weighting strategy the following characteristics are important, where n is the number of components involved:

$$\text{Match Scores}: \; s_1, s_2, \ldots, s_n$$
$$\text{Weights}: \; w_1, w_2, \ldots, w_n$$

Equal weighted fusion

The first strategy is an equal weighting tactic, witch provides all components involved independently of the determined EER with the same weight. In this case the value is 0.25 for each of the four algorithms and 0.5 for the two different semantics, respectively

$$\begin{aligned}\text{Conditions}: \quad & w_1 + w_2 + \cdots + w_n = 1\\ & w_1 = w_2 = \cdots = w_n = n^{-1}\\ \text{Fusion}: \quad & s_{\text{fus}} = w_1 s_1 + w_2 s_2 + \cdots + w_n s_n\end{aligned}$$

Linear weighted fusion 1

With the first linear weighting approach the best fusion component is weighted in dependence to the worst algorithm. That means, the higher the EER of the worst component, the higher we define the weight for the best component. In the first step, components are sorted by their observed EER increasingly order. Then the individual weights are computed according to the following formula:

$$w_i = \frac{\text{eer}_i}{\sum_{m=1}^{n} \text{eer}_m}$$

In the last step, the determined weights are re-ordered and re-assigned in opposite direction, i.e. the fusion component having best EER will be assigned the highest weight and vice versa.

$$\begin{aligned}\text{Conditions}: \quad & w_1 + w_2 + \cdots + w_n = 1\\ \text{Fusion}: \quad & s_{\text{fus}} = w_1 s_1 + w_2 s_2 + \cdots + w_n s_n\end{aligned}$$

Linear weighted fusion 2

The linear strategy 2 depends on the size and the relationship of the EERs from the test of the individual fusion components.

$$\begin{aligned}\text{Conditions}: \quad & w_1 + w_2 + \cdots + w_n = 1\\ & w_i = \frac{\left(\sum_{j=1}^{n} \text{eer}_j\right) - \text{eer}_i}{\sum_{j=1}^{n} \text{eer}_j} \cdot \frac{1}{(n-1)}\\ \text{Fusion}: \quad & s_{\text{fus}} = w_1 s_1 + w_2 s_2 + \cdots + w_n s_n\end{aligned}$$

Quadratic weighted fusion

The quadratic weighted fusion strategy squares the weights determined by the linear weighted fusion strategy 1. The sum of the weights must be again 1.

$$\begin{aligned}\text{Conditions}: \quad & w_1 + w_2 + \cdots + w_n = 1\\ & w_i = \left(\frac{w_{\text{linear}1i}}{\sum_{j=1}^{n} w_{\text{linear}1_j}}\right)^2\\ \text{Fusion}: \quad & s_{\text{fus}} = w_1 s_1 + w_2 s_2 + \cdots + w_n s_n\end{aligned}$$

The set of weighting strategies used is a first selection of many more possibilities as identified in [1]. There are still many other strategies, which may result in better results.

3.3. Multi-algorithmic fusion

The multi-algorithmic system is based on biometric characteristics of the online handwriting, whereby different independent algorithms are consulted for the verification. For this purpose the strategies of the multimodal fusion can be used likewise. The verification decision here is based on a fusion strategy of the respective single results. Our approach combines the four distance measures, mentioned in Section 3.1, within a biometric system and is based on the matching score level strategy. A model for this fusion on matching score level for our system is represented in Fig. 3. The system is based on one single biometric trait and multiple classifiers as already described in Section 2.

The input data for all four algorithms are identical signals of the sensor, a digitizer tablet. They consist of physical characteristics of the specimen of handwriting over time. Each algorithm may use its own feature extraction, but in the current setup, it is identical for all four experts. Basis of the feature extraction of the four algorithms used is the Biometric Hash algorithm. Altogether, we implemented the four distance measures, described in Section 3.1, into the Biometric Hash algorithm, in order to create four different experts.

