De-Duplication Complexity of Fingerprint Data in Large-Scale Applications

Nalla Pattabhi Ramaiah and C. Krishna Mohan

Abstract—De-duplication using biometrics has gained much attention from research communities as it provides a unique identity for each and every individual among the large population. De-duplication is the process of removing the instances of multiple enrollments by the same person using the person's biometric data. An important issue in the large-scale de-duplication applications is the speed of matching and the accuracy of the matching because the number of persons to be enrolled runs into millions. This paper presents an efficient method to improve the accuracy of fingerprint de-duplication in de-centralized manner. De-duplication accuracy decreases because of the noise present in the data, which would cause improper slap fingerprint segmentation. In this paper, an attempt is made to remove the noise present in the data by using binarization of slap fingerprint images and region labeling of desired regions with 8-adjacency neighborhood. The distinct feature of this technique is to remove the noise present in the data for an accurate slap fingerprint segmentation and improve the de-duplication accuracy. Experimental results demonstrate that the fingerprint segmentation rate and de-duplication accuracy are improved significantly.

Index Terms—8-adjacency neighborhood, de-duplication, fingerprints, slap fingerprint segmentation.

1. Introduction

Government of India provides different services and welfare schemes for the benefit of the people in the society. Some of these services include issuance of birth certificate, voter identity card, driving license, passport, etc. Also, welfare schemes include the targeted public distribution system (TPDS), the national rural employment guarantee system (NREGS), health insurance, old age pensions, etc. for the economic and social upliftment of the people. A unique identity (UID) number assigned for every citizen would obviate the need for a person to produce multiple documentary proofs of his/her identity for availing any government service, or private services like opening of a bank account. The UID number would remain a permanent identifier right from the birth to death of every citizen in the country. The UID would enable government to ensure that benefits under various welfare programs reach the intended beneficiaries, prevent cornering of benefits by a few sections of people, and minimize frauds. UIDs are also expected to be of help in law and order enforcement, effective implementation of the public distribution system, define social welfare entitlements and financial inclusion, and improve overall efficiency of the government administration. The biometrics plays a key role in providing unique identity of a person. The one among the most popular biometrics is fingerprints. The fingerprints are more accurate and reliable biometrics in civilian applications.

A fingerprint[1] consists of ridges and valleys on the surface of the finger. The uniqueness of a fingerprint can be determined by the minutiae points. Minutiae points are the local ridge features which are identified by a ridge bifurcation or a ridge ending. One way of acquiring fingerprints is to capture the slap fingerprint image. Slap fingerprints are taken by pressing four fingers simultaneously onto a slap fingerprint scanner. In general, the slap fingerprint image will be captured in the manner of 4-4-2 fingers, which means capturing left four fingers at one time, followed by all right four fingers and then followed by two thumb fingers. The captured slap fingerprint images are then processed for the segmentation[2] of fingerprints into the individual fingers.

The methods described in [3]–[9] used various filtering techniques to enhance the significant details of single-fingerprint images. Fingerprint segmentation using block-wise grey-level variances or local histograms of ridge orientations was described in [10]. In [11], Gabor filters were used to divide fingerprint into foreground and background regions. Edge detection and convex hull calculation were used in [12] for segmenting the image into different disconnected regions. In 2004, NIST (National Institute of Standards and Technology) organized a contest called Slap Fingerprint Evaluation 2004 (SlapSeg04)[13], in which thirteen segmentation algorithms were evaluated. One algorithm which is accessible publicly is the NIST fingerprint segmentation algorithm (NFSEG)[14]. In 2008,

NIST again conducted the contest called Slap Fingerprint Segmentation Evaluation II (SlapSegII). The difference between the two contests is the metrics used for successful slap fingerprint segmentation. SlapSeg04[15] used the fingerprint matching algorithm to determine the accuracy of slap fingerprint segmentation. SlapSegII used the ground truth data which has hand marked segmentation boxes as baseline for using NFSEG algorithm.

This paper presents the de-duplication process of fingerprint images. The targeted public distribution system (TPDS) is an Indian government welfare scheme for ensuring access and availability of food grains and other essential commodities at subsidized prices to the households. Identification of eligible beneficiaries and ensuring delivery commodities to them effectively and efficiently are the basic challenges for TPDS. The main objective of the TPDS is to find the duplicates by de-duplicating the fingerprints using fingerprint matching which requires the individual fingers. These individual fingerprint images are submitted for the de-duplication process which will eliminate the duplicates using fingerprint matching.

