Fingerprint Identification Problem: Using Delaunay Triangulation Technique for Model Database Indexing

Michael Khachay[1,2,3]*, Alexander Dremin[1] and Anton Leshko[1]

[1]Krasovsky Institute of Mathematics and Mechanics,
16 S. Kovalevskoy st., Ekaterinburg, Russia
[2]Ural Federal University, 15 Mira st., Ekaterinburg, Russia
[3]Omsk State Technical University, 11 Mira ave, Omsk, Russia

Abstract. A new modification of Delaunay triangulation of finite minutiae subsets extracted from fingerprints is proposed. Stability examining and numerical analysis results are presented as well.

1 Introduction

Recently developing activity of automatic biometric verification and identification is steadily increasing. A variety of biometric technologies have been developed for the last fifty years. Among these, fingerprints, face, iris, speech, and so on. Each technology has its strength and weakness. The main criteria used for comparative analysis of several biometric technologies are universality, uniqueness, collectability, permanence etc. Fingerprint biometric technology (also known as dactyloscopy) is the oldest and the most popular due to the high personality and stability of fingerprint images. For a given finger and a given person, fingerprint is just a digital gray-scale image obtained by optical scanner and containing a picture of papillary ridges and valleys. So, fingerprint verification / identification systems are based on special image processing and image analysis algorithms. While the former, verification problem seems to be solved for the most part, the later remains a great challenge for researchers and developers. Along with accuracy, scalability becomes one of the first-priority issues of developing identification methods and algorithms.

Although there are known several fingerprint verification systems comparing raw fingerprint images using correlation analysis, usually [1] the verification / identification stage is preceded by some feature extraction one. Among them, minutiae feature extraction technique is the most popular. Geometrically, minutia is an irregular point on a fingerprint image (termination, bifurcation, crossover etc. of papillary lines). These points can be considered as finite collection (template) on the plane, and many modern fingerprint analysis algorithms are based on this consideration. Unfortunately, a regular fingerprint image usually contains several dozens of minutiae and analysis of all their sub-combinations can be very computationally expensive. Several geometric techniques are developed [2] to reduce this combinatorial complexity, and triangulation of the minutiae set is one of them.

* This research was partially supported by RFBR, grants no. 13-07-00181 and 13-01-210, and Ural Branch of RAS, grants no. 12-P-1-1016 and 12-S-1017/1

Khachay M., Dremin A. and Leshko A..
Fingerprint Identification Problem: Using Delaunay Triangulation Technique for Model Database Indexing.
DOI: 10.5220/0004394400940100
In *Proceedings of the 4th International Workshop on Image Mining. Theory and Applications* (IMTA-4-2013), pages 94-100
ISBN: 978-989-8565-50-1
Copyright © 2013 SCITEPRESS (Science and Technology Publications, Lda.)

In the paper, an improved version of the Delaunay triangulation technique for fingerprint identification is presented.

2 A Structure of Identification System

From conceptual point of view, any automated fingerprint identification system consists of two subsystems. First of them (we call it *offline*) is used in the preparation stage during a model database construction, and the second one makes an identification of query fingerprints on the basis of the prepared database.

To describe the first subsystem briefly, it is convenient to use the well known *black box* model. By virtue of any standard minutiae extracting algorithm, an offline subsystem maps the initial model fingerprint set $\mathfrak{I} = \{I_j : j = 1, \ldots, N\}$ into the family of finite subsets of $\mathbb{Z}^3_+$ (cube of the set of nonnegative integer numbers). Actually, any model image $I_j \in \mathfrak{I}$ is mapped to subset $T(I_j) = \{(x_i, y_i, w_i) \in \mathbb{Z}^3_+ : i = 1, \ldots, N_j\}$ that is called *a template*. For any triple, x_i and y_i coordinates coincide with appropriate geometric coordinates of the i-th minutia detected on the image, and w_i is a confidence level of this detection. So, in the beginning of the first stage we have a set $\mathfrak{B} = \{T_j = T(I_j)\}$ of templates of the initial images (which is called *a model database*). Further, any couple (T_j, q) we assign a minutiae subset $T_{j,q} = \{(x_i, y_i, w_i) \in T_j : w_i \geq q\}$ of the model template consisting of minutiae having sufficiently high accuracy level. In the sequel, we consider projections of these subsets onto planes $H_q = \{(x, y, w) : w = q\}$ which are parallel to the coordinate plane xOy.

