Study of Applicability of Virtual Users in Evaluating Multimodal Biometrics

Franziska Wolf, Tobias Scheidat, Claus Vielhauer

Otto-von-Guericke University of Magdeburg, Department of Computer Science, ITI Research Group on Multimedia and Security,
Universitätsplatz 2, 39106 Magdeburg, Germany
{franziska.wolf|tobias.scheidat|claus.vielhauer}
@iti.cs.uni-magdeburg.de

Abstract. A new approach of enlarging fused biometric databases is presented. Fusion strategies based upon matching score are applied on active biometrics verification scenarios. Consistent biometric data of two traits are used in test scenarios of handwriting and speaker verification. The fusion strategies are applied on multimodal biometrics of two different user types. The *real users* represent two biometric traits captured from one person. The *virtual users* are considered as the combination of two traits captured from two discrete users. These virtual users are implemented for database enlargement. In order to investigate the impact of these virtual users, test scenarios using three different semantics of handwriting and speech are accomplished. The results of fused handwriting and speech of exclusively real users and additional virtual users are compared and discussed.

1 Introduction

The need for secure biometric authentication methods arises strongly to give consideration to the growing requirements in automatic user authentication in our today's technical world.

Biometrics can be divided into the passive traits using physiological characteristics and the active traits based on human behavioral specifics for authentication. But the biometric research hasn't achieved an adequate recognition performance yet. The reasons base upon noisy-data, intra-class variations, inter-class similarities and non-universality ([1]). Especially active biometrics show a great amount of variability and therefore lack adequate security levels. In order to solve the mentioned problems, some single modality systems (e.g. based on voice *or* handwriting) have been fused (e.g. using voice *and* handwriting). Ross and Jain differentiate in [1] biometric fusion scenarios by single or multi usage of traits, sensors and classifiers. From the variety of approaches to multimodal biometric systems we will introduce present work. Actually the fusion of passive biometric traits is researched broadly. Such as Jain and Ross present in [2] a biometric system that uses face, fingerprint and hand geometry for authentication. Fierrez-Aguilar et al. combine the active and passive traits of as unimodal systems of face, fingerprint and online signature ([3]). Vielhauer et al. present in [4] a multimodal system where a speech recognition system and a signature recognition system are fused on matching score level. In [5] the multimodal system above is enhanced exchanging a single signature component by a multi-algorithmic handwriting subsystem.

In order to provide statistical reliance in studies of multimodal systems, a large amount of data is essential. In recent years varying multimodal databases for training and testing verification systems have been developed. For example the BANCA database contains face and voice modalities of 208 peoples captured in four European languages ([6]). XM2VTSDB contains sound files of voice along with video sequences,3D models of the speaking and a rotating head shot of 295 subjects taken over a period of four months ([7]). Future projects like MyIDea shall include talking face, audio, fingerprints, signature, handwriting and hand geometry ([8]).

According to our investigations, the fusion of multiple active biometric exclusively has not been undertaken so far. But especially this combination of active biometrics like handwriting and speech offers various scenarios of use, because they comprise authentication, written contracts and options of various controls.

A general application area for handwriting and speech could be for example the automotive domain as well the use for contracts, such as rental agreements. For cars a combination of handwriting and voice recognition could be used for driver authentication in the car environment ([9]).

The data acquisition of active biometrics is tedious. Therefore the practice of enlarging multimodal test databases by combination of differing traits and users has been applied recently in varying biometric works. Multimodal databases have been enlarged by virtual users also, in [10] finger geometry and feature extraction of the palmar flexion creases are integrated in a few number of discrete points for faster and robust processing.

The method of artificial enlargement is justified assuming the independence of biometric traits that hasn't been proven hitherto but partly questioned [11].

Based on our knowledge also no exclusive assumption concerning active traits has been researched. Also combined databases of speech and various handwriting data have not been set up so far. Therefore we will present first approaches to the study of virtual users.

In this work we will investigate an impact of virtual enlargement on fused biometrics. As research subject single active biometric traits (handwriting [12] and voice) of differing persons are combined to one *virtual user*. The fused traits of real persons and the virtual user results will be tested by an authentication system and the results will be compared. We aim to achieve perceptions on the distinct impact of virtual enlargement of multimodal databases and research if an increment of data amount has to be paid with a decrement of quality.

This paper is structured as follows: In the next section the multimodal fusion strategies are presented and the underlying verification algorithm is given. Section 3 gives an overview of the samples of the biometric database and describes the combination strategies of virtual users for enlarging. Section 4 shows the test results and a discussion of their meaning. A short summary of this paper and an outlook of future work are given in section 5.

