A New Indexing Method for Biometric Databases Using Match Scores and Decision Level Fusion

Ilaiah Kavati[1,2], Munaga V.N.K. Prasad[2], and Chakravarthy Bhagvati[1]

[1] University of Hyderabad, Hyderabad-46, India
[2] Institute for Development and Research in Banking Technology, Hyderabad-57, India
kavati089@gmail.com, mvnkprasad@idrbt.ac.in, chakcs@uohyd.ernet.in

Abstract. This paper proposes a new clustering-based indexing technique for large biometric databases. We compute a fixed length index code for each biometric image in the database by computing its similarity against a preselected set of sample images. An efficient clustering algorithm is applied on the database and the representative of each cluster is selected for the sample set. Further, the indices of all individuals are stored in an index table. During retrieval, we calculate the similarity between query image and each of the cluster representative (*i.e.*, query index code) and select the clusters that have similarities to the query image as candidate identities. Further, the candidate identities are also retrieved based on the similarity between index of query image and those of the identities in the index table using voting scheme. Finally, we fuse the candidate identities from clusters as well as index table using decision level fusion. The technique has been tested on benchmark PolyU palm print database consist of 7,752 images and the results show a better performance in terms of response time and search speed compared to the state of art indexing methods.

Keywords: Palm print, Indexing, Clustering, Sample images, Match scores, Decision level fusion.

1 Introduction

Biometric identification refers to automated method of identifying individuals based on their physiological and/or behavioral characteristics. However, in biometric identification systems, the identity corresponding to a query image is typically determined by comparing it against all images in the database [1]. This exhaustive matching process increases the response time and the number of false positives of the system. Therefore, an efficient retrieval technique is required to increase the search speed and reduce the response time of the system. The retrieval technique should be such that, instead of comparing the query image with every image in the database it has to retrieve a small set of images from the database to which the actual matching process is applied. This retrieval can be done by, a) partitioning the database into groups of similar images in order to facilitate and accelerate the search process, b) indexing methods.

M.K. Kundu et al. (eds.), *Advanced Computing, Networking and Informatics - Volume 1*, Smart Innovation, Systems and Technologies 27,
DOI: 10.1007/978-3-319-07353-8_57,

The partitioning of the database into groups can be done by either classification [3,4] or clustering [9]. Authors in [3], [4] partitions the database into predefined groups or classes. The class of the query identity is first calculated and compared only with the entries present in the respective class during the search process. However, the classification methods suffers with, uneven distribution of images in the predefined classes and image rejection [8]. On the other hand, clustering method organizes the database into natural groups (clusters) in the feature space such that images in the same cluster are similar to each other and share certain properties; whereas images in different clusters are dissimilar [9]. Clustering method does not use category labels that tag objects with prior identifiers, *i.e.,* class labels.

The other approach to reduce number of matching for identification is indexing. In traditional databases, records are indexed in an alphabetical or numeric order for efficient retrieval. But biometric data do not have any natural sorting order to arrange the records [9]. Hence, traditional indexing approaches are not suitable for biometric databases. Biometric indexing techniques are broadly categorized into, a) point [5], [6], b) triplet of points [11], [12], [13], and c) match score [14], [15], [16] based approaches. In [5], [6], authors extracted the key feature points of the biometric images and mapped them into a hash table using geometric hashing. Authors in [11], [12], [13], computed the triplets of the key feature points and mapped them into a hash table using some additional information. However, the limitation of these indexing methods is that all of them deal with variable length feature sets which make the identification system statistically unreliable.

In recent years, indexing techniques based on fixed length match scores also investigated for biometric identification. Maeda et al [14], computes a match score vector for each image by comparing it against all the database images and stored these vectors permanently as a matrix. Though, the approach achieves quicker response time, it takes linear time in worst case and also storing of match score matrix leads to increase in the space complexity. Gyaourova et al. [15] improved the work on match scores by choosing a small set of sample images from the database. For every image in the database a match score vector (index code) was computed by matching it against the sample set using a matcher and stored this match score vector as a row in an index table. However, a sequential search is done in the index space for identification of best matches which takes linear time and is prohibitive for a database containing millions of images. Further, authors in [16],used Vector Approximation (VA+) file to store the match score vectors and *k-NN* search, palm print texture to retrieve best matches. However, the performance of VA+ file method generally degrades as dimensionality increases [17]. To address these problems, this paper proposes an efficient clustering-based indexing technique using match scores. We compute a fixed-length index code for each input image based on match scores. Further, we propose an efficient storage and retrieval mechanism using these indices.

