Fingerprint Matching Based on Neighboring Information and Penalized Logistic Regression*

Kai Cao, Xin Yang, Jie Tian, Yangyang Zhang, Peng Li, and Xunqiang Tao

Institute of Automation, Chinese Academy of Sciences, Beijing 100190 China
tian@ieee.org, jie.tian@ia.ac.cn
http://www.fingerpass.net

Abstract. This paper proposes a novel minutiae-based fingerprint matching algorithm. A fingerprint is represented by minutiae set and sampling points on all ridges. Therefore, the foreground of a fingerprint image can be accurately estimated by the sampling points. The similarity between two minutiae is measured by two parts: neighboring minutiae which are different in minutiae pattern and neighboring sampling points which are different in orientation and frequency. After alignment and minutiae pairing, Nine features are extracted to represent the matching status and penalized logistic regression (PLR) is adopted to calculate the matching score. The proposed algorithm is evaluated on fingerprint databases of FVC2002 and compared with the participants in FVC 2002. Experimental results show that the proposed algorithm achieves good performance and ranks 5th according to average equal error rate.

Keywords: Minutia, fingerprint matching, penalized logistic regression, ridge.

1 Introduction

A fingerprint is the pattern of ridges and valleys on the surface of a fingertip. Due to their strongpoint, such as uniqueness and permanence, fingerprint recognition becomes one of the most important biometric technologies. Previous researchers have proposed many kinds of fingerprint matching algorithms [1,2,3]. According to the fingerprint representation, the existing fingerprint algorithms can be classified into the following categories: image or correlation based [12], ridge feature based [13], orientation feature based [14] and minutiae based [15]. The performance of image based algorithm is largely affected by non-linear distortion. Ridge image is an effective and compact representation of the fingerprint image. However, the correspondences between ridges are difficult to establish due to break and wrong connection in low quality fingerprint images. The difference between two orientation fields is not evident if two fingerprints have similar singular locations. A minutia represents local discontinuities and marks

---

position where the ridge comes to an end or bifurcates into two, which is one of the discriminative and reliable features. However, the information provided by only minutiae set is limited since it discards the global ridge information. Therefore, researchers proposed to make use of other discriminatory features to establish the correspondence among minutiae and enhance the performance [1,2,3]. Different from above methods, we combine minutiae and all ridges to represent the fingerprint and propose a novel similarity measure methodology based on neighboring minutiae and neighboring sampling points on the ridges. The neighboring minutiae capture the difference of local minutiae pattern, while the neighboring sampling points capture the difference of ridge orientation and ridge frequency around two minutiae.

After minutiae pairing the matching score computation is another challenging problem. A traditional way to calculate the matching score is $n^2/(N_T \times N_I)$, where N_T and N_I represent the numbers of minutiae in template and query minutiae sets, respectively, and n is the number of matched minutiae in both sets. Bazen and Gerez [4] claimed that using $n^2/(N_T + N_I)$ to compute the matching score will give better results. However, both of these two formulas can not deal with low quality fingerprint images and small overlapping region. Recently, Feng[5] tried to solve the problem of minutiae-based matching as a two class pattern recognition problem and used support vector machine (SVM) to give the classification result. From Bayesian decision theory we can get the least error if we know the probability distribution of the features. In this paper we adopt penalized logistic regression (PLR) [9] to give the matching result, which has the advantage of providing the estimation of the underlying probability. To our knowledge it is the first work that tries to solve the minutiae-based matching as a probability estimation problem.

The rest of this paper is organized as follows. In section 2, fingerprint representation is described. In section 3 similarity measure method between two minutiae is discussed. In section 4 the penalized logistic regression model is introduced. In section 5 experimental results are reported and the conclusion is drawn in section 6.

2   Fingerprint Representation

For an input gray-scale fingerprint image, the method proposed by Hong et al. [6] is used to enhance the image. Then the thinned image is obtained and postprocessed by the knowledge-based enhancement method [7]. The original image and the thinned image are displayed in Fig.1(a)-(b).

