Volume 4, No. 2, Jan-Feb 2013

ISSN No. 0976-5697

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Bayes Classification for the Fingerprint Retrieval

Leelambika.K.V*
Computer Science and Engineering
Vel Tech High Tech Dr.Rangarajan Dr.Sakunthala Engineering College Chennai, India
kvleelambika@gmail.com

A.Rohini
Computer Science and Engineering
Vel Tech High Tech Dr.Rangarajan Dr.Sakunthala Engineering College Chennai, India
rohinaruna@gmail.com

M.Queen Mary Vidya
Computer Science and Engineering
Vel Tech High Tech Dr.Rangarajan Dr.Sakunthala Engineering College
Chennai, India
queenmaryvidya@gmail.com

Abstract: The Fingerprint is the most commonly used biometric property in security, commerce, industrial, civilian and forensic applications. The goal is to raise the recognition rate in the fingerprint retrieval system. In this work, the Bayes classifier is adopted assuming Gaussian statistics. The set of training samples are expanded by spatial modeling technique and implement a variant of the Fisher's Linear Discriminant Analysis (FLDA) for dimension reduction and Quadratic Discriminant Analysis (QDA) for lowering estimation errors. Finally calculating the probabilistic features for Gabor and Minutiae which helps to reduce the error rate about 75% which outperforms the K-NN classifier where the error rate was about 30-60%. The accuracy and Speed are evaluated using FVC2004 database and satisfactory retrieval performance is achieved. Thus the objective of the Fingerprint Retrieval system that is efficient and accurate is build.

Keywords: Fingerprint retrieval, Bayes classifier, Gaussian statistics, training samples, Fishers Linear Discriminant Analysis (FLDA), Quadratic Discriminant Analysis (QDA), minutiae, FVC database.

I. INTRODUCTION

The Fingerprints are most widely used biometric system for the authentication in many applications. Over the years, researchers have proposed different problem formulation to be tackled by their systems. Those systems can be clustered into three major categories namely, fingerprint classification, verification and indexing/retrieval. Fingerprint classification refers to the problem of assigning a given fingerprint into a predefined class based on its global structure and singular points. A fingerprint verification system authenticates a person's identity by comparing the captured fingerprint with his/her own previously enrolled reference template stored in the database. It compares one-to-one to confirm the images from database of identity by the individual is true. In fingerprint indexing/retrieval, the problem of one to one matching is extended to one-to-N matching without requiring the subject's claim of identity.

Given a fingerprint instance of unknown identity, the system searches through the entire database of enrolled templates and returns a list of probable fingers (identifiers of individual) that the fingerprint may belong. The fingerprint features chosen to train the system should be reliable enough to distinguish one finger from all the others in the database. To this aim, some reported works made use of the orientation field features while some explored the features on the complex filtered images. Minutiae features and other transformed features were also adopted.

II. RELATED WORK

One common problem in pattern recognition is the lack of samples in training a classifier. The curse of dimensionality often prohibits classifiers to be sufficiently trained, especially in high dimensional space [2]. Unfortunately, the number of samples per finger that is available for recognition is rather limited. Many publications about fingerprint retrieval resorted to K-nearest neighbor (K -NN) classification, in which was often set to one and the distance metric was often Euclidean. The Fingerprint Verification Competition (FVC) databases (2000, 2002, and 2004) contain only eight instances per finger in each set. Even worse, the National Institute of Standards and Technology database 4 (NIST-4) [3] contains only two instances per finger. This poses a big challenge to researchers trying to solve the problem of fingerprint retrieval or identification.

The problems of insufficient training samples by generating additional samples are being tackled, emulating the genuine samples that would have been captured by the fingerprint sensing device during fingerprint enrolment. In the existing system, more training samples are produced for the NIST-4 database by three kinds of spatial modeling, namely, translational modeling, rotational modeling and distorted sample generation. The distortion model employed has been constructed based on the publication plastic deformation of fingerprint images. The Euclidean 1-NN retrieval performance has been greatly improved with the help of these artificial samples. At the same time, the classification speed and required storage have been compromised.