On the second (online) stage, the query image I is processed by the similar way, and a template $T = T(I)$ is produced, after that the final identification decision is made by one-to-one matching T with corresponding candidates subset $\mathfrak{B}_T \subset \mathfrak{B}$ of the model database. Time complexity of this procedure (for a fixed template T) is $O(M|\mathfrak{B}_T|)$, where M is complexity of the inner matching algorithm. So, the problem is to construct the reducing algorithm R constructing for any T a subset $R(T) = \mathfrak{B}_T$ satisfying the following additional constraints.

1. $|R(T)| \ll N = |\mathfrak{B}|$.
2. Let a fingerprint I producing the query template T belong to some known person and the model database contain templates produced by another him (or her) fingerprints. Denote the subset of these templates by $\mathfrak{B}'_T$ and the conditional probability of the event $\mathfrak{B}'_T \cap R(T) = \varnothing$ by P_T. Then $P_T \leq \alpha$ for some given value $\alpha \in (0, 1)$.

Mathematically, this problem is equivalent to the construction problem of the efficient computable mostly powerful test statistic of the significance level α for the null hypothesis 'known person'. For any query template T the test produces a 'candidates' subset $R_\alpha(T)$ for the subsequent one-to-one matching (with T).

Several ways to constructing a solution of this problem are known. The approach based on preliminary clusterization of the model database by the type of *core* of the initial fingerprint images seems [1] to be the earliest. According to this approach, at the online stage a core type is for the query template is computed after that the search can be narrowed to the corresponding cluster. Unfortunately, number of known core types is small and distribution of the real fingerprints (among them) is far from the uniform.

The another approach based on indexing of the model database supposed [2] to be more promising. Indexing procedures improve the classical two-stage identification scheme in the both stages. At offline stage, a model database is indexed using some hash functions. At the online stage the required subset $R(T)$ is constructed from the models with the most similar hash values to ones calculated from the query template T.

During the indexing substage, for any model template, several *partial invariants* (that of values of the geometrical nature that are almost invariant to the given transformation group on the plane) are computed and quantized. For instance, let some numerical features $f_1, f_2, \ldots, f_k$ of the geometrical shapes of some kind formed by the fingerprint minutiae are used as partial invariants, then for any model T_i and for any shape S of interest the following record $g_1(S), \ldots, g_k(S), r_i$ will be included into the indexing table. Here g_j is quantized value of the feature f_j and r_i is reference to model T_i. Such a way, any model template T_i is transformed to some finite subset in k-dimensional indexing space.

The online identification stage is started on calculating the similar invariant values for the query template. The computed k-dimensional vectors are filtered using some system of the additional constraints that of control parameters of the algorithm. Further, the separated vectors are used for search in index table and posterior probabilities of models T_i computing. The result ordered subset $R(T)$ is constructed from the most probable models according to this posterior measure.

Performance of indexing algorithms is suggested [2] to estimate by values of *correct index power (CIP)*.

Suppose, we have a couple (T_i, T_i') of fingerprints for any respondent obtained from the same finger. Construct the model database $\mathfrak{B}$ from the former elements of each couple, and test database $\mathfrak{C}$ from the later ($|\mathfrak{B}| = |\mathfrak{C}| = N$, by construction). The model $T_i \in \mathfrak{B}$ is said to be *correctly indexed* by the algorithm R if $T_i \in R(T_i')$. Let $N_{ci}(R)$ be the number of correctly indexed elements of $\mathfrak{B}$ then

$$CIP(R) = \frac{N_{ci}(R)}{N}. \tag{1}$$

It's clear, that $CIP(R)$ is a stochastic variable which depend not only on the algorithm R in question but also on random choice of $(\mathfrak{B}, \mathfrak{C})$. Its population value can be statistically estimated on representative fingerprint samples. In this paper the well known '*NIST Special Fingerprint Database 4*' [3] is used for such estimations.

