IET Research Journals

The Institution of Engineering and Technology

# Computational Workload in Biometric Identification Systems: An Overview

ISSN 1751-8644
doi: 0000000000
www.ietdl.org

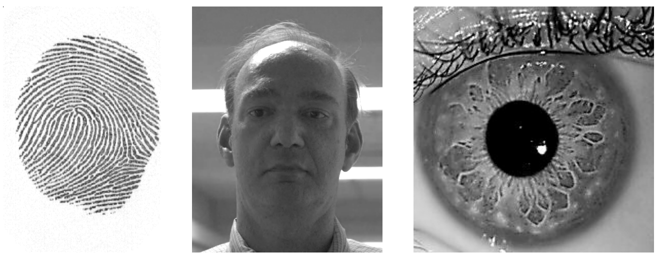
*P. Drozdowski*[1,2] *C. Rathgeb*[1] *C. Busch*[1]

[1] *da/sec – Biometrics and Internet Security Research Group, Hochschule Darmstadt, Germany*
[2] *NBL – Norwegian Biometrics Laboratory, Norwegian University of Science and Technology, Gjøvik, Norway*
* *E-mail: pawel.drozdowski@h-da.de*

**Abstract:** Computational workload is one of the key challenges in biometric identification systems. The naïve retrieval method based on an exhaustive search becomes impractical with the growth of the number of the enrolled data subjects. Consequently, in recent years, many methods with the aim of reducing or optimising the computational workload, and thereby speeding-up the identification transactions, in biometric identification systems have been developed. In this article, a taxonomy for conceptual categorisation of such methods is presented, followed by a comprehensive survey of the relevant academic publications, including computational workload reduction and software/hardware-based acceleration. Lastly, the pertinent technical considerations and trade-offs of the surveyed methods are discussed, along with an industry perspective, and open issues/challenges in the field.

## 1 Introduction

The interest in biometric technologies has been steadily growing in recent years, as evidenced by various market value studies [1–3] and numbers of scientific publications in the area. Many states have utilised biometric technologies for purposes such as forensic investigations and law enforcement, border crossing entry-exit tracking, national citizen inventory (ID systems), and voter registration. By far the largest biometric deployment to date is the Indian Aadhaar national ID system, which, at the time of this writing, accommodates 1.3 billion enrolled subjects – almost the entire Indian population.



(a) Fingerprint (b) Face (c) Iris

**Fig. 1**: Example images of some biometric characteristics commonly used in large-scale biometric identification systems (taken from the MCYT, FRGC, and IITD databases)

Table 1 gives an overview of this and several other examples of operational and planned large-scale biometric systems. The table is non-exhaustive; instead, it seeks to highlight the diversity of the used biometric characteristics, the system purposes, and the geographical locations of some of the largest biometric systems around the world. In figure 1, example images of biometric characteristics most commonly used in large-scale biometric identification systems are shown.

Biometric systems can operate in a broad variety of ways. Two such ways (as defined in the ISO/IEC international standards [4, 5]) 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". Two main scenarios can be distinguished in this case: **closed-set** identification, for which all potential users are enrolled in the system, and **open-set** identification, for which some potential users are not enrolled in the system.

Naturally, the second case (*i.e.* open-set identification, as well as the duplicate-enrolment check) is the most interesting and challenging from the practical point of view for the aforementioned real-world applications. Unfortunately, in the worst case, an exhaustive search (*i.e.* comparing a probe against all the enrolled subjects) is required in order to reach a decision. This naïve approach quickly runs into two non-trivial problems:

**Computational costs** As the number of enrolled subjects increases, the system response times become gradually slower, thus requiring optimisations and/or investment into larger hardware architectures.

**False positives costs** The probability of at least one false positive ($P_N$) occurring in a identification scenario is: $P_N = 1 - (1 - P_1)^N$, where $N$ is the number of enrolled subjects and $P_1$ the false positive probability of a one-to-one template comparison. This relationship is very demanding – even for systems which perform extremely well in verification mode (*i.e.* have low $P_1$), the value of $P_N$ very quickly becomes unacceptably high, as the number of enrolled subjects $N$ increases (see [17]). Note, that this equation ignores other system errors, *e.g.* the failure-to-acquire rate and assumes that at a given threshold all subjects have the same false-match-rate, which likely is not the case. Nonetheless, it is a useful approximation for illustrating this challenge of biometric identification systems.

In a recent interview [18], Daugman, the pioneer of iris recognition (see [19]), has stated that performing accurate and efficient biometric identification (*i.e.* not by an exhaustive search) is one of the important, unsolved issues in the biometrics field in general. Substantial research effort has been devoted to development of workload reduction methods, which seek to alleviate the aforementioned issues (especially the computational cost, since the biometric performance can also be improved through other means, such as increasing data quality and information fusion). Since the overall computational costs in a biometric identification scenario are dominated by

**Table 1** Examples of currently operational and planned large-scale biometric identification systems around the world

| Status | System | Location | Characteristic(s) | Subjects | Purpose |
|---|---|---|---|---|---|
| Operational | Aadhaar [6, 7] | India | Fingerprint and iris (operational), face (potential future use) | 1.3 billion | National ID |
| | EURODAC [8, 9] | EU | Fingerprint | 7 million | Border-control |
| | IDENT/US-VISIT [10] | USA | Fingerprint (operational), face and iris (pilots ongoing) | 200 million | Entry-exit |
| | CODIS [11] | USA | DNA | 17.5 million | Law enforcement |
| | CENI [12] | DR Congo | Fingerprint | 46 million | Voter registration |
| Planned | HART [13, 14] | USA | Fingerprint, Face, Iris | (expected) up to 500 million | Entry-exit |
| | EES [15, 16] | EU | Fingerprint, Face | (expected) up to 200 million | Entry-exit |

performing the biometric comparisons, most approaches are aimed at that step in the system pipeline. Specialised data representations and search algorithms are utilised to reduce the computational effort required for a single template comparison, and/or to reduce the overall number of required template comparisons. However, biometric data exhibits certain properties, which present challenges or outright invalidate many traditional approaches aimed at retrieval speed improvement:

**Ordering** Biometric data has no inherent logical ordering (as opposed to, for example, text data, which can be indexed *e.g.* alphabetically).

**Within-subject variability** The samples acquired from the same subject (even within short time intervals) are almost never exactly identical (*i.e.* they are fuzzy). Some variations are nearly inevitable due to numerous noise sources in the acquisition process (*e.g.* distance and angle from the sensor, environmental conditions, occlusions *etc.*).

**Dimensionality** The biometric feature vectors are typically high-dimensional; many search and indexing methods perform poorly in such spaces [20].

Consequently, computational workload reduction methods tailored specifically to the particular properties of biometric data have been developed in recent years. Such methods will be surveyed in the following sections. For a general overview of search structures and algorithms used for fast similarity searches across various disciplines, the reader is referred to *e.g.* [21–28]. The reader is expected to possess certain background knowledge on biometric recognition systems in general and the typical algorithms used in their signal processing pipelines. For quick primers, the reader is referred to the encyclopedia of biometrics [29], as well as the renowned handbook series: [30] for biometrics in general and [31–34] specifically for fingerprint, face, iris, and vascular characteristics, respectively.

While previously there have been surveys on biometric workload reduction methods (*e.g.* [35] for iris and [36] for fingerprint), they tend to concentrate on particular methods and/or biometric characteristics, rather than the overall spectrum of available research. Although the emphasis of this article is on the *academic* research, a discussion from the industry perspective and the interplay between academia and industry are included. The main contributions of this article are thus as follows:

**Taxonomy** which conceptually categorises the computational workload reduction methods in biometric identification.

**Comprehensive survey** of the methods reported in the scientific literature. It is organised by the relevant concepts, rather than by biometric characteristics. Instead of concentrating on one biometric characteristic only, the (arguably) most popular ones (in terms of actual use in industry and scientific research interest) are surveyed.

**Discussion** of relevant technical considerations and trade-offs, along with an industry perspective, and open issues/challenges pertaining to this research field.

The remainder of this article is organised as follows: section 2 gives an overview of relevant background information; in particular it introduces and defines key concepts used throughout the article, as well as outlines the current methodologies for results reporting and issues associated therewith. Section 3 contains a comprehensive survey of the existing computational workload reduction approaches reported in the scientific literature, conceptually organised within the framework of the proposed taxonomy. Section 4 discusses the topic from the purely academic, as well as industrial perspective, and outlines open issues/challenges. A summary and concluding remarks are given in section 5.

## 2 Background

This section gives an overview of relevant background information. Subsection 2.1 contains a list and short descriptions of the pertinent concepts and nomenclature, whereas in subsection 2.2, the dilemma associated with biometric result reporting and benchmarking is outlined.

### *2.1 Concepts and Nomenclature*

Throughout this article, the nomenclature from the biometric vocabulary [4] and biometric performance testing and reporting [5] ISO/IEC international standards are used whenever applicable. However, as of this writing, many concepts relating to computational workload in biometric systems have not yet been put into standards by ISO/IEC (although efforts in this direction are ongoing, especially as some of the key standards are now/soon up for a revision). In this context, the present standards only defines the terms (quoted directly from the standards):

**Pre-selection algorithm** Referring to the "algorithm to reduce the number of templates that need to be matched in an identification search of the enrolment database".

**Pre-selection error** Referring to "the error that occurs when the corresponding enrolment template is not in the pre-selected subset of candidates when a sample from the same biometric characteristic on the same user is given".

**Baseline performance** Referring to the "performance of a biometric system in a reference evaluation environment".

Those terms are insufficient to capture the whole spectrum of issues and methods relevant in the aforementioned context. Therefore, several key concepts listed below are defined based on their actual use in the surveyed scientific literature:

**Baseline system** A state-of-the-art algorithm performing an exhaustive database search during a biometric identification transaction.

**Computational workload** The total computational effort of a single transaction (or a set of transactions) in a biometric (identification) system, including: number of intrinsic operations, execution time, memory and storage requirements.

**Computational workload reduction** The extent to which a method reduces the computational requirements (workload) of a biometric transaction (in a biometric identification system). See also subsection 2.2.

**Pre-filtering** (also "pre-selection", "cascading algorithms", "serial combination of algorithms", "guided search", "continuous classification") Computationally efficient, but somewhat inaccurate, comparator(s) are used to compare the biometric probe against the enrolled templates to produce successively smaller short-lists

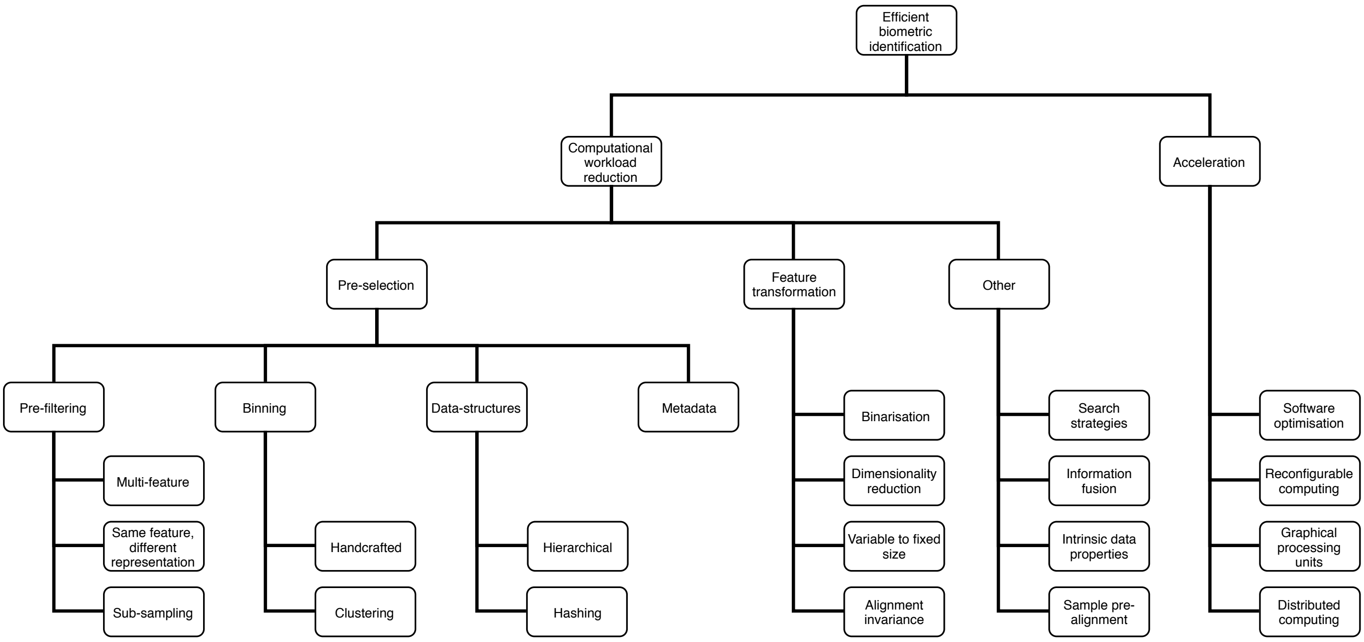



**Fig. 2**: Taxonomy of methods used for the purpose of speeding-up biometric identification

of candidate identities. In the end, the actual accurate, but computationally expensive, comparator is applied only to a fraction of the entries from the candidate short-list.

**Binning** (also "(discrete/exclusive) classification", "clustering") Splitting of the enrolment database into a number of subset (*i.e.* bins) based on coarse-level features. Those features can be tangible sample meta-data (*e.g.* sex, ethnicity, age) or based on intrinsic statistical properties of a template representation. During retrieval, the search space is reduced by only searching within the bins(s) most likely corresponding to the biometric probe.

**Data-structures** Organising the enrolment database to take advantage of efficient ordering principles (*e.g.* based on trees or fuzzy hashing), thus enabling searching in sub-linear/logarithmic time.

**Indexing** An often used umbrella term (in the biometric literature – *e.g.* a recent survey [36] and many individual publications) for all pre-selection methods (*i.e.* pre-filtering, binning, and data-structures). Simultaneously, it also has specific meaning outside the biometrics community. In order to avoid ambiguities, the term is not used in this article. Instead, the publications which present "indexing" methods, are assigned conceptually to one of the aforementioned categories.

**Feature transformation** The act of deriving additional features from a biometric template with the goal of attaining some desirable properties (*e.g.* smaller template size, ability to use a faster comparator, biometric sample alignment invariance).

**Acceleration** (hardware and/or software based) Utilisation of specialised hardware, hardware-software co-design, parallelism, distributed computing, and other methods in order to increase the efficiency/speed of a system compared to execution on standard CPUs.

In section 3, a taxonomy, which encompasses the abovementioned concepts and terms is presented.

### 2.2 Results Reporting

In subsequent subsections, tables which summarise the surveyed publications are presented. They include, among other matters, biometric performance and computational workload details. The metrics used for measuring biometric performance are well-defined and standardised [5]. The most relevant, in the context of this article, is the pre-selection error rate (complement of the hit rate; incidentally the hit rate is preferred in the vast majority of the works referenced later on in this article). While, in theory, this should make it possible to directly compare different methods, the reality is rather disappointing. First of all, some of the listed publications pre-date or ignore this standard, *i.e.* use a wide range of other metrics. Secondly, there inevitably exist other confounding issues and discrepancies in the experimental protocol, such as *e.g.* mode of operation (closed or open set), choice of dataset (hence, crucially, data quality), as well as size and partitioning thereof (*i.e.* training/testing partitions, number of biometric mated and non-mated comparison trials). Furthermore, at the time of this writing, metrics for measuring computational workload and its reduction are not standardised in any way whatsoever; many different metrics do appear in the scientific literature, for example:

- Penetration rate, which measures what fraction of the database is searched during a biometric identification transaction.
- Biometric template and/or model size, which determines how computationally expensive a single biometric comparison is.
- The fraction or percentage between the computational workload of a proposed system and a baseline system.
- Computational time, which measures the average execution time on some specific hardware configuration.

