Semantic Fusion for Biometric User Authentication as Multimodal Signal Processing

Andrea Oermann, Tobias Scheidat, Claus Vielhauer, and Jana Dittmann

Otto-von-Guericke-University of Magdeburg, Universitaetsplatz 2, 39106 Magdeburg, Germany
{andrea.oermann, tobias.scheidat, claus.vielhauer, jana.dittmann}@iti.cs.uni-magdeburg.de

Abstract. Today the application of multimodal biometric systems is a common way to overcome the problems, which come with unimodal systems, such as noisy data, attacks, overlapping of similarities, and non-universality of biometric characteristics. In order to fuse multiple identification sources simultaneously, fusion strategies can be applied on different levels. This paper presents a theoretical concept of a methodology to improve those fusions and strategies independently of their application levels. By extracting and merging certain semantic information and integrating it as additional knowledge (e.g. metadata) into the process the fusion can be potentially improved. Thus, discrepancies and irregularities of one biometric trait can be verified by another one and signal errors can be identified and corrected.

Keywords: Biometrics, Security of Multimedia Content, Identification and Authentication

1 Motivation

Biometrics offers evidently promising techniques in order to determine an individual's identity and authorize an individual in time and location independent communication systems. Biometric systems consider different biological characteristics, also known as traits [1], [2], which identify an individual's uniqueness. Unimodal biometric systems, which consider only one source of information to authenticate an individual, are established in common applications. However, unimodal systems are associated with several problems: background noise, attacks, overlapping of similarities, and non-universality of biometric characteristics [1]. Hence, the development of multimodal biometric systems, which consider multiple sources of information, is a major focus of recent research.

A biometric modality contains several aspects, and can be formally represented as

$$m = \{se, bt, rp, ins, sa\} \tag{1}$$

where se = sensor, bt = biometric trait, rp = representation, ins = instance and sa = particular sample. Fusion strategies are applied in order to increase the performance of the authentication. As described in [1] and formalized in section 2, a fusion can be applied on different levels in a biometric authentication system while varying aspects are considered.

In this paper we present a theoretical methodology to further improve fusion strategies for biometric user authentication. Our approach, as presented in detail in section 3 and 4, is a basic concept to classify certain information features within different levels, differentiated in syntax and semantics. The model's priority is integrating semantic aspects in fusion strategies to detect earlier occurred discrepancies and irregularities. This model provides an approach to evaluate the correctness of one particular biometric system and to verify the occurrence of signal errors. Furthermore, extracting certain semantic information and integrating it as additional knowledge (e.g. metadata [3], [4], [5], [6]) into the process, the fusion accuracy may be improved.

Further, our model enables to analyze and improve the performance of an online synchronization of two parallel biometric systems by merging semantic features of two biometric systems as demonstrated in section 4. Thus, the authenticity of a person can be verified with a higher level of security. Attacks can be more reliably detected. This online synchronization is becoming increasingly important regarding for example the security of automobiles. Here, more than one biometric capturing sensor should be applied for authentication due to many factors such as noise, dirt, different cultural, ethnical and conditional background of frequently changing persons [6], etc. are impacting the signals. Our concept implies a scalable methodology which can be applied for any media or biometric system, and whose practical evaluation is subject of ongoing research.

2 Fusion Strategies for Biometric User Authentication

Multimodal biometric systems join several biometric subsystems for different modalities (e.g. voice and handwriting [7]). There are different strategies such as [2] and [8] when, how, and where to apply a fusion in order to increase the performance of a biometric user authentication process in a multimodal biometric system, also known as a multibiometric system. Basically three fusion levels can be applied: feature level, matching score level or decision level [2]. In addition to these three fusion levels, two fusion levels, sensor level and rank level have been introduced by [9] and [10], as it is shown in Fig. 1.

A fusion on sensor level indicates that biometric raw data captured by different sensors are consolidated to generate new raw data. Given a set of k different biometric modalities $\{m_1, ..., m_k\}$, a fusion function FU on sensor level sl can be formally represented as

$$\mathbf{FU}_{sl}(m_1, ..., m_k) = rd_{m_1} \diamond \ldots \diamond rd_{m_k}; k \in \mathbf{N} \qquad (2)$$

where rd stands for raw data. The result type of the fusion function FU_{sl} and the generic fusion operator $\diamond$ for the sensor level depends on the specific modality and sensor, and can for example be a digital image representation.