3.4. Multi-semantic fusion

The idea of the multi-semantic fusion bases on the combination of different semantics classes to improve the individual results. Four semantics, which have different attributes, are selected and combined in pairs: the *Signature* represents a traditional and well-accepted feature for the user authentication and it has individual and creative qualities. The *Symbol* also holds individual and creative features, however, it has an additionally knowledge based component in form of the sketched object. Further, all users have written the same predefined numeric string for the personal identification number (*PIN*, given as 77993). The dynamics of the handwriting can be analyzed mainly without a reference to secret knowledge here. The fourth semantic class is the answer to the question, where the test person comes from. Since the type of the answer among different test persons turns out variably (place of birth, place of domicile, native country), we denote this semantics generally as *Place*. This answer includes personal knowledge in a certain degree which, however, is not absolutely secret.

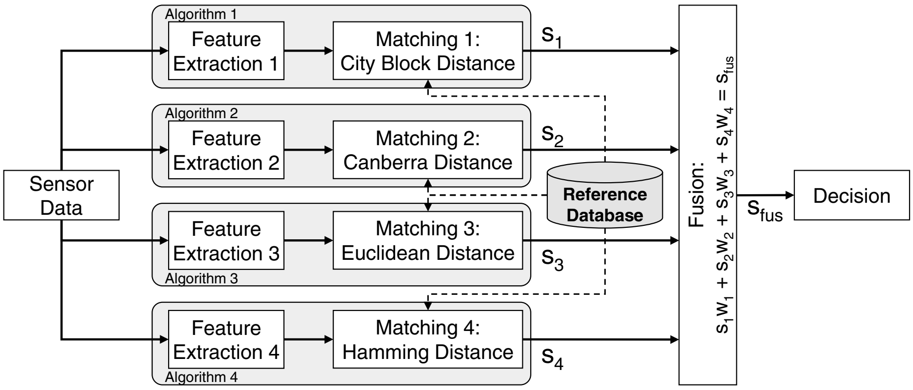

Fig. 3. Matching score level fusion of four biometric algorithms.

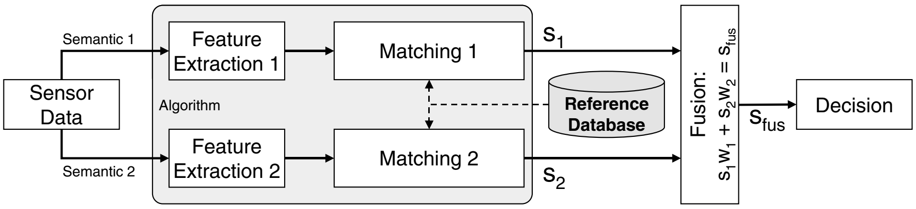

Fig. 4. Matching score level fusion of two biometric handwriting semantics.

Fig. 4 shows a general scheme of the multi-semantic fusion process on matching score level. For the two different semantics taken by the sensor, the matching scores are calculated by the same biometric algorithm, separately. In the fusion step the scores get weighted according to the chosen weighting parameter estimation strategy and totalized. The resulting sum is the input matching score for the decision module.

4. Experimental setup

This section describes the evaluation process, which is used to compare the authentication performance of the fusion of multiple algorithms or multiple semantics with those of the single fusion components. Therefore the underlying database and the evaluation methodology are presented.

4.1. Database

Our entire evaluation database of handwriting samples is structured in about 50 semantics classes captured by various graphic/signature tablets. Out of this entire dataset we choose the handwriting samples donated in the semantics classes Signature, Symbol, PIN and Place, acquired on two devices, the Wacom Cintiq l5 tablet and the Toshiba Portégé M200 tablet PC. Both of them are based on so-called active displays, where the tablet functionality is integrated into a computer display. Contrary to common graphical tablets, on active displays the digital representation of writing appears directly on the pen tip during the writing process, which leads to a higher quality of the resulting reference and verification data. The data were acquired by the three international partners of the EU project CultureTech [19]. The test persons from Germany, Italy and India followed a guideline that structured the 48 English tasks, which can be grouped in individual, creative and given tasks. The tasks can be subdivided in the signature, an arbitrary symbol, individual answers to questions, predefined numbers and given sentences. A summary of the semantics which are used in the database can be seen from [17].