Additionally, it is often the case, that the publications present various parameter configurations with different trade-off spectra *etc.* for the proposed systems. It is therefore not always clear, which result to choose to present in a survey table, and how to select the single operational point which best encompasses all the aspects of the proposed systems. As such, the choices in this survey were made as follows:

1. If the authors have provided a single representative result (operational point) in the publication text (*e.g.* in the abstract or summary) for the biometric performance and/or computational workload, those values are taken directly.
2. Otherwise, a single operational point is chosen in good faith from the presented plots and tables. If possible, this is done based on what is commonly reported elsewhere in the literature, *e.g.* equal-error-rate or other recognised metric. For the sake of consistency, if results for multiple ranks (*e.g.* CMC curve) are available, rank-1 results are preferred.
3. Computational time results are not reproduced, since they depend on a specific hardware configuration (which is most likely obsolete anyway). Where possible, the relative speed-up between the baseline and the proposed method is (calculated and) reported.

Due to the aforementioned issues, directly comparing the results from the surveyed publications is problematic, if not impossible. Furthermore, different systems require different considerations and trade-offs w.r.t. the biometric performance and the computational workload, as well as additional matters such as user convenience, software and hardware infrastructure, financial costs, and others. Consequently, the readers interested in benchmarking and/or utilising the surveyed methods are strongly recommended to investigate the relevant publications by themselves in order to obtain full-picture information of the proposed methods along with the biometric performance and computational workload trade-offs associated therewith.

### 2.3 Feature Extraction

Extracting sufficiently discriminative features is a critical prerequisite for any biometric system. This is especially a concern in biometric identification systems, due to the significantly increased risk of false positive errors (see section 1). Over time, various general purpose and biometric characteristic specific feature extraction methods have been proposed and used in this context. However, comprehensively surveying and comparing those would tremendously extend the already significant scope of this article. Therefore, the reader interested in a detailed treatment of this subject is referred to a recently published comprehensive survey of general purpose texture based feature extraction methods [37], as well as the handbook series: [30] for biometrics in general and [31–34] specifically for fingerprint, face, iris, and vascular characteristics, respectively.

## 3 Computational Workload Reduction Approaches

In this section, the current state-of-the-art is presented. Firstly, the proposed taxonomy around which this section is structured is introduced and described below. Thereafter, a comprehensive survey of existing methods is given and put in the context of the taxonomy.

Figure 2 shows the proposed taxonomy under which the existing approaches to speeding-up the biometric identification can be categorised. Note, that in many cases the approaches can be combined into multi-level frameworks, *e.g.* a binning followed by tree-based hierarchical retrieval, implemented utilising hardware acceleration or pre-selection based on multiple levels of complementary features. Two main approaches to improving the computational efficiency of biometric identification can be distinguished: workload reduction (subsections 3.1 to 3.5) and acceleration (subsection 3.6). The latter does not reduce the computational workload *per se* – instead, it seeks to perform the same amount of computations in a more efficient manner (*e.g.* by utilising specialised hardware or optimising the software implementation). The goal of the former is to reduce the amount of computations necessary to perform a biometric identification transaction. For those approaches, two main categories can be distinguished: concentrating on reducing the penetration rate, the aim of the pre-selection approaches (subsections 3.1 to 3.3) is to narrow down the search space by taking advantage of auxiliary features, metadata, or search structures, which can be extracted or created from the samples. On the other hand, the goal of feature transformation approaches (subsection 3.4) is to reduce the computational cost of individual template comparisons, *e.g.* by reducing their dimensionality or utilising more computationally efficient comparators. The vast majority of the approaches can be assigned to one of those categories. The remaining few ones (subsection 3.5) are based *e.g.* on augmenting the search strategy of the retrieval algorithm or rely on certain intrinsic properties of specific biometric data.

This section is organised to facilitate selective reading: firstly, a very broad overview of the efficient biometric identification research areas has been given above by introducing and describing the proposed taxonomy. The following subsections' text outlines the relevant high-level concepts and ideas, while the tables contain more detailed information w.r.t. specific tools, algorithms, and datasets used, as well as the achieved results. Finally, the considerations and trade-offs associated with the different approach categories are discussed in subsection 4.1.

### 3.1 Pre-filtering

Figure 3 shows a conceptual overview of pre-filtering approaches, while table 2 summarises the surveyed methods.

*3.1.1 Multi-Feature:* The key idea behind the multi-feature approaches is the extraction of one or several auxiliary features, which in themselves do not have sufficient discriminative power for unique identification, but can nonetheless significantly reduce the search space (*i.e.* by acting as an index, which allows to determine a candidate short-list).

Auxiliary features such as orientation field, ridge density, local (minutiae-based) and global (*e.g.* fingerprint types, which have been in use for decades for the purposes of manual indexing of analog ten-fingerprint records with the Henry Classification System, see *e.g.* [72, 73], and subsection 3.2) can be extracted from fingerprint images; some of them also pertain to other characteristics, such as vascular and palmprint patterns. Several authors (*e.g.* [39–44, 46, 57]) utilise such coarse features as an index in a pre-filtering step. In other cases, the methods proposed in the scientific literature do not rely on specific, biometric characteristic-dependent features as above; instead, to create an index, they utilise general-purpose algorithms, such as texture extractors (*e.g.* [51, 56]), principal component analysis (*e.g.* [50]), and, more recently, deep learning (*e.g.* [52, 53]). It should be noted, that the pre-filtering can happen in a cascading manner, over two (*e.g.* [48, 49, 54]) or multiple (*e.g.* [38, 45, 58]) levels, which successively produce smaller candidate lists, or through direct application of information fusion strategies to the extracted features (*e.g.* [47]). However, an in-depth analysis and evaluation concerning which of the methods (cascades or fusion) performs better has not yet been reported in the scientific literature.

*3.1.2 Same feature, different representation:* The key idea behind this category of approaches is transformation of the original feature representation into a more compact one, whereby the computational costs of comparisons are decreased (often at the cost of losing some discriminative power). The compact templates can then be used to reduce the search space (*i.e.* by acting as an index, which allows to determine a candidate short-list).

Conceptually similar approaches, where binarised (see also subsection 3.4.1) and/or shortened feature vectors are used as an index in the pre-filtering step, have been proposed *e.g.* in [59–63] for iris, face, fingervein, voice, and ear, respectively.

The difference between the key idea in this and previous subsection is subtle – here, the same feature is used to create the index template (*e.g.* through binarisation), whereas in the multi-feature concept, additional features are extracted from the sample (*e.g.* through texture or keypoint descriptors or high level geometric features).

*3.1.3 Sub-sampling:* The key idea behind sub-sampling is to utilise partial information from the original feature vectors once or in an incremental manner to facilitate search space reduction via accurate early rejection of unlikely candidates. In other words, parts of the original feature vector itself act as an index in this case. This can be done trivially by deterministically or randomly selecting the partial information or, in more advanced approaches, by reorganising the feature vectors based on reliability and discriminative power (see *e.g.* [74]), as well as utilising other heuristics. In the literature, numerous conceptually similar approaches have been presented *e.g.* in [64–66, 68–71] for various biometric characteristics, including fingerprints, face, iris, and fingervein. In all the aforementioned publications, the computational workload is shown to be substantially reduced without causing degradation of the biometric performance. In [67] a more sophisticated approach, which relies on creating an auxiliary search guiding structure and an early search termination strategy, was presented with impressive results, albeit on proprietary data.

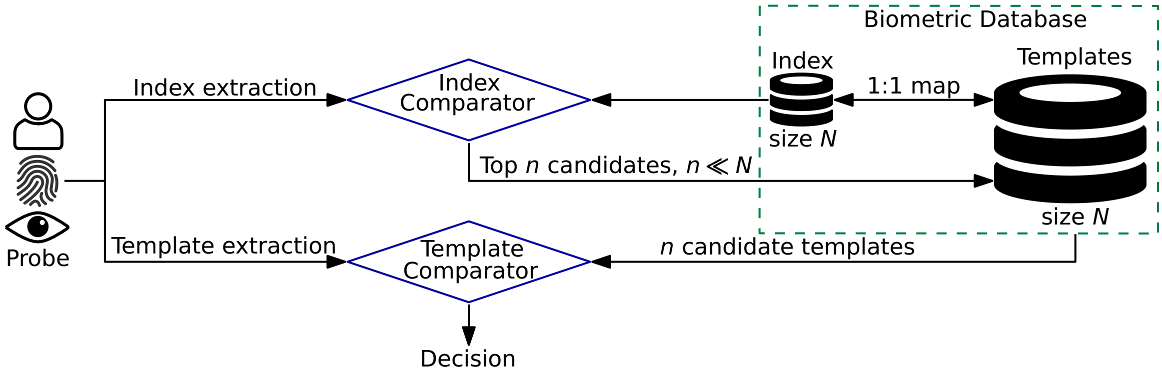



**Fig. 3**: Conceptual view of pre-filtering approaches

**Table 2** Pre-filtering approaches

| Taxonomy | Characteristic | Publication | Method | Database | Biometric Performance | Computational Workload |
|---|---|---|---|---|---|---|
| Multi-Feature | Fingerprint | Ratha *et al.* [38] | Metadata (only conceptual), fingerprint type, ridge density | NIST-9 subset | 80% accuracy, 10% reject rate | 12.5% search space |
| | | De Boer *et al.* [39] | Directional field, FingerCode, and minutiae triplets | FVC2000 | 100% hit rate | 18% penetration rate |
| | | Bhanu [40] | Minutiae triplets, geometric features | NIST SD4 | 85% hit rate | 10% penetration rate |
| | | Feng *et al.* [41] | Minutiae points and types, local ridge structures | FVC2002 | 100% hit rate | 22% penetration rate |
| | | Li *et al.* [42] | Ridge structure, symmetrical filters | NIST DB4 | 98% hit rate | 32.7% penetration rate |
| | | Liang *et al.* [43] | Minutiae neighbourhoods, Delaunay triangulation | FVC2002, FVC2004 | 100% hit rate | 18.1%, 20.9% penetration rate |
| | | Wang *et al.* [44] | 2D Fourier expansion coefficients | NIST SD 14 | 100% hit rate | 10% penetration rate |
| | | Feng *et al.* [45] | Fingerprint type, singular points, orientation field | NIST SD27 (search attempts), NIST SD4, SD14 and SD27 (background) | 97.3% accuracy | 39% penetration rate |
| | | Cappelli [46] | Ridge-line orientations and frequencies | NIST SD4, SD14, FVC2000 (DB2, DB3), FVC2002 (DB1) | 96.5%, 96.5%, 99%, 93.5%, 99% hit rate | 10% penetration rate |
| | | Paulino *et al.* [47] | Orientation field, ridge period, singular points, minutiae triplets, simplified MCC | NIST SD27 (search attempts), in-house (background) | 90.3% hit rate | 20% penetration rate |
| | Fingerprint, face | Gyaourova *et al.* [48, 49] | Index-codes from non-mated comparison trials, fusion | FERET, FRGC, WVU | 100% hit rate | 84% reduction |
| | Face | Mohanty *et al.* [50] | Affine approximation, PCA | FERET | — | 20-fold reduction |
| | | Chen *et al.* [51] | LBP, semantic codewords from metadata | LFW, Pubfig | 18.6%, 21.0% MAP | — |
| | | Wang *et al.* [52, 53] | Deep features and COTS | LFW, IJB-A | 0.25 MAP at 1% FAR, 0.175 MAP at 1% FAR | 30-fold time reduction |
| | Iris | Konrad *et al.* [54] | Rotationally invariant representation | CASIA-V1, CASIA-V3 Interval, MMU | 92% IR, 0% FAR; 89% IR, 0.85% FAR; 79% IR, 0.85% FAR | 70-80% time reduction |
| | | Gadde *et al.* [55] | BWT | CASIA-V3-Interval | 99.83% hit rate | 17.23 % penetration rate |
| | | Dey *et al.* [56] | Gabor energy features, multi-sample enrolment | Bath, CASIA-V3-Interval, CASIA-V4-Thousand, MMU2, WVU | 98.2%, 91.1%, 90.7%, 85.2%, 96% hit rate | 11.3%, 14.5%, 16.3%, 13.5%, 10.3% penetration rate |
| | Fingervein | Kavati *et al.* [57] | Delaunay triangulation | NTU NIR, NTU FIR | 100% hit rate | 17.99%, 11.75% penetration rate |
| | Palmprint | You *et al.* [58] | Global geometry, global texture energy, fuzzy “interest” line, local texture | In-house | 6.13% FRR at 11.77% FAR | 2-fold speed-up |
| Same feature, different representations | Face | Wu *et al.* [59] | Binary template pre-screening | In-house | Better than baseline | ∼10-fold reduction |
| | Iris | Gentile *et al.* [60] | Short-length Iris-Codes | MMU | 7% pre-selection error | 12-fold reduction |
| | Fingervein | Tang *et al.* [61] | Binary vein encoding | PKU | 98.4% hit rate | 250-fold time reduction |
| | Voice | Billeb *et al.* [62] | Binary template pre-screening | Unknown, text-independent | same or better than baseline | 95% speed-up |
| | Ear | Pflug *et al.* [63] | Binary template pre-screening | PolyU, UND-J2 | 100% hit rate | 30% penetration rate |
| Sub-sampling | Fingerprint | Iqbal *et al.* [64] | Incremental matching | FVC2002 | 99% hit rate | 26% penetration rate |
| | Fingerprint, palmprint | Chen *et al.* [65] | Incremental matching | THU, NIST SD 4, in-house | 90.4% IR, 85.3% IR, 75.3% IR | ∼50% reduction |
| | Face | Yi *et al.* [66] | Incremental matching | FERET, in-house | Same as baseline | 7.5-fold speed-up |
| | Iris | Hao *et al.* [67] | BGS, incremental matching | UAE | 0% FAR, 0.64% FRR | 0.006% penetration rate |
| | | Ross *et al.* [68] | Partial matching | UPOL | 0.62% EER | 10% of baseline |
| | | Hämmerle-Uhl *et al.* [69] | Partial matching | CASIA-V3 Interval | Same as baseline | 1 order of magnitude reduction |
| | | Rathgeb *et al.* [70] | Incremental matching | CASIA-V3 Interval | Same as baseline | 95% fewer bit comparisons |
| | Fingervein | Surbiryala *et al.* [71] | Partial matching | Combined 7 fingervein DBs | 8.05% pre-selection error | ∼3-fold reduction |

### *3.2 Binning*

Figure 4 shows a conceptual overview of binning approaches, while table 3 summarises the surveyed methods.

*3.2.1 Handcrafted:* Depending on the observed biometric characteristic, there exist classification approaches designed to reliably extract human understandable attributes from a biometric sample, *e.g.* sex or ethnicity for face, or fingerprint types. Such attributes are called “soft biometrics” (see *e.g.* [75] for a comprehensive survey).

