Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 585139, 15 pages
http://dx.doi.org/10.1155/2014/585139

Research Article

Strategies for Exploiting Independent Cloud Implementations of Biometric Experts in Multibiometric Scenarios

P. Peer,[1] Ž. Emeršič,[1] J. Bule,[1] J. Žganec-Gros,[2] and V. Štruc[3]

[1] *Faculty of Computer and Information Science, University of Ljubljana, Tržaška cesta 25, 1000 Ljubljana, Slovenia*
[2] *Alpineon d.o.o., Ulica Iga Grudna 15, 1000 Ljubljana, Slovenia*
[3] *Faculty of Electrical Engineering, University of Ljubljana, Tržaška cesta 25, 1000 Ljubljana, Slovenia*

Correspondence should be addressed to V. Štruc; vitomir.struc@fe.uni-lj.si

Received 18 October 2013; Revised 8 January 2014; Accepted 22 January 2014; Published 13 March 2014

Academic Editor: Yue Wu

Cloud computing represents one of the fastest growing areas of technology and offers a new computing model for various applications and services. This model is particularly interesting for the area of biometric recognition, where scalability, processing power, and storage requirements are becoming a bigger and bigger issue with each new generation of recognition technology. Next to the availability of computing resources, another important aspect of cloud computing with respect to biometrics is accessibility. Since biometric cloud services are easily accessible, it is possible to combine different existing implementations and design new multibiometric services that next to almost unlimited resources also offer superior recognition performance and, consequently, ensure improved security to its client applications. Unfortunately, the literature on the best strategies of how to combine existing implementations of cloud-based biometric experts into a multibiometric service is virtually nonexistent. In this paper, we try to close this gap and evaluate different strategies for combining existing biometric experts into a multibiometric cloud service. We analyze the (fusion) strategies from different perspectives such as performance gains, training complexity, or resource consumption and present results and findings important to software developers and other researchers working in the areas of biometrics and cloud computing. The analysis is conducted based on two biometric cloud services, which are also presented in the paper.

1. Introduction

Biometric technology is slowly gaining ground and is making its way into our daily lives. This development is exemplified best by the last generation of smart-phones, which is starting to adopt fingerprint technology as means of improving security and is bringing biometrics closer to our minds than ever. While biometric technology for personal devices, such as notebooks and mobile phones, is slowly gaining traction, its broader use on the Internet is still quite modest. The main reason for this setting pertains mainly to open issues with respect to the accessibility and scalability of the existing biometric technology [1]. Scalability issues are also of relevance to other deployment domains of biometrics, such as forensics or law-enforcement, where biometric databases are expected to grow significantly over the next few years to accommodate several hundred millions (or even billions) of identities [2]. To meet these demands, it is necessary to develop scalable biometric technology, capable of operating on large amounts of data, and to ensure sufficient storage capacity and processing power [1].

A possible solution for the outlined issues is the development of biometric technology for the cloud, where the cloud platform ensures appropriate scalability, sufficient amount of storage, and parallel processing capabilities. With the widespread availability of mobile devices, the cloud also provides an accessible entry point for various applications and services relying on mobile clients [1]. The enormous potential of cloud-based biometric solutions was also identified by various companies which are currently developing or have only recently released their biometric cloud services to the market.

While a cloud platform can ensure the necessary infrastructure and resources for the next generation of biometric technology, the technology itself must ensure the best possible recognition (e.g., verification) performance. In this

respect, it is necessary to stress that biometric techniques relying on a single biometric trait (i.e., unimodal biometric experts) can only be improved to a certain extent in terms of performance. From a certain point forward, it may be either too costly or not yet feasible to further improve their performance. However, if performance is of paramount importance, the use of multibiometrics may represent a feasible solution.

The term multibiometrics refers to biometric technology that exploits several biometric experts and, hence, relies on several biometric traits of the same individual for identity inference. Multibiometric systems can offer substantial improvements in terms of accuracy and improvements in terms of flexibility and resistance to spoofing attacks. They also introduce higher tolerance to noise and data corruption and also reduce the failure-to-enroll rate [3].

In this paper, we address the problem of building (cloud-based) multibiometric systems based on existing implementations of unimodal biometric experts. Building (cloud-based) multibiometrics systems from existing implementations of biometric experts, instead of developing the system from scratch, has several advantages. The most obvious advantage is the reduction in effort needed to implement a multibiometric system. Furthermore, it is possible to choose the single expert systems from different vendors according to the desired specifications and performance capabilities. On the other hand, it is necessary to understand the process of combining the single expert systems from the perspectives of potential performance gains, additional resources needed, implementation complexity, and the like. Ideally, we would like to combine different (existing) biometric cloud services into a multibiometric service with significant performance gains and hence large improvements in security, but without the need for large modifications of existing client applications. Such multibiometric services would be of great interest to existing end-users of biometric cloud services and would exhibit significant market value.

To better understand the problem outlined above, we present in this paper an analysis of different (fusion) strategies for combining existing cloud-based biometric experts. To be as thorough as possible, we conduct the analysis under the assumption that only limited information, such as classification decisions or similarity scores, can be obtained from the existing cloud services (i.e., from the unimodal biometric experts). The fusion strategies are analyzed from different perspectives such as performance gains, training complexity, or resource consumption. The analysis is carried out based on two biometric cloud services developed in the scope of the KC CLASS project [1], the first being a face recognition cloud service and the second being a fingerprint recognition cloud service. The results of the analysis are important to engineers, software developers, and other researchers working in the areas of biometrics, cloud computing, and other related areas.

The rest of the paper is structured as follows. In Section 2, prior work in the area of multibiometrics is surveyed. In Section 3, the baseline (cloud-based) unimodal biometric experts are introduced. In Section 4, different strategies for combining the biometric experts are presented and their characteristics are discussed. In Section 5, a detailed analysis of all fusion strategies is presented and nonperformance related characteristics of the fusion strategies are also presented and elaborated on. The paper is concluded with some final comments and directions for future work in Section 6.

2. Related Work

The problem of combining unimodal biometric experts into multibiometric systems has been studied extensively in the literature (see, e.g., [3–9]). In general, the process of combining unimodal systems (usually referred to as fusion) can be conducted at the following [3].

(i) *The Signal or Sensor Level*. Sensor level fusion can benefit multisample systems which capture multiple snapshots of the same biometric. The process commonly referred to as mosaicing, for example, captures two or more impressions of the same biometric trait and creates an enhanced composite biometric sample that is better suited for recognition [3, 10].

(ii) *The Feature Level*. Fusion at the feature level involves integrating evidence of several biometric feature vectors of the same individual [3] obtained from multiple information sources. It is generally believed that fusion at this level ensures better recognition results than fusion at the later levels (i.e., the decision or matching score levels) as the features sets typically contain richer information about the raw biometric samples [3, 11].

(iii) *The Matching Score Level*. The matching scores still contain relatively rich information about the input biometric samples and it is also rather easy to combine matching scores of different experts. Consequently, information fusion at the matching score level is the most commonly used approach in multibiometric systems [3]. The matching score data from different biometric experts may not be homogeneous, may not be on the same numerical scale, or do not follow the same probability distribution [3]. These reasons make score level fusion a demanding problem.

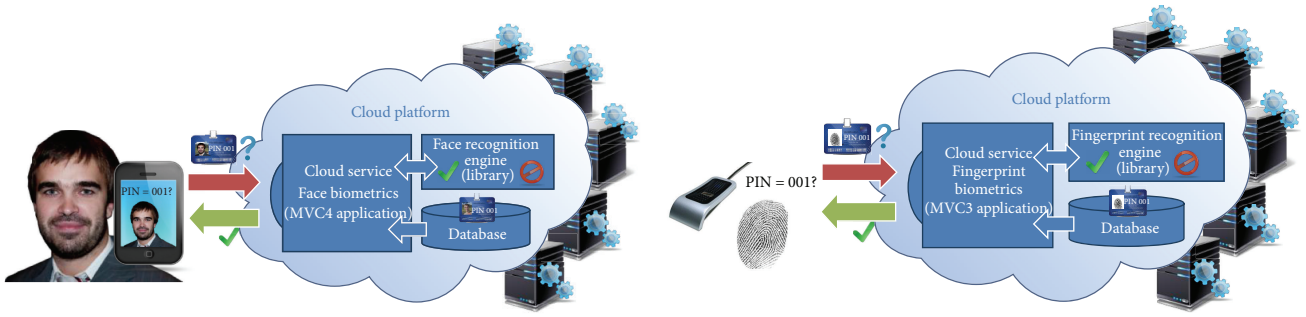
(iv) *The Decision Level*. This type of fusion is sensible when the unimodal biometric experts provide access only to the final stage in the process of biometric recognition, namely, the final classification result [3]. Different techniques can be considered at this level, for example, the AND- and OR-rules, majority voting, weighted majority voting, and others [3, 9, 12].

Among the different types of fusion techniques studied in the literature, fusion techniques applied at the matching score level are by far the most popular. This is also evidenced by Table 1, where a short overview of recent studies on biometric fusion is presented. Note that matching score level fusion techniques clearly dominate the research in this area.