4.2. Methodology

In order to compare the authentication performance of the individual components involved, as well as for their fusion, we use biometric error rates: the false non-match rate (FNMR) specifies, how often authentic persons are rejected from the system. How frequently non-authentic persons are accepted by the system, is indicated by the false match rate (FMR). The point of our interest in the error rates characteristics is the equal error rate (EER), where the values of both, FNMR and FMR, are identical. It needs to be stated however, that the EER do not represent the optimal operating point of the biometric system. The optimal operating point depends on other factors such as the desired level of security and/or comfort of the planned biometric system. The minimum and maximum values, which can be achieved by the error rates, are 0 and 1, respectively. A value of 0 means that no match error occurred during an evaluation, while 1 denotes in the worst case, that only match errors occurred.

In our evaluation protocol, 10 sequentially acquired handwriting samples ($H = \{H_1, \ldots, H_{10}\}$) are used, for each user in each of the selected 4 semantics classes.

From these sample set H, we build different subsets in order to generate references, estimate weighting parameters and carry out fusion based verifications as described follows.

4.2.1. Creation of the reference datasets

The first 4 samples $H_1,\ldots,H_4$ are taken from H to generate 4 reference data sets using a leave-one-out strategy, based on a combination of 4 choose 3. This means, we create 4 different reference datasets $R_1,\ldots,R_4$, which containing 3 handwriting samples each. The references created in such way, are used for both: the estimation of fusion weighting parameters and the determination of fusion verification performance.

4.2.2. Estimation of fusion weighting parameters

In order to determine the fusion weighting parameters based on the estimation strategies presented in Section 3.2, the reference datasets $R_1,\ldots,R_4$ and the samples H_5 and H_6 are used to calculate the EERs of each component (algorithm and/or semantic) involved into the fusion. Based on these EERs the corresponding weights are determined.

4.2.3. Determination of fusion verification performance

The FNMRs are calculated by the comparison of each reference dataset $R_1,\ldots,R_4$ of a user to the samples $H_7,\ldots,H_{10}$ of the same user, based on the corresponding component. The FMRs are determined based on the comparison of each reference set $R_1,\ldots,R_4$ of a user with samples $H_7,\ldots,H_{10}$ of all other users in the same semantics class, respectively. In this article, we do not study the influence of skilled forgeries on the multi-algorithmic or multi-semantic fusion. We focus on a closed verification scenario, containing only registered persons.

5. Results

In this section, we present and discuss the results of the individual fusion components and their multi-algorithmic or multi-semantic fusion, and we compare the results of our multi-algorithmic fusion strategy with those of a similar fusion approach also based on multiple handwriting verification experts.

5.1. Multi-algorithmic fusion

We have introduced initial results for the multi-algorithmic fusion of our algorithms in [1]. The differences between this first evaluation and the results presented here are the used databases. Here all data were taken under supervision and the same number of references or verifications were used for every user in addition in this publication. In our first tests [1] the data were acquired in unsupervised sampling sessions. At the evaluation all available references and verifications of each person were used, leading to great differences for single users with respect to the number of the samples. The initial database contains 1761 reference datasets (with 4 signatures per reference) and 1101 verification signatures. The best fusion strategy results in a decrease of the EER of 12.1% in comparison to the best individual algorithm (here the Canberra Distance).

In our new investigations, first we determined the error rates of the individual algorithms in order to determine the fusion weights as described in Section 4.2 and based on the weighting parameter estimation strategies introduced in Section 3.2. The authentication performance of the individual algorithms and their fusion are shown in Tables 1–4 in the columns titled *Single algorithms* and *Fusion Strategies*. The best result of each table is printed bold. Here we observe that the algorithms based on Canberra Distance

Table 1
EER of single biometric tests for Signature and of fusion of four algorithms

Single algorithms		Fusion strategies							
		Equal		Linear1		Linear2		Quadratic	
Name	EER	Weights	EER	Weights	EER	Weights	EER	Weights	EER
CBD	0.1689	0.2500	0.0328	0.1240	0.0255	0.2210	0.0308	0.0410	0.0224
CD	**0.0218**	0.2500		0.4990		0.3200		0.6550	
ED	0.2118	0.2500		0.0400		0.1670		0.0040	
HD	0.0759	0.2500		0.3370		0.2920		0.3000	

