Appeared in Proc. of IEEE Computer Society Workshop on Biometrics at the CVPR Conference, (Colorado Springs, USA), June 2011

Indexing Fingerprints using Minutiae Quadruplets

Ogechukwu Iloanusi
University of Nigeria, Nsukka
oniloanusi@gmail.com

Aglika Gyaourova and Arun Ross
West Virginia University
http://www.csee.wvu.edu/~ross

Abstract

The computational complexity of matching an input fingerprint against every entry in a large-scale fingerprint database can be prohibitive. In fingerprint indexing, a small set of candidate fingerprints is selected from the database and only images in this set are compared against the input probe fingerprint thereby avoiding an exhaustive matching process. In this paper, a new structure named "minutiae quadruplet" is proposed for indexing fingerprints and is used in combination with a clustering technique to filter a fingerprint database. The proposed indexing algorithm is evaluated on all datasets in the Fingerprint Verification Competition (FVC) 2000, 2002 and 2004 databases. The high hit rates achieved at low penetration rates suggest that the proposed algorithm is beneficial for indexing. Indeed, it was observed that for 50% of the fingerprints, in most of the datasets, the penetration rate was less than 5.5% at a 100% hit rate. The robust performance across different databases suggests that the indexing algorithm can be adapted for use in large-scale databases.

1. Introduction

Two problems associated with human identification in large-scale fingerprint databases are (a) a long response time due to high computational complexity, and (b) a potential increase in false matches with increasing database size. Approaches taken to solve these problems include fingerprint classification, sub-classification and indexing [1]. Fingerprint indexing refers to the assignment of a numerical index vector to a fingerprint. The indexing mechanism includes a retrieval strategy that is invoked for selecting candidate fingerprints from the database (or gallery) based on their similarity with the input probe in the index vector space. The selected fingerprints constitute the candidate list of fingerprints for the given probe. By limiting the matching operation to those fingerprints in the candidate list and eliminating the need for exhaustive matching of the probe with every fingerprint in the gallery, indexing methods are able to reduce the response time in large-scale fingerprint databases.

In the literature, techniques employed for fingerprint indexing are based on ridge features, ridge pattern or structures derived from ridge features [3, 4], minutiae features [5, 6, 7], a combination of features [2, 8] or matching scores [9].

One of the earliest techniques for indexing fingerprints based on minutiae triplets [5] uses features such as sides of the triangle and minutiae orientation that are highly sensitive to fingerprint distortion. This feature set was improved in [6] but with the additional cost of introducing multiple thresholds that had to be trained on the images in the gallery.

In this work, a new topological structure based on minutiae quadruplets is proposed for indexing fingerprints. A minutia quadruplet is a quadrilateral formed from a set of 4 minutiae points as shown in Figure 1. Compared to minutiae triplets, minutiae quadruplets allow for the use of features that are less sensitive to distortion.

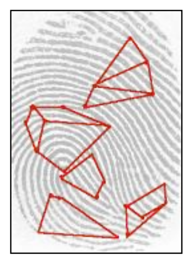

Figure 1: Sample minutiae quadruplets in a fingerprint image.

2. Features for Indexing

Given a quadrilateral, several different geometric features can be extracted. Initially, a set of 17 different features was considered. Based on a systematic evaluation process, some of these features were eliminated. In this work, seven features, $F = \{\varphi_1, \varphi_2, \delta_1, \delta_2, \rho_1, \rho_2, \eta\}$, from a minutiae quadruplet are proposed for indexing fingerprints. The geometrical interpretation of δ_1, δ_2, ρ_1, and ρ_2 is shown in Figure 2.

