DISTANCE-LEVEL FUSION STRATEGIES FOR ONLINE SIGNATURE VERIFICATION

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

In this paper an approach for combining online signature authentication experts will be proposed. The different experts are based on one feature extraction method presented in our earlier work, the Biometric Hash algorithm [1], to which different distance measurement functions are applied. We will show that by the fusion of several algorithms with an appropriately parameterized strategy an improvement of the recognition accuracy can be achieved. The best fusion strategy results in a decrease of the EER of 12.1% in comparison to the best individual algorithm. The database we used contains 1761 genuine enrollments (with 4 signatures per enrollment), 1101 genuine verification signatures and 431 well skilled forgeries (so-called "brute force attack") by 22 persons. Based on our experimental results, we further discuss usability of alternative handwriting semantics such as pass phrases or PIN.

1. INTRODUCTION

In order to increase the security of biometric systems some approaches attempt to reach a better performance by combination of various biometric modalities. Such multibiometric systems consist of several biometric subsystems for different modalities (e.g. fingerprint and iris). In general a multibiometric system is based on one of three fusion levels, feature extraction level, matching score level or decision level [2]. In the feature extraction level the information extracted from the different sensors 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. The fusion on matching score level is based on the combination of matching scores after the comparison of reference data and test data. Additionally, matching scores of the different modalities may be weighted. 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. Jain and Ross for example presented a multibiometric system that uses face, fingerprint and hand geometry characteristics of a person for authentication [2]. This system is based on the matching score level strategy.

Another possibility to increase the performance of biometric systems is the combination of several experts of one single individual modality. Our system is based on the handwriting modality and four different distance measures in connection with one particular feature extraction algorithm, the Biometric Hash, as introduced in [1]. A distance measure determines similarity between reference data and test data.

2. ADAPTATION TO DISTANCE-LEVEL FUSION

Our system is based on biometric characteristics of only one modality (the handwriting) whereby different independent settlement proceedings are consulted for the verification. For this purpose the strategies of the multibiometric fusion can be used likewise. The verification decision here is based on a fusion strategy of the respective single results. Our approach combines four distance measures within a biometric system and is based on the matching score level strategy. Because in this case the matching score is a distance value, we call the procedure distance-level fusion. A fusion on distance level is represented in figure 1. In contrast to the multibiometric fusion, our procedures involved use reference data from the same sensor.

The input data for all four algorithms are identical. They consist of physical characteristics of the specimen of handwriting over time. Each expert may use its own feature extraction, but in the current setup, it is the identical for all experts. For future experiments, it will be possible that the experts use different feature extraction algorithms.

Basis of the feature extraction of the four used algorithms is the Biometric Hash algorithm, which was initially reintroduced in [1]. The Biometric Hash algorithm converts the signals of one actually acquired handwriting sample to a unique hash value as a feature vector of fixed dimensionality. Altogether, we implemented four distance measures into the Biometric Hash algorithm, in order to create four different experts.

0-7803-9332-5/05/$20.00 ©2005 IEEE

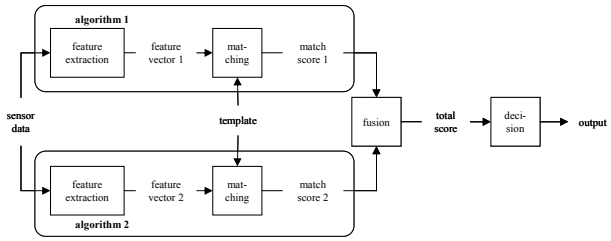

Figure 1. Distance level fusion

2.1 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 define two Biometric Hash vectors *x* and *y*, each of integer value and dimensionality *n*. Smaller distance between any two vectors *x* and *y* denotes greater similarity than larger.

2.1.1. Canberra Distance

The Canberra Distance calculates the sum of a set of ratios between appropriate values.

$$cd(x,y)=\sum_{i=0}^{n}\frac{|x_i-y_i|}{|x_i|+|y_i|}$$

The result is in the interval [0, n]. In our system *n* is equal to 68 for 69 statistical features of one handwriting sample.

