Multi-Layered Multimodal Biometric Authentication for Smartphone Devices

https://doi.org/10.3991/ijim.v14i15.15825

Qurban A. Memon
UAE University, Al Ain, UAE
qurban.memon@uaeu.ac.ae

Abstract—As technological advances in smartphone domain increase, so are the issues that pertain to security and privacy. In current literature, multimodal biometric approach is addressed at length (for banks as an example) to improve secured access into personal devices. However, personal devices currently do not support enforcing multilayered access to its different domains/regions of data. In this paper, a multilayered multimodal biometric approach using three biometric methods (such as fingerprint, face and voice) is proposed for smartphones. It is shown that fusion of biometric methods can be layered to enforce private data security on smartphone. The experimental results are presented.

Keywords—Multimodal Biometric, Security, Privacy, Data Partition.

1 Introduction

For security reasons, user authentication has turned out to be important tool especially in current smartphone devices, which contain a lot of persona and private data. Biometric-based techniques are an alternative to passwords, smart cards, tokens etc., to authenticate persons, and may replace knowledge-based user authentication methods. The strength of these systems frees user from recalling passwords or PINs and at the same time increases the binding of recognition process to persons. The face recognition approach, being one of them, has received much interest in developing commercial products. Other competing approaches involve iris, voice, fingerprint, hand images, etc. Face recognition has turned out to be frequently used and preferred one in today's emerging video surveillance and biometrics based products, as it depends on identifying a person based on a face image. The thrust in adopting this face recognition approach has resulted in in development of many algorithms and commercial products.

A lot of work has been reported on face recognition for biometric based security [1-4]. For example, few face recognition algorithms are evaluated in [1], where authors compare two well-known dimensionality reduction algorithms such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) by evaluating the performance of respective algorithms based on the basis of recognition rate, and claim that LDA outperforms PCA, for a large size training data set. In [2], the authors present independent and comprehensive comparative analysis (involving performance and computational complexity) of six subspace face recognition algorithms tested on three

databases (i.e., FERET, ORL and YALE) with four popular distance metrics, using FERET evaluations methodology that closely simulates real life conditions. Beside facial images, the authors in [3] also investigate subject characteristics and respective environment in relation to face recognition algorithms performance, by using statistical methods to biometric performance evaluation to help developers in the related field to improve on recognition performance of face recognition algorithms. For real time performance of facial recognition algorithm, the authors [4] use TMS320C64x platform to evaluate memory requirements and power efficiency and conclude that well-optimized implementation may provide an effective design choice for embedded products. The interested reader is also referred to [5] for further reading on face recognition approach.

The research is also reported on fingerprint-based approach towards biometric based recognition. One such work [6] evaluates fingerprint identification, where discrete cosine transform (DCT) is applied on each segment of the fingerprint image to extract directional information features. The final decision is based on comparison to database image features. In another work [7], an experimental system that combines fingerprint and PIN verification using a double random phase encoding scheme is developed that employs a template to check estimated position difference. The security of fingerprint recognition is examined in [8], where authors present the methodology to evaluate the security of a fingerprint recognition system using an attack tree that corresponds to multitude of vulnerabilities. Similarly, a research on feature detection algorithms for fingerprint recognition systems is presented [9], where a detailed work is done on various factors that include image quality, image enhancement, segmentation, feature detection and verification, etc. Summarizing, the contributed work [9] includes fingerprint classification metrics, corner point feature-based segmentation method, two thinning-free feature detection algorithms to produce high accuracy feature sets, etc.

Research is also reported on voice biometric for authentication purposes. For example, the authors in [10] highlight major developments in automatic speech recognition in different areas such as knowledge representation, models, infrastructure, algorithms, search, metadata, etc. In [11], the author comprehensively discusses speech recognition, speaker recognition and verification, then identify limitations in typical smartphone implementation. The author proposes several methods to speed up speaker recognition on fixed point processor with maximum possible accuracy. For speaker verification, several databases have been built [12], where authors have analyzed speech databases to facilitate research and evaluation for speaker recognition. In another research, the authors [13] highlight innovations in voice biometrics and discuss how these innovations support the biometrics community to adopt speaker recognition systems.