Appeared in Proc. of IEEE Computer Society Workshop on Biometrics at the CVPR Conference, (Colorado Springs, USA), June 2011

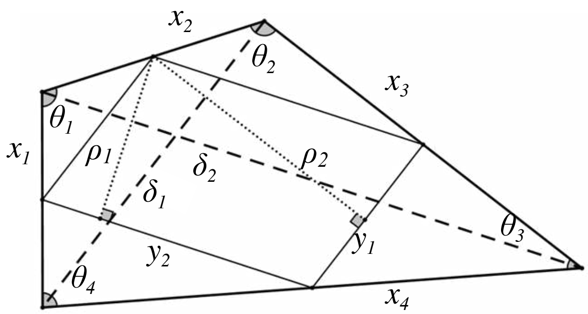

Figure 2: A minutiae quadruplet

2.1. Difference of Internal Angles

The first two features, φ_1 and φ_2, are the differences of two opposite angles in the quadruplet:

$$\varphi_1 = \theta_1 - \theta_3 \tag{1}$$
$$\varphi_2 = \theta_2 - \theta_4, \tag{2}$$

where θ_1, θ_2, θ_3 and θ_4 are the four internal angles of the quadruplet. These differences are more robust to distortion compared to the interior angles themselves. When one of the vertices changes its position due to distortion in the fingerprint image, two of the interior angles change in a similar manner and, therefore, their difference may change by a very small amount or may even remain unchanged.

2.2. Diagonals of the Quadruplet

The second pair of features, δ_1 and δ_2, are the diagonals of the quadruplet. The two diagonals, δ_1 and δ_2, can tolerate distortions due to one or two minutiae points. If a single point or two opposite points are distorted, the length of the diagonal opposite to these points does not change. Figure 3 shows an example where two opposite minutiae points, θ_1 and θ_3, have changed positions and the length of the diagonal δ_1 remains the same.

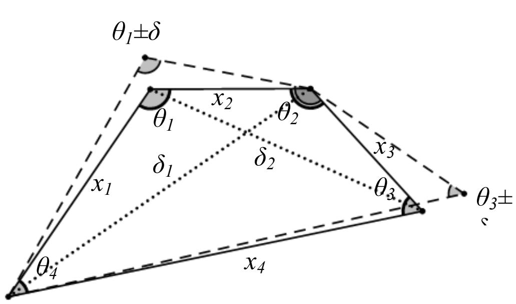

Figure 3: Change in the position of one or two opposite vertices does not affect the length of the diagonal connecting the other two vertices. In this figure, the length of diagonal δ_1 stays the same.

2.3. Heights of the Parallelogram

The next pair of features, ρ_1 and ρ_2, are the heights of the inner parallelogram, whose vertices are the midpoints of the sides of the quadruplet.

2.4. Global Feature η

The last feature is a composite global feature that combines the sides and the areas of the quadruplet and the parallelogram:

$$\eta = 100\log_{10}(\tau\upsilon)\,, \tag{3}$$

where

$$\tau = \sqrt{A_P} + \sqrt[4]{x_1 \times x_2 \times x_3 \times x_4} \tag{4}$$

and

$$\upsilon = \sqrt{A_q} + \sqrt{y_1 \times y_2} \tag{5}$$

A_p is the area of the parallelogram, x_1, x_2, x_3 and x_4 are the lengths of the sides of the quadruplet, A_q is the area of the quadruplet, and y_1 and y_2 are the lengths of the sides of the parallelogram.

The heights of the parallelogram, ρ_1 and ρ_2, together with the lengths of the diagonals of the quadruplet define the general shape of the quadruplet. Therefore, η is a general description of the shape and the size of the quadruplet and is only slightly affected by minutiae distortions.