Rest of the paper is organized as follows. The proposed indexing technique has been discussed in Section 2. Section 3 describes the proposed retrieval technique. Section 4 presents the experimental results and performance of the proposed system against other indexing methods in the literature. Conclusions are given in Section 5.

2 Proposed Indexing Methodology

This section discusses our proposed methodology for indexing the biometric databases. Let $S = \{s_1, s_2, ..., s_k\}$ be the sample image set, and $M_x = \{m(x,s_1), m(x,s_2), \ldots m(x,s_k)\}$ be the set of match scores obtained for an input image x against each sample image in S. We describe M_x as the index code of image x i.e., the index code of an image is the set of its match scores against the sample set. The match score between two images is computed by comparing their key features in Euclidean space. The match scores obtained are usually in the range 0-100.

Further, we store the index code of each individual in a 2D Index table A Fig. 1. Each column of the table corresponds to one sample image in the sample set. If image x has a match score value $m(x,s_i)$ with sample image s_i its identity (say x) is put in location $A(m(x,s_i), s_i)$. It can be seen from Fig. 1 that, each entry of the table $A(m(x,s_i), s_i)$ contains a list of image identities (*i.e.*, *IidList*) from the database whose match score is $m(x,s_i)$ against sample image s_i.

The motivation behind this concept is that, images that belong to a same user will have approximately similar match scores against a third image (say sample image s_i). Let q be a query image, we can thus determine all similar images from this index table by computing its match score against the sample image and selecting all images (*i.e.*,*IidList*) that have approximately similar match score against the sample image.

Sample image / Match score	s_1	....	s_i	....	s_k
0	*IidList*	....	*IidList*	....	*IidList*
1	*IidList*		*IidList*		*IidList*
.	.		.		.
.	.		.		.
100	*IidList*		*IidList*		*IidList*

Fig. 1. Index table organization

2.1 Selection of Sample Image Set

The selection of sample images from the database plays a crucial role in the performance of the system. Images which are more generic (*i.e.,* very different from one another) and represent the qualities of the entire database should be selected for sample set. In this paper, we use an efficient dynamic clustering algorithm (*i.e.,*leader algorithm) for the selection of sample image set.

Leader clustering algorithm [7] makes only a single pass through the database and finds a set of leaders as the cluster representatives (which we call sample images). In this work, we use the match score between the images to determine the cluster similarity. The motivation of using match score as similarity measure is that, usually similar images will have almost same features and so their match score is high *i.e.,* images in the same cluster will have high match score between them. Leader clustering algorithm uses a user specified similarity threshold and one of the image as the starting leader. At any step, the algorithm assigns the current image to the most similar cluster (leader) or the image itself may get added as a leader if its match score similarity with the current set of leaders does not qualify based on the user specified threshold. Finally, each cluster will have a listing of similar biometric identities and is represented with an image called leader. The found set of leaders acts as sample set of the database. The major advantage of dynamic clustering (such as leader algorithm) is that, new enrollments can be done with a single database scan and without affecting the existing clusters which is useful for clustering and indexing large databases.

3 Retrieval of Best Matches (Identification)

This section proposes an efficient retrieval system to identify a query image. Fig. 2 shows the proposed method of identification. When a query image is presented to the identification system, the technique retrieves the candidate identities from the clusters as well as from the 2D Index table which are similar to the query. Finally, the proposed system fuses the candidate identities (evidences) of both strategies to achieve better performance.