2 Fusion strategies

The underlying verification algorithm for the handwriting modality is based on the Biometric Hash algorithm, as described in [12]. The speaker verification bases upon Mel-Frequency Cepstrum Coefficients (MFCC) using frequency in a logarithmic scale, the mel scale ([13]) and the log spectrum, the cepstrum ([14]) for distinguishing sounds.

2.1 Description of Fusion

In general a multimodal system is based on one of three fusion levels ([15]) depending on the point of fusion within the single systems involved (see Figure 1): feature extraction level, matching score level or decision level. The data itself or the extracted features are fused on the *feature extraction level*. At the *matching score level* the matching scores of all subsystems involved are combined by the multimodal system. In order to parameterize the subsystems, matching scores of the different modalities may be weighted and the decision will be determined. For a fusion on *decision level* each subsystem involved is processed completely and the individual decisions are fused to a final decision, i.e. by Boolean operations.

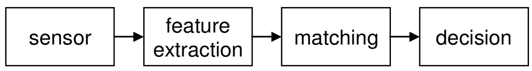

Fig. 1. Modules of a biometric system

In our approach for every modality (handwriting and speech) the matching scores are calculated separately and fused on the matching score level. With this serial method the Equal Error Rate (EER) for every modality can be obtained and used to calculate a fusion weight for this modality based upon a certain fusion strategy. The Equal Error Rate (EER) denotes the point where FRR and FAR yield identical values and is used to compare the results of different tests. The False Rejection Rate (FRR) indicates how frequently authentic persons are rejected from the system whereas the acceptance rate of non-authentic subjects is represented by the False Acceptance Rate (FAR).

2.2 Combining handwriting and speech

In this work the fusion bases on two active biometric traits. The subsystems involved are a handwriting recognition system and a speaker verification system. The combination of the *n* subsystems of the multimodal fusion suggested in this work is accomplished on the matching score level. Input from different modalities requires different feature extraction algorithms so that normalization is necessary. Details of the underlying feature extracting algorithms can be found in [16].

A matching score s of one modality is normalized (s_{norm}) to a range of three times the standard deviation σ_{verif} around the mean $\bar{s}_{verif}$ of the verification as shown in following equation:

$$s'_{norm} = \frac{1}{2} + \frac{s - \bar{s}_{verif}}{3\sigma_{verif}}$$
$$s_{norm} = \begin{cases} 0 \text{ if } s'_{norm} < 0, \\ 1 \text{ if } s'_{norm} > 1, \\ s'_{norm} \text{ otherwise} \end{cases} \tag{1}$$

The value 3σ was chosen to represent approximately 99% on a normalized scale. After the normalization of the matching scores the fusion module combines them to a joint matching score. For the fusion we use one of five weighting strategies presented in previous work ([15]) for multi-algorithmic fusion:

$$\begin{array}{ll} \textit{Match Scores}: & s_1, s_2, ..., s_n \\ \textit{Weights}: & w_1, w_2, ..., w_n \\ & n = \textit{number of systems involved} \end{array} \tag{2}$$

Linear weighted fusion. With this strategy the systems are weighted by the relations of the EERs. The system, which received the highest EER, gets the smaller weight and contrary. The individual weights are determined according to the following formula:

$$\begin{array}{ll} & w_i = \dfrac{eer_i}{\sum_{m=1}^{n} eer_m} \\ \textit{Conditions}: & w_1 + w_2 + ... + w_n = 1 \\ \textit{Fusion}: & s_{fus} = w_1 s_1 + w_2 s_2 + ... + w_{n-1} s_{n-1} + w_n s_n \end{array} \tag{3}$$

A generalized form is presented here by the formula. In this work we will focus on the limit of $n = 2$ modalities. The joint matching score of the fusion is used by the decision module to determine the final authentication result of the whole system.

3 Methodology

In this section the test-database and the captured data are presented. Furthermore we describe the methodology of building virtual users. In order to evaluate the virtual enlargement of the test-database the results based on EER of the virtual database are compared to the original.

3.1 Test-database

The biometric database was captured as follows: For each modality three alternative semantics that are described in table 1 were chosen out of 48 semantics collected following the test plan described in [17]. The *Signature* represents a traditional and accepted feature for the written user authentication and has *Name* as counterpart for the speech trait. A predefined *PIN* given as “7-79-93” in both handwriting and speech such as the given *Sentence* “Hello, how are you?”.