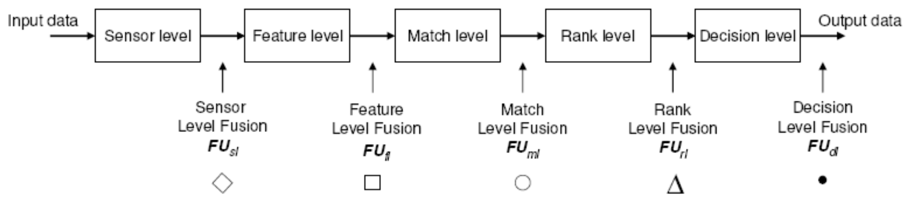

Fig. 1. Overview of 5 different fusion levels FU_{sl}, FU_{fl}, FU_{ml}, FU_{rl}, FU_{dl} and the respective generic fusion symbols

If a fusion is placed on feature level, joint feature vectors will be compared with feature vectors stored in a reference database. During this process vectors can also be weighted. Given a set of k different biometric modalities $\{m_1, ..., m_k\}$, the fusion FU function on feature level fl can be formally represented as

$$\mathbf{FU}_{fl}(m_1, ..., m_k) = \begin{pmatrix} w_{m_{1,1}} \\ \cdots \\ w_{m_{1,l}} \end{pmatrix} \begin{pmatrix} fv_{m_{1,1}} \\ \cdots \\ fv_{m_{1,l}} \end{pmatrix} \Box \ldots \Box \begin{pmatrix} w_{m_{k,1}} \\ \cdots \\ w_{m_{k,l}} \end{pmatrix} \begin{pmatrix} fv_{m_{k,1}} \\ \cdots \\ fv_{m_{k,l}} \end{pmatrix} \quad (3)$$

$$k, l \in \mathbf{N}, w = [0, 1], w \in \mathbf{R}$$

where *fv* is a specific extracted feature of a modality, w is applied as a weight quantifier, l is the number of extracted features and determines the dimension of each feature vector. For simplification all feature vectors have the same dimension. The generic fusion operator for the feature extraction level $\Box$ may be represented for example by concatenation of feature vectors as a fusion option and the result type of the fusion function FU_{fl} and the generic fusion operator $\Box$ for the feature level is a new feature vector.

A fusion on matching score level implies a consolidation of matching scores gained from separate comparisons of reference data and test data for each modality. Additionally, matching scores can be weighted. Given a set of k different biometric modalities $\{m_1, ..., m_k\}$, the fusion FU on matching score level ml can be formally represented as

$$\mathbf{FU}_{ml}(m_1, ..., m_k) = s_{m_1} w_{m_1} \bigcirc \ldots \bigcirc s_{m_k} w_{m_k} \quad (4)$$

$$k \in \mathbf{N}, w = [0, 1]$$

where w denotes a quantifier. The generic fusion operator for the matching level $\bigcirc$ may be represented by mathematical operations such as SUM, MULT, MEAN, or MEDIAN; the result type of the fusion function FU_{ml} and the generic fusion operator $\bigcirc$ for the matching score level may be a scalar value.

Following [9], a fusion, which is applied at ranking level, includes the consolidation of the multiple ranks associated with an identity into a new rank the final decision can rely on. Given a set of k different biometric modalities $\{m_1, ..., m_k\}$, the fusion FU on rank level rl can be formally represented as

$$\mathbf{FU}_{rl}(m_1, ..., m_k) = r_{u_{x_{m_1}}} \triangle \ldots \triangle r_{u_{x_{m_k}}} \tag{5}$$

$x = 1, ..., n; k, n \in \mathbf{N}$

where x indicates arbitrary user u for whose the rankings of each k modality are consolidated. As a result type of the fusion function FU_{rl} and the generic fusion operator $\triangle$ for the rank level, a new ranking list will impact the decision about the identity of an user.

In case a fusion is applied on decision level, each biometric subsystem draws a completely autonomous decision. The multimodal decision combines each of these individual decisions by boolean operations. Given a set of k different biometric modalities $\{m_1, ..., m_k\}$, the fusion FU on decision level dl can be formally represented as

$$\mathbf{FU}_{dl}(m_1, ..., m_k) = d_{m_1} \bullet \ldots \bullet d_{m_k}; k \in \mathbf{N}, d = 0, 1 \tag{6}$$

The generic fusion operator for the decision level $\bullet$ may be represented for example by conjunction $\wedge$, disjunction $\vee$, or XOR. The result type of the fusion function FU_{dl} and the generic fusion operator $\bullet$ for the decision level is a boolean value.

