Storage Approaches for Biometric Database: A Review

Ashish Dabas\textsuperscript{1}, Ms. Shalini Bhadola\textsuperscript{2}, Ms. Kirti Bhatia\textsuperscript{3}

\textsuperscript{1}M.Tech Student, Computer science & Engineering, Sat Kabir Institute of technology and Management Bahadurgarh (HR) Affiliated by Maharshi Dayanand University(Rohtak), Haryana, India
\textsuperscript{2,3}Assistant Professor, Computer science & Engineering, Sat Kabir Institute of technology and Management Bahadurgarh (HR) Affiliated by Maharshi Dayanand University(Rohtak), Haryana, India

Abstract- Due to the uniqueness and persistent properties of the biometric fingerprint characteristics, large scale in border control and governmental applications such as the Aadhaar project in India, the Visa Information System (VIS) in Europe and US-VISIT / IDENT system in the USA are based on fingerprint recognition and generally contain millions of fingerprint samples. The study of fingerprint indexing techniques is inevitable, to improve the effectiveness in searching for satisfactory candidate reference list in such large biometric databases. In systems using biometric identification, the identity associated with the input data is decided by its comparison against every single entry in the database. This matching process is exhaustive which leads to increase in the rate of erroneous identification and the response time of the system. This paper presents a survey on the fingerprint indexing methods that are currently available and some of them are presented. Fingerprint indexing is based on the local ridge line orientation, global feature, minutiae and other features.

Index Terms- Biometrics, Indexing, Fingerprint, minutiae

1. INTRODUCTION

Biometrics are automated methods of recognizing an individual based on their physiological (e.g., fingerprints, face, retina, iris) or behavioral characteristics (e.g., gait, signature). Due to the properties regarding uniqueness and permanence of fingerprints, the use of fingerprint identification is an essential component in governmental applications [1]. It is the ultimate form of electronic verification of physical attribute of a person to make a positive identification [2]. Biometric identification is the process of associating an identity to the input biometric data by comparing it against the enrolled identities in a database [3]. In many systems, this comparison is undertaken in a time consuming manner, i.e., the input data of an individual is compared against each and every enrolled data in order to decide the identity of the individual. The false positive identification rate and the matching time are directly proportional with the size of database. These biometric systems generally contain millions or even billions of fingerprints samples. The biometric database may be very large and hence in such case the identification process will take a long response time to identify the individual. However, the identification process can be accelerated by reduction of the search space. There are two reasons for necessity of a fingerprint indexing scheme for a large scale biometric database: (i) slow process involved in identifying a correct person, as all the entries in a database have to be searched and compared; (ii) chance of increase in ‘false accept error’ as the database grows in size.

A database can be searched on three distinct schemes: classification, clustering and indexing. There are two types of machine learning namely supervised and unsupervised learning. Classification requires supervised learning where labels or classes are learned during the process and the input set is categorized on to these labels. On the contrary in unsupervised learning, prior knowledge about labels is not required. Clustering is an unsupervised learning. In clustering dataset is partitioned into consistent groups called clusters. Fingerprint is the most widely used biometric ‘trait’ in large scale biometric database systems, for identification purposes. Obviously, it is important to improve the efficiency while seeking for suitable references in such large-scale database, ensuring at the same time high recognition accuracy. In order to achieve this goal, fingerprint indexing techniques, which reduce the response time by selecting a number of candidates from the reference database for further
comparison, have been studied during the last two decades. A detailed survey is carried out on the various fingerprint indexing techniques.

The rest of this paper is organized as follows. Section 2 describes Fingerprint Indexing and its types based on Local features (Section 2.1), Global features (Section 2.2) and some other features (Section 2.3). Finally, Section 3 draws the conclusion.

2. FINGERPRINT INDEXING

In order to reduce the search time for finding a matching identity to a query sample, fingerprint indexing techniques are discovered. Fingerprint Indexing methods associate a probe or a query sample with index codes or feature vectors that best describe the features of the fingerprint. Indexing techniques aim to select a few fingerprints in the database, called the candidate list, and the probe is compared with the fingerprints in the candidate list only by a matching algorithm. The probe should be contained by the candidate list. In a 10-print identification system, by limiting all query samples searches to multiple candidate lists of fingerprints that include the genuine prints to the probes the search time for a match would be reduced. If the indexing technique is efficient the size of a candidate list should be remarkably less than the total database size.

