Multi-biometric Identification with Cascading Database Filtering

Pawel Drozdowski, Christian Rathgeb, Benedikt-Alexander Mokroß, Christoph Busch

Abstract—The growing scale and number of biometric deployments around the world necessitates research into technologies which facilitate fast identification queries and high discriminative power. In this context, this article presents a biometric identification system which relies on a successive pre-filtering of the potential candidate list using multiple biometric modalities, coupled with a weighted score-level information fusion. The proposed system is evaluated in a series of experiments using a compound dataset constructed from several publicly available datasets; an open-set identification scenario is considered with the enrolment database containing 1,000 chimeric instances. This evaluation shows that the proposed system exhibits a significantly increased biometric performance w.r.t. a weighted score-level or rank-level fusion based baseline, while simultaneously providing a consequential computational workload reduction in terms of penetration rate. Lastly, it is worth noting that the proposed system could be flexibly employed in any multi-biometric identification system, irrespective of the chosen types of biometric characteristics and the encoding of their extracted features.

Index Terms—Biometric Identification, Information Fusion, Computational Workload Reduction

1 Introduction

Various market value studies (see *e.g*. [1], [2], [3]) evince the rapid growth of interest and investment in biometric technologies. Biometrics are being used by various governmental organisations around the world for purposes such as law enforcement and forensic investigations (see *e.g*. [4], [5], [6]), border control (see *e.g*. [7], [8], [9], [10]), national ID systems (see *e.g*. [11], [12]), as well as during elections for voter registration (see *e.g*. [13], [14]). The largest of such deployments to date is located in India, where the Unique Identification Authority of India operates a national ID system (Aadhaar) which accommodates, at the time of this writing, almost 1.3 billion enrolled subjects (see *e.g*. the online dashboard [15]). Additionally, the prevalence and computing power of mobile devices (especially smartphones) has been steadily increasing. Together with the advances in embeddable high-quality sensors, those trends have sparked interest in (single and multi modal) mobile biometrics, which has become an active area of research and product development (see *e.g*. [16], [17], [18], [19], [20]).

With the aforementioned increase of the popularity and sizes of biometric systems in the governmental and commercial sectors alike, it is important to develop technologies which facilitate accurate and efficient processing of large amounts of biometric data. In particular, guaranteeing practical system response times by means of algorithmic solutions, rather than merely the scaling of the hardware architecture is of utmost interest. Those considerations are especially important for biometric identification (and duplicate enrolment check) scenarios, where the conventional biometric systems typically conduct an exhaustive search (entailing one-to-many comparison) to identify the biometric probes. Daugman, the pioneer of iris recognition, stated (in a recent interview) that performing accurate and efficient biometric identification (*i.e*. without an exhaustive search) is one of the most important, unsolved issues in biometrics in general. From the governmental side there exists a strong interest for computationally efficient biometric algorithms, as evidenced by multiple competitions and benchmarks (*e.g*. 1:N Evaluation under Face Recognition Vendor Test (FRVT) [21], one-to-many evaluations under Iris Exchange (IREX) [22], and Biometric Technology Rally [23]).

In recent years, a significant research effort has been devoted to addressing this topic by developing methods for computational workload reduction in biometric systems (see subsection 2.2 and a recent survey of Drozdowski *et al*. [24] for more details). The contribution of this work in this context is a proposal of an information fusion scheme, as well as an experimental evaluation thereof on a large compound dataset in the biometric open-set identification scenario. The scheme is based on a successive filtering of candidate shortlists coupled with information fusion on score level. It is shown that the proposed scheme increases the biometric performance w.r.t. the weighted score-level or rank-level fusion based baseline by an order of magnitude, while simultaneously significantly reducing the computational workload (in terms of penetration rate) of the biometric identification transactions. In related works, several authors utilised dimensionality reduction and/or binarisation to create short-length templates, which are used to pre-filter the enrolment database in a two-stage framework (see *e.g*. Gentile *et al*. [25], Billeb *et al*. [26], and Pflug *et al*. [27]), whereas Drozdowski *et al*. [28] used biometric image morphing in a similar manner. All of those methods considered single-modal systems. A decision-based cascade operating on the principle of sequential fusion of fingerprint

- *Pawel Drozdowski and Christoph Busch are with da/sec – Biometrics and Internet Security Research Group, Hochschule Darmstadt, Germany, and NBL – Norwegian Biometrics Laboratory, Norwegian University of Science and Technology, Gjøvik, Norway.*
- *Benedikt-Alexander Mokroß is with iCOGNIZE GmbH, Dietzenbach, Germany.*
- *Christian Rathgeb is with da/sec – Biometrics and Internet Security Research Group, Hochschule Darmstadt, Germany.*
  E-mail: {name}.{lastname}@h-da.de

and iris recognition systems was presented by Elhoseny *et al*. [29]. Lastly, database pre-filtering based on demographic and geographic metadata is a known and widely used method of searching in large-scale databases (see *e.g*. Gehrmann *et al*. [30]). Soft biometrics (see *e.g*. Dantcheva *et al*. [31]) can also be used in an analogous manner. Best to the authors' knowledge, previous research has not considered a multi-modal fusion utilised in a cascading manner for the simultaneous purpose of computational workload reduction and biometric performance improvement.

The remainder of this article is organised as follows: section 2 provides a background overview of the two relevant related work areas – biometric information fusion and computational workload reduction. In section 3, the proposed system is described. The details of the experimental setup are outlined in section 4, while the results of the experiments are presented in section 5 and discussed in section 6. A summary and concluding remarks are given in section 7.

2 Background and Related Work

In this section, relevant background information and related work w.r.t. the two main topics of this article are outlined. Specifically, subsection 2.1 addresses biometric information fusion, while subsection 2.2 deals with computational workload reduction in biometric systems. Furthermore, the scope of this article and the proposed system is demonstrated within the overall overview of those research areas.

2.1 Biometric Information Fusion

One of the key goals of biometric information fusion is to increase the overall discriminative power of a biometric recognition system. Systems where biometric information fusion is utilised are referred to as multi-biometric systems. In such systems, multiple information sources are considered and combined (fused) with each other. In the context of biometrics, following main fusion categories can be distinguished (see *e.g*. Ross *et al*. [32] and ISO/IEC TR 24722 [33]):

Multi-type Information from multiple biometric characteristics (*e.g*. face and fingerprint) is used.

Multi-sensorial Biometric data is acquired using several complementary sensors (*e.g*. visible wavelength and near-infrared camera).

Multi-algorithm Biometric samples are processed using multiple complementary algorithms (*e.g*. texture and keypoint based image descriptors for feature extraction and/or different concepts for comparison).

Multi-instance Information from multiple instances of the same characteristic is used (*e.g*. left and right iris).

Multi-sample Multiple samples (acquisitions) of the same characteristic are used (*e.g*. for sample quality assurance or detection of reliable regions).

The system proposed in this article pertains to the first scenario. Specifically, three types of biometric characteristics are chosen and are subsequently used in a pre-filtering and fusion scheme. In addition to the coarse categories above, several levels of the biometric processing pipeline can be distinguished where information fusion can be performed (see *e.g*. Ross *et al*. [32]):

Sensor Information from multiple sensors or multiple samples (*e.g*. on the pixel level for images or phase level for audio/video signals) is combined prior to any other processing steps. See *e.g*. Jain *et al*. [34] and Kusuma *et al*. [35].

Feature Information from multiple extracted feature sets is consolidated. The data could come from the same biometric characteristic (*e.g*. multiple, complementary feature extractors are used) or different biometric characteristics (*e.g*. a common feature representation is used for the fusion). See *e.g*. Kanhangad *et al*. [36] and Yan *et al*. [37].

Score The comparison scores acquired from multiple information channels are combined (*e.g*. summed or averaged). Depending on the used biometric comparators, this often requires normalisation of the scores to a common domain. See *e.g*. Snelick *et al*. [38] and Jain *et al*. [39].

Rank First, the ranks (order) of potential matches of a probe against an enrolment database are established. Subsequently, heuristics (*e.g*. choosing the best rank or majority vote) are used to consolidate the information from multiple systems. See *e.g*. Abaza *et al*. [40] and Kumar *et al*. [41].

Decision The decisions (*i.e*. accept/reject) reached by multiple systems are combined using heuristics (*e.g*. majority voting or statistics-based rulesets). See *e.g*. Prabhakar *et al*. [42] and Paul *et al*. [43].

In the context of this work, information fusion on score and rank level is of most interest. This is partially because score-level fusion is amongst the most popular and best performing of the aforementioned methods (see Ross *et al*. [32]), and partially because the proposed system (see section 3) is designed to work at those levels of the biometric pipeline, *i.e*. irrespective of the chosen biometric characteristics, acquisition methods, and feature extraction algorithms.

Several extensive works and surveys on the topic of biometric information fusion have been published in the scientific literature. The interested reader is therefore referred to *e.g*. Ross *et al*. [32] for a comprehensive general introduction to this topic, Jain *et al*. [39] and Snelick *et al*. [38] for score-level fusion specifically, as well as Radu *et al*. [44], Dinca *et al*. [45], and ISO/IEC TR 24722 [33] for more recent works concerning the overall topic of biometric information fusion.

2.2 Computational Workload Reduction

There exists a broad variety of ways in which biometric systems can operate. The main two of them (quoted directly from the ISO/IEC international standards [33], [46], [47]) are:

Biometric verification Referring to the "process of confirming a biometric claim through biometric comparison".

Biometric identification Referring to the "process of searching against a biometric enrolment database to find and return the biometric reference identifier(s) attributable to a single individual".

In the context of biometric identification, two main scenarios can be distinguished, namely *closed-set* identification, where it is known that the enrolment database contains all the potential system users as data subjects, and *open-set* identification, where it is possible that some potential users (impostors) are not enrolled in the system.

Open-set biometric identification, which is, arguably, the most challenging from the practical point of view, is the focus of this article. Due to the necessity of protecting against impostors not enrolled with the system, as well as the lack of an identity claim during a transaction, in the worst case an exhaustive search (*i.e.* comparisons between the probe and the entire enrolment database) is required in order to make a decision. Unfortunately, two non-trivial problems are quickly encountered by this naïve approach:

Computational costs With an increasing size of the enrolment database, the response times become proportionally slower, hence requiring hardware investment and/or software optimisations to facilitate the growing number of the data subjects.

