International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 6 Issue 2, January-February 2022 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Implementation of Multimodal Biometrics Systems in Handling Security of Mobile Application Based on Voice and Face Biometrics

Gusti Made Arya Sasmita, I Putu Arya Dharmaadi

Department of Information Technology, Faculty of Engineering Udayana University Badung, Bali, Indonesia

ABSTRACT

The use of mobile applications nowadays has been implemented in various fields. Especially in online transactions via smartphone devices, business people have decreased quite a lot because they provide benefits in transactions. Through the mobile application, businesses can process transactions anywhere and anytime. One security aspect regarding authentication is an important concern for consumer who use mobile devices. So far, most of the authentication is done by only using password security, but the level of security is still quite easy to bypass the security. A better level of authentication can be done by utilizing biometric technology into the application security system. Several studies that have been conducted still use unimodal biometric systems to provide good authentication guarantees. This investigation is carried out by research that can improve the authentication of the existing models by applying multimodal biometric authentication based on voice and face images. In voice biometric authentication, the method of MFCC (Mel Frequency Cepstrum Coefficients) and DTW (Dynamic Time Warping) will be used, while facial authentication will apply the PCA (Principle Component Analysis) method. The results obtained show that the user authentication success rate reaches 90%. Therefore, biometric multimodal security can begin to be used for security systems in electronic transactions via smartphone devices.

KEYWORDS: *authentication, biometrics, multimodal, MFCC, PCA*

How to cite this paper: Gusti Made Arya Sasmita | I Putu Arya Dharmaadi "Implementation of Multimodal Biometrics Systems in Handling Security of Mobile Application Based on Voice and Face Biometrics" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-6 | Issue-2, February 2022, pp.683-690, URL: www.ijtsrd.com/papers/ijtsrd49289.pdf

IJTSRD49289

1. INTRODUCTION

The use of smartphones in online transactions is quite in demand by business people because it makes transactions easier. Through the mobile application, businesses can process transactions anywhere and anytime. However, the security aspect regarding authentication is an important concern for transactors who use mobile devices. So far, most of the authentication is done by only using password security, which is still quite easy to bypass the security. A better level of authentication can be done by leveraging biometric technology into the existing security system. Several studies that have been conducted still use unimodal biometric systems to provide good authentication guarantees. In fact, the weakness of unimodal biometrics, which only utilizes one of the existing biometrics, still allows it to be penetrated more easily. After seeing the weaknesses of unimodal authentication, a research was finally carried out that could improve the existing authentication model by implementing multimodal biometric authentication based on voice and face images. In voice biometric authentication using the MFCC (Mel Frequency Cepstrum Coefficients) and DTW (Dynamic Time Warping) methods, while for facial image verification the PCA (Principle Component Analysis) method is applied. The results of this study can form a security system with strong enough authentication, so that only actors who have authority can make transactions electronically through their mobile devices / smartphones.

2. Research Method / Proposed Method

Research methodology is a basic process in a system. The research methodology contains stages or an overview of making the system. These stages include;

A. Defining the problem and limiting the problem so that it can be raised in a study.

B. Literature study by collecting various references that are useful as a theoretical basis for the problem being investigated.

C. Design a system whose contents are operational steps in the data processing and process procedures to support an operation.

D. Test the image classification system that has been designed, the system is tested using wood texture images from database under certain conditions. It aims to obtain data on the accuracy, precision, recall, f1-score, and effect of preprocessing process, then the data obtained from the test results is used as an analysis.

E. Integrating all systems that have been developed in the testing phase and implementing applications on Android.

2.1. System Overview

An overview of the Mobile Application Security System by Implementing Face Image Biometrics can be described as follows.