The objective of this research is to build a fingerprint retrieval system that is accurate and efficient, with the goal to raise the recognition rate for the first few top rank candidates. This is achieved by extending the work in which additional training samples are generated by spatial modeling. With a lot more training samples, it becomes possible to train up a more sophisticated classifier instead of the 1-NN classifier. In this work, the Bayes classifier with the assumption of Gaussian statistics is adopted.

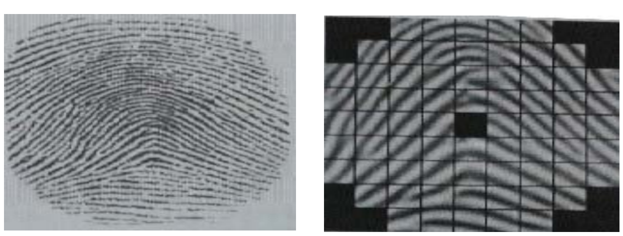
Extensive simulations with the FVC 2000, FVC 2002 and NIST-4 databases have confirmed the usefulness of the proposed approach. The first rank identification error has been reduced by about 30% to 60%, compared with 1-NN classification without additional samples. In most of the time, the system offers comparable performance among other published works with similar goal and database under test. The gain in computational efficiency of Bayes classification (assuming Gaussian statistics) is also significant when more samples are available.

III. BASELINE RETRIVAL SYSTEM

The proposed system is built on top of a baseline system. The baseline system is implemented according to the baseline descriptions in the recent work focusing on artificial sample generation, unless otherwise specified below. The baseline system is built up from a number of existing algorithms and techniques. It is composed of the following major modules

A. Feature extraction:

The modified Finger Code is employed as features in the experiments. The enhanced fingerprint image is filtered by Gabor filters in M(=12) directions to capture finer details of the ridge structures [5].The filtering orientations are (0, Π/12, 2Π/12,…..,11Π/12). Unlike the original Finger-Code, each feature cell is defined as a square sector of size *v*×*v* instead of the shape of radial sector.



Original Image

Enhanced, Normalized & Cropped Image

Figure. 1 Feature Extraction by Gabor Filter on an Image from FVC 2004

The weighted features are not be used as described in the previous work to attenuate features situating far away from the registration point, mainly because the FVC databases do not provide enough fingers for accurate estimation of the weights [7], [8]. Variable transformation that takes the power *PV* of each feature value is also not used. A graphical illustration for defining the blocks of interest for feature extraction is shown in Fig. 1 and an example set of Gabor-filtered images in 12 directions is given in Fig. 2.

B. Fingerprint Retrieval:

The Euclidean 1-NN classification approach in the baseline system implementation is approached. Given a feature vector X = $(x_1, x_2, \ldots\ldots x_d)^2$ for a test sample, a feature vector Y = $(y_1, y_2, \ldots\ldots y_d)^2$ for a sample in the database and the two binary vectors $Z_x = (Z_{x1}, Z_{x2}, \ldots\ldots Z_{xd})^2$ and $Z_y = (Z_{y1}, Z_{y2}, \ldots\ldots Z_{yd})^2$ defining the validity of the features (with value of 0 for background or missing feature, and 1 for foreground feature), the distance D_{xy} of the two fingerprints is computed by

$$D_{xy} = \sqrt{\frac{1}{\sum_{m=1}^{d} z_{xm}\, z_{ym}} \sum_{k=1}^{d} Z_{xk} Z_{yk}\, (Y_K - X_K)^{2}}$$