False positives costs Daugman [48] has pointed to a demanding relationship facing biometric identification systems:

$$P_N = 1 - (1 - P_1)^N \tag{1}$$

This equation denotes the probability (P_N) of at least one false positive occurrence in an identification transaction within a system which comprises N enrolled users and has a P_1 false positive probability of a one-to-one template comparison. Even when P_1 is very low (*i.e.* the system would exhibit good biometric performance in verification mode), P_N raises very quickly to unacceptable levels as N increases[1].

Since the overall computational costs in a biometric identification scenario are dominated by performing the biometric comparisons (see *e.g.* Drozdowski *et al.* [24]), most computational workload reduction approaches are aimed at that step in the system pipeline. It should be noted, that due to certain properties of biometric data (*i.e.* lack of inherent ordering, within-subject variability, and high dimensionality), many traditional approaches (such as normal database indexing) are often unsuitable or perform poorly (see Hao *et al.* [49]). Therefore, approaches specifically tailored to those properties have been developed. In particular, two main approach classes can be distinguished:

Pre-selection Approaches in this category concentrate on reduction of the potential search space, *i.e.* the number of necessary template comparisons (penetration rate) during a biometric identification transaction. Three principal sub-categories can be distinguished here:

1) **Pre-filtering** Multiple algorithms or feature representations are used. The idea is to first use computationally efficient (but somewhat inaccurate) methods to create a candidate shortlist, whereupon a computationally expensive (but accurate) method is used on this small, pre-filtered subset of the database (see *e.g.* Ratha *et al.* [50], Gentile *et al.* [25], and Billeb *et al.* [26]).
2) **Binning** The database is split into distinct bins/partitions based on some coarse auxiliary features. Examples include metadata (such as demographic and geographic attributes, see *e.g.* Gehrmann *et al.* [30]) or biometric characteristic specific features, such as fingerprint classes (see *e.g.* Drozdowski *et al.* [51]). During a biometric identification only the bins corresponding to the sample are considered, thereby reducing the search space. As an alternative to such handcrafted features, unsupervised clustering can also used (see *e.g.* Ross *et al.* [52] and Pflug *et al.* [53]).
3) **Datastructures** The enrolment database is reorganised to take advantage of efficient ordering principles, for example based on search trees (see *e.g.* Proença [54] and Rathgeb *et al.* [55]) or fuzzy hashing (see *e.g.* Cappelli *et al.* [56] and Kaushik *et al.* [57]), thereby enabling sub-linear/logarithmic search time.

Feature transformation Approaches in this category concentrate on reducing the computational cost of the individual template comparisons. Typical approaches in this category accomplish this by reducing the dimensionality of the biometric templates by extracting the most discriminative parts (see *e.g.* Gentile *et al.* [58] and Rathgeb *et al.* [59]), utilising more efficient comparators such as integer/bit-based instead of float-based (see *e.g.* Lim *et al.* [60] and Drozdowski *et al.* [61]), or providing sample alignment invariance (see *e.g.* Rathgeb *et al.* [62] and Damer *et al.* [63]).

An exhaustive survey of this research area is out of scope for this article – for more details, the interested reader is referred to other works on this topic. Specifically, in a recently published work of Drozdowski *et al.* [24], a biometric characteristic-agnostic, concept-based taxonomy of computational workload reduction approaches in biometrics has been proposed. Additionally, the authors conducted a comprehensive survey of computational workload reduction in biometric identification systems in the context of said taxonomy. For biometric characteristic-specific works, surveys by Schuch [64] (fingerprint), Proença *et al.* [65] (iris), and Kavati *et al.* [66] (fingerprint, face, iris) are of interest.

In the context of the above categories of computational workload reduction approaches, the pre-selection (more specifically, pre-filtering) one is most relevant to this work. This is because, as previously mentioned, this article presents (see section 3) a method which relies on a successive candidate shortlist filtering, and works irrespective of the chosen type of biometric characteristics and their feature representations, thereby precluding any approaches which rely on specific feature transformations.

3 Proposed System

Consider a biometric enrolment database with references of N data subjects for K different biometric modalities (*i.e.* types of biometric characteristics). A standard approach for a biometric identification transaction would be to conduct the comparisons (C) exhaustively (*i.e.* $\#\mathrm{C}_{\text{baseline}} = K \cdot N$

1. Although this equation ignores other system errors, such as the failure-to-acquire rate and also assumes that at a given threshold all subjects have the same false-match-rate (which likely is not the case), it nonetheless is a useful approximation through which the challenges of the biometric identification systems can be illustrated quantitatively.

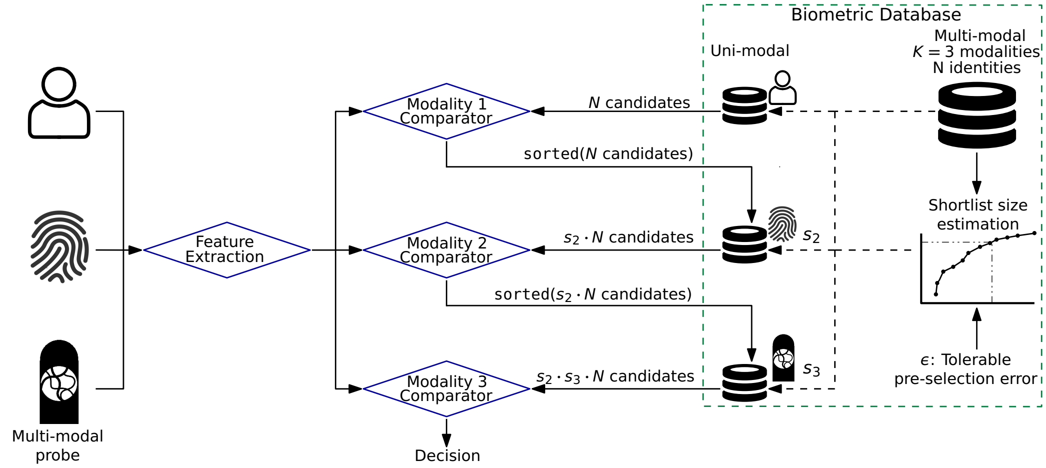


Fig. 1: Overview of the proposed system

comparisons) for all the modalities and to fuse the scores using one of the traditional strategies (such as score or rank level fusion) described in subsection 2.1. This approach will serve as a baseline later on in the experiments. Here, an alternative method is proposed with the aim of improving the biometric performance and reducing the computational workload.

The conceptual overview of the proposed system (for $K = 3$) is shown in figure 1. On the biometric database side, the stored modalities are given a specific order (see subsection 3.2 and section 5 for more details on the ordering of the chosen types of biometric characteristics). The key idea is to successively filter the list of potential candidates based on the comparisons within the individual modalities, thus creating a multi-stage (K-stage), cascading filtering system. In the illustrated example, a biometric identification transaction would proceed as follows:

1) Features are extracted for each of the probe sample of each biometric type (*i.e.* modality).
2) Modality 1 (face) probe is compared exhaustively (N comparisons) against the enrolment database. Based on the sorted comparison scores, a certain fraction (denoted s_2) of the most promising candidates (a candidate shortlist) is passed onto the next level.
3) Modality 2 (fingerprint) probe is compared against the ($s_2 \cdot N$) most promising candidates. A fraction of those (s_3) is then passed onto the last level.
4) Modality 3 (fingervein) probe is compared against the ($s_2 \cdot s_3 \cdot N$) most promising candidates to reach the final identification decision.

The types of biometric characteristics (face, fingerprint, and fingervein) were chosen based on the criteria that the three types be not correlated, that they are widely deployed (in various operational systems), and exhibit desirable properties w.r.t. presentation attack detection (the latter especially concerning the fingervein). However, it should be noted that the system is not in any way reliant on those specific characteristics or this particular ordering thereof – the system design is applicable irrespective of the participating biometric characteristics and their feature representations. The order of the characteristics in the cascade is also flexible – more on this topic in subsection 3.2 and the experimental evaluation in section 5.

The sizes of the candidate lists passed between the levels of the cascade (values in s_i) are estimated empirically in a training step, see subsection 3.1 for more details. Since at the first level the whole database is used for the comparisons, s_1 would equal 1.0 and is not depicted in figure 1. The theoretical impact of the proposed system on the biometric performance and computational workload is described in subsection 3.2.

3.1 Shortlist Size Estimation

For each considered type of biometric characteristic, a tolerable pre-selection margin of error in terms of false negative identification rate is determined. This margin is denoted as $\epsilon \in [0\% \ldots 100\%]$ and can be set arbitrarily low or high by the system operator depending on the system policy. The extreme values are unlikely in a practical scenario and are listed for the purposes of the mathematical definition only. This parameter is used for the purpose of shortlist size estimation for the pre-selection algorithm described in the previous subsection. The goal is to find the minimum fraction (denoted s) of candidate identities to pass between two levels of the cascade, so that the selected tolerable margin of error is not violated. Estimating s for a biometric type happens in a dedicated training step on a disjoint dataset, where a closed-set identification experiment is carried out and a cumulative match characteristic (CMC) curve is computed. Using the CMC curve, one first needs to calculate the minimum rank (r), so that $IR \geq 100\% - \epsilon$; then, r is expressed relative to the size (N_{train}) of the training enrolment database (E_{train}). An abstract, formal description of this concept is given in algorithm 1.