When exploiting existing implementations of biometric experts, such as in our case, not all of the listed fusion levels are possible. Fusion at the first two levels requires data to be extracted right after the sample acquisition or the feature

Table 1: A few multibiometric systems discussed in the recent literature.

Author and year	Biometric modalities	Fusion level	Approach used
Nandakumar et al., 2008 [4]	Face, fingerprint, speech, iris	Matching score level	Likelihood ratio-based fusion
Maurer and Baker, 2008 [38]	Fingerprint, speech	Matching score level, quality-based fusion	Quality estimates via a Bayesian belief network (modified sum-rule)
Poh et al., 2009 [39]	Face, fingerprint, iris	Matching score level	Benchmarking 22 different biometric fusion algorithms
Lin and Yang, 2012 [40]	Face	Matching score level	Enhanced score-level fusion based on boosting
Tao and Veldhuis, 2009 [21]	Face (two face recognition algorithms)	Matching score level, decision level	Optimal fusion scheme at decision level by AND- or OR-rule (score levels: sum-rule, likelihood ratio, SVM)
Vatsa et al., 2010 [41]	Face (two face recognition algorithms)	Matching score level	Sequential fusion algorithm (likelihood ratio test + SVM)
Poh et al., 2010 [42]	Face, fingerprint	Matching score level	Quality-based score normalization
Poh et al., 2010 [43]	Face, fingerprint, iris	Matching score level	Addressing missing values in multimodal system with neutral point method
Nanni et al., 2011 [44]	Fingerprint, palm print, face	Matching score level	Likelihood ratio, SVM, AdaBoost of neural networks
Poh and Kittler, 2012 [45]	Face, fingerprint	Matching score level, quality-based fusion	A general Bayesian framework
Nagar et al., 2012 [46]	Face, fingerprint, iris	Feature level	Feature level fusion framework using biometric cryptosystems
Tao and Veldhuis, 2013 [47]	Face, speech	Matching score level	Native likelihood ratio via ROC



(a) Cloud implementation of face recognition service

(b) Cloud implementation of fingerprint recognition service

Figure 1: Illustration of basic architecture of the biometric cloud services.

extraction process, which is usually not possible, as existing (e.g., commercial) services commonly do not expose APIs for accessing the required data (i.e., signals or features). Existing cloud services typically only allow access to the decision and/or the matching score level. Hence, these two levels also form the basis for our assessment presented in the experimental section.

3. Baseline Systems

To be able to evaluate different strategies for combining independent implementations of biometric experts into a multibiometric cloud service, we first require access to unimodal biometric cloud services. In this section, we briefly introduce the basics of the unimodal face and fingerprint services that were used for the evaluation presented in the experimental section.

3.1. The Cloud Implementations. As we can see from Figure 1, both (i.e., face and fingerprint) services share a similar architecture, which is more or less a characteristic for biometric cloud services. Both services feature a background worker (i.e., typically implemented in the form of a programming library), which represents the recognition engine of the cloud service. The biometric database is implemented in the form of an SQL database, while the communication with potential clients of the services is conducted through a RESTful Interface. Note that both services are implemented and optimized for the task of biometric verification (and not biometric identification) and as such are capable of

returning either the verification result (i.e., the class label) or a matching score indicating the similarity between the input biometric sample and the template of the claimed identity. These characteristics are common to most biometric cloud-based verification systems and define the boundaries for possible strategies that may be explored for combining existing implementations of biometric cloud services.

3.2. The Face Recognition Engine. The core component of the face recognition service is the face recognition engine, which relies on Gabor-LBP features (LBP-Local Binary Patterns). Below, we briefly summarize the basics.

(i) *Face Detection, Localization, and Preprocessing*. Facial images are first preprocessed to remove illumination artifacts and then subjected to the Viola-Jones face detector to extract the facial region [13]. Next, facial landmark localization with PSEF correlation filters is performed to find anchor points in the faces that serve as the basis for geometrical normalization of the images [14]. The normalized images are rescaled to a fixed size of 128×128 pixels and finally subjected to the photometric normalization technique presented by Tan and Triggs in [15].

(ii) *Feature Extraction and Supporting Representation*. The main feature representation used by the face recognition service relies on Gabor magnitude responses and Local Binary Patterns (LBPs). Here, the normalized facial images are first filtered with a bank of 40 Gabor filters. The Gabor magnitude responses are then encoded with the LBP operator and local LBP histograms are computed from patches of all computed responses. The local histograms are ultimately concatenated into a global feature vector that forms the template for the given identity. To improve recognition performance, a vector of the first few DCT coefficients of the normalized facial image is also added to the template.

(iii) *Verification*. In the verification stage, the claim of identity is validated by comparing the template computed from the test/live/input image to the template of the claimed identity. Here, the Bhattacharyya distance is used to measure the similarity between the histograms of the LBP encoded Gabor magnitude responses and a simple whitened cosine similarity measure is used to match the DCT coefficients. Both similarity scores are then stacked together with image-quality measures (see [16] for details—Q-stack) and the newly combined feature vector is subjected to an AdaBoost classifier to obtain the final matching score based on which identity inference is conducted.

3.3. The Fingerprint Recognition Engine. The core component of the fingerprint recognition service is the minutiae-based fingerprint recognition engine first presented in [17]. Below, we briefly summarize the basics.

(i) *Segmentation and Image Enhancement*. Fingerprint scans are first subjected to a segmentation procedure, where the fingerprint pattern is separated from the background. Through the segmentation procedure, the processing time is shortened and the matching accuracy is increased. Since fingerprints are often degraded due to various external factors, the fingerprint patterns are enhanced by binarizing the captured fingerprint samples and ridge profiling [18].

(ii) *Minutiae Extraction*. The minutiae pattern is obtained from the binarized profiled image by thinning of the ridge structures, removal of structure imperfections from the thinned image, and the final process of minutiae extraction. For each detected minutia, its type (bifurcation or ending), spatial coordinates (x, y), and the orientation of the ridge containing the minutia are stored as the templates for each given identity [18].

(iii) *Matching and Verification*. Given a claimed identity and an input fingerprint sample, the claim of identity is validated by comparing the template computed from the test/live/input sample and the template corresponding to the claimed identity using a minutiae-based matching algorithm. Here, two fingerprints match when a sufficient number of minutiae match by type, location, and orientation.

4. Fusion Strategies

The task of combining different experts into a multiexpert system is common to many problems in the areas of pattern recognition and machine learning and is not restricted solely to the area of biometric recognition. Nevertheless, each problem has its specifics and it is important to understand the fusion task in the context of the specific problem one is trying to solve. The cloud implementations of the two biometric experts presented in the previous section were designed for the problem of biometric verification. We, therefore, commence this section by formalizing the problem of biometric verification and introducing the fusion task with respect to the presented formalization. In the second part of the section, we introduce different fusion strategies and elaborate on their characteristics.

4.1. Prerequisites. Let us assume that there are N identities registered in the given biometric system and that these identities are labeled with $\omega_1, \omega_2, \ldots, \omega_i, \ldots, \omega_N$ and that there are a total of J biometric experts at our disposal. Furthermore, let us assume that we are given a feature vector $\mathbf{x}^{(j)}$ of the jth expert, where $j \in \{1, 2, \ldots, J\}$ and a claimed identity ω_i from the pool of the N enrolled identities (in general, there could be a different number of identities enrolled in each of the J biometric experts, but for the sake of simplicity, we assume that this number (i.e., N) is the same for all experts). The aim of biometric verification is to assign the pair $(\omega_i, \mathbf{x}^{(j)})$ to class C_1 (a genuine/client claim) if the claim of identity is found to be genuine and to class C_2 (an illegitimate/impostor claim) otherwise. Commonly, the validity of the identity

claim is determined based on the so-called matching score $d^{(j)}$, which is generated by comparing the feature vector $\mathbf{x}^{(j)}$ to the template corresponding to the claimed identity ω_i [19, 20]; that is,

$$\left(\omega_i, \mathbf{x}^{(j)}\right) = \begin{cases} C_1, & \text{if, } d^{(j)} \leq \theta \text{ for } j \in \{1, 2, \ldots, J\} \\ C_2, & \text{otherwise,} \end{cases} \quad (1)$$

where θ stands for the decision threshold. Here, we assume that small matching scores correspond to large similarities and large matching scores correspond to small similarities.

In multibiometric systems, several (i.e., J) biometric experts are available for classifying the given pair $(\omega_i, \mathbf{x}^{(j)})$, where $i \in \{1, 2, \ldots, N\}$ and $j \in \{1, 2, \ldots, J\}$, with respect to (1). Thus, after the verification process, the following families of results are typically available:

$$\begin{aligned} \mathscr{C} &= \left\{C_k^{(j)} \mid j = 1, 2, \ldots, J; k \in \{1, 2\}\right\}, \\ \mathscr{D} &= \left\{d^{(j)} \mid j = 1, 2, \ldots, J\right\}, \end{aligned} \quad (2)$$

where $C_k^{(j)}$ denotes the classification result and $d^{(j)}$ is the matching score produced by the jth expert; $k \in \{1, 2\}$.

