Database Binning and Retrieval in Multi-Fingerprint Identification Systems

P. Drozdowski*†, D. Fischer*, C. Rathgeb*, C. Schiel‡, and C. Busch*
* da/sec - Biometrics and Internet Security Research Group, Hochschule Darmstadt, Germany
† Norwegian Biometrics Laboratory, NTNU, Gjøvik, Norway
‡ Bundeskriminalamt (BKA), Wiesbaden, Germany

{pawel.drozdowski,daniel.fischer,christian.rathgeb,christoph.busch}@h-da.de
Christopher.Schiel@bka.bund.de

Abstract

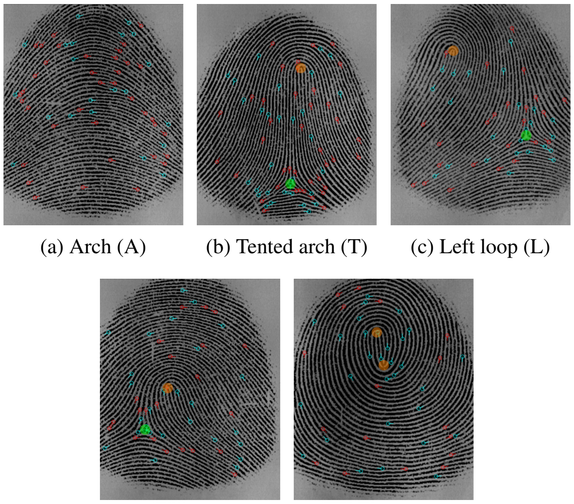
The increasingly large scale of deployed biometric systems necessitates approaches for computational workload reduction in order to perform identification queries efficiently. Simple database binning based on classification of features in biometric samples is amongst the most frequently used and researched methods for achieving said goal. However, multi-instance database binning appears to be a neglected topic in the scientific literature: best to the authors' knowledge, for fingerprints there exists only one, entirely theoretical, study on this subject. In this paper, we propose a retrieval algorithm based on multi-instance binning of fingerprint databases, along with usage of statistical information on fingerprint classes and their correlations.

The aforementioned statistics are obtained from NIST SD9 database and data obtained from the German Federal Criminal Police Office. Subsequently, the experimental evaluation of the proposed algorithm is performed on the NIST SD9 database. The proposed system is evaluated using a classifier based on the PCASYS tool and neuronal networks. The results show a significant workload reduction from a baseline exhaustive search scenario – down to 12.7% for this particular classifier and 5.8% for a theoretical perfect (completely accurate) classifier. The proposed method could be seamlessly integrated into operational systems, as it relies on well-established features and compatibility with the current acquisition methods.

1. Introduction

Nowadays, biometric technologies are already deployed in numerous nation-wide large-scale applications, such as the Indian Aadhaar project [21]. With the rapid growth of biometric systems' sizes and popularity, technologies supporting efficient and accurate processing of large amounts of biometric data are vital in order to guarantee practical response times. Conventional biometric systems require exhaustive one-to-many comparisons in order to identify biometric probes, *i.e.* comparison time frequently dominates the overall computational workload of an identification attempt. In past years, researchers have invested significant efforts to tackle the challenge of computational workload reduction in biometric identification systems. Basically, four different key concepts can be distinguished: *classification* or "binning", *indexing*, a *serial combination* of a computationally efficient and an accurate (but more complex) algorithm and *hardware-based acceleration*. Depending on the used biometric characteristic, the vast majority of classification approaches are designed to reliably extract human understandable attributes from a biometric sample, *e.g.* sex or ethnicity for face. While not necessarily unique to an individual, those attributes allow for a binning of biometric databases according to a predefined number of classes, *i.e.* the search space ($\widehat{=}$ computational workload) for a given biometric probe can be reduced to one (or a few) bin(s). In contrast, biometric indexing approaches introduce hierarchical search structures (tolerating a certain amount of biometric variance), where the process of search space reduction might not be reproducible by human experts. Lastly, the latter two categories do not aim at reducing the complexity of an identification attempt but response times.

Focusing on fingerprint recognition systems, the classification model of Henry [12] has been widely used by researchers, as well as commercial vendors, for computational workload reduction in identification scenarios. The five fingerprint classes (or types), *i.e. arch*, *tented arch*, *right loop*, *left loop* and *whorl*, which are depicted in figure 1, are unevenly distributed in the population. Fingerprint classes are mainly determined based on the global (level-1) features, in particular ridge line flow (orientation map) and the singular points, *i.e.* core and delta, derived from it. Numerous approaches, which either directly employ or further process those features, have been proposed for the purpose of distinguishing between said classes. For more



(a) Arch (A) (b) Tented arch (T) (c) Left loop (L)

(d) Right loop (R) (e) Whorl (W)