Based on the global pattern formed by the ridge lines, fingerprints can be classified into a number of classes/types initially proposed by Galton [76] and Henry [72] (currently typically 4 or 5, *i.e.* whorls, right and left loops, and (tented) arches, sometimes extended with additional sub-types). Over time, numerous approaches to automated fingerprint type classification have been proposed (see *e.g.* [77, 78] for a comprehensive survey). The classification accuracy on data of reasonable quality is near-optimal; however, it tends to vary somewhat across the different fingerprint types. Binning based on fingerprint classes has been evaluated for single fingerprints by *e.g.* [38, 79] and for multi-instance data in [80, 81]. Attributes extracted from iris data can also be used in this manner. Conceptually similar systems are presented in [82], [83], and [84], where binning based on biometric characteristic-specific geometric/texture features is proposed for irs, palmprint, and palmvein data, respectively. In [85–87], it has been demonstrated, that ethnicity and gender information can be extracted from iris images. When reliably extracted, such features could be used for simple database binning. Binning based on iris colour has been performed *e.g.* in [88, 89]. Although the vast majority of the human population has brown eyes, for certain population groups, the eye colour can be used as a somewhat distinguishing soft biometric trait. Currently, all practical iris recognition systems operate within the near-infrared (NIR) light spectrum. In recent years, significant advances in the visible-wavelength (VW) iris recognition have been made, hence potentially making it an emerging technology. See *e.g.* [90] for an investigation of the reliability of the iris colour as a soft biometric trait. Facial region is a rich source of potential soft biometric attributes. In addition to simple approaches based on sex, age, or ethnicity classification, binning based on marks, scars, and tattoos has been proposed [91].

While the aforementioned attributes are not discriminative enough to be directly used in biometric identification, they allow for a relatively straightforward binning of biometric databases according to a predefined number of classes. In other words, the potential search space for a given biometric probe can be narrowed down to one (or a few) bin(s), thereby reducing the penetration rate, and hence the computational workload.

*3.2.2 Clustering:* Cluster analysis or clustering refers to the unsupervised or semi-supervised classification of patterns (*i.e.* feature vectors, data items, or observations) into groups (referred to as clusters), wherein the items are, in some sense, similar to each other.

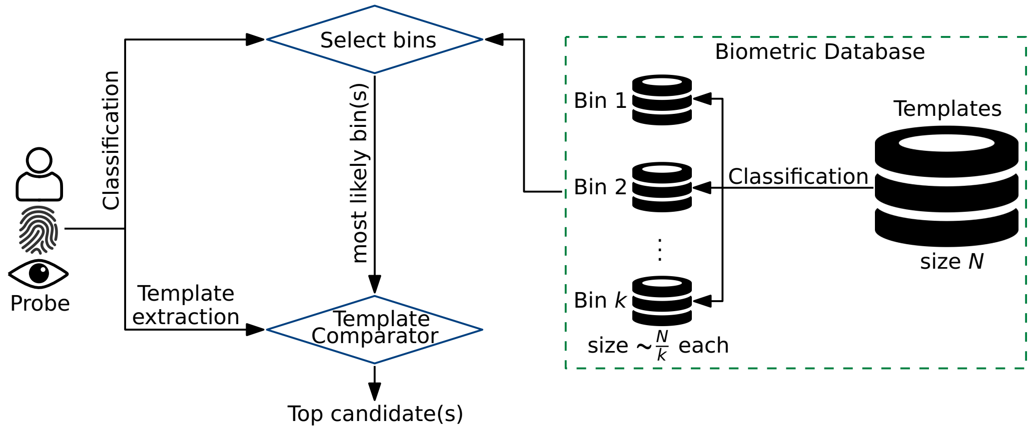



**Fig. 4**: Conceptual view of binning approaches

**Table 3** Binning approaches

| Taxonomy | Characteristic | Publication | Method | Database | Biometric Performance | Computational Workload |
|---|---|---|---|---|---|---|
| Handcrafted | Fingerprint | Zheng *et al.* [79] | Classification, coarse-level matching, class-jumping, SURF | NIST DB 4 | 100% hit rate | 15% penetration rate |
| | | Drozdowski *et al.* [81] | Fingerprint types, multi-instance, variable search order | NIST DB 9, in-house Bundeskriminalamt (BKA) DB | Same as an exhaustive search | 5-15% of an exhaustive search |
| | Face | Park *et al.* [91] | Facial marks, scars, and tattoos | PCSO (police mugshots) | 7.1%, 0.5% rank-1 accuracy loss | 7%,20% speed-up |
| | Iris | Yu *et al.* [82] | Box-counting, fractal dimensions | In-house | 1.72% pre-selection error | Less than 40% time |
| | | Puhan *et al.* [88] | Colour information in YCbCr space, set intersection | UBIRIS | 97% hit rate | 25% penetration rate |
| | | Zhao [89] | Average RGB colour components, set union | UBIRIS | 92.35% hit rate | 28.28% penetration rate |
| | Palmprint | Palla *et al.* [83] | Geometric features, codebook vectors, Voronoi regions | In-house | 100% hit rate rate | 30% penetration rate |
| | Palmvein | Zhou *et al.* [84] | Principal orientation features | PolyU, CASIA, in-house | 96.67%, 96.00%, 97.71% retrieval accuracy | 14.29%, 14.50%, 14.28% penetration rate |
| Clustering | Fingerprint | Germain *et al.* [92] | Minutiae triplets, ridge skeleton, Flash algorithm | In-house | 3.5% FNMR at 0.01% FMR | — |
| | | Ross *et al.* [93] | Delaunay triangulation, geometric and ridge features, k-means clustering | FVC2002, FVC2004 | 100% hit rate | ∼50% av. penetration rate |
| | | Liu *et al.* [94] | Orientation field, average ridge distance, k-means clustering | NIST-DB 4 | 95.8% hit rate | 20% penetration rate |
| | | Biswas *et al.* [95] | Curvature, minutiae geometry, k-means clustering | IBM proprietary | 90% rank-1 accuracy | 5-fold decrease |
| | | Iloanusi *et al.* [96, 97] | Minutiae quadruplets, k-means clustering | FVC2002, FVC2004 | 100% av. hit rate | ∼12% av. penetration rate |
| | Face | Perronnin *et al.* [98] | Expectation maximisation clustering, anchor modelling | FERET | ∼95% IR | 6-7-fold reduction |
| | | Chaari *et al.* [99] | Eigenfaces and Fisherfaces, k-means clustering | XM2VTS | 87.5% IR at rank-1 | 40% penetration rate |
| | | Klare *et al.* [100] | Spectral clustering, k-means and k-medoids clustering | LFW, PCSO | 85% IR | 50% reduction |
| | Iris | Mukherjee *et al.* [101] | Iris-Code, PCA, k-means clustering | CASIA-V3-Interval | 80% hit rate | 8% penetration rate |
| | | Ross *et al.* [68] | Statistical texture features, Principal Direction Divisive Partitioning | UPOL | 100% CCR | 3-5-fold reduction |
| | | Sun *et al.* [102] | Ordinal measures, hierarchical visual codebook, k-means clustering, SVM | CASIA Thousand | ∼2% EER | less than 30% |
| | | Nalla *et al.* [103] | Online dictionary learning, k-means clustering | UPOL | 100% CCR | 3-4-fold reduction |
| | Fingervein | Surbiryala *et al.* [71] | Maximum curvature, k-means clustering | Combined 7 fingervein DBs | 97.47% hit rate | 86.43% penetration rate |
| | | Raghavendra *et al.* [104] | Self Organizing Map neural network, k-means or k-medoids clustering, multi-cluster search | Combined 7 fingervein DBs | 92.42%; 99.02% hit rate | 42.48%; 52.88% penetration rate |
| | Palmprint and signature | Mhatre *et al.* [105] | K-means clustering | Unknown | 0% FRR, — FAR | 5% penetration rate |
| | Ear | Pflug *et al.* [106] | K-means clustering, texture descriptors | UND-J2, AMI, IITK | 3.11% pre-selection error rate | 31.7% penetration rate |

With applications across many different disciplines, k-means clustering is currently one of the most popular and effective algorithms used in data mining [107].

Likewise, in the surveyed literature, k-means clustering (and its various extensions/derivatives) is by far the most popular method, used in *e.g.* [71, 93–97, 99–106]. Other methods include *e.g.* multimap clustering [92], expectation maximisation clustering [98], and principal direction divisive partitioning [68]. Comparing the various clustering methods is out of scope for this article. For more details regarding this field of research, the reader is referred to surveys, *e.g.* [108, 109]. Generally, the approaches referenced in this subsection extract certain biometric characteristic-specific features (*e.g.* orientation field or Delaunay triangles for fingerprint, or general-purpose texture descriptors for iris) to facilitate the clustering or apply it directly with the feature vectors (*e.g.* minutiae points) themselves. As a result, the search space is separated into distinct bins, whereby during biometric identification, candidates only from the most likely one(s) are retrieved. Hence, the penetration rate (and thereby the computational workload) is significantly reduced.

### *3.3 Data-Structures*

Figure 5 shows a conceptual overview of hierarchical retrieval approaches, while table 4 summarises the surveyed methods. A multitude of methods, algorithms, and data-structures (whose detailed descriptions are out of scope for this article) has been used in the surveyed approaches. For a general introduction to on approximate searching, relevant concepts, and most commonly used data-structures, the reader is referred to existing surveys, *e.g.* [21, 22, 25] for theoretical, practical, and easily digestible perspectives, respectively.

*3.3.1 Hierarchical:* Approaches in this category are most often tree-based, most prominently utilising k-d trees (*e.g.* [111, 114, 115, 122]), b or b+ trees (*e.g.* [112, 113, 124, 125]), other tree-like search structures (*e.g.* [35, 101, 110, 116, 121, 123]), and forests thereof (*e.g.* [117–120]). The differences between the various types of used trees (some of which are each other's generalisations) are out of scope for this article; instead, the reader is referred to *e.g.* [138, 139]. The key idea is to create a search structure, which repeatedly partitions the data (*i.e.* the search space – the enrolment database) into successively smaller subsets. For this partitioning, the highly discriminative (and high-dimensional) feature vectors themselves and/or the more coarse auxiliary features can be used. By doing so, sub-linear or even logarithmic lookup complexity can be achieved, thereby substantially reducing the computational workload of biometric identification.

*3.3.2 Hashing:* Hashing makes it possible to map the highly-dimensional biometric feature vectors into compact hashtables or similar data-structures, which facilitate efficient retrieval. Since biometric data is inherently fuzzy (recall section 1), many traditional hashing approaches are not suitable. Nevertheless, there exist methods, which can deal with fuzzy data. One of such method is

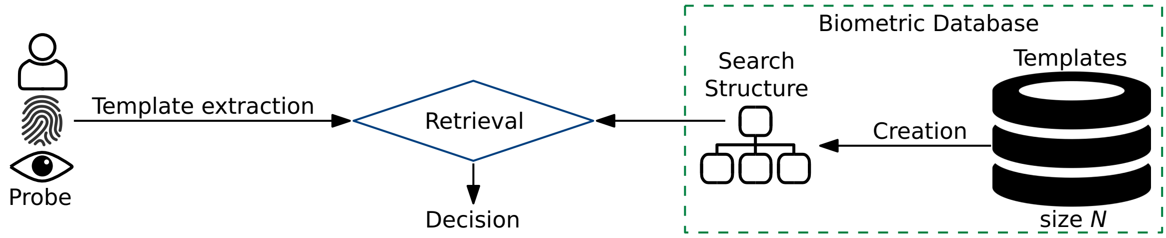



**Fig. 5**: Conceptual view of data-structures approaches

**Table 4** Data-structures approaches

| Taxonomy | Characteristic | Publication | Method | Database | Biometric Performance | Computational Workload |
|---|---|---|---|---|---|---|
| Hierarchical | Fingerprint | Mansukhani *et al.* [110] | Local minutiae neighbourhoods, unbalanced tree | FVC2002, FVC2004 | 81% accuracy | Almost constant w.r.t. enrolment DB size |
| | Face | Dewangan *et al.* [111] | SURF, kd-tree | FERET, FRGC, CalTech | 95.57%, 97.00%, 92.31% hit rate | 7.90%, 12.55% and 23.72% penetration rate |
| | Iris | Mukherjee *et al.* [101] | Blockwise texture SPLDH, tree-like structure | CASIA-V3-Interval | 84% hit rate | 30% penetration rate |
| | | Mehrotra *et al.* [112] | DCT, subband coding, energy histogram, b-tree | CASIA Interval, BATH, IITK | 95% hit rate | 25% penetration rate |
| | | Khalaf *et al.* [113] | DCT, DWT, SVD, subband coding, energy histogram, b-tree | CASIA Interval, BATH, IITK | ∼97.5%,∼97.5%, 95% hit-rate | 20% penetration rate |
| | | Jayaraman *et al.* [114] | Iris colour, SURF, kd-tree | UBIRISv2, UPOL | 98.7%, 98.5% av. hit rate | 7.5%, 1.5% av. penetration rate |
| | | Barbu *et al.* [115] | HOG, kd-tree | UPOL | 85% precision and recall | — |
| | | Rathgeb *et al.* [116] | Bloom filters, binary search trees | IITD | same or better than baseline | 6% penetration rate |
| | | Drozdowski *et al.* [117] | Bloom filters, binary search trees, multi-instance fusion | Combination of 4 iris datasets | 99.41% TPIR at 0.01% FPIR | <1% of baseline |
| | | Drozdowski *et al.* [118] | Bloom filters, binary search trees | Combination of 4 iris datasets | 98% TPIR at 0.1% FPIR | 10% of baseline |
| | | Damer *et al.* [119] | LSH-forest | ISYN1 | 99.85% single instance, 99.99% multi instance hit rate | 0.4% penetration rate |
| | | Damer *et al.* [120] | General Borda count, LSH-forest, multi-instance | ISYN1 | >99.5% hit rate | 0.1% penetration rate |
| | | Proença *et al.* [35, 121] | Multi-resolution decomposition, n-ary trees | CASIA-V4-Thousand, UBIRISv2 | 95% hit rate | 20%, 80% penetration rate |
| | Iris, Signature, Face, Ear | Jayaraman *et al.* [122] | Dimensionality reduction, feature-level fusion, kd-tree | IITK | 97.33% hit-rate at 0.66% FRR | — |
| | Fingervein | Wang *et al.* [123] | Local textons, vocabulary tree | PolyU, SDUMLA, MMCBNU, FV-USM | ∼99% hit rate at rank-5 | Up to 5-fold speedup |
| | Palmprint | Mhatre *et al.* [124] | Geometric features, spatial hashing, b-tree | unknown | 0% FRR, — FAR | 8.86% penetration rate |
| | Ear | Gupta *et al.* [125] | Division into quadrants, wavelet decomposition, b-tree | IITK | 95.8% accuracy | 34% penetration rate |
| Hashing | Fingerprint | Shuai *et al.* [126] | SIFT, LSH | FVC2000, FVC2002 | 98%, 96% hit rate | 10% penetration rate |
| | | He *et al.* [127] | SIFT, SURF, DAISY, LSH | FVC2000, FVC2002 | 99%, 90% hit rate | 10% penetration rate |
| | | Capelli *et al.* [128] | MCC, LSH, voting | NIST SD4, 14, FVC2000, 2002 | 95% hit rate | <10% penetration rate |
| | | Yuan *et al.* [129] | Minutiae triplets, two-level hashtable | FVC2000, 2002, 2004 | 100% hit rate | 22%, 9.9%, 11.7% av. penetration rate |
| | | Wang *et al.* [130] | MCC, Markov random field theory, geometric dictionary | FVC2002 DB1 | 100% hit rate | 10% penetration rate |
| | | Li *et al.* [131] | MCC, binarisation, LSH | FVC2002, FVC2004, FVC2006 | 7.5%, 22.5%, 4% pre-selection error rate | 10%, 10%, 5% penetration rate |
| | Face | Kaushik *et al.* [132] | SURF, geometric hashing, voting | FERET | 100% hit rate | 4% penetration rate |
| | Iris | Mehrotra *et al.* [133] | SIFT, geometric hashing, voting | BATH, CASIA-V3-Interval, IITK, UBIRIS | 98.29%, 98.55%, 99.61%, 97.57% EER | Order of magnitude faster than baseline |
| | | Rathgeb *et al.* [134] | Iris texture hashes, Karnaugh map | CASIA-V3 Interval | 90% accuracy | 3% of baseline |
| | | Jayaraman *et al.* [135] | Iris-Code, LSH, voting | CASIA-V3-Interval | 94.07% hit rate | 10.63% penetration rate |
| | | Panda *et al.* [136] | SIFT, geometric hashing | CASIA-V3-Interval, UBIRISv1 | 98.25%, 97.62% accuracy | ∼75% of baseline time |
| | Palmprint | Badrinath *et al.* [137] | SURF, geometric hashing | IITK, CASIA, PolyU | 100% hit rate | 22.5%, 22.8%, 31.9% penetration rate |

locality-sensitive hashing (LSH) [20], which refers to a family of functions, which can be used to map data points into buckets in such a way, that it is highly probable for data points which are close to each to be located in the same buckets; conversely, data points which are distant from each other, are likely located in different buckets. Several authors utilised LSH and variations/extensions thereof to facilitate efficient retrieval of (in most cases) fingerprint data [126–129, 131, 135]. Geometric hashing [140], which was originally developed for object recognition (matching similar geometric shapes irrespective of translation, rotation, and scaling), has also been applied in the context of biometrics by coupling it with general-purpose keypoint detectors [132, 133, 136, 137].

