# SINGLE-SEMANTIC MULTI-INSTANCE FUSION OF HANDWRITING BASED BIOMETRIC AUTHENTICATION SYSTEMS

*Tobias Scheidat, Claus Vielhauer, Jana Dittmann*

Otto-von-Guericke University Magdeburg, D-39106 Magdeburg, Germany
{tobias.scheidat, claus.vielhauer, jana.dittmann}@iti.cs.uni-magdeburg.de

## ABSTRACT

The fusion of biometric systems, algorithms and/or traits is a well known solution to improve authentication performance of biometric systems. In this article the fusion of two instances of the same semantic is suggested, where semantics are alternative handwritten contents such as numbers or sentences, in addition to commonly used signature. In order to fuse two instances of one semantic, a biometric authentication is carried out on both by Biometric Hash algorithm up to matching score computation. The fusion is done by combination of matching scores to a joint score as basis for authentication decision. Three individual fusion strategies are used to study to which degree the authentication performance can be improved or degraded. Therefore one pragmatic and two optimistically weighting approaches for biometric fusion are used. The best fusion result is even better than the corresponding best individual result by approximately 17%.

***Index Terms***— Biometrics, fusion, handwriting, semantics

## 1. INTRODUCTION

The need for automatic authentication of information and persons is a requirement in our today's society. In order to ensure the authenticity of subjects there are three methods: secret knowledge, personal possession and biometrics. While the approaches of secret knowledge are based on information reserved to the owner, the methods of personal possession are based on the principal if items remaining in possession of the authorized person. Problems of both methods include the potential for theft, lost or hand over to unauthorized persons of the authentication object (secret information or personal possession). Biometric authentication is based on a physical (e.g. iris) or behavioral (e.g. handwriting) trait of a user. The authentication object of biometric approaches is directly linked to the body or to the behavior of its bearer, and theft, lost and hand over to unauthorized persons is not possible in an easy way.

However, one general problem of biometric systems is the natural variability of the biometric trait: Contrary to knowledge and possession, where presented data have to be identical with stored data, the individual biometric samples are not identical for each measurement. This intra-class variability is caused by several reasons such as natural biometric variability, changing sensors, different environments or aging of the bearer of biometric information. Inter-class similarity is achieved by high degree of identicalness of the same biometric trait and/or its feature representation between different persons. Intra-class variability and inter-class similarity may lead to false classifications regarding authentication attempts at the one hand: *False positives* are the authentication results of persons which are wrongly recognized by a biometric system as another person. On the other hand authentication results of persons, which are rejected from the system although they are authorized are denoted as *false negatives*. Both error classes are combined into an equal error rate (EER, see section 3.2).

Possible solutions to compensate for the false classification problem due to intra-class variability and inter-class similarity can be found in the fusion of biometric systems or experts. Ross and Jain present a classification for biometric fusion based on number of sensors, algorithms, systems and traits involved in fusion process [1]: single biometric trait - multiple sensors, single biometric trait - multiple classifiers, single biometric trait - multiple units and multiple biometric traits.

Based on a test database of 160 persons, Jain et al. show that by the combination of prints of two fingers or two versions of one finger, improvements are possible [2]. A so-called multi-semantic approach based on handwriting biometrics is proposed by Scheidat et al., where biometric fusion approach uses a pair wise combination of four different handwritten contents (e.g. signature, symbol). The highest relative improvement achieved amounts approximately *55%* [3].

One advantage of handwriting biometrics is the usage of different content of writing due to its behavioral nature. It has been shown that alternative handwritten contents can be used also for handwriting authentication as signature [4]. These alternatives are called *semantics*. Further studies show that the concept of semantics is also applicable for speech based authentication [5]. A second advantage of handwriting modality is the seamless integration of biometric recognition in pen-based Human-to-Computer

Interaction (HCI) applications, which may be found increasing frequently with the use of pen-based computers such as Personal Digital Assistants (PDA) and Tablet-PCs.

The aim of this article is to study another strategy of multi-biometric fusion by combining two instances of the same handwritten semantic in order to improve the authentication performance. Therefore three different semantics were captured and fused by a biometric system based on three different fusion strategies. The fusion is carried out by combination of results of two identical algorithms based on different semantic instances each.

This article is structured as follows: Section 2 gives a general overview, section 3 describes the evaluation methodology, database and biometric error rates, section 4 presents test results and section 5 closes with conclusions and future work.