In this paper we combine minutiae with all sampling points on the ridges to represent a fingerprint. The minutiae points are detected from the thinned image by Hong's method [6]. For poor fingerprint images, some spurious minutiae may exist. Thus, it is necessary to deal with the spurious minutiae. A minutia that satisfies one of the following two conditions will be regarded as a spurious minutia and discarded: (1) the ridge (s) associated with the minutia is (are) shorter than a threshold; (2) the number of minutiae circled the minutia within a small radius is larger than a threshold. The valid minutiae set is represented as $\{(x_i, y_i, \theta_i, t_i)\}_{i=1}^n$, where i is the index of the minutia, (x_i, y_i) is its coordinate, θ_i is the direction in the range $(-\pi, \pi]$, and t_i is the minutia type (termination or bifurcation).

Each ridge in the thinned image is traced. The ridges associated with a bifurcation are treated as three different ridges. To make the tracing process efficient, we use the following tracing order: (1) ridges associated with a bifurcation; (2) ridges associated with a termination; (3) ridges without associated minutiae. If the length of a ridge is shorter than a threshold (shorter than 12 pixels in our experiment) then the ridge is regarded as an invalid one and discarded. Each valid ridge is sampled at a constant interval. Then all of the ridges in a fingerprint are represented as $SP = \{\{x_{ij}, y_{ij}, \theta_{ij}\}_{j=1}^{nr_i}\}_{i=1}^{N_r}$, where x_{ij}, y_{ij} and θ_{ij} denote the x, y coordinates and tangential direction in the range $(-\pi/2, \pi/2]$ of the jth point on the ith ridge respectively, N_r denotes the number of valid ridges in the thinned image, nr_i denotes the number of sampling points of the ith ridge. An example of ridge tracing and sampling is illustrated in Fig.1(c)-(d).

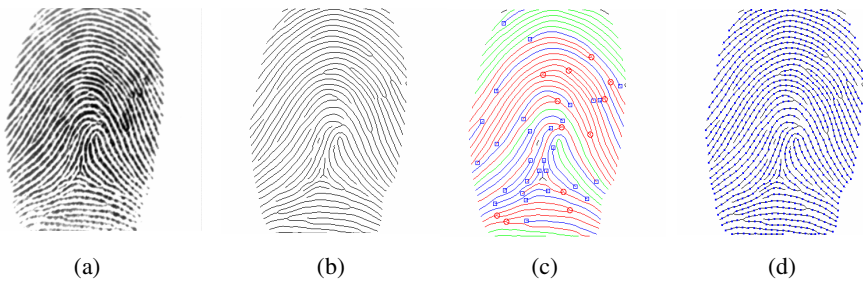| | | | |
| (a) | (b) | (c) | (d) |

Fig. 1. (a) Intensity image (500 dpi); (b) thinned image; (c) minutiae and an example of ridge tracing (red curves denote ridges associated with a bifurcation, blue curves denote ridges associated with a termination, green curves denote ridges without associated minutiae and black curves denote invalid ridges); (d) thinned image and sampling points

3   Similarity Measure between Two Minutiae

In this section, we propose an approach to obtain the foreground of the fingerprint from the sampling points and a novel similarity measure between two minutiae using neighboring minutiae and neighboring sampling points.