Figure 1: Example fingerprints for each of the five classes displaying minutiae, core and delta points (images generated using Synthetic Fingerprint Generator (SFinGe) [5])

details on the topic of fingerprint classification and a comprehensive survey of proposed approaches, the reader is referred to [9, 10]. State-of-the-art fingerprint classification schemes obtain near-optimal classification accuracy. Table 1 summarises most notable approaches and reported results in terms of Correct Classification Rate (CCR) of the last five years. Note, that all of these classification approaches aim to determine the class of a *single* fingerprint. As opposed to the existing literature, this paper investigates fingerprint classification in multi-finger identification systems. This is motivated by the facts that large-scale identification systems leverage the information of multiple fingerprints of data subjects, *e.g*. [21], and modern fingerprint capture devices can acquire multiple fingerprints of data subject's hand simultaneously, *e.g*. [13]. It is well-known that the classes of fingerprints obtained from one hand are highly correlated. Nevertheless, to the best of the authors' knowledge, the potential of multi-fingerprint database binning has only been theoretically analysed by Wayman [23]. It was confirmed that bins formed by combinations of fingerprint classes highly vary in probability. Moreover, theoretical estimations about expected penetration rates are reported. However, so far the potential of improving the overall fingerprint retrieval accuracy by consolidating information obtained from single fingerprint classification scores has been neglected. In this work, we obtain universally valid statistics of fingerprint class distributions and correlations from two datasets, namely the NIST SD9 [17] and an in-house database of the German Federal Criminal Police Office (BKA). Those statistics are used to effectively retrieve bins representing combinations of fingerprint classes according to their likelihood. In experiments on the SD9 database the well-established, publicly available Pattern-level Classification Automation SYStem (PCASYS) tool [3] in conjunction with a neuronal network-based classifier are employed for the purpose of fingerprint classification. The proposed approach is shown to substantially reduce the computational workload by combining the classifier scores obtained from up to five fingers of a data subject's hand.

Table 1: Most relevant fingerprint classification approaches proposed in the last five years

Ref.	Year	Method	Database(s)	Classes	CCR	Reject
[4]	2013	MCC	SD4	4 / 5	97.2% / 95.9%	-
[11]	2014	FCA	FVC00/02/04	4	92.74%	-
[15]	2014	KNN	SD4	4 / 5	96.8% / 94.6%	-
[24]	2014	FCA	FVC02-1	5	91.1%	-
			FVC04-1	5	91.8%	-
[8]	2015	FCA	SD4	5	80.51%	12%
			FVC02-1	5	90.11%	-
			FVC04-1	5	88.98%	-
[14]	2015	RDM	FVC00	4	91.1%	-
			FVC02	4	97.8%	-
			FVC04	4	97.3%	-
[7]	2016	FCA	SD4	4 / 5	88.3% / 92.13%	-
[22]	2016	ANN	SD4	4	91.4% / 93.1%	-
[2]	2017	ANN	FVC2000	-	97.56%	-
[19]	2017	MCC	SD4	5	92.97%	-
			SD14	5	93.76%	-
			SFinGe	5	94.38%	-

MCC ... multiple classifier combination
FCA ... fixed classifier approach
KNN ... k-nearest neighbour
RDM ... ridge distribution models
ANN ... artificial neuronal networks

The remainder of this paper is organised as follows: fingerprint class distributions and correlations are analysed in section 2. Section 3 describes the proposed multi-finger binning and retrieval approach. Experimental results are reported in section 4. Conclusions are drawn in section 5.

2. Fingerprint class statistics

In the following subsections, the used databases are presented, along with statistical distributions of fingerprint classes and their correlations.

2.1. Databases

Two databases were used for experiments in this paper:

SD9 NIST Special Database 9 [17], containing fingerprint images from scanned/rolled-ink ten-print cards. 2 samples per finger are available for each of the 2,700 subjects, hence the total number of images is 54,000. Fingerprint class annotations made by professional forensic examiners are included. Example images from the database are shown in figure 2.

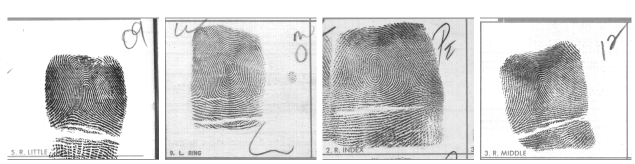


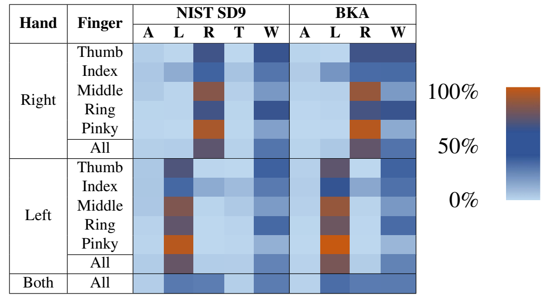
Figure 2: Sample images from the SD9 database