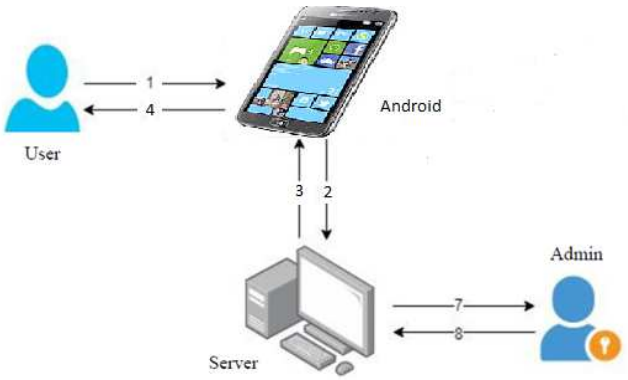

Figure 1. System Overview

Figure 1 illustrates how the mobile application security system works. In the early stages of voting carried out in making a mobile application security system, namely by using the microphone found on the user's smartphone. Then proceed with the second stage by taking facial data using the camera on the smartphone. The captured voice data and face images will then be directly processed by the available modules in the application for matching to the system.

2.2. System Design

System design is a stage for transforming various system requirements into data and program architecture which will be implemented at the system creation stage. The design includes making an overview of the system, explaining the system in the form of a process flow chart.

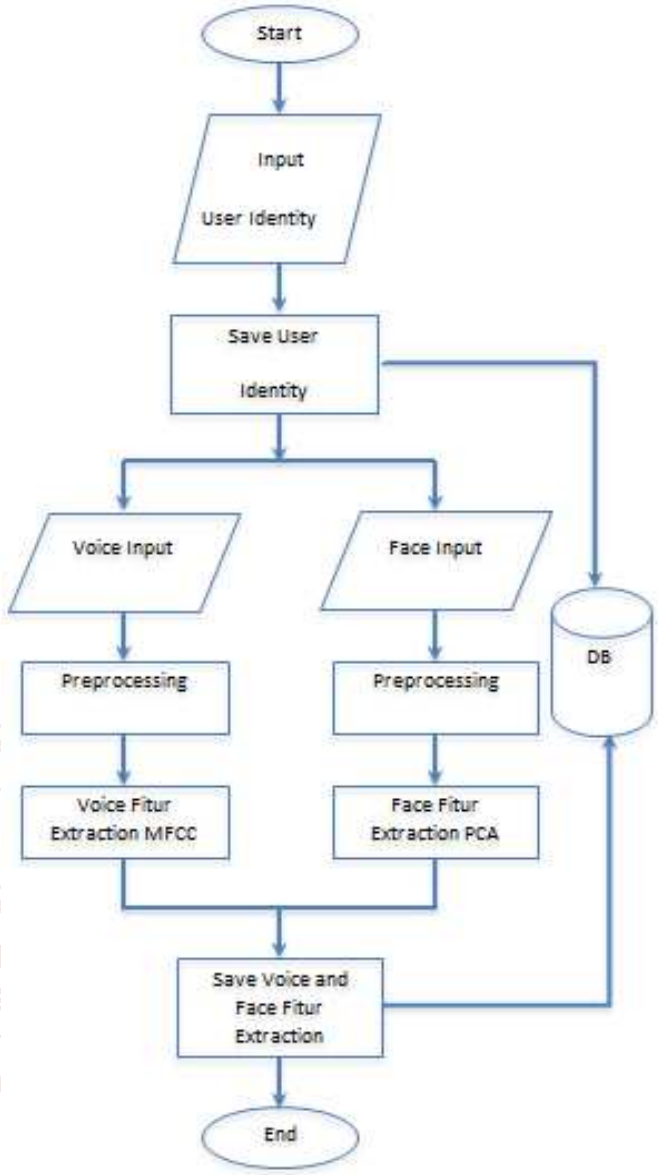

Figure 2. User data input process, voice and face image

Figure 2 is an overview of the user data input process, voice and face images using several processes, namely:

A. Input user identity and save it to the database.

B. Voice data and face image acquisition.

1. Voice data acquisition is done using a microphone available on a smartphone device.
2. Face image acquisition is a face image acquisition from the user. Face image acquisition is carried out using a camera available on a smartphone device.