3.1   Foreground of Fingerprint

Tico and Kuosmanen [1] proposed an effective matching score calculation approach by only considering the minutiae that fall inside the common region of both fingerprints. The similarity calculation between two minutiae based descriptors defined by Feng[5] was also confined to the common region of two fingerprints. The common region plays an important role in fingerprint matching and it can be regarded as the intersection of two fingerprint foregrounds after registration. However, the foreground is hard to estimate by the minutiae set. In this paper, we use Graham's algorithm [8] to determining the convex hull of the sampling set which can be viewed as the boundary of the fingerprint. The points inside the convex hull are regarded as the foreground of the fingerprint. Fig. 2 illustrates the process.
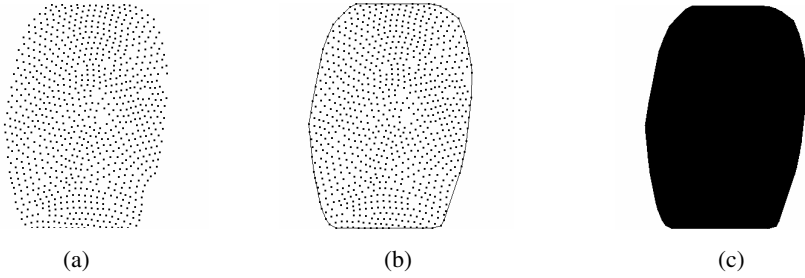
     (a)   (b)   (c)

Fig. 2. (a) Sampling points; (b) convex hull; (c) foreground of fingerprint

3.2 Similarity Measure Based on the Neighboring Minutiae

Suppose p is a minutia in the template fingerprint and q is a minutia in the input fingerprint, the similarity calculation process can be divided into two stages: (1) minutia p and its neighbors are mapped to the coordinate system of q, (2) minutia q and its neighbors are mapped to the coordinate system of p.

In the stage 1, let $N(p, r) = \{p_i\}_{i=1}^{n_p}$ denote the set of the neighboring minutiae circled p within r radius, $N(q, r + \Delta r) = \{q_i\}_{i=1}^{n_q}$ denote the set of the neighboring minutiae circled q within $r + \Delta r$ radius and T represent the corresponding rigid transformation from p to q. Each minutia p_i in $N(p, r)$ is mapped to p_i' using T. The distance difference between p_i' and q_j is calculated as follows,

$$dLen_1(p_i', q_j) = \sqrt{(x_{p_i'} - x_{q_j})^2 + (y_{p_i'} - y_{q_j})^2}. \tag{1}$$

The orientation difference between p_i' and q_j is calculated as

$$d\phi_1(p_i', q_j) = \begin{cases} 2\pi - \left| \theta_{p_i'} - \theta_{q_j} \right| & \text{if } \left| \theta_{p_i'} - \theta_{q_j} \right| > \pi, \\ \left| \theta_{p_i'} - \theta_{q_j} \right| & \text{otherwise.} \end{cases} \tag{2}$$

Then the contribution of p_i with respect to the minutia p is computed as

$$C_{p_i} = \max_{q_j} f(dLen_1(p_i', q_j), disThr1, disThr2) \cdot f(d\phi_1(p_i', q_j), oriThr1, oriThr2), \tag{3}$$

where $disThr1$ and $disThr2$ are two distance threshold, $oriThr1$ and $oriThr2$ are two orientation distance threshold, and function f is defined as

$$f(x, th1, th2) = \begin{cases} 1 & \text{if } x \leq th1, \\ 0 & \text{if } x > th2, \\ \frac{x - th1}{th2 - th1} & \text{otherwise.} \end{cases} \tag{4}$$

In the stage 2, we define another two neighboring structures: $N(q, r)$ and $N(p, r + \Delta r)$, the meaning of which is similar as stage 1. We use the same symbol T to represent the relative rigid transformation from q to p. Each minutia q_j is mapped to q_j' using T. Similar to the stage 1, the contribution of q_j to the minutia q is computed as follows

$$C_{q_j} = \max_{p_i} f(dLen(p_i, q'_j), disThr1, disThr2) \cdot f(d\phi_1(p_i, q'_j), oriThr1, oriThr2). \quad (5)$$

The similarity based on neighboring minutiae between p and q is measured using the following formula

$$MS_{pq} = \frac{1 + \sum_{p_i \in N(p,r)} C_{p_i}}{M_p + bias} \cdot \frac{1 + \sum_{q_j \in N(q,r)} C_{q_j}}{M_q + bias}, \quad (6)$$

where M_p is the number of the minutiae in $N(p, r)$ whose transformed version are in the foreground of input fingerprint, M_q is the number of the minutiae in $N(q, r)$ whose transformed version are in the foreground of the template fingerprint and $bias$ is a parameter.