Table 2: Fingerprint class distributions

(a) Percentages

Hand	Finger	NIST SD9					BKA			
		A	L	R	T	W	A	L	R	W
Right	Thumb	3.49%	0.71%	48.94%	0.22%	46.64%	1.49%	0.57%	49.02%	48.92%
	Index	5.61%	14.72%	39.43%	7.06%	33.18%	4.23%	22.41%	36.62%	36.74%
	Middle	4.76%	1.30%	69.48%	2.94%	21.52%	2.17%	2.66%	74.24%	20.93%
	Ring	1.19%	1.41%	49.61%	1.19%	46.60%	0.65%	1.56%	50.84%	46.95%
	Pinky	0.93%	0.19%	79.41%	0.82%	18.65%	0.38%	0.56%	83.47%	15.59%
	All	3.20%	3.66%	57.37%	2.45%	33.32%	1.79%	5.55%	58.84%	33.82%
Left	Thumb	5.50%	53.31%	0.93%	0.48%	39.78%	2.60%	57.75%	0.42%	39.23%
	Index	5.84%	37.72%	15.17%	9.63%	31.64%	3.60%	45.48%	16.35%	34.57%
	Middle	5.61%	67.36%	1.49%	5.06%	20.48%	2.66%	74.01%	1.70%	21.63%
	Ring	1.90%	58.66%	0.48%	1.67%	37.29%	0.81%	62.00%	0.62%	36.57%
	Pinky	1.26%	83.95%	0.22%	1.08%	13.49%	0.47%	88.07%	0.19%	11.27%
	All	4.02%	60.20%	3.66%	3.58%	28.54%	2.03%	65.46%	3.86%	28.65%
Both	All	3.61%	31.93%	30.52%	3.01%	30.93%	1.91%	35.76%	31.11%	31.22%

(b) Heatmap



BKA A subset of the Automated Fingerprint Identification Systems (AFIS) data of the BKA consisting of fingerprint type statistical data from around 26,000 randomly selected subjects. Due to lack of actual images (data protection restrictions), this dataset was only used to validate the statistical results obtained on SD9 and not the computational workload reduction experiments. The data does not distinguish between arches and tented arches; instead classifying them together into one class.

The subjects in both databases were selected from their respective AFIS' randomly, hence ensuring a natural distribution of the fingerprint classes.

2.2. Distributions and Correlations

For the statistical analysis, only a single (first) sample from each finger is considered in order to avoid using redundant information. The class distributions for the SD9 and BKA datasets can be seen in table 2a, while table 2b presents the same information graphically in a heatmap format. It can be observed, that the loop classes are the most prevalent; overwhelmingly, their direction corresponds to the hand of the given finger (left loops on left hand and analogously for the right hand). Together with whorls, they account for around 95% of the total samples. The fingerprint class distributions (both for overall percentages and individual fingers) obtained from both datasets tend to coincide. The largest (relative) discrepancies can be seen for the arches and tented arches. Overall, however, the findings from both datasets are very similar, which suggests the generality of the results.

As previously mentioned, the classes of fingers of one hand are also known to be correlated. Tables 3 to 6 show the occurrence frequency for the most prevalent (top 10 from SD9) class combinations between contiguous sequences of two to five adjacent fingers of each hand. For example, RL corresponds to a right and left loop fingerprint class (recall figure 1), for the pair of fingers noted in the table header. It can be observed, that combinations of loops and whorls again account for the vast majority of cases; furthermore, it is very often the case that same fingerprint classes are seen across multiple or even all fingers of a given hand. As was the case for the class distributions described earlier, the results for the class combinations within both datasets tend to largely coincide.

Table 3: Distributions of fingerprint class combinations for two contiguous fingers

Hand	Thumb, index			Index, middle			Middle, ring			Ring, pinky		
	Cls.	SD9	BKA	Cls.	SD9	BKA	Cls.	SD9	BKA	Cls.	SD9	BKA
Right	WW	26.53%	25.35%	RR	35.12%	36.66%	RR	41.66%	46.58%	RR	46.93%	49.70%
	RR	25.94%	26.31%	WW	16.69%	17.04%	RW	25.75%	26.00%	WR	29.73%	30.93%
	WR	11.82%	12.33%	WR	15.98%	19.14%	WW	19.47%	18.74%	WW	16.76%	14.69%
	RL	8.21%	11.57%	LR	10.33%	14.89%	AR	2.90%	1.76%	RW	1.67%	1.47%
	RW	6.43%	11.11%	TR	5.65%	—	TR	2.30%	—	TR	1.11%	—
	WL	6.02%	7.14%	LW	2.45%	2.07%	WR	2.04%	1.81%	LR	1.04%	1.13 %
	RT	5.13%	—	AA	2.45%	1.28%	RL	1.15%	0.94%	AR	0.59%	0.54%
	RA	3.23%	2.64%	AR	2.42%	3.01%	AA	0.93%	0.54%	RT	0.52%	—
	WT	1.64%	—	RW	2.16%	1.53%	RT	0.74%	—	AA	0.41%	0.11%
	AA	1.45%	0.71%	RT	1.00%	—	LR	0.71%	1.79%	RA	0.37%	0.37%
	Other	3.60%	2.84%	Other	5.75%	4.38%	Other	2.35%	1.84%	Other	0.87%	1.06%
Left	LL	24.05%	32.50%	LL	32.42%	40.77%	LL	46.51%	52.45%	LL	55.58%	58.05%
	WW	21.56%	20.83%	WW	16.17%	17.17%	LW	19.41%	18.74%	WL	25.69%	27.18%
	WL	11.34%	11.99%	WL	14.91%	16.60%	WW	17.10%	18.34%	WW	11.38%	10.41%
	LR	9.59%	9.44%	RL	10.71%	12.53%	TL	4.13%	—	LW	2.04%	1.60%
	LW	9.48%	13.13%	TL	7.21%	—	AL	3.42%	2.69%	TL	1.49%	—
	LT	7.10%	—	AA	2.94%	1.83%	WL	3.38%	4.01%	AL	0.93%	0.63%
	WR	4.91%	5.69%	LW	2.64%	3.32%	AA	1.38%	0.69%	AA	0.71%	0.31%
	LA	3.09%	2.58%	AL	2.12%	2.32%	RL	1.23%	1.09%	LT	0.48%	—
	AA	2.04%	0.86%	RW	1.45%	1.95%	LT	0.71%	—	LA	0.45%	0.43%
	AL	1.93%	1.32%	RT	1.45%	—	AT	0.56%	—	AT	0.26%	—
	Other	4.91%	1.66%	Other	7.98%	3.51%	Other	2.17%	1.99%	Other	0.99%	1.39%