The slap fingerprint images have some noisy data due to some external factors which affect the calibration process of the fingerprint device. In the process of slap fingerprint segmentation, some of the total slap fingerprint images are improperly segmented because of noise present in the data. Moreover, the noise present in the slap fingerprint images are segmented as individual fingers instead of splitting the actual finger. In this paper, an efficient fingerprint de-noising algorithm is proposed to remove the noise present in the images for accurate slap fingerprint segmentation and to improve the de-duplication accuracy. The fingerprint de-duplication and NIST fingerprint image quality (NFIQ) scores are used as the baseline for determining the successful slap fingerprint segmentation. NFIQ score ranges on the scale 1 to 5, where lower quality score represents good quality and higher quality score represents poor quality. Fig. 1 illustrates the noisy fingerprint image.

This paper is organized as follows. In Section 2, the de-duplication process and the enrollments using centralized manner and decentralized manner are explained. In Section 3, various steps involved in the slap fingerprint noise removal algorithm are discussed. Experimental results are given in Section 4. Conclusions are explained in Section 5.

Fig. 1. Noisy slap fingerprint image.

Table 1: De-duplication complexity in a centralized manner

Scenario	Number of matches (millions)	Time taken (seconds)
If 1 finger (say left thumb) is matched against all left thumbs of previously enrolled persons	200	40
If all 10 fingers are matched against the respective fingers of all the previously enrolled persons	2000	400
If all the 10 fingers are matched against all the fingers of all the previously enrolled persons	20000	4000

2. De-Duplication

De-duplication is the process of removing the instances of multiple enrollments by the same person using the person's biometric data. During the de-duplication process, matching the biometrics of a citizen is done against the biometrics of other citizens to ensure that the same person is not enrolled more than once. This will ensure that each person will have a unique identity. The de-duplication complexity is demonstrated by using two different enrollment scenarios, i.e., enrollment using a centralized manner and enrollment using a decentralized manner.

2.1 Enrollment Using a Centralized Manner

In the case of enrollment using a centralized manner, the fingerprints of the citizen have to be matched against the fingerprints of all the previously enrolled citizens. The matching has to be done soon after the fingerprints are captured to check whether the same citizen has been enrolled earlier. In case a match is found, the citizen will not be enrolled into the system.

To illustrate the de-duplication complexity in the centralized manner, let us consider an example where 200 million citizens have already been enrolled, and a new citizen is now waiting to be enrolled into the system at the enrollment station. Also it is assumed that there are 10 blade servers with a total matching capacity of 5 million per second. The number of matches to be performed across different fingerprints and the time taken for the matching process is shown in Table 1.

2.2 Enrollment Using a Decentralized Manner

In the case of enrollment using a decentralized manner, the biometrics of citizens captured during a certain period have to be matched against the unique identity enrollment database of all the previously enrolled citizens. The matching has to be done by aggregating the data from each of the decentralized enrollment stations and matching against the de-duplicated biometrics of all the previously enrolled citizens. To illustrate the de-duplication complexity in a decentralized manner, we consider the case that 200 million citizens have already been enrolled, and data of 1 million citizens has been aggregated from the enrollment stations. The data of the 1 million citizens will have to be matched against the 200 million citizens to avoid multiple enrollments. With the assumption to assess the de-duplication complexity, we assume 10 blade servers with

Table 2: De-duplication complexity in a de-centralized manner

Scenario	Number of matches (trillions)	Time taken (days)
If 1 finger (say left thumb) is matched against all left thumbs of previously enrolled persons	200	463
If all 10 fingers are matched against the respective fingers of all the previously enrolled persons	2000	4630
If all the 10 fingers are matched against all the fingers of all the previously enrolled persons	20000	46300

a total matching capacity of 5 million per second. The number of fingerprint matches across different fingerprints to be performed and the time taken for matching the fingerprints are shown in Table 2.

3.　Noise Removal Method for Slap Fingerprint Image Segmentation

The noise removal method uses the noisy slap images. Fig. 1 represents the sample noisy four-finger slap image with dimensions of 500 dpi (dots per inch) resolution and 1600×1500 size. The steps involved in the noise removal process[16] includes the binarization of slap image, foreground and background segmentation of slap image, resampling and region labeling of slap image, and finally reconstruction of the original data for the larger labeled regions of the slap image

4.　Experimental Results

The database consists of 1.8 million slap fingerprint images. Each slap image has the dimensions of 1600×1500 size and 500 dpi resolution. It is observed that the correct segmentation rate before the noise removal process is 78%, and after the noise removal process, the correct segment-ation rate is 89% in Phase-I and 99% in Phase-II. These results are presented in Table 3.