C. Extract voice data and face image features on the client side

1. The extraction of voice data features was carried out using the MFCC method.
2. Facial image feature extraction is a process to obtain facial image features by applying the PCA method.

D. Voice feature and face image storage to database server.

Voice features and facial images that have been obtained are stored in a database. The results of voice extraction and facial images will be used as a reference pattern for authentication.

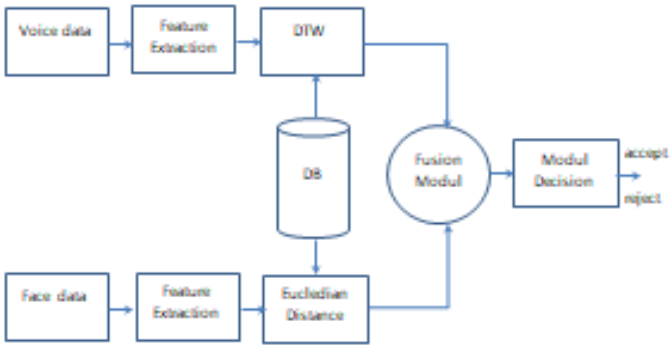

Figure 3. Voice and Face Authentication Process

Furthermore, an overview of the authentication process for facial and voice image data will be carried out as shown in Figure 3.

A. Input user id.
B. Voice data and face image acquisition.
1. Voice data acquisition is done using a microphone available on a smartphone device.
2. Face image acquisition is a face image acquisition from the user. Face image acquisition is carried out using a camera available on a smartphone device.

C. Extract voice data and face image features on the client side
1. The extraction of voice data features was carried out using the MFCC method.
2. The extraction of facial data features was carried out using the PCA method.

D. Verify voice data and face image features according to user identity on the server
1. The matching of voice data features was carried out using the DTW method.
2. Matching facial image features is done by using the Eucledian Distance method.

E. If the verification of both features is successful, the transaction process is declared successful, but if it fails, the voice and face data acquisition process can be repeated.

3. Literature Study

Biometrics comes from Greek which consists of two basic words, namely bio which means life and also metric which means measurement. Roughly speaking, biometrics are measurements made using the life characteristics of the person. Behavioral characteristics are easy to change because they are influenced by human psychological conditions, while physical characteristics have the advantage that they cannot be removed, forgotten or transferred from one person to another, and are also difficult to imitate or falsify. Biometric systems use a person's physical characteristics (such as fingerprints, irises or veins) or behavioral characteristics (such as voice, handwriting or typing rhythm) to determine identity or to confirm that their claims are true. Self-recognition system is a system for recognizing a person's identity automatically using computer technology. The use of the biometric system as an identification and verification system is actually not a new thing. The system will search for and match a person's identity with a reference database that has been previously prepared through the registration process. A biometric system is basically a pattern recognition system that operates by obtaining biometric data from individuals, extracting a set of features from the data obtained, and comparing these features against templates in the database.

3.1. Voice Recognition

Voice is a combination of various signals, but theoretically pure voice can be explained by the oscillation speed or frequency measured in Hertz (Hz) and the amplitude or loudness of the voice by being measured in decibels (dB). Voice recognition first appeared in 1952 and consists of a device for recognizing one digit of spoken words. Then in 1964, IBM Shoebox appeared, one of the well-known technologies in America in the medical field is Medical Transcriptionist (MT) which is a commercial application that uses speech recognition. Voice recognition is divided into two types, namely speech recognition and speaker recognition. Speech recognition is a voice identification process based on the words spoken. The parameter being compared is the level of voice emphasis which will then be matched with the available database templates. While the voice recognition system based on the person speaking is called speaker recognition [15]. An explanation of the classification of the sound signal processing system is shown in Figure 4.