The improvement over existing biometric methods can be achieved if biometric approach is layered [14] or fusion of different biometric methods is created [15-17], since face, voice or fingerprint are prone to individual duplication and that each one could be intolerant to respective changes due to aging, mood level or physical condition. The authors in [15] present supervised learning-based fusion of different biometric techniques to improve performance of overall biometric authentication. The fusion is made at each unimodal score level to develop a multimodal biometric system in order to improve accuracy over any existing unimodal biometric system. Similarly, the authors in [16] conclude that score level fusion is superior to decision level or rank level fusion at

the cost of complexity, while authors in [17] study and evaluate different multimodal biometric techniques using fusion at score, feature and decision level. For further related works on secured data access, interested readers are referred to [18-22].

In this paper, a multi-step/multi-layered multimodal biometric for smartphone devices is presented to investigate authentication with full or partial privileges depending on work environments. The paper is structured as follows. In the next section, proposed approach for biometric fusion is presented, followed by experimental setup in section 3. The results are presented in section 4, biometric data storage is discussed in section 5, and conclusions are in section 6.

2 Proposed Approach

In order to have simple and easy to use authentication scheme, we propose to divide access into three data domains: general, semi-private, private. In general domain, for example, limited Internet and voice calls may be allowed, whereas in semi-private domain, full Internet access, voice calls and messaging (without financial and social media access) may be provided. For private domain, full access to the device is allowed. Each data domain is triggered by either one or fusion of more than one biometric methods. In this work, we propose access to general domain by fingerprint biometric method, whereas semi-private domain is accessed using fusion score (f_u) of fingerprint and face recognition approaches. The score based fusion is easier to implement since most biometric modules generate scores. For private (full) access, speaker verification score (S_s) is conducted to compute fusion score (f_v) after successful fusion of fingerprint (S_f) and face recognition (S_F) scores. In this step, fusion has three-dimensional vector as an input. Thus, for full access, the proposed approach involves three step multi-modal biometric approach, where success of a biometric authentication step triggers another one, and so on. Each step computes a score compared to a threshold to determine failure, safe and success. The score above a threshold is divided into two ranges: safe mode and success mode. In safe mode, privileges are assigned up to the current domain, whereas success leads to another biometric authentication for greater access. For general access, the safe and success mode are defined in the interval $T_{f1} \leq$ safe mode $\leq T_{f2}$ and success mode $> T_{f2}$ respectively, whereas for semi private, modes are defined as $T_{u1} \leq$ safe mode $\leq T_{u2}$ and success mode $> T_{u2}$ respectively. The proposed approach is illustrated as shown in flow diagram (in Figure 1).

In this research, a supervised fusion algorithm using support vector machine (SVM) is used to train and test the network. The available database (developed from 200 subjects: 160 users and 40 imposters) is recorded in a month time interval and is split into two sets: training set (150) and validation set (50). Though, the dataset may look small for deep learning applications, but this set seemed enough for classifying into two or three classes. Since, there are two fusion steps, hence two separate SVM networks are trained to classify data into three classes for semi private access, and two classes for private access. Each SVM network takes training data and throws into higher dimensional space (using a 2^{nd} order polynomial kernel) to be linearly separable into general, semi-private, and full access domains [23], though a radial basis function kernel can

also be used. The performance for this type of supervised fusion depends on tuning of training parameters, such as kernel function used, parameter value C, etc. [23].