Table 4: Distributions of fingerprint class combinations for three contiguous fingers

Hand	Thumb, index, middle			Index, middle, ring			Middle, ring, pinky		
	Cls.	SD9	BKA	Cls.	SD9	BKA	Cls.	SD9	BKA
Right	RRR	23.78%	24.47%	RRR	24.19%	25.43%	RRR	39.69%	44.77%
	WWW	15.01%	14.09%	WWW	15.46%	15.76%	RWR	18.06%	19.05%
	WWR	11.11%	11.00%	WRW	10.89%	11.14%	WWR	10.78%	10.94%
	WRR	10.00%	11.34%	RRW	9.92%	10.63%	WWW	8.70%	7.77%
	RLR	6.61%	9.84%	LRR	6.54%	10.86%	RWW	7.69%	6.72%
	RWR	4.72%	8.00%	WRR	5.02%	7.91%	ARR	2.42%	1.62%
	RTR	4.27%	—	TRR	4.46%	—	TRR	2.19%	—
	WLR	3.42%	4.82%	LRW	3.49%	3.84%	WRR	1.97%	1.67%
	WLW	2.04%	1.36%	LWW	2.04%	1.70%	RRW	1.52%	1.30%
	RWW	1.60%	2.89%	RWW	1.82%	1.28%	RLR	0.82%	0.77%
	Other	17.44%	12.19%	Other	16.17%	11.45%	Other	6.16%	5.39%
Left	LLL	21.26%	29.64%	LLL	23.35%	31.50%	LLL	44.24%	50.82%
	WWW	12.08%	12.36%	WWW	14.13%	14.54%	LWL	14.28%	14.85%
	WLL	9.37%	9.99%	RLL	8.85%	10.27%	WWL	10.97%	11.87%
	WWL	9.11%	8.27%	LLW	8.62%	8.67%	WWW	6.13%	6.47%
	LRL	7.10%	8.07%	WLW	8.40%	7.87%	LWW	4.94%	3.83%
	LWL	5.43%	8.13%	WLL	6.51%	8.64%	TLL	3.98%	—
	LTL	5.32%	—	TLL	5.99%	—	WLL	3.20%	3.75%
	LWW	3.87%	4.81%	LWW	2.12%	2.40%	ALL	3.01%	2.46%
	WRL	3.20%	4.18%	WWL	2.04%	2.55%	LLW	1.71%	1.34%
	ALL	1.52%	2.46%	ALL	1.82%	2.03%	RLL	1.15%	1.03%
	Other	21.74%	12.09%	Other	18.17%	11.53%	Other	6.39%	3.58%

Table 5: Distributions of fingerprint class combinations for four contiguous fingers