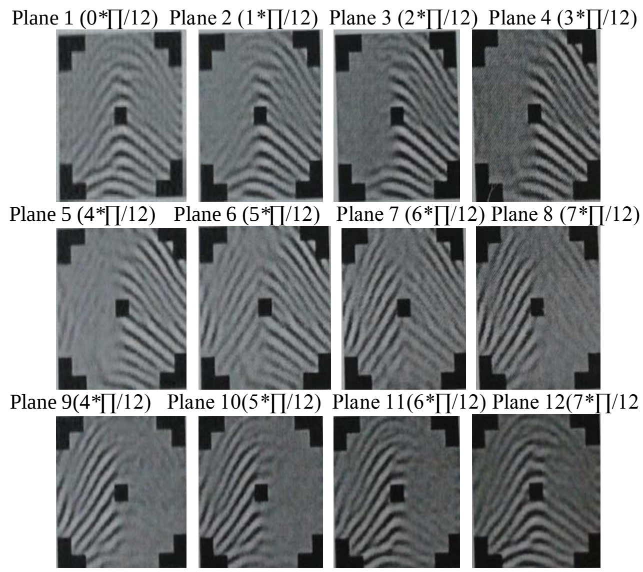


Figure. 2 Set of Gabor Filtered fingerprint images in 12 directions

C. Fingerprint Retrieval:

The Euclidean 1-NN classification approach in the baseline system implementation is approached. Given a feature vector X = $(x_1, x_2, \ldots\ldots x_d)^2$ for a test sample, a feature vector Y = $(y_1, y_2, \ldots\ldots y_d)^2$ for a sample in the database and the two binary vectors $Z_x = (Z_{x1}, Z_{x2}, \ldots\ldots Z_{xd})^2$ and $Z_y = (Z_{y1}, Z_{y2}, \ldots\ldots Z_{yd})^2$ defining the validity of the features (with value of 0 for background or missing feature, and 1 for foreground feature), the distance D_{xy} of the two fingerprints is computed by

$$D_{xy} = \sqrt{\frac{1}{\sum_{m=1}^{d} z_{xm}\, z_{ym}} \sum_{k=1}^{d} Z_{xk} Z_{yk}\, (Y_K - X_K)^{2}}$$

A commonly used retrieval method in pattern classification is the fixed order retrieval. Given a target penetration rate and a test image from an unknown finger, a list of probable fingers with the list length proportional to the penetration rate is produced based on the distance of the test sample to different fingers[9], [10]. It is noted that some other fingerprint retrieval methods with better nominal performance were proposed. In the current work, bayes classifier is used which calculates probability value given by,

$$P(Ci / X) = \frac{P\left(\frac{X}{C_i}\right) P(C_i)}{P(X)}$$

IV. BAYES CLASSIFICATION

Fingerprint retrieval systems mostly adopt the 1-NN classifier due to the fact that there are only two and eight samples per finger in the NIST-4 and FVC databases, respectively, which are far from sufficient for training up more sophisticated classifiers. Despite its simplicity, theN1-NN classifier is inferior to many classifiers such as the Bayes classifier. Cover has proven that the risk of 1-NN classification is bounded by twice the risk of the Bayes classifier and it approaches the performance of the Bayes classifier asymptotically only when there is an infinite number of training samples.

In this paper, the proposed Bayes classifier assuming Gaussian statistics and this requires the following issues to be addressed: shortage of training samples; curse of dimensionality leading to inaccurate statistical estimation and over fitting; and remedy of missing feature values [13].

The proposed fingerprint spatial modeling techniques allows generating several tenths of additional training samples from one genuine sample, directly solving the first issue partly. Regarding the second issue, it is realized that the original feature dimension is very high, ranging from a few hundred to a thousand. A variant of the Fisher's Linear Discriminant for dimension reduction and a variant of the quadratic discriminant function (Bayes decision function assuming Gaussian statistics) for lowering estimation errors are implemented. Instead of simply adopting the standard formulations, we regularize some of the estimated matrices. In the Fisher's Linear Discriminant, the within-class scatter matrix S_w is regularized by adding a constant hw^2 to the diagonal elements of the matrix, while keeping the between class scatter matrix S_B intact. In the Quadratic Discriminant function, the covariance matrix Σi is regularized in each class (the term "class" here means the "identity of a finger," but *not* Henry class) by using a class-independent constant h^2, for which the technique is termed modified Quadratic Discriminant function 1 (MQDF1).