Although there are other strategies like multi-biometrics [18], [19] (such as multi-sensor, multi-algorithm, multi-sample, etc.) to retrieve multiple evidences for personal identification, we want to make a full use of the intermediate results in the process of computing index code in order to reduce the computational cost. It is easy to see from the Fig. 2 that, when computing the index code for a query image to identify the possible matches (Candidate list2) from the index table, we can get set of match scores against cluster representatives. Using them, we can also retrieve the candidate identities (Candidate list1) as additional evidence from the selected clusters whose representative match score greater than a threshold.

Let $G = \{g_1, g_2,..., g_k\}$ be the set of clusters, and $M_q = \{m(q,s_1), m(q,s_2), \ldots m(q,s_k)\}$ be the index code of a query image q where $m(q,s_i)$ be the match score value of q against sample image s_i. The retrieval algorithm use $(m(q,s_i), s_i)$ as index to the Index table and retrieve all the images (*IidList*) found in that location as similar images to the query into a Temporary list. In other words we retrieve all the images from the index table whose match score value against the sample image is equal to the query image. We also retrieve images from the predefined neighborhood of the selected location into the Temporary list. Finally, we give a vote to each retrieved image. Further, we also retrieve images from cluster g_i as similar images to the query, if $m(q,s_i) \geq$ similarity threshold *i.e.*, clusters are selected whose representative is similar to the query image. We store the retrieved cluster images into Candidate list 1. We repeat this process for each match score value of the query index code. In our next step, we accumulate and count the number of votes of each name in Temporary list. Finally, we sort the all the individuals in descending order based on the number of votes received and select the individuals whose vote score greater than a predefined threshold into Candidate list 2.

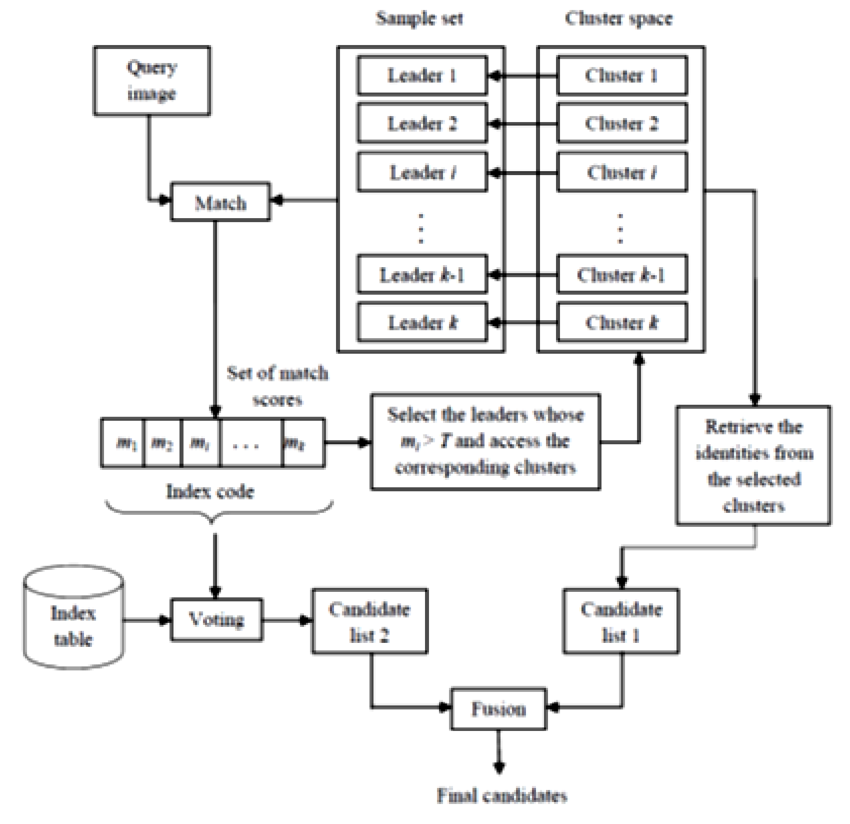


Fig. 2. Block diagram of the proposed identification system

3.1 Fusion of Decisions Output

The performance of uni-modal biometric systems may suffer due to issues such as limited population coverage, less accuracy, noisy data and matcher limitations [18]. To overcome the limitations of uni-modal biometrics and improve the performance, fusion of multiple pieces of biometric information has been proposed. Fusion can be performed at different levels such as data, feature, match score and decision level. In this paper, we use decision level fusion method. With this method, the decisions output (candidate identities) obtained from the cluster space and Index table are combined using, a) union of candidate lists, b) intersection of candidate lists.