Hand	Thumb, index, middle, ring			Index, middle, ring, pinky		
	Cls.	**SD9**	**BKA**	**Cls.**	**SD9**	**BKA**
Right	RRRR	18.99%	18.34%	RRRR	23.19%	24.35%
	WWWW	14.12%	13.21%	WWWR	8.18%	8.70%
	WWRW	8.10%	6.72%	WWWW	7.28%	7.03%
	WRRW	5.46%	4.82%	RRWR	6.91%	7.97%
	RLRR	5.02%	7.80%	WRWR	6.76%	7.57%
	WRRR	4.31%	6.44%	LRRR	6.21%	10.69%
	RRRW	4.12%	5.61%	WRRR	4.50%	7.54%
	RTRR	3.53%	—	TRRR	4.38%	—
	WWRR	3.01%	4.05%	WRWW	4.12%	3.52%
	RWRW	2.71%	4.14%	LRWR	3.08%	3.09%
	Other	30.63%	28.87%	Other	25.39%	19.54%
Left	LLLL	16.51%	24.14%	LLLL	22.08%	30.70%
	WWWW	10.86%	10.79%	WWWL	8.85%	9.36%
	LRLL	6.32%	6.87%	RLLL	8.70%	9.99%
	WWLW	5.87%	4.29%	LLWL	6.39%	7.18%
	WLLL	5.61%	6.44%	WLLL	6.10%	8.21%
	LTLL	4.65%	—	WLWL	5.91%	5.89%
	LLLW	4.54%	5.04%	TLLL	5.76%	—
	WLLW	3.64%	3.49%	WWWW	5.28%	5.18%
	WWLL	3.23%	3.95%	WLWW	2.45%	1.95%
	LWWW	3.12%	3.75%	LLWW	2.16%	1.46%
	Other	35.65%	31.24%	Other	26.32%	20.08%

Table 6: Distributions of fingerprint class combinations for five contiguous fingers

Hand	Thumb, index, middle, ring, pinky		
	Cls.	**SD9**	**BKA**
Right	RRRRR	18.43%	17.75%
	WWWWR	7.28%	6.97%
	WWWWW	6.84%	6.21%
	WWRWR	4.76%	4.28%
	RLRRR	4.76%	7.68%
	WRRRR	3.86%	6.10%
	RTRRR	3.53%	—
	RRRWR	3.42%	4.20%
	WWRWW	3.34%	2.61%
	WRRWR	3.23%	3.57%
	Other	40.55%	40.63%
Left	LLLLL	16.06%	23.58 %
	WWWWL	6.51%	6.61%
	LRLLL	6.25%	6.72%
	WLLLL	4.87%	6.21%
	LTLLL	4.42%	—
	WWWWW	4.35%	4.18%
	WWLWL	3.90%	3.29%
	LLLWL	3.61%	4.29%
	WWLLL	2.94%	3.69%
	LWLLL	2.94%	4.41%
	Other	44.15%	37.02%

3. Multi-fingerprint binning and retrieval

Figure 3 shows the overview of the proposed multi-fingerprint binning and retrieval algorithm. The binning step consists of enumerating all possible bins based on the combinations of fingerprint classes and accordingly assigning the data from the enrolled subjects to the bins. Subsections 3.1 and 3.2 describe the combination and adjustment of classifier outputs, as well as the utilised retrieval strategy.

3.1. Combining classifier outputs

1. For each of the $\{n|2 \leq n \leq 5\}$ considered fingers of a hand, the classifier produces a list of k classification probabilities (p), where k is the number of possible fingerprint classes and $\sum_{i=0}^{k} p_i = 1$. In other words, a

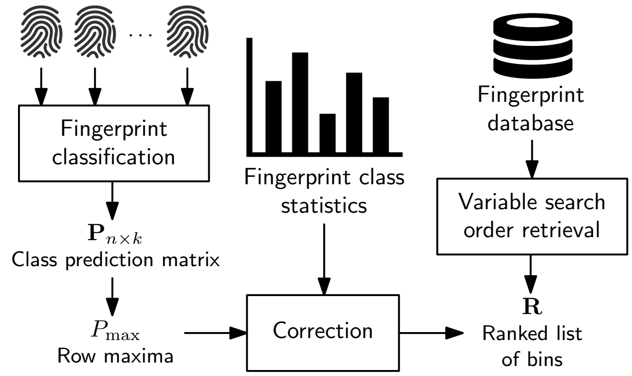


Figure 3: System overview

$\mathbf{P}_{n\times k}$ matrix of class predictions is obtained for the given hand, which may, for example, look as follows for $n = 5$ and $k = 4$:

$$\mathbf{P} = \begin{array}{c} \\ \text{Thumb} \\ \text{Index} \\ \text{Middle} \\ \text{Ring} \\ \text{Pinky} \end{array} \begin{array}{c} \begin{array}{cccc} \text{A} & \text{L} & \text{R} & \text{W} \end{array} \\ \begin{bmatrix} 4\% & 92\% & 1\% & 3\% \\ 1\% & 2\% & 1\% & 96\% \\ 5\% & 76\% & 2\% & 17\% \\ 1\% & 0\% & 0\% & 99\% \\ 1\% & 12\% & 2\% & 85\% \end{bmatrix} \end{array} \tag{1}$$

2. For each finger, the probability of the most probable class is determined (*i.e.* the row-wise maximum values). In this case:

$$P_{max} = \begin{array}{c} \begin{array}{ccccc} \text{Thumb} & \text{Index} & \text{Middle} & \text{Ring} & \text{Pinky} \end{array} \\ \begin{bmatrix} 92\% & 96\% & 76\% & 99\% & 85\% \end{bmatrix} \end{array} \tag{2}$$