As it can be seen, the complexity and the output of a fusion differs by its level of application. This paper presents a methodology how to additionally increase the performance independent from the fusion level by a semantic fusion, which will be introduced in the following sections 3 and 4.

3 Reference Model

This reference model is a so-called "Verifier-Tuple" to classify information in order to cluster specific information features. "Verifier-Tuple" is derived from a general concept of the explanation of programming languages as it is presented in [11]. It describes a combination of syntax and semantics, as introduced in [12] and further applied in [13]. According to [14], we now additionally differentiate three interdependent levels of syntax as an extension of our basic "Verifier-Tuple". Instead of only four levels of information we now distinguish between six levels of information, as it can be seen from equation 7.

$$V = \{SY_P, SY_L, SY_C, SE_E, SE_F, SE_A\} \tag{7}$$

These six levels, which are divided in two main domains - syntax and semantics, are the following:

Syntactic domain SY:

1. Syntax (SY_P) - physical level (location and characteristics of storage)
2. Syntax (SY_L) - logical level (bit-streams, formates)
3. Syntax (SY_C) - conceptual level (information)

Semantic domain SE:

1. Semantics (SE_S) - structural level
2. Semantics (SE_F) - functional level
3. Semantics (SE_A) - analytical level

As a result of this "Verifier-Tuple", a more precise information analysis can be outlined. By this methodology information features can be extracted and structurally analyzed in order to detect signal errors. This classification of information is required to be able to efficiently analyze information and to capture the whole context. The specific classification of features for different modalities such as voice, handwriting and video-based face is presented in [15].

According to equation 7, the extraction of syntactic and semantic features of an arbitrary modality m_i can be formally represented in vectors as follows:

$$\overrightarrow{sfv}_{m_i} = \{\overrightarrow{sfv}_{SY_P}, \overrightarrow{sfv}_{SY_L}, \overrightarrow{sfv}_{SY_C}, \overrightarrow{sfv}_{SE_S}, \overrightarrow{sfv}_{SE_F}, \overrightarrow{sfv}_{SE_A}\} \qquad (8)$$

$i = 1, ..., k, k \in \mathbf{N}$

$$\overrightarrow{sfv}_{m_i} = \{\begin{pmatrix} psy_1 \\ \cdots \\ psy_a \end{pmatrix}, \begin{pmatrix} lsy_1 \\ \cdots \\ lsy_b \end{pmatrix}, \begin{pmatrix} csy_1 \\ \cdots \\ csy_c \end{pmatrix}, \begin{pmatrix} sse_1 \\ \cdots \\ sse_d \end{pmatrix}, \begin{pmatrix} fse_1 \\ \cdots \\ fse_e \end{pmatrix}, \begin{pmatrix} ase_1 \\ \cdots \\ ase_f \end{pmatrix}\} \qquad (9)$$

$a, b, c, d, e, f \in \mathbf{N}$

where $\overrightarrow{sfv}_{SY_P}$ represents the vector of features on the physical syntactic level and $\{psy_1, ..., psy_a\}$ is the set of specific features, $\overrightarrow{sfv}_{SY_L}$ represents the vector of features on the logical syntactic level and $\{lsy_1, ..., lsy_b\}$ is the set of specific features, $\overrightarrow{sfv}_{SY_C}$ represents the vector of features on the conceptual syntactic level and $\{csy_1, ..., csy_c\}$ is the set of specific features, $\overrightarrow{sfv}_{SE_S}$ represents the vector of features on the structural semantic level and $\{sse_1, ..., sse_d\}$ is the set of specific features, $\overrightarrow{sfv}_{SE_F}$ represents the vector of features on the functional semantic level and $\{fse_1, ..., fse_e\}$ is the set of specific features and $\overrightarrow{sfv}_{SE_A}$ represents the vector of features on the analytical semantic level and $\{ase_1, ..., ase_f\}$ is the set of specific features. Vectors can have varying dimensions, which is indicated through *a,b,c,d,e,* and *f*. For analyzing information not necessarily all information levels need to be integrated or can be considered in the fusion levels. In this case the corresponding vector will be set to 0.