Applying different functions on the results of the verification procedure from (2) gives rise to different fusion procedures. Some of these procedures that also represent valid fusion strategies with respect to the two cloud implementations presented in one of the previous sections are presented in the remainder of the paper. Note also that we will assume that $J = 2$ from this point on, since we only have two cloud services at our disposal. All presented strategies are, however, easily extended for the case, where $J > 2$.

4.2. Decision-Level Fusion Rules. The first strategy for combining the verification results of two independent cloud implementations of biometric experts one may consider is to combine the results at the decision level. The decision level represents the most basic way of combining expert opinions of several biometric systems. The experts are simply queried for the classification result $C_k^{(j)}$ (for $j = 1, 2, \ldots, J$ and $k \in \{1, 2\}$) and the results are then combined into the final decision:

$$\psi : \left\{C_k^{(1)}, C_k^{(2)}, \ldots, C_k^{(J)}\right\} \longrightarrow C_k^{\text{fused}}, \quad \text{where } k \in \{1, 2\}, \quad (3)$$

where C_k^{fused} is the combined classification result and $k \in \{1, 2\}$.

Several options are in general available for choosing the fusion function ψ, but two of the most common are the AND- and OR-rules [21]. In the context of the two cloud-based biometric experts at our disposal, the two rules, which assume that the class labels C_1 and C_2 are binary encoded, that is, $C_1 = 1$ and $C_2 = 0$, are defined as

$$\begin{aligned} \psi_{\text{AND}}\left(C_k^{(1)}, C_k^{(2)}\right) &= C_k^{\text{fused}} = C_k^{(1)} \,\&\, C_k^{(2)}, \\ \psi_{\text{OR}}\left(C_k^{(1)}, C_k^{(2)}\right) &= C_k^{\text{fused}} = C_k^{(1)} \mid C_k^{(2)}, \end{aligned} \quad (4)$$

where & and | denote the logical AND and OR operators and the superscript indices $^{(1)}$ and $^{(2)}$ stand for the face and fingerprint experts, respectively, and $k \in \{1, 2\}$.

While the decision level fusion strategies are easy to implement and offer a straightforward way of consolidating experts opinions, potential client applications relying on these strategies do not possess the flexibility of freely choosing the operating point of the combined biometric system. Instead, the operating point is determined by the operating points of the single expert systems.

4.3. Matching-Score-Level Fusion Rules. The second strategy that can be exploited to combine the verification results of several biometric experts is fusion at the matching score level using fixed fusion rules [5]. Most cloud-based biometric services (including our two) can be queried for a similarity score rather than the final classification decision. The client application then implements the classification procedure (see (1)) using a desired value for the decision threshold θ. Such an operating mode is implemented in most biometric services as it gives the client applications the possibility of choosing their own operating points and, hence, selecting a trade-off between security and user-convenience.

The general form for consolidating several expert opinions at the matching score level is

$$\phi : \left\{d^{(1)}, d^{(2)}, \ldots, d^{(J)}\right\} \longrightarrow d^{\text{fused}}, \quad (5)$$

where ϕ is the fusion function and $d^{\text{fused}} \in \mathbb{R}$ represents the combined matching score that can be exploited for the final identity inference using (1). Note that the decision threshold for the fused scores needs to be recalculated for all desired operating points and cannot be found in the specifications of the cloud services anymore.

For our assessment presented in the experimental section, we implemented two fixed matching-level fusion rules, namely, the weighted sum-rule and the weighted product-rule. The two rules are defined as follows:

$$\phi_{\text{SUM}}\left(d^{(1)}, d^{(2)}\right) = d^{\text{fused}} = w d^{(1)} + (1 - w)\, d^{(2)}, \quad (6)$$

$$\phi_{\text{PRO}}\left(d^{(1)}, d^{(2)}\right) = d^{\text{fused}} = \left(d^{(1)}\right)^{w} \cdot \left(d^{(2)}\right)^{(1-w)}, \quad (7)$$

where the superscript indices $^{(1)}$ and $^{(2)}$ again denote the face and fingerprint experts, respectively, d^{fused} represents the combined matching score, and the real-valued $w \in [0, 1]$ stands for the weighting factor balancing the relative importance of the face and fingerprint scores. Note here that the weighted product-rule in (7) could also be represented as a weighted log-sum fusion rule making it very similar to the weighted sum-rule in (6). However, as shown in [9], the two fusion rules are based on different assumptions. The interested reader is referred to [9] for more information on this topic.

It should be emphasized here that the matching scores of independent biometric systems are typically of a heterogeneous nature—they are not on the same numerical

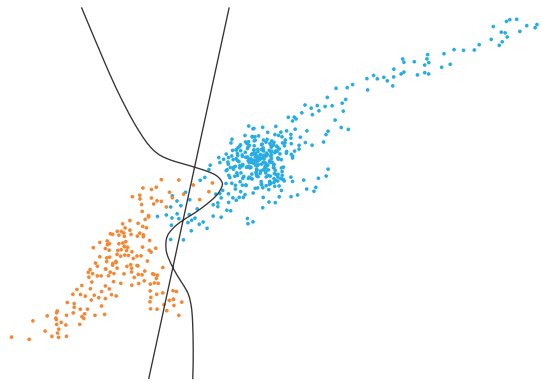


Figure 2: The linear and nonlinear decision boundaries are displayed, dividing genuine (in blue) and impostor (in red) classes.

range. Score normalization is, therefore, used to transform the scores to a common range prior to combining them [22, 23]. In this work, min-max normalization is used as it is quite simple and typically gives satisfactory results. Min-max normalization transforms all the scores to a common range of [0, 1].

4.4. Fusion with Classifiers. The third strategy one may consider when combining biometric experts (again at the matching score level) is to use pattern classifiers. Similar to the fixed fusion rules presented in the previous section, it is first necessary to obtain similarity scores from the cloud services rather than classification labels. Rather than combining the scores to a single scalar value using fixed rules, the matching scores are concatenated into "new" feature vectors which are then classified into one of two classes: "genuine" or "impostor" (i.e., classes C_1 and C_2 in (1)) [3]. In this setting, the classifier is actually used to indirectly learn the relationship between the vector of matching scores provided by the biometric experts and the a posteriori probabilities of the genuine and the impostor classes [3]. Once trained, the discriminant function associated with the given classifier can be used to produce combined matching scores.

The described procedure can be formalized as follows:

$$\xi : \left\{d^{(1)}, d^{(2)}, \ldots, d^{(J)}\right\} \longrightarrow d^{\text{fused}} = \delta\left(\mathbf{x}'\right), \tag{8}$$

where $\mathbf{x}' = [d^{(1)}, d^{(2)}, \ldots, d^{(J)}]^T$ denotes the new feature vector and $\delta(\cdot)$ stands for the discriminant function of the given classifier.

The classifier learns a decision boundary between the two classes, which can be either linear or nonlinear, depending on the choice of classifier. In Figure 2, where a toy example is presented, the impostor class is represented with red color and the genuine class with blue. The straight line represents a linear decision boundary between the two classes, whereas the curved line represents a nonlinear boundary. Note that during verification, any new matching score vector is classified into the genuine/impostor class depending on which side of the decision boundary it falls. Thus, most existing implementations of the most common classifiers return the class label instead of the output of their discriminant functions. However, with a little bit of tweaking, most existing implementations can be altered to return the output of the discriminant function as well.

Different from the fixed fusion rules, classifiers are capable of learning the decision boundary irrespective of how the feature vectors are generated. Hence, the output scores of the different experts can be nonhomogeneous (distance or similarity metric, different numerical ranges, etc.) and no processing is required (in theory) prior to training the classifier [3].

In the experimental section, we assess the relative usefulness of the classifier-based fusion strategy based on two classifiers: a Support Vector Machine (SVM) (with a linear kernel) [24, 25] and a Multilayer Perceptron (MLP) [26] classifier. The former falls into the group of linear classifiers (in our case), while the latter represents an example of a nonlinear classifier.

5. Experiments

5.1. Database, Protocol, and Performance Measures. To evaluate the different fusion strategies, a bimodal chimeric database is constructed from the XM2VTS and FVC2002 databases [27, 28]. A chimeric database represents a database in which biometric modalities from different databases are combined and assigned common identities. Since the biometric samples in the initial databases are not taken from the same identities, this procedure creates artificial (chimeric) subjects. Note that such a procedure is reasonable due to the fact that biometric modalities are generally considered to be independent one from another (e.g., a facial image says nothing about the fingerprint of the subject and vice versa) [29]. The constructed chimeric database consists of facial imagery and fingerprint data of 200 subjects with each subject having a total of 8 biometric samples for each modality. A few sample images from the chimeric database are presented in Figure 3.

For the experiments, the data is divided into two disjoint parts of 100 subjects (with 8 biometric samples per each modality). The first part is used for learning open hyperparameters of the fusion procedures (e.g., fusion weights, decision thresholds, etc.), while the the second is used for evaluating the fusion techniques on unseen testing data with fixed hyperparameters. Each of the experimental runs consists of enrolling each of the 800 biometric samples (i.e., face and fingerprint samples) from the given part into the corresponding (biometric) cloud service and matching the same 800 samples against all enrolled samples. This experimental setup results in 640, 000 matching scores (800× 800) for the training and testing parts, out of which 6400 correspond to genuine verification attempts and 633600 correspond to illegitimate verification attempts. Note that prior to the experiments the matching scores are normalized using min-max score normalization [23].