3. All possible combinations (Cartesian product) of class labels for all fingers and corresponding probabilities taken from $\mathbf{P}$ are recorded in matrix $\mathbf{B}_{k^n \times n}$:

$$\mathbf{B} = \begin{array}{c} \text{Bin} \\ AAAAA \\ AAAAL \\ ... \\ WWWWW \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Thumb} & \text{Index} & \text{Middle} & \text{Ring} & \text{Pinky} \end{array} \\ \begin{bmatrix} 4\% & 1\% & 5\% & 1\% & 1\% \\ 4\% & 1\% & 5\% & 1\% & 12\% \\ \dots & \dots & \dots & \dots & \dots \\ 3\% & 96\% & 17\% & 99\% & 85\% \end{bmatrix} \end{array} \tag{3}$$

For example, the second bin ($AAAAL$) consists of the A class probabilities for the thumb, index, middle, and ring fingers, along with the L class probability for the pinky finger taken from $\mathbf{P}$.

4. In order to fix unreliable classifier outputs, for each row (*i.e.* possible bin) in the above matrix, a classification correction algorithm is ran. It works based on the previously described correlation statistics (subsection 2.2). For every classification probability in the given bin, if the value is below a threshold (previously estimated on a disjoint training set), the statistical data is used to adjust it. In the current case ($n = 5$), data from

table 6 is retrieved. For instance, if the algorithm is correcting the second (index finger) probability in the bin $AAAAA$:

(a) The statistics for itself, along with $ALAAA$, $ARAAA$, and $AWAAA$ would be retrieved.

(b) Subsequently, the sum of the statistical probabilities (p) for each of those bins is computed, *i.e.* $s = \text{sum}\,(p_{AAAAA}, p_{ALAAA}, p_{ARAAA}, p_{AWAAA})$.

(c) Finally, the probability for the bin is divided by said sum and normalised by the maximum probability for the finger currently under processing (derived in step 2), *i.e.* $\frac{p_{AAAAA}}{s} * P_{max}(index)$.

In other words, the probabilities below the threshold are considered to contain no useful/significant information regarding the classification output, so the global (statistical) information is incorporated in a normalised manner to complement the classifier.

5. The probabilities are summed row-wise, and normalised to the interval $[0, 1] \in \mathbb{R}$. Thus, for each bin in the final list, the overall probability that it matches the fingerprint classes of the probe is recorded.

$$\mathbf{O} = \begin{array}{c} \text{Bin} \\ AAAAA \\ AAAAL \\ \dots \\ WWWWW \end{array} \begin{array}{c} \text{Probability} \\ \left[\begin{array}{c} 2\% \\ 0.5\% \\ \dots \\ 50\% \end{array}\right] \end{array} \tag{4}$$

3.2. Retrieval strategy

The *variable search order* strategy [16] is employed in the retrieval step. The previously acquired list of bin probabilities ($\mathbf{O}$) for the probe is first sorted in descending order of bin occurrence probability, thus producing a ranked list of bins ($\mathbf{R}$). Subsequently, the corresponding bins in the enrolment database are successively searched using the one-to-first strategy, *i.e.* until a match is found, whereupon the retrieval is concluded immediately.

4. Performance evaluation

The following subsections describe the experimental setup, the used fingerprint classification method, and the obtained results.

4.1. Experimental setup

Performance evaluations are conducted on the previously described SD9 database. A ten-fold cross-validation with randomly chosen disjoint training (20%) and test (80%) sets is performed using scikit-learn [18]. Classification accuracy is measured in terms of CCR, while the computational workload reduction is estimated in terms of the number of visited database bins and corresponding subjects for the identification transactions.

4.2. Feature extractor and classifiers

To facilitate the reproducibility of presented results, the publicly available PCASYS tool is employed for fingerprint classification. Extracted feature vectors comprise 128 elements, which are further processed using the Keras Framework [6] with Tensorflow 1.7 [1]. In order to obtain a suitable classifier input feature vector, the elements are normalised and scaled to the range $[-2, 2] \in \mathbb{R}$ using the Keras MinMaxScaler function. For the classification task, a neuronal network-based classifier[1], *i.e.* a Multi-Layer Perceptron, is trained. The network consists of three (hidden) dense layers (192/64/32 nodes), each with a ReLU activation kernel, which are initialised with the RandomNormal initialiser. The output layer comprises four nodes and is initialised with zeroes. Four fingerprint classes are used, *i.e.* arch and tented arch are represented as one class due to their rare occurrences. The output of the classifier is determined by the Sigmoid activation function. In the training (learning) step, a stochastic gradient descent is used with a learning rate of 0.005, a beta1 of 0.95 and a beta2 of 0.999. Training feature vectors are shuffled once and subsequently 150 epochs are performed with a batch size of 64.