Deeper descriptions of the various hashing algorithms and their extensions are out of scope for this article – the reader is referred to *e.g.* [24, 141]. Generally, by significantly reducing the dimensionality of the data and facilitating retrieval of a subset of candidate identities, general purpose fuzzy hashing methods adapted to the biometric data can be used to greatly reduce the computational workload. Aside from potential biometric performance degradation due to hashtable/bucket misses, the storage requirements of the system (especially in the case of geometric hashing) are typically increased.

### 3.4 *Feature Transformation*

This subsection surveys methods based on creating efficient representations of biometric templates, which reduce the computational cost of a single template comparison. This can typically be achieved through *e.g.* reducing the template dimensionality, creating fully or partially alignment invariant representations, or utilising more efficient template comparators (for instance, based on bit instead of floating-point operations). In other words, the goal is often to transform the original template (or create an unrelated alternative representation), so that it obtains certain desirable properties, while predominantly maintaining the discriminative power. Templates utilising such alternative or transformed representations can then be used on their own in an exhaustive search, or in more advanced approaches, *e.g.* act as a pre-selector (see subsection 3.1) in a multi-stage retrieval system. Table 5 summarises the surveyed methods.

*3.4.1* *Binarisation:* Comparison of float-based feature vectors is relatively expensive computationally, due to use of comparators based on *e.g.* Euclidean or $\chi^2$ distances. In many cases, such feature vectors can be quantised and encoded into binary strings, whereby utilisation of comparators based on *e.g.* Hamming distance is possible. Such comparators can take advantage of the more efficient bitwise operators, thereby reducing the computational workload. An illustrative example can be seen in [144] (and a simpler one in [143]), where various bit allocation schemes for float-based feature vectors generated by neural network-based systems are benchmarked. In [142], a new representation is extracted from minutiae points, which can be further binarised to accelerate the biometric template comparisons. Although some information is lost through the binarisation process, both publications show only negligible biometric performance loss in relation to their respective baselines, while achieving a significant speed-up. Finally, binarised feature vectors are an essential component in the context of many template protection schemes (see *e.g.* [157] for more details on this subject).

*3.4.2* *Dimensionality reduction:* Templates produced through dimensionality reduction can be used directly as a replacement for the full-sized templates (*e.g.* through PCA). Additionally, they can serve as a first pre-filtering step in a two-stage system (see subsection 3.1 for examples). An illustrative example is[145] (and a similar approach in [146]), where the so-called "short-length Iris-Codes', which comprise the most discriminative parts of the normal Iris-Codes, are presented. The transformed templates are an order of magnitude smaller than the original ones, and exhibit somewhat impaired biometric performance when benchmarked against the original templates, thereby making them good candidates for a pre-filtering step.

**Table 5** Feature transformation approaches

| Taxonomy | Characteristic | Publication | Method | Database | Biometric Performance | Computational Workload |
|---|---|---|---|---|---|---|
| Binarisation | Fingerprint | Capelli *et al.* [142] | Binarised minutia cylinder-code | FVC2006 | <1% average EER | At least an order of magnitude faster |
| | Face | Schlett *et al.* [143] | Multi-scale block LBP, binarisation | FERET, Extended-Yale-B | 15% FNMR at 10% FMR | 20-fold speed-up |
| | | Drozdowski *et al.* [144] | Benchmark of various quantisation and encoding methods | FERET, FRGC | 0.3% EER, 2.3% EER | An order of magnitude fewer CPU operations required |
| Dimensionality reduction | Iris | Gentile *et al.* [145] | Short-length Iris-Codes | MMU | 79.4% FNR at 1% FPR | 12-fold size reduction |
| | | Rathgeb *et al.* [146] | Most discriminative bits, selective algorithm fusion | CASIA-V3-Interval | 1.15% EER | ∼50% fewer bits |
| Variable to fixed size | Fingerprint | Jain *et al.* [147] | FingerCode | NIST SD9, MSU_DBI | ∼15% FRR at 1% FAR; ∼8% FRR at 1% FAR | — |
| | | Xu *et al.* [148] | Spectral minutiae | MCYT | 3.13% EER | — |
| | | Yang *et al.* [149] | Tessellated invariant moment features | FVC2002 | 3.57% average EER | 3-fold reduction |
| Alignment invariance | Iris | Rathgeb *et al.* [150] | Bloom filters | CASIA-V3 interval | 1.5% EER | 20% of baseline |
| | | Damer *et al.* [151] | Translation-invariant transformation | SYN1 | 0.646%EER, 1.213% EER | 6.56% of baseline, 2.45% of baseline |

**Table 6** Other approaches

| Taxonomy | Characteristic | Publication | Method | Database | Biometric Performance | Computational Workload |
|---|---|---|---|---|---|---|
| Search strategies | Iris | Kuehlkamp *et al.* [152] | 1-to-first search | Notre-Dame | see paper | 50-70% of baseline |
| | Fingerprint | Cappelli *et al.* [153] | Analysis of comparison scores, ruleset/criteria | FVC | 1% average error rate | 27% penetration rate (from indexing) reduced to 3.9% |
| Intrinsic data properties | Iris | Rathgeb *et al.* [154] | Iris-Code analysis, fewer relative shifting positions at comparison | CASIA-V4 interval | <1% EER | 4-fold reduction |
| Sample pre-alignment | Iris | Drozdowski *et al.* [155] | Pre-alignment of raw samples based on eye corner and pupil center locations | BioSecure | ∼2.5% EER | 2-fold reduction |
| Information fusion | Face | Drozdowski *et al.* [156] | Morphing | FERET | 98.82% RR-1 | 52.5% penetration rate |

*3.4.3 Variable to fixed size:* Comparisons of variable-size feature vectors are computationally demanding and often suffer from other domain-specific drawbacks. In biometrics, most prominently used variable-sized feature representation is that of fingerprint minutiae. The number of minutiae points can be inherently different between different data subjects and can further be augmented depending on the sample acquisition conditions (*i.e.* the so-called missing and spurious minutiae). In the literature, a number of alternative approaches to the traditional minutiae-based fingerprint comparison algorithm has been proposed by several authors [147–149]. All of those methods achieve biometric performance and computational workload results competitive with those of the traditional variable-size, minutiae-based algorithm.

*3.4.4 Alignment Invariance:* An important issue in biometrics, and especially fingerprint and iris recognition is the necessity of compensating for the relative sample misalignment caused by roll pose variations. This is typically done by considering multiple relative shifting positions of the Iris-Codes matrix and choosing the one with best comparison score, thereby increasing the computational cost of a single template comparison. In [150] and [151] feature transformations are presented, which ensure that sample misalignment (to a certain degree, reasonable from practical point of view) is intrinsically compensated for by the resulting feature vectors. Both approaches achieve substantial speed-up in an exhaustive search without significantly impairing the baseline biometric performance. Several other (not feature transformation based) approaches tackling the issue of iris alignment are also listed in subsection 3.5.

### *3.5 Other*

This subsection presents computational workload reduction approaches which do not fit into the previous categories. Table 6 summarises the surveyed methods. A simple method of reducing the computational workload in an exhaustive search is performing an early exit strategy, *i.e.* finishing the search once first (not necessarily best) suitable candidate is found. This is sometimes referred to as "one-to-first" search. In [152] this search strategy is analysed extensively for iris recognition in order to assess potential degradation of biometric performance. It is discovered, that the biometric performance degradation is strongly dependent on the decision thresholds (accuracy target) and size of the enrolment database. For some parameters, the biometric performance of an exhaustive search can be maintained, while the computational workload is significantly reduced. In [153] several strategies were proposed, which reduce candidate lists (produced by other methods) through analysis of comparison scores. In [154] an approach to reduce the number of relative shifting positions of the Iris-Codes which need to be considered in a template comparison was presented. The method is based on an analysis of the intrinsic properties of the iris data and achieves a considerable speed-up without impairing the biometric performance. In [155] a pre-alignment of raw iris images is performed. The method is based on automatic detection of eye corners and several other points in raw iris images, and subsequently aligning the eye corners onto a horizontal line. Thus, at a later point, once features are extracted, fewer relative shifting positions need to be considered during template comparisons. The approach of [156] relies on morphing (signal-level fusion). The facial images from the enrolment database are morphed (in 2s, 4s, or 8s), whereby biometric information from multiple subjects is fused into one image. The morphed images are then utilised for pre-filtering (see subsection 3.1). In addition to being explicitly used in some computational workload reduction schemes surveyed in this article, information fusion is an important aspect in ensuring the scalability of biometric systems in terms of biometric performance. For a comprehensive survey of this topic, the reader is referred to *e.g.* [158].

### *3.6 Acceleration*

Hardware acceleration can facilitate massive execution speed gains for certain types of computations. In the following subsections, the use of reconfigurable computing (subsection 3.6.1) and graphical processing units (subsection 3.6.2) in biometric systems is surveyed. The references in those two subsections are by no means exhaustive, due to the focus of this article being elsewhere. Instead, they outline the relevant concepts and highlight a few systems created for the different biometric characteristics. Lastly, they focus on the more recent publications due to the fast pace of developments within hardware components. For a quick general comparison of the capabilities, along with the advantages and disadvantages of those two types of hardware, the reader is referred to *e.g.* an industry white paper in [159], or a general survey of various Big Data analysis platforms and methods [160].

Although hardware acceleration cannot be strictly considered a method of workload reduction (since the *amount* of computations is not reduced – it is merely parallelised, distributed, or executed more efficiently), it is also mentioned here as an important aspect of speeding-up transactions in large-scale biometric identification systems. There appears to be a substantial research interest in the area of hardware-based acceleration utilising FPGAs and GPUs. Some of the existing publications present convincing and well-substantiated results, whereby massive speed gains (up to two orders of magnitude) are achieved in the benchmarks. It should be noted, however, that in some cases the experimental protocols of the benchmarks are questionable; in particular, it is not always clear if the external latency factors (unrelated to the algorithms themselves) have been accounted for in the evaluation. Furthermore, the degree of the CPU-based baseline algorithm optimisation is often not clearly outlined. The results must therefore be closely scrutinised, as it could be that the speed gains result merely from a poor baseline implementation.

This caveat notwithstanding, using reconfigurable computing and/or graphical processing units could be a promising avenue for speeding up the execution of various components (or even entire pipelines) in many different biometric modalities. On the other hand, factors such as difficulty of implementation, as well as purchase and maintenance cost have to be taken into consideration for real-world systems.

Lastly, software acceleration and optimisation are also worth mentioning in this context; although there does not seem to be many scientific publications on the topic. In [161], an extensive analysis of possible speed-ups in CPU-based Iris-Code comparisons is presented. The authors consider possible improvements through low-level implementations, manual loop unrolling, caching and pre-computing certain parts of data, analysis of memory access bottlenecks, multi-threading, as well as statistical optimisation of micro-operations. In [162], a hardware-software co-design of iris recognition pipeline is proposed. The authors benchmark highly optimised software code, coupled with a hardware-based implementation of several of the pipeline components. Both publications show that substantial speed-ups (but not computational workload reduction) can be achieved through code optimisations, which do not in themselves change the underlying algorithms or biometric feature representations.

*3.6.1 Reconfigurable Computing:* Field Programmable Gate Arrays (FPGAs) are integrated circuits containing an array/matrix of programmable logic blocks (of different types, *e.g.* general logic, memory, arithmetic), which can be programmably interconnected with each other and with input/output blocks. The programming/configuring is generally done using a hardware description language (*e.g.* VHDL or Verilog) or (nowadays rarely) circuit diagrams, and takes place *after* the chip has been manufactured. In other words, the FPGAs can be configured and re-configured to execute arbitrary digital circuits, and thus are capable of solving any computable problem. FPGAs can utilise hardware parallelism and deep pipelining extensively, thereby completing many more computations per clock cycle as opposed to a normal sequential execution. Additionally, they rely on much fewer layers of abstraction than the general purpose CPUs, thus facilitating lower-level programming, as well as custom memory and I/O interfaces. Those properties can be exploited to yield potentially massive speed-ups for certain applications (see *e.g.* [163, 164]). For a more detailed view of the current FPGA state-of-the-art, advantages and disatvantages, as well as future outlook and challenges, the reader is referred to fundamentals, *e.g.* [165]. Due to the abovementioned advantages, reconfigurable computing has been extensively applied to solve a variety of problems in many fields (see *e.g.* [166] for a survey), including computer vision, signal processing and pattern matching (see *e.g.* [167]), and neural networks (see *e.g.* [168]). Algorithms from those domains are cornerstones of various biometric systems; hence, substantial research effort has also been devoted to development of FPGA-based processing of biometric data.

FPGA based implementations of biometric systems' components or complete data processing pipelines were published *e.g.* for iris [169], fingerprint [170], face [171], (finger)vein [172], retina [173], and voice [174]. There, speed-ups over traditional CPU-based algorithms of up to two orders of magnitude were reported.