For evaluation purposes, standard performance measures typically used in conjunction with two-class recognition problems are adopted, namely, the false acceptance error rate

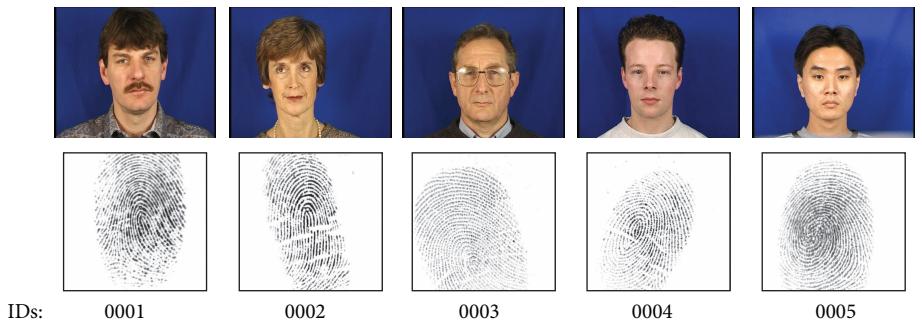


Figure 3: Sample images from the constructed chimeric database.

(FAR) and the false rejection error rate (FRR). The two error rates are defined as [30–33]

$$\mathrm{FAR}(\theta) = \frac{\left|\left\{d_{\mathrm{imp}} \mid d_{\mathrm{imp}} \leq \theta\right\}\right|}{\left|\left\{d_{\mathrm{imp}}\right\}\right|}, \qquad \mathrm{FRR}(\theta) = \frac{\left|\left\{d_{\mathrm{cli}} \mid d_{\mathrm{cli}} > \theta\right\}\right|}{\left|\left\{d_{\mathrm{cli}}\right\}\right|}, \tag{9}$$

where $\{d_{\mathrm{cli}}\}$ and $\{d_{\mathrm{imp}}\}$ represent sets of client and impostor scores generated during the experiments, $|\cdot|$ denotes a cardinality measure, and θ represents the decision threshold, and the inequalities assume that dissimilarity measures were used to produce the matching scores (it is assumed that large similarities between biometric samples result in small values of the matching scores and vice versa).

Note that both error rates, FAR and FRR, represent functions of the decision threshold θ. Selecting different values of the decision threshold, therefore, results in different error rates that form the basis for various performance metrics. In this paper, three such metrics are used, namely, the equal error rate (EER), which is defined with the decision threshold that ensures equal values of the FAR and FRR on the training set, that is,

$$\mathrm{EER} = \frac{1}{2}\left(\mathrm{FAR}(\theta_{\mathrm{eer}}) + \mathrm{FRR}(\theta_{\mathrm{eer}})\right), \tag{10}$$

where

$$\theta_{\mathrm{eer}} = \underset{\theta}{\operatorname{argmin}} \left|\mathrm{FAR}(\theta) - \mathrm{FRR}(\theta)\right|, \tag{11}$$

the verification rate at the false acceptance error rate of 0.1% (VER@0.1FAR), which is defined as

$$\mathrm{VER@0.1FAR} = 1 - \mathrm{FRR}(\theta_{\mathrm{ver01}}), \tag{12}$$

where

$$\theta_{\mathrm{ver01}} = \underset{\theta}{\operatorname{argmin}} \left|\mathrm{FAR}(\theta) - 0.001\right|, \tag{13}$$

and the verification rate at the false acceptance error rate of 0.01% (VER@0.01FAR):

$$\mathrm{VER@0.01FAR} = 1 - \mathrm{FRR}(\theta_{\mathrm{ver001}}), \tag{14}$$

where

$$\theta_{\mathrm{ver001}} = \underset{\theta}{\operatorname{argmin}} \left|\mathrm{FAR}(\theta) - 0.0001\right|. \tag{15}$$

The presented performance metrics are typically computed based on client and impostor score populations generated on the training data. To obtain an estimate of the generalization capabilities of a given fusion technique on unseen data, the thresholds θ_{eer}, θ_{ver01}, and θ_{ver001} are applied to client and impostor score populations generated on the evaluation data. Thus, during test time, the FAR and FRR defined in (9) are computed based on the fixed thresholds and then combined into the half-total error rate (HTER) as follows:

$$\mathrm{HTER}(\theta_k) = \frac{1}{2}\left(\mathrm{FAR}_e(\theta_k) + \mathrm{FRR}_e(\theta_k)\right), \tag{16}$$

where $k \in \{\mathrm{eer}, \mathrm{ver01}, \mathrm{ver001}\}$ and the subscript index $_e$ indicates that the error rates FAR and FRR were computed on the evaluation set. Alternatively, it is also possible to evaluate the verification rate and the false acceptance error rate at a specific decision threshold set during training; that is,

$$\mathrm{VER}_e(\theta_k) = 1 - \mathrm{FRR}_e(\theta_k), \quad \text{with } \mathrm{FAR}_e(\theta_k), \tag{17}$$

where, in our case, k again stands for $k \in \{\mathrm{eer}, \mathrm{ver01}, \mathrm{ver001}\}$.

In addition to the quantitative performance metrics, performance curves are also used to present the results of the experiments. Specifically, Receiver Operating Characteristic (ROC) Curves and Expected Performance Curves (EPC) are generated during the experiments to better highlight the differences among the assessed techniques [34]. ROC curves plot the dependency of the verification rate (VER) and the false acceptance rate (FAR) with respect to varying values of the decision threshold θ. ROC curves are usually plotted using a linear scale on the x- and y-axes; however, to better highlight the difference among the assessed procedures at the lower values of the false acceptance rate, a log scale for the x-axis is used in this paper.

To generate EPC, two separate image sets are needed. The first image set, that is, the training set, is used to find a decision threshold that minimizes the following weighted error (WER) for different values of α:

$$\mathrm{WER}(\theta, \alpha) = \alpha\mathrm{FAR}(\theta) + (1 - \alpha)\,\mathrm{FRR}(\theta), \tag{18}$$

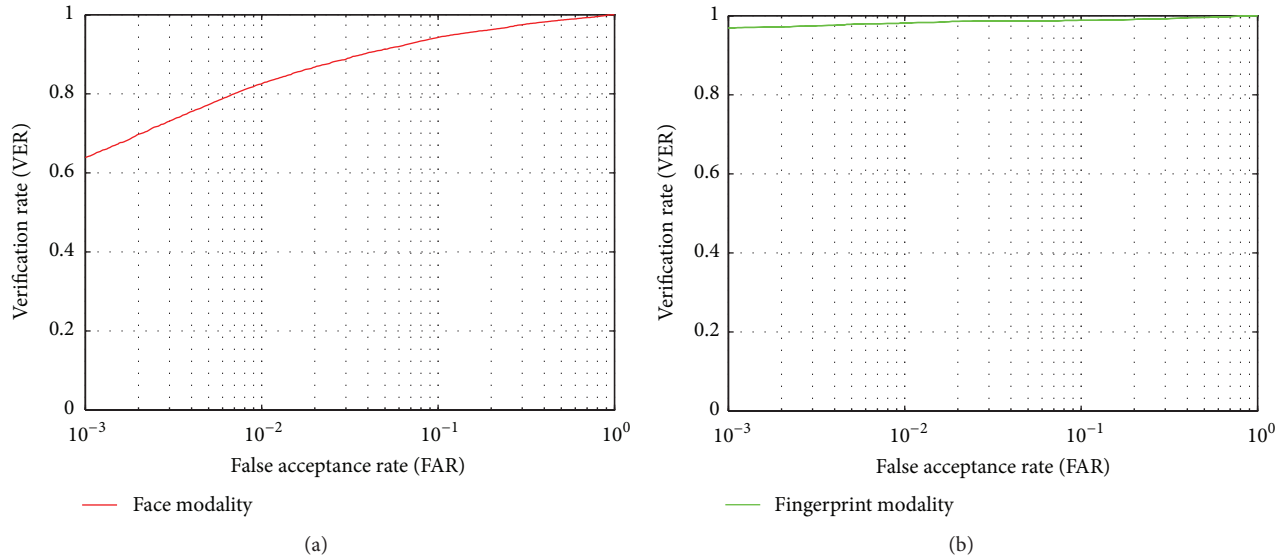


Figure 4: ROC curves of the experiments: face recognition (a) and fingerprint recognition (b).

Table 2: Quantitative comparison of the biometric modalities.

Procedure	EER	VER@0.1FAR	VER@0.01FAR
Face	0.0720	0.6394	0.3748
Fingerprint	0.0163	0.9691	0.9556

where α denotes a weighting factor that controls the relative importance of the FAR and FRR in the above expression. Next, the second image set, that is, the testing/evaluation image set, is employed to estimate the value of the HTER at the given α and with the estimated value of the decision threshold θ. When plotting the HTER (obtained on the testing/evaluation image sets) against different values of the weighting factor α, an example of an EPC is generated [35].