Table 7: CCR at a confidence interval of 95% for the classification of single fingerprints

Class	Mean	Lower bound	Upper bound
A	63.50%	61.02%	65.98%
L	90.95%	90.11%	91.78%
R	90.19%	89.50%	90.88%
W	86.73%	85.69%	87.78%

Table 8: Single-finger binning and retrieval results

Hand	Finger	Visited bins	Visited subjects	% of naïve	Best possible
Right	Thumb	1.13	725.3	53.7%	45.9%
	Index	1.19	524.6	38.8%	30.7%
	Middle	1.14	821.7	60.8%	53.6%
	Ring	1.12	731.1	54.1%	46.5%
	Pinky	1.11	995.3	73.7%	66.6%
Left	Thumb	1.12	728.0	53.9%	44.7%
	Index	1.22	544.6	40.3%	29.1%
	Middle	1.15	806.7	59.7%	50.9%
	Ring	1.16	792.7	58.7%	48.6%
	Pinky	1.14	1,102.1	81.6%	72.4%

4.3. Results

The performance of the employed method for the single fingerprint classification task is summarised in table 7.

[1] Parameters of the DNN-based classifier are summarised according to the guidelines provided by the IEEE Signal Processing Society.

Compared to the current state-of-the-art (*c.f.* table 1), the applied fingerprint classification scheme achieves a moderate accuracy. Particularly, a significantly lower CCR can be observed for the arch class, which results from natural (unbalanced) fingerprint class distribution in the training data. The resulting workload reduction obtained in a single-finger binning and retrieval strategy is listed in table 8 (average values are given for "visited" bins and "visited subjects").

With respect to the estimation of the maximum computational workload reduction (denoted "best possible"), a perfect fingerprint recognition system is assumed. This means, the retrieval is considered successful, when the fingerprints of the correct identity are reached in a closed set identification. This assumption is reasonable considering the accuracy reported for multi-fingerprint recognition systems [20]. Obtained workload reduction (denoted "% of naïve") for the proposed multi-finger binning and retrieval strategy for different number of contiguous fingers used (as described in subsection 2.2), and combinations thereof is summarised in table 9. The computational workload reduction is estimated by comparing the proposed scheme to a naïve system performing an exhaustive one-to-many search. The results for multi-fingerprint binning represent a significant improvement over a conventional single-finger binning, *c.f.* table 8. Additionally, the computational workload could be further reduced, by employing a more accurate classifier

Table 9: Multi-finger binning and retrieval results

Nr. fingers	Hand	First finger	Visited bins	Visited subjects	% of naïve	Best possible
2	Right	Thumb	1.47	349.4	25.9%	17.9%
		Index	1.49	384.9	28.5%	20.2%
		Middle	1.38	499.5	37.0%	28.3%
		Ring	1.33	584.3	43.3%	33.8%
	Left	Thumb	1.50	345.0	25.5%	15.4%
		Index	1.56	392.4	29.1%	18.3%
		Middle	1.46	545.2	40.4%	29.4%
		Ring	1.43	696.7	51.6%	39.1%
3	Right	Thumb	2.12	261.9	19.4%	12.1%
		Index	2.14	266.0	19.7%	12.4%
		Middle	1.84	419.7	31.1%	22.0%
	Left	Thumb	2.17	254.2	18.8%	10.0%
		Index	2.35	289.6	21.4%	11.6%
		Middle	2.09	493.4	36.5%	24.4%
4	Right	Thumb	3.81	199.1	14.7%	8.5%
		Index	3.69	227.1	16.8%	9.5%
	Left	Thumb	4.08	197.7	14.6%	7.0%
		index	4.21	261.5	19.4%	9.3%
5	Right	Thumb	8.35	171.6	12.7%	6.6%
	Left	Thumb	9.41	182.8	13.5%	5.8%

5. Conclusion

With the consistently growing size of deployed biometric databases, the need to reduce the computational requirements of the biometric identification scenario is clear. One of the popularly employed methods is arranging the enrolled database into bins based on the samples' tangible features (such as fingerprint classes). By doing so, during an identification transaction, only a small fraction of bins (and thereby biometric references) needs to be visited during the retrieval step. In this paper, the idea of multi-instance fingerprint binning is explored. Best to the author's knowledge, this is a neglected topic in the scientific literature, with only a single theoretical analysis done in the past.

In the proposed system, the classifier outputs for multiple fingers of one hand are combined and adjusted with statistical information about the occurrences of fingerprint classes and correlations among them obtained from two large databases (NIST SD9 and an in-house dataset of the German Federal Police). Subsequently, a variable search order strategy is applied to conduct a one-to-first search of the enrolled database. The experiments were conducted using the publicly available and well-established PCASYS tool and a neuronal network classifier. The results convince by significant reduction of the computational workload: for instance, when using all five fingers of a hand, it is reduced to less than 15% of the naïve exhaustive search. Additionally, the theoretical limits of the approach are established for a perfect (always accurate) classifier, with which the computational workload could be brought down to approximately 5% of the naïve exhaustive search.