One subset of the test subjects contributes biometric data of both traits and will be used building up the database of real users. The second subset provides one trait and will be used for combining the virtual users. Both subsets are presented in table 1. Each semantic of each trait has been captured ten times and five samples are used for enrollment creation and the remaining five samples for verification creation

Table 1. Number of test subjects

Semantic of Handwriting / Speech	Subset providing both traits	Subset providing one trait	
		Handwriting	Speech
Signature / Name	27	63	39
PIN	21	68	32
Sentence	23	47	38

3.2 How to create Virtual Users

In order to enlarge the group of investigated collected handwriting and speech data was build up to new virtual identities combining identities and single modality data of different users as showed in Figure 1.

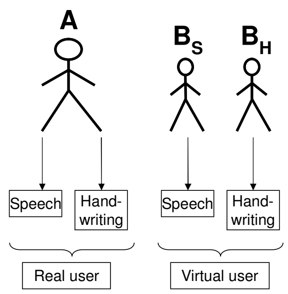

Fig 1. Methodology of combining virtual users

A user of type A is defined by the fact that he or she has donated both required traits to the database, handwriting and speech data, and is therefore designated as *real user.* User of type B, who provided only one biometric trait, either handwriting data (B_H) *or* speech data (B_S), are combined and designated as *virtual user.* The results are two test sets of three semantics each, one holding only real users and their test data, the other holding real users plus virtual combined users. This strategy of building virtual multimodal users resulted in a total test set size of 39 (plus 44.4%) for the *Signature/Name* scenario, 32 (plus 52.4%) for the *PIN* and 38 (plus 65.2%) for the *Sentence* scenario.

4. Experimental results

The results for user groups are shown in Table 2 that holds the number of 21-27 real users depending on the test scenario. Table 3 represents the combined database of the real users and 11-15 additional virtual users. Table 4 shows the relative change of the EER (column 1-3) in percent and number of real to virtual users (column 4) comparing Table 2 and 3. The weights are calculated dynamically in order to consider the best quota of the systems involved.

Table 2. Test-database of real users

	No. of Persons	Handwriting EER	Handwriting weights	Speech EER	Speech weights	Fusion
Sign./Name	27	0.0131	0.998	0.3084	0.002	0.0133
PIN	21	0.0353	0.989	0.3276	0.011	0.0317
Sentence	23	0.0248	0.990	0.2476	0.010	0.0214

Table 3. Test results for real and virtual users fused using distinct weights

	No. of Persons	Handwriting EER	Handwriting weights	Speech EER	Speech weights	Fusion
Sign./Name	39 (27+12)	0.0151	0.953	0.3059	0.047	0.0142
PIN	32 (21+11)	0.0344	0.909	0.3438	0.091	0.0323
Sentence	38 (23+15)	0.0292	0.903	0.2720	0.097	0.0284

Table 4. Relative change of EER and rise of test subjects of real and virtual users

Handwriting	Speech	Fusion	Increase subject No.
13.24%	-0.82%	6.34%	44.44%
-2.62%	4.71%	1.86%	52.38%
15.07%	8.97%	24.65%	65.22%

The number of additional virtual users depends on the single-modalities provided by the test-database. Comparing the results of original and enlarged databases using Table 4 (column 1 and 2) the single modalities of handwriting in average show little degradation with respect to recognition accuracy whereas the speech modality shows trends of improvement. Comparing the change of the fusion results (column 3) degradation can be measured compared to the original database. A dependency of additional virtual users and the decreasing quality measures can be assumed from table 4 (column 4). The error rates of fused experiments altogether show encouraging results, as indicated for example by the EER in a range of 1% for the *Signature/Name* scenario. However, the virtually enlarged databases appear to lack confidence compared to of the original ones. This could reveal a dependency of active biometrics that hasn't been researched so far. Future work should study these aspects of active biometric fusion more thoroughly.

5. Conclusions and future work

In this paper we have presented a first approach to study the effects of virtual enlargement of active biometric databases. It could be shown that an enlargement of databases holding biometric data of handwriting and speech can be done by combining single biometric traits of different subjects to so-called virtual users. Our experimental evaluations have shown that an enlargement by approximately 50% leads to degradations of up to approximately 25% with respect to equal error rates.

Nevertheless the results reveal encouraging possibilities and provide considerably further research: In order to achieve a higher statistical confidence, more biometric data has to be captured. Databases containing only virtual users should be investigated concerning grouping metadata like gender or nationality and compared to multimodal databases of real users on a larger scale. On the other hand the fusion should be extended to combinations of differing semantics in handwriting and speech. Other fusion strategies such as equal or quadratic fusion (see [15]) should be researched at alternative weightings of distance level fusion.

