Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 3 (2024)

Multifunctional Biometric Authentication with FaceNet or Gaussian Mixture Model for Face and Voice Recognition

Abhishek Kumar Agrahari[1], Priya Chauhan[2], Pooja Jaiswal[3], Smriti[4], Dheeraj namdev[5]

[1]*Assistant Professor of SCSE at Galgotias University, Greater Noida Uttar Pradesh India*

[2]*Assistent Professor of Computer Science and Engineering Department at IIMT College of Engineering, Greater Noida India*

[3]*Assistant Professor of Computer Science and Engineering Department at Mahatma Gandhi Mission's College of Engineering Technology, Noida India*

[4]*Teaching Assistant of Computer Application Department at Invertis University, Bareilly India*

[5]*Assistant Professor of AI&DS Department IIMT College of Engineering, Greater Noida India*

Abstract: - Advancements in information technology have made information security a crucial aspect of the field. Authentication plays a key role in maintaining security, requiring users to be identified through biometrics that analyze specific physiological and behavioral traits. Reliable personal recognition systems are essential for verifying the identity of individuals accessing various services, ensuring that only authorized users can utilize these services. This case study focuses on enhanced accuracy in multisensory biometric identification, namely voice and face recognition, which effectively reduces the equal mistake rate. The suggested solution uses a Gaussian mixture model for voice recognition, a FaceNet model for facial identification, and score-level fusion to determine user identity. The findings show that this new strategy has the lowest equal error rate when compared to existing methodologies.

Keywords*:* Biometric Authentication, Face Recognition, Voice Recognition, Mobile Security, Multi-Modal Authentication, Machine Learning, Data Privacy, Secure Access, User Experience, Artificial Intelligence (AI).

1. Introduction

Ensuring the integrity, availability, and confidentiality of information in all its forms is a critical aspect of information security. There are various technologies and strategies that contribute to effective information security management, with biometric systems playing a significant role in certain areas. Biometric authentication is vital for identification, authentication, and non-repudiation within the field of information security. Biometric authentication has become increasingly popular for personal identification. Given the rising concerns about identity theft and credit card fraud, accurately verifying a person's identity has become a critical issue in society. Traditional methods such as PINs, token-based systems, and passwords have limitations due to their inherent weaknesses. In contrast, biometrics involve matching a captured sample to a stored template, making it an effective way to identify individuals based on unique traits. Authentication can be performed using three primary methods: knowledge-based security (e.g., passwords), possession-based security (e.g., security tokens or cards), and biometric-based security. Effective security systems often combine multiple inputs, such as security tags, codes, and biometric samples, to enhance security through diverse sample requirements. Biometric authentication aims to establish a reliable one-to-one link between an individual and their information. This paper introduces an improved face-voice multimodal biometric authentication scheme designed to reduce the Equal Error Rate (EER).

Types of Biometrics

Biometric systems can be classified into two main categories:

1. **Unimodal Biometric System:** This system relies on a single biometric trait—whether physical (like face, palmprint) or behavioral (such as voice or gait)—to identify individuals[1]. An example of a unimodal biometric system is one that uses iris recognition for authentication, as illustrated in Figure 1.

