



**HAL**  
open science

# Comparative Study of Fingerprint Database Indexing Methods

Joannes Falade, Sandra Cremer, Christophe Rosenberger

► **To cite this version:**

Joannes Falade, Sandra Cremer, Christophe Rosenberger. Comparative Study of Fingerprint Database Indexing Methods. CyberWorlds conference, Oct 2019, Kyoto, Japan. hal-02387781

**HAL Id: hal-02387781**

**<https://hal.science/hal-02387781>**

Submitted on 30 Nov 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# Comparative Study of Fingerprint Database Indexing Methods

Joannes Falade, Sandra Cremer, Christophe Rosenberger

► **To cite this version:**

Joannes Falade, Sandra Cremer, Christophe Rosenberger. Comparative Study of Fingerprint Database Indexing Methods. CyberWorlds conference, Oct 2019, Kyoto, France. hal-02387781

**HAL Id: hal-02387781**

**<https://hal.archives-ouvertes.fr/hal-02387781>**

Submitted on 30 Nov 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Comparative Study of Fingerprint Database Indexing Methods

Joannes Falade, Sandra Cremer, Christophe Rosenberger

**Abstract**—Nowadays, there are large country-sized fingerprint databases for identification purposes, for border access controls and also for Visa issuance procedures around the world. Fingerprint indexing techniques aim to speed up the research process in automatic fingerprint identification systems. Therefore, several preselection, classification and indexing techniques have been proposed in the literature. However, the proposed systems have been evaluated with different experimental protocols, that makes it difficult to assess their performances. The main objective of this paper is to provide a comparative study of fingerprint indexing methods using a common experimental protocol. Four fingerprint indexing methods, using naive, cascade, matcher and Minutiae Cylinder Code (MCC) approaches are evaluated on FVC databases from the Fingerprint Verification Competition (FVC) using the Cumulative Matches Curve (CMC) and for the first time using also the computing time required. Our study shows that MCC gives the best compromise between identification accuracy and computation time.

**Keywords**-Biometrics, matching, indexing, fingerprint, minutiae, quality metrics, NFIQ2, Locality-sensitive Hashing.

## I. INTRODUCTION

Fingerprint biometric systems use fingerprint characteristics (unique, permanent and easy-to-acquire properties) for people recognition. Fingerprint databases are widely used for border control and administrative procedures. For example, Aadhaar is the largest biometric database, with more than 1 billion persons, built for indian population [16].

Two tasks are mainly achieved by biometric systems: verification and identification. Verification task, also referred as 1-1 matching, consists on deciding whether a person, who claims an identity, is or is not the owner of that identity. Verification systems are evaluated with two accuracy metrics: False Acceptance Rate (FAR) and False Reject Rate (FRR). The FAR is inversely related to FRR. Identification task, also referred as 1-N matching aims to find a target person in a biometric database. Identification task requires a long computation time to carry out the comparisons against all database entries. Moreover, performing all comparisons increases the identification error rates [9].

Therefore, classification, clustering and indexing strategies are prominent to achieve the identification task in limited time. The first objective of indexing method is to split and filter a large database, to reduce the list of candidates. This is measured by the Penetration Rate ( $PR$ ) computed as  $PR = \frac{\#candidates*100}{database\ Size}$ . The second objective is to ensure that the list of candidates

contains the target. This is measured by the Hit Rate ( $HR$ ) computed as  $HR = \frac{\#Well\ Identified\ Candidates*100}{\#Total\ Identified\ Candidates\ Test}$ . The objective for an indexing method applied to fingerprint identification is to tend  $PR$  to 0 and  $HR$  to 100.

The main challenge in fingerprint identification is to achieve both high speed and accuracy. The recently proposed algorithms used different experimental protocols and have been evaluated on different fingerprint databases [18], [5]. Therefore, it is not simple to objectively compare the performance of the proposed algorithms. In this paper, we provide a comparative study of various state-of-the-art indexing algorithms. To our knowledge, such a comparative study was never proposed in previous works. This is our main contribution to an important topic in biometrics.