3.3 Similarity Measure Based on the Neighboring Sampling Points

In this section, we propose a similarity measure method which can capture both ridge orientation and ridge frequency difference. Similar to the section 3.2, there are two stages in this similarity measure process for minutiae pair p and q from template and input fingerprint, respectively.

In the stage 1, let $SN(p, sr) = \{s_i\}_{i=1}^{sn_p}$ denote the sn_p sampling points circled p within rs radius, $SN(q, sr + \Delta sr) = \{r_j\}_{j=1}^{sn_q}$ denote the sn_q sampling points circled q within $sr + \Delta sr$ radius and T denote the relative transformation. Each sampling point s_i in $SN(p, sr)$ is mapped to s'_i using T. The Euclidean distance is not appropriate to this condition because position of the sampling points will be slight different due to the different initial tracing point. In this work, we adopt digital distance map [16] to measure the distance between the sampling point pair, formally

$$dLen_2(s'_i, r_j) = \left| (s'_s - r_j)^T \eta_{s'_i} \right|, \quad (7)$$

where $\eta_{s'_i}$ is the normal to the direction of sampling point s'_i. Here we use s'_i and r_j to represent their coordinates for convenience. This formula gives a "point to line" distance, which is appropriate to reflect the ridge frequency. Fig.3 illustrates this case. The orientation distance between s'_i and r_j is computed using the following formula:

$$d\phi_2(\theta_1, \theta_2) = \begin{cases} \pi - |\theta_1 - \theta_2| & \text{if } |\theta_1 - \theta_2| > \frac{\pi}{2}, \\ |\theta_1 - \theta_2| & \text{otherwise.} \end{cases} \quad (8)$$

Then the contribution of s_i with respect to the minutia p is computed as

$$C_{s_i} = f(dLen_2(s'_i, r_{j*}), sdThr1, sdThr2) \cdot f(d\phi_2(s'_i, r_{j*}), soThr1, soThr2), \quad (9)$$

where

$$j^* = \arg\min_{j, dLen_1(s'_i, r_j)) < Thr} dLen_2(s'_i, r_j) \quad (10)$$

In stage 2 the calculation process is similar to the stage 1. Then the similarity based on neighboring sampling points between p and q can calculated as the following

$$SS_{pq} = \frac{\sum_{s_i \in SN(p,sr)} C_{s_i}}{s_p+1} \cdot \frac{\sum_{r_j \in SN(q,sr)} C_{r_j}}{s_q+1}, \tag{11}$$

where S_p is the number of the sampling points in $SN(p, sr)$ whose transformed version are in the foreground of input fingerprint, S_q is the number of the sampling points in $SN(q, sr)$ whose transformed version are in the foreground of the template fingerprint.
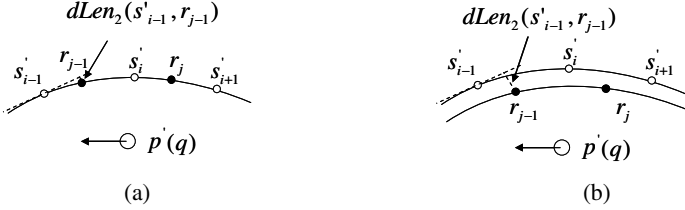

(a)   (b)

Fig. 3. (a) Frequency around p and q is same, then $dLen_2(s'_{i-1}, r_{j-1})$ is close to 0; (b) frequency around p and q is different then $dLen_2(s'_{i-1}, r_{j-1})$ is larger than that in (a)

3.4 Combined Similarity Measure

The similarity measure based on the neighboring minutiae capture local minutiae pattern while the similarity based on the neighboring sampling point capture the local information of the orientation and the location distribution. Therefore, the two similarity functions can be combined to measure the similarity between the minutiae pair. Weighted mean rule and product rule are two widely used fusion rules, shown as follows:

$$S_{pq} = \eta \cdot SS_{pq} + (1 - \eta) \cdot MS_{pq}, \tag{12}$$

$$S_{pq} = SS_{pq} \cdot MS_{pq}, \tag{13}$$

where η is a weighting coefficient. Experiment in section 5.1 will give the performances of the two formulas.