The images shown in Fig. 2 (a), (b), (c), (d), (e), and (f), and Fig. 3 (a), (b), (c), (d), and (e) belong to Slap-Group-1 and It is observed that in the Slap-Group-1, Fig. 2 (a) represents the slap with high noise. Fig. 2 (b), (c), (d), and (e) are the segmented fingers of the slap with high noise, which has the NIST fingerprint image quality (NFIQ) scores 3, 5, 4, and 5, respectively. NFIQ score ranges on the scale 1 to 5, where lower quality score represents good quality and higher quality score represents poor quality. The resultant image after binarization as well as foreground and background segmentation of the slap fingerprint image is shown in Fig. 2 (f). Fig. 3 (a) shows the noise-free fingerprint image with recovered original data, and Fig. 3 (b), (c), (d), and (e) are the corresponding segmented fingers with NFIQ scores 1, 1, 1, and 3, respectively.

Table 3: Segmentation statistics

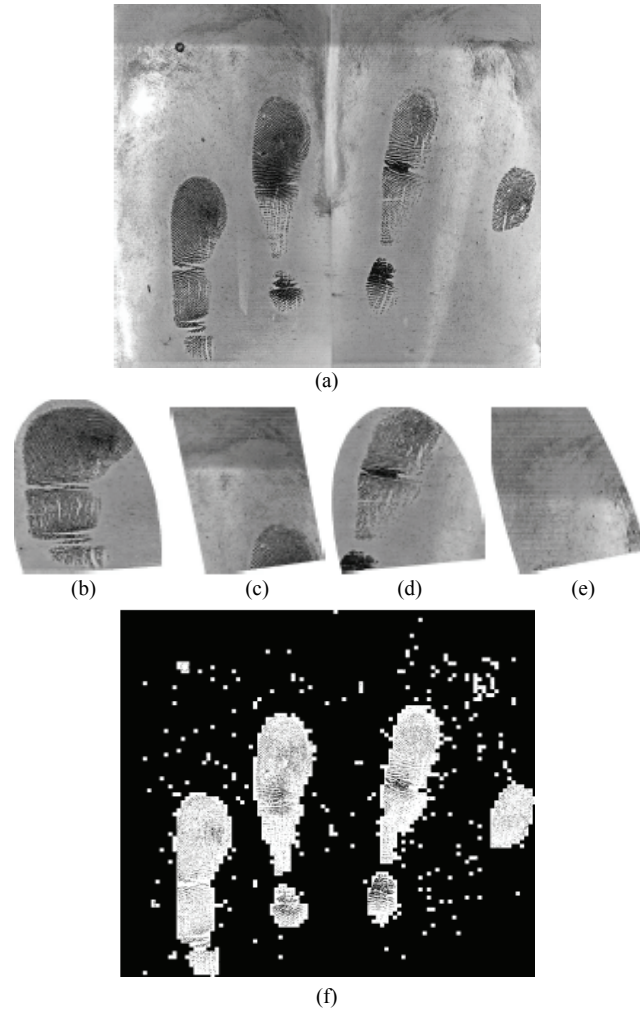| Scenario	Segmentation accuracy (%)
Before noise removal process	78
After noise removal process (Phase-I)	89
After noise removal process (Phase-II)	99

(a)

(b)　　　(c)　　　(d)　　　(e)

(f)

Fig. 2. Illustration of sequence of steps of de-noising slap fingerprints for accurate slap fingerprint segmentation: (a) slap with high noise, (b), (c), (d), and (e) are segmented fingers of the slap with high noise, and (f) foreground-background separation of high noisy slap.
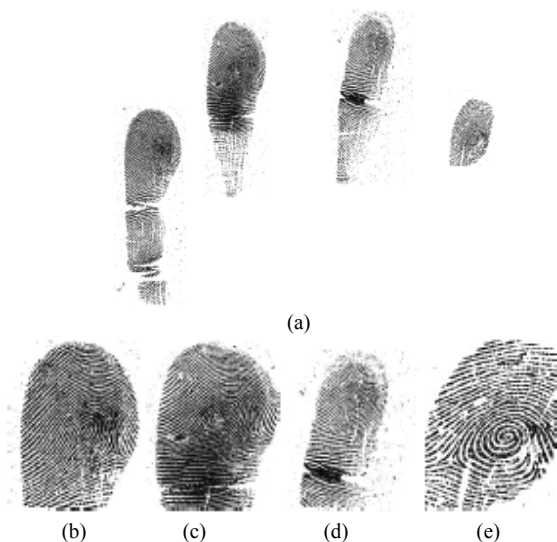