The rest of this paper is organized as follows. In section II, various indexing algorithms applied to fingerprint identification are presented. Section III describes the four selected methods for our comparative study. In section IV, a common experimental protocol is proposed and evaluation performances are exposed. Conclusions and perspectives are given in Section V.

## II. RELATED WORKS

There are two main categories for fingerprint features: global and local features. For global features, fingerprint representation uses the core point, ridges shapes, ridges number and singular points such as delta and loop [13]. On the other hand, local representation uses fingerprint minutiae points and pores to build a solid geometric neighborhood features (Triplets, quadruplets, circle or cylinder around minutiae) [9].

Generally, the problem of database searching is addressed using three strategies: classification, clustering and indexing. Classification is a supervised learning approach while clustering uses an unsupervised learning technique. Indexing is the combination of classification and clustering by assigning an index number to each database entry. Figure 1 shows the identification process using a classification/indexing strategy. The first classification method was proposed by Henry [6] in 1900 based on global features extraction. He proposed partitioning fingerprint databases into five classes: right loop, left loop, whorl, arch and tented arch. In [8], the authors proposed a five-class based classification. In order to improve the accuracy rate, they used the core point and built 48 areas of interest around this point. In [17], [1], [10],

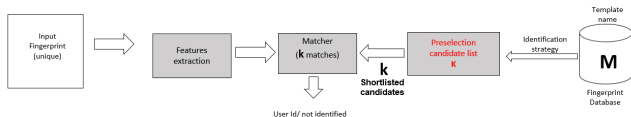


Figure 1. Fingerprint identification process of an input fingerprint with a large database of size  $M$ . Identification strategies such as classification or indexing are used to retrieve a shortlist of  $k$  candidates for next verification step (on Matcher box). Note's,  $k \ll M$

authors proposed to increase the number of classes (from five to nine) to achieve better performances. However, exclusive classification into five to nine fingerprint classes is limited by the small number of classes [6]. Thus, fingerprint indexing methods are used in recent works to improve the PR and HR.

Although many approaches used indexing strategy, they focused on fingerprint features representation as a discriminatory factor. In [15], a valuable review of fingerprint indexing methods was presented. In [7], an original indexing fingerprint method based on minutiae quadruplets was described. In [12] and [11] an expanded Delaunay triangulation (triplets) and minutiae pairs was proposed to develop a fingerprint indexing method with good performances. A new triplets scheme for fingerprint indexing was detailed in [18] while in [2], the authors proposed a comprehensive approach to reduce the list of possible candidates index. Note that indexing methods are generally based on the local features because they seem to be the most discriminative features.

### III. SELECTED METHODS

In this part, we describe the selected methods for the comparative study. These methods are based on basic, cascade, matcher and MCC approaches. We choose these methods for their simplicity, good performances and the fact that our fingerprint representation is appropriate for their implementation. These methods are evaluated in terms of  $HR$  and  $PR$  along with the computing time.

Let's  $D$  be a database of size  $M$  and  $m_i$  is a matcher.  $m_i$  is characterized by  $FAR$ ,  $FRR$  and a computation time for matching  $t_i$ . We note  $Time$  the time required for a fingerprint searching in  $D$ .

#### A. Basic method

The basic method for the identification task is also known as the naive one. It consists on considering identification as multiple successive verifications. In the database  $D$ , an input fingerprint  $I$  is compared to all  $M$  fingerprints to return a shortlist of best scores. The time required to compute basic is  $Time = M * t_i$ .



Figure 2. Cascade chain for fingerprint identification

#### B. Cascade method

Cascade method is a combination of successive naive methods. Generally, the faster a matching algorithm is, the less precise it is. Therefore, the cascade is forged by the fastest matching algorithm at the front and the most accurate algorithm at the end (see figure 2). The time required to compute cascade is  $Time = \sum_{i=1}^l k_{i-1} * t_i$ .