3 Partial Invariants

The system of invariants constructed in this paper generalizes the system developed in [4] and contains quantities that are invariant to the similarity transformations group (over the plane). For a fixed accuracy value q of detected minutiae and any template T a projection $\Pi_q(T)$ of the set $T_q = \{(x_i, y_i, w_i) \in T : \ w_i \geq q\}$ onto the plane $H_q = \{(w, y, w) : \ w = q\}$ is assigned and Delaunay triangulation [5] of the set $\Pi_q(T)$ is constructed. The choice of Delaunay triangulation method is due to the following reasons

– such a triangulation is unique for any non degenerate finite set on the plane;

- the resulting triangulation consists of $O(n)$ facets that is substantially smaller than the number $O(n^3)$ of all possible triangles with the vertices of the given n-point set;
- this triangulation can be constructed efficiently, we use an algorithm [6] with time-complexity $O(n \log n)$;
- a topological structure of the resulting triangulation is stable [7] to small perturbations of the initial data.

Suppose, a triangle $\varDelta$ is a triangulation facet with edges $a \leq b \leq c$. Assign to this triangle a vector $\nu(\varDelta) = [\alpha, \beta, \gamma]$ by the formulas $\alpha = b/c$, $\beta = a/b$, and $\gamma = \cos C$. This vector is invariant to any similarity (e.g. translation, rotation and scaling) transform on the plane and satisfies the following inequalities

$$\frac{1}{2} < \alpha \leq 1,\ 0 < \beta \leq 1,\ -1 < \gamma \leq \frac{1}{2}.$$

Suitable discretized (particularly, to automatically distinguish the triangles-isomers) these parameters are used in both stages, offline and online.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

4 Proposed Algorithm

4.1 Indexing Stage

Input.

1. Model database $\mathfrak{B} = \{T_j :\ j = 1, \ldots, N\}$.
2. Minimum accuracy level q for detected minutiae.
3. Maximum index values n_1, n_2, n_3.

Output. Set-valued map $h : \mathbb{Z}_q^3 \to 2^{\mathfrak{B} \times \mathfrak{D}}$ (index table) defined on integer lattice

$$[0, \ldots, n_1] \times [0, \ldots, n_2] \times [0, \ldots, n_3]$$

as follows: any triple (i, j, k) is assigned to set of couples $(T_o, \varDelta_t)$, where T_p is some model template and the triangle $\varDelta_t$ is a facet of Delaunay triangulation of $\varPi_q(T_p)$ such that discretized value of the vector $\nu(\varDelta)$ is equal to (i, j, k).

4.2 Identification (Query) Stage

Input.

1. A query template T and minutiae extracting accuracy level q.
2. Index (hash) table h.
3. Maximum number L of the resulting hypotheses.
4. Discretization parameters $n_1, \ldots, n_4$ for similarity transformations.

Output. A sequence of couples (T_p, r_p, S_p), where T_p is extracted model (identification hypothesis), r_p is its rating, and S_p is affine similarity transform assigning T to T_p. The sequence is ordered by r_p by descending.

Scheme.

1. Similar to the above considerations, for any Delaunay triangulation facet Δ_t of the projection $\Pi_q(T)$ (for the query template T) an appropriate index cell (i_t, j_t, k_t) and the set $h(i_t, j_t, k_t)$ are assigned.
2. For any triangle $\delta_v \in h(i_t, j_t, k_t)$ an appropriate similarity transform S_{tv} (mapping the vertexes of Δ_t into corresponding vertexes of δ_v) is computed. Scaling parameter λ_{tv}, cosine $\cos\varphi_{tv}$ of the rotation angle, and translation vector b_{tv} are discretized, and the corresponding model template T_v is added as entry of the secondary index table (by the increasing of its rating).
3. The required L hypotheses and the appropriate similarity transforms are found by searching over the secondary table (Fig. 1).

5 Testing of the Algorithm

Training (tuning the controllable parameters) and testing of the algorithm were made on the well known NIST-4 Special Fingerprint Database, the respectable testing polygon for modern fingerprint verification/identification heuristics. By structure, this database consists of 2K fingerprint pairs, for each of them both images (denoted by 'fXXXX' and 'sXXXX') are obtained from the same finger.

