

# Multimodal Biometrics for Voice and Handwriting

Claus Vielhauer and Tobias Scheidat

School of Computer Science, Department of Technical and Business Information Systems,  
Advanced Multimedia and Security Lab, Otto-von-Guericke University Magdeburg,  
Universitätsplatz 2, D-39106 Magdeburg, Germany  
{claus.vielhauer, tobias.scheidat}@iti.cs.uni-magdeburg.de  
[http://www.iti.cs.uni-magdeburg.de/iti\\_ams1/](http://www.iti.cs.uni-magdeburg.de/iti_ams1/)

**Abstract.** In this paper a novel fusion approach for combining voice and online signature verification will be introduced. While the matching algorithm for the speaker identification modality is based on a single Gaussian Mixture Model (GMM) algorithm, the signature verification strategy is based on four different distance measurement functions, combined by multialgorithmic fusion. Together with a feature extraction method presented in our earlier work, the Biometric Hash algorithm, they result in four verification experts for the handwriting subsystem. The fusion results of our new subsystem on the multimodal level are elaborated by enhancements to a system, which was previously introduced by us for biometric authentication in HCI scenarios. Tests have been performed on identical data sets for the original and the enhanced system and the first results presented in this paper show that an increase of recognition accuracy can be achieved by our new multialgorithmic approach for the handwriting modality.

**Keywords:** biometrics, combination, distance, fusion, handwriting, identification, matching score level, multialgorithmic, multimodal, voice.

## 1 Introduction

The necessity for user authentication rose strongly in the last years. In the today's digital world it is no longer possible for humans to determine the identity of the other one mutually from face to face, for example due to the distance between two parties, which may be virtually linked by a computer network (e.g. the World-Wide Web). The task of ensuring the identity of participants of a process is made increasingly often by automatic systems, e.g. by user verification. Verification is the confirmation of the identity of a person. The three fundamental methods of user verification are based on secret knowledge, personal possession and biometrics. An important advantage of biometrics is that it identifies the person neither by knowledge, nor by an object, which can be lost or handed over to other persons. In contrast to knowledge and possession, biometric characteristics are intrinsically conjoined to their owners. Prominent modalities for biometrics are passive traits like iris and fingerprint on one side and behavioral properties such as voice and handwriting on the other. Voice and handwriting, especially signature, are very intuitive behavioral and ubiquitous

biometrics. They can be captured by modern personal computers, as well as by Personal Digital Assistants (PDA) or even some pen-enabled smart phones. Further, no expensive special hardware is required. Only a microphone and a graphical tablet or touch sensitive display are necessary.

One of the problems with biometrics is the lack of recognition accuracy of single systems, reflected for example by the error characteristics of false identifications. In order to increase the accuracy of biometric systems, some approaches try to reach a better performance by combination of various biometric modalities (e.g. fingerprint and iris). These approaches are called multimodal or multibiometric verification systems. These multimodal biometrics can be advantageous also for persons, who cannot exhibit one or several of the required characteristics. For example, a missing modality could be ignored and those characteristics available could be increasingly significant for subjects lacking one or more features.

Given a biometric system of only one single modality, another possibility to improve the verification performance is the fusion of different algorithms of this individual biometrics. Systems of this category are denoted as multialgorithmic systems. In this paper we analyze the effects of replacement of a subsystem of an existing multimodal system to the recognition accuracy. While the original subsystem is based on a single distance measurement algorithm for handwriting, the new subsystem is a multialgorithmic signature verification expert. For our analysis, we first give an overview of the original multimodal system and the underlying fusion strategy. We then summarize the multialgorithmic approach for signature verification. This multialgorithmic approach is then used as a replacement for the subsystem for handwriting, which is introduced by our novel fusion model. In our experimental evaluation we then compare the multimodal recognition results of the original and the new subsystem.

This paper is structured as follows. In section 2, we give a short description of the original multimodal system. Section 3 provides an overview of the fusion methods for combining multimodal or multialgorithmic biometric systems. Further, it describes our new approach based on the fusion of four signature verification experts. We present first experimental results of this new subsystem alone, and at the end of the third section for the entire multimodal system with and without the new subsystem. In section 4 we summarize this article, draw some conclusions for our research and discuss further activities in this area.

## **2 Multimodal Fusion on the Example of Voice and Handwriting**

In the multimodal approach presented in [2], the biometrics speech and signature are fused with one another. The focus here is on the use of pen-based mobile devices for Human to Computer Interaction (HCI), where the authors concentrate on spoken and hand-written input. The problem with the use of speech is the influence of the results by noises and consequently, the fusion idea is to compensate this influence by a complimentary biometric modality, the handwriting.