For a concrete example of the concept, see figure 2. There, a CMC curve for an example system (purely for illustrative purposes; the chosen type of biometric characteristic does not matter in this case) with 30 enrollees has been computed. Two operational points (with different ϵ values) are considered, expressing different system policies: a liberal one, wherein some pre-selection (false negative) errors are acceptable (depicted with the orange line), and a

Algorithm 1 Shortlist size estimation

Input: E_{train}, ϵ
Output: s_ϵ
1: $CMC \leftarrow$ COMPUTECMC(E_{train})
2: $r_\epsilon \leftarrow \min\{r \in \{1 \ldots N\} \mid CMC(r) \geq 100 - \epsilon\}$
3: $N_{train} \leftarrow$ LENGTH(E_{train})
4: $s_\epsilon \leftarrow \frac{r_\epsilon}{N_{train}}$
5: **return** s_ϵ

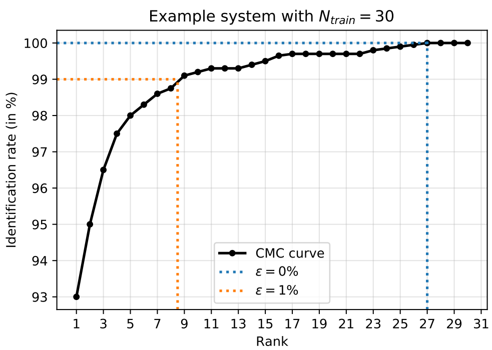


Fig. 2: Determining s_ϵ based on a training CMC curve and ϵ

stringent one, which seeks to minimise pre-selection errors (depicted with the blue line). Consequently, in the case of the $\epsilon = 1\%$ policy, the lowest rank satisfying the IR constraint (see line 2 in algorithm 1) is 9, meaning that the candidate shortlist passed onto the next level would be around $s_{\epsilon=1\%} = \frac{9}{30} = 30\%$ of the enrolment database. On the other hand, for the $\epsilon = 0\%$ policy, the constraint is satisfied only at rank 27, thus making the passed candidate shortlist $s_{\epsilon=0\%} = \frac{27}{30} = 90\%$ of the enrolment database.

Thus, for each considered modality and ϵ, a theoretically optimal candidate shortlist size, expressed as a fraction of the enrolment database (s, $s \in \{x \in \mathbb{R} \mid 0.0 < x \leq 1.0\}$), can be ascertained w.r.t. the system policy. A multi-stage system with K modalities would then have $S_\epsilon = [s_1 \ldots s_K]$ shortlist sizes, with s_1 always equal to 1.0. Generally, more liberal (*i.e.* higher) values of ϵ mean smaller candidate shortlists (*i.e.* lower penetration rate), but increased potential of pre-selection (false negative) errors. On the other hand, the reduction of penetration rate contributes to reducing the false positive error rate. This is because the reduction in penetration rate corresponds to the factor N in equation 1 being reduced, *i.e.* there being fewer potential comparisons in a biometric identification transaction where an impostor could, just by chance, get a better comparison score against a reference in the enrolment database. Note, that this is true insofar there exists no correlation between the comparison scores of the biometric characteristics in the system, *i.e.* that they be statistically independent. Certain biometric modalities can exhibit explicit or hidden symmetries and correlations (see *e.g.* Gomez-Barrero *et al.* [67] and *e.g.* Kumar *et al.* [68]). Such correlations have a non-trivial impact on the biometric performance of information fusion schemes (see *e.g.* Ulery *et al.* [69]). Generally speaking, the utility (in terms of increase of information entropy, see *e.g.* Adler *et al.* [70]) of correlated modalities may be lower than that of uncorrelated ones. Furthermore, specifically for the proposed scheme, using correlated modalities for the pre-filtering stage would be counter-productive, as computational workload would have to be expended on performing the comparisons, but little or no additional information would have been gained for the pre-selection of candidates. In other words, while the proposed scheme technically supports any combination of biometric modalities by the virtue of operating at the level of comparison scores, some attention is nevertheless required w.r.t. the choice of the modalities participating in the scheme. Ideally, completely uncorrelated modalities should be used. If correlated modalities are chosen, the results in terms of biometric performance and computational workload reduction may be degraded. Therefore, care and awareness is advised w.r.t. the choice of biometric modalities for the proposed scheme. Note, that this caveat of correlated data is also applicable to other existing biometric information fusion schemes. In this article, three uncorrelated biometric characteristics have been chosen for the experiments (see subsection 4.1).

3.2 System Ordering

The modalities participating in the cascade can be ordered arbitrarily – the number of possible permutations for a system with K modalities is $K!$. The ordering is expected to have a non-trivial impact on the computational workload and biometric performance. If the computational cost of individual template comparisons is also considered (see section 6), the system ordering has an impact not only on the biometric performance, but also the overall computational workload. The total number of comparisons in the proposed system is:

$$\#\mathrm{C}_{\mathrm{proposed}} = N + \sum_{k=2}^{K} \prod_{i=1}^{k} s_i \cdot N \tag{2}$$

The key idea behind equation 2 being that $\#\mathrm{C}_{\mathrm{proposed}} \ll \#\mathrm{C}_{\mathrm{baseline}}$, *i.e.* reducing the penetration rate of the search. The lower bound of the penetration rate is then $p = \frac{1}{K} + \frac{1}{N}$, *i.e.* in the case of a 3-level system the minimum penetration rate could be around 33.(3)%. This limit is due to exhaustive search always having to be be conducted for the first modality in the cascade. A potential extension of the proposed system could consider another scheme of computational workload reduction (*e.g.* binning) to be used prior to the first level of the cascade in order to avoid the necessity of conducting an exhaustive search there.

3.3 Combination with Weighted Score-level Fusion

The system proposed in the previous subsections uses the comparison scores in the shortlist from the final level of the cascade to make a decision. It is, however, also possible to combine the traditional weighted score-level fusion with the proposed scheme. Specifically, such a combined scheme would work as follows:

1) Conduct the cascading filtering with K modalities as described in the previous subsections.
2) Retrieve the identities of the subjects in the candidate shortlist produced at the last level of the cascade.

3) Retrieve the comparison scores corresponding to the candidate shortlist for the modality at the final level of the cascade *and* the previous levels of the cascade.
4) Fuse the scores.

In other words, the database is first filtered to find the most likely candidates using the individual modalities in the cascade, whereupon the comparison scores (for all the modalities) of the candidates in the shortlist are fused. In theory, such a system should exhibit a decreased computational workload, as well as an increased biometric performance. This idea is evaluated experimentally in addition to the system proposed in the previous subsections.

4 Experimental Setup

The following subsections outline the details of the experimental setup. The chosen datasets and processing pipelines are described in subsections 4.1 and 4.2, respectively. The baseline and proposed system configurations, as well as the evaluation metrics are described in subsections 4.3 and 4.4.

4.1 Datasets

The research conducted in this paper is aimed at cooperative systems, *i.e.* ones where biometric samples of reasonably good quality can be expected. Hence, in-the-wild, large time-scale, and occluded facial datasets (or parts thereof), as well as latent fingerprint datasets were not considered. Since none large-scale multi-modal datasets were available to the authors, it was decided to create a virtual dataset from existing single-modal ones.

While there exist several datasets with very large numbers of biometric samples, their size in terms of data subjects is typically much smaller. As such, several facial and fingervein datasets had to be considered in order to obtain a suitable number of data subjects. For the fingerprint and fingervein datasets, the individual instances (fingers) are not correlated and can therefore be treated as separate subjects.

The fingerprint database was used directly without any filtering. For the facial datasets, frontal images without intentional occlusions (*e.g.* scarves or sunglasses) were chosen, while some images with exceedingly poor quality were removed from the fingervein and facial datasets (to facilitate reproducible research, the lists of chosen images and other experimental setup details will be made available online after this article is accepted for publication). It should be noted that the facial and especially fingervein data is extremely inhomogeneous across the chosen datasets. The images were acquired using different cameras/sensors, under varying lighting conditions, and the images have been saved in several distinct resolutions. The chosen datasets are listed in table 1 (the numbers given in the table are after the filtering was applied). Example images from the datasets are shown in figure 3.

Following the selection, the datasets of the same biometric characteristic have been merged and a compound dataset was constructed. This was done by repeatedly (ten times) shuffling the instances and samples from the original datasets to construct new chimeric instances. The ten copies of the dataset enable a tenfold cross-validation in the experimental evaluation. Each of the copies consists of 2,500 instances and approximately 15,000 samples (depending on availability, since different datasets contain different number of samples per instance). Finally, the resulting compound dataset has been split into two partitions:

Training Consists of 1,000 instances. Used for computing CMC curves in a closed-set identification scenario to approximate the appropriate shortlist sizes for each modality. Additionally used for computing information necessary for comparison score normalisation.

Testing Consists of 1,500 instances. Used for evaluating the baselines (for each modality individually and for several popular information fusion schemes) and the proposed system in an open-set identification scenario.

4.2 Processing Pipelines

The images were processed using exclusively open-source frameworks. While commercial frameworks may have offered a better or even errorless biometric performance on

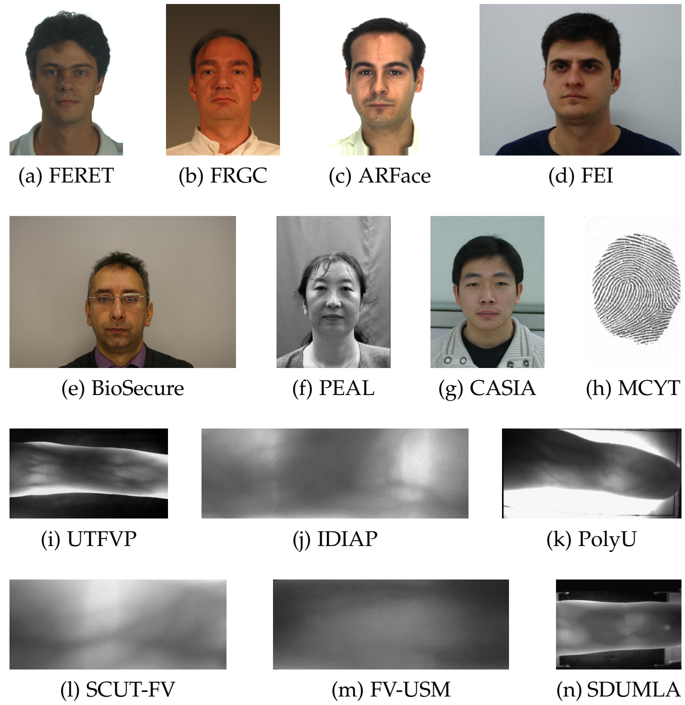


Fig. 3: Example images from the selected datasets

TABLE 1: Used datasets

Characteristic	Dataset	Instances	Samples
Face	FERET	994	2,716
	FRGC	453	2,754
	ARFace	136	1,526
	FEI	200	600
	BioSecure	210	840
	PEAL	429	3,274
	CASIA	725	3,072
Fingerprint	MCYT	3,300	39,600
Fingervein	UTFVP	360	1,440
	IDIAP	220	440
	PolyU	312	3,132
	SCUT-FV	600	3,600
	FV-USM	492	5,904
	SDUMLA	636	3,816

the dataset used in the experiments (see subsection 4.1), facilitating reproducible research has been deemed a higher priority, hence favouring the open-source frameworks. Furthermore, it has been shown both theoretically and in practice (see *e.g*. Daugman *et al*. [71] and Grother *et al*. [72]), that the biometric performance in the identification scenario decreases with the growing size of the enrolment database. In other words, in large-scale systems, optimal biometric performance is not to be expected, even from the commercial systems. Lastly, to evaluate the proposed fusion methods, the key metric is the *relative* biometric performance gain/loss w.r.t. to a baseline and not the *absolute* biometric performance achieved. Following tools and frameworks were used to extract features from the images and compare the resulting templates:

Face A neural-network based approach is used. Specifically, the FaceNet CNN of Schroff *et al*. [73] is used with a pre-trained model made available by the authors[2]. The network learns to map facial images to Euclidean space, whereby the produced templates (embeddings) can be directly compared using Euclidean distance.