The two important measures used for measuring the performance of fingerprint indexing are Penetration rate and Hit rate. The performance of a fingerprint indexing technique is determined by the penetration rates and hit rates of all the query samples used in the process of evaluation. The penetration rate is the portion of the candidate identities retrieved from the database upon presentation of the query sample and the probability that the genuine fingerprint is retrieved is called the hit rate.

For the efficient performance of the fingerprint indexing scheme the penetration rate should be low. Therefore, the aim is to have penetration rates as low as possible and a 100% hit rate. A hit occurs when the candidate list contains the corresponding match of the probe sample. Fingerprint indexing methods currently available can be classified into three categories based on the features used in their approaches:

- Local feature based: Features extracted from the ridge-line orientation, fingerprint’s minutiae using Delaunay triangulation, minutiae location and direction information, ridge density, local ridge-line orientation, etc. are known as local features.
- Global feature based: Features extracted from the whole orientation field, ridgeline frequency and singular points have been used to construct feature vectors are called global features.
- Other features based: Features extracted from the symmetric filters, scale invariant feature transformation (SIFT) also has been used to develop a fingerprint indexing.

![Fig 1: Fingerprint Indexing including the enrollment and retrieval processes.](image)

3. FINGERPRINT INDEXING BASED ON LOCAL FEATURES

In [4] J. De Boer et al. have proposed an approach using three fingerprint features; Finger Code, minutiae triplets and the registered directional field estimate.

![Fig 2: Fingerprint with directional field, singular points and minutiae.](image)

It has been proved that indexing methods that are based on these features search the biometric database more productively than done by a simple linear scan. The directional field (DF) is defined as the local orientation of the ridge-valley structures. The singular points (SPs) are the discontinuities in the directional field [5]. The details of the ridge-valley structures are provided by the minutiae. Next, by combination of these three features a new indexing...
scheme has been proposed. It has been shown that the new approach results in a better performance than the approaches that are based on the individual features, thereby allowing a wider range of fingerprint databases to be searched. Compared to a simple linear search, this allows the size of databases to be 100 times as large, while maintaining the same FAR (False Acceptance Rate) and FRR (False Rejection Rate).

In [6] Bir Bhanu and Xuejun Tan present a model-based approach, which retrieves correct candidate list using features of triangles formed by the minutiae triplets as the basic representation unit. Angles, type, direction, handedness, and maximum side triangle are the features used. Moreover used to eliminate false correspondences, geometric features depending on other attributes of minutiae are used. It has been shown show this indexing method effectively reduces down the number of candidate list in the presence of scale, shear, translation, rotation, occlusion, and clutter by experimental results on NIST special database 4 (NIST-4) and live-scan fingerprint images of different qualities. In various comparisons with the prominent indexing method proposed by Germain et al., this approach has performed better.

In [7] X. Liang et al. concentrate on a more accurate fingerprint indexing method that retrieves the top N possible matching candidate list from a large biometric database. Based on minutia neighborhood and a more secure triangulation algorithm (low-order Delaunay triangles, consisting of order 0 and 1 Delaunay triangles), are used which are both insensitive to fingerprint distortion. The features used for indexing include minutiae detail and attributes of low-order Delaunay triangle (its angles, handedness, maximum edge, and related angles between edges and orientation field). The above approach has proved to be better in searching large databases stably than previous approaches for two reasons: (i) Minutia detail provides more minutiae classes than minutia types, and further reduces search space. Since they can be obtained during the minutiae extraction step, the proposed indexing algorithm does not require additional computing time. (ii) Low order Delaunay triangles include Delaunay and order 1-Delaunay triangles. These small number of additional order 1-Delaunay triangles increase the quantity of genuine matched triangles, especially on distorted fingerprints. Thus, it is more insensitive to elastic distortion. Since the proposed triangulation algorithm creates triangles compared to triangles for an all-triplets-based algorithm, many redundant or incorrectly matched triangles are avoided. Simultaneously, the search space in a database is reduced when using the same. The proposed fingerprint indexing approach has better abilities to tolerate distortion and reduce the possible matching fingerprints for the next step, which is shown by experiments on database FVC2002 and FVC2004.