(a)

(b)　　　(c)　　　(d)　　　(e)

Fig. 3. Illustration of sequence of steps of de-noising slap fingerprints for accurate slap fingerprint segmentation: (a) noise-free slap of high noise, (b), (c), (d), and (e) are segmented fingers of the noise-free slap of high noise.
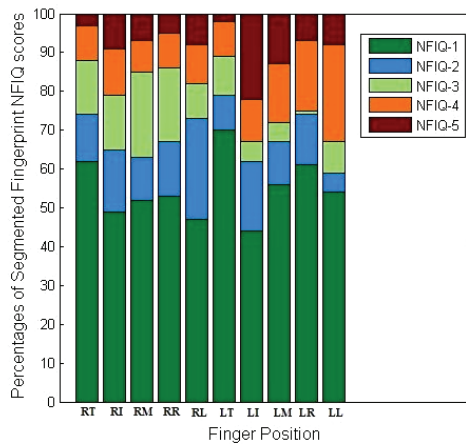
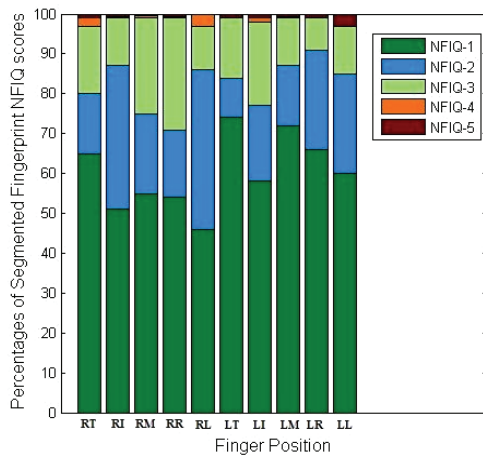Fig. 4. Levels of fingerprint image NFIQ scores for the entire dataset before the noise removal process.

Fig. 5. Levels of fingerprint image NFIQ scores for the entire dataset after the noise removal process.

The results shown in Fig. 4 and Fig. 5 illustrate the levels of segmented fingerprint image NFIQ scores before and after the noise removal process of the entire dataset respectively. The NFIQ scale values from 1 to 5 are represented as NFIQ-1, NFIQ-2, NFIQ-3, NFIQ-4, and NFIQ-5, respectively. The X-axis shows the finger positions in the sequence of right thumb (RT), right index (RI), right middle (RM), right ring (RR), right little (RL), left thumb (LT), left index (LI), left middle (LM), left ring (LR), and left little (LL). The Y-axis represents the percentages of segmented fingers with NFIQ scores. The correct segmentation is defined by fixing the NFIQ scores of all the respective slap segmented fingers less than 4. It is observed that the segmentation failure before the noise removal process for the two thumbs is 4%, the failure for the left four fingers slap is 11%, and it is 7% for the right four fingers slap as shown in Fig. 4. The correct segmentation rate for the entire dataset is significantly improved to 99% as shown in Fig. 5.

5. Conclusions

In this paper, we address a few issues to improve the de-duplication accuracy in large scale de-duplication applications. The large scale de-duplication applications need a lot of enhancements in different phases of recognition process to achieve high speed and good accuracy. The noise removal method is proposed to segment the individual fingers accurately from the slap fingerprint images using binarization of slap fingerprint image, foreground and background segmentation of slap image and region labeling of desired regions with 8-adjacency neighborhood. De-duplication accuracy and slap fingerprint segmentation rate are improved significantly.

References

[1] The Science of Fingerprints. [Online]. Available: http://www.fun-science-project-ideas.com/The-science-of-fingerprints.html

[2] P. Gupta and P. Gupta, "Slap fingerprint segmentation," in *Proc. of IEEE 5th Int. Conf. on Biometrics: Theory, Applications and Systems*, Arlington, 2012, pp. 189−194.