Figure 1 outlines the multimodal biometric system model from [2]. The Fusion is accomplished on the matching score level (see chapter 3). In this perspective, the multimodal biometric system contains two separate biometric subsystems until the

fusion process. Both subsystems have their own modality dependent input data. As from this point of view it is irrelevant whether the subsystems consist of one or more algorithms, the usage of multialgorithmic schemes for one or more of the subsystems generally fits in this multimodal layout.

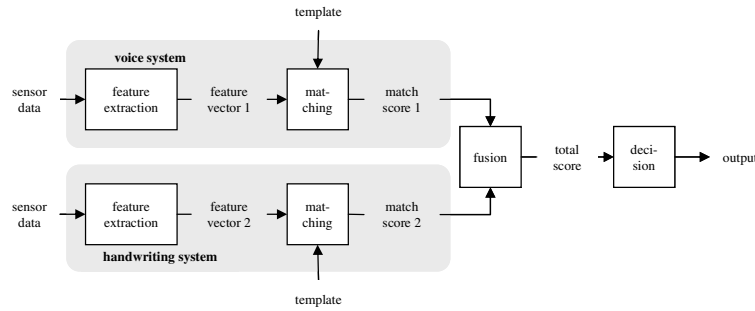


Fig. 1. Multimodal fusion for voice and handwriting on matching score level

For the actual design of the fusion strategy, a variety of alternatives exist, based on linear and non-linear weighting, user-specific and global weighting models et cetera. For the sake of simplicity, in [2] a global, linear weighting based on the modified z-score distance measures of each modality has been implemented. Here, separately for each modality the non-normalized distance measure  $x$  is normalized to the modified z-score  $z$  according to the following equation:

$$z = \frac{x - \min(x)}{\sigma} . \tag{1}$$

$\sigma$  and  $\min(x)$  denote the standard deviation and the minimum above all observed distance measures in a test. The fusion of z-scores is then simply given by the summation of the individual scores for the two modalities,  $z_{HW}$  for handwriting and  $z_{SR}$  for speaker recognition:

$$z_{final} = z_{SR} + z_{HW} . \tag{2}$$

The reference data and test data descended from ten persons for both, speech and handwriting. The spoken inputs are German. They were captured in a soundproofed environment. Later two kinds of noises, generated white Gaussian noise and recorded laptop fan noise, were added in order to simulate a mobile setup. Each person had to read 15 sentences for training and one different sentence for testing and for the signature part, each person had to write her or his signature six to eleven times. One of these samples was used as test sample. The samples remained were used for the reference data set. The handwriting data were acquired on a graphical tablet, Wacom Cintiq15, which output the same kind of signals as those digitizers used in tablet PCs.

A disadvantage of this original system is the missing weighting. These could use advantages of a person during speaking or writing by higher weight.

### 3 Fusion of Handwriting Algorithms

In order to increase the verification performance of the multimodal system described in section 2, we propose to use a multialgorithmic combination of handwriting verification methods in the signature part in order to achieve overall recognition improvements on the multimodal level.

Fusion strategies for biometrics have raised increasing interest by science in the recent past and a diversity of publications on this subject can be found. From the variety we want to briefly give reference to some examples, which appear particularly interesting in context of our work, without neglecting other contributions.

As described by Jain and Ross in [3] a multibiometric system is generally based on one of three fusion levels: feature extraction level, matching score level or decision level. In the feature extraction level all systems involved separately extract the information from the different sensors. The determined data are stored in separate feature vectors. During the fusion process, the feature vectors are combined to a joint feature vector, which is used for the matching score process. Dependent on the number of subsystems involved and the dimension of each individual feature vector, the resulting joint feature vector may be high dimensional, which can make its further processing cumbersome. The fusion on matching score level is based on the mixture of matching scores after the comparison between reference data and test data. Additionally, a normalization and weighting of the matching scores of the different modalities is possible, for example by relevance. The fusion results in a new matching score, which is the basis for decision. In decision based systems, each biometric subsystem involved is completed separately. Here, the individual decisions are combined to a final decision, e.g. by boolean operations like AND/OR. Because this fusion is accomplished at the latest point in time of the overall process, it cannot be controlled and parameterized as granularly as the other two approaches.