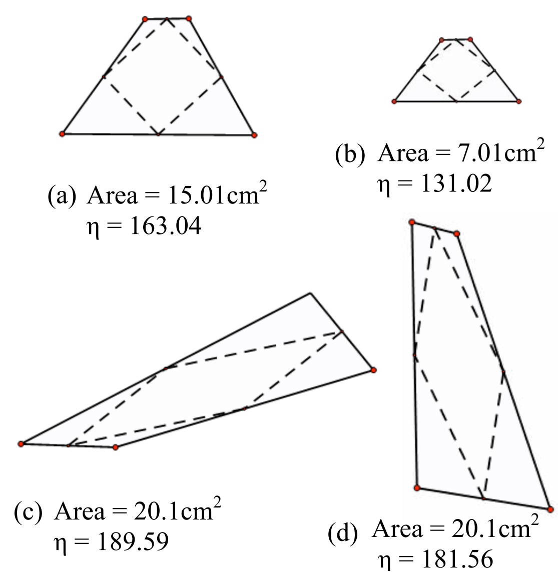

Figure 4: (a) and (b) are quadruplets with similar shapes but different sizes, while (c) and (d) are quadruplets with different shapes but same sizes.

Appeared in Proc. of IEEE Computer Society Workshop on Biometrics at the CVPR Conference, (Colorado Springs, USA), June 2011

Figure 4 shows four examples in which the quadruplets are drawn to scale. The quadruplets in Figure 4 (a) and (b) have similar shapes but different sizes and their η varies according to their respective sizes. In Figure 4 (c) and (d), the quadruplets are very different - though they have the same size (quadruplet area = 20.1cm^2), the values of η are different.

As shown in our experiments, the global feature, η, can be used as a single feature for indexing leading to relatively good results.

2.5. Types of Quadruplets

Three types of irregular quadruplets, i.e., convex, concave and reflex (crossed), can be formed from four vertex points as shown in Figure 5. Concave quadruplets were discarded and not used in the indexing experiments while all reflex quadruplets were converted to convex.

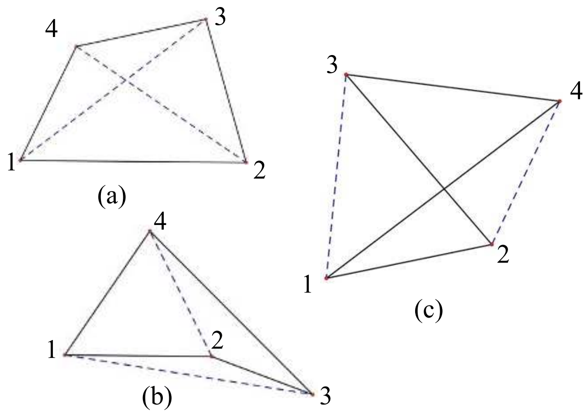

Figure 5: Irregular quadruplets. The sides are indicated by solid lines and the diagonals are indicated by dashed lines. (a) Convex. (b) Concave. (c) Reflex.

3. Proposed Indexing Algorithm

In the proposed approach, each fingerprint image is viewed as a set of quadruplets with each quadruplet being represented by the 7 aforementioned features. In the discussion below, the term feature vector is used to denote the 7 features pertaining to a single quadruplet. First, we use a training set of fingerprint images to construct an *index space*. The feature vectors extracted from all these images are clustered using the k-means algorithm [10]. The centroids of the clusters, $c_1, c_2, \ldots c_k$, computed using the arithmetic mean, define the index space.

An indexing string is computed for each image in the gallery and stored as a row in a table **T**. The length of the indexing string is the number of clusters in the index space. Thus, each column in **T** corresponds to a cluster and each row corresponds to an image from the gallery.

To construct the indexing string of an input image, its quadruplets $\{q_w \mid w = 1, 2, \ldots Q\}$, are assigned to the clusters generated in the index space using the minimum distance rule based on the Euclidean distance:

Assign q_w *to* c_k, if $k = \arg\min \{l_2(q_w, n_j), j = 1..k\}$ (6)

Here, n_j is the centroid of the jth cluster, l_2 is the Euclidean distance and c_k is the cluster id. Each quadruplet is assigned to a single cluster. The indexing string is constructed by counting the *number* of quadruplets assigned to each cluster. Thus, the indexing string of image *x* is $S(x) = \{a_1^x, a_2^x, \ldots, a_k^x\}$ where a_i^x is the number of quadruplets from image *x* which are assigned to cluster *i*, and *k* is the total number of clusters. We will refer to a_i as an accumulator for centroid *i*.