We use this database for solving the following additional problems.

1. Proving the stability of the proposed indexing scheme to small perturbations of the initial data, such as addition (deletion) of minutiae and modifications of their geometrical locations.
2. Discretization parameters tuning for primal and secondary indexing tables.

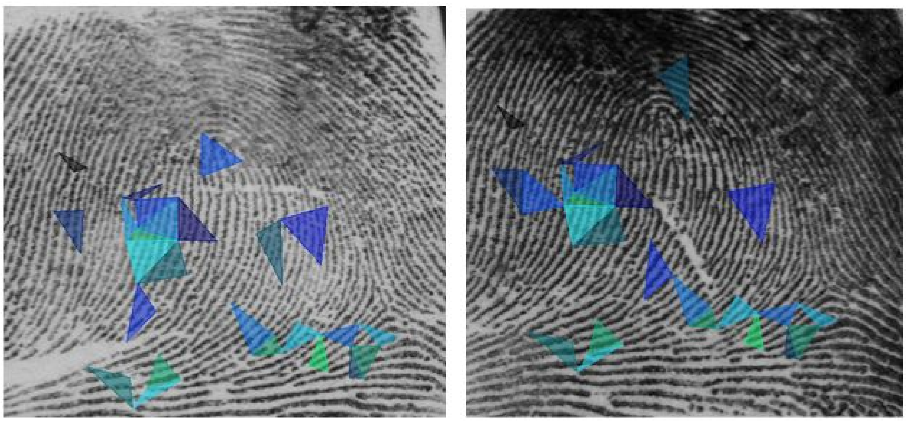

Fig. 1. Query fingerprint S1678 and the most valuable hypothesis F1678.

Accuracy level for detected minutiae was fixed to $q = 64$. In both problems, the subset the dataset of 1923 pairs (96%), where f-image produced a template with at least 50 minutiae, was chosen.

5.1 Stability Examining

This kind of testing proceed on special synthetic dataset obtained from the mentioned above NIST-4 database. According to the well known "white noise" model, for any f-image from the initial dataset, geometrical locations of the minutiae were modified using independent identically, $N(0, \sigma)$-distibuted normal noise

$$x'_i = x_i + \xi_i,\ y'_i = y_i + \eta_i.$$

For an additional parameter $r \in [1, 2]$, a minutiae (x'_i, y'_i, w_i) was included to perturbed template iff $\xi_i^2 + \eta_i^2 \leq r^2\sigma^2$ Such a way, for any initial model template a 20-element stochastic perturbed sample was assigned.

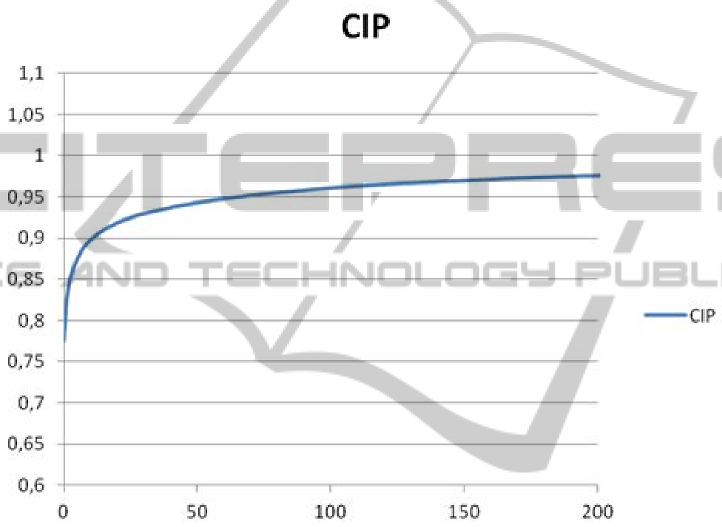

Fig. 2. CIP-index for $\sigma = 3$ and $r = 1.5$.