4 Matching Score Computation

Let $\{p_i\}_{i=1}^{N_p}$ and $\{q_j\}_{j=1}^{N_q}$ denote two minutiae sets from template and input fingerprint, respectively, $\{s_{ij}\}_{i=1,j=1}^{N_p,N_q}$ denote the similarities between minutiae pair. The minutiae matching algorithm proposed in Ref.[5] is adopted to establish the correspondences. It contains the following steps: (1) normalize the similarities and sort them in decreasing order; (2) one of the top N_α minutiae pairs is chosen as reference pairs, using which two minutiae sets are aligned; (3) Nine features in section 4.1 are extracted and PLR is used to compute the score;(4) this process repeat N_α times and the largest score is used as the matching score. In the following, we describe the feature extraction and PLR.

4.1 Feature Representation

Nine features are selected to represent the matching status:

(1) The number of matched minutiae pairs (n);

(2) The ratio of matched minutiae ($mRatio$) in the common region, which is defined as:

$$mRatio = (2n)/(IN_T + IN_I),$$

where IN_T and IN_I denote the number of minutiae from the template and input fingerprint located inside the common region, respectively.

(3) The mean similarity ($simi$) of the matched minutiae pair.

(4) The mean difference of the length ($lenDiff$) between two minutiae pairs.

(5) The mean difference of the minutiae direction between two minutiae pairs

(6) The average number of sampling points ($aveSNum$) in the common region, which is defined as

$$aveSNum = (S_T + S_I)/2.$$

where the S_T and S_I are the number of sampling points from the template and input fingerprint, which are located in the common region, respectively. This feature describes the size of the common region.

(7) The ratio of the difference $SNumDiff$ of the sampling points, which describes the frequency difference in the common region and is defined as

$$SNumDiff = |S_T - S_I| / (S_T + S_I).$$

(8) The corresponding sampling points are considered to be matched if their associated minutiae are matched. If the number of the sampling points associated with one minutia is difference from the other of the pair. We choose the smaller number as the number of the matched sampling points. Suppose there are si matched sampling points for minutiae pair i. Then the total number of the matched sampling points ($spNum$) is calculated as follows:

$$spNum = \sum_{i=1}^{n} s_i.$$

(9) The mean difference between the minutiae and the sampling points ($MSLenDiff$) is adopted to represent the matching status of the texture. Suppose that q_i and p_i are the ith minutiae pair, sq_j and sp_j are the last matched sampling points for the jth minutiae pair q_j and p_j, then the $MSLenDiff$ can be calculated as follows:

$$MSLenDiff = \frac{1}{n} \sum_{i=1}^{n} |dLen_1(q_i, sq_j) - dLen_2(p_i, sp_j)|.$$

4.2 Penalized Logistic Regression

The PLR can be divided into two processes: training and validation. We assume there are N samples in the training set. Each sample can be in one of two classes: class 0 and class 1. Class 0 means the two fingerprints are from two different fingers while class 1 means the two fingerprints are from the same finger. Let

$x_i^T = (x_{i_0}, x_{i_1}, x_{i_2}, \cdots, x_{i_d})$ be the ith feature vector with $d+1$ features, where x_{i_0} is the constant term with the value 1, y be the class label of the feature vector. The task of the training is to estimate parameters β in the probability model $p(x; \beta)$, where $p(x; \beta) = P(y = 1|X = x; \beta)$ is the conditional probability of a feature vector belonging to class 1 given $X = x$. Since only two classes are considered, the conditional probability of x representing class 0 is consequently $1 - p(x; \beta)$. The normal logistic regression model for a two class problem is