Fingerprint The features (minutiae triplets, *i.e*. 2-D location and angle) are extracted using a neural-network based approach. In particular, the FingerNet CNN of Tang *et al*. [74] is used with a pre-trained model made available by the authors[3]. To compare such templates, a minutiae pairing and scoring algorithm of the sourceAFIS system of Važan [75] is used[4].

Fingervein A minutiae based approach is used. Specifically, the maximum curvature algorithm of Miura *et al*. [76] is used to extract the skeleton of the fingervein patterns, which is subsequently thinned using the method presented by Guo *et al*. [77]. The minutiae are retrieved from the vein skeleton with a convolution kernel proposed by Olsen *et al*. [78]. Lastly, using the method of Xu *et al*. [79] and Hartung *et al*. [80], [81], the variable-sized minutiae vector is translated into the Spectral Minutiae Representation (SMR), which is fixed-length and additionally offers certain implicit rotation and scaling invariance. Such templates can be compared using a simple correlation measure (likewise presented in [79]), which is a common approach in image processing.

Table 2 summarises the information about the utilised data processing pipelines.

TABLE 2: Data processing pipelines

Characteristic	Extraction	Representation	Size	Comparison
Face	FaceNet	1-D embedding	512 floats	Euclidean distance
Fingerprint	FingerNet	Minutiae triplets set	Variable	Minutiae pairing
Fingervein	Spectral minutiae	2-D matrix	256×128 floats	Correlation

For the biometric fusion, two scenarios were considered (see *e.g*. Jain *et al*. [39], Snelick *et al*. [38], and Ho *et al*. [82] for details):

Score level The scores were normalised using Z-score method, which is one of the most commonly used score normalisation methods and relies on the arithmetic mean and standard deviation of the scores data. This method is expected to perform well when prior knowledge about the score distributions is available – which is the case in this experimental setup (see subsection 4.1). Subsequently, the normalised scores were fused with a sum-rule method (using those methods, very good biometric performance has been observed in general, see *e.g*. Jain *et al*. [83] and ISO/IEC TR 24722 [33]).

Rank level A Borda count based method (see Black [84]), which is a group consensus function and a generalisation of the majority vote, was used. The method relies on summing the ranks assigned to the probe-reference pairs based on the comparison scores during a biometric identification transaction and requires no prior training.

In both cases, a weighted variant was also included, whereby the individual types of biometric characteristics are assigned relative weights, which are multiplied with the normalised scores prior to the fusion. The optimal weights' combinations were estimated experimentally (see subsection 4.3).

4.3 Baseline and Proposed System Configurations

To establish the baseline, against which the results of the proposed methods can be benchmarked, the following experiments were conducted on the testing subset of the compound dataset in an open-set identification scenario:

- Each of the 3 modalities individually.
- Weighted score-level and rank-level fusion (see subsection 4.2) of all possible combinations of 2 modalities and of all 3 modalities.

Pairs of weights in the interval $[0.05 \dots 0.95]$ with a step size of 0.05 were considered for the score and rank level fusion, thus yielding a total of 19 and 171 weights combinations for the fusion of 2 and 3 modalities, respectively.

Two versions of the proposed system were evaluated:

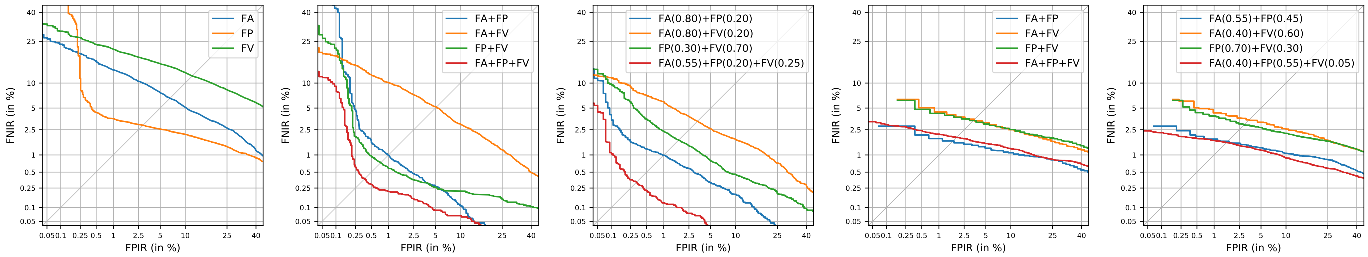
- Cascading filtering.
- Cascading filtering + weighted score-level fusion with a sum-rule.

For the second item above, same combinations of weights as in the baseline were used. Furthermore, all possible orderings of the modalities in the cascade were evaluated. All the experiments with the baseline and the proposed system were conducted using a tenfold cross-validation, as mentioned in subsection 4.1. Table 3 lists the number of configurations in each of the experiment types.

TABLE 3: Configurations per experiment

Experiment	Modalities	Orderings	Weights	Epsilons	Total
Individual baseline	1	3	—	—	3
Weighted fusion baseline	2	1	19	—	19
Weighted fusion baseline	3	1	171	—	171
Cascading filtering	2	6	—	7	42
Cascading filtering	3	6	—	7	42
Cascading filtering + weighted score fusion	2	6	19	7	798
Cascading filtering + weighted score fusion	3	6	171	7	7,182

2. https://github.com/davidsandberg/facenet
3. https://github.com/felixTY/FingerNet
4. The original algorithm uses minutiae quadruplets, *i.e*. additionally considers the minutiae type (*e.g*. ridge ending or bifurcation). Since minutiae triplets are extracted by FingerNet, the algorithm has been modified to ignore the type information. Using FingerNet instead of the native minutiae extractor provided by sourceAFIS is preferred, as it has yielded higher biometric performance.



(a) Single modality (b) Score fusion (c) Weighted score fusion (d) Rank fusion (e) Weighted rank fusion

Fig. 4: Baseline results

TABLE 4: Baseline results (with 95% CI)

Method	**Modality**	**PR**	τ	**EER (in %)**	**FPIR0.1 (in %)**	**d'**
Individual	FA	1.000	22.240 ± 0.773	6.648 ± 0.473	22.216 ± 0.583	3.153 ± 0.079
	FP	1.000	47.234 ± 3.435	2.864 ± 0.238	47.202 ± 2.592	3.638 ± 0.066
	FV	1.000	29.204 ± 0.401	11.958 ± 0.508	29.186 ± 0.302	2.218 ± 0.044
Rank fusion	FA+FP	1.000	2.784 ± 0.373	1.491 ± 0.199	2.594 ± 0.300	2.540 ± 0.038
	FA+FV	1.000	6.360 ± 0.544	3.081 ± 0.250	6.280 ± 0.416	2.404 ± 0.047
	FP+FV	1.000	6.332 ± 0.501	3.212 ± 0.314	6.252 ± 0.383	2.478 ± 0.034
	FA+FP+FV	1.000	3.127 ± 0.374	1.862 ± 0.200	2.960 ± 0.297	3.502 ± 0.059
Score fusion	FA+FP	1.000	25.340 ± 2.157	0.959 ± 0.152	25.253 ± 1.632	4.561 ± 0.082
	FA+FV	1.000	17.830 ± 0.336	5.049 ± 0.303	17.802 ± 0.254	3.033 ± 0.053
	FP+FV	1.000	14.587 ± 1.347	0.732 ± 0.092	14.430 ± 1.026	4.267 ± 0.048
	FA+FP+FV	1.000	5.615 ± 0.481	0.379 ± 0.112	5.295 ± 0.382	4.851 ± 0.054
Rank fusion weighted	FA(0.55)+FP(0.45)	1.000	2.743 ± 0.393	1.582 ± 0.195	2.550 ± 0.317	2.540 ± 0.039
	FA(0.40)+FV(0.60)	1.000	6.153 ± 0.573	3.342 ± 0.265	6.070 ± 0.438	2.391 ± 0.051
	FP(0.70)+FV(0.30)	1.000	5.970 ± 0.835	2.921 ± 0.294	5.884 ± 0.640	2.510 ± 0.043
	FA(0.40)+FP(0.55)+FV(0.05)	1.000	2.349 ± 0.303	1.564 ± 0.117	2.121 ± 0.254	3.532 ± 0.049
Score fusion weighted	FA(0.80)+FP(0.20)	1.000	5.107 ± 0.574	0.992 ± 0.146	4.901 ± 0.440	4.739 ± 0.079
	FA(0.80)+FV(0.20)	1.000	10.511 ± 0.309	3.311 ± 0.367	10.459 ± 0.234	3.632 ± 0.072
	FP(0.30)+FV(0.70)	1.000	8.533 ± 0.394	1.704 ± 0.103	8.459 ± 0.299	3.541 ± 0.049
	FA(0.55)+FP(0.20)+FV(0.25)	1.000	1.986 ± 0.160	0.324 ± 0.063	1.504 ± 0.136	4.985 ± 0.063

4.4 Evaluation Metrics

The systems were evaluated on two key aspects, using ISO/IEC standard methods and metrics [47] as well as additional, commonly used ones:

Biometric performance DET curves, equal-error-rate (EER), and false negative identification rate at a certain (here 0.1%) false positive identification rate (denoted FPIR0.1). Additionally, the decidability index over the genuine and impostor score distributions (defined as: $d' = \frac{|\mu_g - \mu_i|}{\sqrt{\frac{1}{2}(\sigma_g^2 + \sigma_i^2)}}$, where μ and σ stand for the means and standard deviations of the genuine and impostor score distributions, respectively) is reported.