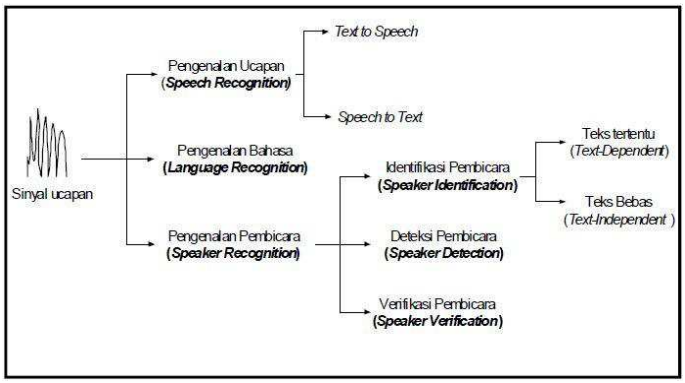

Figure 4 Classification of Voice Signal Processing Systems (Source: Agustini 2007)

3.2. Mel Frequency Cepstrum Coefficients (MFCC) Method

Mel Frequency Cepstrum Coefficients (MFCC) are one of the most widely used methods in the field of speech processing, both speech recognition and speaker recognition which are used to perform feature extraction. This method adopts the workings of the human hearing organ, so that it is able to capture very important sound characteristics which are used to perform parameter extraction, a process that converts sound signals into several parameters. Mel Frequency Cepstrum Coefficients are a technique that takes sound samples as input. After processing, MFCC calculates the unique coefficients for a given sample. MFCC takes the sensitivity of human perception related to frequency into consideration and is therefore best suited for speech recognition. Some of the advantages of this method are:

A. Capable of capturing voice characteristics which are very important for speech recognition or in other words can capture important information contained in the voice signal.

B. Produce minimal data, without eliminating the important information it contains.

C. Replicate the human hearing organ in perceiving sound signals.

MFCC is actually an adaptation of the human hearing system, where the sound signal is filtered linearly for low frequencies (below 1000 Hz) and logarithmic for high frequencies (above 1000 Hz). The parameter extraction stage using the MFCC method is as follows:

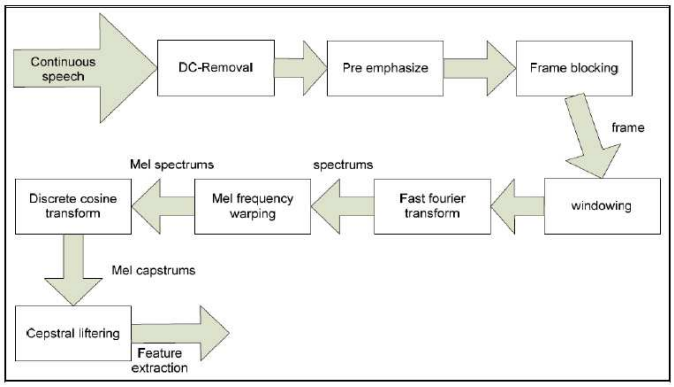

Figure 5 Block Diagram MFCC (Source: Darma Putra, 2009)

3.3. Face Processing

Face detection can be viewed as a pattern classification problem where the input is the input image and the output class label of the image will be determined. In this case there are two class labels, namely face and non-face. Face detection is one of the most important initial stages before the face recognition process is carried out. Research fields related to face processing are:

A. Face recognition is comparing the input face image with a face database and finding the face that best matches the input image.

B. Face authentication, which is testing the authenticity or similarity of a face with face data that has been previously inputted.

C. Face localization, namely the detection of faces, but assuming there is only one face in the image

D. Face tracking is estimating the location of a face in the video in real time.

Facial expression recognition to recognize human emotional conditions.

3.4. Eigenface Facial Recognition Method

Eigenface is a collection of eigenvectors that are used in the field of artificial intelligence to address the problem of human facial recognition. These eigenvectors are derived from the covariance matrix of the random data distribution of human faces at high dimensions in the vector space. This method transforms the face image into a collection of characteristic image features called eigenface, using Principal Component Analysis for the image training process (Turk and Pentland, 1991).