The union fusion scheme combines the candidate list of the individual techniques. This fusion scheme has the potential to increase the chance of finding correct identity even if the correct identity is not retrieved by some of the techniques *i.e.*, the poor retrieval performance of one technique will not affect the overall performance. However, this scheme often increases the search space of the database. With intersection fusion scheme, the final decision output is the intersection of the candidate lists of the individual techniques. This type of fusion can further reduce the size of the search space. However, the poor retrieval performance of one technique will affect the overall performance of the system.

4 Experimental Results

We experimented our approach on benchmark PolyU palm print database [10] consist of 7,752 gray-scale images, approximately 20 prints each of 386 different palms. The first ten images of these prints is used to construct the database, while the other ten images are used to test the indexing performance. We segment the palm images to 151×151 pixels and use the Scale Invariant Feature Transform (SIFT features [2]) to compute the match score between images. Samples of some segmented images from PolyU database is shown in Fig. 3.We evaluated the performance of the system with different sample set sizes and chosen the optimum value as 1/3rd of the database.

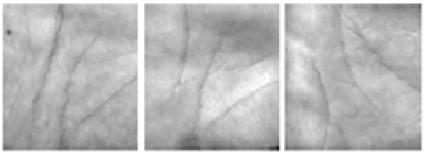


Fig. 3. Samples of segmented palm print images of PolyU database

The performance of the proposed technique is evaluated using two measures, namely Hit Rate (HR), Penetration Rate (PR); where HR is defined as the percentage of test images for which the corresponding genuine match is present in the candidate list and PR is the average search space in the database to identify a test image (*i.e.*, average candidate list size). The performance curves plotting the HR against PR at various thresholds are shown in Fig. 4(a). It is observed that the proposed fusion techniques performs well (as their PR is very less) compared to individual techniques. Further it can be seen that, the union fusion performs well to the intersection fusion. Table 1 shows the performance of the proposed techniques for PolyU database.

Table 1. PR (where HR=100%) of our proposed techniques for PolyU database

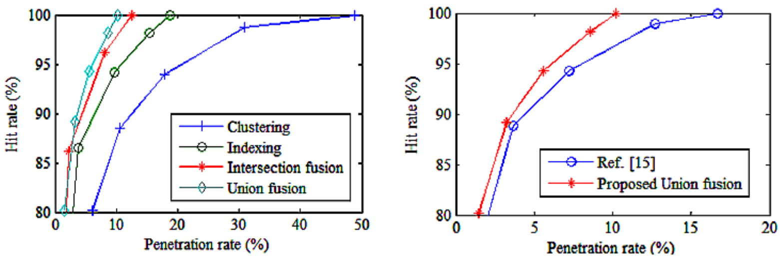
Clustering	Indexing	Intersection fusion	Union fusion
48.7%	18.8%	12.4%	10.2%

4.1 Retrieval Time

We analyze the retrieval time of our algorithm with big-O notation. Let q be the query image, k be the number of sample images chosen, and N be the number of enrolled user in the database. To retrieve the best matches for a query image, our algorithm computes the match score of query against the each sample image and retrieves the images, a. from the index table whose match scores against that sample image are nearer to query, b. as well as from respective cluster of the sample image if its match score is greater than similarity threshold. This process takes O(1) time. However, there are k sample images, so the time complexity our algorithm is O(k). On the hand linear search methods requires O(N). Thus our approach takes less time than the linear search approach as k<<N.

4.2 Comparison with Other Indexing Techniques

The proposed technique has been compared with existing match score base indexing techniques [15] [16]. The performance of proposed technique against technique in [15] can be seen from the Fig. 4(b). It is seen that our algorithm performs with less PR compared to the [15]. Further, authors in [15] performed linear search over the index space to retrieve the best matches. This process takes considerable amount of time i.e. O(N). Finally, the system in [16] achieves only a maximum of 98.28% HR for PolyU database. It can be inferred that the proposed system enhanced the indexing performance.