5.2. Single Expert Assessment. Before evaluating the feasibility and efficiency of different strategies for combining the cloud implementations of the biometric experts, it is necessary to establish the baseline performance of the unimodal biometric systems, that is, the face and fingerprint systems. To this end, the training data (note that the data used during this series of experiments represents the training data for the fusion techniques; for the unimodal systems, data is still valid testing/evaluation data) from our chimeric database is used to generate all of the relevant performance metrics and performance curves introduced in the previous section. The results of the experiments are presented in the form of ROC curves in Figure 4 and with quantitative performance metrics in Table 2.

As expected, the fingerprint recognition system performs much better than the face recognition system, especially at the lower values of the false acceptance error rates. At the equal error rate, for which the cloud-based biometric experts were also optimized, the face modality results in an error of around 7%, while the fingerprint modality ensures an error rate of a little more than 1.5%.

It is interesting to look at the distribution of the client and impostor similarity scores of the single experts in the *fingerprint—versus face-score*—space; see Figure 5. Since the optimal decision boundary appears to be different from a horizontal or vertical line (this would correspond to conducting identity inference based on one of the biometric experts), performance gains (at least on the matching score level) can be expected by combining the two experts. Different strategies for doing so are evaluated in the remainder.

5.3. Assessing Decision-Level Strategies. One possible strategy for combining the outputs of the cloud implementations of the face and fingerprint recognition experts is to consolidate the opinions of the two experts at the decision level. In this setting, the cloud services are asked to make a decision regarding the validity of the identity claim made with the given biometric sample. Since no similarity scores are sent to the client application, the operating point (i.e., the ratio of the FAR and FRR) of the cloud recognition service cannot be changed and is determined by the settings on the service side. In our case, the operating point of the cloud services is set to the equal error rate (EER).

Two decision-level fusion schemes are implemented for the experiments, namely, the AND- and OR-rules, as described in Section 4. The results of the experiments (on the training data) are shown in Table 3 in the form of various performance metrics. Note that it is not possible to generate ROC curves for this series of experiments, since no similarity scores are available.

Several observations can be made based on the presented results. Both fusion strategies result in similar performance in terms of the HTER with the difference that the AND-rule favors small FARs, while the OR-rule favors small FRRs. When compared to the performance of the single

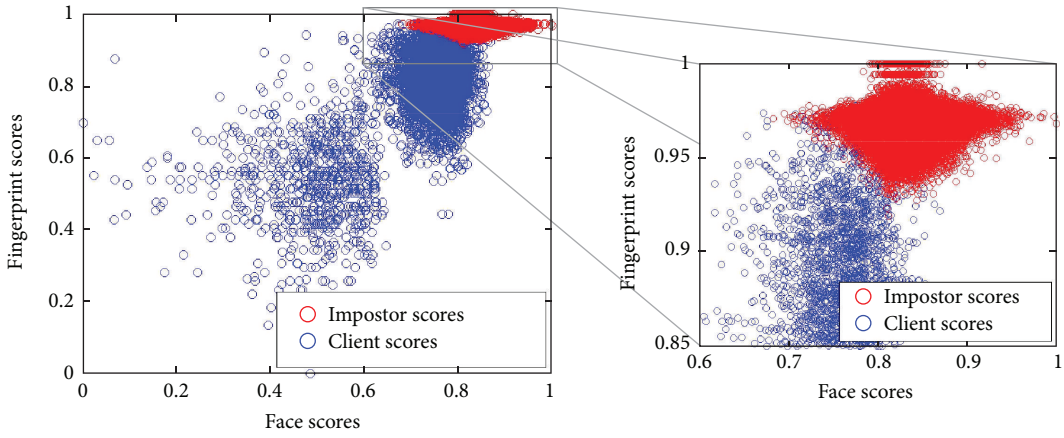


Figure 5: Plot of scores in the face-fingerprint-score plane.

Table 3: Quantitative comparison of the decision-level fusion rules (training data).

Procedure	HTER	FAR	FRR
AND-rule	0.0440	0.0011	0.0869
OR-rule	0.0440	0.0862	0.0014

expert systems presented in Table 2, the decision-level fusion schemes performed better than the face expert but a little worse than the fingerprint expert. All in all, the decision-level fusion rules did not prove to be too useful for improving the recognition performance of the single expert systems but should rather be considered as a way of changing the operating point of the combined system toward lower FARs or FRRs in cases where only decision-level outputs can be obtained from cloud implementations of biometric experts.

5.4. Assessing Matching-Score-Level Strategies. When assessing matching-score-level strategies for combining the two biometric experts, we first focus on the fixed fusion rules and present experiments related to classifier fusion strategies in the second part of this section.

The performance of the sum and product fusion rules is first examined on the training part of the constructed chimeric database. Before reporting the final performance, it is necessary to find (or learn) appropriate values for the open hyperparameter w of the sum and product fusion rules ((6) and (7)). To this end, the value of w is gradually increased from 0 to 1 with a step size of 0.1 and the values of EER, VER@0.1FAR, and VER@0.01FAR are computed for each value of w. The results of this series of experiments are shown in Figure 6. Note that both the sum and product fusion rules peak in their performance at a value of $w = 0.3$. This value is, therefore, selected for both fusion rules for all the following experiments.

To compare the performance of the sum and product fusion rules with fixed hyperparameters to that of the single expert systems, we generate ROC curves from the scores obtained on the training part of the chimeric database. The performance curves are shown in Figure 7 and the corresponding performance measures in Table 4. Note that the sum and product fusion rules perform very similarly; both are significantly better than the unimodal (single expert) systems. The EER, for example, falls by more than 50% with both fusion rules when compared to the better performing single expert system. While these results are encouraging, it needs to be taken into account that the ROC curves for the fusion techniques shown in Figure 7 were generated by optimizing the open hyperparameter w on the same data that was used for constructing the curves in the first place. This means that the performance of the fusion techniques may be somewhat biased. To analyze this issue, we present comparative experiments on the evaluation/testing data of the constructed chimeric database in Section 5.5.

Next to fixed fusion rules, the second possibility for combining the similarity scores of the single expert systems is to stack the similarity scores into a two-dimensional feature vector and use the constructed vector with a classifier. To evaluate this possibility, the training part of the chimeric database is used to train SVM (Support Vector Machine [24, 25]) and MLP (Multilayer Perceptron [26]) classifiers. For the SVM classifier, a linear kernel is selected, and for the MLP classifier, an architecture with two hidden layers (each with 5 neurons) is chosen. This setting results in a classifier capable of finding linear decision boundary between the client and impostor classes (i.e., the SVM) and a classifier capable of setting the decision boundary in a nonlinear manner (i.e., the MLP). Once trained, both classifiers are applied to the training data to compute performance metrics and construct performance curves. The results of this series of experiments are shown in Figure 8 and Table 5. Note that with most existing software solutions for training SVM and MLP classifiers (see, e.g., [36, 37]) a little of tweaking is needed to obtain the similarity scores required for constructing ROC curves, as the classifiers usually output only class labels. When looking at the performance of the SVM and MLP classifiers, it is obvious that they did not ensure any additional performance improvements when compared to the fixed fusion rules. While this could be expected for the linear SVM classifier, it is somehow unexpected that the MLP classifier did not improve the performance over