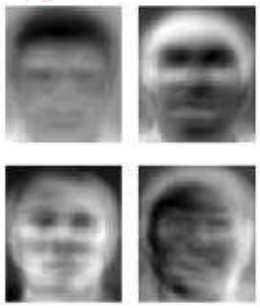

Figure 6 Eigenfaces Image
(http://en.wikipedia.org/wiki/Eigenface)

3.5. Principal Component Analisys (PCA)

Eigenface is a series of eigenvectors used to recognize human faces in computer vision. The approach to using eigenface as a means of facial recognition was developed by Sirovich and Kirby (1987) and used by Matthew Turk and Alex Pentland for classification on faces. And it is considered successful as the first example of facial recognition technology. The eigenvector comes from the covariance matrix which has a high probability distribution and dimensional space vector to identify possible faces.

To produce a set of eigenfaces, the size of a set of calculated facial images, taken under the same lighting conditions, are normalized from the top row of the eyes and mouth. Then all of them are sampled in the same resolution. Eigenface can be extracted from image data. The results of the face data that

have been extracted by eigenface will appear as dark light and the areas are arranged in a certain pattern. This pattern is how to evaluate and assess the various features of each face.

Hotelling proposed a technique to reduce the dimensions of a space represented by the statistical variables x1, x2,…., Xn, where these variables are usually correlated with one another so that there is a new set of variables that has relatively the same properties as the previous variable where it is desired. The new variable set has fewer variables (dimensions) than the previous variable. Hotelling calls this method the Principal Component Analysis (PCA) or Hotelling Transformation and Karhunen-Loeve Transformation. Karhunen-Loeve transformation is widely used to project or convert a large data set into another form of data representation with a smaller size. The Karhunen-Loeve transformation of a large data space will generate a number of ortonormal base vectors into a collection of eigenvectors from a certain covariance matrix, which can optimally represent the data distribution.

4. Result and Discussion

Multimodal biometrics systems is an application on the Android platform. The system that has been designed, then tested with a number of conditions that have been determined, so as to provide the following results.

4.1. Voice Authentication

The low noise condition test is a matching result when the noise condition is heard at a low level, such as the screaming of small children from a considerable distance, people chatting from a distance, etc. At the time of registration and testing as well as recording the results, from 10 users. In voice recognition, the test is repeated 3 times with the same word for each user and matching results are recorded. The following is the user test score data in table form.

Table 1 Voice authentication results

Id User	Identified Data	Unidentified Data
1	1	0
2	1	0
3	1	0
4	0	1
5	1	0
6	1	0
7	1	0
8	0	1
9	1	0
10	0	1
Total	7	3

4.2. Face Authentication

The facial authentication test is performed using an image with a lux value of 600 to 1200 lux. The face detection process uses the OpenCV library, the detected face is then carried out by cutting the image on the face according to the coordinates obtained from the previous detection process and scaling the image size to 150x150 pixels, the normalization process and the process of storing the resulting face.

Normalization using the Histogram Equalization method aims to uniform the brightness levels of faces due to differences in lighting when taking face data. The process of image normalization and scaling is an important part of the recognition process because to produce a collection of eigenfaces requires a collection of human face images that have the same characteristics such as having the same lighting level and having the same size. Face recognition using the PCA method. The recognition system will compare the tested image features with the sample image features by looking for the drinking weight and then this weight value is compared with a predetermined threshold value, if this weight value is less or equal to the threshold value then the test image is successfully recognized and will be given Access rights to access a able behind the online verification able using the face ic able. The determination of the threshold value is very important in able recognition because this value will affect the success rate of the able. The selection of the actual threshold depends largely on what areas the self-recognition able application is applied to. For security applications, the threshold value selected is the threshold value which gives the smallest FAR and FRR values. The following are the results of user testing in table form.

Table 2 Face authentication results

Id User	Identified Data	Unidentified Data
1	1	0
2	0	1
3	1	0
4	1	0
5	1	0
6	1	0
7	1	0
8	0	1
9	1	0
10	1	0
Total	8	2

4.3. Voice and Face Authentication

At this stage of testing, the integration of voice and face authentication is carried out. Based on the authentication design carried out, the fusion technique is applied to the decision module part. This is

intended to provide simplicity of integration and maximize authentication results according to the level of accuracy of each biometric. To verify voice data features and face images according to user identity on the server, voice data features are matched using the DTW method, while facial image features are matched using the Eucledian Distance method.