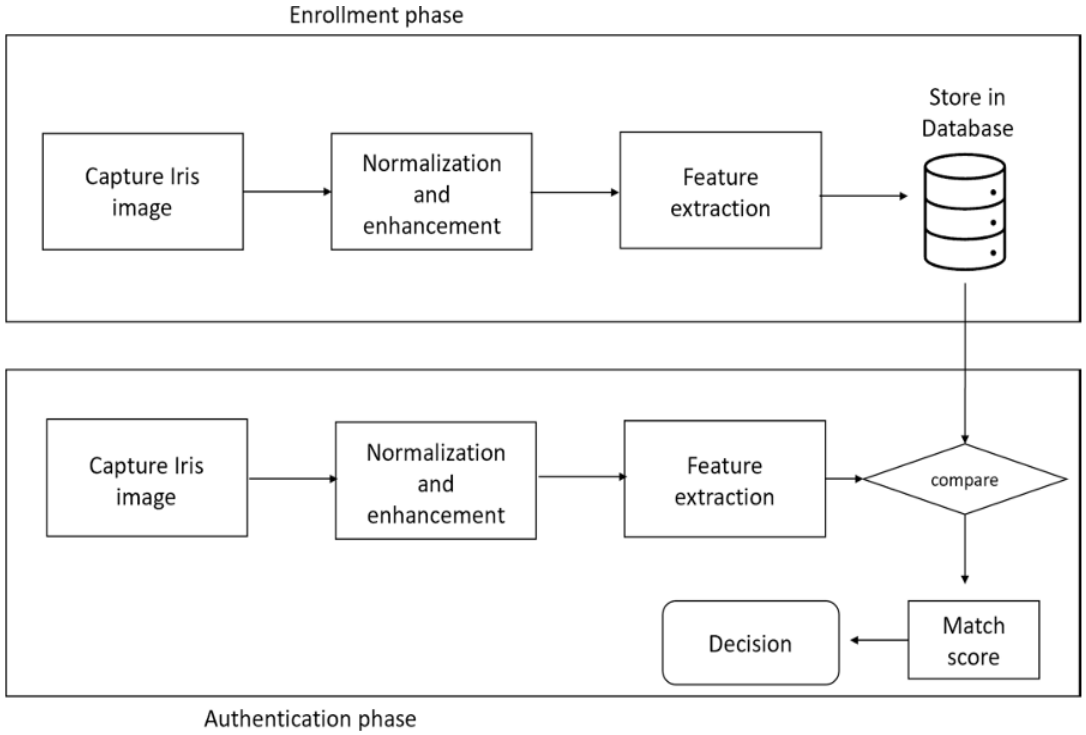

Figure 1: Block diagram of a unimodal biometric system utilizing iris authentication.

2. **Multimodal Biometric System:** This system integrates data from multiple biometric sources. For example, it could combine facial recognition with gait analysis or voice recognition to enhance identification accuracy.

Unimodal biometric systems focus on measuring and evaluating a single attribute of the human body. However, they have several limitations:

- **Data Noise:** The performance of a biometric system is highly dependent on the quality of the biometric sample, which can be affected by noise.
- **Non-universality:** Not all biometric modalities are universal, meaning they cannot be provided by every individual in a given population.
- **Lack of Individual Uniqueness:** Similarities in biometric data among individuals can sometimes reduce the system's ability to distinguish between them (Ammour, Bouden & Boubchir, 2018).
- **Intra-class Variation:** Biometric data collected during the training phase to create a user template can differ from data collected during testing. This variation may result from the user's inconsistent interaction with the sensor (Kabir, Ahmad & Swamy, 2018).
- **Spoofing:** Although it might seem difficult to replicate biometric traits, spoofing techniques can sometimes bypass biometric systems[2].

To address these issues, integrating multiple biometric modalities into a single system, known as a multimodal biometric system, can be an effective solution (Kabir, Ahmad & Swamy, 2018; Matin et al., 2017).

Multimodal Biometric System

Multimodal biometric systems enhance person identification by utilizing information from various biometric features. Figure 2 provides a block diagram of such a system.