Further, the initial templates were used as models in the indexing stage and any constructed template was identified using the algorithm proposed. Obtained numerical data confirm the known theoretically proved [7] stability result. Particularly, for $\sigma = 1$ and $r \in [1.5, 2]$ (from 33% to 13% of excluded minutiae in average) 100% perturbed templates were classified correctly within $L = 1$. Increasing of σ leads to increasing of the L-value, as expected. But stability of the entire algorithm remains high. For instance, for $\sigma = 3$ and $r = 1.5$, CIP-value for $L = 1$ was 77%, and for $L = 10$ (0.5% of the initial database) - more then 89% (Fig. 2).

5.2 Tuning and the Final Testing

For training (parameter tuning) a subset of 430 (21.%) fingerprint pairs is used, where f-image possessed at least 100 minutiae on the accuracy level 64, while the for the testing the complement (of this subset) to the initial database were taken. During the training stage the parameters were tuned by local search heuristic. The optimal values of parameters are $8 \times 8 \times 8$ for a primal index table (hash) and $17 \times 17 \times 17 \times 47$ for the secondary. The final results for several voting strategies are presented in Fig. 3.

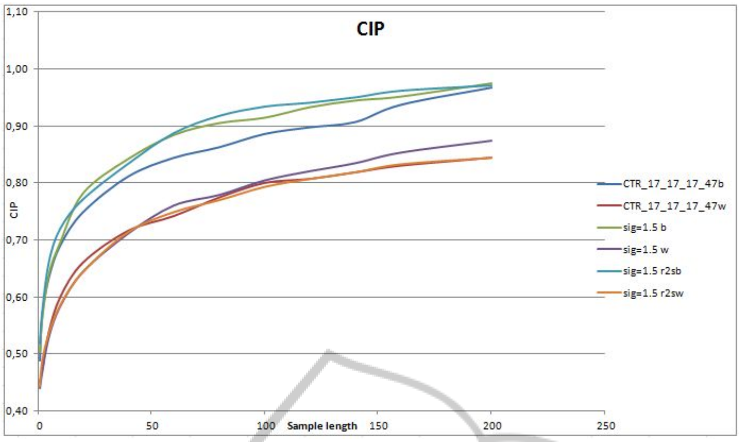

Fig. 3. CIP-analysis of several voting strategies.

6 Discussion

A detailed comparative analysis of the known indexing techniques for fingerprint identification problem is presented in the book [2]. The algorithm with 85% CIP value for $L = 0.1N$ is recognized optimal among them. The indexing scheme used in this algorithm was based on considering the all possible triangles with vertexes in minutiae locations, and time-complexity of its online stage was $O(n^3)$. Our method has CIP value of $82 \pm 5\%$ for the same L, while its time-complexity is $O(n \log n)$ thanks to Delaunay triangulation.

7 Conclusions

A new indexing technique for the model database in fingerprint identification is proposed. The technique is based on Delaunay triangulation of finite minutiae set on the plane and original indexing parameters adjusting heuristic. Computational experiments proved high performance of the proposed algorithm in relation to the known analogs.

References

1. Maltoni D., Maio D., Jain A.K., Prabhakar S. Handbook of Fingerprint Recognition. New York. Springer. 2005.
2. Bhanu B., Tan X. Computational Algorithms for Fingerprint Recognition. Boston. Kluwer Academic Publishers, 2004.
3. NIST Special Database 4. http:// http://www.nist.gov/srd/nistsd4.cfm
4. Bebis G., Deaconu T., Georgiopulos M. Fingerprint Identification Using Delaunay Triangulation IEEE International Conference on Intelligence, Information, and Systems, ICIIS, 1999.
5. Du Q., Wang D. Recent Progress in Robust and Quality Delaunay Mesh Generation. Journal of Computational and Applied Mathematics. 195 (2006) no. 1-2. 8–23.
6. Skvortsov A.V. A survey of algorithms fro constructing a Delaunay triangulation. Numerical Methods and Programming. 3 (2002) Sect. 1. 14–39.
7. Guibas L., Knuth D., and Sharir M. Randomized Incremental Construction of Delaunay and Voronoi Diagrams. Algorithmica. 7 (1992) 381–413.