Considering equation 1 in the introduction, the basic definition of m can now be replaced by the modality specific feature vector $\overrightarrow{sfv}_{m_i}$. It not only includes the basic definition but also goes beyond and provides a more detailed representation of a modality.

4 Semantic Fusion Approach

Certain biometric features can be better captured in one modality than in another one. For example emotions, conditions, cultural and ethnical background of individuals can be better traced in a video stream of the face than in an audio stream of the voice. Thus, semantic aspects especially of higher levels (functional and analytical level) as additional components of a fusion approach are promising to integrate, since two different modalities are hardly comparable on syntactical levels or the structural semantic level. Especially when synchronizing varying modalities online, the integration of additional knowledge can improve the performance of an authentication technique. Depending on the fusion level, varying levels of information can be considered to be integrated into the fusion process in order to increase the performance.

As shown in equation 10, we introduce the semantic fusion function $\overrightarrow{SFU}$ with the fusion operator $\bigodot$ as a consolidation of specific feature vectors of each modality in addition to the basic biometric fusion. Hence, the biometric fusion can be controlled as presented in Fig. 2.

$$\overrightarrow{SFU} = \overrightarrow{sfv}_{m_1} \bigodot \ldots \bigodot \overrightarrow{sfv}_{m_k} \tag{10}$$

Semantic fusion operator $\bigodot$ can be represented for example by consolidation of specific feature vectors introduced in section 3. An exemplary outcome can be found in [15].

Our introduced Verifier-Tuple functions as an additional analysis to control the fusion process. Each captured modality data itself will be syntactically and structural semantically analyzed. In addition to the basic definition of m in the introduction, especially features on the functional and analytical semantic level can now be structurally extracted and additionally be compared in order to evaluate and improve the accuracy of a fusion strategy and to verify occurring signal errors on the syntactical levels. Thus, an efficient fusion of two different biometric modalities, either synchronous or asynchronous, may be achieved.

By applying our methodology, an error analysis can be performed. The signal can be corrected and background noises can be eliminated in the syntactic domains of each biometric modality. In other words, the best biometric modality can be identified. The fusion can be potentially improved by applying our model. Further, the accuracy of the synchronization of the two signals can be systematically tested. Thus, attacks can be more reliably detected. For example, the visual analysis of the mouth movement considering the corresponding voice can discover an attack which would not have been identified as an attack only by analyzing the voice stream. To summarize the application of methodology, it can be said that strategies for fusions can be systemized by our theoretical approach.

5 Summary and Outlook

In this paper, a theoretical concept of a new information classification model is presented in order to improve fusion strategies for multimodal biometric user authentication systems. This model as a so-called "Verifier-Tuple" impacts the fusion process in a controlling function. It provides a methodology to efficiently analyze information, differentiated in a syntactic and semantic domain. Six different classification levels are available to extract certain syntactic and semantic information. Features on the semantic levels are extracted and compared in order to verify occurring signal errors on the syntactical levels and consequently to evaluate the accuracy of a fusion strategy and.

By doing so, our "Verifier-Tuple" controls the outcome of a fusion. Hence, irregularities can be detected. Especially the integration of semantics into the fusion process is the challenge of our model. The presented methodology can be applied independently of the level a fusion which is established in a multimodal biometric user authentication system.

The focus of future work has to be the transfer the theory into practice, in particular to to evaluate methods, which detect and analyze semantic features. This is not only a challenge in itself for a single modality, e.g. speech recognition, emotion detection, biometric voice identification. Questions such as: "What are good features to detect semantic irregularities?" or: "Which extracted semantic information is additionally and most suitably influencing the fusion in order to get the most accurate result?" need to be answered. But it is also a challenge considering the synchronization of varying biometric modalities. The goal of this paper was to provide the theoretical basis for successive practical tests and evaluations.

6 Acknowledgements

The information in this document is provided as is, and no guarantee or warranty is given or implied that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability. The work described in this paper has been supported in part by the Federal Office for Information Security (BSI), Germany, in particular the development of the semantic classification levels, and partly by the EU-India project CultureTech, especially the influences of metadata. 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 views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Federal Office for Information Security (BSI), or the German Government, or the European Union.