In [8] Ogechukwu Iloanusi et al., propose a new structure named “minutiae quadruplet” for the purpose of indexing fingerprints is used in with a clustering method to filter a fingerprint database. A set of 4 minutiae points forms a quadrilateral as shown in Figure 3 is called a minutiae quadruplet.

![Fingerprint image with minutiae quadruplets.](image)

Fig 3: Fingerprint image with minutiae quadruplets.

Minutiae quadruplets allow the use of features that are less sensitive to deformation as compared to minutiae triplets. The above indexing algorithm is employed on all database in the Fingerprint Verification Competition (FVC) 2000, 2002 and 2004 databases. The high hit rates are obtained at low penetration rates which prove that the proposed method is beneficial for fingerprint indexing. Moreover, it shows that for 50% of the fingerprint images, the penetration rate was smaller than 5.5% at a 100% hit rate, in most of the datasets. The robust performance against various databases suggests that the fingerprint indexing scheme can be employed for use in large-scale biometric databases. Experiments on fingerprints with missing minutiae and fingerprints with spurious minutiae points and show that the method is remarkably robust. The retrieval strategy is computationally cheaper and the proposed algorithm has fewer requirements for storage.

In [9] Cappelli et al. have proposed a new hash-based indexing method to accelerate fingerprint identification process in large biometric databases. A fixed length bit vector is formed by encoding the neighbor of each minutia point using minutiae cylinder code (MCC). Using Locality- Sensitive
Hashing (LSH) the bit vector is indexed. Each minutiae (m) is a triplet: m = {xm, ym, qm} where xm & ym are minutia location and qm is the minutia direction in the range %0, 2p%. A binary vector set vi was created using MCC from the given template and for each vector vj the hashing function is applied. During the creation of the index any binary vector vj of each template ti is given as input to all the hash function and the pair (i,j) identifies ti. The similarity between two templates is counted by the number of collision of each pair of binary vectors. The binary vector obtained with MCC has only a few 1 bits then it is more likely to contain a larger no of pairs[9].

Relying on Minutiae Cylinder-Code (MCC), a Locality-Sensitive Hashing (LSH) method has been discovered, which proved to be very beneficial in mapping a minutiae-based representation into a set of fixed-length transformation-invariant binary vectors. A new search method has been designed thanks to the derivation of a numerical approximation for the similarity between MCC vectors. Experiments have been carried out to compare the proposed algorithm across 15 existing methods over all the benchmarks typically used for fingerprint indexing. The durability of the proposed method is confirmed by the fact that exactly the same parameter values (for both LSH and MCC) have been used against a set of heterogeneous data sets, including offline-scanned, rolled, flat and live-scan fingerprints. When the indexed fingerprints grow from 2,700 to 24,000, the proposed algorithm seems to scale nicely with the increase in database size since it obtains the same accuracy with an almost linear increase in search time.

In [10] Ogechukwu N. Iloanusi, has proposed a multi-finger fingerprint indexing approach using minutiae quadruplets along with a clustering scheme. Four, five and ten fingerprints from an individual are fused at the rank level using the highest rank rule. Rank level fusion is the combination of the ranks output by each classifier’s one-to-many comparisons in a multi-biometric identification system. The minutiae quadruplet features are robust and the clustering method assists in quickly identifying a potential candidate list from the biometric database. It has shown from the result of experiments that the fusion scheme with minutiae quadruplet features gives good performance in fingerprint indexing. Thus, the search time for the candidate list for a query sample is less despite the number of finger types to be fused. The penetration rates for all query samples used in a test set are less than the database size; hence, minutiae quadruplet features using the clustering retrieval strategy have been proved to have better performances in fused fingerprints.

In [11] Wei Zhou et al. have investigated the indexing performance based on some commonly used features of minutiae triplets and proposed to fuse these features with some novel features of minutiae triplets for fingerprint indexing. Experiments have been employed on FVC 2000 DB2a and 2002 DB1a which show that the proposed fingerprint indexing algorithm can perform better for full fingerprint indexing and experimental results on NIST SD 14 show that the performance is improved noticeable after the new features were fused to the feature space. Experimental evaluation on full fingerprint databases FVC 2000 DB2a and 2002 DB1a show that the proposed indexing approach can achieve better performance.