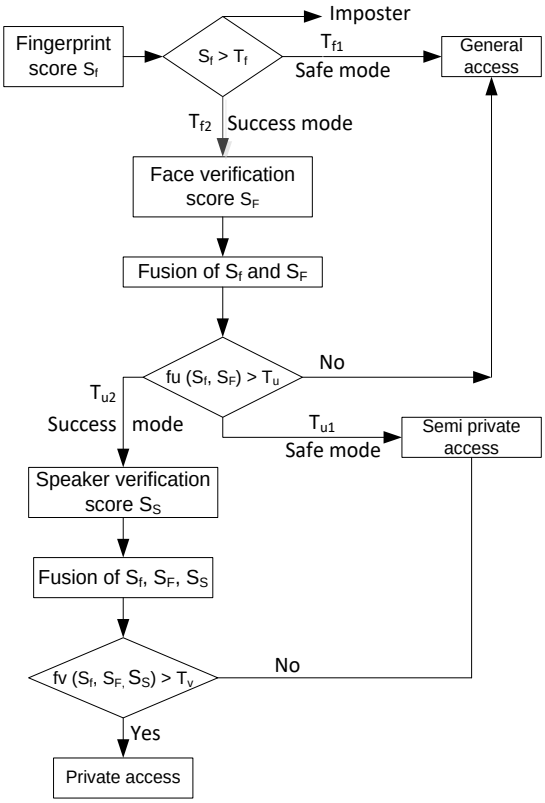

Fig. 1. Flow diagram of data access

The success of the testing is determined by calculating False Acceptance Rate (FAR), False Rejection rate (FRR) and Total Error Rate (TER). The purpose of FAR is to estimate impostor attempts that are accepted as correct, while FRR estimates genuine attempts that are rejected falsely. In the next section, the implementation details are presented:

$$FAR = \frac{Number\ of\ accepted\ imposters}{Number\ of\ imposter\ transactions} \quad (1)$$

$$FRR = \frac{Number\ of\ rejected\ users}{Number\ of\ rejected\ users} \quad (2)$$

$$TER = \frac{FAR+FRR}{2} \quad (3)$$

3 Experimental Setup

In this section, fingerprint, face authentication and speaker verification steps are discussed followed by experimental results about corresponding fusion steps. For multimodal biometric, the first mode is finger-print authentication, where an approach is used that detects minutiae and traces the ridge in region of interest. Once detected, each ridge is labeled for post processing by low pass filters at regions that need smoothing [24]. This adaptation is done due to presence of variance in the ridge at different points, and thus approximates the ridge using piecewise lines. The tracing results in forming of the skeleton image. The block diagram for this step is shown in Figure 2(a). The performance efficiency of finger print recognition can be objectively measured [24] by calculating mean values of error index (E_I) over all minutiae of the image [24]:

$$E_{Im} = \sqrt{\frac{\sum_{m=1}^{M}(e_i^m)^2}{Mxd^2}} \tag{4}$$

where E_{Im} is the average normalized location error, *M* is the number of minutiae, e_i^m is the location error of a minutiae, and *d* is constant and set to 10.

In case of success in first step, the system leads to second mode of authentication, which is face recognition. There are two modes of operation (i.e., authentication and identification) for face recognition. As a first step, the system processes the individual's face (gray) image and then rejects/accepts the claimed identity. In this, the system compares the given face image to a database of registered persons, and returns the confidence score or most likely identities, as shown in Figure 2(b). In terms of implementation, an ORL database is made available. The training images go through feature extraction and then these vectors are projected into a subspace. The feature extraction is done using Linear Discriminant Analysis (LDA) [25], which transforms high-dimensionality data into lower-dimensionality space to minimize data dispersion from the same class and maximizes the data-points separation from different classes. For testing, features are extracted using the same approach and then matched with stored vectors from the database using a distance measure such as cosine similarity as:

$$Cos\ \theta = \frac{\vec{x}.\vec{y}}{\|\vec{x}\|\ \|\vec{y}\|} \tag{5}$$

where *x* and *y* represent features of testing and database samples respectively.