*3.6.2 Graphical Processing Units:* As the name suggests, traditional Graphical Processing Units (GPUs) were designed for very efficient processing of two and three dimensional graphics and have a rigid set of functions and programmable features. Over time, the ease of use/programmability and the range of applications for GPUs have steadily increased, especially with the introduction of general purpose frameworks for GPU programming such as CUDA [175] and OpenCL [176]. Taking advantage of the single program multiple-data (SPMD) programming model, the data can be processed in highly parallel ways. Thus, adapting code to run on GPUs can yield massive execution speed gains for many applications, *e.g.* linear algebra, sorting and searching, differential equations, or more generally floating-point operations on vectorisable data. For a general introduction to GPU computing, the reader is referred to *e.g.* [177]. Some tasks at which GPUs excel are important in typical biometric processing pipelines. Hence, there has been interest in the scientific community to leverage the power of GPUs in this domain as well. A good general introduction to usage of GPUs in biometrics, along with a brief survey of applications for fingerprint-based systems can be found in [178].

GPU based implementations of biometric systems' components or complete data processing pipelines were reported. It should also be noted, that GPUs (and more recently, specialised tensor processing units (TPUs) [179]), have also been utilised extensively in problems involving machine learning and deep neural networks, see *e.g.* [180]. In recent years, those technologies have also been applied to biometrics (*e.g.* facial recognition deep neural networks [181]), highlighting possibilities of hardware-acceleration use beyond efficient biometric identification, more specifically in the algorithm training phase. *E.g.* for iris [182], fingerprint [183], face [184], and sclera-vein [185], similarly to FPGAs (see subsection 3.6.1), speed-ups over traditional CPU-based algorithms of up to two orders of magnitude were reported.

## 4 Discussion

In this section, several matters relevant to the topic of this article are discussed, namely: the considerations and trade-offs of computational workload reduction approaches (subsection 4.1), a brief digression into data security (subsection 4.2), a perspective on how real large-scale biometric systems deal with large-scale biometric identification (subsection 4.3), and finally an outline of open issues and challenges in this research field (subsection 4.4).

### 4.1 Considerations and Trade-offs

As evidenced by previous sections, there exists a plethora of approaches which seek to reduce the computational workload requirements in biometric identification systems. Below, a systematic (qualitative, due to the infeasibility of directly comparing the results – recall subsection 2.2) discussion of noteworthy matters w.r.t. the different approach categories is given, concentrating on their general impact on: 1) computational workload, 2) biometric performance, and 3) disk/memory storage.

**Pre-filtering**

- **Computational workload** The potential speed-up depends on the discriminative power and size of the index templates. Given strongly discriminative index templates, a much smaller short-list of candidates can be produced, thereby minimising the number of the necessary template comparisons with the expensive (and accurate) comparator. On the other hand, the size of the index templates determines the computational cost of the pre-filtering step, as the probe index is compared exhaustively against the index templates. Naturally, those two parameters typically counterbalance each other – smaller size of the index templates typically entails lower discriminative power.
- **Biometric performance** Since the features used for pre-filtering typically have limited discriminative power, errors may occur, so that the sought identity is not among the returned candidate short-list, thereby increasing the false-negative rates. CMC curves are useful in assessing the efficacy of such features and can help decide on a reasonable size of the candidate short-list.
- **Storage** Since additional information (index) is stored in order to facilitate the pre-filtering step, the storage requirements are increased.

**Binning**

- **Computational workload** The potential speed-up benefits are limited by the number of bins. It tends to be rather small, especially for the handcrafted classes/types. Additionally, the handcrafted classes/types are very often inherently unevenly distributed (due to genetics and environmental influences). Consequently, computational workload reduction obtained through binning varies accordingly with the relative frequencies of the bins. A good example is binning based on fingerprint types.

Whorls and loops generally exhibit the highest prevalence and there are variations across different ethnic groups (see Rife [186]). Nevertheless, it is still a feasible approach, and it has been used in operational systems, *e.g.* in AFIS' (see *e.g.* [187, 188]) – initially using the explicit classes, more recently utilising machine learning to develop non-exclusive classes that lead to more balanced bin sizes. In some cases, however, a severely non-uniform distribution across the bins can invalidate the binning approach entirely. For instance, people of many ethnic groups (or entire countries) have predominantly brown eyes, thus little to no speed-up can be achieved by binning using eye colour in such systems.

**Biometric performance** In order for the system to be viable, the classification accuracy must be near-optimal. Otherwise, the probability of false-negative errors increases due to misclassification and consequently searching in the wrong bin(s) (*i.e.* pre-selection errors).

**Storage** Typically not significantly increased, since only the metadata (*e.g.* the fingerprint types) need to be stored.

**Data-structures**

**Computational workload** By often relying on divide-and-conquer approaches, the complexity of the retrieval algorithm can often be reduced from the linear complexity down to the (near-)logarithmic complexity.

**Biometric performance** Due to wrong paths being taken during the search structure traversal, the potential for making false-negative error increases. For many approaches this is especially relevant for the higher levels of the structure, where the information stored by the nodes is denser than near the leaves. On the other hand, the potential for false positive errors is typically reduced due to the lower penetration rate.

**Storage** The storage requirements are typically increased, since additional hashtables and/or tree-like data-structures have to be maintained.

**Feature transformation**

**Computational workload** Although the individual template comparisons are computed much more efficiently, the identification is still carried out over the entire search space (exhaustive search), thereby severely limiting the potential computational workload reduction.

**Biometric performance** The more compact template representations and/or more efficient comparators may suffer from a decrease in discriminative power and hence a lower biometric performance.

**Storage** Typically decreased, due to more compact template representations.

**Acceleration**

**Computational workload** Not reduced *per se*, merely computed more efficiently (*e.g.* paralellised, distributed, or otherwise optimised).

**Biometric performance** Typically unaffected, as functionally equivalent algorithms are somehow implemented or optimised to achieve faster computation speeds.

**Storage** Possibly increased, as it may be necessary to port the biometric data to the specifics of the utilised system (*e.g.* CUDA) or distribute them across a network.

Due to varying system requirements and policies, it is important to enable the biometric systems' operators to make well-informed decisions w.r.t. the used algorithms. Therefore, for any proposed computational workload reduction methods, it is crucial to include the above information, as well as benchmarks against a current state-of-the-art algorithm performing an exhaustive search (baseline). By doing so, the trade-offs (biometric performance, computational workload, storage) of the proposed methods can be evaluated, thereby facilitating informed decisions on the systems' design and policies. In some cases, it may even be possible to probabilistically model the impact of the proposed methods on the biometric performance, which could potentially be very useful in establishing the pertinent trade-offs even prior to the experimental evaluations on real data. However, such models rarely appear in the surveyed literature. Examples include, *e.g.* [17], which discusses the binning approaches in general and [118], where a statistical model for the proposed Bloom filter-based hierarchical retrieval method is included. Hence, the development of such models could be an interesting avenue of future research.

### 4.2 Data Security

In addition to the need for computational workload reduction, which was the core topic of this article, the potential of data exposure is a large concern in biometric system deployments, where the stored data is, in most cases, secured using traditional encryption algorithms (see *e.g.* [189]). From the technical point of view, this means that should the data be compromised, serious problems such as identity theft, cross-matching without consent arise, furthermore the renewability of such biometric templates is severely limited. Additionally, the centralised storage of sensitive personal and biometric data has increasingly been under scrutiny, both by the general public and various non-governmental organisations, which recently has led to widened legislation against privacy violations (*e.g.* GDPR in Europe [190]). The ISO/IEC international standard on biometric information protection [191] stipulates several properties required for biometric template protection schemes. While many approaches have been proposed for normal biometric systems (see *e.g.* [192] for a survey), template protection coupled together with computational workload reduction has received relatively little attention in the scientific literature. Some early proof-of-concept works and trade-off analyses have been carried out *e.g.* in [193–195].

### 4.3 Real-World Systems

Due to confidentiality constraints (*i.e.* company or state secrets), the availability of details for real-world systems is not nearly as abundant as that of scientific publications. Nevertheless, this subsection will give several examples, based on the existing literature and personal communications.

**Aadhaar** As of this writing, the Indian National ID Programme (see *e.g.* [6, 196]) encompasses acquisition and usage of the largest biometric system in the world in terms of number of the enrolled data subjects. During enrolment, a de-duplication check must be carried out, in order to avoid issuing multiple unique identification numbers to the same individual. The de-duplication proceeds in several steps: first, obvious duplicates are pre-filtered out based on metadata (demographic information), either through exact matches or fuzzy matches followed by a preemptive biometric check. Subsequently, an exhaustive search of the biometric database is performed by one of the COTS systems provided by three different vendors. Each system uses a different implementation and information fusion (data from two irides and ten fingerprints is used) strategies. Any potential duplicate is verified by another system, and if the need arises, it is also adjudicated manually by a trained biometric system operator. The whole system design is distributed and massively parallelisable, and claimed to be scalable through use of commodity hardware and enterprise Big Data solutions.

**UAE** The border control system in United Arab Emirates (see *e.g.* [197, 198]) takes advantage of the intrinsically efficient iris representation (Iris-Codes, see [199]) and distributed architecture of COTS components for quick biometric identification queries. It is reported, that an exhaustive search against the database of close to 1 million subjects can be executed within a couple seconds. I/O latency issues are avoided by pre-loading the entire enrolment database into random access memory.

**NEXUS** In Canada, automated self-service kiosks for selected airports and (frequent) travellers are offered in order to expedite the border control process. The system uses iris as the biometric characteristic, and performs 1-to-first searches on the database of over 0.5 million enrolled data subjects for each biometric identification (see *e.g.* [200]).

**EES** Due to the specifics of the legal mandate [15], the biometric systems for the EU visa and entry-exit system will be forced to perform exhaustive database searches. Due to the operational scenarios (such as border control), stringent requirements for quick (real-time) query responses have been imposed on the potential biometric systems vendors. It can therefore be expected that efficient data representations, information fusion schemes, as well as parallel/distributed design will be essential components of the forthcoming infrastructure solutions.

**AFIS** Deployments of the Automated Fingerprint Identification Systems (see *e.g.* [187, 201, 202]) are ubiquitous around the world and used for instance in the context of criminal investigations. Such systems are known to utilise computational workload reduction methods based on demographics, coarse fingerprint data (such as fingerprint type), along with highly optimised software algorithms to facilitate fast response times. Prominent examples include the Integrated AFIS (IAFIS) ran by the Federal Bureau of Investigation (FBI) in the USA and the database of the German Bundeskriminalamt (BKA).

**Industry** In order to provide competitive search speeds for large biometric identification systems, commercial vendors of biometric recognition technologies, for example the German company Dermalog (information acquired through personal communication), are known to utilise methods of workload reduction in their products. National Institute of Standards and Technology (NIST) carried out an evaluation of 1:N face recognition vendors [203]. Among the details, it has been stated that several submitted algorithms take the expense of constructing fast search data-structures at enrolment, in order to achieve sub-linear search duration growth (with respect to the size of the biometric reference database).

From the above examples, it is clear that computational workload is a critical consideration in operational systems and certain methods used to expedite the high number of queries handled by the existing large-scale systems around the world. To summarise in the context of the proposed taxonomy (recall figure 2), the following methods are represented in the list above:

- Pre-filtering by metadata (*e.g.* demographic and geographic).
- Search strategies (*e.g.* 1-to-first search).
- Binning with coarse features (*e.g.* fingerprint types).
- Intrinsically efficient feature representations (*e.g.* IrisCodes).
- Software optimisation.
- Massive parallelisation and distribution of computations.

Notably absent are the methods of hardware acceleration (*i.e.* reconfigurable computing and graphical processing units) – it could be, that the practical matters (*e.g.* monetary implementation and maintenance costs) outweigh the benefits of potential speed improvements. Another potential important issue is vendor lock-in and the necessity of tailoring for specific software/hardware combination, which was deliberately avoided in *e.g.* the UAE and Aadhaar systems by making the design decision to use commodity CPU-based hardware (see *e.g.* [6, 197]). The pre-selection methods, which are heavily researched, seem to be seldom used in those deployments. In other words, in this case there does appear to be (or perhaps is perpetuated by the information scarcity in this area), at least to a certain degree, a mismatch between what is researched in academia and what is actually used by the industry. This and other open issues are discussed in more detail in the next subsection.

### 4.4 Open Issues

Several open issues/challenges remain in research related to computational workload reduction in biometric identification systems:

**Standardisation** As described in subsection 2.2, as of yet there exists no standardised way of reporting results for biometric computational workload and its reduction. This leads to a multitude of methodologies and metrics in the scientific literature, thereby making direct benchmarks and comparative assessment of the proposed methods extremely cumbersome to carry out. Moreover, in many publications, even the baseline results (*i.e.* exhaustive search with a state-of-the-art algorithm) are not reported, which further exacerbates this issue. In [118], experimental prerequisites and metrics for such evaluations are proposed; as of this writing, there is an ongoing effort to include metrics to measure computational workload and its reduction in the current revision of the ISO/IEC IS 19795-1. This effort notwithstanding, the standardisation in this research area remains an open debate subject within the standardisation committee and in general.

**Scalability** Due to limited availability of large-scale biometric data for academic research, many, if not most, of the surveyed approaches were tested on relatively small databases (mostly up to hundreds or thousands of subjects; tens of thousands of samples). Hence, the scalability of many of the proposed methods remains questionable or unproven in practice (*cf.* table 1 w.r.t. the sizes of large real-world deployments of biometric systems). Some authors are fortunate enough to receive access to large-scale, sequestered databases (*e.g.* law enforcement) in order to evaluate and validate their approaches (*e.g.* [53, 67]), but such cases are rare in the surveyed literature, likely as a consequence of the significant practical and legal hurdles associated with accessing the sensitive biometric/personal data.

**Biometric performance trade-off** It appears that many approaches are capable of delivering significant (*e.g.* one or two orders of magnitude) decrease in computational workload requirements. However, further reductions prove elusive for most methods due to rapidly degenerating biometric performance. Some of the surveyed methods incorporate in their designs information fusion (from multiple biometric characteristics or multiple instances of the same characteristics) to mitigate this issue to a certain degree.

**Dissonance between academia and industry** There seem to exist, to a certain degree, discrepancies between methods and goals of the research conducted within the academia, and the actual practical use cases in the industry. Several examples (also partially related to the aforementioned issues of standardisation and scalability) relevant to the topic of this article are:

**Evaluation protocol** Whereas a substantial number of the surveyed publications perform their experiments using the closed-set identification, the real-world systems are essentially universally required to perform the much more challenging open-set identification.

**Decision thresholds** Almost all of the surveyed publications do not report decision thresholds at which the given biometric performance was achieved. Furthermore, evaluations on different datasets with fixed decision thresholds are rarely performed. This is in stark contrast with the industry/law enforcement practices, where decision thresholds and fixed operational points are used extensively in the operational systems.

**Results reporting** The surveyed publications often report results using metrics, which are of limited value in the industry practice. For instance, operational systems would not typically operate at EER (even if it ever was an operational point at all), but rather at fixed false-positive identification-error rates acceptable within their respective system policies. Many publications use rank-based reporting and CMC curves, which imply the less interesting (from the industry point of view) closed-set evaluation protocol, as mentioned above.

**Acceleration** As evidenced by subsection 3.6, significant research efforts have been devoted to develop biometric algorithms suitable for FPGA or GPU computations. However, as mentioned in subsection 4.3, the existing deployments of large-scale biometric systems lean towards distributed architectures of commodity CPU hardware due to other practical considerations; additionally, software optimisations, which play an important role in the commercial systems, are only superficially treated in the scientific literature.