#### C. Matcher method

The matcher method was introduced by Arun Ross [5] for fingerprint indexing. This method was proposed to build a new input fingerprint index by using other specific fingerprints as references. Fingerprints which produce a high variance of matching scores against other fingers are chosen as references (discriminative property). These  $n$  references are at most 5% of  $M$ . Then, a quantification process used a threshold value to discretize each matching score against the references. The size of the created index is equal to the number of the chosen references. To retrieve the list of candidates for an input fingerprint  $I$ , the matching scores are only computed against references and then, the candidates list is the closest set of fingerprints to  $I$  according to the Hamming distance. This representation depends on the number of references  $n$ , discretization interval  $b$  and maximum Hamming distance, considered as valid,  $d$ . We compute the time as  $Time = n * t_i$

#### D. Mcc method

Minutiae Cylinder Code (MCC) is a binary representation of a fingerprint proposed by Capelli et al. in [3]. Authors described each minutiae of a fingerprint as a discretized binary cylinder which represents its neighborhood. Locality-sensitive Hashing (LSH) shows good performances for high binary dimension reduction. It is performed with MCC to create an indexing method as explained in [4].  $l$  multiple LSH functions are executed to index each minutia of each fingerprint template. The intuition is that the minutiae of similar templates have a good chance of being under the same index value for a reduced  $h$  dimension ( $h \ll n$ ). To retrieve a shortlist for an input query fingerprint, the identifiers of template models that share the maximal indexes with the query are selected. We roughly assume MCC identification time is given as  $Time = l * t_i$ .

### IV. EXPERIMENTAL PROTOCOL

In this section, fingerprint databases are presented. Then parameters of the selected methods are fixed. Finally, exper-

imental results in terms of CMC and computing time are exposed.

### A. Fingerprint Databases

To evaluate the four selected methods, we show the result on Fingerprint Verification Competition (FVC) databases FVC 2002DB1. This database contains 100 individuals and 8 samples per person. We also combined eleven FVC databases (2000DB1 to 2006DB4) to build challenging large database. This database is ten times larger than FVC2000DB1. We use the best sample, provided by NFIQ2 metric [14], for enrolment and the rest as queries.

### B. Parameters Setups

Table I shows the performance of the different matching algorithms in terms of Equal Error Rate (EER) and computing time. EER is obtained with the threshold where FAR=FRR. The algorithms are commercial.<sup>1</sup> Basic method is applied with *algo4*. For cascade approach, *algo1*, *algo2* and *alog4* are executed in this order (see Figure 2). Matcher method is performed with *algo3* using parameters such as:  $b = 2$ ,  $n = 5$  and  $d = 3$ . As previously explained, MCC method does not require a matching algorithm and its parameters are given such as:  $n = 1532$ ,  $h = 32$ ,  $l = 25$ ,  $min = 5$ . Note that if the number of bit '1' in the minutiae hash is less than *min*, the minutiae is rejected.

Matcher	<i>algo1</i>	<i>algo2</i>	<i>algo3</i>	<i>algo4</i>
EER	0.095	0.078	0.04	$10^{-2}$
$t_i$	$10^{-1}$	$1.1 * 10^{-1}$	$2 * 10^{-1}$	$6.7 * 10^{-1}$

Table I  
PERFORMANCE EVALUATION OF MATCHING ALGORITHMS IN TERMS OF EER AND COMPUTING TIME  $t_i$  IN MILLISECOND ON EACH SIMPLE FVC DATABASE.

### C. Experimental results

The main contribution of the proposed study is to evaluate different indexing methods, with the same databases and ressources. The CMC (Cumulative Matches Curve), which combines *PR* and *HR*, is computed. We also add the computing time for each indexing method performance metrics for indexing methods. These conditions allow us to provide a solid and common environment benchmark for the different pre selection methods.