$$f(x; \beta) = \log \frac{p(x;\beta)}{1 - p(x;\beta^T)} = \beta^T x, \tag{14}$$

where $\beta = [\beta_0, \beta_1, \cdots, \beta_d]^T$ are the $d+1$ parameters that need to be estimated and $f(x; \beta)$ is formally, Then the penalized binomial log-likelihood for N observation is

$$L(\beta) = \sum_{i=1}^N y_i \log p(x_i; \beta) + \sum_{i=1}^N (1 - y_i) \log(1 - p(x_i; \beta))) - \frac{\lambda}{2} \sum_{j=1}^d \beta_j^2. \tag{15}$$

Since the sizes of the samples from two classes is greatly different, we modify the objective function as follows to reduce the impact of the imbalance

$$L'(\beta) = C_1 \sum_{i=1}^N y_i \log p(x_i; \beta) + C_2 \sum_{i=1}^N (1 - y_i) \log(1 - p(x_i; \beta)) - \frac{\lambda}{2} \sum_{j=1}^d \beta_j^2 \tag{16}$$

where $C_1 = 1/\sum_{i=1}^n y_i$ and $C_2 = 1/\sum_{i=1}^n (1 - y_i)$ are two weighting parameters. The Newton-Raphson algorithm [10] is used to find the optimal value of β which minimizes the objective function (16). The parameter λ is selected by cross-validation.

5   Experimental Results

Experiments are performed over the fingerprint databases provided by FVC2002. Our experiments check the discriminating ability of the similarity measure between minutiae pairs and evaluate the final matching performance.

5.1   Validation of Similarity Measure between Minutiae Pairs

In this experiment two kinds of minutiae pairs are selected to evaluate the performance of the similarity measure: corresponding minutiae pairs and non corresponding minutiae pairs. The minutiae pairs matched in the genuine matches are called corresponding minutiae pairs otherwise are called non-corresponding minutiae pairs. This experiment is conducted over FVC2002 DB1_B. Totally 6992 corresponding minutiae pairs are obtained by minutiae matching algorithm proposed by He et al [2]. The same number of non-corresponding minutiae pairs are randomly selected from the imposter matches. Four kinds of similarity degrees are calculated for all minutiae pairs. For each similarity degree false match rate (FMR) and false non-match rate (FNMR) at a fixed threshold t are recorded as FMR(t) and FNMR(t). By varying the threshold t from 0 to 1, a Receiver Operating Characteristics (ROC) curve is obtained. Fig.4 shows that the performance of two combined similarity measures is similar and better than other two. In the following experiment we use the product rule as the similarity measure for convenience.
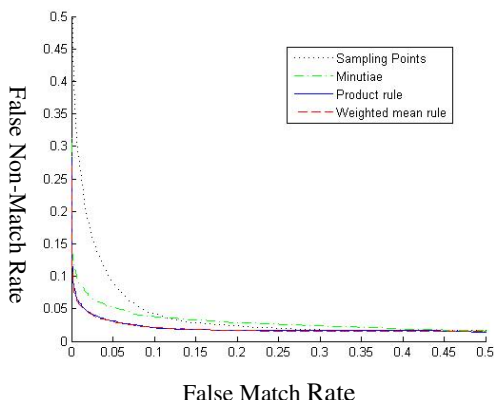
Fig. 4. The ROC curves of various similarity measures including neighboring minutiae based, neighboring sampling points based, combined by product rule and combined by weighted mean rule

5.2 Performance on FVC2002

The experiments reported in this section have been conducted on the four databases of FVC 2002. Database B is used to select penalized λ and train the parameter β while database A is used to validate the proposed algorithm. For a genuine match or imposter match from database B, totally N_α attempts are performed, and the feature vector corresponding to the attempt that maximizes $mRatio \cdot simi$ is selected as a training feature vector. We compare our result with that of algorithm PB15 and PB05 in Table 1. Both of PB15 and PB05 have participated in the competition of FVC2002, and got the 4th and 5th place ranked by the average equal error rate (AveEER), respectively. According to the ranking rule our algorithm is in the 5th place.