Table 2
EER of single biometric tests for Symbol and of fusion of four algorithms

Single algorithms		Fusion strategies							
		Equal		Linear1		Linear2		Quadratic	
Name	EER	Weights	EER	Weights	EER	Weights	EER	Weights	EER
CBD	0.2010	0.2500	0.0393	0.1430	0.0308	0.2140	0.0356	0.0550	0.252
CD	**0.0234**	0.2500		0.4710		0.3240		0.5960	
ED	0.2406	0.2500		0.0270		0.1760		0.0020	
HD	0.0825	0.2500		0.3590		0.2860		0.3470	

Table 3
EER of single biometric tests for PIN semantic class and fusion of four algorithms

Single algorithms		Fusion strategies							
		Equal		Linear1		Linear2		Quadratic	
Name	EER	Weights	EER	Weights	EER	Weights	EER	Weights	EER
CBD	0.2537	0.2500	0.0456	0.1860	0.0398	0.2230	0.0423	0.1050	**0.0372**
CD	0.0458	0.2500		0.4260		0.3150		0.5500	
ED	0.2775	0.2500		0.0560		0.1910		0.0100	
HD	0.0936	0.2500		0.3320		0.2710		0.3350	

Table 4
EER of single biometric tests for Place semantic class and fusion of four algorithms

Single algorithms		Fusion strategies							
		Equal		Linear1		Linear2		Quadratic	
Name	EER	Weights	EER	Weights	EER	Weights	EER	Weights	EER
CBD	0.1819	0.2500	0.0344	0.1410	0.0268	0.2210	0.0334	0.0520	0.0220
CD	**0.0217**	0.2500		0.4940		0.3240		0.6460	
ED	0.2180	0.2500		0.0290		0.1690		0.0030	
HD	0.0810	0.2500		0.3360		0.2860		0.2990	

(CD) and Hamming Distance (HD) calculate the best results for every semantics. The columns, containing weights and EERs of the four fusion strategies (*equal, linear1, linear2, quadratic)*, show, that the quadratic fusion strategy determines the best EER in comparison to all fusion strategies. In addition, for the semantics *PIN*, the results of the quadratic fusion approach is better than the result of the best single algorithm based on Canberra Distance (see Table 3). In comparison to Canberra and Hamming Distance the City Block (CBD) and Euclidian Distances (ED) yield worse results. For example, in Table 3 the worst result for the Euclidian Distance ($\text{EER}_{\text{ED}}(\text{PIN}) = 0.2775$) is more than six times higher than the value of the Canberra Distance ($\text{EER}_{\text{CD}}(\text{PIN}) = 0.0458$) for the same semantics.

In order to compare our actual results to a related method, we compare our results to a similar multi-algorithmic approach. Such a multi-algorithmic system, which is also based on the fusion on matching score level and four single systems for online handwriting is introduced in [7]. The authors describe fusion tests based on three reference systems (Sysl, Sys2, Sys3), delivered by three universities in the context of the BioSecure Network of Excellence [20] and one additional system (Sys 4), which comes from a fourth university. The fusion results are based on a mean rule using the matching scores after normalization. Table 5 shows a comparison of our multimodal system presented in this paper and the multi-algorithmic system described in [7]. In this table, the best EER of each approach is printed bold. Both approaches are based on four single biometric systems based on online signature features. The results of the single subsystems differ considerably from each other for both multi-algorithmic systems. For the BioSecure system the fusion determines a better overall equal error rate ($\text{EER}_{\text{fusion}}(\text{Signature}) = 0.0122$) than the best single subsystem ($\text{EER}_{\text{Sys1}}(\text{Signature}) = 0.0291$). On the other side, the fusion ($\text{ERR}_{\text{fusion}}(\text{Signature}) = 0.0224$) does not provide the best result for our system, however, it is ranked second and is quite close to the result of the best single algorithm ($\text{EER}_{\text{CD}}(\text{Signature}) = 0.0218$, see column CD). A possible explanation for the difference between the two methods is, that in our approach all algorithms are based on the same feature set, the Biometric Hash, while at the BioSecure approach different types of feature sets are extracted by the individual algorithms. However it can be mentioned that the equal error rates lie relative close to each other. Both approaches result in EERs of 0.0122 or 0.0224, which corresponds to a relative difference of approximately 53%. They are thus encouraging results for further research.