Then, if the verification of both features is successful, the transaction process is declared successful, but if it fails, the voice and face data acquisition process can be repeated. Increasing the success rate can be done by adjusting the weighting level of each biometric. In this test, face authentication is given a higher weight than voice authentication, with the consideration that the success value of face authentication is better than voice authentication. After testing it appears that the results obtained are as in table 3.

Table 3 Voice and face authentication results

Id User	Identified Data	Unidentified Data
1	1	0
2	1	0
3	1	0
4	1	0
5	1	0
6	1	0
7	1	0
8	0	1
9	1	0
10	1	0
Total	9	1

4.4. Multimodal Biometric Analisys

By looking at the overall results of the trials of each biometrics that exist, the results of the percentage of successful authentication are obtained as in Table 4

Table 4 Results of the percentage of successful biometric authentication

Biometric Types	Total Data	Identified Data	Unidentified Data	Success Rate (%)
Voice	10	7	3	70
Face	10	8	2	80
Voice and Face	10	9	1	90

From the table, it can be seen that the modal union authentication system, both face and voice, each provides quite good authentication results. However, by implementing multimodal biometrics, the authentication success rate will be even better with a success rate of up to 90%. This comparison is better shown in the test result graph in Figure 7.

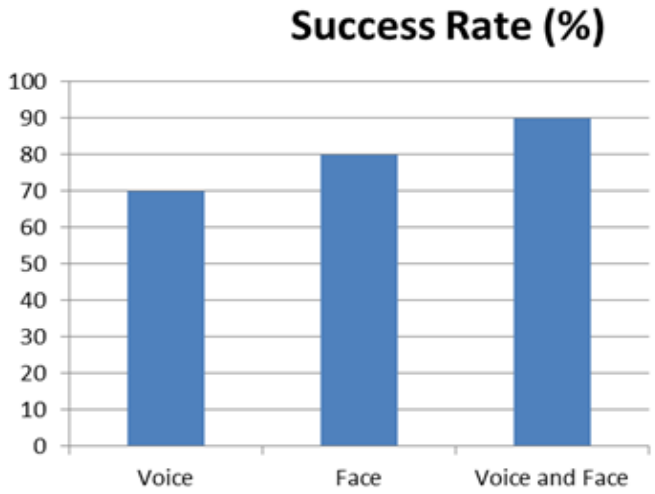

Figure 7 Success Rate Graph

4.5. Implementation of Application

The system that has been designed then implemented into an application that can run on an Android device with the following results.