Regularization in both Fisher's Linear Discriminant and Quadratic Discriminant function may be required with reasonable constants or else the identification and retrieval rates could drop. For optimal performance, these regularization constants have to be adjusted according to the number of training samples per finger and feature dimension.

The final issue to be resolved is the treatment of missing feature values. The segmentation algorithm defines the foreground and background areas of a fingerprint image and the background features should not be used in training a classifier. There also exist quite a lot of fingerprint images in the FVC 2000 and FVC 2002 FVC 2004 databases with their registration point locating very close to the edge of the image frame. In such cases, almost half of the feature values would be missing when the Gabor features around the registration point are extracted. The missing features can be ignored in probability-based bayes classification—only those features available on both the training and test sample will be included in calculating the probability.

In Bayes classification along with Fisher's linear discriminant, however, all features must be present in all training samples for the calculation of class statistics and feature projection, and in all test samples for feature projection and Bayes classification.

The mean imputation approach to get around this pitfall in the current attempt is taken. Given a set of training samples $\{Yi^1, Yi^2, \ldots Yi^q\}$ of class, with each feature vector specified by $Yi^{q0} = \{Y1^{q0}, Y2^{q0}, \ldots Yd^{q0}\}$ and validity of feature specified by $Zi^{q0} = \{Z1^{q0}, Z2^{q0}, \ldots Zd^{q0}\}$, lets construct a relaxed class mean vector m_i given by

$$m_i = (m_1^i, m_2^i, \ldots m_d^i)^T$$

$$= \left(\frac{1}{\sum_{t=1}^{q} z_1^t} \sum_{u=1}^{q} y_1^u z_1^u, \frac{1}{\sum_{t=1}^{q} z_2^t} \sum_{u=1}^{q} y_2^u z_2^u \ldots, \frac{1}{\sum_{t=1}^{q} z_d^t} \sum_{u=1}^{q} y_d^u z_d^u\right)^T$$

The training samples a global relaxed mean m vector over all m_{d0}^i is given by

$$m = (m1, m2, \ldots md) = (m_1^i, m_2^i, \ldots m_d^i)^T$$

In the case where $\sum_{t=1}^{q} z_{d0}^t = 0$, the mean value of the feature is set as (m^i_{d0}) to be m_{d0} which is the global mean feature value over all classes in the database. With the relaxed mean vector m_i, for all y_i^{q0} with zero(s) in with zero(s) in $(Z_1^{q0}, Z_2^{q0}, \ldots Z_d^{q0})^T$, the value of y_{d0}^{q0} for any feature with $Z_{d0}^{q0} = 0$ is substituted by the corresponding value of m_{d0}^i. By this substitution, we can apply on the training samples the Fisher's linear discriminant, feature projection and train the Bayes classifier as usual.

Given a test sample $X = (x_1, x_2, \ldots x_d)^T$ with validity specified by $Z = (Z_1, Z_2, \ldots Z_d)^T$, if any of the $Z'_{d0}\, s$ is zero, then the corresponding X_{d0} is assumed to be m_{d0}. This substitution is inevitable since we do not know what finger the unknown sample belongs to prior to classification It should be pointed out that substituting a mean class feature value to the missing feature on training samples would flatten the shape of the cluster for that class in the feature space. In particular, the variance along the dimension of missing feature, hence the related eigen values, would be smaller than what is expected when the "real" feature value is available. Other sophisticated imputation methods may be adopted to diminish such errors.