Algorithm for indexing an input image

1. Let x be the input image, $\{q_w \mid w=1\ldots Q\}$ be the quadruplets extracted from X, and $\{a_i, i=1\ldots k\}$ be the accumulators for each centroid.

2. Set the accumulators for all centroids to zero.
3. For each q_w
   Find the closest centroid k using (6)
   Increment the accumulator for k, e.g., $a_k = a_k + 1$.
4. Construct an indexing string $S(x) = \{a_1^x, a_2^x, \ldots, a_k^x\}$.
5. Insert $S(x)$ and the id of x into **T**.

During retrieval, an indexing string is created for the probe image. Next, a sorted list of clusters in descending order of the corresponding accumulators is generated. The cluster id on the top of the list corresponds to that cluster which was assigned the largest number of quadruplets from the probe. This sorted list is used to identify a small number of clusters. Specifically, the clusters in the sorted list that were assigned 60% of the quadruplets from the probe are considered.

Once the desired clusters are identified, the corresponding accumulators from table **T** are used as votes for the gallery image they belong to. The top *M* gallery images having the largest number of votes constitute the candidate list.

Table 1: Databases used to create the index space for each evaluation experiment.

Database used to create the index space	Database used for evaluation
FVC 2000 DB1	FVC 2000 DB3, FVC 2004 DB1, FVC 2002 DB1 and DB2
FVC 2000 DB2	FVC 2002 DB3
FVC 2002 DB2	FVC 2000 DB1 and DB2
FVC 2000 DB4	FVC 2002 DB4
FVC 2002 DB4	FVC 2004 DB4, FVC 2000 DB4

Algorithm for retrieving candidate images

Let $S(x) = \{a_1, a_2, \ldots, a_k\}$ be the index string of the probe p, Q be the number of quadruplets in the probe and $\mathbf{T}_{G \cdot k}$ be the table storing the accumulators for the images in the gallery.

1. Sort $S(p)$ in descending order resulting in a sorted list $\{a_{i1}, a_{i2}, \ldots, a_{ik}\}$.
2. Find the top r clusters of the probe that contain at least 60% of the quadruplets.
3. For each gallery image g sum the accumulators in $\mathbf{T}$ corresponding to the r clusters. Let the sum be C^g.
4. Sort C^g for $g = 1 \ldots G$ in descending order.
5. Retrieve the gallery images corresponding to the top n C^g values.

Step 2 in the retrieval algorithm greatly reduces the number of clusters considered during retrieval, i.e., $r \leq 7$ for most cases. Furthermore, not every identity is assigned to every cluster and, therefore, many of the C^g sums will be equal to zero. An efficient implementation of Step 4 should discard all identities for which $C^g = 0$, prior to sorting.

4. Experiments

4.1. Experimental Setup

The databases used for the experiment were the Fingerprint Verification Databases (FVC) 2000, 2002 and 2004 [11, 12, 13] each of which has four datasets - DB1A, DB2A, DB3A and DB4A. Each dataset contains 8 images for each of 100 subjects making a total of 800 images.

The dataset used to create the index space was different from the dataset used for evaluating the proposed scheme. Images of the first 25 subjects in a dataset were used for creating the index space. For each indexing experiment, the dataset used for creating the index space was chosen based on the following criteria: the image resolution was the same as the dataset used for evaluation and the scanners had similar properties. Five different index spaces were created for the 12 databases used in the indexing experiments, as shown in Table 1.

For each subject, 4 impressions were placed in the gallery while the remaining 4 impressions were used as probes. The 4 gallery impressions for a subject were selected at random. The VeriFinger SDK was used to extract minutiae points from the images.