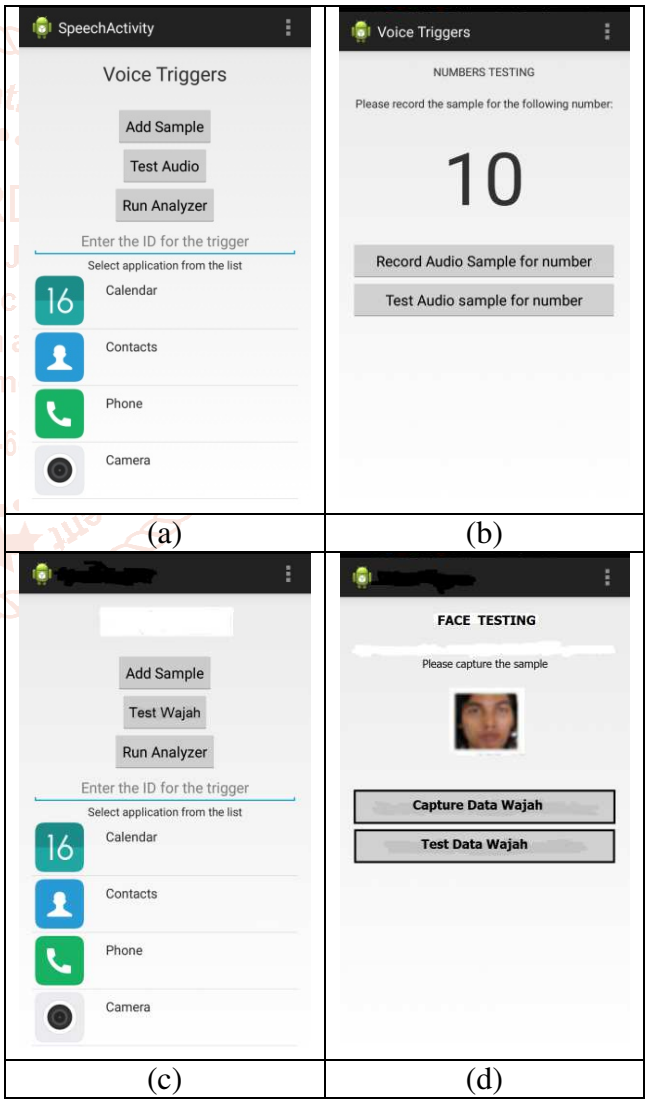

Figure 8 (a) Dashboard Voice (b) Voice Testing (c) Dashboard Face (d) Face Testing

Figures 8 (a), (b), (c), and (d) are application interfaces. There are 3 menus, the first is the menu of adding data, then the menu for testing and running the analysis. To add data, the first step that must be done

is data acquisition by recording voice or capturing faces, after that upload the voice and face image to the server and the server will return an authentication result.

5. Conclusion

By paying attention to the results of the authentication test, both facial and voice biometrics, it seems that it is good enough to identify user identities, but to achieve higher authentication results the application of multimodal biometrics by combining voice and face biometrics can provide a better success rate of up to 90%. So that the application of face and voice multimodal biometrics is still relevant for mobile application authentication today.

References

[1] Agustini Ketut. 2007. "Biometrik Suara Dengan Transformasi Wavelet Berbasis Orthogonal Daubenchies". *Gematek Jurnal Teknik Komputer, September 2007, Vol. 9, No. 2.*

[2] Amtul Fatima, 2011., "E-Banking Security Issues – Is There A Solution in Biometriks?", Journal of Internet Banking and Commerce, August 2011, vol. 16, no. 2.

[3] Chakraborty, Koustav, Asmita Talele, and Prof. Savitha Upadhya. 2014. "Voice Recognition Using MFCC Algorithm." *International Journal of InnovativeResearch in Advanced Engineering (IJIRAE)* 1(10):2349–2163.

[4] Dinata, Candra, Diyah Puspitaningrum, and Ernawati. 2017. "Implementasi Teknik Dynamic Time Warping (DTW) Pada Aplikasi Speech to Text." *Jurnal Teknik Informatika* (April):49–58.

[5] H, Nazruddin Safaat. 2012. *Pemrograman Aplikasi Mobile Smartphone Dan Tablet PC Berbasis Android*. Bandung: Penerbit Informatika. Kumar, A. Vija., Aruna, and M. Vijayapa. Reddy. 2011. "A Fuzzy Neural Networkfor Speech Recognition." *ARPN Journal of System and Software* 1(9):284–90.

[6] Latha P., 2012, "Face Recognition Using Neural Network", Signal Processing An International Journal.

[7] Mansour, Abdelmajid H., Gafar Zen Alabdeen Salh, and Khalid A. Mohammed. 2015. "Voice Recognition Using Dynamic Time Warping and Mel- Frequency Cepstral Coefficients Algorithms." *International Journal of Computer Applications (0975-8887)* 116(2):34–41.

[8] Mayank Agarwal, 2010, "Face Recognition Using Eigen Faces and Artificial Neural Network", International Journal of Computer Theory and Engineering, August 2010, Vol. 2, No. 4, 624-629.

[9] Melissa, Gressia. 2008. "Pencocokan Pola Suara (Speech Recognition) Dengan Algoritma FFT Dan Divide and Conquer." *Makalah If2251 StrategiAlgoritmik*.