## 2. BIOMETRIC FUSION STRATEGIES

In general a biometric system works in two modes: enrollment and authentication. During the enrollment process a user will be registered within the system by presenting his/her physiological or behavioral biometric trait to acquisition module as shown in figure 1. The feature extraction process step determines a feature vector describing the biometric characteristic within the system. As last step of the registration the feature vector is linked with identity of the person and stored in the system's database as reference data. Verification and identification are the two modes an authentication can be carried out: While during verification the system checks the claimed identity of a person, at identification the system determines the identity of a person, if he/she is enrolled in the system. Firstly, the biometric trait is acquired and the features are extracted. Secondly, the matching module compares the feature vectors of currently presented data and reference data of claimed identity (verification) or all registered identities (identification) and calculates a matching score describing similarity/dissimilarity. Finally, this score value is basis for final authentication decision.

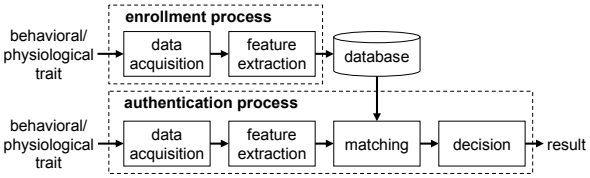


**Figure 1.** General scheme of a biometric process

### 2.1. Level of fusion

Based on the structure of a biometric authentication system as shown in figure 1 there are three main points within the process to fuse biometric systems [1]: on feature extraction level, on matching score level or on decision level. At *feature extraction level* the feature vectors of the systems involved are combined with each other to built a joint feature vector as basis for score determination. The individual matching scores are fused at the *matching score level* to one combined value for decision. At fusion on decision level each subsystem determines its own authentication decision and all individual results are combined to a common decision of the fusion system.

The fusion studied in this article is carried out on matching score level. One important advantage of this kind of fusion is one single scalar value each subsystem determines which can be weighted regarding the authentication performance of causing system.

### 2.2. Fusion weighting parameter estimation strategies

Based on previous work the fusion is carried out on matching score level [6], whereby the fusion matching score $m_{fus}$ is determined by the weighted sum of the individual matching scores of the $n$ subsystems, $m_1,\ldots,m_n$, whereby $w_1,\ldots,w_n$ denote the weights for each of the experts, normalized such that:

$$w_1 + w_2 + \ldots + w_n = 1 \qquad (1)$$

The selection of three approaches as presented in the coming subsections is chosen from the huge number of possibilities for our experiments in order to represent the classes of equal, linear and super linear weighting estimation strategies. In our test we want to investigate if the usage of different weighting approaches, as suggested in [6], might have effects to single-semantic multi-instance fusion, for example caused by training effects of the user (like sequential writing of the semantics several times, e.g. *10* times).

#### *2.2.1. Equal weighted fusion*

In the equal weighted fusion, all weights used are equal, independent of the single methods' performance:

$$w_1 = w_2 = \ldots = w_n = n^{-1} \qquad (2)$$

The equal weighting strategy uses no a-priori knowledge about the subsystems involved and could be integrated in any existing system without closed-user group optimizing effort. Since our system consists of two discrete verification processes up to fusion step, this leads to weights of $w_1=w_2=0.5$.

$$w_{linear\,i} = \frac{eer_i}{\sum_{m=1}^{n} eer_m} \qquad (3) \qquad w_i = \frac{w_{linear\,i}^{\;2}}{\sum_{m=1}^{n} w_{linear\,m}^{\;2}} \qquad (4)$$

#### *2.2.2. Linear weighted fusion*

In linear weighted fusion, the weights are determined from the authentication performance of the single methods displayed utilizing the equal error rate (EER), which has been determined on a closed-group scope, based on an a-priori experimental evaluation of three different semantics. The weights of matching scores are determined by dividing the single EER with sum of all EER (see equation 3). A property of this weighting scheme is that the matching

scores of the system, which received the highest EER are multiplied with the smallest weight and vice versa.

*2.2.3. Quadratic weighted fusion*

As for the linear weighted fusion, weights are related to a-priori studies of the EER performance of each subsystem. In order to reward the method having the best individual performance in a super linear manner, the weights of linear weighting strategy are squared and used for the quadratic weighted fusion. Then the weights are again normalized to have a summed up value of *1* (see equation 4).

In our study, for both, the linear and quadratic weighting schemes, we have applied the same experimental data for the a-priori study of individual EER characteristics and the actual evaluation of the fusion system. We have chosen this optimistic approach to study a theoretical upper bound with respect to recognition performance in comparison to the equally weighted fusion, which represents a scenario with no a-priori knowledge at all (thus a lower bound of the achievable improvements).