Matching score level based approaches have been successfully applied for a number of multimodal systems, for example in [3], a multibiometric system is presented by Jain and Ross, that uses face, fingerprint and hand geometry characteristics of a person for authentication. This system applies an user adapted weighting tactic. Ly-Van et al. [4] combine signature verification (based on HMM's) with text dependent (based on DTW) and text independent (based on Gaussian Mixture Model) speech verification, at a time. They report that fusion increases the performance by a factor 2 relatively to the best single system. Czyz et al. ([5]) propose combination strategies of face verification algorithms. The authors show that the combination based on simple sum rule can reach a better result than the best individual expert.

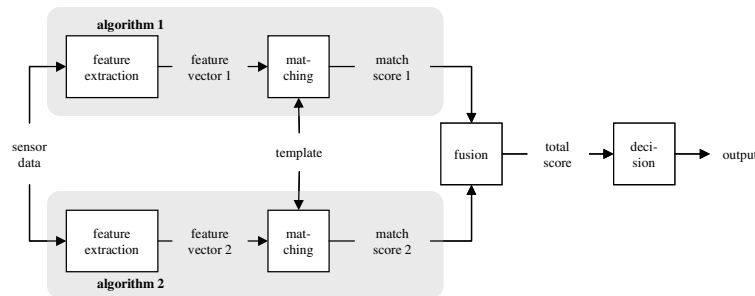
Because of the good characteristics, like simple normalization and weighting, and the encouraging results subscribed in [3] and [5] we decided for a fusion on matching score level in our multialgorithmic system for the handwriting modality.

#### 3.1 New Approach

Our goal is to improve the verification performance of the multimodal system described in section 2 by use of multialgorithmic handwriting verification algorithm. If the handwriting modality itself reaches a better identification rate, the performance

of the entire multimodal system should become better, which we want to demonstrate experimentally.

In [6] we have shown that in principle, multialgorithmic fusion can be achieved by multimodal fusion methods and because of the very encouraging test results in our work on handwriting, we choose the matching score level for combining the individual handwriting algorithms. Another argument for the matching score is that normalization and weighting can be accomplished here relatively simple. Each algorithm (expert) produces a distance value, which expresses the similarity of reference data and test data. Normalization then makes the values of the different experts comparable to each other. In the last step before the decision process, weighting is applied to each matching score, where the definition of the weight parameters is part of the system configuration. Such a multialgorithmic fusion on matching score level is shown in figure 2. In difference to the multimodal fusion, the procedures involved use the same sensor data and reference data.



**Fig. 2.** Multialgorithmic matching score level fusion

At this time all algorithms in our system use the Biometric Hash method introduced in [1] for feature extraction from the handwriting samples. Only the similarity of input data and reference data is determined by different distance measurement functions. The used alternative distance measures are Canberra Distance, City Block (Manhattan) Distance, Euclidean Distance and Hamming Distance. Note that with a small distance, the feature vectors are each other more similar than with a larger.

In [7] we have shown that a well weighted fusion of different distance measure algorithms can result in a better verification performance than the best individual algorithm. These results were experimentally determined from a database of 1761 genuine enrollments (with 4 signatures per enrollment), 1101 genuine verification signatures and 431 well skilled forgeries by 22 persons. All samples have been captured on the same device, a Wacom Cintiq15. We have chosen this tablet since it has an active display. Through this not only the quality of the enrollments and verifications improves but also the quality of the forgeries becomes better. The reason for it is that the written text appears in the place in which it is produced. This corresponds to the natural writing behavior of human beings. We were able to show that the best fusion strategy of signatures results in a decrease of the EER of 12.1% in

comparison to the best individual algorithm. Additional investigations with other trays and other semantic classes, as published in [6], led to similar results.

### 3.2 Multialgorithmic Fusion: Experimental Results

In our new approach we now want to use these improved concepts for the combination of voice and signature. In the first step we created a fusion of four distance measure algorithms within a biometric system as described above. In the second we combined the voice system and the handwriting system by the matching score level.

In order to show the increase in identification performance of the system described in [2], our tests based on the same signature test sets. Due to the functional properties of our evaluation program we used four out of five to ten signature samples for enrollments and the remaining one to six samples are used as test data for each user. In case that more than one test data are available for one enrollment per user, our system selects those enrollments having the smallest distance value. In our earlier work [7] in single tests of the four distance functions we have created five weighting strategies, based on the respective value of the individual distances for the test set. The weighting strategy, which led to the best results, was adapted also on the handwriting data from [2]. By using the described transformation function we determined a modified z-score. The identification rate for the signature amounts to 80% and the rate amounted to 50% before.