Table 1. Equal error rate on four databased of PB15, PB05 and ours

	DB1 (%)	DB2 (%)	DB3 (%)	DB4 (%)	Average (%)
PB15	0.63	1.03	0.81	0.61	0.77
PB05	0.52	0.69	1.48	0.98	0.92
ours	0.34	0.51	1.70	0.88	0.86

6 Conclusion

In this paper we represent the fingerprint by minutiae and ridge sampling points. This representation holds the uniqueness of the fingerprint and is compact and effective. In the light of this representation we propose a novel similarity measure method to capture neighboring minutiae pattern, ridge orientation and ridge frequency. After the alignment and minutiae pairing nine features are extracted to represent the matching status of the input and query fingerprint, and penalized logistic regression is adopted

to estimate the probability of the two fingerprints coming from a same finger. Experimental results on FVC2002 show that the proposed algorithm achieves good performance and ranks 5th place according to the average equal error rate.

References

[1] Tico, M., Kuosmanen, P.: Fingerprint Matching Using an Orientation-Based Minutia Descriptor. IEEE Trans. Pattern Anal. Mach. Intell. 25(8), 1009–1014 (2003)

[2] He, Y., Tian, J., Luo, X., Zhang, T.: Image Enhancement and Minutiae Matching in Fingerprint Verification. Pattern Recognition Letters 24(9), 1349–1360 (2003)

[3] He, Y., Tian, J., Li, L., Chen, H., Yang, X.: Fingerprint Matching Based on Global Comprehensive Similarity. IEEE Trans. Pattern Anal. Mach. Intell. 28(6), 850–862 (2006)

[4] Bazen, A.M., Gerez, S.H.: Fingerprint Matching by Thin-Plate Spline Modelling of Elastic Deformations. Pattern Recognition 36(8), 1859–1867 (2003)

[5] Feng, J.: Combining Minutiae Descriptors for Fingerprint Matching. Pattern Recognition 41(1), 342–352 (2008)

[6] Lin, H., Wan, Y., Jain, A.K.: Fingerprint Image Enhancement: Algorithm and Performance Evaluation. IEEE Trans. Pattern Anal. Mach. Intell. 20(8), 777–789 (1998)

[7] Luo, X., Tian, J.: Knowledge Based Fingerprint Image Enhancement. In: Proc. 15th ICPR, pp. 4783–4786 (2000)

[8] Graham, R.L.: An Efficient Algorithm for Determining The Convex Hull of a Finite Planar Set. Information Processing Letters 26, 132–133 (1972)

[9] Shen, L., Tan, E.T.: Dimension Reduction-Based Penalized Logistic Regression for Cancer Classification Using Microarray Data. IEEE/ACM Trans. On Computational Biology and Bioinformatics 2(2), 166–175 (2005)

[10] Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Springer, Heidelberg (2001)

[11] The 2nd fingerprint verification competition,
     http://bias.csr.unibo.it/fvc2002/

[12] Wilson, C.L., Watson, C.I., Paek, E.G.: Effect of Resolution and Image Quality on Combined Optical and Neural Network Fingerprint Matching. Pattern Recognition 33(2), 317–331 (2000)

[13] Feng, J., Ouyang, Z., Cai, A.: Fingerprint Matching Using Ridges. Pattern Recognition 39(11), 2131–2140 (2006)

[14] Gu, J., Zhou, J., Yang, C.: Fingerprint Recognition by Combining Global Structure and Local Cues. IEEE Trans. On Image Processing 15(7), 1952–1964 (2006)

[15] Jain, A.K., Hong, L., Bolle, R.: On-Line Fingerprint Verification. IEEE Trans.Pattern Anal. Mach. Intell. 19(4), 302–314 (1997)

[16] Borgefors, G.: Distance Transformations in Digital Images. Computer Vision, Graphics, and Image Processing 34(3), 344–371 (1986)