Furthermore, several interesting observations can be made regarding the choice of the binning parameters:

- One or few fingers: Few bins (each containing many subjects) have to be visited, but more subjects have to be considered, *i.e.* the overall search space is larger.
- Many fingers: Many bins (each containing few subjects) have to be visited, but fewer subjects have to be considered, *i.e.* the overall search space is smaller.'
- Choice of fingers: the thumb appears to exhibit less correlation to other fingers, which makes its inclusion in the binning scheme beneficial, *c.f.* table 9, where the lowest workload is always achieved by including the thumb.

By using data from both hands (*i.e.* ten fingers instead of five), the computational workload could presumably be further reduced, but this scenario would be less practical for operational deployments, which typically perform acquisition for a single hand or individual fingers only. Since the proposed approach utilises well-known and understood features, and is readily compatible with the current fingerprint sample acquisition methods, it could be seamlessly integrated into operational biometric deployments.

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).

References

[1] M. Abadi et al. TensorFlow: Large-scale machine learning on heterogeneous systems. https://www.tensorflow.org/, 2015.

[2] S. Borra, G. Reddy, and E. Reddy. Classification of fingerprint images with the aid of morphological operation and AGNN classifier. *Applied Computing and Informatics*, 2017.

[3] G. Candela, P. Grother, C. Watson, R. Wilkinson, and C. Wilson. PCASYS – A Pattern-Level Classification Automation System for Fingerprints, NIST Interagency/Internal Report (NISTIR) - 5647. Technical report, National Institute of Standards and Technology, 1995.

[4] K. Cao, L. Pang, J. Liang, and J. Tian. Fingerprint classification by a hierarchical classifier. *Pattern Recognition*, 46(12):3186–3197, 2013.

[5] R. Cappelli, D. Maio, and D. Maltoni. SFinGe: an approach to synthetic fingerprint generation. In *Intl. Workshop on Biometric Technologies*, 2004.

[6] F. Chollet et al. Keras: The Python deep learning library. `https://keras.io`, 2015.

[7] S. Chua, E. Wong, and A. Tan. A fuzzy rule-based fingerprint image classification. *Intl. J. of Applied Engineering Research*, 11:7920–7925, 2016.

[8] K. Dorasamy, L. Webb, J. Tapamo, and N. Khanyile. Fingerprint classification using a simplified rule-set based on directional patterns and singularity features. In *Intl. Conf. on Biometrics*, pages 400–407, 2015.

[9] M. Galar, J. Derrac, D. Peralta, I. Triguero, et al. A survey of fingerprint classification part I: Taxonomies on feature extraction methods and learning models. *Knowledge-Based Systems*, 81:76–97, 2015.

[10] M. Galar, J. Derrac, D. Peralta, I. Triguero, et al. A survey of fingerprint classification part II: Experimental analysis and ensemble proposal. *Knowledge-Based Systems*, 81:98–116, 2015.

[11] J.-M. Guo, Y.-F. Liu, J.-Y. Chang, and J.-D. Lee. Fingerprint classification based on decision tree from singular points and orientation field. *Expert Systems with Applications*, 41(2):752–764, 2014.

[12] E. Henry. *Classification and uses of finger prints*. HM Stationery Office, 1900.

[13] IDEMIA. MorphoWave Desktop: On the fly biometric acquisition. `https://www.morpho.com/en/fly-biometric-acquisition`.

[14] H.-W. Jung and J.-H. Lee. Noisy and incomplete fingerprint classification using local ridge distribution models. *Pattern Recognition*, 48(2):473–484, 2015.

[15] J. Luo, D. Song, C. Xiu, S. Geng, and T. Dong. Fingerprint classification combining curvelet transform and gray-level cooccurrence matrix. *Mathematical Problems in Engineering*, pages 1–15, 2014.

[16] D. Maltoni, D. Maio, A. Jain, and S. Prabhakar. *Handbook of Fingerprint Recognition*. Springer, 2009.

[17] NIST. NIST Special Database 9. `https://www.nist.gov/srd/nist-special-database-9`, 2010.

[18] F. Pedregosa et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[19] D. Peralta, I. Triguero, S. Garca, Y. Saeys, et al. Distributed incremental fingerprint identification with reduced database penetration rate using a hierarchical classification based on feature fusion and selection. *Knowledge-Based Systems*, 126(Supplement C):91–103, 2017.

[20] A. Ross, K. Nandakumar, and A. Jain. *Handbook of Multibiometrics*. Springer, 1st edition, 2011.

[21] Unique Identification Authority of India. Aadhaar dashboard: https://uidai.gov.in/aadhaar. `https://uidai.gov.in/aadhaar`.

[22] R. Wang, C. Han, and T. Guo. A novel fingerprint classification method based on deep learning. In *Intl. Conf. on Pattern Recognition*, pages 931–936, 2016.

[23] J. Wayman. *Multifinger Penetration Rate and ROC Variability for Automatic Fingerprint Identification Systems*, pages 305–316. Springer, 2004.

[24] L. Webb and M. Mathekga. Towards a complete rule-based classification approach for flat fingerprints. In *Intl. Symp. on Computing and Networking*, pages 549–555, 2014.