For voice recognition, the Mel frequency Cepstral Coefficients (MFCC) are computed, followed by matching, as illustrated in Figure 2(c). Once speech signal is digitized into 12-bit resolution, the silence from speech is removed, followed by pre-emphasis to boost high frequency values. The MFCC frame duration is set typically between 20ms and 40ms,and is overlapped to offset windowing. The frame size is computed by setting frame duration (say 32ms) with sampling frequency of 16.384 kHz, as:

$$Frame\_size = frame\_duration \times f_s = 0.032\text{x}16384 = 512 \text{ samples} \tag{6}$$

Further, windowing equal to frame size is used to smooth edges generated due to cutoff during. After this, fast Fourier transform is applied to estimate power in different

frequencies present in the speech, followed by its energy calculation. In order to mimic human behavior, the Mel frequency scale is used. The Mel-filter bank is designed to calculate total frequencies energy under each frame, to reflect total energy for each frame. Then, its log is obtained to undergo discrete cosine transformation to obtain final coefficients. Only the first 12 of these coefficients are taken and stored as a reference features because they contain the essential voice characteristics needed for recognition.

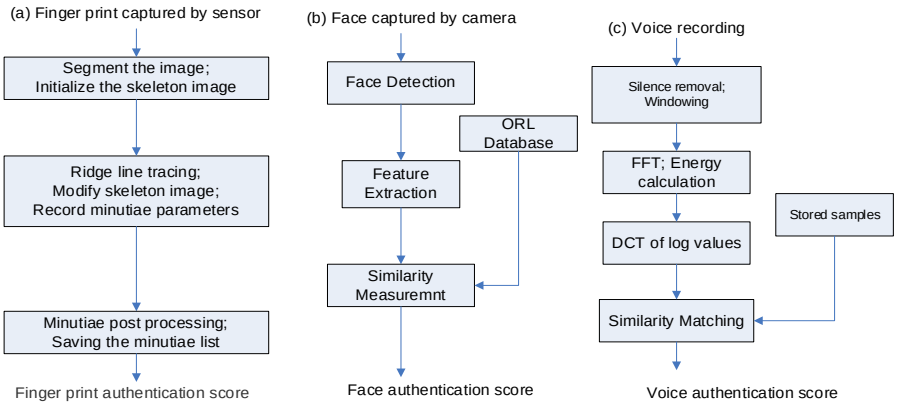

Fig. 2. Individual biometric authentication scores

For robustness, five recorded samples of the speech signal of the same subject are processed and stored to account for slightest variation in the speech frames. For matching, a similarity matrix is developed by multiplying each column in one matrix with corresponding column in another and then normalizing the result by product of square root of sum of each individual matrix values. Finally, testing score against all five stored samples of the same person is calculated using minimum cost function. The minimum cost function is distance from the first cell in the matrix (1, 1) to the last cell (n, m), such that it results in the total smallest cost possible. Using various experiments, a suitable decision was estimated for authorization, and was determined to be below 2.

4 Results

For each independent biometric approach, testing was carried out using Huawei Mate 10 smartphone with front-mounted fingerprint sensor on 3000 fingerprint images each of size 512x512 from NIST database 4. For unimodal fingerprint recognition, the mean value of error index (E_I) was recorded and found to 0.23. Furthermore, the values of FAR and FRR were measured. The results are shown in Table 1. For unimodal face recognition, all testing images were collected in same lighting condition using the smartphone camera. The total images tested were 80 belonging to 10 subjects with each having 8 face images. The resulting FAR and FRR were recorded and are shown in Table 1. For unimodal voice recognition testing, five speech samples of ten subjects were recorded and stored on smartphone for testing self and imposter user. The results

are shown Table 1. For fusion, two separately trained SVM networks were tested using 50 subjects (i.e. 38 users and 12 imposters), and FAR and FRR parameters recorded. The results are displayed in Table 1. From Table 1, the results show that multimodal biometric authentication improves with fusion.