### 2.3. Single-semantic multi-instance fusion

Based on the fusion strategies described above one fusion scenario is created. It is a single-semantic multi-instance system, which uses the same algorithm for authentication of both instances of the same semantic (see figure 2). The underlying authentication algorithm is the Biometric Hash method presented by Vielhauer et al. in [4] and [7].

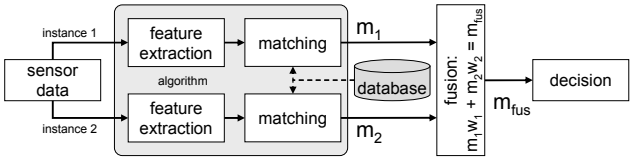


**Figure 2.** Multi-instance uni-algorithmic system

## 3. METHODOLOGY

In this section the structure of the evaluation data and methodology are described. In addition a short explanation of the equal error rate used as authentication performance measurement is given.

### 3.1 Evaluation Setup

Contrary to other biometric modalities the handwriting provides possibilities to change the content of authentication object. Therefore we have decided to study different semantics: The database utilized for the evaluation of the systems described in section 2 contains handwriting data of the semantics *Signature* (*62* users), *PIN* (*63* users) and *Sentence* (*45* users). Although the numbers of donors are not representative, our goal is to present a general methodology to study fusion techniques for uni-semantic fusion with respect to increase or decrease in false recognition rates.

### 3.2. Biometric error rates

The authentication performance of a biometric system has to be determined empirically, due to the fact that it cannot be measured directly. Therefore biometric error rates are used: The false non match rate (FNMR) calculates the ratio between the number of rejected authorized persons and the entire number of authentication attempts. The false match rate (FMR) describes the ratio between accepted non-authorized users and the entire number of authentication attempts. A common measurement in biometrics is the equal error rate (EER) where FNMR and FMR yield the same value. It can be used as normalized reference point for comparison in terms of one scalar value of biometric algorithms.

### 3.2 Evaluation Methodology

Our evaluation protocol is based on collections of *10* sequentially acquired handwriting samples ($S=S_1,...,S_{10}$) for each user in each semantic class. From these samples $S$, we construct test sets for building references, weighting parameters and fusion based verifications as follows:

*Reference set:* From $S$ we take the first *4* samples ($S_1,...,S_4$) to generate *4* references in a leave-one-out strategy. This means a combination of *4* choose *3*, i.e. *4* different references ($R=R_1,...,R_4$) are created, containing *3* handwriting samples each. These references are used for both: the estimation of weighting parameters and the determination of fusion verification performance.

*Estimation of weighting parameters:* The individual weighting parameters for the two instances are determined using $S_7$ for the first and $S_8$ for the second one. In order to determine the weights, $EER_1$ and $EER_2$ are calculated based on comparison of the *4* references $R_1,...,R_4$ and $S_7$ or $S_8$ respectively. The weights $w_1$ and $w_2$ are calculated based on $EER_1$ and $EER_2$ as described in section 2.2.

*Determination of fusion verification performance:* In order to measure the FNMR each $R_1,...,R_4$ is compared to $S_5$ and $S_6$ for the determination of the first matching score $m_1$. For $m_2$ the FNMR is calculated based on comparison of each $R_1,...,R_4$ and $S_9$ and $S_{10}$. The FMR is determined based on the comparison of each $R_1,...,R_4$ of a user with samples of all other users in the same semantic class, $S_5$ and $S_6$ or $S_9$ and $S_{10}$ respectively. Note that $S_5$ and $S_6$ have been originally acquired after $S_8$ and $S_9$ and we assume that a possible training effect may lead to a higher quality for $S_9$ and $S_{10}$.

## 4. TEST RESULTS

This section describes the results of the tests based on methodology presented in section 3. Table 1 shows EERs determined on the single semantics and on their fusion based on equal, linear and quadratic weighted fusion. For fusion the cells also show the weights ($w_1$, $w_2$). A general

observation is that the *Signature*-based results are better than results of *Sentence* and *PIN* for both, single tests and fusion tests. One reason for worse result of *PIN* can be the fact, that all users write the same combination of digits (77993) out of a small set of characters (0-9). Thus the inter-class similarity is higher than for individual semantics such as *Signature* or for given semantics with a higher number of characters out of a larger set of characters such as *Sentence*.