(a) (b)

Fig. 4. HRVs PR of different techniques on PolyU database, (a) Our proposed methods, (b) Comparison with Ref [15]

5 Conclusions

In this paper, we propose a new clustering based indexing technique for identification in large biometric databases. We compute a fixed length index code for each biometric image using the sample images. Further, we propose an efficient storing and searching method for the biometric database using these index codes. We efficiently used the intermediate results in the process of computing index code that retrieve multiple evidences which improves the identification performance without increasing computational cost. Finally, the results shows the efficacy of our approach against state of art indexing methods. Our technique is easy to implement and can be applied to any large biometric database.

References

1. Jain, A.K., Pankanti, S.: Automated fingerprint identification and imaging systems. In: Advances in Fingerprint Technology, 2nd edn. Elsevier Science (2001)
2. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal on Computer Vision 60, 91–110 (2004)
3. Henry Classification System International Biometric Group (2003), http://www.biometricgroup.com/HenryFingerprintClassification.pdf
4. Wu, X., Zhang, D., Wang, K., Huang, B.: Palmprint classification using principal lines. Pattern Recognition 37, 1987–1998 (2004)
5. Boro, R., Roy, S.D.: Fast and Robust Projective Matching for Finger prints using Geometric Hashing. In: Indian Conference on Computer Vision, Graphics and Image Processing, pp. 681–686 (2004)
6. Mehrotra, H., Majhi, B., Gupta, P.: Robust iris indexing scheme using geometric hashing of SIFT keypoints. Journal of Network and Computer Applications 33, 300–313 (2010)
7. Jain, A.K., Murty, M.N., Flynn, P.J.: Data Clustering: A Review. ACM Computing Surveys 31(3) (1999)
8. Maltoni, D., Maio, D., Jain, A.K., Prabhakar, S.: Handbook of Fingerprint Recognition. Springer (2003)
9. Mhatre, A., Palla, S., Chikkerur, S., Govindaraju, V.: Efficient search and retrieval in biometric databases. Biometric Technology for Human Identification II 5779, 265–273 (2005)
10. The PolyU palmprint database, http://www.comp.polyu.edu.hk/biometrics
11. Bhanu, B., Tan, X.: Fingerprint indexing based on novel features of minutiae triplets. IEEE Transactions on Pattern Analysis and Machine Intelligence 25, 616–622 (2003)
12. Kavati, I., Prasad, M.V.N.K., Bhagvati, C.: Vein Pattern Indexing Using Texture and Hierarchical Decomposition of Delaunay Triangulation. In: Thampi, S.M., Atrey, P.K., Fan, C.-I., Perez, G.M. (eds.) SSCC 2013. CCIS, vol. 377, pp. 213–222. Springer, Heidelberg (2013)
13. Jayaraman, U., Prakash, S., Gupta, P.: Use of geometric features of principal components for indexing a biometric database. Mathematical and Computer Modelling 58, 147–164 (2013)
14. Maeda, T., Matsushita, M., Sasakawa, K.: Identification algorithm using a matching score matrix. IEICE Transactions in Information and Systems 84, 819–824 (2001)
15. Gyaourova, A., Ross, A.: Index Codes for Multi biometric Pattern retrieval. IEEE Transactions on Information Forensics and Security 7, 518–529 (2012)
16. Paliwal, A., Jayaraman, U., Gupta, P.: A score based indexing scheme for palmprint databases. In: International Conference on Image Processing, pp. 2377–2380 (2010)
17. Weber, R., Schek, H., Blott, S.: A quantative analysis and performance study for similarity search methods in high-dimensional spaces. In: Proceedings of the 24th Very Large Database Conference, pp. 194–205 (1998)
18. Ross, A., Nandakumar, K., Jain, A.K.: Handbook of Multibiometrics. Springer, New York (2006)
19. Kumar, A., Shekhar, S.: Personal identification using multibiometrics rank-level fusion. IEEE Transactions on Systems, Man and Cybernetics 41, 743–752 (2011)