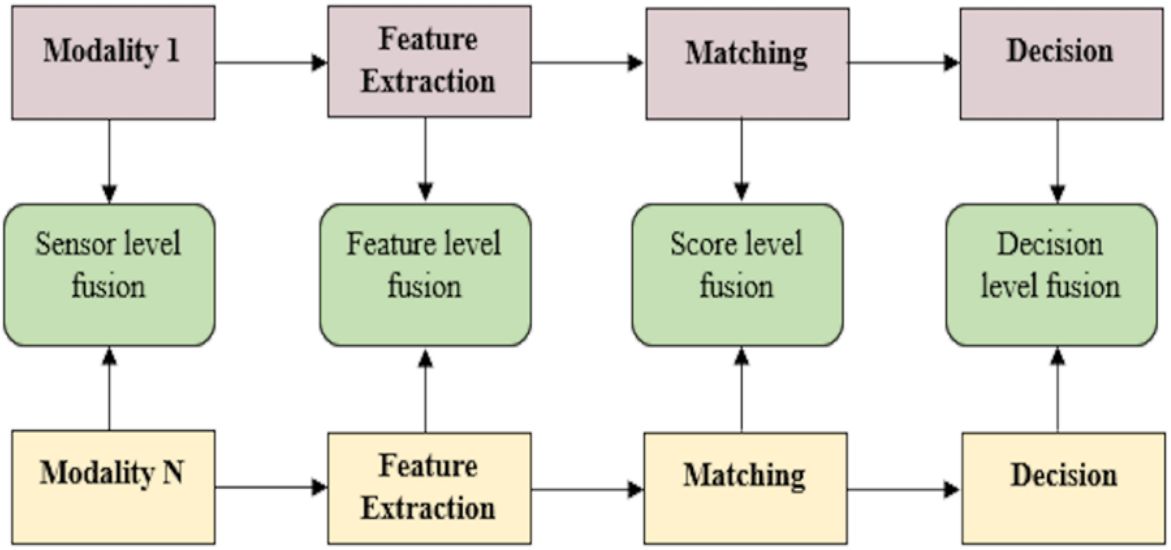

Figure 2: Block diagram of a multimodal biometric system.

Compared to unimodal systems, multimodal biometric systems offer several advantages:

1. They collect multiple forms of biometric information, leading to a significant increase in matching accuracy.

2. Addressing Non-universality: Multimodal biometric systems can effectively overcome the issue of non-universality. For instance, while 2% of individuals might not have usable fingerprints with a unimodal system (Ross, Nandakumar & Jain, 2006), multimodal systems can accommodate a larger user base. If a user lacks one valid biometric trait, they can still be enrolled using another valid trait. This flexibility is further enhanced by registering multiple attributes of a user and verifying only a subset of those features.

3. Reduced Susceptibility to Spoofing: Multimodal biometric systems are less vulnerable to spoofing. The complexity of replicating multiple biometric traits makes it significantly harder to deceive the system compared to unimodal systems.

4. Resilience to Data Noise: These systems are more robust against noise in biometric data. If one biometric trait is affected by noise, the system can use another trait from the same individual for verification

5. Enhanced Monitoring and Tracking: Multimodal systems are beneficial in scenarios where a single biometric feature is insufficient. They enable continuous monitoring or tracking by using multiple traits simultaneously, such as tracking both facial features and gait.

2. Related Work

This section summarizes past research and recent advances in biometric recognition technologies designed to address weak and hackable passwords. Initially, single-feature biometric systems, such as fingerprint recognition, were used, but they showed varied accuracy. As technology progressed, multimodal systems combining multiple biometric traits, like facial recognition with fingerprints, emerged to improve security.

Notable systems include Brunelli & Falavigna's facial and fingerprint system, Kittler & Messer's combination of face, fingerprint, and hand geometry, and the BioID system integrating facial recognition, voice, and lip movement. Advances also include Joseph et al.'s use of fingerprints, palm prints, and iris scans for cloud computing, and Sarier's fingerprint and facial recognition system for mobile computing[3].

Other research explored facial and voice biometrics for their affordability and ease of use. Techniques like fusion, Fast Fourier Transform (FFT), and Discrete Wavelet Transform (DWT) have been applied to enhance performance. Biometric security in wireless body area networks (WBAN) and hybrid systems combining facial

and voice biometrics have also been developed, achieving various equal error rates (EERs). For example, Poh & Korczak's hybrid system achieved EERs of 0.15% for face and 0.07% for voice recognition.

More recent work includes Kasban's use of Mel Frequency Cepstral Coefficients (MFCCs) and Principal Components Analysis (PCA) for feature extraction, and Soltane's use of Gaussian Mixture Models (GMM) for combining features, with notable EERs for multimodal recognition.

Table 1 presents various multimodal biometric schemes involving face and voice.