Thus, aside from the technical challenges (and potential limits) of improving on the trade-offs between biometric performance and computational workload reduction, there are two key areas that should be considered by the large-scale biometric identification practitioners from academia and industry alike:

**Academia and industry cooperation** Much tighter integration between the academic research and industry requirements is needed. The academics should seek out industry partners to validate their proposed systems and solutions outside of a lab setting; conversely, the industry should engage in outreach initiatives to academia in order to promote the actual prerequisites, requirements, and challenges of the commercial systems. More frequent and deeper joint (research or otherwise) projects and partnerships between academia and industry could potentially help to ameliorate the aforementioned dissonances. Jain *et al*. [204] recently published the short "Guidelines for Best Practices in Biometric Research". This document can serve as a good starting point outlining the absolute essentials for legitimate and practical reporting of results within biometrics research. Furthermore, the representatives from all the stakeholders should begin or continue to actively engage in the international standardisation efforts, for instance the ones by the ISO/IEC JTC 1/SC 37 (who are responsible for, among others, the biometric performance evaluation and harmonized biometric vocabulary standards [4, 5]) and/or of other national agencies, such as NIST in the USA and BSI in Germany. This way, meaningful consensus w.r.t. evaluation protocols, metrics, and benchmarks that reflect the real use cases can be attained.

**Practical evaluations** Entities (*e.g*. governmental agencies, universities, companies) in possession of large amounts of biometric data should support large-scale evaluations, thus facilitating scalability assessment and fair benchmarks between the systems developed by the academic researchers and the commercial vendors. Such benchmarks make available an API against which algorithms can be coded, then submitted to a central server, and finally ran and evaluated there using the same experimental protocol and metrics. Such tests offer an additional advantage, in that the vast majority of the image data used for evaluation remains unseen by the algorithm authors, which in turn facilitates higher generalisably of the proposed algorithms. Lastly, synthetically generated data could be used to some extent in the context of biometric scalability testing (see *e.g*. [205]). Interesting existing initiatives in this area are, for example: the BEAT platform [206] with the aim of developing a general standard framework biometric technologies' evaluation, along with several more constrained initiatives such as the indexing competition under Fingerprint Verification Competition (FVC-onGoing) [207], the 1:N Evaluation under Face Recognition Vendor Test (FRVT) [208], one-to-many evaluations under Iris Exchange (IREX) – for fingerprint, face, and iris characteristics, respectively.

## 5 Summary

Large-scale biometric identification systems are confronted with high computational workload. This is especially the case for an exhaustive search, where the computational effort required during retrieval grows linearly with the number of enrollees. Methods that seek to alleviate this issue aim at reducing the number of template comparisons necessary per retrieval (penetration rate), the computational costs associated with individual template comparisons, or at optimising the software/hardware system implementations. In this article, a taxonomy for conceptual categorisation of such methods is presented, followed by a comprehensive survey of publications pertaining therewith. The article is concluded with a discussion of matters to take note of with the various categories of approaches, a digression on usage of such methods in real-world systems, as well as an outline of remaining relevant challenges.

As the number and scale of the biometric systems deployments worldwide steadily increases, computational workload reduction in biometric identification systems can be expected to remain an active field of research, especially since a number of open issues/challenges remains unresolved regardless of the significant advances made through the academic and commercial research. To solve said challenges, standardisation and much tighter cooperation between the academia, industry, governmental agencies, and other concerned parties is necessary. There exists an urgent need of a unified methodology for reporting of computational workload and its reduction. Furthermore, another important matter is the development of experimental protocols, benchmarks, and metrics which closely correspond with the actual prerequisites and use cases of the real-world deployments. To accomplish this, the international standardisation efforts are a promising avenue, albeit continuous engagement from all the concerned stakeholders is necessary to establish a suitable and broad consensus.

## Acknowledgements

This work was partially supported by the German Federal Ministry of Education and Research (BMBF), by the Hessen State Ministry for Higher Education, Research and the Arts (HMWK) within Center for Research in Security and Privacy (CRISP), and the LOEWE-3 BioBiDa Project (594/18-17).

## 6 References

1 Thakkar, D.. 'Global biometric market analysis: Trends and future prospects'. (Bayometric, 2018. last accessed: 11/6/2019. https://www.bayometric.com/global-biometric-market-analysis/
2 Bhutani, A., Bhardwaj, P. 'Biometrics market size by application'. (Global Market Insights, 2017. GMI493
3 Markets and Markets. 'Biometric system market by authentication type - global forecast to 2023'. (Markets and Markets, 2018. SE 3449
4 ISO/IEC JTC1 SC37 Biometrics. 'ISO/IEC 2382-37:2017. information technology – vocabulary – part 37: Biometrics'. (International Organization for Standardization, 2017
5 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, 2006
6 Unique Identification Authority of India. 'Role of biometric technology in Aadhaar enrollment'. (UIDAI, 2012.
7 Unique Identification Authority of India. 'Aadhaar dashboard'. (UIDAI, 2018. last accessed: 11/6/2019. https://www.uidai.gov.in/aadhaar_dashboard/
8 European Union. 'Regulation (EU) no 603/2013 of the European Parliament and of the Council'. (Official Journal of the European Union, 2013. last accessed: 11/6/2019. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32013R0603
9 European Union Agency for the Operational Management of Large-Scale IT Systems in the Area of Freedom, Security and Justice. 'Eurodac storage capacity increased'. (eu-LISA, 2016. last accessed: 11/6/2019. https://www.eulisa.europa.eu/Newsroom/News/Pages/Eurodac-storage-capacity-increased.aspx
10 Gemalto. 'DHS's automated biometric identification system IDENT - the heart of biometric visitor identification in the USA'. (Gemalto Case Studies, 2019. last accessed: 11/6/2019. https://www.gemalto.com/govt/customer-cases/ident-automated-biometric-identification-system
11 Federal Bureau of Investigation. 'CODIS - NDIS statistics'. (FBI, 2018. last accessed: 11/6/2019. https://www.fbi.gov/services/laboratory/biometric-analysis/codis/ndis-statistics
12 Consortium for Elections and Political Process Strengthening. 'Assessment of electoral preparations in the Democratic Republic of the Congo'. (CEPPS, 2018.
13 Paynter, T.. 'Northrop grumman wins $95 million award from department of homeland security to develop next-generation biometric identification services system'. (Northrop Grumman, 2018. last accessed: 11/6/2019. 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
14 Department of Homeland Security. 'DHS/ALL-041 external biometric records (EBR) system of records'. (Federal Register, 2018. last accessed: 11/6/2019. https://www.regulations.gov/docket?D=DHS-2017-0039
15 European Union. 'Regulation (EU) 2017/2226 of the European Parliament and of the Council'. (Official Journal of the European Union, 2017. last accessed: 11/6/2019. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32017R2226
16 European Commission. 'Smart borders'. (EU Migration and Home Affairs, 2018. last accessed: 11/6/2019. https://ec.europa.eu/home-affairs/what-we-do/policies/borders-and-visas/smart-borders_en
17 Daugman, J. 'Biometric decision landscapes'. (University of Cambridge - Computer Laboratory, 2000. UCAM-CL-TR-482
18 Institute of Electrical and Electronics Engineers. 'Biometric council newsletter'. (IEEE, 2015. last accessed: 11/6/2019. http://ieee-biometrics.org/images/pdf/Newsletter_Nov_2015_corrected.pdf

19 Daugman, J.. 'History of iris recognition'. (University of Cambridge - Computer Laboratory, 2019. last accessed: 11/6/2019. https://www.cl.cam.ac.uk/ jgd1000/history.html
20 Gionis, A., Indyk, P., Motwani, R. 'Similarity search in high dimensions via hashing'. In: International Conference on Very Large Data Bases (VLDB). (Morgan Kaufmann, 1999. pp. 518–529
21 Chávez, E., Navarro, G., Baeza-Yates, R., Marroquín, J.L.: 'Searching in metric spaces', *Computing Surveys (CSUR)*, 2001, **33**, (3), pp. 273–321
22 Hjaltason, G.R., Samet, H.: 'Index-driven similarity search in metric spaces (survey article)', *Transactions on Database Systems (TODS)*, 2003, **28**, (4), pp. 517–580
23 Zezula, P., Amato, G., Dohnal, V., Batko, M.: 'Similarity search: the metric space approach'. (Springer, 2006)
24 Wang, J., Shen, H.T., Song, J., Ji, J.: 'Hashing for similarity search: A survey', *arXiv preprint arXiv:14082927*, 2014, pp. 1–29
25 Abbasifard, M.R., Ghahremani, B., Naderi, H.: 'A survey on nearest neighbor search methods', *International Journal of Computer Applications*, 2014, **95**, (25)
26 Rachkovskij, D.A.: 'Index structures for fast similarity search for binary vectors', *Cybernetics and Systems Analysis*, 2017, **53**, (5), pp. 799–820
27 Rachkovskij, D.A.: 'Index structures for fast similarity search for real-valued vectors I', *Cybernetics and Systems Analysis*, 2018, **54**, (1), pp. 152–164
28 Rachkovskij, D.A.: 'Index structures for fast similarity search for real vectors II', *Cybernetics and Systems Analysis*, 2018, **54**, (2), pp. 320–335
29 Li, S.Z., Jain, A.K.: 'Encyclopedia of biometrics'. (Springer, 2015)
30 Jain, A.K., Flynn, P., Ross, A.: 'Handbook of biometrics'. (Springer, 2007)
31 Maltoni, D., Maio, D., Jain, A.K., Prabhakar, S.: 'Handbook of fingerprint recognition'. (Springer, 2009)
32 Jain, A.K., Li, S.Z.: 'Handbook of face recognition'. (Springer, 2011)
33 Bowyer, K.W., Burge, M.J.: 'Handbook of iris recognition'. (Springer, 2016)
34 Uhl, A., Marcel, S., Busch, C., Veldhuis, R.N.J.: 'Handbook of Vascular Biometrics'. (Springer, 2019). forthcoming
35 Proença, H., Neves, J.C. 5. In: Rathgeb, C., Busch, C., editors. 'Iris biometric indexing'. Security. (Institution of Engineering and Technology, 2017. pp. 101–124
36 Schuch, P.: 'Survey on features for fingerprint indexing', *IET Biometrics*, 2018,
37 Liu, L., Chen, J., Fieguth, P., Zhao, G., Chellappa, R., Pietikäinen, M.: 'From BoW to CNN: Two decades of texture representation for texture classification', *International Journal of Computer Vision*, 2019, **127**, (1), pp. 74–109
38 Ratha, N.K., Karu, K., Chen, S., Jain, A.K.: 'A real-time matching system for large fingerprint databases', *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 1996, **18**, (8), pp. 799–813
39 de Boer, J., Bazen, A.M., Gerez, S.H. 'Indexing fingerprint databases based on multiple features'. In: Annual Workshop on Circuits, Systems and Signal Processing. (Technology Foundation (STW), 2001. pp. 300–306
40 Bhanu, B., Tan, X.: 'Fingerprint indexing based on novel features of minutiae triplets', *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2003, **25**, (5), pp. 616–622
41 Feng, J., Cai, A. 'Fingerprint indexing using ridge invariants'. In: International Conference on Pattern Recognition (ICPR). vol. 4. (IEEE, 2006. pp. 433–436
42 Li, J., Yau, W.Y., Wang, H. 'Fingerprint indexing based on symmetrical measurement'. In: International Conference on Pattern Recognition (ICPR). vol. 1. (IEEE, 2006. pp. 1038–1041
43 Liang, X., Bishnu, A., Asano, T.: 'A robust fingerprint indexing scheme using minutia neighborhood structure and low-order delaunay triangles', *Transactions on Information Forensics and Security (TIFS)*, 2007, **2**, (4), pp. 721–733
44 Wang, Y., Hu, J., Phillips, D.: 'A fingerprint orientation model based on 2D fourier expansion (FOMFE) and its application to singular-point detection and fingerprint indexing', *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2007, **29**, (4), pp. 573–585
45 Feng, J., Jain, A.K. 'Filtering large fingerprint database for latent matching'. In: International Conference on Pattern Recognition (ICPR). (IEEE, 2008. pp. 1–4
46 Cappelli, R.: 'Fast and accurate fingerprint indexing based on ridge orientation and frequency', *Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 2011, **41**, (6), pp. 1511–1521
47 Paulino, A.A., Liu, E., Cao, K., Jain, A.K. 'Latent fingerprint indexing: Fusion of level 1 and level 2 features'. In: International Conference on Biometrics: Theory, Applications and Systems (BTAS). (IEEE, 2013. pp. 1–8
48 Gyaourova, A., Ross, A. 'A coding scheme for indexing multimodal biometric databases'. In: Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). (IEEE, 2009. pp. 93–98
49 Gyaourova, A., Ross, A.: 'Index codes for multibiometric pattern retrieval', *Transactions on Information Forensics and Security (TIFS)*, 2012, **7**, (2), pp. 518–529
50 Mohanty, P., Sarkar, S., Kasturi, R., Phillips, P.J.: 'Subspace approximation of face recognition algorithms: An empirical study', *Transactions on Information Forensics and Security (TIFS)*, 2008, **3**, (4), pp. 734–748
51 Chen, B.C., Chen, Y.Y., Kuo, Y.H., Hsu, W.H.: 'Scalable face image retrieval using attribute-enhanced sparse codewords', *Transactions on Multimedia*, 2013, **15**, (5), pp. 1163–1173
52 Wang, D., Jain, A.K. 'Face retriever: Pre-filtering the gallery via deep neural net'. In: International Conference on Biometrics (ICB). (IEEE, 2015. pp. 473–480
53 Wang, D., Otto, C., Jain, A.K.: 'Face search at scale', *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2017, **39**, (6), pp. 1122–1136
54 Konrad, M., Stögner, H., Uhl, A., Wild, P.: 'Computationally efficient serial combination of rotation-invariant and rotation compensating iris recognition algorithms', *International Conference on Computer Vision Theory and Applications (VISAPP)*, 2010, **1**, pp. 85–90
55 Gadde, R.B., Adjeroh, D., Ross, A. 'Indexing iris images using the burrows-wheeler transform'. In: International Workshop on Information Forensics and Security (WIFS). (IEEE, 2010. pp. 1–6
56 Dey, S., Samanta, D.: 'Iris data indexing method using Gabor energy features', *Transactions on Information Forensics and Security (TIFS)*, 2012, **7**, (4), pp. 1192–1203
57 Kavati, I., Prasad, M.V.N.K., Bhagvati, C. 'Vein pattern indexing using texture and hierarchical decomposition of Delaunay triangulation'. In: International Symposium on Security in Computing and Communication. (Springer, 2013. pp. 213–222
58 You, J., Kong, W.K., Zhang, D., Cheung, K.H.: 'On hierarchical palmprint coding with multiple features for personal identification in large databases', *Transactions on Circuits and Systems for Video Technology (TCSVT)*, 2004, **14**, (2), pp. 234–243
59 Wu, Z., Ke, Q., Sun, J., Shum, H.Y. 'Scalable face image retrieval with identity-based quantization and multi-reference re-ranking'. In: Conference on. Computer Vision and Pattern Recognition (CVPR). (IEEE, 2010. pp. 3469–3476
60 Gentile, J.E., Ratha, N., Connell, J. 'An efficient, two-stage iris recognition system'. In: International Conference on Biometrics: Theory, Applications and Systems (BTAS). (IEEE, 2009. pp. 211–215
61 Tang, D., Huang, B., Li, R., Li, W. 'A person retrieval solution using finger vein patterns'. In: International Conference on Pattern Recognition (ICPR). (IEEE, 2010. pp. 1306–1309
62 Billeb, S., Rathgeb, C., Buschbeck, M., Reininger, H., Kasper, K. 'Efficient two-stage speaker identification based on universal background models'. In: International Conference of the Biometrics Special Interest Group (BIOSIG). (IEEE, 2014. pp. 1–6
63 Pflug, A., Rathgeb, C., Scherhag, U., Busch, C. 'Binarization of spectral histogram models: An application to efficient biometric identification'. In: International Conference on Cybernetics (CYBCONF). (IEEE, 2015. pp. 501–506
64 Iqbal, A., Namboodiri, A. 'Cascaded filtering for fingerprint identification using random projections'. In: Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). (IEEE, 2012. pp. 77–82
65 Chen, F., Huang, X., Zhou, J.: 'Hierarchical minutiae matching for fingerprint and palmprint identification', *Transactions on Image Processing*, 2013, **22**, (12), pp. 4964–4971
66 Yi, D., Lei, Z., Hu, Y., Li, S.Z.: 'Fast matching by 2 lines of code for large scale face recognition systems', *arXiv preprint arXiv:13027180*, 2013,
67 Hao, F., Daugman, J., Zielinski, P.: 'A fast search algorithm for a large fuzzy database', *Transactions on Information Forensics and Security (TIFS)*, 2008, **3**, (2), pp. 203–212
68 Ross, A., Sunder, M.S. 'Block based texture analysis for iris classification and matching'. In: Conference on Computer Vision and Pattern Recognition - Workshops (CVPRW). (IEEE, 2010. pp. 30–37
69 Hämmerle-Uhl, J., Penn, G., Pötzelsberger, G., Uhl, A.: 'Size-reduction strategies for iris codes', *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, 2015, **9**, (1), pp. 290–293
70 Rathgeb, C., Uhl, A., Wild, P.: 'Incremental iris recognition: A single-algorithm serial fusion strategy to optimize time complexity', *International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2010,
71 Surbiryala, J., Raghavendra, R., Busch, C. 'Finger vein indexing based on binary features'. In: Colour and Visual Computing Symposium (CVCS). (IEEE, 2015. pp. 1–6
72 Henry, E.R.: 'Classification and uses of finger prints'. (HM Stationery Office, 1900)
73 Heindl, R.: 'Daktyloskopie'. (W. de Gruyter & Company, 1927)
74 Hollingsworth, K.P., Bowyer, K.W., Flynn, P.J.: 'The best bits in an iris code', *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2009, **31**, (6), pp. 964–973
75 Dantcheva, A., Elia, P., Ross, A.: 'What else does your biometric data reveal? A survey on soft biometrics', *Transactions on Information Forensics and Security (TIFS)*, 2016, **11**, (3), pp. 441–467
76 Galton, F.: 'Fingerprint directories'. (Macmillan and Company, 1895)
77 Galar, M., Derrac, J., Peralta, D., Triguero, I., Paternain, D., Lopez-Molina, C., et al.: 'A survey of fingerprint classification part I: Taxonomies on feature extraction methods and learning models', *Knowledge-based systems*, 2015, **81**, pp. 76–97
78 Galar, M., Derrac, J., Peralta, D., Triguero, I., Paternain, D., Lopez-Molina, C., et al.: 'A survey of fingerprint classification part II: experimental analysis and ensemble proposal', *Knowledge-Based Systems*, 2015, **81**, pp. 98–116
79 Zheng, R., Zhang, C., He, S., Hao, P. 'A novel composite framework for large-scale fingerprint database indexing and fast retrieval'. In: International Conference on Hand-Based Biometrics (ICHB). (IEEE, 2011. pp. 1–6
80 Wayman, J.L. 'Multifinger penetration rate and ROC variability for automatic fingerprint identification systems'. In: Automatic Fingerprint Recognition Systems. (Springer, 2004. pp. 305–316
81 Drozdowski, P., Fischer, D., Rathgeb, C., Schiel, C., Busch, C. 'Database binning and retrieval in multi-fingerprint identification systems'. In: International Workshop on Information Forensics and Security (WIFS). (IEEE, 2018. pp. 1–7
82 Yu, L., Zhang, D., Wang, K., Yang, W.: 'Coarse iris classification using box-counting to estimate fractal dimensions', *Pattern Recognition*, 2005, **38**, (11), pp. 1791–1798
83 Palla, S., Chikkerur, S., Govindaraju, V., Rudravaram, P. 'Classification and indexing in large biometric databases'. In: Biometrics Consortium Conference. (NIST, 2004. pp. 1–3
84 Zhou, Y., Liu, Y., Feng, Q., Yang, F., Huang, J., Nie, X.: 'Palm-vein classification based on principal orientation features', *PLOS ONE*, 2014, **9**, (11), pp. 1–12
85 Qiu, X., Sun, Z., Tan, T.: 'Global texture analysis of iris images for ethnic classification', *Advances in Biometrics*, 2005, **3832**, pp. 411–418
86 Tapia, J.E., Perez, C.A., Bowyer, K.W. 'Gender classification from iris images using fusion of uniform local binary patterns'. In: European Conference on