Figure 3 shows the CMC of basic, cascade, matcher and MCC methods. Furthermore, the computation time for each method is respectively 67, 21, 1 and 5 milliseconds for this database of 100 individuals. According to this CMC,

<sup>1</sup>Due to confidence issues, we are not allowed to communicate the name of commercial systems.

the basic method seems to perform as the best. This is explained by two reasons. Firstly, the fingerprint query is compared to all database entries. Secondly, the best matching 1-1 algorithm (*algo4*) is used. However, the basic method gives the longest computing time. Therefore, this method is not appropriate for large database (as Aadhaar one as for example).

The cascade method shows similar performances as the basic method. However, for the figure 3, the CMC shows worse performances. As previously mentioned, the front matching algorithm for the cascade approach is *algo1*. The errors made by this algorithm cannot be recovered by the following matching algorithms (*algo2* and *algo4*), which explains the difference with the basic approach results.

The matcher method presents the worst performances. For example, to obtain a *HR* greater than 75%, the *PR* need to exceed 50%. On the other hand, for this method the computing time is the shortest because the indexes are computed from few matching references (just five matching 1-1 in our case and only 1millisecond required).

Figure 4 shows the CMC of basic, cascade, matcher and MCC methods, computed on the combined FVC databases. This explains the accuracy results under real conditions of challenging and large database. The MCC method seems to be the best identification strategy. Although the database is more complex, It achieves the best performance. The CMC curve is as good as the CMC of a simple FVC database which contains just 100 individuals (Green curves are the same for figures 3 and 4). The accuracy of basic and cascade methods has dropped sharply and show that these methods are not adapted to large database.

So, the MCC approach presents the best compromise between performances and computing time. For a challenging and very short database, such as FVC2002DB1 (fig.3), the MCC obtains similar result as basic approach and outperforms cascade and matcher approaches. We notice that for a  $PR = 10\%$ , the  $HR > 95\%$ . Moreover, for a large databases (more than 1000 individuals), MCC keeps the same CMC accuracy shape. This confirms that the MCC indexing strategy is independant from database complexity.

This study allows us to point out the following conclusions: The basic and cascade are not suited for very large databases more than 1000 individuals. These approaches are totally rely on the performances of the 1 – 1 matching algorithm and are only appropriate for small-sized database (at most hundred of people). The Matcher method is appropriate for real-time application but depends on the performances of the discretization technique. The MCC technique is resistant to the quality of the fngerprints and gives the best compromise between accuracy and computing time.

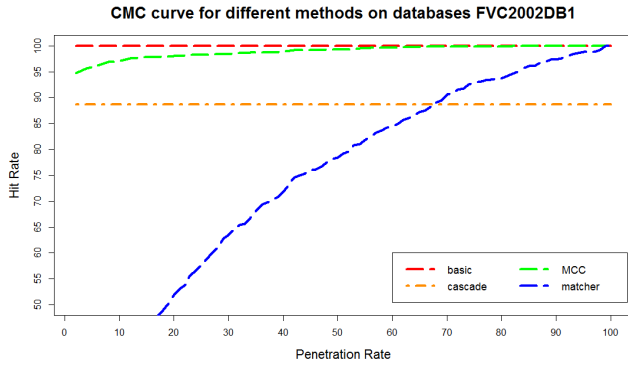


Figure 3. CMC identification performance on database FVC2002DB1

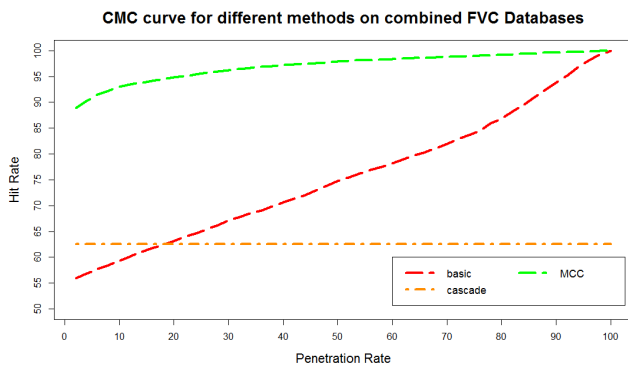


Figure 4. CMC identification performance on combined FVC databases

## V. CONCLUSION AND PERSPECTIVES

This paper provides a comparative study of four fingerprint indexing methods. Basic, cascade, matcher and MCC approaches are selected to conduct this study. The main contribution of this paper is to compare the indexing methods under a common experimental protocol using FVC databases and also proposed a challenging large database for assessing indexing methods. Our study shows that MCC gives the best results in terms of CMC and computing time. In addition, we noticed that indexing techniques based on cascade approach depends on the 1-1 matching algorithms and the quality of fingerprint databases. Future works will be dedicated to extend this paper and details about each implemented method against more databases.