[3] (May 2007). American National Standard for Information Systems Data Format for the Interchange of Fingerprint, Facial, & Other Biometric Information Part 1, ANSI/NIST-ITL 1-2007. [Online]. Available: http://fingerprint.nist.gov/standard/index.html.

[4] M. U. Akram, S. Nasir, A. Tariq, I. Zafar, and W. S. Khan, "Improved fingerprint image segmentation using new modified gradient based technique," in *Proc. of IEEE Canadian Conf. on Electrical and Computer Engineering*, Niagara Falls, 2008, pp. 001967−001972.

[5] J.-J. Gao, and M. Xie, "The layered segmentation, gabor filtering and binarization based on orientation for fingerprint preprocessing," presented at the 8th IEEE Int. Conf. on Signal Processing, Banff, 2006.

[6] J.-Z. Cao and Q.-Y. Dai, "A novel online fingerprint segmentation method based on frame-difference," in *Proc. of IEEE Int. Conf. on Image Analysis and Signal Processing*, Kuala Lumpur , 2009, pp. 57−60.

[7] X. Guo, G. Yang, and Y. Yin, "Sensor interoperability of fingerprint segmentation: An empirical study," in *Proc. of IEEE Int. Conf. on Information Engineering and Computer Science*, Wuhan, 2009, pp. 1−4.

[8] J. Qi and M. Xie, "Segmentation of fingerprint images using the gradient vector field," in *Proc. of IEEE Int. Conf. on Cybernetics and Intelligent Systems*, Chengdu, 2008, pp. 543−545.

[9] Z. Ma, M. Xie, and C. Yu, "Fingerprint segmentation based on PCNN and morphology," in *Proc. of IEEE Int. Conf. on Communications, Circuits and Systems*, San Jose, 2009, pp. 566−568.

[10] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar *Handbook of Fingerprint Recognition*, New York: Springer-Verlag New York Inc., 2003.

[11] F. Alonso-Fernandez, J. Fierrez-Aguilar, J. Ortega-Garcia, "An enhanced gabor filter-based segmentation algorithm for fingerprint recognition systems," in *Proc. of the 4th IEEE*

Int. Symposium on Image and Signal Processing and Analysis, Dubrovnik, 2005, pp. 239–244.

[12] P. Z.-P. Lo and, P. V. Sankar, "Slap print segmentation system and method," in *Google Patents*, 2006, US Patent 7,072,496.

[13] B. Ulery, A. Hicklin, C. Watson, M. Indovina, and K. Kwong, "Slap fingerprint segmentation evaluation 2004 analysis report," Technical report, National Institute of Standards and Technology, 2005.

[14] S. Maddala, S. R. Tangellapally, J. S. Bartuněk, and M. Nilsson "Implementation and evaluation of NIST biometric image software for fingerprint recognition," in *Proc. of the 4th IEEE Int. Conf. on Signal and Image Processing*, Québec, 2010, pp. 207–211.

[15] B. Ulery, A. Hickline, C. Watson, M. Indovina, and K. Kwong. (March 2005). Slap fingerprint segmentation evaluation 2004. SlapSeg04 analysis report. [Online]. Available: http://www.nist.gov/itl/iad/ig/upload/ir_7209.pdf

[16] N. P. Ramaiah and C. K. Mohan, "De-noising Slap Fingerprint Images for Accurate Slap Fingerprint Segmentation," in *Proc. of the 10th IEEE Int. Conf. on Machine Learning and Applications*, Honolulu, 2011, pp. 208–211.

Nalla Pattabhi Ramaiah was born in Andhrapradesh, India in 1981. He received the B.Tech. degree from the Jawaharlal Nehru Technological University, Hyderabad in 2004 in computer science and information technology and the M.Tech. degree from the University of Hyderabad in 2007 in artificial intelligence. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, Indian Institute of Technology Hyderabad (IITH). His research interests include biometrics, image processing, and pattern recognition.

C. Krishna Mohan was born in Andhrapradesh, India in 1967. He received the Bachelor of Science Education (B.Sc.Ed) degree from Regional Institute of Education, University of Mysore in 1988 and the Master of Computer Applications (M.C.A) degree from S. J. College of Engineering, Mysore in 1991. He received the Master of Technology in system analysis and computer applications from IITH, Surathkal in 2000. He received the Ph.D. degree from Indian Institute of Technology Madras in 2007. He is an assistant professor with the Department of Computer Science and Engineering, IITH. His research interests include video content analysis, pattern recognition, and neural networks.

The author's photograph is not available at the time of publication.

.