5.2. Multi-semantic fusion

Since the examination of our multi-algorithmic fusion has yielded, that the Hamming Distance and the Canberra

Table 5
Comparison of single and fusion results for signature of our multi-algorithmic system and the multi-algorithmic system presented in [7] by Garcia-Salicetti et al.

Strategy	Signature (our approach)				Signature (from [7])			
	CBD	CD	ED	HD	Sys1	Sys2	Sys3	Sys4
Single	0.1689	**0.0218**	0.2118	0.0759	0.0291	0.0918	0.1150	0.0430
Fusion	0.0224				0.0122			

Distance provide the best individual results, these two distances were used for the further examinations. Also the quadratic strategy for the weighting parameter estimation was used for the multi-semantic fusion. The idea of this fusion is the possibility of saving references of more than two semantics. During authentication, the system may then challenge two semantics randomly chosen. This way the security of the system can be increased since an attacker must predict the required combination.

The results of our EER tests for the single semantics and their fusions are shown in Table 6. In the first column the pair wise combinations of the semantic classes are shown, followed by the corresponding single results (EER_{single}), fusion weights *(Weights)* and the fusion result (EER_{fusion}) for the two algorithms, based on Canberra Distance and Hamming Distance, respectively. The last column shows the number of users, which are taken in consideration for result's determination. These values are different, because the groups of users, which have donated handwriting samples for the different semantics, are not fully identical. On the one side, the best result for the multi-semantic fusion using the Canberra Distance based version of the Biometric Hash algorithm was obtained by the combination of the Signature (EER_{CD}(Signature) = 0.0224) and Symbol (EER_{CD} (Symbol) = 0.0233), here the EER for the fusion is 0.0095. On the other side, Symbol is also involved in the worst fusion result (EER_{CD}(Symbol|PIN) = 0.0155) together with the PIN (EER_{CD} (PIN) = 0.0459). As in all other cases (see bold printed values in Table 6), here the fusion also improves the two corresponding individual results.

The best output for the single semantics and their fusions based on the Biometric Hash approach using Hamming Distance was reached by the fusion of the semantics Signature and Place with an EER of 0.0298, where the single results are EER_{HD}(Signature) = 0.0782 and EER_{HD}(Place) = 0.0835.

The analysis of the individual results shows that the PIN yields the worst rate. The reason for this observation could be the fact that the written number sequence is same for all subjects of a test thus likely to be visually and dynamically more similar, and may frequently result in false acceptances.

5.3. Comparison of the multi-algorithmic and the multi-semantic approach

Both methods of the fusion, multi-algorithmic and multi-semantic, obtain improvements in comparison with the results of the single algorithms involved. On one side, the multi-algorithmic fusion based on the quadratic fusion strategy obtains the best result in one of four cases. An improvement can be achieved by the fusion using the PIN (EER_{fusion}(PIN) = 0.0372) of at most 19% relatively to the best individual result (EER_{CD}(PIN) = 0.0458) here. On the other side, using the multi-semantic fusion, also based on quadratic fusion strategy, for every pair wise combination of semantics a more significant improvement could be reached. The best fusion result was achieved by fusion of semantics Signature (EER_{CD}(Signature) = 0.0224) and Symbol (EER_{CD}(Symbol) = 0.0233) using Canberra Distance (EER_{CD}(Signature|Symbol) = 0.0095). This corresponds to a relative reduction by more than 57%.