## REFERENCES

- [1] Ebtesam Najim Abdullah AlShemmary. *Classification of Fingerprint Images Using Neural Networks Technique*. Journal of Engineering (JOE) Vol 1, 2012.
- [2] Raffaele Cappelli, Matteo Ferrara, and Dario Maio. *Candidate List Reduction Based on the analysis of Fingerprint Indexing Scores*. IEEE Transactions on information Forensics and Security, 2011.
- [3] Raffaele Cappelli, Matteo Ferrara, and Davide Maltoni. *BMinutia Cylinder-Code: A New Representation and Matching Technique for Fingerprint Recognition*. IEEE transaction on Pattern Analysis and Machine Intelligence, vol 32, No12, 2010.
- [4] Raffaele Cappelli, Matteo Ferrara, and Davide Maltoni. *Short papers: Fingerprint Indexing Based on Minutia Cylinder-Code*. IEEE transaction on Pattern Analysis and Machine Intelligence, vol 33, No 5, 2011.
- [5] Aglika Gyaourova and Arun Ross. *A Novel Coding Scheme for Indexing Fingerprint Patterns*. Proceedings of S+SSPR Workshop, LNCS 5342, pp 765-774, (Orlando, USA), 2008.
- [6] A.R. Henry. *Classification and uses of finger prints*. London, Routledge, 1900.
- [7] Ogechukwu Iloanusi, Aglika Gyaourova, and Arun Ross. *Indexing fingerprints using minutiae quadruplets*. 2014.
- [8] Anil Jain, Salil Probhakar, and Lin Hong. *A multichannel approach to fingerprint classification*. 1999.
- [9] Asker M. Bazen Johan de Boer and Sabih H. Gerez. *Indexing Fingerprint Databases Based on Multiple Features*. ProRISC Workshop on Circuits, Signal, Processing, Veldhoven, The Netherlands, 2001.
- [10] Kalle Karu and Anil.K. Jain. *Fingerprint Classification*. Pattern Recognition Vol 29, N3, pp: 389-404, 1996.
- [11] Javad Khodadoust and Ali Mohammad Khodadoust. *Fingerprint indexing based on expanded Delaunay triangulation*. Expert Systems with Applications, 2017.
- [12] Javad Khodadoust and Ali Mohammad Khodadoust. *Fingerprint indexing based on minutiae pairs and convex core point*. Pattern Recognition, 2017.
- [13] Davide Maltoni, Dario Maio, Anil K. Jain, and Salil Prabhakar. *Handbook of Fingerprint Recognition*. Springer, 2003.
- [14] NIST. *NIST Fingerprint Image Quality*. NIST Report, 2016.
- [15] A. Pooja Parmar and D.Sheshgang Degadwala. *Fingerprint Indexing Approches for Biometric Database: A Review*. International Journal of Computer Applications (0975-8887) Volume 130 - N13, 2015.
- [16] Unique Identification Authority of India UIDAI. *Role of biometric technology in Aadhaar authentication*. Planning Commission, Govt. of India, 2012.
- [17] Qinzhi Zhang, Kai Huang, and Hong Yan. *Fingerprint classification based on extraction and analysis of singularities and pseudoridges*. 2006.
- [18] Wei Zhou, Jiankun Hu, Song Wang, Ian Petersen, and Mohammed Bennamoun. *Fingerprint indexing based on combination of novel minutiae triplet features*. 2014.