V. MINUTIAE EXTRACTION

Minutiae are major features of a fingering, using which comparisons of one print with another can be made. Minutiae include: Ridge and Bifurcation. The number and locations of the minutiae vary from finger to finger in any particular person, and from person to person for any particular finger [1] (for example, the thumb on the left hand). To extract the minutiae features first image enhancement process is carried out. The next step after enhancement of the image is the extraction of minutiae. The enhanced image is binarised first in this step. The skeleton of the image is then formed. The minutiae points are then extracted by the following method. The binary image is thinned as a result of which a ridge is only one pixel wide [6], [8]. Then crossing number algorithm is used to extract minutiae features. Crossing number algorithm is defined as half the sum of the differences between pairs of adjacent pixels defining the 8-neighborhood of 'p'. Mathematically in equation,

$$Cn(p)=\frac{1}{2}\sum_{i=1,2..8}|val(P_i\ mod 8)-val(P_i-1)|$$

Where p0 to p7 are the pixels belonging to an ordered sequence of pixels defining the 8-neighborhood of p and val(p) is the pixel value.

Crossing numbers 1 and 3 correspond to ridge endings and ridge bifurcations respectively. The minutiae obtained from this algorithm must be filtered to preserve only the true minutiae. The different types of false minutiae introduced during minutiae extraction include spike, bridge, break, Spur and Misclassified Border areas. The extracted minutiae feature is shown in fig.3.

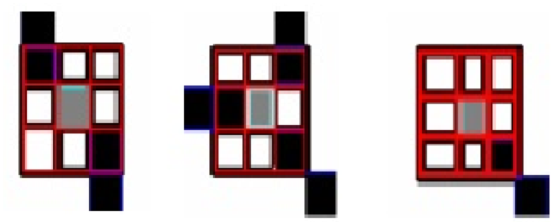


Figure. 3 cn(p)=2, cn(p)=3 and cn(p)=1 representing anon minutiae region, a bifurcation and a ridge ending

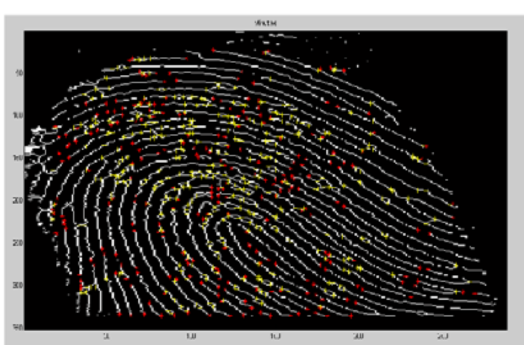


Figure. 4 The extracted minutiae feature

VI. EXPERIMENTAL RESULTS

The proposed method is made possible after undergoing Training and Testing process. Training is the process in which fingerprints are saved i.e., enrolment and in Testing process the fingerprint's are compared and retrieved. The new recognition flowchart with Fisher's Linear Discriminant and Bayes classification is drawn in fig.5.

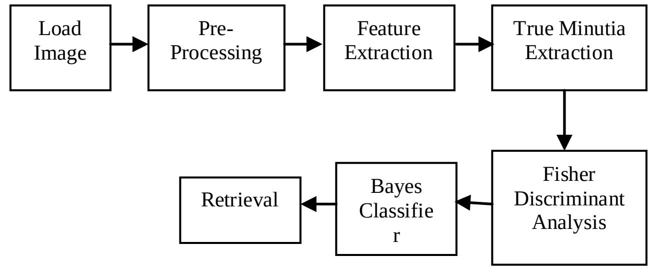


Figure. 5 Flow Processing

The modules used for training and testing are similar. They are composed of the following major modules:

i. Pre-processing
ii. Feature extraction
iii. Image retrieval
iv. Minutiae extraction
v. False minutiae removal
vi. Bayes classifier for retrieval.