Computer Vision (ECCV). Springer. (Springer, 2014. pp. 751–763
87 Singh, M., Nagpal, S., Vatsa, M., Singh, R., Noore, A., Majumdar, A. 'Gender and ethnicity classification of iris images using deep class-encoder'. In: International Joint Conference on Biometrics (IJCB). (IEEE, 2017. pp. 666–673
88 Puhan, N.B., Sudha, N. 'A novel iris database indexing method using the iris color'. In: Conference on Industrial Electronics and Applications. (IEEE, 2008. pp. 1886–1891
89 Zhao, Q. 'A new approach for noisy iris database indexing based on color information'. In: International Conference on Computer Science Education (ICCSE). (IEEE, 2011. pp. 28–31
90 Dantcheva, A., Erdogmus, N., Dugelay, J.L. 'On the reliability of eye color as a soft biometric trait'. In: Workshop on Applications of Computer Vision (WACV). (IEEE, 2011. pp. 227–231
91 Park, U., Liao, S., Klare, B., Voss, J., Jain, A.K. 'Face finder: Filtering a large face database using scars, marks and tattoos'. (Michigan State University, 2011. MSU-CSE-11-15
92 Germain, R.S., Califano, A., Colville, S.: 'Fingerprint matching using transformation parameter clustering', *Computational Science and Engineering*, 1997, **4**, (4), pp. 42–49
93 Ross, A., Mukherjee, R. 'Augmenting ridge curves with minutiae triplets for fingerprint indexing'. In: Biometric Technology for Human Identification IV. vol. 6539. International Society for Optics and Photonics. (SPIE, 2007. pp. 1–12
94 Liu, M., Jiang, X., Kot, A.C.: 'Efficient fingerprint search based on database clustering', *Pattern Recognition*, 2007, **40**, (6), pp. 1793–1803
95 Biswas, S., Ratha, N.K., Aggarwal, G., Connell, J. 'Exploring ridge curvature for fingerprint indexing'. In: International Conference on Biometrics: Theory, Applications and Systems (BTAS). (IEEE, 2008. pp. 1–6
96 Iloanusi, O.N., Gyaourova, A., Ross, A. 'Indexing fingerprints using minutiae quadruplets'. In: Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). (IEEE, 2011. pp. 127–133
97 Iloanusi, O.N.: 'Fusion of finger types for fingerprint indexing using minutiae quadruplets', *Pattern Recognition Letters*, 2014, **38**, pp. 8–14
98 Perronnin, F., Dugelay, J.L. 'Clustering face images with application to image retrieval in large databases'. In: Biometric Technology for Human Identification II. vol. 5779. (International Society for Optics and Photonics, 2005. pp. 256–265
99 Chaari, A., Lelandais, S., Ahmed, M.B. 'A pruning approach improving face identification systems'. In: International Conference on Advanced Video and Signal Based Surveillance. (IEEE, 2009. pp. 85–90
100 Klare, B.F., Blanton, A., Klein, B. 'Efficient face retrieval using synecdoches'. In: International Joint Conference on Biometrics (IJCB). (IEEE, 2014. pp. 1–7
101 Mukherjee, R., Ross, A. 'Indexing iris images'. In: International Conference on Pattern Recognition (ICPR). (IEEE, 2008. pp. 1–3
102 Sun, Z., Zhang, H., Tan, T., Wang, J.: 'Iris image classification based on hierarchical visual codebook', *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2014, **36**, (6)
103 Nalla, P.R., Chalavadi, K.M.: 'Iris classification based on sparse representations using on-line dictionary learning for large-scale de-duplication applications', *SpringerPlus*, 2015, **4**, (1)
104 Raghavendra, R., Surbiryala, J., Busch, C. 'An efficient finger vein indexing scheme based on unsupervised clustering'. In: International Conference on Identity, Security and Behavior Analysis (ISBA). (IEEE, 2015. pp. 1–8
105 Mhatre, A.J., Palla, S., Chikkerur, S., Govindaraju, V. 'Efficient search and retrieval in biometric databases'. In: Biometric technology for human identification II. vol. 5779. (International Society for Optics and Photonics, 2005. pp. 265–274
106 Pflug, A., Busch, C., Ross, A. '2D ear classification based on unsupervised clustering'. In: International Joint Conference on Biometrics (IJCB). (IEEE, 2014. pp. 1–8
107 Wu, X., Kumar, V., Quinlan, J.R., Ghosh, J., Yang, Q., Motoda, H., et al.: 'Top 10 algorithms in data mining', *Knowledge and information systems*, 2008, **14**, (1), pp. 1–37
108 Jain, A.K., Murty, M.N., Flynn, P.J.: 'Data clustering: A review', *Computing Surveys (CSUR)*, 1999, **31**, (3), pp. 264–323
109 Aggarwal, C.C., Reddy, C.K.: 'Data Clustering: Algorithms and Applications'. (Chapman & Hall/CRC, 2013)
110 Mansukhani, P., Tulyakov, S., Govindaraju, V.: 'A framework for efficient fingerprint identification using a minutiae tree', *Systems Journal*, 2010, **4**, (2), pp. 126–137
111 Dewangan, J., Dey, S., Samanta, D.: 'Face images database indexing for person identification problem', *International Journal of Biometrics and Bioinformatics*, 2013, **7**, (2), pp. 93–122
112 Mehrotra, H., Srinivas, B.G., Majhi, B.: 'Indexing iris biometric database using energy histogram of DCT subbands', *Journal of Communications in Computer and Information Science*, 2009, **40**, pp. 194–204
113 Khalaf, E.T., Mohammed, M., Moorthy, K.: 'Robust partitioning and indexing for iris biometric database based on local features', *IET Biometrics*, 2018,
114 Jayaraman, U., Prakash, S., Gupta, P.: 'An efficient color and texture based iris image retrieval technique', *Expert Systems with Applications*, 2012, **39**, (5), pp. 4915–4926
115 Barbu, T., Luca, M. 'Content-based iris indexing and retrieval model using spatial acces methods'. In: International Symposium on Signals, Circuits and Systems (ISSCS). (IEEE, 2015. pp. 1–4
116 Rathgeb, C., Breitinger, F., Baier, H., Busch, C. 'Towards bloom filter-based indexing of iris biometric data'. In: International Conference on Biometrics (ICB). (IEEE, 2015. pp. 422–429
117 Drozdowski, P., Rathgeb, C., Busch, C. 'Multi-iris indexing and retrieval: Fusion strategies for Bloom filter-based search structures'. In: International Joint Conference on Biometrics (IJCB). (IEEE, 2017. pp. 1–8
118 Drozdowski, P., Rathgeb, C., Busch, C.: 'Bloom filter-based search structures for indexing and retrieving iris-codes', *IET Biometrics*, 2018, **7**, pp. 260–268
119 Damer, N., Terhörst, P., Braun, A., Kuijper, A. 'Indexing of single and multi-instance iris data based on LSH-forest and rotation invariant representation'. In: International Conference on Computer Analysis of Images and Patterns (CAIP). Springer. (IEEE, 2017. pp. 190–201
120 Damer, N., Terhörst, P., Braun, A., Kuijper, A. 'General Borda count for multi-biometric retrieval'. In: International Joint Conference on Biometrics (IJCB). (IEEE, 2017. pp. 420–428
121 Proença, H.: 'Iris biometrics: Indexing and retrieving heavily degraded data', *Transactions on Information Forensics and Security (TIFS)*, 2013, **8**, (12), pp. 1975–1985
122 Jayaraman, U., Prakash, S., Gupta, P. 'Indexing multimodal biometric databases using kd-tree with feature level fusion'. In: International Conference on Information Systems Security. (Springer, 2008. pp. 221–234
123 Wang, K., Yang, L., Su, K., Yang, G., Yin, Y. 'Binary search path of vocabulary tree based finger vein image retrieval'. In: International Conference on Biometrics (ICB). IEEE. (IEEE, 2016. pp. 1–8
124 Mhatre, A., Chikkerur, S., Govindaraju, V. 'Indexing biometric databases using pyramid technique'. In: International Conference on Audio-and Video-Based Biometric Person Authentication. (Springer, 2005. pp. 841–849
125 Gupta, P., Sana, A., Mehrotra, H., Hwang, C.J. 'An efficient indexing scheme for binary feature based biometric database'. In: Biometric technology for human identification IV. vol. 6539. (International Society for Optics and Photonics, 2007. pp. 1–10
126 Shuai, X., Zhang, C., Hao, P. 'Fingerprint indexing based on composite set of reduced SIFT features'. In: International Conference on Pattern Recognition (ICPR). (IEEE, 2008. pp. 1–4
127 He, S., Zhang, C., Hao, P. 'Comparative study of features for fingerprint indexing'. In: International Conference on Image Processing (ICIP). (IEEE, 2009. pp. 2749–2752
128 Cappelli, R., Ferrara, M., Maltoni, D.: 'Fingerprint indexing based on minutia cylinder-code', *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2011, **33**, (5), pp. 1051–1057
129 Yuan, B., Su, F., Cai, A. 'Fingerprint retrieval approach based on novel minutiae triplet features'. In: International Conference on Biometrics: Theory, Applications and Systems (BTAS). (IEEE, 2012. pp. 170–175
130 Wang, Y., Wang, L., Cheung, Y.M., Yuen, P.C.: 'Learning compact binary codes for hash-based fingerprint indexing', *Transactions on Information Forensics and Security (TIFS)*, 2015, **10**, (8), pp. 1603–1616
131 Li, G., Yang, B., Busch, C. 'A fingerprint indexing scheme with robustness against sample translation and rotation'. In: International Conference of the Biometrics Special Interest Group (BIOSIG). (IEEE, 2015. pp. 1–8
132 Kaushik, V.D., Umarani, J., Gupta, A.K., Gupta, P.: 'An efficient indexing scheme for face database using modified geometric hashing', *Neurocomputing*, 2013, **116**, pp. 208–221
133 Mehrotra, H., Majhi, B., Gupta, P.: 'Robust iris indexing scheme using geometric hashing of sift keypoints', *Journal of Network and Computer Applications*, 2010, **33**, (3), pp. 300–313
134 Rathgeb, C., Uhl, A. 'Iris-biometric hash generation for biometric database indexing'. In: International Conference on Pattern Recognition (ICPR). (IEEE, 2010. pp. 2848–2851
135 Jayaraman, U., Gupta, P. 'Iris code hashing'. In: International Conference on Communications (ICC). (IEEE, 2013. pp. 2123–2127
136 Panda, A.K., Mehrotra, H., Majhi, B.: 'Parallel geometric hashing for robust iris indexing', *Journal of real-time image processing*, 2013, **8**, (3), pp. 341–349
137 Badrinath, G.S., Gupta, P., Mehrotra, H.: 'Score level fusion of voting strategy of geometric hashing and SURF for an efficient palmprint-based identification', *Journal of real-time image processing*, 2013, **8**, (3), pp. 265–284
138 Knuth, D.: 'Sorting and searching', *The art of computer programming*, 1998, **3**
139 Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C.: 'Introduction to algorithms'. (MIT press, 2009)
140 Lamdan, Y., Wolfson, H.J. 'Geometric hashing: A general and efficient model-based recognition scheme'. In: International Conference on Computer Vision (ICCV). (IEEE, 1988. pp. 238–249
141 Wolfson, H.J., Rigoutsos, I.: 'Geometric hashing: An overview', *Computational science and engineering*, 1997, **4**, (4), pp. 10–21
142 Cappelli, R., Ferrara, M., Maltoni, D.: 'Minutia cylinder-code: A new representation and matching technique for fingerprint recognition', *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2010, **32**, (12), pp. 2128–2141
143 Schlett, T., Rathgeb, C., Busch, C. 'A binarization scheme for face recognition based on multi-scale block local binary patterns'. In: International Conference of the Biometrics Special Interest Group (BIOSIG). (IEEE, 2016. pp. 1–4
144 Drozdowski, P., Struck, F., Rathgeb, C., Busch, C. 'Benchmarking binarisation schemes for deep face templates'. In: International Conference on Image Processing (ICIP). (IEEE, 2018. pp. 1–5
145 Gentile, J.E., Ratha, N., Connell, J. 'SLIC: short-length iris codes'. In: International Conference on Biometrics: Theory, Applications, and Systems (BTAS). (IEEE, 2009. pp. 1–5
146 Rathgeb, C., Uhl, A., Wild, P. 'On combining selective best bits of iris-codes'. In: European Workshop on Biometrics and Identity Management (BioID). (Springer, 2011. pp. 227–237
147 Jain, A.K., Prabhakar, S., Hong, L., Pankanti, S.: 'Filterbank-based fingerprint matching', *Transactions on Image Processing*, 2000, **9**, (5), pp. 846–859
148 Xu, H., Veldhuis, R.N.J., Kevenaar, T.A.M., Akkermans, A.H.M., Bazen, A.M. 'Spectral minutiae: A fixed-length representation of a minutiae set'. In: Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). (IEEE, 2008. pp. 1–6