2.1.2. City Block Distance

The City Block Distance is the sum of the single distances along each dimension.

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

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

2.1.3. Euclidian Distance

The Euclidian Distance is general the shortest connection between two points.

$$ed(x,y)=\sqrt{\sum_{i=0}^{n}(x_i-y_i)^2}$$

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

2.1.4. Hamming Distance

With the Hamming Distance the elements of the two Biometric Hash vectors, which are on the same index, are compared with each other. If they are identical, the result of the comparison is 0, in the other case 1. The distance is the sum of the single results. For this reason the distance is at least 0 and at the most n.

2.2 Combining Experts

After creating the individual experts we developed several weighting strategies for combining their results. Five tactics for weighting the match scores was developed. Four are based on the EERs of the tests of the individual algorithms. For each weighting strategy the following characteristics are important:

Match Scores: $s_1, s_2, ..., s_n$

Weights: $w_1, w_2, ..., w_n$

2.2.1. Binary weighted fusion

First tactic simulates a fusion on decision level, because it weights the best algorithm with 1 and ignores the others with a weight of 0:

Conditions: $w_1 + w_2 + ... + w_n = 1$

$w_i = 1, if\ s_i = \max(s_1, s_2, ..., s_n) = s_{\max}, w_i = 0, else$

Fusion: $s_{fus} = w_1 s_1 + w_2 s_2 + ... + w_n s_n = s_{\max}$

2.2.2. Equal weighted fusion

The second strategy is an equal weighting tactic, witch it provides all procedures involved independently of the determined EER with the same weight. In this case the value is 0.25 for each algorithm.

Conditions: $w_1 + w_2 + ... + w_n = 1$

$w_1 = w_2 = ... = w_n = n^{-1}$

Fusion: $s_{fus} = w_1 s_1 + w_2 s_2 + ... + w_n s_n$

2.2.3. Linear weighted fusion 1

With the first linear weighting strategy the best algorithm is weighted in dependence to the worst algorithm. That means, the more largely the EER of the worst algorithm, the more largely is the weight for the best algorithm. In the first step the EERs of advice of the algorithms is sorted according to the size. Then the individual weights are computed according to the following formula:

$$w_i = \frac{eer_i}{\sum_{m=1}^{n} eer_m}$$

In the last step the determined weights descending that ascending sorted EERs assigned.

Conditions: $w_1 + w_2 + ... + w_n = 1$

$eer_{w_1} < eer_{w_2} < \ldots < eer_{w_a}$

$eer_{s_1} > eer_{s_2} > \ldots > eer_{s_b}$

Fusion: $s_{fus} = w_a s_1 + w_{a-1} s_2 + ... + w_2 s_{b-1} + w_1 s_b$

2.2.4. Linear weighted fusion 2

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

Conditions: $w_1 + w_2 + ... + w_n = 1$

$$w_i = \frac{\left(\sum_{j=1}^{n} eer_j\right) - eer_i}{\sum_{j=1}^{n} eer_j} \cdot \frac{1}{(n-1)}$$

Fusion : $s_{fus} = w_1 s_1 + w_2 s_2 + \ldots + w_n s_n$

2.2.5. Quadratic weighted fusion

Because the linear weighting strategy 1 was in most cases the best, the quadratic fusion strategy is based on it. It is the square of the value determined there. The sum of the weights must be again 1.

Conditions : $w_1 + w_2 + \ldots + w_n = 1$

$$w_i = \frac{w_{linear1\,i}^{\;2}}{\sum_{j=1}^{n} w_{linear1\,j}^{\;2}}$$

Fusion : $s_{fus} = w_1 s_1 + w_2 s_2 + \ldots + w_n s_n$

The set of weighting strategies used is a first selection of many more possibilities. There are still many other strategies, from which are surely some better than these.