Pre-processing consists of two main processes that are Point transformation and filtering. In point transformation, the pixel value distribution of an image is expanded so as to increase the perception information. The original histogram of a fingerprint image has the bimodal type, the histogram after the histogram equalization occupies all the range from 0 to 255 and the visualization effect is enhanced. In Filtering, the image is divided into small processing blocks (32x32 pixels) and perform the Fourier Transform, Thus the image is converted to frequency domain as well as filtered (i.e., noises are removed).

Feature extraction has three important processes in it. They are binarization, Gabor feature and thinning. In binarization process the 8-bit Gray fingerprint image is transformed to a 1-bit image with 0-value for ridges and 1-value for furrows.

Gabor feature extracts translational and rotational models. Thinning of fingerprint is done for making the terminations and bifurcations to be clearly visible Minutiae features are extracted by crossing number algorithm. In False minutiae removal the outside region of interest are eliminated as false minutiae through distance formulae.

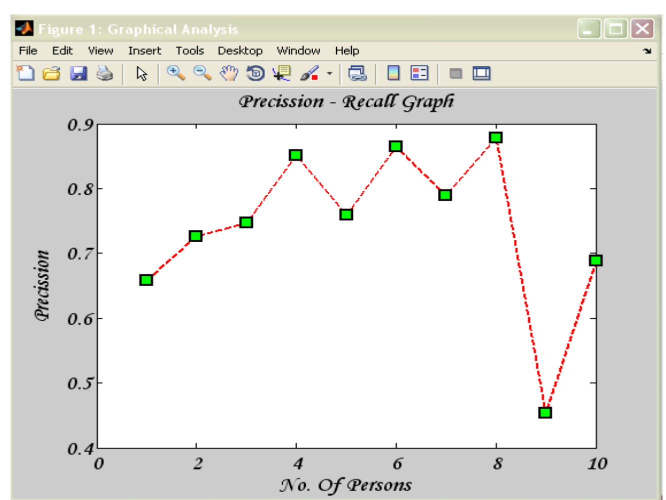


Figure. 6 Retrieval Performance on the FVC 2004 Database by Bayes Classification with different reduced dimensions

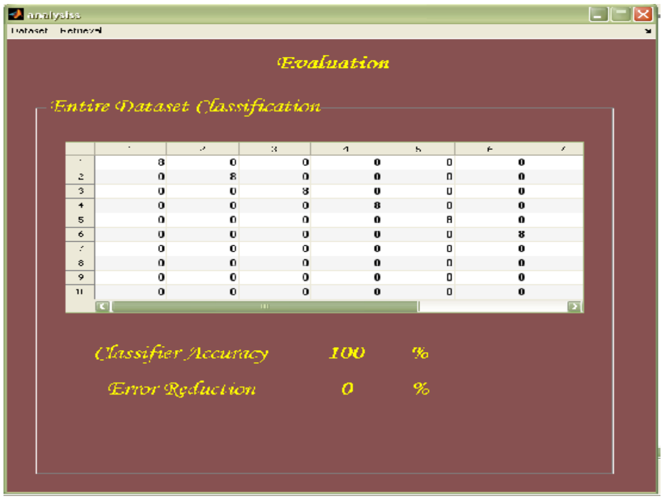


Figure. 7 Evaluation for entire data set showing classifier accuracy and error rate in %

Fisher's Linear Discriminant Analysis: The features extracted are projected in feature space using this algorithm. These features are subjected to QDA (Quadratic Discriminant Analysis) for satisfying the conditions of statistics. The result is set for training Bayes Classifier. In Bayes classifier the computed features are set to define their probability. Fig.6 shows the retrieval performance on FVC2004 database by Bayes classification. The graph is plotted between precision and number of persons. Fig.7 shows the evaluation result obtained for entire data set showing classifier accuracy as 100% and error reduction as 0%.