Computational workload Penetration rate (PR), *i.e.* the number of the pre-selected candidate templates as a fraction of the total number of templates in the enrolment database.

Additionally, a metric which brings the two aspects together (adapted from Proença *et al.* [65]) is used. The metric (τ) calculates the Euclidean distance from the optimal operating point (*i.e.* FPIR0.1 = 0 and PR $\approx$ 0) and is defined as follows: $\tau = \sqrt{(\text{FPIR0.1})^2 + \text{PR}^2}$.

5 Results

In this section, the experimental results are presented. First, in subsection 5.1, the baseline is established. Subsequently, subsection 5.2 shows the empirical shortlist sizes estimation for the proposed system, while its results are presented in subsection 5.3. All the tables and figures in this section use a short notation for the biometric characteristics: FA (face), FP (fingerprint), and FV (fingervein). For the weighted fusion variants, the relative weights are written in parentheses immediately following their corresponding biometric characteristics.

5.1 Baseline

The results of the baseline experiments are shown in figure 4 and table 4. All possible combinations of modalities are shown; whereas for the weighted scenario, the results of the configuration with the lowest FPIR0.1 value for each modality combination are given. Looking at the baseline results, following conclusions can be reached:

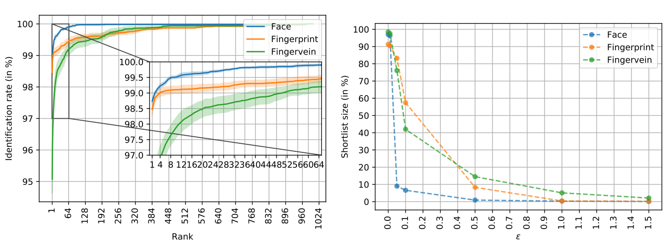
- The biometric performance of the individual modalities is only moderate. This is to be expected due to open-source tools being used, as well as the high degree of homogeneity and sometimes poor quality of the facial and fingervein data. However, it is also demonstrated that the biometric performance can be improved to useful levels by applying information fusion.
- The score-level fusion performs better than the rank-level fusion.
- The results can be further improved by applying relative weighting of the modalities. This is especially the case in terms of FPIR0.1 for the score-level fusion and less so for the rank-level fusion. It should be noted

that the exact optimal weights are only pertinent for a particular experimental setup (*i.e.* the specific databases, algorithms, *etc.*) and should not be used to reach general conclusions about weighted biometric fusion.

- The biometric performance in terms of EER of the best combination of 2 modalities and weights is around 1%, whereas using all 3 modalities reduces the EER down to around 0.35%. However, it should be noted that the FPIR0.1 is relatively high in both cases – around 5% and 1.5% for 2 and 3 modalities, respectively.
- Since the baseline setup relies on an exhaustive search method, the penetration rate is 1.0 and τ depends solely on the values of FPIR0.1.

5.2 Shortlist Size Estimation

To estimate the shortlist sizes, the methodology outlined in subsection 3.1 is followed. Accordingly, CMC curves are computed on the training partition of the dataset and then used to estimate the shortlist size for several ϵ values. In figure 5, the CMC curves are shown, along with the relation between the ϵ value and the shortlist size. It can be seen, that relatively high identification rate is achieved at very low ranks; however, it takes a while before 100% is reached, especially for the fingerprint and fingervein modalities. It should be noted, that those CMC curves do not provide a general statement w.r.t. to the relative strength of the chosen types of biometric characteristics; they merely provide a benchmark and overview relevant to the particular experimental setup (*i.e.* the specific databases, algorithms, *etc.*) used in this work.



(a) CMC curves (with 95% CI)

(b) The relation between ϵ value and the shortlist size as a percentage of the enrolment DB size

Fig. 5: Estimation of the shortlist sizes

5.3 Proposed

The results of the experiments with the proposed system are shown in figure 6 and table 5. All possible modality combinations and orderings are shown; whereas for the weighted scenario, the results of the configuration with the lowest τ value for each modality combination and ordering are given. Looking at the results of the proposed system, following conclusions can be reached:

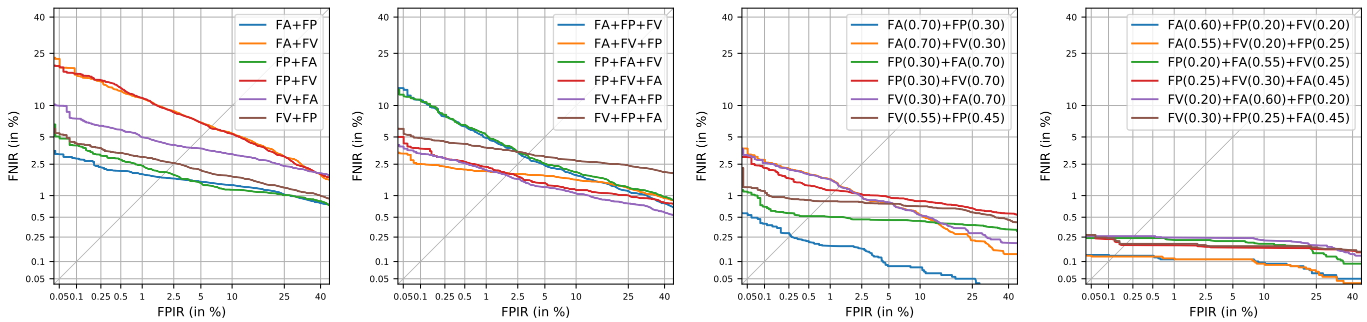
- Using the proposed technique alone improves the biometric performance for 2 modalities. For 3 modalities, a relatively good biometric performance is reached, albeit it is somewhat lower than the baseline.
- By combining the proposed technique with a weighted score-level fusion, the biometric performance is significantly improved (by an order of magnitude in some cases, *c.f.* table 6). The best baseline weighted score-level fusion configuration achieves approximately 0.992% and 0.324% EER for 2 and 3 modalities, respectively. The best configuration of the proposed scheme achieves approximately 0.254% and 0.109% EER for 2 and 3 modalities, respectively. Even more significant improvements can be seen in the higher security region of the error curves. The best baseline weighted score-level fusion configuration achieves approximately 4.901% and 1.504% FPIR0.1 for 2 and 3 modalities, respectively. The best configuration of the proposed scheme achieves approximately 0.333% and 0.125%. It should be noted that the exact optimal weights are only pertinent for a particular experimental setup (*i.e.* the specific databases, algorithms, *etc.*) and should not be used to reach general conclusions about weighted biometric fusion.
- In all the cases, the penetration rate (and hence the computational workload) is significantly reduced – down to 0.545 and 0.388 for 2 and 3 modalities, respectively. Those results are close to the theoretical maximum reduction (*i.e.* down to $\sim\frac{1}{K}$) for the proposed scheme as described in subsection 3.2.

6 Discussion

This section expands on the discussion items provided directly with the results in the previous section. Specifically, the results in terms of biometric performance and computational workload reduction are addressed in subsections 6.1 and 6.2, respectively. Lastly, subsection 6.3 outlines and discusses the potential limitations of this work.

6.1 Biometric Performance

It appears that the most successful system ordering follows the training CMC curves, *i.e.* preferring the type of biometric characteristic with the highest identification rate at low ranks to be used first. Accordingly, the best orderings (in terms of the τ metric) in the experiments were Face-Fingerprint and Face-Fingerprint-Fingervein for 2 and 3 modalities, respectively. In general, as has been demonstrated in the previous section, the proposed system increases the biometric performance when benchmarked against the baseline. This increase happens both in terms of FPIR0.1, as well as EER. Although the FNIR can be somewhat higher than that of the baseline (*c.f.* figure 7), this happens at values of FPIR which are considered impractical for operational systems. Those errors occur due to the pre-filtering – if, for example, at the first level of the cascade a sample of bad quality is filtered out, the proposed system cannot recover, whereas a score-fusion based system might, provided excellent scores for the other modalities. On the other hand, by the act of pre-filtering the database, the potential for false positives is decreased (recall subsection



(a) Cascading filtering, 2 modalities (b) Cascading filtering, 3 modalities (c) Cascading filtering and weighted score-level fusion, 2 modalities (d) Cascading filtering and weighted score-level fusion, 3 modalities

Fig. 6: Proposed system's results

TABLE 5: Proposed system's results (with 95% CI)