For images containing a large number of minutiae points, the number of quadruplets may be prohibitively large. Therefore, the number of quadruplets for each image used in the experiments was empirically limited to 1200 by removing the largest quadruplets. This was done by removing the quadruplets having a diagonal larger than a threshold until the number of the remaining quadruplets reached 1200. Furthermore, concave quadruplets were not used and all reflex quadruplets were converted to convex. Reducing the number of the quadruplets is performed offline for the images in the gallery.

The number of clusters, k, used in the experiments was 50 for each dataset. This number was chosen empirically as a compromise between high penetration rates (for k~30) and low hit rates for (k~100).

4.2. Evaluation of Indexing Performance

Indexing performance can be measured using two factors: the hit rate and the penetration rate. The hit rate denotes the fraction of probes for which the selected candidate list contains the correct identity, and the penetration rate denotes the average length of the candidate list retrieved for each probe.

5. Results

The results of the experiments on the 12 databases are reported in Figures 6, 7 and 8. For each experiment, the following three scenarios were considered: datasets in their original form; datasets with 20% spurious minutiae; and datasets with 20% missing minutiae. In each plot, the labels ending with N, M and S, represent the original, the missing and the spurious minutiae sets, respectively.

An important result of the proposed approach is the consistently low penetration rate at a hit rate of 100% which varies from 18.25% to 35.25% in the original minutiae sets, from 24.50% to 39.00% in the 20% missing minutiae sets, and from 34.75% to 57.5% for the 20% spurious minutiae sets.

Furthermore, deleting 20% of the minutiae points did not degrade the performance substantially. This shows that the minutiae quadruplets are robust to low quality images.

The results of the experiments with spurious minutiae are acceptable despite the generous quantity of randomly generated spurious data that were added to all the fingerprints. This shows that the minutiae quadruplet features are robust to a certain degree of noise in the fingerprints.

Finally, the results are similar across all databases showing the robustness of the proposed technique to different scanners.

An experiment in which the global feature, η, alone was used for indexing, led to penetration rates of 12.25% and 38% at 80% and 100% hit rates, respectively, on FVC 2000 DB1.

Appeared in Proc. of IEEE Computer Society Workshop on Biometrics at the CVPR Conference, (Colorado Springs, USA), June 2011