3. EXPERIMENTAL RESULTS

3.1 Test Database

Our evaluation database of handwriting samples is structured in five semantic classes on various graphic and signature tablets. Semantic class denotes handwriting alternatives to signatures, which in our case consist of the semantics PIN (predefined for all users as '8710'), user-defined pass phrase, password (given for all users as the German word 'Sauerstoffgefäß') and a user-defined symbol in addition to signature. In this paper we present our initial results based on samples obtained from one selected graphic tablet, the Wacom Cintiq15. It consists of 1761 genuine enrollments (with 4 signatures per enrollment), 1101 genuine verification signatures and 431 brute force forgeries of 22 users. Brute force forgeries have been generated with the highest level of knowledge of the genuine signature. For further information about our evaluation methodology based on semantic classes, hardware dependency and attack strength see [3]. We divided our tests into three scenarios: First we examined verifications and blind forgeries of the signatures. Secondly we compared verifications and brute force attacks of the signatures. In the last step we determined the best verification strategy separately for each semantic class, for comparison to signatures.

3.2 Methodology and Metric

In our investigations we determined the error rates of the individual algorithms and fusion strategies. The false non match rate (FNMR) indicates, how frequently authentic persons are rejected. The acceptance rate of non-authentic subjects is represented by the false match rate (FMR). For the comparison of accuracy of the individual algorithms as well as those of the fusion results, the equal error rate (EER) has been used, where EER denotes the point in the error characteristics, where FNMR and the FMR yield identical value. Although we are aware that the EER does not represent the optimal operating point of our algorithm, we assume that it offers itself a reference point for comparison of the different procedures.

Our evaluation methodology is based on the concept of skilled forgery strength as introduced in [4]. In the first step the algorithms were examined individually. We used only the semantic class of signatures captured on the Wacom Cintiq15 graphic tablet. The EERs are determined by comparisons of verifications on one side and random attacks respectively brute force attacks on the other. Random attacks denote verification attempts of samples of all users except the actually enrolled user. The comparison of verification and random attack is a simulation of a best case scenario, with only genuine users and the system is not exposed to any skilled forgeries. The worst case is simulated by the verification and brute force attack comparison, where test subjects have been asked to produce skilled forgeries after observation of the original writer's behavior. After determining the EERs of the individual algorithms we calculated the fusion weights according to section 2.2. In a second test run, we determined the matching scores of the comparisons of each verification, random attack and brute force attack of each person for the weighted fusion.

3.3 Experimental Results

Table 1 shows for the signatures the EERs and the weights for verifications/random attacks, derived from the EERs. The first row shows the EERs of the tests of the individual algorithms. The "Weights" rows contain the weights corresponding to the weighting strategies. In the last column the results of the separately weighted fusion strategies are shown. Here, the binary weighting strategy selects the Canberra Distance, as we have observed that this has been the best algorithm in the majority of tests. From Table 1, we can see that the quadratic fusion achieves an improvement in relation to the best single algorithm: EER of the Canberra Distance of 0,091 (first line, column "Canberra") could be improved by the quadratic fusion on 0,080 (last line, column "EER fusion"). The worst results in Table 1 result from the equal weighted strategy and the linear weighted strategy 2. These two tactics alone do not seem to be suitable, in order to accomplish reasonable fusion results.

For further evaluation, we determined the weights additionally for the comparison of verification and brute force attack for signatures. The EERs and weights for this test are presented in table 2. In this case, no improvement was reached for the quadratic fusion. However, apart from the binary strategy, it showed the second best result. Here it can be stated again that the equal weighted fusion and the linear weighted fusion 2 yield the two worst EERs.

Table 1. Weights and EERs of verification/random attack

	Algorithm	City-Block	Canberra	Euclid	Hamming	EER Fusion
	EER	0,388	**0,091**	0,376	0,092	
Weights	binary	0,000	1,000	0,000	0,000	0,091
	equal	0,250	0,250	0,250	0,250	0,276
	linear1	0,096	0,410	0,097	0,397	0,122
	linear2	0,197	0,301	0,201	0,301	0,235
	quadratic	0,027	0,488	0,027	0,458	**0,080**