For an application scenario the administrator may have to decide between both approaches. The decision of the multi-algorithmic system based on four experts for handwriting verification. This fact can make the result more trustworthy. However, for all systems, process steps have to carry out up to the matching score calculation. This procedure is from the point of view of computational resources expensive. This results in higher energy consumption and may complicate the use in a mobile environment. Using the multi-semantic fusion for user authentication only one algorithm has to run two times and reaches significantly lower error probability. Through this, energy can be saved and the system is suitable better for mobile equipment than the multi-algorithmic approach. A problem here can be the acceptance of the users, since they need to carry out the writing process twice to be authenticated by the system.

Table 6
EERs of single biometric tests per semantic class and fusion of two semantic classes using Canberra Distance and Hamming Distance

Semantic class	Canberra Distance			Hamming Distance			Number of users
	EER_{single}	Weights	EER_{fusion}	EER_{single}	Weights	EER_{fusion}	
Signature Symbol	0.0224 0.0233	0.3870 0.6130	**0.0095**	0.0764 0.0802	0.6590 0.3410	**0.0365**	59
Signature PIN	0.0196 0.0450	0.6620 0.3380	**0.0112**	0.0818 0.0704	0.7580 0.2420	**0.0460**	56
Signature Place	0.0228 0.0238	0.2840 0.7160	**0.0117**	0.0782 0.0835	0.4730 0.5270	**0.0298**	59
Symbol PIN	0.0231 0.0459	0.7770 0.2230	**0.0155**	0.0820 0.0837	0.5810 0.4190	**0.0357**	61
Symbol Place	0.0233 0.0234	0.4290 0.5710	**0.0109**	0.0784 0.0793	0.3520 0.6480	**0.0383**	64
PIN Place	0.0429 0.0223	0.1750 0.8250	**0.0152**	0.0797 0.0765	0.3120 0.6880	**0.0472**	60

6. Conclusions and future work

The re-evaluation of our method suggested in [1] has again shown that the multi-algorithmic fusion reaches an improvement of the recognition performance. In one case the result of the multi-algorithmic fusion was better than the result of the best individual algorithm involved. The relative improvement was 19% in this best case for the quadratic fusion of all four algorithms using the semantics PIN. In addition we have observed for all four semantics that in comparison to the other three strategies the quadratic weighting strategy provides the best result at the multi-algorithmic fusion.

The results of the multi-semantic fusion show that for every possible pair wise combination of the semantics significantly better equal error rates are reached by the fusion than for every single semantics. The best result is an equal error rate of 0.0095 for the fusion of the semantics Signature and Symbol, this corresponds to a relative improvement of over 57%, compared to a single semantics.

An aim of our further work will be the expansion of our database by further dynamic biometric modalities. This will allow to examining the fusion of different semantics for other biometric methods. Such a fusion would be conceivable at the speech recognition and the keystroke dynamics, for example. Further we will examine where the boundaries of the improvement of the multi-semantic concept lie for more than two semantics. On the other side the impact of fusion also shall be analyzed for two or more instances of the same semantics. Different examinations are further necessary. Measuring and comparisons have to be carried out regarding the duration of a single authentication process and the time requirement for the creation of reference and verification data. The acceptance of the users for taking more than one handwritten semantics for an authentication process should also be studied, for example, by subjective tests.

Acknowledgements

The work described in this paper has been supported in part by the European Commission through the IST Programme under Contract IST-2002-507634 BIOSECURE. The work on biometric hashes for verification purposes is partly supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation). The database acquisition used in this publication has been produced with the assistance of the European Union (project Culture-Tech). The content of this publication is the sole responsibility of the University Magdeburg and their co-authors and can in no way be taken to reflect the views of the European Union.

References

[1] T. Scheidat, C. Vielhauer, J. Dittmann, Distance-level fusion strategies for online signature verification, in: Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), Amsterdam, The Netherlands, 2005.

[2] A. Ross, A.K. Jain, Multimodal biometrics: an overview, in: Proceedings of the 12th European Signal Processing Conference (EUSIPCO), Vienna, Austria, 2004, pp. 1221–1224.

[3] K. Chang, K.W. Bowyer, P.J. Flynn, Face recognition using 2D and 3D facial data, in: Proceedings of Workshop on Multimodal User Authentication (MMUA), 2003.