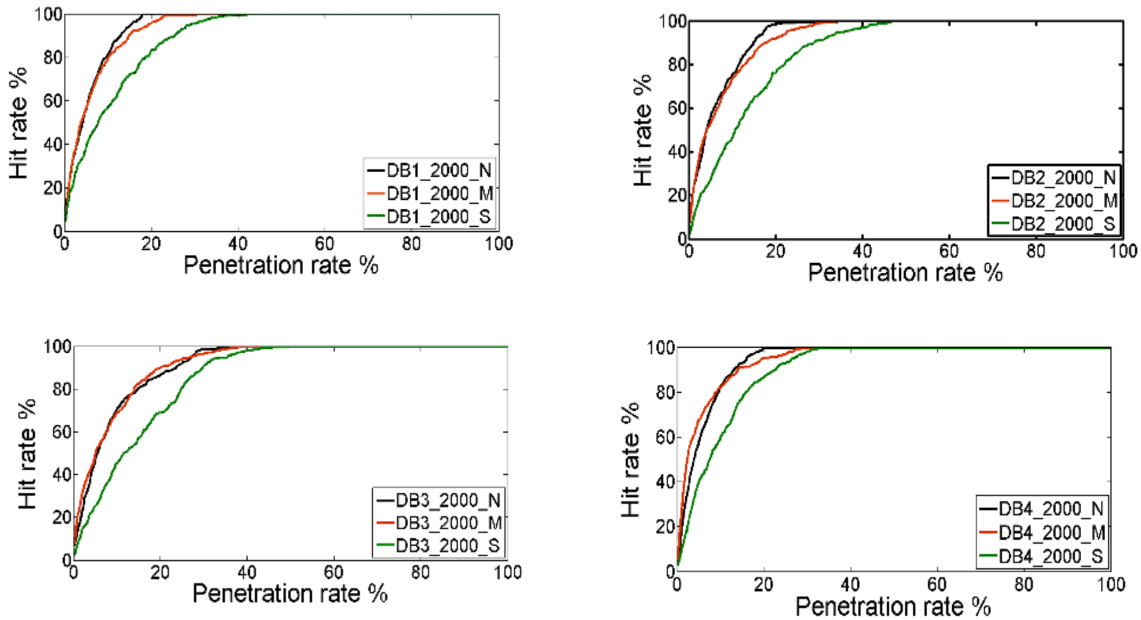

Figure 6: Performance on FVC 2000 DB1, DB2, DB3 and DB4 databases using the original minutiae data (N), 20% missing minutiae (M) and 20% spurious minutiae (S).

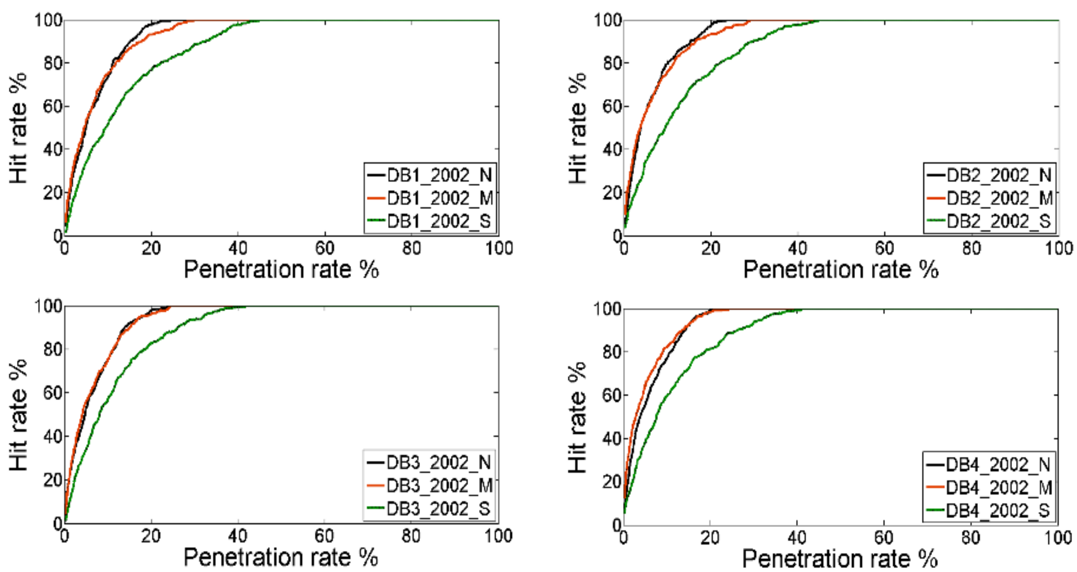

Figure 7: Performance on FVC 2002 DB1, DB2, DB3 and DB4 databases using the original minutiae data (N), 20% missing minutiae (M) and 20% spurious minutiae (S).

Appeared in Proc. of IEEE Computer Society Workshop on Biometrics at the CVPR Conference, (Colorado Springs, USA), June 2011