**Table 1.** Modified z-scores of the multialgorithmic method

	1	2	3	4	5	6	7	8	9	10
1	0,6	3,7	2,5	2,4	3,0	2,1	1,7	0,0	3,0	3,7
2	1,3	1,0	2,3	1,9	1,7	1,3	1,5	0,7	2,2	3,1
3	0,6	2,5	0,3	1,1	1,4	0,9	1,6	0,2	2,2	3,2
4	0,5	3,2	2,0	0,5	1,9	0,5	0,9	0,3	1,5	3,2
5	1,9	2,5	1,3	1,7	1,1	1,4	1,9	1,7	2,6	2,9
6	1,6	2,9	2,6	1,8	2,0	0,0	0,9	0,5	0,8	3,4
7	2,6	2,9	1,6	2,2	2,3	1,7	0,4	1,9	3,3	3,5
8	1,4	4,0	2,6	2,3	2,8	1,9	1,8	0,0	1,5	3,5
9	1,1	3,2	2,3	1,9	2,2	1,7	1,3	0,4	0,4	3,4
10	1,3	3,5	2,9	1,3	1,6	0,8	1,0	0,7	1,3	0,7

Table 1 shows the results of the identification tests. For each user an identification attempt was accomplished. The similarity of the test data of a person was determined in each case to their reference data and the reference data of all other persons. The matching scores of the individual algorithms were normalized, if necessary, to the interval [0-68]. The number results from the number of 69 of statistical features extracted by the Biometric Hash algorithm as suggested in [8]. In the next step, we determined the modified z-score as described in section 2. Identification is then performed by the nearest neighbor strategy. In case there are more than one matches for an assignment of a reference data to a test data, we consider an identification

failure. With our new subsystem, we reached an identification rate of 80%. Table 1 shows the observed z-scores after the fusion. Columns show the enrollments for each user and rows show the verification data for each user, consequently, the marked diagonal shows the genuine z-scores of each user. The first column and the first row represent the user IDs.

We are confident that with a larger number of persons and/or test data and optimized parameterization of the fusion weights, identification rates could be improved, but we assume that with these first test results, we may conceptually prove the qualification of our approach for usage in multimodal systems.

### 3.3 Multimodal Fusion: Experimental Results

The next step is to bring together the speech-based subsystem and the multialgorithmic signature-based subsystem to form the multimodal system. To ensure that the results are comparable, we have selected the same proceeding for the multimodal fusion of the speech and the handwriting subsystems, which was used also with the original system, as described in [2]. This fusion consists of a non-weighted addition of the z-scores. In addition, we assume an uniform distribution of z-scores.

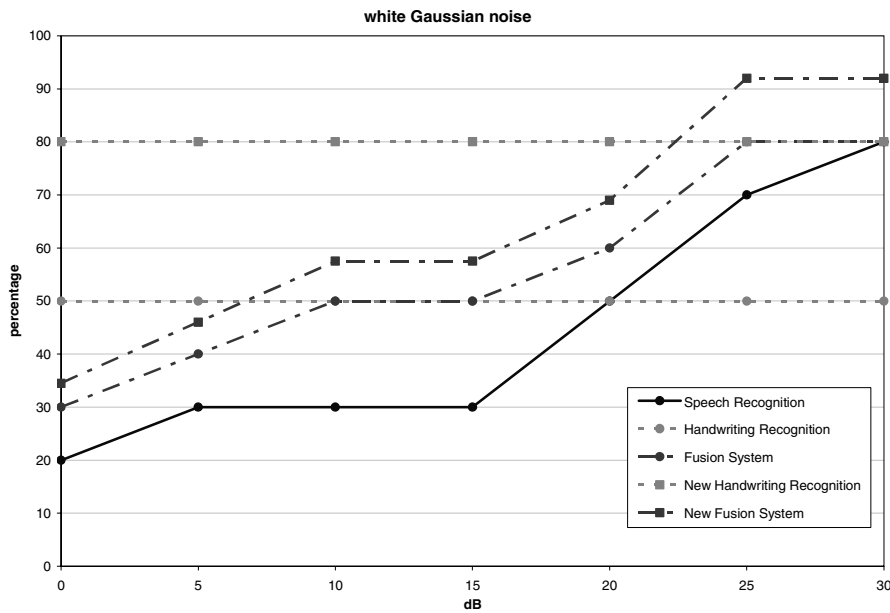


Fig. 3. Adapted identification rates and original identification rates as function of noise