References

1. A. Ross, A.K. Jain: Multimodal Biometrics: An Overview. Proceedings of the 12th Euro-pean Signal Processing Conference (EUSIPCO), Vienna, Austria, (2004) pp. 1221 - 1224
2. A.K. Jain, A. Ross: Multibiometric Systems, Communications of the ACM, Vol. 47, No. 1, (2004) pp. 34 - 40
3. C. Vielhauer, T. Basu, J. Dittmann, P.K. Dutta: Finding Meta Data in Speech and Handwriting Biometrics. In Proceedings of SPIE-IS&T. 5681, (2005) pp. 504-515

4. A. K. Jain, S. C. Dass and K. Nandakumar: Soft Biometric Traits for Personal Recognition Systems. In Proceedings of International Conference on Biometric Authentication (ICBA), LNCS 3072, Hong Kong, July (2004) pp. 731 - 738
5. A. K. Jain, S. C. Dass and K. Nandakumar: Can soft biometric traits assist user recognition? In Proceedings of SPIE Vol. 5404, Biometric Technology for Human Identification, Orlando, FL, April (2004) pp. 561 - 572
6. F. Wolf, T.K. Basu, P.K. Dutta, C. Vielhauer, A. Oermann, B. Yegnanarayana: A Cross-Cultural Evaluation Framework for Behavioral Biometric User Authentication.: From Data and Information Analysis to Knowledge Engineering. In: Proceedings of 29 Annual Conference of the Gesellschaft fr Klassifikation e. V., GfKl 2005, University of Magdeburg, Germany. Springer-Verlag, (2006) pp. 654-661
7. C. Vielhauer, T. Scheidat: Multimodal Biometrics for Voice and Handwriting. In: Jana Dittmann, Stefan Katzenbeisser, Andreas Uhl (Eds.), Communications and Multimedia Se-curity: 9th IFIP TC-6 TC-11International Conference, CMS 2005, Proceedings, LNCS 3677, Salzburg, Austria, September 19 - 21, (2005) pp. 191 - 199
8. C. Vielhauer, T. Scheidat: Fusion von biometrischen Verfahren zur Benutzerauthentifikation. In: P. Horster (Ed.), D-A-CH Security 2005 - Bestandsaufnahme, Konzepte, Anwendungen, Perspektiven (2005) pp. 82 - 97
9. A. Ross, R. Govindarajan: Feature Level Fusion Using Hand and Face Biometrics. In Proceedings of SPIE Conference on Biometric Technologie for Human Identification II, Vol. 5779, Orlando, USA (2005) pp. 196 - 204
10. : Y. Lee, K. Lee, H. Jee, Y. Gil, W. Choi, D. Ahn, S. Pan: Fusion for Multimodal Biometric Identification. In: Proceedings of Audio- and Video-Based Biometric Person Authentication: 5th International Conference, AVBPA 2005, Hilton Rye Town, NY, USA, July 20-22, 2005. Ed.: T. Kanade, A. Jain, N. K. Ratha, LNCS, Springer Heidelberg Berlin, Vol. 3546, (2005) p. 1071
11. H.R. Nielson, F. Nielson: Semantics with Applications: A Formal Introduction, revised edition, John Wiley & Sons, original 1992 (1999)
12. A. Oermann, A. Lang, J. Dittmann: Verifyer-Tupel for Audio-Forensic to Determine Speaker Environment. City University of New York (Organizer): Multimedia and security, MM & Sec'05, Proceedings ACM, Workshop New York, NY, USA (2005) pp. 57 - 62
13. A. Oermann, J. Dittmann, C. Vielhauer: Verifier-Tuple as a Classifier of Biometric Hand-writing Authentication - Combination of Syntax and Semantics. Dittmann, Jana (Ed.), Katzenbeisser, Stefan (Ed.), Uhl, Andreas (Ed.): Communications and multimedia security, Berlin: Springer, Lecture notes in com, CMS 2005 9th IFIP TC-6 TC-11 international conference Salzburg, Austria, (September 2005) pp. 170 - 179
14. K. Thibodeau: Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years. Council on Library and Information Resources: The State of Digi-tal Preservation: An International Perspective, Conference Proceedings, July 2002 http://www.clir.org/pubs/reports/pub107/thibodeau.html , last requested on 24.06.2006 (2002)
15. A. Oermann: Feature extraction of different media applying "Verfier-Tuple". http://wwwiti.cs.uni-magdeburg.de/ schauen/featextr_vt.htm , last requested on 24.06.2006 (2006)