In [1] Guoqiang Li et al., proposed a score-level fusion fingerprint indexing method by combing a new indexing method and minutia cylinder-code indexing method. The new designed indexing approach extracts a 9 dimensional feature vector including 9 components from minutiae details and a triplet which is obtained from minutiae vicinity. K-means learning method has been employed to cluster the training features. A binary string which has fixed length can be created by allotting each features vector from the subject’s samples to its nearest cluster. This binary string is stored to generate the index space and to represent the subject. Experiments have been employed on number of public databases from FVC 2004 and FVC 2006. The result proves that the proposed approach shows improvement among the other approaches.

In [12] Ntethelelo et al. have proposed a fingerprint indexing algorithm based on minutiae points to reduce biometric database search space.

Fig 4: Classes; Top Left: Left Loop, Top Right: Right Loop, Bottom Left: Arch, Bottom Right: Whorl
This work is motivated by the predefined classes as shown in figure 4 (Left Loop, Right Loop, Whorl, Tented Arch, Plain Arch) are not always equally distributed in the search space i.e. some classes are dominant than others. The matching process can be time consuming to find an exact match, in such cases. In the above approach this problem is solved by constructing a rotational, scale and translation (RST) invariant fingerprint feature based on minutiae points. Then RST invariant feature descriptor dimensions are reduced and passed to a spectral clustering scheme which automatically creates 50 classes. Each class from these 50 classes are then represented with a B+- Tree data structure for fast fingerprint indexing. Distances of feature vectors from the center of the cluster are used as keys in each cluster where they belong.

The authors also proposed to search for match in other clusters by employing triangle inequality rule, instead of searching a probe to only a predicted cluster. The proposed fingerprint indexing approach is 81.4443% accurate on the NIST 4 special database. The results show good performance because NIST 4 special database contains a lot of partial fingerprint. In [13] Cappelli, describes a new fingerprint indexing algorithm based on scalar and vector features, obtained from frequencies and ridge-line orientations. A carefully designed ad-hoc score measures and set of features allow the proposed indexing approach to be extremely efficient and effective, as confirmed by the results of experiments. The new method noticeably outperforms competing techniques over the six publicly available data sets. Moreover, it scales to large biometric databases without degrading accuracy; on a standard PC, a search over one million fingerprints takes less than 1 second. Once the features have been extracted, searching for a fingerprint from a gallery requires just 2.3 ms over the NIST DB14 data set.

Based on the critical assessment of reported research based on local features, Table 1 shows the advantages and disadvantages of the state of the art of algorithms for fingerprint indexing.

4. FINGERPRINT INDEXING BASED ON GLOBAL FEATURES

In [14] Liu et al. have propose a continuous fingerprint indexing approach based on direction estimation, location and correlation of fingerprint singular points. By application of a T-shape model to directional field of fingerprint images, location and direction estimation are obtained. To find lateral-axes and further the main-axes, the T-shape model analyzes homocentric sectors around the singular points of the candidate. Then to obtain a correlation-based similarity measure which gives the evidence of searching priority, a distortion tolerant filter of Minimum Average Correlation Energy is used. The experiment is evaluated by retrieval of 400-fingerprint from 10,000 templates and the mean search space is only 3.46% of the whole database. Singular points are robustly identified and consist of fingerprint intrinsic features. A continuous retrieval approach based on information of all detected singular points including not only cores but deltas also which were eliminated in previous approaches.

In [15] Jun Li et al. have an algorithm based on three symmetrical filters of different orientation structures. The three kinds of symmetrical filters are the delta type filter, the core type filter and the parallel type filter, to map the various structures of fingerprints into three different feature spaces. The uniqueness of the fingerprint orientation image is captured more effectively with these three fingerprint orientation structures. An assumption is made in this approach that the reference point should be available when acquiring a fingerprint image if not; the fingerprint might not possess enough information for the indexing approach. The performance of the indexing approach is also affected by noise, normalization i.e. the reference point and its direction are not always correctly obtained for some image and when resolution is not good. It has been shown by conducting experiments on the NIST database 4 that the proposed approach is effective.