In future work we also will consider metadata clustering. A dependency of active biometrics shall be studied in this context. Finally, dependencies of language based biometrics therefore could reveal limits of fusion and virtual combination.

6. Acknowledgment

This work has been partly supported by the EU Networks of Excellence SIMILAR (Proposal Reference Number: FP6−507609) in regards to fusion strategies and BIOSECURE (Contract IST-2002-507634 BIOSECURE) for evaluation and fusion methodologies. With respect to the experimental evaluations, this publication has been produced partly with the assistance of the EU-India cross cultural program (project CultureTech, see [18]). 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. Ross, A., Jain, A.K.: Multimodal Biometrics: An Overview. Proc. Of the 12th European Signal Processing Conference (EUSIPCO), Vienna, Austria (2004) 1221 – 1224

2. Jain, A.K., Ross A.: Multibiometric Systems. Communications Of The ACM, Vol. 47, No. 1 (2004) 34-40
3. Fierrez-Aguilar, J., Ortega-Garcia J., Gonzalez-Rodriguez, J.: Fusion strategies in multimodal biometric verification. Proc. IEEE Intl. Conf. on Multimedia and Expo, ICME, Vol. 3, Baltimore, USA (2003) 5-8
4. Vielhauer, C., Schimke, S., Valsamakis, A., Stylianou, Y.: Fusion Strategies for Speech and Handwriting Modalities in HCI. Proceedings of SPIE-IS&T Electronic Imaging, Vol. 5684 (2005) 63-71
5. Vielhauer, C., Scheidat, T.: Multimodal Biometrics for Voice and Handwriting. Proceedings of the 9th IFIP Conference on Communications and Multimedia Security, Salzburg, Austria (2005) 191-199
6. The BANCA Database. http://www.ee.surrey.ac.uk/banca/, requested April 2006
7. The XM2VTS Database. http://www.ee.surrey.ac.uk/Research/VSSP/xm2vtsdb/, requested April 2006
8. DIUF - BIOMETRICS – MyIDea. http://diuf.unifr.ch/diva/biometrics/MyIdea/, requested April 2006
9. Moreno, A., Lindberg, B., Draxler, C., Richard, G., Choukri, K., Euler, S., Allen, J.: SPEECH DAT CAR. A Large Speech Database For Automotive Environments. Proceedings of the II Language Resources European Conference, Athens, Greece (2000)
10. Ross, A., Jain, A.K.: Information Fusion in Biometrics. Pattern Recognition Letters, Vol. 24, Issue 13 (2003) 2115-2125
11. Poh, N., Bengio, S.: Can Chimeric Persons Be Used in Multimodal Biometric Authentication Experiments?, MLMI, (2005) 87-100
12. Vielhauer, C.: Biometric User Authentication for IT Security: From Fundamentals to Handwriting, Springer, New York (2005)
13. Stevens, S. S., Volkmann, J., Newman, E. B.: A Scale for the Measurement of the Psychological Magnitude Pitch. Journal of the Acoustic Society of America, Vol. 8 (1937) 185-190
14. Tukey, J. W., Bogert, B. P., Healy, J. R.: The quefrency alanysis of time series for echoes: cepstrum, pseudo-autovariance, cross-cepstrum and saphe cracking, Proceedings of the Symposium on Time Series Analysis (1963) 209-243
15. Scheidat, T., Vielhauer, C., Dittmann, J.: Distance-Level Fusion Strategies for Online Signature Verification. Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), Amsterdam, The Netherlands (2005)
16. Vielhauer, C., Scheidat, T., Lang, A., Schott, M., Dittmann, J., Basu, T.K., Dutta, P.K.: Multimodal Speaker Authentication – Evaluation of Recognition Performance of Watermarked References, to appear in Proceedings of MMUA, Toulouse, France (2006)
17. Scheidat, T., Wolf, F., Vielhauer, C.: Analyzing Handwriting Biometrics in Metadata Context. Proceedings SPIE conference at the Security, Steganography, and Watermarking of Multimedia Contents VIII, IS&T/SPIE Symposium on Electronic Imaging, San Jose, USA (2006)
18. The Culture Tech Project, Cultural Dimensions in digital Multimedia Security Technology, a project funded under *the EU-India Economic Cross Cultural Program*, http://amsl-smb.cs.uni-magdeburg.de/culturetech/, requested March 2006