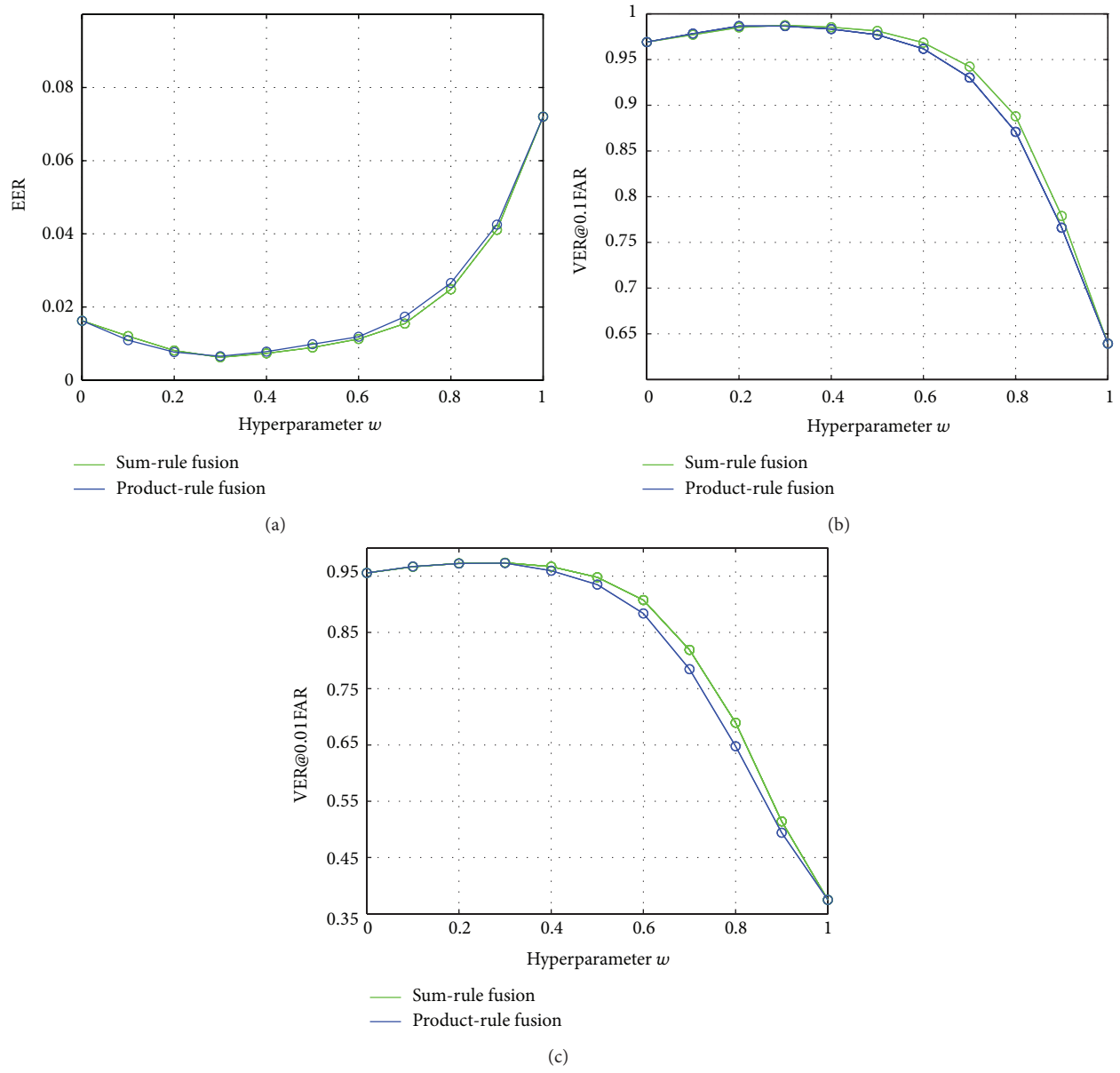


Figure 6: EERs, VER@0.1FARs, and VER@0.01FARs for the sum and product fusion rules for different values of the hyperparameter w.

the fixed sum and product fusion rules. It seems that, without any additional information such as quality or confidence measures, it is extremely difficult to significantly improve upon the performance fixed fusion rules on our chimeric database.

5.5. Generalization Capabilities of Different Strategies. The last series of verification experiments aimed at examining the generalization capabilities of the fusion strategies on the evaluation/testing part of the chimeric database. Here, all decision thresholds and hyperparameters of all assessed fusion strategies are fixed on the training part of the data. The testing/evaluation data is then used to generate performance metrics and curves, which are shown in Figure 9 and Table 6 for this series of experiments.

The first thing to notice from the presented results is the fact that, similar to the training data, all fusion strategies (except for the decision-level fusion techniques) result in performance improvements when compared to either of the two single expert systems. Next, the performance achieved on the training data is also achieved on the evaluation/testing data, which suggests that no overfitting took place during training. Especially important here is also the fact that all results are very well calibrated indicating that a desired operating point (i.e., the ratio between the FAR and FRR) can easily be selected and maintained even after the fusion process.

Last but not least, it is also important to stress that, among the matching-score-level fusion strategies, no particular strategy has a clear advantage in terms of performance on

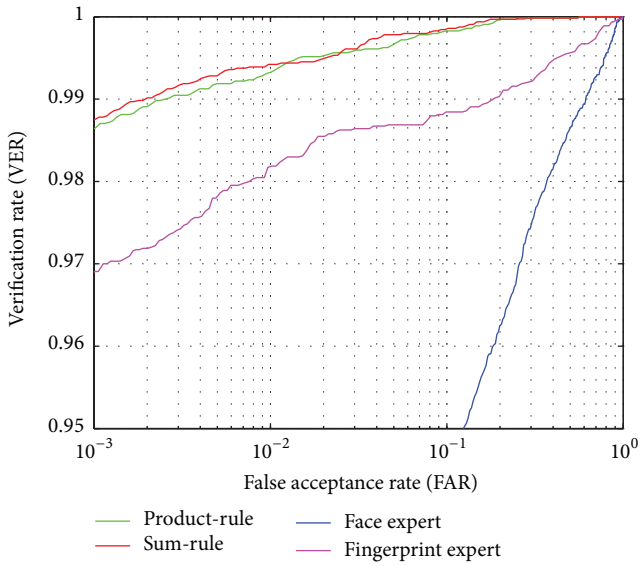


Figure 7: ROC curves for the fusion rules (training data).

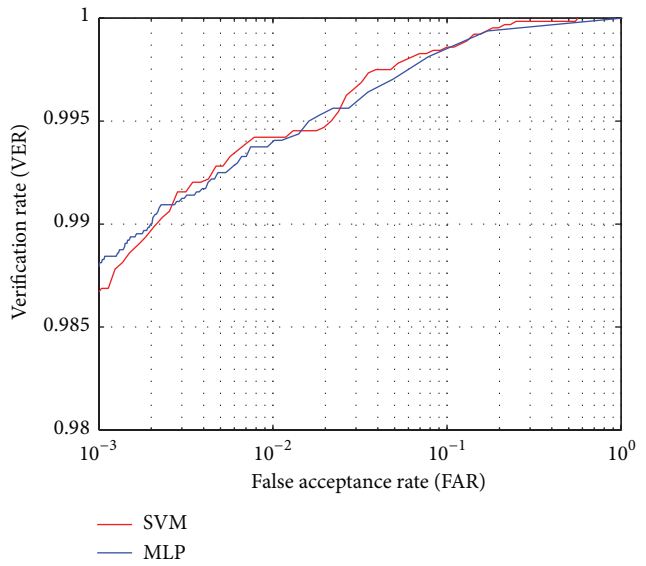


Figure 8: ROC curves for fusion techniques with classifiers (training data).

Table 4: Quantitative comparison of the fusion rules with learned parameter w ($w = 0.3$ for both techniques)—training data.

Procedure	EER	VER@0.1FAR	VER@0.01FAR
Product-rule	0.0066	0.9866	0.9733
Sum-rule	0.0063	0.9875	0.9736

Table 5: Quantitative comparison of fusion techniques with classifiers (training data).

Procedure	EER	VER@0.1FAR	VER@0.01FAR
SVM	0.0064	0.9869	0.9733
MLP	0.0063	0.9881	0.9731

our chimeric database. This suggests that other criteria next to performance should be taken into account when selecting the best strategy for combining different cloud-based biometric experts.

5.6. Subjective Analysis. In the previous sections, different fusion strategies for combining cloud implementations of biometric experts were assessed only from the perspective of performance. However, when combining different biometric experts into a multibiometric system, other criteria are important as well. One may, for example, be interested in how existing client applications need to be modified to enable multiexpert biometric recognition, how difficult it is to reach a specific operating point in the multibiometric system, whether additional libraries need to be included in the client applications, and so forth. To evaluate the fusion strategies based on other (nonperformance related) criteria as well, a grade (low—L, medium—M, or high—H) was assigned to each strategy in seven different categories (note that these grades are of a subjective nature and reflect the perception of the authors). These categories include the following.

(i) Training complexity: the complexity of the training procedure for the given fusion strategy (e.g., training the classifier, setting hyperparameters, etc.).

(ii) Run-time complexity: the run-time complexity required to apply the given fusion strategy to the data (e.g., applying the trained classifier to the data, etc.).

(iii) Storage requirements: memory requirements for storing metadata needed by the given fusion strategy (e.g., for storing support vectors, hyperparameters, etc.).

(iv) Performance gains: performance related criterion that reflects the results of the experiments conducted in the previous sections.

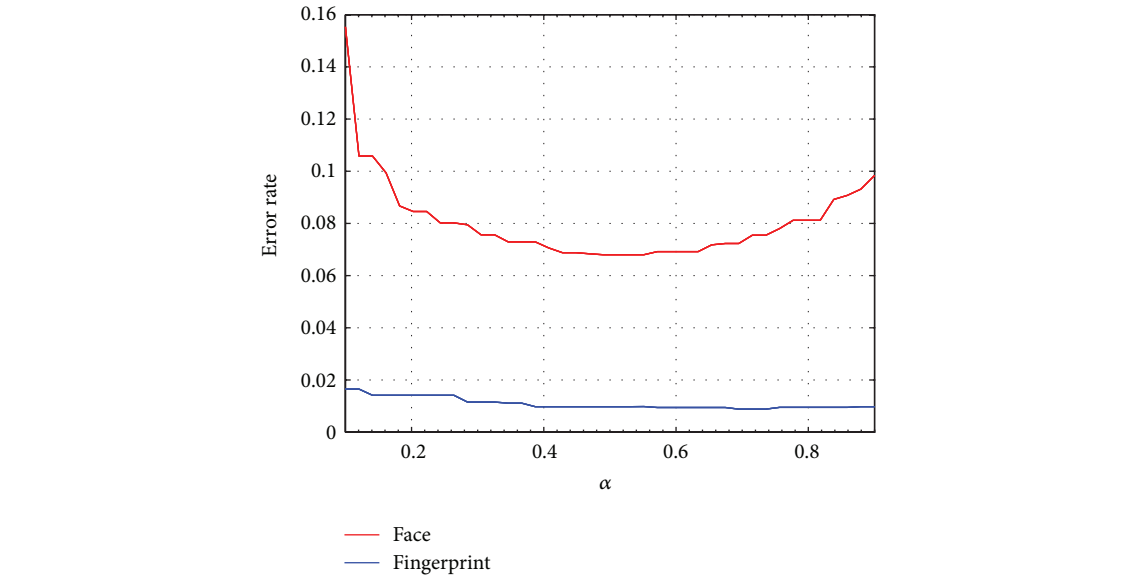
(v) Client disturbance: relating to the amount of work needed to rewrite existing client applications and the need for including additional external resources (e.g., programming libraries, etc.).