149 Yang, J.C., Park, D.S.: 'A fingerprint verification algorithm using tessellated invariant moment features', *Neurocomputing*, 2008, **71**, (10-12), pp. 1939–1946
150 Rathgeb, C., Breitinger, F., Busch, C., Baier, H.: 'On application of bloom filters to iris biometrics', *IET Biometrics*, 2014, **3**, (4), pp. 207–218
151 Damer, N., Terhörst, P., Braun, A., Kuijper, A.: 'Efficient, accurate, and rotation-invariant iris code', *Signal Processing Letters*, 2017, **24**, (8), pp. 1233–1237
152 Kuehlkamp, A., Bowyer, K.: 'Found a good match: Should i keep searching? - accuracy and performance in iris matching using 1-to-first search', *Image and Vision Computing*, 2018, **73**, pp. 17–27
153 Cappelli, R., Ferrara, M., Maio, D.: 'Candidate list reduction based on the analysis of fingerprint indexing scores', *Transactions on Information Forensics and Security (TIFS)*, 2011, **6**, (3), pp. 1160–1164
154 Rathgeb, C., Hofbauer, H., Uhl, A., Busch, C. 'TripleA: Accelerated accuracy-preserving alignment for iris-codes'. In: International Conference on Biometrics (ICB). (IEEE, 2016. pp. 1–8
155 Drozdowski, P., Rathgeb, C., Hofbauer, H., Wagner, J., Uhl, A., Busch, C. 'Towards pre-alignment of near-infrared iris images'. In: International Joint Conference on Biometrics (IJCB). (IEEE, 2017. pp. 359–366
156 Drozdowski, P., Rathgeb, C., Busch, C. 'Turning a vulnerability into an asset: Accelerating facial identification with morphing'. In: International Conference on Acoustics, Speech, and Signal Processing (ICASSP). (IEEE, 2019. pp. 1–5
157 Lim, M.H., Teoh, A.B.J., Kim, J.: 'Biometric feature-type transformation: Making templates compatible for secret protection', *IEEE Signal Processing Magazine*, 2015, **32**, (5), pp. 77–87
158 Singh, M., Singh, R., Ross, A.: 'A comprehensive overview of biometric fusion', *Information Fusion*, 2019, **52**, pp. 187–205
159 Berten Digital Signal Processing. 'GPU vs FPGA performance comparison'. (Berten DSP, 2016. BWP001 v1.0
160 Singh, D., Reddy, C.K.: 'A survey on platforms for big data analytics', *Journal of big data*, 2015, **2**, (1), pp. 1–20
161 Rathgeb, C., Buchmann, N., Hofbauer, H., Baier, H., Uhl, A., Busch, C.: 'Methods for accuracy-preserving acceleration of large-scale comparisons in CPU-based iris ecognition systems', *IET Biometrics*, 2018, **7**, (4), pp. 356–364
162 López, M., Daugman, J., Cantó, E.: 'Hardware–software co-design of an iris recognition algorithm', *IET Information Security*, 2011, **5**, (1), pp. 60–68
163 Saegusa, T., Maruyama, T., Yamaguchi, Y. 'How fast is an FPGA in image processing?'. In: International Conference on Field Programmable Logic and Applications. (IEEE, 2008. pp. 77–82
164 Sirowy, S., Forin, A. 'Where's the beef? why FPGAs are so fast'. (Microsoft Research, 2008. MSR-TR-2008-130
165 Andina, J.J.R., De la Torre Arnanz, E., Valdes, M.D.: 'FPGAs: Fundamentals, Advanced Features, and Applications in Industrial Electronics'. (CRC Press, 2017)
166 Tessier, R., Pocek, K., DeHon, A.: 'Reconfigurable computing architectures', *Proc of the IEEE*, 2015, **103**, (3), pp. 332–354
167 García, G.J., Jara, C.A., Pomares, J., Alabdo, A., Poggi, L.M., Torres, F.: 'A survey on FPGA-based sensor systems: towards intelligent and reconfigurable low-power sensors for computer vision, control and signal processing', *Sensors*, 2014, **14**, (4), pp. 6247–6278
168 Chen, A.T.J., Biglari-Abhari, M., Wang, K.I.K., , Bouzerdoum, A., Tivive, F.H.C.: 'Convolutional neural network acceleration with hardware/software co-design', *Applied Intelligence*, 2018, **48**, (5), pp. 1288–1301
169 Rakvic, R.N., Ulis, B.J., Broussard, R.P., Ives, R.W., Steiner, N.: 'Parallelizing iris recognition', *Transactions on Information Forensics and Security (TIFS)*, 2009, **4**, (4), pp. 812–823
170 Fons, M., Fons, F., Cantó, E.: 'Fingerprint image processing acceleration through run-time reconfigurable hardware', *Transactions on Circuits and Systems II: Express Briefs*, 2010, **57**, (12), pp. 991–995
171 Stekas, N., v. d. Heuvel, D. 'Face recognition using local binary patterns histograms (LBPH) on an FPGA-based system on chip (SoC)'. In: International Parallel and Distributed Processing Symposium Workshops (IPDPSW). (IEEE, 2016. pp. 300–304
172 Jadhav, M., Nerkar, P.M. 'Implementation of an embedded hardware of FVRS on FPGA'. In: International Conference on Information Processing (ICIP). (IEEE, 2015. pp. 48–53
173 Koukounis, D., Ttofis, C., Papadopoulos, A., Theocharides, T.: 'A high performance hardware architecture for portable, low-power retinal vessel segmentation', *Integration, the VLSI journal*, 2014, **47**, (3), pp. 377–386
174 Bouraoui, H., Jerad, C., Chattopadhyay, A., Hadj-Alouane, N.B.: 'Hardware architectures for embedded speaker recognition applications: A survey', *Transactions on Embedded Computing Systems (TECS)*, 2017, **16**, (3), pp. 78
175 Nickolls, J., Buck, I., Garland, M., Skadron, K. 'Scalable parallel programming with CUDA'. In: SIGGRAPH classes. (ACM, 2008. pp. 42–53
176 Stone, J.E., Gohara, D., Shi, G.: 'OpenCL: A parallel programming standard for heterogeneous computing systems', *Computing in Science Engineering*, 2010, **12**, (3), pp. 66–73
177 Mittal, S., Vetter, J.S.: 'A survey of CPU-GPU heterogeneous computing techniques', *Computing Surveys (CSUR)*, 2015, **47**, (4), pp. 1–35
178 Lastra, M., Gutiérrez, P.D., Benítez, J.M., Herrera, F.: 'GPU processing for biometric big data based identification. why and what for?', *Biostatistics and Biometrics Open Access Journal*, 2017, **2**, pp. 1–4
179 Barrus, J.. 'Cloud TPU machine learning accelerators now available in beta'. (Google, 2018. last accessed: 11/6/2019. `https://cloud.google.com/blog/products/gcp/cloud-tpu-machine-learning-accelerators-now-available-in-beta`
180 Li, X., Zhang, G., Huang, H.H., Wang, Z., Zheng, W. 'Performance analysis of GPU-based convolutional neural networks'. In: International Conference on Parallel Processing (ICPP). (IEEE, 2016. pp. 67–76
181 Schroff, F., Kalenichenko, D., Philbin, J. 'FaceNet: A unified embedding for face recognition and clustering'. In: Conference on Computer Vision and Pattern Recognition. (IEEE, 2015. pp. 815–823
182 Vandal, N.A., Savvides, M. 'CUDA accelerated iris template matching on graphics processing units (GPUs)'. In: International Conference on Biometrics: Theory, Applications and Systems (BTAS). (IEEE, 2010. pp. 1–7
183 Ghafoor, M., Iqbal, S., Tariq, S.A., Taj, I.A., Jafri, N.M.: 'Efficient fingerprint matching using GPU', *IET Image Processing*, 2018, **12**, (2), pp. 274–284
184 Yi, S., Yoon, I., Oh, C., Yi, Y. 'Real-time integrated face detection and recognition on embedded GPGPUs'. In: Symposium on Embedded Systems for Real-time Multimedia (ESTIMedia). (IEEE, 2014. pp. 98–107
185 Lin, Y., Du, E.Y., Zhou, Z., Thomas, N.L.: 'An efficient parallel approach for sclera vein recognition', *Transactions on Information Forensics and Security (TIFS)*, 2014, **9**, (2), pp. 147–157
186 Rife, D.C.: 'Finger prints as criteria of ethnic relationship', *American journal of human genetics*, 1953, **5**, (4), pp. 389
187 Moses, K.R., Higgins, P., McCabe, M., Probhakar, S., Swann, S. Automated Fingerprint Identification System (AFIS). In: McRoberts, A., editor. 'Fingerprint sourcebook'. (US Department of Justice, 2010. pp. 1–33
188 Federal Bureau of Investigation: 'The Science of Fingerprints: Classification and Uses'. (General Press, 2013)
189 Nandakumar, K., Jain, A.K.: 'Biometric template protection: Bridging the performance gap between theory and practice', *Signal Processing Magazine*, 2015, **32**, (5), pp. 88–100
190 European Parliament: 'Regulation (EU) 2016/679', *Official Journal of the European Union*, 2016, **L119**, pp. 1–88
191 ISO/IEC JTC1 SC27 IT Security techniques. 'ISO/IEC 24745:2011. information technology – security techniques – biometric information protection'. (International Organization for Standardization, 2011
192 Rathgeb, C., Uhl, A.: 'A survey on biometric cryptosystems and cancelable biometrics', *EURASIP J on Information Security*, 2011,
193 Li, G., Yang, B., Busch, C. 'A fingerprint indexing algorithm on encrypted domain'. In: Trustcom/BigDataSE/ISPA. (IEEE, 2016. pp. 1030–1037
194 Drozdowski, P., Garg, S., Rathgeb, C., Gomez-Barrero, M., Chang, D., Busch, C. 'Privacy-preserving indexing of Iris-Codes with cancelable Bloom filter-based search structures'. In: European Signal Processing Conference (EUSIPCO). (IEEE, 2018. pp. 1–5
195 Wang, Y., Wan, J., Guo, J., Cheung, Y., Yuen, P.C.: 'Inference-based similarity search in randomized Montgomery domains for privacy-preserving biometric identification', *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2018, **40**, (7), pp. 1611–1624
196 Simmhan, Y., Shukla, A., Verma, A. 'Benchmarking fast-data platforms for the Aadhaar biometric database'. In: Big Data Benchmarking. (Springer, 2015. pp. 21–39
197 Al-Raisi, A.N., Al-Khouri, A.M.: 'Iris recognition and the challenge of homeland and border control security in UAE', *Telematics and Informatics*, 2008, **25**, (2), pp. 117–132
198 Daugman, J., Malhas, I.: 'Iris recognition border-crossing system in the UAE', *International Airport Review*, 2004, **8**, (2), pp. 1–5
199 Daugman, J.: 'How iris recognition works', *Transactions on Circuits and Systems for Video Technology (TCSVT)*, 2004, **14**, (1), pp. 21–30
200 Gorodnichy, D.O., Chumakov, M.P.: 'Analysis of the effect of ageing, age, and other factors on iris recognition performance using NEXUS scores dataset', *IET Biometrics*, 2018,
201 Komarinski, P.: 'Automated Fingerprint Identification Systems (AFIS)'. (Elsevier, 2005)
202 Gemalto. 'Automated Fingerprint Identification System (AFIS) - a short history'. (Gemalto Case Studies, 2019. last accessed: 11/6/2019. `https://www.gemalto.com/govt/biometrics/afis-history`
203 Grother, P., Ngan, M., Hanaoka, K. 'Ongoing face recognition vendor test (FRVT) part 2: Identification'. (National Institute of Standards and Technology, 2018. NISTIR 8238
204 Jain, A., Klare, B., Ross, A. 'Guidelines for best practices in biometrics research'. In: International Conference on Biometrics (ICB). (IEEE, 2015. pp. 541–545
205 Murphy, T.M., Broussard, R., Rakvic, R., Ngo, H., Ives, R.W., Schultz, R., et al.: 'Use of synthetic data to test biometric algorithms', *Journal of Electronic Imaging*, 2016, **25**, (4), pp. 043023
206 Marcel, S.: 'BEAT – biometrics evaluation and testing', *Biometric technology today*, 2013, **2013**, (1), pp. 5–7
207 Cappelli, R., Ferrara, M., Maltoni, D.. 'FIDXICB-2013'. (University of Bologna Biometric System Laboratory, 2013. last accessed: 11/6/2019. `https://biolab.csr.unibo.it/fvcongoing/UI/Form/ICB2013FIDX.aspx`
208 National Institute of Standards and Technology. 'Face Recognition Vendor Test (FRVT) 1:N Evaluation'. (NIST, 2017. last accessed: 11/6/2019. `https://www.nist.gov/programs-projects/face-recognition-vendor-test-frvt-1n-2018-evaluation`