In [16] Liu et al. proposed a fingerprint indexing approach based on fingerprint registration with a novel feature called local axial symmetry (LAS). To measure the LAS of a given region, they first cut out a circle in pixel-wise computed orientation field (POF), which is the doubled angle of orientation and is computed using squared gradient averaging method in, centered at \((x, y)\) with a certain radius as the region of interest (ROI). With a series of equally distributed radii which are in the set of candidate symmetry axes of the region, the ROI is divided into \(N\) homocentric sectors. Once the LAS field is
achieved, the location and direction estimation of reference point are obtained in a straightforward way. Then the registered orientation field is used as a feature vector for employing the following indexing approach. A novel scheme of the experiment is introduced and suffice results of experiments are achieved on FVC2000 DB2 that the average search space is only 2.34% of all fingers in the condition of Equal-sized testing set and training set. It shows that the indexing approach can noticeably reduce the burden of the fingerprint authentication system. The size of training set is also analyzed through experimental results of different NSL value. It is proved that the efficiency can be promoted when the training set grows larger.

Fig 5: Example of local axial symmetry: the blue lines in the circles represent their optimized symmetry axes.

In [17] Yi Wang et al have proposed a fingerprint orientation model based on 2D Fourier expansions (FOMFE) in the phase plane. The FOMFE does not need any prior knowledge of singular points (SPs). It is able to describe the overall ridge topology easily, including the Singular Point regions, for noisy fingerprints also. Experiments have been conducted on a public databases and it shows that the proposed FOMFE can remarkably increase the accuracy of fingerprint feature extraction and thus that of fingerprint matching. Furthermore, the FOMFE has a less-computational cost and can work very effectively on large fingerprint databases. The FOMFE provides a detail description for orientation features, which have enabled its profitable use in feature-related applications such as fingerprint indexing. FOMFE model coefficients are exploited to generate the feature vector, whereas most indexing schemes use raw orientation data. Experiments show noticeable results using different fingerprint databases for the proposed fingerprint indexing approach.

In [18] Paulino et al., have proposed an indexing approach, primarily for latent fingerprints, which combines multiple, level 1 and level 2 features to filter out a large portion of the background database while maintaining the accuracy of latent matching. Rolled fingerprints are achieved by rolling a finger from nail to nail to capture the complete ridge details of a finger, while plain or slap fingerprints are obtained by pressing a finger onto a flat surface without movement. Another important type of fingerprint in process of identification is latent fingerprints, which refer to fingerprints lifted from the surfaces of objects at crime scenes, and are regarded as an extremely important source of evidence to identify suspects. Compared to rolled and plain fingerprints that are obtained from cooperative subjects in an attended mode, latent prints are left by individuals in unconstrained scenarios and therefore have poor quality in terms of ridge clarity and complex background noise, and consists only a small part of a finger and large non-linear skin distortion.

Fig 6: Fingerprint acquisition; Left: Rolled print (Michigan State Police database), Middle: plain or slap print (FVC 2002) and Right: latent print (NIST SD27).

Examples of the three types of fingerprints are shown in Figure 6. Different levels: level 1 features such as singular points, orientation field and type, level 2 features such as minutiae and level 3 features such as ridge contours and pores. This approach consists of combination of singular points, orientation field, minutiae and frequency information. Experiments have been carried out on 258 latent fingerprints in NIST SD27 across a large background database (267K)

5. CONCLUSION

This paper gives a concise description of the main indexing approaches currently available. An analysis of the advantages and disadvantages of the analyzed fingerprint indexing approaches, based on the recovery process and the reliability of the extracted features. Finally a comparison of experimental results reported by the authors in the FVC and NIST databases is made. It is important to note that other algorithms have played an important role in the accuracy of fingerprint indexing approaches. Some of these processes are: enhancement of the fingerprints,
features extraction and the location of the centers of the fingerprints. Efforts of triplet based algorithms are focused on the selection of the triplets and the extracted features. Finally, it can be concluded that algorithms that use features based on triplets of minutiae and ridges have the best performances. Future work may be including new approaches that allow the enhancement of the generated triangulations. Hence, better accuracy can be achieved in cases where some minutiae are not detected.

6. ACKNOWLEDGMENT

This work is fully funded by Ashish Dabas under the supervision of Ms. Kirti Bhatia & Ms. Shalini Bhadola (Assistant Professor Computer science & Engg.) Sat Kabir Institute of technology and Management Bahadurgarh (HR)

REFERENCES