(vi) Calibration: referring to the generalization capabilities of the given fusion strategy and the capability of ensuring the same operating point across different data sets.

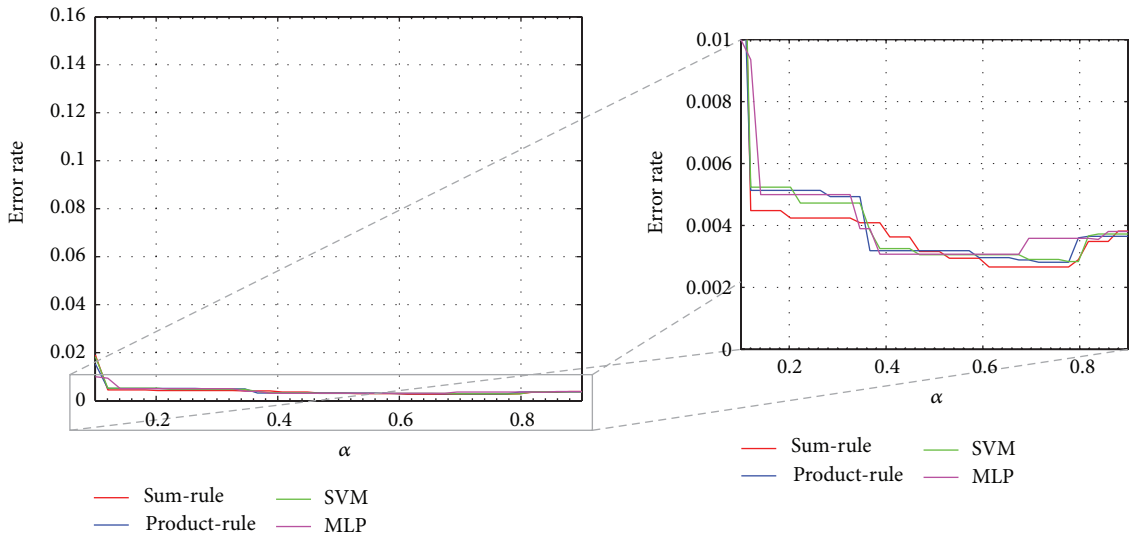
(vii) OP complexity: relating to the complexity of setting a specific operating point for the multibiometric system (e.g., the EER operating point, etc.).

Our ratings are presented in Table 7 in the form of grades and in Figure 10 in the form of Kiviat graphs. With the generated graphs, a larger area represents a more suitable fusion strategy according to the selected criteria (note that the same weight has been given here to all criteria. If a certain criterion is considered more important than others, this could be reflected in the final Kiviat graphs).

Note that the fixed fusion rules (i.e., the sum- and product-rules) turned out to be suited best for combining



(a) EPC-curves for the unimodal experts



(b) EPC-curves for the fusion techniques

Figure 9: EPC-curves of the experiments on the evaluation data.

Table 6: Quantitative comparison of the fusion rules on the evaluation/testing data. Here, the symbol n/a stands for the fact that this value is not computable.

	$@\theta_{\text{eer}}$			$@\theta_{\text{ver01}}$			$@\theta_{\text{ver001}}$		
	HTER_e	VER_e	FAR_e	HTER_e	VER_e	FAR_e	HTER_e	VER_e	FAR_e
Face	0.0716	0.9280	0.0712	0.1808	0.6394	0.0010	0.3126	0.3748	9.95E − 5
Fingerprint	0.0133	0.9897	0.0162	0.0096	0.9819	0.0012	0.0129	0.9744	1.78E − 4
Sum-rule	0.0045	0.9973	0.0064	0.0034	0.9944	0.0011	0.0061	0.9880	1.14E − 4
Product-rule	0.0046	0.9972	0.0065	0.0033	0.9945	0.0011	0.0063	0.9875	1.25E − 4
AND-rule	0.0411	0.0811	0.0012	n/a	n/a	n/a	n/a	n/a	n/a
OR-rule	0.0411	0.0013	0.0863	n/a	n/a	n/a	n/a	n/a	n/a
SVM	0.0046	0.9972	0.0063	0.0032	0.9947	0.0011	0.0060	0.9881	1.36E − 4
MLP	0.0046	0.9973	0.0066	0.0036	0.9939	0.0012	0.0062	0.9877	1.10E − 4

Table 7: Comparison of the fusion strategies based on the perception of the authors and conducted experimentation. High, medium, and low are denoted by H, M, and L, respectively.

Fusion technique	Training complexity	Run-time complexity	Storage requirements	Performance gains	Client disturbance	Calibration	OP complexity
Sum-rule	L	L	L	M	L	H	L
Product-rule	L	L	L	M	L	H	L
AND-rule	L	L	L	L	L	L	H
OR-rule	L	L	L	L	L	L	H
SVM	H	M	M	M	M	H	M
Neural network	M	L	L	M	M	H	L

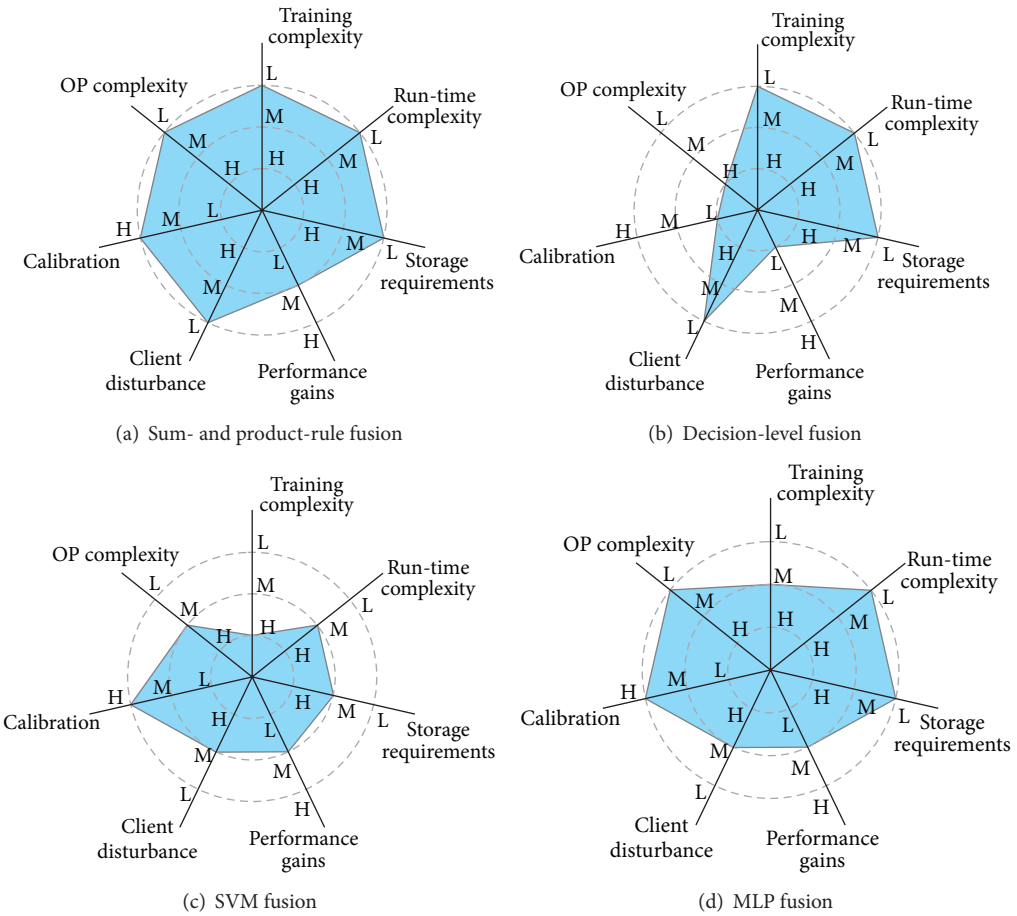


Figure 10: Kiviat graphs of the fusion techniques generated based on the selected evaluation criteria.

different cloud implementations of biometric experts into a multibiometric system as they provide a good trade-off between the complexity of the training and run-time procedures, expected performance gains, flexibility in setting the operating point, calibration, and the need for modifying existing client applications.