Method	Modality	ϵ	PR	τ	EER (in %)	FPIR0.1 (in %)	d′
Cascading	FA+FP	1.0	0.501	2.769 ± 0.602	1.719 ± 0.172	2.721 ± 0.093	4.408 ± 0.093
	FA+FV	1.5	0.500	18.015 ± 2.582	6.365 ± 0.446	18.008 ± 0.390	2.777 ± 0.050
	FP+FA	1.5	0.501	4.333 ± 1.387	1.983 ± 0.162	4.302 ± 0.210	4.929 ± 0.118
	FP+FV	1.5	0.501	17.518 ± 1.622	6.226 ± 0.356	17.511 ± 0.245	2.787 ± 0.044
	FV+FA	1.0	0.525	7.252 ± 1.176	4.308 ± 0.348	7.232 ± 0.178	3.747 ± 0.141
	FV+FP	1.0	0.525	5.015 ± 0.749	3.816 ± 0.342	4.987 ± 0.114	3.941 ± 0.082
	FA+FP+FV	1.0	0.335	10.461 ± 2.695	3.011 ± 0.220	10.456 ± 2.033	3.426 ± 0.054
	FA+FV+FP	1.5	0.334	2.652 ± 0.524	1.920 ± 0.177	2.630 ± 0.398	4.513 ± 0.093
	FP+FA+FV	1.5	0.334	9.829 ± 1.967	3.069 ± 0.212	9.823 ± 1.484	3.409 ± 0.047
	FP+FV+FA	1.0	0.335	3.673 ± 0.990	1.913 ± 0.125	3.657 ± 0.749	4.951 ± 0.098
	FV+FA+FP	0.5	0.382	3.017 ± 0.424	2.172 ± 0.215	2.992 ± 0.322	4.444 ± 0.094
	FV+FP+FA	0.5	0.386	5.028 ± 0.809	3.269 ± 0.267	5.013 ± 0.611	4.243 ± 0.140
Cascading + score fusion weighted	FA(0.70)+FP(0.30)	0.05	0.545	0.648 ± 0.079	0.254 ± 0.051	0.333 ± 0.102	5.702 ± 0.083
	FA(0.70)+FV(0.30)	0.05	0.545	2.713 ± 0.770	1.335 ± 0.171	2.654 ± 0.591	4.293 ± 0.076
	FP(0.30)+FA(0.70)	0.5	0.541	0.838 ± 0.168	0.509 ± 0.101	0.626 ± 0.159	5.640 ± 0.099
	FP(0.30)+FV(0.70)	1.5	0.501	2.440 ± 1.035	1.163 ± 0.144	2.380 ± 0.794	4.415 ± 0.061
	FV(0.30)+FA(0.70)	0.1	0.710	2.685 ± 0.742	1.327 ± 0.172	2.583 ± 0.576	4.275 ± 0.080
	FV(0.55)+FP(0.45)	0.5	0.573	1.482 ± 0.881	0.832 ± 0.134	1.338 ± 0.697	4.583 ± 0.064
	FA(0.60)+FP(0.20)+FV(0.20)	0.05	0.388	0.409 ± 0.012	0.109 ± 0.022	0.125 ± 0.027	6.356 ± 0.076
	FA(0.55)+FV(0.20)+FP(0.25)	0.05	0.386	0.407 ± 0.013	0.111 ± 0.027	0.121 ± 0.036	6.522 ± 0.073
	FP(0.20)+FA(0.55)+FV(0.25)	0.1	0.537	0.592 ± 0.022	0.215 ± 0.041	0.242 ± 0.042	6.215 ± 0.071
	FP(0.25)+FV(0.30)+FA(0.45)	0.1	0.605	0.642 ± 0.020	0.176 ± 0.047	0.204 ± 0.047	6.186 ± 0.077
	FV(0.20)+FA(0.60)+FP(0.20)	0.1	0.483	0.553 ± 0.036	0.239 ± 0.065	0.259 ± 0.064	6.243 ± 0.085
	FV(0.30)+FP(0.25)+FA(0.45)	0.1	0.554	0.597 ± 0.026	0.185 ± 0.048	0.213 ± 0.053	6.184 ± 0.077

3.2), thus yielding better results in terms of FPIR0.1. In other words, the proposed scheme can be used to increase the security of biometric identification systems which already employ information fusion of multiple biometric modalities.

6.2 Computational Workload Reduction

In addition to the aforementioned biometric performance improvement, the proposed system has an impact on the computational complexity of a biometric identification transaction. In this context, two scenarios can be distinguished depending on the *cost of template comparisons* for the used modalities:

Same cost irrespective of the modality In this case, the computational workload depends exclusively on the penetration rate (recall equation 2). To minimise it, the modalities should be ordered corresponding to the ascending order of their respective shortlist sizes, *i.e.* $S = \{s_1 \dots s_K \mid s_i \leq s_j, \forall i < j\}$. The computational workload (W) of an identification transaction in such a setup would be equal to the total number of comparisons, *i.e.* $W = \#\mathrm{C}_{\text{proposed}}$.

Different cost This case adds an extra factor (w_k) in the equations, representing the cost of the template comparison for the k'th modality, to be multiplied with the shortlist and enrolment database sizes. To minimise the computational workload, the ordering of the system would be $S' = \{s'_1 \dots s'_K \mid s'_i {*} w_i \leq s'_j {*} w_j, \forall i < j\}$, and the total computational workload for a biometric identification transaction $W = N \cdot w_1 + \sum_{k=2}^{K} \prod_{i=1}^{k} S'_i \cdot N \cdot w_k$.

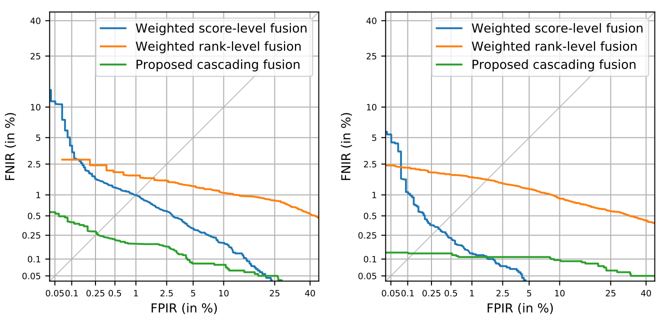
In this work, exclusively the first scenario was considered, due to the difficulty of consistently estimating the computational cost of individual template comparisons (see *e.g.* Drozdowski *et al.* [24] for a more detailed discussion on this topic). The main reason for this are the different feature representations and comparators across the modalities. One could, in theory, measure the execution time; however, this effectively amounts to measuring the efficiency of the software implementation and/or the underlying hardware architecture. This limited general use notwithstanding, such experiments would be useful for a specific system implementation (*e.g.* a commercial deployment).

6.3 Limitations

In terms of computational workload reduction, the main limitation of the proposed system is a hard limit of the potential penetration rate reduction, as described in subsection 3.2. Specifically, the biometric comparisons need to be conducted exhaustively on the first level of the cascade, thereby effectively limiting the minimum penetration rate to $\frac{1}{K}$, where K is the number of modalities in the cascade. Indeed, as reported in subsection 5.3, the results of the proposed system closely approach this maximal penetration

TABLE 6: Summary of the results – best configuration for each of the tested fusion methods (with 95% CI)

Method	Modality	ϵ	PR	τ	EER (in %)	FPIR0.1 (in %)	d′
Rank fusion weighted	FA(0.55)+FP(0.45)	—	1.000	2.743 ± 0.393	1.582 ± 0.195	2.550 ± 0.317	2.540 ± 0.039
	FA(0.40)+FP(0.55)+FV(0.05)	—	1.000	2.349 ± 0.303	1.564 ± 0.117	2.121 ± 0.254	3.532 ± 0.049
Score fusion weighted	FA(0.80)+FP(0.20)	—	1.000	5.107 ± 0.574	0.992 ± 0.146	4.901 ± 0.440	4.739 ± 0.079
	FA(0.55)+FP(0.20)+FV(0.25)	—	1.000	1.986 ± 0.160	0.324 ± 0.063	1.504 ± 0.136	4.985 ± 0.063
Proposed cascading fusion	FA(0.70)+FP(0.30)	0.05	0.545	0.648 ± 0.079	0.254 ± 0.051	0.333 ± 0.102	5.702 ± 0.083
	FA(0.60)+FP(0.20)+FV(0.20)	0.05	0.388	0.409 ± 0.012	0.109 ± 0.022	0.125 ± 0.027	6.356 ± 0.076



(a) 2 modalities

(b) 3 modalities

Fig. 7: Summary of the results – best configuration for each of the tested fusion methods

rate reduction, while simultaneously improving the biometric performance. The proposed scheme could, however, be extended by considering another method of computational workload reduction (*e.g.* binning) prior to the cascade in order to further reduce the penetration rate and avoid the exhaustive search at the first level of the cascade.

Another potential limitation is the necessity of the training step, in order to facilitate the shortlist sizes estimation, as well as score normalisation. This, however, is a common property of many (if not most) effective biometric information fusion systems.

In terms of a practical implementation, it should be noted that fully parallelised computations of the comparison scores across all the cascade levels are not possible. Specifically, while the computations on the individual cascade levels are, naturally, trivially parallelisable, it is not possible to compute all the cascade levels simultaneously. This is because the computations at each subsequent level of the cascade need to wait for the completion of the previous level, *i.e.* the creation of the candidate shortlist.

7 Summary

This article presented a biometric information fusion-based system which addresses two of the main challenges associated with biometric identification: biometric performance and computational workload. By successively filtering the candidate lists using the individual modalities and subsequently fusing the remaining comparison scores, the biometric performance in the region of the DET curve which is relevant for security sensitive applications, can be significantly improved, while simultaneously reducing the penetration rate (computational workload) of a biometric identification transaction. The proposed method could be seamlessly integrated into many operational multi-modal biometric identification systems, as it is designed to work irrespective of the chosen biometric characteristics or their respective feature representations, and only requires a straightforward training step for the purpose of parameter estimation.

A summary of the results (best configurations in terms of τ) for each of the fusion methods is shown in figure 7 and table 6. It can be seen that, w.r.t. using the weighted score-level or rank-level fusion alone, the proposed system has the following effects:

Biometric performance is improved in terms of of EER and FPIR0.1 – by an order of magnitude.

Computational workload is reduced in terms of penetration rate – down to around 55% and 39% for 2 and 3 modal system, respectively.

Operational flexibility is retained due to lack of dependence on specific biometric characteristics or template representations.

Future work in this area could consist of, for example, testing the proposed system with an even larger database (albeit those are difficult to come by in the research context), as well as using commercial off-the-shelf biometric recognition systems to assess the practicability of the proposed concept in the context of real biometric applications and operational (not virtual) datasets.

Acknowledgements

This research work has been funded by the German Federal Ministry of Education and Research and the Hessen State Ministry for Higher Education, Research and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE, and the LOEWE-3 BioBiDa Project (594/18-17).

References

[1] A. Bhutani and P. Bhardwaj, "Biometrics market size by application," Tech. Rep. GMI493, Global Market Insights, August 2017.
[2] Markets and Markets, "Biometric system market by authentication type - global forecast to 2023," Tech. Rep. SE 3449, Markets and Markets, July 2018.
[3] D. Thakkar, "Global biometric market analysis: Trends and future prospects," https://www.bayometric.com/global-biometric-market-analysis/, August 2018, Last accessed: February 12, 2020.
[4] K. R. Moses, P. Higgins, M. McCabe, S. Probhakar, and S. Swann, *Fingerprint Sourcebook*, chapter Automated Fingerprint Identification System (AFIS), pp. 1–33, US Department of Justice, 2010.
[5] Gemalto, "Automated Fingerprint Identification System (AFIS) - a short history," https://www.gemalto.com/govt/biometrics/afis-history, April 2019, Last accessed: February 12, 2020.