VII. CONCLUSION

The retrieval performance of the baseline system is satisfactory on the FVC 2004 databases. When compared with orientation field features and complex filtered features, it is evident that Gabor features give better results as far as fingerprint retrieval is concerned. The proposed spatial modeling techniques have made the retrieval process more reliable. The problem of speed is tackled by proposing the use of Fisher's Linear Discriminant and Bayes Classification with the methods for regularization and handling of missing features. Moreover, Bayes classification with any number of training samples per fingerprint runs faster than, 1-NN classification with eight training samples per finger. The execution time of 1-NN classification increases linearly with the number of training samples per finger, but remains the same for Bayes classification. As the speed of Bayes classification does not depend on the number of training samples per fingerprint, one can capture many instances of the same fingerprint for building up a large but comprehensive database without the fear of lowering the retrieval speed and accuracy.

The experimental results also show that regularizing the covariance matrices in the Fisher's Linear Discriminant and Bayes Classifier is a useful strategy. Taking half of the samples for training and the remaining half for testing, Bayes classification has helped to reduce the first rank error rates about 75%. It is probably an excellent choice only when more training samples are available. The proposed techniques can be of practical uses in security, civilian, and forensic applications. Moreover, when a large number of features are missing, the performance of the probabilistic classifier remains the same.

VIII. REFERENCES

[1]. X. Liang, A. Bishnu, and T. Asano, "A robust fingerprint indexing scheme using minutia neighbourhood structure and low-order Delaunay triangles," IEEE Trans. Inf. Forensics Security, vol. 2, no. 4, pp. 721–733, Dec. 2007.

[2]. A. Senior and R. Bolle, "Improved fingerprint matching by distortion removal," IEICE Trans. Inf. Systems E84-D, no. 7, pp. 825–832, 2001.

[3]. X. Jiang, M. Liu, and A. C. Kot, "Fingerprint retrieval for identification," IEEE Trans. Inf. Forensics Security, vol. 1, no. 4, pp. 532–542, Dec. 2006.

[4]. D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain, "FVC2004: Third fingerprint verification competition," in Proc. 17th Int. Conf. Biometric Authentication, Cambridge, U.K., 2004, pp. 1–7.

[5]. B. Bhanu and X. Tan, "Fingerprint indexing based on novel features of minutiae triplets," IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 5, pp. 616–622, May 2003.

[6]. A. K. Jain and L. Hong, "A multichannel approach to fingerprint classification,"IEEE Trans. Pattern Anal. Mach. Intell., vol. 21, no. 4, pp.348–359, Apr. 1999.

[7]. A. K. Jain, S. Prabhakar, L. Hong, and S.Pankanti, "Filter bank-based fingerprint matching," IEEE Trans. Image Process., vol. 9, no. 5, pp. 846–859, May 2000.

[8]. R. Cappelli, D. Maio, D. Maltoni, and L. Nanni, "A two-stage fingerprint classification system," in Proc. 2003 ACM SIGMM Multimedia Biometrics Methods and Applications Workshop, Berkeley, CA, 2003, pp. 95–99.

[9]. M. Liu, X. Jiang, and A. C. Kot, "Fingerprint retrieval by complex filter responses," in Proc. 18th Int. Conf. Pattern Recognition, Hong Kong, 2006, pp. 1042–1045.

[10]. A. Gyaourova and A. Ross, "A novel coding scheme for indexing fingerprint patterns," Lecture Notes Comput. Sci., vol. 5342, pp. 755–764, 2008.

[11]. X. Shuai, C. Zhang, and P. Hao, "Fingerprint indexing based on composite set of reduced SIFT features," in Proc. 18th Int. Conf. Pattern Recognition, Florida, 2008, pp. 1–4.

[12]. K. Nilsson and J. Bigun, "Localization of corresponding points in fingerprints by complex filtering," Pattern Recognit. Lett., vol. 24, no. 13, pp. 2135–2144, 2003.

[13]. F. Kimura, K. Takashina, S. Tsuruoka, and Y.Miyake, "Modified quadratic discriminant functions and the application to Chinese character recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 9, no. 1, pp. 149–153, Jan. 1987.