6. Conclusion

We have presented an analysis of different strategies for combining independent cloud implementations of biometric experts into a multibiometric recognition system. For the analysis, we used our own implementations of cloud-based face and fingerprint verification services and a specially constructed chimeric database. The results of our analysis suggest that fixed fusion rules that combine the single expert systems at the matching score level are the most suitable for the studied task as they provide a good trade-off between expected performance gains and other important factors such as training complexity, run-time complexity, calibration, and client disturbance. As part of our future work, we plan to examine possibilities of including confidence measures in

the fusion strategies, as these have the potential to further improve the recognition performance of the combined multi-biometric system. We also plan to develop biometric cloud services combining more than two single expert systems.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The work presented in this paper was supported in part by The National Research Program P2-0250(C) Metrology and Biometric Systems, by the National Research Program P2-0214 Computer Vision, and by the EU, European Regional Fund, in scope of the framework of the Operational Programme for Strengthening Regional Development Potentials for the Period 2007–2013, Contract no. 3211-10-000467 (KC Class), the European Union's Seventh Framework Programme (FP7-SEC-2011.20.6) under Grant agreement no. 285582 (RESPECT), and the European Union's Seventh Framework Programme (FP7-SEC-2010-1) under Grant agreement no. 261727 (SMART). The authors additionally appreciate the support of COST Actions, IC1106 and IC1206.

References

[1] P. Peer, J. Bule, J. Zganec Gros, and V. Štruc, "Building cloud-based biometric services," *Informatica*, vol. 37, no. 1, pp. 115–122, 2013.

[2] E. Kohlwey, A. Sussman, J. Trost, and A. Maurer, "Leveraging the cloud for big data biometrics: meeting the performance requirements of the next generation biometric systems," in *Proceedings of the IEEE World Congress on Services (SERVICES '11)*, vol. 1, pp. 597–601, July 2011.

[3] A. Ross, K. Nandakumar, and A. Jain, *Handbook of Multibiometrics*, Springer Science+Business Media, LLC, New York, NY, USA, 2006.

[4] K. Nandakumar, Y. Chen, S. C. Dass, and A. K. Jain, "Likelihood ratio-based biometric score fusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 342–347, 2008.

[5] L. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*, Wiley-Interscience, Hoboken, NJ, USA, 2004.

[6] A. B. Khalifa and N. B. Amara, "Bimodal biometric verification with different fusion levels," in *Proceedings of the 6th International Multi-Conference on Systems, Signals and Devices (SSD '09)*, vol. 1, pp. 1–6, March 2009.

[7] L. I. Kuncheva, "A theoretical study on six classifier fusion strategies," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 2, pp. 281–286, 2002.

[8] F. Alkoot and J. Kittler, "Experimental evaluation of expert fusion strategies," *Pattern Recognition Letters*, vol. 20, no. 11–13, pp. 1361–1369, 1999.

[9] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–239, 1998.

[10] X. Xia and L. O'Gorman, "Innovations in fingerprint capture devices," *Pattern Recognition*, vol. 36, no. 2, pp. 361–369, 2003.

[11] A. Ross and R. Govindarajan, "Feature level fusion using hand and face biometrics," in *Proceedings of the SPIE Conference on Biometric Technology for Human Identification*, vol. 5779, pp. 196–204, March 2005.

[12] L. Lam and C. Y. Suen, "Application of majority voting to pattern recognition: an analysis of its behavior and performance," *IEEE Transactions on Systems, Man, and Cybernetics A*, vol. 27, no. 5, pp. 553–568, 1997.

[13] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[14] V. Štruc, J. Zganec Gros, and N. Pavešić, "Principal directions of synthetic exact filters for robust real-time eye localization," in *Lecture Notes in Computer Science*, vol. 6583, pp. 180–192, 20112011.

[15] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1635–1650, 2010.

[16] K. Kryszczuk and A. Drygajlo, "Improving biometric verification with class-independent quality information," *IET Signal Processing*, vol. 3, no. 4, pp. 310–321, 2009.

[17] U. Klopčiv and P. Peer, "Fingerprint-based verification system: a research prototype," in *Proceedings of the 17th International Conference on Systems, Signals and Image Processing (IWSSIP '10)*, A. Conci and F. Leta, Eds., vol. 1, pp. 150–153, 2010.

[18] J. Fieriez-Aguilar, J. Ortega-Garcia, J. Gonzalez-Rodriguez, and J. Bigun, "Kernel-based multimodal biometric verification using quality signals," in *Proceedings of the Biometric Technology for Human Identification Conference*, pp. 544–554, April 2004.

[19] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4–20, 2004.

[20] V. Štruc and N. Pavešić, "The corrected normalized correlation coefficient: a novel way of matching score calculation for LDA-based face verification," in *Proceedings of the 5th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD '08)*, pp. 110–115, Shandong, China, October 2008.

[21] Q. Tao and R. Veldhuis, "Threshold-optimized decision-level fusion and its application to biometrics," *Pattern Recognition*, vol. 42, no. 5, pp. 823–836, 2009.

[22] S. Chaudhary and R. Nath, "A multimodal biometric recognition system based on fusion of palmprint, fingerprint and face," in *Proceedings of the International Conference on Advances in Recent Technologies in Communication and Computing (ARTCom '09)*, pp. 596–600, October 2009.

[23] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognition*, vol. 38, no. 12, pp. 2270–2285, 2005.

[24] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[25] V. Vapnik, *Statistical Learning Theory*, Wiley-Interscience, New York, NY, USA, 1998.

[26] C. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, Oxford, UK, 1995.

[27] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, "Xm2vtsdb: the extended m2vts database," in *Proceedings of the 2nd International Conference on Audio and Video-based Biometric Person Authentication (AVBPA '99)*, vol. 1, pp. 72–77, 1999.

[28] D. Maio, D. Maltoni, R. Cappelli, J. Wayman, and A. Jain, "Fvc 2002: second fingerprint verification competition," in *Proceedings of the 16th International Conference on Pattern Recognition*, vol. 1, pp. 811–814, 2002.

[29] N. Poh and S. Bengio, "Using chimeric users to construct fusion classifiers in biometric authentication tasks: an investigation," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP '06)*, vol. 1, pp. V1077–V1080, May 2006.

[30] R. Gajsek, F. Mihelic, and S. Dobrisek, "Speaker state recognition using an hmm-based feature extraction method," *Computer Speech & Language*, vol. 27, no. 1, pp. 135–150, 2013.

[31] R. Gajšek, S. Dobrišek, and F. Mihelič, *Analysis and Assessment of State Relevance in Hmm-Based Feature Extraction Method*, Lecture Notes in Computer Science, Springer, New York, NY, USA, 2012.

[32] B. Vesnicer and F. Mihelic, "The likelihood ratio decision criterion for nuisance attribute projection in gmm speaker verification," *EURASIP Journal of Advances in Signal Processing*, vol. 2008, Article ID 786431, 11 pages, 2008.

[33] M. Günther, A. Costa-Pazo, C. Ding et al., "The 2013 face recognition evaluation in mobile environment," in *Proceedings of the 6th IAPR International Conference on Biometrics*, June 2013.

[34] S. Bengio and J. Marithoz, "The expected performance curve: a new assessment measure for person authentication," in *Proceedings of the Speaker and Language Recognition Workshop (Odyssey)*, pp. 279–284, Toledo, Spain, 2004.

[35] V. Štruc and N. Pavešić, "The complete gabor-fisher classifier for face recognition under variable lighting," *EURASIP Journal of Advances in Signal Processing*, vol. 2010, no. 31, pp. 1–26, 2010.

[36] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1–27, 2011.

[37] S. Nissen, "Implementation of a fast artificial neural network library (fann)," Tech. Rep., Department of Computer Science, University of Copenhagen, Kbenhavn, Denmark, 2003.

[38] D. E. Maurer and J. P. Baker, "Fusing multimodal biometrics with quality estimates via a Bayesian belief network," *Pattern Recognition*, vol. 41, no. 3, pp. 821–832, 2008.

[39] N. Poh, T. Bourlai, J. Kittler et al., "Benchmarking quality-dependent and cost-sensitive score-level multimodal biometric fusion algorithms," *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 4, pp. 849–866, 2009.

[40] W.-Y. Lin and C.-J. Yang, "An enhanced biometric score fusion scheme based on the adaboost algorithm," in *Proceedings of the International Conference on Information Security and Intelligence Control (ISIC '12)*, pp. 262–265, 2012.

[41] M. Vatsa, R. Singh, A. Noore, and A. Ross, "On the dynamic selection of biometric fusion algorithms," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 3, pp. 470–479, 2010.

[42] N. Poh, J. Kittler, and T. Bourlai, "Quality-based score normalization with device qualitative information for multimodal biometric fusion," *IEEE Transactions on Systems, Man, and Cybernetics A*, vol. 40, no. 3, pp. 539–554, 2010.

[43] N. Poh, D. Windridge, V. Mottl, A. Tatarchuk, and A. Eliseyev, "Addressing missing values in kernel-based multimodal biometric fusion using neutral point substitution," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 3, pp. 461–469, 2010.

[44] L. Nanni, A. Lumini, and S. Brahnam, "Likelihood ratio based features for a trained biometric score fusion," *Expert Systems with Applications*, vol. 38, no. 1, pp. 58–63, 2011.

[45] N. Poh and J. Kittler, "A Unified framework for biometric expert fusion incorporating quality measures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 3–18, 2012.

[46] A. Nagar, K. Nandakumar, and A. K. Jain, "Multibiometric cryptosystems based on feature-level fusion," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 1, pp. 255–268, 2012.

[47] Q. Tao and R. Veldhuis, "Robust biometric score fusion by naive likelihood ratio via receiver operating characteristics," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 2, pp. 305–313, 2013.