[4] A. Kumar, D.C.M. Wong, H. Shen, A.K. Jain, Personal verification using palmprint and hand geometry biometric, in Proceedings of International Conference on Audio- and Video-based Person Authentication, 2003, pp. 668–675.

[5] J. Czyz, J. Kittler, L. Vandendorpe, Combining face verification experts, in: Proceedings of ICPR, 2002, pp. 28–31.

[6] B. Ly Van, S. Garcia-Salicetti, B. Dorizzi, Fusion of HMM's likelihood and Viterbi path for on-line signature verification, in: Proceedings of BioAW 2004, Lecture Notes in Computer Science, LNCS 3087, 2004, pp. 318–331.

[7] S. Garcia-Salicetti, J. Fierrez-Aguilar, F. Alonso-Fernandez, C. Vielhauer, R. Guest, L. Allano, T. Doan Trung, T. Scheidat, B. Ly Van, J. Dittmann, B. Dorizzi, J. Ortega-Garcia, J. Gonzalez-Rodriguez, M. Bacile di Castiglione, M. Fairhurst, Biosecure reference systems for on-line signature verification: a study of complementarity, in: Annals of Telecommunications, Special Issue on Multimodal Biometrics, vol. 62, 1–2, 2007, pp. 36–61.

[8] A.K. Jain, S. Prabhakar, A. Ross, Fingerprint matching: data acquisition and performance evaluation, MSU Technical Report TR99-14, 1999.

[9] B. Yanikoglu, A. Kholmatov, Combining multiple biometrics to protect privacy, in: Proceedings of ICPR- BCTP Workshop, 2004.

[10] B. Ly-Van, R. Blouet, S. Renouard, S. Garcia-Salicetti, B. Dorizzi, G. Chollet, Signature with text-dependent and text-independent speech for robust identity verification, in: Proceedings of IEEE Workshop on Multimodal User Authentication, 2003, pp. 13–18.

[11] A.K. Jain, A. Ross, Multibiometric systems, Communications of The ACM 47 (1) (2004) 34–40.

[12] C. Vielhauer, S. Schimke, A. Valsamakis, Y. Stylianou, Fusion strategies for speech and handwriting modalities in HCI, Proceedings of SPIE-IS&T Electronic Imaging, vol. 5684, 2005, pp. 63–71.

[13] C. Vielhauer, T. Scheidat, Multimodal biometrics for voice and handwriting, in: J. Dittmann, S. Katzenbeisser, A. Uhl (Eds.), Communications and Multimedia Security: 9th IFIP TC-6 TC-11 International Conference, CMS 2005, Proceedings, LNCS 3677, 2005, pp. 191–199.

[14] C. Schmidt, On-line Unterschriftenanalyse zur Benutzerverifikation, Shaker Verlag, Aachen/Germany, 1999 (in German).

[15] Y. Kato, T. Hamanoto, S. Hangai, A proposal of writer verification of hand written objects, in: Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), Lausanne, Switzerland, 2002, pp. 585–588.

[16] C. Vielhauer, Biometric User Authentication for IT Security: From Fundamentals to Handwriting, Springer, New York, 2006.

[17] T. Scheidat, F. Wolf, C. Vielhauer, Analyzing handwriting biometrics in metadata context, in: Edward J. Delp III, Ping Wah Wong(Eds.), Security, Steganography, and Watermarking of Multimedia Contents VIII, Proceedings of SPIE-IS&T Electronic Imaging, SPIE, vol. 6072, 2006.

[18] C. Vielhauer, R. Steinmetz, A. Mayerhöfer, Biometric Hash based on statistical features of online signature, in: Proceedings of the International Conference on Pattern Recognition (ICPR), Conference on Pattern Recognition (ICPR), Quebec City, Canada, vol. 1, 2002, pp. 123–126.

[19] The Culture Tech Project, Cultural dimensions in digital multimedia security technology, a project funded under the EU-India Economic Cross Cultural Program, Available from: <http://amsl-smb.cs.uni-magdeburg.de/culturetech/>, requested February 2006.

[20] BioSecure, Available from: <http://www.biosecure.info/>, requested April 2006.