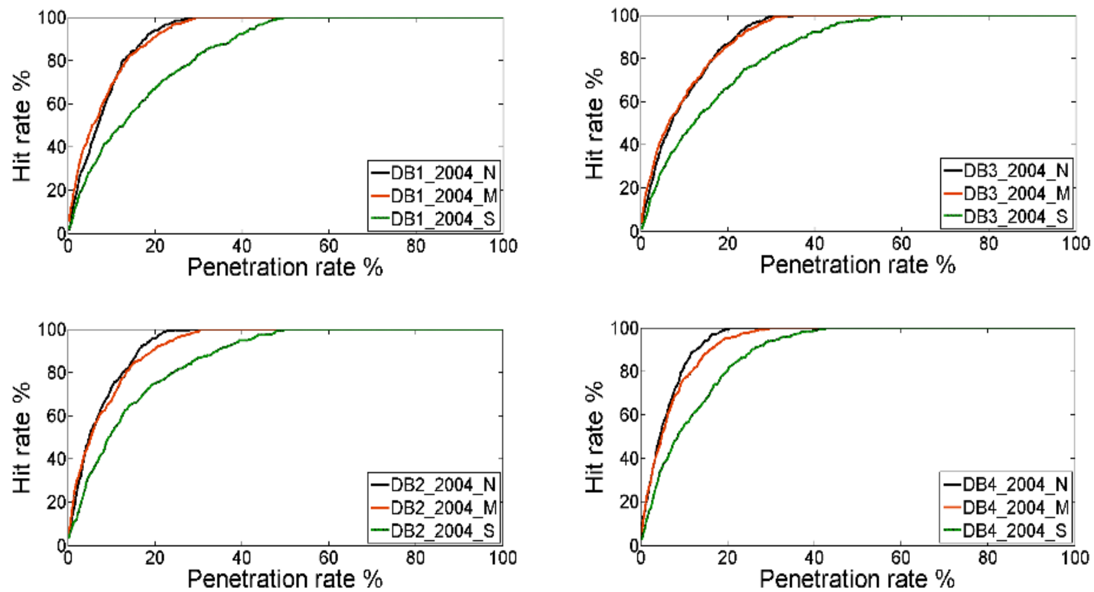

Figure 8: Performance on FVC 2004 DB1, DB2, DB3 and DB4 databases using the original minutiae data (N), 20% missing minutiae (M) and 20% spurious minutiae (S).

Table 2: Average penetration rates when using Minutiae Quadruplets, Low-order Delaunay Triangle (LoD) [7] and Minutiae Triplets [6] at hit rates of 99% and 100%. The evaluation protocol was based on [7].

	99% hit rate		100% hit rate	
FVC databases (DB1)	2002	2004	2002	2004
Minutiae Quadruplets	11.2%	11.8%	**11.8%**	**12.0%**
Low-order Delaunay Triangle [7]	**8.1%**	**10.0%**	18.1%	20.9%
Minutiae Triplets [7] (based on [6])	23.6%	27.2%	38.1%	40.9%

6. Comparison of Minutiae Quadruplets with Other Fingerprint Indexing Techniques

The proposed technique is compared with three other fingerprint indexing techniques; Minutiae triplets [6], Low-order Delaunay triangle [7] and Composite sets of reduced SIFT features [14].

The proposed technique based on quadruplets was evaluated on FVC 2002 DB1 and 2004 DB1, and compared with Liang et al.'s [7] results on minutiae triplets and low-order Delaunay triangles. For this comparison, we follow the testing scenario of Liang et al [7] and use the first three images for each subject as gallery images and the rest as probes. Table 2 shows the average penetration rates of minutiae triplets, low-order Delaunay triangle [7] and minutiae quadruplets at 99% and 100% hit rates.

The proposed technique was also compared with the method based on composite sets of reduced SIFT features [14]. For this comparison on the FVC 2000 DB2 database, the first image of each subject was enrolled in the gallery and the rest of the images were used as probes, as done in [14].

Table 3 shows the average penetration rate of Shuai et al.'s method [14] compared with minutiae quadruplets at a 100% hit rate.

Table 3: Average penetration rates when using Minutiae Quadruplets and SIFT features on FVC 2000 DB2. The evaluation protocol was based on [14].

	99% hit rate	100% hit rate
Minutiae Quadruplets	**19%**	**26%**
SIFT Features [14]	21%	91%

In Table 3, the average penetration at a hit rate of 100% is 26.33% for Minutiae Quadruplets and 91% for SIFT Features.