Table 1. Experimental Results

	Fingerprint (1)	**Face (2)**	**Fusion (1 & 2)**	**Voice (3)**	**Fusion (1, 2, & 3)**
FAR	1.7%	1.9%	0.22%	1.8%	0.014%
FRR	1.5%	1.8%	0.21%	1.7%	0.015%
TER	1.6%	1.85%	0.215%	1.75%	0.0145%

5 Biometric Data Storage

Generally, there are five ways to store biometric data:

a) **Recognition system on hardware**: In this approach, data is stored locally on a hardware and works with the device for recognition for a faster response.
b) **Portable token**: This ensures that biometric data is stored on portable device like a smart card. This way, biometric data is captured, template created and then stored on the card to avoid network related vulnerabilities.
c) **On-device**: Being most common, the biometric data is stored on the end-user's device storage through a chip that protects the data separately to the device's network.
d) **Biometric database server**: This is one of the most cost-effective way for biometric data storage but is more susceptible to network related vulnerabilities. However, this approach offers multi-location verification process.
e) **Distributed data storage**: This way, biometric templates are stored both on a server and a device to provide distributed security. In fact, the data is divided into smaller, encrypted files stored separately on server and storage area of the authentication device.

For smartphones, an isolated area in the phone's hardware like a separate processor with its own memory and operating system (commonly known as TEE – Trusted execution environment) can be used to analyze and store biometric data. This provides protection from rooting and with bootloader locked. Typically, bootloader's encryption keys are stored in TEE. This isolated area can work for biometric data the same way it does for encryption keys. Once a fingerprint is registered on Android phone, the captured data is analyzed by operating system in TEE, and then creates validation data and an encrypted template. This will look junk to anyone except the TEE, which holds the encryption key. The encrypted template is stored in another encrypted container, thus securing the biometric data in a three-layer encryption.

6 Conclusion

A multilayered multimodal biometric authentication system was presented for smartphone devices to protect data. The biometric methods used were fingerprint, face and voice, though other methods can also be used. The approach adopted convenient three-step (multilayered) multimodal method, and it was noted that increasing the number of biometric methods is likely to increase inconvenience to the user. The results show that multilayered approach is robust in terms of security as compared to single layered multimodal access. With maturity of unimodal biometric methods, the multi-layered is likely to succeed to the convenience of the user. The only inconvenience to the user is initial storing of credentials such as fingerprint, voice, and facial template in private data area. The other notable challenge to implementation is non-availability of an application that can allow partitioning of data into public, semi-private and private domains. Hopefully, researchers will address this problem in near future.

7 References

[1] Suganya, S., Menaka, D. (2014), "Performance Evaluation of Face Recognition Algorithms," *International Journal on Recent and Innovation Trends in Computing and Communication*, 2(1) pp. 135–140.

[2] Bajwa, U., et al., "A multifaceted independent performance analysis of facial subspace recognition algorithms," PLoS ONE 8(2): e56510. https://doi.org/10.1371/journal.pone.0056510

[3] .Givens, G., *et al*. (2013), "Introduction to face recognition and evaluation of algorithm performance," *Computational Statistics and Data Analysis*, Vol. 67, pp. 236–247. https://doi.org/10.1016/j.csda.2013.05.025

[4] Batur, A., Flinchbaugh, B. (2002), "Performance Analysis of Face Recognition Algorithms on TMS320C64x," *Application Report*, SPRA874 – December.

[5] F. Li, H. Wechsler, (2009), "Face Authentication Using Recognition-by-Parts, Boosting and Transduction," *IJPRAI* 23(3), pp. 545-573. https://doi.org/10.1142/s0218001409007193

[6] Lavanya, B., Raja, K., (2011), "Performance Evaluation of Fingerprint Identification Based on DCT and DWT using Multiple Matching Techniques," *IJCSI*, 8(6), pp. 275-283.

[7] Suzuki, H., *et al*. (2006), "Experimental evaluation of fingerprint verification system based on double random phase encoding," *Optics Express*, 14(5). https://doi.org/10.1364/oe.14.001755

[8] O. Henniger, D. Scheuermann, T. Kniess, (2010), "On security evaluation of fingerprint recognition systems," *International Biometric Performance Conference*, March, pp. 1–5.

[9] Wu, C., (2007), "Advanced Feature Extraction Algorithms for Automatic Fingerprint Recognition Systems," *PhD Thesis*, University of New York at Buffalo.

[10] Baker, J.; *et al*. (2009), "Developments and directions in speech recognition and understanding, Part 1 [DSP Education]," *IEEE Signal Processing Magazine*, 26(3), pp.75-80.