Multimodal biometric approach	**Extracted features**		**Fusion technique**	**Database**		**Results (EER%)**	
	Face	Voice			Face	Voice	Fusion
Poh & Korczak (2001)	Moments	Wavelet	No Fusion	Persons	0.15	0.07	-
Elmir, Elberrichi & Adjoudj (2014)	Gabor filter	MFCC	CMD	VidTIMIT	1.02	22.37	0.39
Soltane (2015)	Eigenfaces	MFCC	GMM	eNTERFACE	0.399	0.0054	0.281
Kasban (2017)	PCA, LDA, Gabor filter	MFCCs, LPCs, LPCCs	LLR	PROPOSED	1.95	2.24	0.64
Abozaid et al. (2019)	Eigenfaces, PCA	MFCC	LLR	PROPOSED	2.98	1.43	0.62

Table 1 shows some multimodal biometric systems that use speech and recognition of faces.

Based on our review of existing research, we propose a scheme that integrates the GMM model with FaceNet. FaceNet, known for its high accuracy in extracting face embeddings, utilizes a Siamese neural network, which excels in user authentication by managing class imbalances and learning embeddings that group similar classes together, thereby capturing semantic similarity. According to Table 1, the GMM model is highly effective for voice recognition. This paper explores the combination of GMM and FaceNet to create a multimodal biometric authentication system.

Methodology

This section describes the design and structure of our proposed voice and face authentication system. It consists of three main components: (i) voice recognition, (ii) face recognition, and (iii) score-level fusion, as shown in Figure 3.

The system enhances security by using collaborative and federated learning (FL) techniques, which analyze decentralized data to improve data security, privacy, and confidentiality. As data protection becomes more crucial, decentralizing the training process is an effective way to maintain privacy and secrecy (Lu et al., 2022)[4]. The process starts with collecting voice and facial samples, from which features are extracted and stored in a database. During authentication, new samples are collected and compared to the database features. The final decision to grant or deny access is made through score fusion.

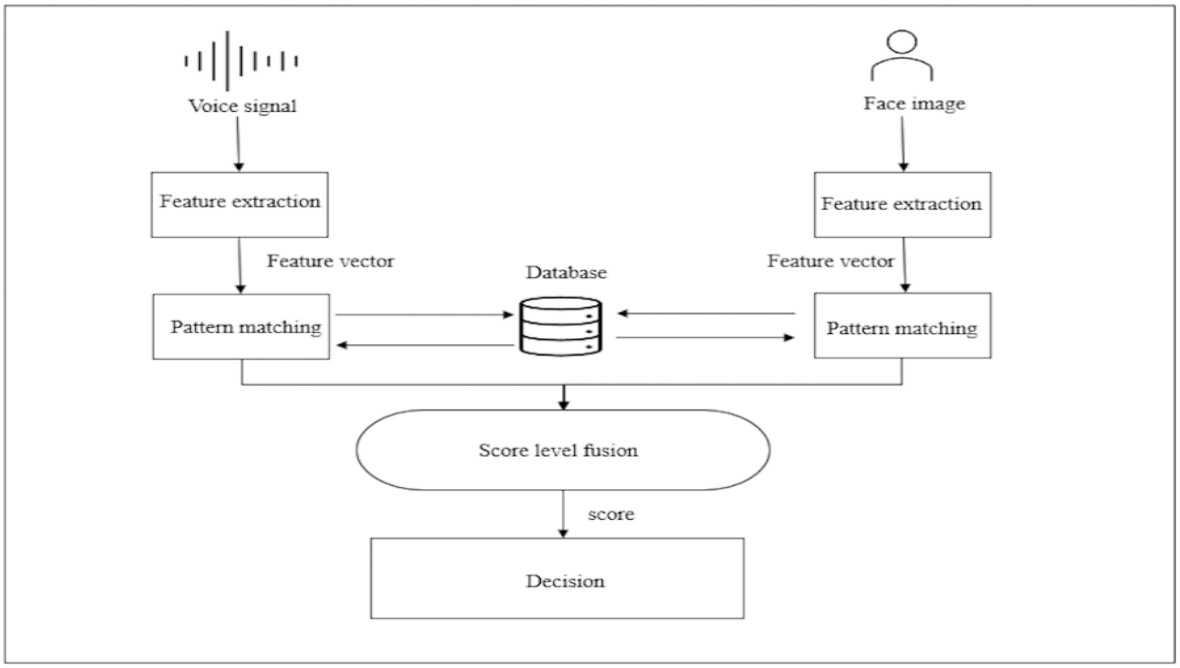

Figure 3: Block diagram of the proposed multimodal biometric fusion scheme.