7. Conclusions

In this paper, the use of minutiae quadruplets has been proposed for indexing fingerprints. The consistent performance of the proposed method on the FVC 2000, 2002 and 2004 databases (set A) indicates that the proposed technique is database-independent. Experiments on fingerprints with spurious minutiae points and fingerprints with missing minutiae show that the technique is reasonably robust. The retrieval strategy is computationally inexpensive and the proposed method has small storage requirements. Further analysis is necessary to leverage the proposed technique into operational systems. Based on the experiments conducted in this work, it is apparent that minutiae quadruplets are a viable alternative to minutiae triplets for indexing.

8. Acknowledgments

This work was done when Iloanusi was a Visiting Scholar at West Virginia University. Iloanusi was supported by grants from the University of Nigeria, Nsukka. Gyaourova and Ross were supported by US NSF CAREER grant number IIS 0642554.

References

[1] D. Maltoni, D. Mario, A.K. Jain and S. Prabhakar. Handbook of Fingerprint Recognition. 2nd Ed. Springer-Verlag London Limited, pp 236 – 262, 2009.

[2] A. Ross and R. Mukherjee. Augmenting Ridge Curves with Minutiae Triplets for Fingerprint Indexing. In: Proc. of SPIE Conference on Biometric Technology for Human Identification IV, Orlando, USA. April 2007.

[3] J. Feng and A. Cai. Fingerprint Indexing using Ridge Invariants. In: Proc. of the 18th International Conference on Pattern Recognition, 2006.

[4] J. Li, W. Yau and H. Wang. Fingerprint Indexing Based on Symmetrical Measurement. In: Proc of the 18th International Conference on Pattern Recognition, 2006.

[5] R. S. Germain, A. Califano and S. Colville. Fingerprint Matching using Transformation Parameter Clustering. IEEE Computational Science and Engineering. pp. 42-49, 1997.

[6] B. Bhanu and X. Tan. Fingerprint Indexing Based on Novel Features of Minutiae Triplets. IEEE Transactions on Pattern Analysis and Machine Intelligence. 25(5): 616-622, 2003.

[7] X. Liang, A. Bishnu and T. Asano. A Robust Fingerprint Indexing Scheme Using Minutia Neighborhood Structure and Low-Order Delaunay Triangles. IEEE Transactions on Information Forensics and Security, 2(4):721-733, 2007.

[8] J. Boer, A. M. Bazen and S. H. Gerez. Indexing Fingerprint Databases Based on Multiple Features. In: ProRISC 2001 Workshop on Circuits, Systems and Signal Processing, 2001.

[9] A. Gyaourova and A. Ross. A Novel Coding Scheme for Indexing Fingerprint Patterns. In: Proc. Joint IAPR International Workshop on Structural, Syntactic, and Statistical Pattern Recognition. Springer-Verlag, pp. 755 - 764, 2008.

[10] A.K. Jain. Data Clustering: 50 Years Beyond K-Means. Pattern Recognition Letters, 31(8): 651-666, 2010.

[11] D. Maio, D. Maltoni, R. J. Cappelli, L. Wayman and A. K. Jain, "FVC2000: Fingerprint Verification Competition". In: Proc. 15th International Conference Pattern Recognition, Barcelona, September 3-8, 2000.

[12] D. Maio, D. Maltoni, R. J. Cappelli, L. Wayman and A. K. Jain, "FVC2002: Second Fingerprint Verification Competition", Proc. International Conference on Pattern Recognition, Quebec City, August 11-15, pp. 811-814, 2002.

[13] D. Maio, D. Maltoni, R. J. Cappelli, L. Wayman and A. K. Jain, "FVC2004: Third Fingerprint Verification Competition", Proc. International Conference on Biometric Authentication (ICBA), Hong Kong, pp. 1-7, July 2004.

[14] X. Shuai, C. Zhang and P. Hao. Fingerprint Indexing Based on Composite Set of Reduced SIFT Features, In: IEEE International Conference on Pattern Recognition (ICPR), Tampa, FL, USA, December 8-11, pp. 1-4, 2008.