[11] Karpov, E., (2011), "Efficient Speaker Recognition for Mobile Devices," *PhD Thesis*, University of Eastern Finland.

[12] Feng, L., Hansen, L. (2005),"A New Database for Speaker Recognition", *Technical Report*: IMM Informatik og Matematisk Modelling, DTU.

[13] N. Scheffer, *et al*. (2013), "Recent developments in voice biometrics: Robustness and high accuracy," *IEEE International Conference on Technologies for Homeland Security*, Waltham, MA, pp. 447-452. https://doi.org/10.1109/ths.2013.6699046
[14] M. El Beqqal, M. Azizi, J. Lanet, (2018), "A Novel Approach for an Interoperable Biometric Verification," *International Journal of Interactive Mobile Technologies*, 12(6). https://doi.org/10.3991/ijim.v12i6.9528
[15] Damousis, I., S. Argyropoulos, S. (2012), "Four Machine Learning Algorithms for Biometrics Fusion: A Comparative Study," *Applied Computational Intelligence and Soft Computing*, https://doi.org/10.1155/2012/242401
[16] Divyakant T., Meva, D., Kumbharana, C. (2013), "Comparative Study of Different Fusion Techniques in Multimodal Biometric Authentication," *IJCA*, 66(19), March, pp. 16-19.
[17] Soruba Sree, S., Radha, N. (2014), "A Survey on Fusion Techniques for Multimodal Biometric Identification*," IJ RCCE*, 2 (12), pp. 7493-7497, December.
[18] M. A. Dar, J. Parvez, (2016), "Novel Techniques to Enhance the Security of Smartphone Applications," *International Journal of Interactive Mobile Technologies*, 10(4). https://doi.org/10.3991/ijim.v10i4.5869
[19] Q. Memon, S. Khoja, (2010), "Semantic web for program administration," *International Journal of Emerging Technologies in Learning*, 5(4), pp. 31-40.
[20] Q. Memon, et al., (2017), "Audio-Visual Biometric Authentication for Secured Access into Personal Devices, "Proceedings of the 6th International Conference on Bioinformatics and Biomedical Science, pp: 85-89. June, Singapore https://doi.org/10.1145/3121138.3121165
[21] Aliyu Abubakar, *et al*., (2019), "Discrimination of Healthy Skin, Superficial Epidermal Burns and Full-thickness Burns from 2D-Coloured Images Using Machine Learning," *Data Science: Theory, Analysis and Applications .pp* 201-224, CRC Press https://doi.org/10.1201/9780429263798-9
[22] S. Reiff-Marganiec, *et al*., (2013), "User Activity Recognition through Software Sensors," *Distributed Networks: Intelligence, Security, and Applications*, CRC Press. https://doi.org/10.1201/b15282-10
[23] N. Christianini and J. Shawe-Taylor, (2000), *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, Cambridge University Press. https://doi.org/10.1108/k.2001.30.1.103.6
[24] X. Jiang, W.-Y. Yau, and W. Ser., (2001), "Detecting the fingerprint minutiae by adaptive tracing the gray-level ridge," *Pattern Recognition*, 34(5): pp. 999–1013. https://doi.org/10.1016/s0031-3203(00)00050-9
[25] F. Chelali, A. Djeradi and R. Djeradi, (2009), "Linear discriminant analysis for face recognition," *International Conference on Multimedia* Computing *and Systems*, pp. 1-10. https://doi.org/10.1109/mmcs.2009.5256630

8 Author

Qurban A. Memon graduated from University of Central Florida, Orlando, n 1996. Currently, he is working as an Associate Professor at UAE University, College of Engineering. He has authored over hundred publications in his career. He has executed research grants and projects in the area of intelligent systems security and networks. He has served as a reviewer of many international journals and conferences, as well as session chair at various conferences. Email: qurban.memonuaeu.ac.ae.

Article submitted 2020-05-28. Resubmitted 2020-06-23. Final acceptance 2020-06-24. Final version published as submitted by the authors.