[10] Mohamad Abdolahi, Majid Mohamadi, Mehdi Jafari, 2013, "Multimodal Biometric System Fusion Using Fingerprint and Iris with Fuzzy Logic", International Journal of Soft Computing and Engineering (IJCSE), 2013, Vol. 2, No. 6, 504-510.

[11] Mohan, Bhadragiri Jagan and N. Ramesh Babu. 2014. *Speech Recognition Using MFCC and DTW*.

[12] Muhammad, Ardiansyah. 2005. "Penggunaan Jarak Dynamic Time Warping (DTW) Pada Analisis Cluster Data Deret Waktu (Studi Kasus Pada Dana Pihak Ketiga Provinsi Seindonesia)." 277–80.

[13] Natalia, De, Bambang Hidayat Dea, and Unang Sunarya. 2013. "Perancangan Dan Implementasi Speech Recognition Sistem Sebagai Fungsi Unlock Pada Handset Android."

[14] Priambodo, Dimas, Maman Abdurohman, and Iwan Iwut Tirtoasmoro. 2013. "Analisis Dan Implementasi Penggunaan Biometrik Suara SebagaiPembangkit Kunci Kriptografi AES 128 Bit." *Teknik Informatika, FakultasTeknik Informatika, Universitas Telkom*.

[15] Putra, Darma and Adi Resmawan. 2011. "Verifikasi Biometrika Suara Menggunakan Metode MFCC Dan DTW." *Lontar Komputer* 2(1):8–21.

[16] R, Layansan, Aravinth S, Sarmilan S, Banujan C, and G. Fernando. 2015. "Android Speech-to-Speech Translation System for Sinhala." *International Journal of Scientific & Engineering Research* 6(10):1660–64.

[17] Reddy, B. Raghavendhar and E. Mahender. 2013. "Speech to Text Conversion Using Android Platform." *International Journal of Engineering Research and Applications (IJERA)* 3(1):253–58. Retrieved (www.ijera.com).

[18] Savita Choudhary. 2014. "Design of Biometric Based Transaction System Using Open Source

Software Development Environment." *IJIRCCE* 2(2):37–42.

[19] Setiawan, Angga, Achmad Hidayatno, and R. Rizal Isnanto. 2011. "Aplikasi Pengenalan Ucapan Dengan Ekstraksi Mel-Frequency Cepstrum Coefficients (MFCC) Melalui Jaringan Syaraf Tiruan (JST) Learning Vector Quantization (LVQ) Untuk Mengoperasikan Kursor Komputer." *TRANSMISI* 13(3):82–86.

[20] Shweta Mehta et. all, "Face Recognition using Neuro-Fuzzy Inference System", International Journal of Signal Processing, Image Processing and Pattern Recognition, 2014, Vol. 7, No. 1, 331-344.

[21] Sunny, Ananti Selaras. 2009. "Speech Recognition Menggunakan Algoritma Program Dinamis." *STRATEGI ALGORITMIK*.

[22] Wardana, I. Nyoman Kusuma and I. Gede Harsemadi. 2014. "Identifikasi Biometrik Intonasi Suara Untuk Sistem Keamanan Berbasis Mikrokomputer." *JURNAL SISTEM DAN INFORMATIKA* 9(1):29–39.

[23] Warno. 2011. "Pembahasan Bahasa Java Pada Jantungnya Pemprograman."*Jurnal Ilmu Komputer* 7:48–59. Retrieved (http://ejurnal.esaunggul.ac.id/index.php/Komp/article/view/470).

[24] Wicaksono, Anggoro, Sukmawati NE, Satriyo Adhy, and Sutikno. 2014. "Aplikasi Speech Recognition Bahasa Indonesia Dengan Metode Mel- Frequency Cepstral Coefficient Dan Linear Vector Quantization Untuk Pengendalian Gerak Robot." *Prosiding Seminar Nasional Ilmu KomputerUndip* 61–66.