[6] Federal Bureau of Investigation, "CODIS - NDIS statistics," https://www.fbi.gov/services/laboratory/biometric-analysis/codis/ndis-statistics, June 2018, Last accessed: February 12, 2020.

[7] European Union Agency for the Operational Management of Large-Scale IT Systems in the Area of Freedom, Security and Justice, "Eurodac storage capacity increased," https://www.eulisa.europa.eu/Newsroom/News/Pages/Eurodac-storage-capacity-increased.aspx, April 2016, Last accessed: February 12, 2020.

[8] European Commission, "Smart borders," https://ec.europa.eu/home-affairs/what-we-do/policies/borders-and-visas/smart-borders_en, 2018, Last accessed: February 12, 2020.

[9] Gemalto, "DHS's automated biometric identification system IDENT - the heart of biometric visitor identification in the USA," https://www.gemalto.com/govt/customer-cases/ident-automated-biometric-identification-system, March 2019, Last accessed: February 12, 2020.

[10] T. Paynter, "Northrop Grumman wins $95 million award from Department of Homeland Security to develop next-generation biometric identification services system," https://news.northropgrumman.com/news/releases/northrop-grumman-wins-95-million-award-from-department-of-homeland-security-to-develop-next-generation-biometric-identification-services-system, February 2018, Last accessed: February 12, 2020.

[11] Unique Identification Authority of India, "Role of biometric technology in Aadhaar enrollment," Tech. Rep., UIDAI, January 2012.

[12] A. Dalwai, "Aadhaar technology and architecture: principles, design, best practices and key lessons," Tech. Rep., Unique Identification Authority of India (UIDAI), March 2014.

[13] K. W. Bowyer, E. Ortiz, and A. Sgroi, "Iris recognition technology evaluated for voter registration in Somaliland," *Biometric Technology Today*, vol. 2015, no. 2, pp. 5–8, February 2015.

[14] Consortium for Elections and Political Process Strengthening, "Assessment of electoral preparations in the Democratic Republic of the Congo," Tech. Rep., CEPPS, May 2018.

[15] Unique Identification Authority of India, "Aadhaar dashboard," https://www.uidai.gov.in/aadhaar_dashboard/, Last accessed: February 12, 2020.

[16] A. Das, C. Galdi, H. Han, R. Ramachandra, J.-L. Dugelay, and A. Dantcheva, "Recent advances in biometric technology for mobile devices," in *International Conference on Biometrics Theory, Applications and Systems (BTAS)*. October 2018, pp. 1–11, IEEE.

[17] A. Rattani, R. Derakhshani, and A. Ross, *Selfie Biometrics: Advances and Challenges*, Springer, 2019.

[18] M. I. Gofman, S. Mitra, T.-H. K. Cheng, and N. T. Smith, "Multimodal biometrics for enhanced mobile device security," *Communications of the ACM*, vol. 59, no. 4, pp. 58–65, April 2016.

[19] H. Aronowitz, M. Li, O. Toledo-Ronen, S. Harary, A. Geva, S. Ben-David, A. Rendel, R. Hoory, N. Ratha, S. Pankanti, and D. Nahamoo, "Multi-modal biometrics for mobile authentication," in *IEEE International Joint Conference on Biometrics*. September 2014, pp. 1–8, IEEE.

[20] M. Gofman, S. Mitra, K. Cheng, and N. Smith, "Quality-based score-level fusion for secure and robust multimodal biometrics-based authentication on consumer mobile devices," in *International Conference on Software Engineering Advances (ICSEA)*. November 2015, pp. 274–276, IARIA.

[21] National Institute of Standards and Technology, "Face Recognition Vendor Test (FRVT) 1:N Evaluation," https://www.nist.gov/programs-projects/face-recognition-vendor-test-frvt-1n-2018-evaluation, October 2017, Last accessed: February 12, 2020.

[22] National Institute of Standards and Technology (NIST), "Iris Exchange (IREX)," https://www.nist.gov/programs-projects/iris-exchange-irex-overview, 2018, Last accessed: February 12, 2020.

[23] C. Burt, "DHS S&T biometric technology rally results suggest face best for fast processing," https://www.biometricupdate.com/201909/dhs-st-biometric-technology-rally-results-suggest-face-best-for-fast-processing, September 2019, Last accessed: February 12, 2020.

[24] P. Drozdowski, C. Rathgeb, and C. Busch, "Computational workload in biometric identification systems: An overview," *IET Biometrics*, vol. 8, no. 6, pp. 351–368, November 2019.

[25] J. E. Gentile, N. Ratha, and J. Connell, "An efficient, two-stage iris recognition system," in *International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. September 2009, pp. 211–215, IEEE.

[26] S. Billeb, C. Rathgeb, M. Buschbeck, H. Reininger, and K. Kasper, "Efficient two-stage speaker identification based on universal background models," in *International Conference of the Biometrics Special Interest Group (BIOSIG)*. September 2014, pp. 1–6, IEEE.

[27] A. Pflug, C. Rathgeb, U. Scherhag, and C. Busch, "Binarization of spectral histogram models: An application to efficient biometric identification," in *International Conference on Cybernetics (CYBCONF)*. June 2015, pp. 501–506, IEEE.

[28] P. Drozdowski, C. Rathgeb, and C. Busch, "Turning a vulnerability into an asset: Accelerating facial identification with morphing," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. May 2019, pp. 2582–2586, IEEE.

[29] M. Elhoseny, E. Essa, A. Elkhateb, A. E. Hassanien, and A. Hamad, "Cascade multimodal biometric system using fingerprint and iris patterns," in *International Conference on Advanced Intelligent Systems and Informatics (AISI)*. August 2017, pp. 590–599, Springer.

[30] C. Gehrmann, M. Rodan, and N. Jönsson, "Metadata filtering for user-friendly centralized biometric authentication," *EURASIP Journal on Information Security*, vol. 2019, no. 1, pp. 7, June 2019.

[31] A. Dantcheva, P. Elia, and A. Ross, "What else does your biometric data reveal? A survey on soft biometrics," *Transactions on Information Forensics and Security (TIFS)*, vol. 11, no. 3, pp. 441–467, March 2016.

[32] A. Ross, K. Nandakumar, and A. K. Jain, *Handbook of multibiometrics*, Springer, 2006.

[33] ISO/IEC JTC1 SC37 Biometrics, *ISO/IEC TR 24722:2015. Information technology – Biometrics – Multimodal and other multibiometric fusion*, 2 edition, December 2015.

[34] A. Jain and A. Ross, "Fingerprint mosaicking," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. May 2002, vol. 4, pp. IV–4064–IV–4067, IEEE.

[35] G. P. Kusuma and C.-S. Chua, "PCA-based image recombination for multimodal 2D + 3D face recognition," *Image and Vision Computing*, vol. 29, no. 5, pp. 306–316, April 2011.

[36] V. Kanhangad, A. Kumar, and D. Zhang, "Contactless and pose invariant biometric identification using hand surface," *Transactions on Image Processing (TIP)*, vol. 20, no. 5, pp. 1415–1424, May 2011.

[37] X. Yan, W. Kang, F. Deng, and Q. Wu, "Palm vein recognition based on multi-sampling and feature-level fusion," *Neurocomputing*, vol. 151, no. 2, pp. 798–807, March 2015.

[38] R. Snelick, M. Indovina, J. Yen, and A. Mink, "Multimodal biometrics: issues in design and testing," in *International Conference on Multimodal Interfaces*. November 2003, pp. 68–72, ACM.

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

[40] A. Abaza and A. Ross, "Quality based rank-level fusion in multibiometric systems," in *International Conference on Biometrics: Theory, Applications, and Systems (BTAS)*. September 2009, pp. 1–6, IEEE.

[41] A. Kumar and S. Shekhar, "Personal identification using multibiometrics rank-level fusion," *Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 41, no. 5, pp. 743–752, September 2011.

[42] S. Prabhakar and A. K. Jain, "Decision-level fusion in fingerprint verification," *Pattern Recognition*, vol. 35, no. 4, pp. 861–874, April 2002.

[43] P. P. Paul, M. L. Gavrilova, and R. Alhajj, "Decision fusion for multimodal biometrics using social network analysis," *Transactions on Systems, Man, and Cybernetics: Systems*, vol. 44, no. 11, pp. 1522–1533, November 2014.

[44] P. Radu, K. Sirlantzis, G. Howells, F. Deravi, and S. Hoque, "A review of information fusion techniques employed in iris recognition systems," *International Journal of Advanced Intelligence Paradigms*, vol. 4, no. 3/4, pp. 211–240, February 2012.

[45] L. M. Dinca and G. P. Hancke, "The fall of one, the rise of many: A survey on multi-biometric fusion methods," *IEEE Access*, vol. 5, pp. 6247–6289, April 2017.

[46] ISO/IEC JTC1 SC37 Biometrics, *ISO/IEC 2382-37:2017. Information technology – Vocabulary – Part 37: Biometrics*, International Organization for Standardization and International Electrotechnical Committee, 2 edition, February 2017.

[47] ISO/IEC JTC1 SC37 Biometrics, *ISO/IEC 19795-1:2006. Information Technology – Biometric Performance Testing and Reporting – Part 1: Principles and Framework*, International Organization for Standardization and International Electrotechnical Committee, April 2006.

[48] J. Daugman, "Biometric decision landscapes," Tech. Rep. UCAM-CL-TR-482, University of Cambridge - Computer Laboratory, January 2000.

[49] F. Hao, J. Daugman, and P. Zielinski, "A fast search algorithm for a large fuzzy database," *Transactions on Information Forensics and Security (TIFS)*, vol. 3, no. 2, pp. 203–212, June 2008.

[50] N. K. Ratha, K. Karu, S. Chen, and A. K. Jain, "A real-time matching system for large fingerprint databases," *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 18, no. 8, pp. 799–813, August 1996.

[51] P. Drozdowski, D. Fischer, C. Rathgeb, C. Schiel, and C. Busch, "Database binning and retrieval in multi-fingerprint identification systems," in *International Workshop on Information Forensics and Security (WIFS)*. December 2018, pp. 1–7, IEEE.