**Table 1.** EER of single semantics and their fusion

| | Single | | Fusion | | |
|---|---|---|---|---|---|
| | | | equal | linear | quadr. |
| Semantic | $EER_{HW1}$ | $EER_{HW2}$ | $w_1,w_2,EER$ | $w_1,w_2,EER$ | $w_1,w_2,EER$ |
| **Signature** | 0.0500 | 0.0831 | 0.500 \| 0.500 | 0.602 \| 0.398 | 0.697 \| 0.303 |
| | | | 0.0442 | **0.0432** | 0.0437 |
| **PIN** | 0.0832 | 0.0859 | 0.500 \| 0.500 | 0.510 \| 0.490 | 0.520 \| 0.480 |
| | | | **0.0687** | 0.0690 | 0.0690 |
| **Sentence** | 0.0559 | 0.0874 | 0.500 \| 0.500 | 0.534 \| 0.466 | 0.586 \| 0.414 |
| | | | 0.0528 | 0.0498 | **0.0494** |

The single results and the derived weights show no significant performance difference for single verifications and training effects might no have occurred in our test setup (of course this might be caused by the limited size of samples considered).

Table 1 shows that by the different fusions of the single semantics improvements can be reached in each case. The best overall result was reached by the fusion of two *Signature* instances with an EER of *0.0432*. Although in the worst case the fusion system calculates an EER of *0.0690* for the linear and quadratic strategies of the *PIN* instances, these strategies improve the best single result ($EER_{HW1}$=*0.0832*) of this semantic class.

The results in Table 1 show also, that it is possible to improve the verification performance by the fusion of two instances of the same semantic. The relative improvement reaches from *11.6%* up to *17.4%* for the verification.

## 5. CONCLUTIONS AND FUTURE WORK

A biometric single-semantic multi-instance fusion based on handwriting is suggested. The fusion is carried out by the combination of the matching scores of two instances of one handwritten semantic. The underlying authentication algorithm is the same for both instances, the Biometric Hash method.

In our evaluation improvements can be observed for all three semantics and fusion strategies in comparison to the best individual results. The signature based multi-instance scenario reaches the highest verification performance with an EER of *0.0432* for linear weighted fusion strategy. Although the equal weighted fusion strategy is not based on a-priori knowledge of the underlying subsystems, it determines similar good results as the other strategies, linear and quadratic. These two strategies use information of the individual verification performance of the single subsystems based on underlying database as a-priori knowledge to estimate fusion weights. Remarkable is the fact that an additional instance of the authentication object leads to a relative decrease of the EER by *13.6%* for *Signature*, *17.4%* for *PIN* and *11.6%* for *Sentence*.

For future work it appears necessary to create and evaluate fusion strategies without a-priori knowledge in order to optimize multi-biometric systems for upcoming unknown reference and/or authentication data. This could be for example a scenario based on two disjunctive databases, one for weighting parameter estimation and one for evaluation purposes. Also, a study of other single-semantic multi-instance biometric approaches should be carried out to proof the concept presented in this article. This could be behavioral biometrics such speech or keystroke dynamics as well as physiological biometric such iris or fingerprint.

## 6. 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 (biometric fusion), SIMILAR (Project Number: FP6-507609, pen-based modalities) and the European Union (project CultureTech, data acquisition). 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.

## 7. REFERENCES

[1] A. Ross, A.K. Jain, "Multimodal Biometrics: An Overview", Proc. Of the 12th European Signal Processing Conference (EUSIPCO), Vienna, Austria, pp. 1221 - 1224, 2004.

[2] A.K. Jain, S. Prabhakar, A. Ross. "Fingerprint Matching: Data Acquisition and Performance Evaluation", MSU Technical Report TR99-14, 1999.

[3] T. Scheidat, C. Vielhauer, J. Dittmann, "Handwriting Verification - Comparison of a multi-algorithmic and a multi-semantic Approach", To appear in Image and Vision Computing, Multimodal Biometrics Special Issue, 2007.

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

[5] C. Vielhauer; T. Scheidat; A. Lang; M. Schott; J. Dittmann; T.K. Basu; P.K. Dutta, "Multimodal Speaker Authentication - Evaluation of Recognition Performance of Watermarked References"; In: Proceedings of the 2nd Workshop on Multimodal User Authentication (MMUA), Toulouse, France, 2006.

[6] 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 2005 (ICME), Amsterdam, The Netherlands, 2005.

[7] C. Vielhauer, R. Steinmetz, A. Mayerhöfer, "Biometric Hash based on Statistical Features of Online Signatures", In: Proceedings of the IEEE International Conference on Pattern Recognition (ICPR), Quebec City, Canada, Vol. 1, pp. 123 – 126, 2002.