As an overall improvement of 30% for the identification rate for the subsystem has been observed (originally 50%, now 80%), we can estimate the effect of accuracy of the multialgorithmic subsystem to 15% on the entire multimodal system. The

improvement arises from the participation of both subsystems with equal weights. In figure 3 the improvement of the modified system is graphically represented in the comparison to the original system. In comparison to the original fusion, our improved approach has shown for example an identification rate improvement from 30% to 34.5% at a zero Signal-to-Noise-Ratio (SNR) and from 80% to 92% for an SNR of 25dB. Similarly, for all intermediate SNR values, we observe a significant improvement of the recognition rate.

## 4 Conclusions and Future Work

To best of our knowledge, so far no research results on the combination of multimodal and multialgorithmic experts concerning speech and handwriting has been published and in this work we implemented and evaluated this approach by enhancing an existing system [2]. The original system consisted of one speech-based subsystem and one signature-based subsystem and in our modification the signature-based subsystem was replaced by a multialgorithmic subsystem. Our examinations, based on identical data set as in the original publication, show that an improvement of the verification performance of the originally multimodal approach is feasible. By exploitation of a multialgorithmic signature verification system, an increase in recognition accuracy of 15% could be observed for the whole system.

Since the weight parameters used for our multialgorithmic subsystems have been estimated based on entirely different data sets, we can truly state that parameters and test results are uncorrelated. On the other side, this implies that further improvement can be achieved by optimization towards the actual test set.

Although we can derive some initial conclusions on our new concept, it needs to be stated that the size of the used test sets is not statistically representative. Therefore one of our next aims will be the further collection of data of both, voice and handwriting, in order to carry out more significant tests. Besides the possibility of determination of data dependent weights towards optimized recognition accuracy, we further plan to conduct test in verification mode as well, where the multimodal biometric systems supposed to verify an identity claim rather than determining the actual identity.

## Acknowledgements

This work has been partly supported by the following EU Networks of Excellence: BioSecure (Proposal/Contract Reference number IST-2002-507634) supported the work on handwriting modality and SIMILAR (Proposal Reference Number: FP6-507609) the Human to Computer Interaction part activities on multimodal fusion. The contents of this publication are the sole responsibility of the authors and can in no way be taken to reflect the views of the European Union.

We would particularly like to thank Jana Dittmann for her support and fruitful discussions in the context of our work.



## References

1. Vielhauer, C., Steinmetz, R., Mayerhöfer, A.: Biometric Hash based on Statistical Features of Online Signature, Proceedings of the International Conference on Pattern Recognition (ICPR), Conference on Pattern Recognition (ICPR), August, Quebec City, Canada, Vol. 1, (2002) 123-126, ISBN 0-7695-1696-3
2. Vielhauer, C., Schimke, S., Thanassis, V., Stylianou, Y.: Fusion Strategies for Speech and Handwriting Modalities in HCI, to appear in: SPIE Proceedings - Electronic Imaging, Security and Watermarking of Multimedia Contents VII (2005)
3. Jain, A.K., Ross, A.: Multibiometric Systems, Communications Of The ACM, Vol. 47, No. 1 (2004) 34-40
4. Ly-Van, B., Blouet, R., Renouard, S., Garcia-Salicetti, S., Dorizzi, B., Chollet, G.: Signature with Text-Dependent and Text-Independent Speech for Robust Identity Verification, Proceedings, IEEE Workshop on Multimodal User Authentication (2003) 13-18
5. Czyz, J., Kittler, J., Vandendorpe, L.: Combining face verification experts, Proceedings of ICPR 2002 – Intl. Conference on Pattern Recognition; Quebec; Canada (2002)
6. Scheidat, T.: Fusion von Klassifikationsverfahren für handschriftbasierte Benutzerauthentifikation, M. Sc. Thesis, University Magdeburg (2005) (in German)
7. Vielhauer, C., Scheidat, T.: Fusion von biometrischen Verfahren zur Benutzerauthentifikation, In: P. Horster (Ed.), D-A-CH Security 2005 – Bestandsaufnahme, Konzepte, Anwendungen, Perspektiven (2005) 82-97, ISBN 3-00-015548-1 (in German)
8. Vielhauer, C.: Biometric User Authentication for IT Security: From Fundamentals to Handwriting, Springer New York, to appear 2006