[52] A. Ross and M. S. Sunder, "Block based texture analysis for iris classification and matching," in *Conference on Computer Vision and Pattern Recognition - Workshops (CVPRW)*. June 2010, pp. 30–37, IEEE.

[53] A. Pflug, C. Busch, and A. Ross, "2D ear classification based on unsupervised clustering," in *International Joint Conference on Biometrics (IJCB)*. September 2014, pp. 1–8, IEEE.

[54] H. Proença, "Iris biometrics: Indexing and retrieving heavily degraded data," *Transactions on Information Forensics and Security (TIFS)*, vol. 8, no. 12, pp. 1975–1985, December 2013.

[55] C. Rathgeb, F. Breitinger, H. Baier, and C. Busch, "Towards bloom filter-based indexing of iris biometric data," in *International Conference on Biometrics (ICB)*. May 2015, pp. 422–429, IEEE.

[56] R. Cappelli, M. Ferrara, and D. Maltoni, "Fingerprint indexing based on minutia cylinder-code," *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 33, no. 5, pp. 1051–1057, May 2011.

[57] V. D. Kaushik, J. Umarani, A. K. Gupta, and P. Gupta, "An efficient indexing scheme for face database using modified geometric hashing," *Neurocomputing*, vol. 116, pp. 208–221, 2013.

[58] J. E. Gentile, N. Ratha, and J. Connell, "SLIC: short-length iris codes," in *International Conference on Biometrics: Theory, Applications, and Systems (BTAS)*. 2009, pp. 1–5, IEEE.

[59] C. Rathgeb, A. Uhl, and P. Wild, "On combining selective best bits of iris-codes," in *European Workshop on Biometrics and Identity Management (BioID)*. March 2011, pp. 227–237, Springer.

[60] M.-H. Lim, A. B. J. Teoh, and J. Kim, "Biometric feature-type transformation: Making templates compatible for secret protection," *Signal Processing Magazine*, vol. 32, no. 5, pp. 77–87, September 2015.

[61] P. Drozdowski, F. Struck, C. Rathgeb, and C. Busch, "Benchmarking binarisation schemes for deep face templates," in *International Conference on Image Processing (ICIP)*. October 2018, pp. 191–195, IEEE.

[62] C. Rathgeb, F. Breitinger, C. Busch, and H. Baier, "On application of Bloom filters to iris biometrics," *IET Biometrics*, vol. 3, no. 4, pp. 207–218, December 2014.

[63] N. Damer, P. Terhörst, A. Braun, and A. Kuijper, "Efficient, accurate, and rotation-invariant iris code," *Signal Processing Letters*, vol. 24, no. 8, pp. 1233–1237, August 2017.

[64] P. Schuch, "Survey on features for fingerprint indexing," *IET Biometrics*, vol. 8, no. 1, pp. 1–13, January 2019.

[65] H. Proença and J.C. Neves, "Iris biometric indexing," in *Iris and Periocular Biometric Recognition*, pp. 101–124. Institution of Engineering and Technology, July 2017.

[66] I. Kavati, M. Prasad, and C. Bhagvati, "Search space reduction in biometric databases: a review," in *Computer Vision: Concepts, Methodologies, Tools, and Applications*, pp. 1600–1626. IGI Global, 2018.

[67] M. Gomez-Barrero, C. Rathgeb, K. B. Raja, R. Raghavendra, and C. Busch, "Biometric symmetry: Implications on template protection," in *European Signal Processing Conference (EUSIPCO)*. August 2017, pp. 941–945, IEEE.

[68] A. Kumar and K. Wang, "Identifying humans by matching their left palmprint with right palmprint images using convolutional neural network," in *International Workshop on Deep Learning and Pattern Recognition*. July 2016, SPIE.

[69] B. Ulery, A. Hicklin, C. Watson, W. Fellner, and P. Hallinan, "Studies of biometric fusion," Tech. Rep. NISTIR 7346, National Institute of Standards and Technology, September 2006.

[70] A. Adler, R. Youmaran, and S. Loyka, "Towards a measure of biometric feature information," *Pattern Analysis and Applications*, vol. 12, no. 3, pp. 261–270, September 2009.

[71] J. Daugman and C. Downing, "Searching for doppelgängers: assessing the universality of the IrisCode impostors distribution," *IET Biometrics*, vol. 5, no. 2, pp. 65–75, June 2016.

[72] P. Grother, M. Ngan, and K. Hanaoka, "Ongoing face recognition vendor test (FRVT) part 2: Identification," Tech. Rep. NISTIR 8238, National Institute of Standards and Technology, November 2018.

[73] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2015, pp. 815–823, IEEE.

[74] Y. Tang, F. Gao, J. Feng, and Y. Liu, "FingerNet: An unified deep network for fingerprint minutiae extraction," in *International Joint Conference on Biometrics (IJCB)*. October 2017, pp. 108–116, IEEE.

[75] R. Važan, "SourceAFIS – opensource fingerprint matcher," https://sourceafis.machinezoo.com/, 2019, Last accessed: February 12, 2020.

[76] N. Miura, A. Nagasaka, and T. Miyatake, "Extraction of finger-vein patterns using maximum curvature points in image profiles," *Transactions on Information and Systems*, vol. 90, no. 8, pp. 1185–1194, August 2007.

[77] Z. Guo and R. W. Hall, "Parallel thinning with two-subiteration algorithms," *Communications of the ACM*, vol. 32, no. 3, pp. 359–373, March 1989.

[78] M. Aastrup Olsen, D. Hartung, C. Busch, and R. Larsen, "Convolution approach for feature detection in topological skeletons obtained from vascular patterns," in *Workshop on Computational Intelligence in Biometrics and Identity Management (CIBIM)*. April 2011, pp. 163–167, IEEE.

[79] H. Xu, R. N. J. Veldhuis, T. A. M. Kevenaar, A. H. M. Akkermans, and A. M. Bazen, "Spectral minutiae: A fixed-length representation of a minutiae set," in *Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. June 2008, pp. 1–6, IEEE.

[80] D. Hartung, M. A. Olsen, H. Xu, and C. Busch, "Spectral minutiae for vein pattern recognition," in *International Joint Conference on Biometrics (IJCB)*. October 2011, pp. 1–7, IEEE.

[81] D. Hartung, M. A. Olsen, H. Xu, H. T. Nguyen, and C. Busch, "Comprehensive analysis of spectral minutiae for vein pattern recognition," *IET Biometrics*, vol. 1, no. 1, pp. 25–36, March 2012.

[82] T. K. Ho, J. J. Hull, and S. N. Srihari, "Decision combination in multiple classifier systems," *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 1, no. 1, pp. 66–75, January 1994.

[83] A. Jain, B. Klare, and A. Ross, "Guidelines for best practices in biometrics research," in *International Conference on Biometrics (ICB)*. May 2015, pp. 541–545, IEEE.

[84] D. Black, *The theory of committees and elections*, Cambridge University Press, 1958.

Pawel Drozdowski is a Researcher with the Faculty of Computer Science, Hochschule Darmstadt (HDA), Germany. He currently pursues a Ph.D. degree at Norwegian Biometrics Laboratory (NBL) within the Norwegian University of Science and Technology (NTNU). He co-authored over 15 technical publications in the field of biometrics. He won the Best Student Paper Runner-Up Award (WIFS'18) and Best Poster Award (BIOSIG'19). He is a member of the European Association for Biometrics (EAB) and serves as a reviewer for IET Biometrics (IET BMT) and IEEE Transactions on Information Forensics and Security (IEEE TIFS) journals, as well as the International Conference of the Biometrics Special Interest Group (BIOSIG).

Dr. Christian Rathgeb is a Senior Researcher with the Faculty of Computer Science, Hochschule Darmstadt (HDA), Germany. He is a Principal Investigator in the National Research Center for Applied Cybersecurity ATHENE. His research includes pattern recognition, iris and face recognition, security aspects of biometric systems, secure process design and privacy enhancing technologies for biometric systems. He co-authored over 100 technical publications in the field of biometrics. He is a winner of the EAB - European Biometrics Research Award 2012, the Austrian Award of Excellence 2012, Best Poster Paper Awards (IJCB'11, IJCB'14, ICB'15) and the Best Paper Award Bronze (ICB'18). He is a member of the European Association for Biometrics (EAB), a Program Chair of the International Conference of the Biometrics Special Interest Group (BIOSIG) and a editorial board member of IET Biometrics (IET BMT). He has served for various program committees and conferences (*e.g.* ICB, IJCB, BIOSIG, IWBF) and journals as a reviewer (e.g. IEEE TIFS, IEEE TBIOM, IET BMT).

Benedikt-Alexander Mokroß Benedikt-Alexander Mokroß is a head of research and developer at iCOGNIZE GmbH, Germany, where he specialises in computer vision and vascular biometrics. He holds a master's degree in computer science from the Hochschule Darmstadt, Germany. He received one national and one faculty award for his master's thesis. Furthermore, he serves as a reviewer for IEEE Transactions on Information Forensics and Security (IEEE TIFS) journals, as well as the International Conference of the Biometrics Special Interest Group (BIOSIG).

Prof. Dr. Christoph Busch is member of the Department of Information Security and Communication Technology (IIK) at the Norwegian University of Science and Technology (NTNU), Norway. He holds a joint appointment with the computer science faculty at Hochschule Darmstadt (HDA), Germany. Further he lectures Biometric Systems at Technical University of Denmark (DTU) since 2007. Christoph Busch co-authored more than 400 technical publications and has been a speaker at international conferences. He served for various program committees (NIST IBPC, ICB, ICHB, BSI-Congress, GI-Congress, DACH, WEDELMUSIC, EUROGRAPHICS) and served for several conferences, journals and magazines as reviewer (*e.g.* ACM-SIGGRAPH, ACM-TISSEC, IEEE CG&A, IEEE Transactions on Signal Processing, on Information Forensics and Security, on Pattern Analysis and Machine Intelligence and the Elsevier Journal Computers & Security). He is also an appointed member of the editorial board of the IET journal on Biometrics and of IEEE TIFS journal. Furthermore, on behalf of Fraunhofer, he chairs the biometrics working group of the TeleTrusT association as well as the German standardization body on Biometrics (DIN-NIA37). He is convenor of WG3 in ISO/IEC JTC1 SC37 on Biometrics and active member of CEN TC 224 WG18.