Table 2. Weights and EERs of verification/brute force attack

	Algorithm	City-Block	Canberra	Euclid	Hamming	EER Fusion
	EER	0,642	**0,230**	0,655	0,301	
Weights	binary	0,000	1,000	0,000	0,000	**0,230**
	equal	0,250	0,250	0,250	0,250	0,512
	linear1	0,165	0,358	0,126	0,351	0,465
	linear2	0,216	0,291	0,214	0,278	0,525
	quadratic	0,092	0,436	0,054	0,419	0,379

Table 3. List of the best classification strategies

	Verification/Random Attack		Verification/Brute Force Attack	
Semantic	best Algorithm	EER	best Algorithm	EER
8710	Canberra	0,253	Canberra	0,289
Passphrase	Canberra	0,073	City Block/Fusion	0,250
Sauerstoffgefäß	Hamming	0,312	Fusion	0,494
Signature	Fusion	0,080	Canberra	0,230
Symbol	Hamming/Fusion	0,064	Canberra	0,228

Table 3 summarizes over the entire set of five semantic classes that the quadratic fusion has shown to be at least as good as the best single algorithm in four out of ten cases. However, it is to be seen that the fusion is not successful with verification/random attack and verification/brute force attack with same semantics. Furthermore, it can be also seen that the results of semantics with given textual contents ('8710' and 'Sauerstoffgefäß') are worse than the others. Apparently, the use of same textual content in the handwriting semantic seems to reduce the discriminatory power and thus results in a degradation of the classification of the persons involved.

4. CONCLUSIONS AND FUTURE WORK

In this paper, we have reviewed four different distance measure functions and evaluated five alternative strategies for a matching score level fusion. Based on a reasonably large database of more than 3000 samples of different semantic content, our first observation is that not one particular distance measure could be identified as the single best for each semantic class. Secondly, we were able to show that in many cases one of our suggested fusion strategies, the quadratic weighted fusion, yields equal or better result than the best single distance algorithm.

Further improvements by the fusion appear feasible, if the weights are determined not globally for a whole semantic class and/or a tray. Rather than that, the enrollments and verifications of individual persons could be observed. From this, individual weights can be considered for each individual person (user-adaptive weighting).

With each new verification, the weights could be further adapted within the system. However, this needs to be examined and confirmed by further tests. Further, additional parametric distance measures, like for example the Mahalanobis or Correlation Distance, can be included in future evaluations. Also, it is possible to include different verification algorithms, which are not only based on alternative distance measures. In this case, it may become necessary to change of the fusion strategy, for example by choice of another fusion level.

The results of this work open the possibility for single or multi biometric procedures, the examined distance measures could also be applied in the future for feature representations from other biometric modalities, such as the iris code [5].

5. ACKNOWLEDGEMENT

The work on fusion strategies described in this paper has been supported in part by the European Commission through the IST Programme under Contract IST-2002-507634 BIOSECURE. 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. Efforts for the research on distance measures are partly sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant number FA8655-04-1-3010. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Office of Scientific Research or the U.S. Government.

6. REFERENCES

[1] C. Vielhauer, R. Steinmetz, A. Mayerhöfer, "Biometric Hash based on Statistical Features of Online Signature", Proc. of the Intern. Conf. on Pattern Recognition (ICPR), Conference on Pattern Recognition (ICPR), Quebec City, Canada, Vol. 1, pp. 123-126, 2002

[2] A.K. Jain, A. Ross, "Multibiometric Systems", Communications Of The ACM, Vol. 47, No. 1, pp. 34-40, 2004

[3] C. Vielhauer, "Biometric User Authentication For IT Security: From Fundamentals to Handwriting", Springer, New York, U.S.A., to appear 2006.

[4] F. Zöbisch, C. Vielhauer: "A Test Tool to support Brut-Force Online and Offline Signature Forgery Tests on Mobile Devices", Proc. of IEEE International Conference on Multimedia and Expo 2003 (ICME), Baltimore, U.S.A., Vol. 3, pp. 225-228, 2003

[5] J. Daugman: "The importance of being random: Statistical principles of iris re-cognition", Pattern Recognition, Vol. 36, No. 2, pp. 279 - 291, 2003