Voice Recognition

The human voice is unique due to the distinct arrangement of the teeth, trachea, nasal passages, vocal cords, and the individual's way of amplifying sound. These elements combine to create a voiceprint as individual as a fingerprint, which cannot be replicated or transferred (Khitrov, 2013). Unlike other biometric methods, speech biometrics have the advantage of being contact-free. Voiceprints can be captured from a distance, making them useful in various situations, such as while driving, from another room, or on mobile devices (Khitrov, 2013)[5].

A speech biometric system functions by having the user speak a passphrase, which is then recorded and compared to a previously stored voiceprint. The system scores how closely the new speech matches the stored voiceprint. To ensure security, access thresholds can be set, and if the match score is too low, access will be denied (Khitrov, 2013). Figure 4 illustrates the voice recognition process, which includes capturing a voice sample, extracting features, removing noise, and verifying the user.

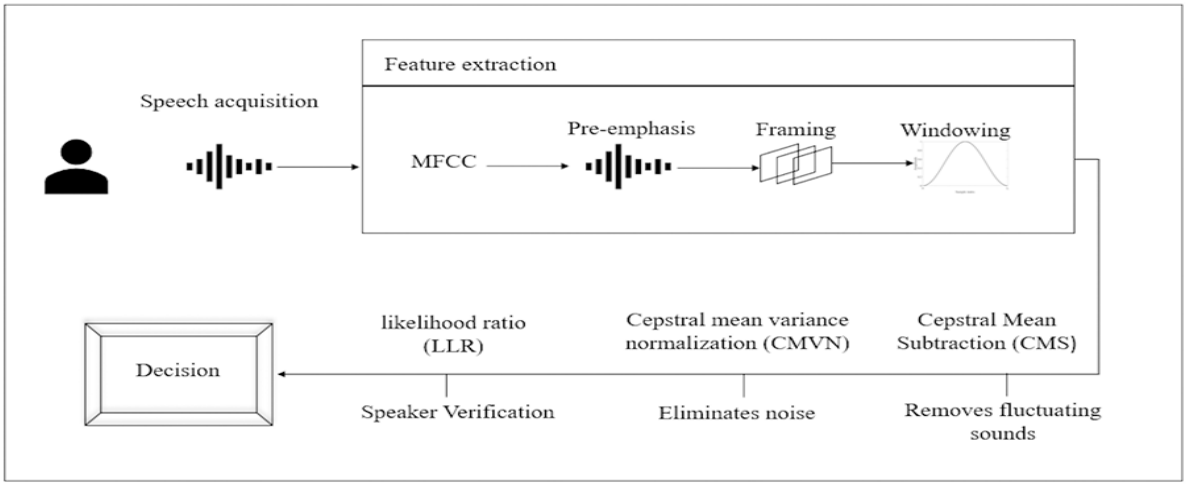

Figure 4: Proposed voice recognition process.

Face Recognition

Face recognition is popular due to its ease of use and societal acceptance. It allows for remote and non-consensual face image capture, making it useful for surveillance. However, many current facial-recognition algorithms struggle with discrimination and are affected by changes over time. Identical appearances in some individuals, such as identical twins, can further complicate recognition.

The FaceNet model, developed by Schroff, Kalenichenko, and Philbin (2015), is employed for face identification, verification, and clustering. The process starts with detecting the face using the Haar Cascade classifier, followed by resizing the face region and generating 128-dimensional facial embeddings. Figure 5 shows the face recognition process, which involves detecting the face, extracting embeddings, predicting identity, and verifying the user[5].

Face Detection

Face detection involves identifying and localizing faces in images using a bounding box. The Haar Cascade classifier, an object detection algorithm introduced by Viola & Jones (2001)[6], is used for this purpose. It works by finding Haar features, creating integral images, training Adaboost, and using a cascading classifier to detect faces (Priambodo et al., 2021)[7]. For effective training, the algorithm requires a dataset of positive facial images and negative non-face images. After detection, the FaceNet model creates face embeddings, and a linear support vector machine (SVM) classifier predicts the face's identity. Detected faces are resized to 96x96 pixels[8].

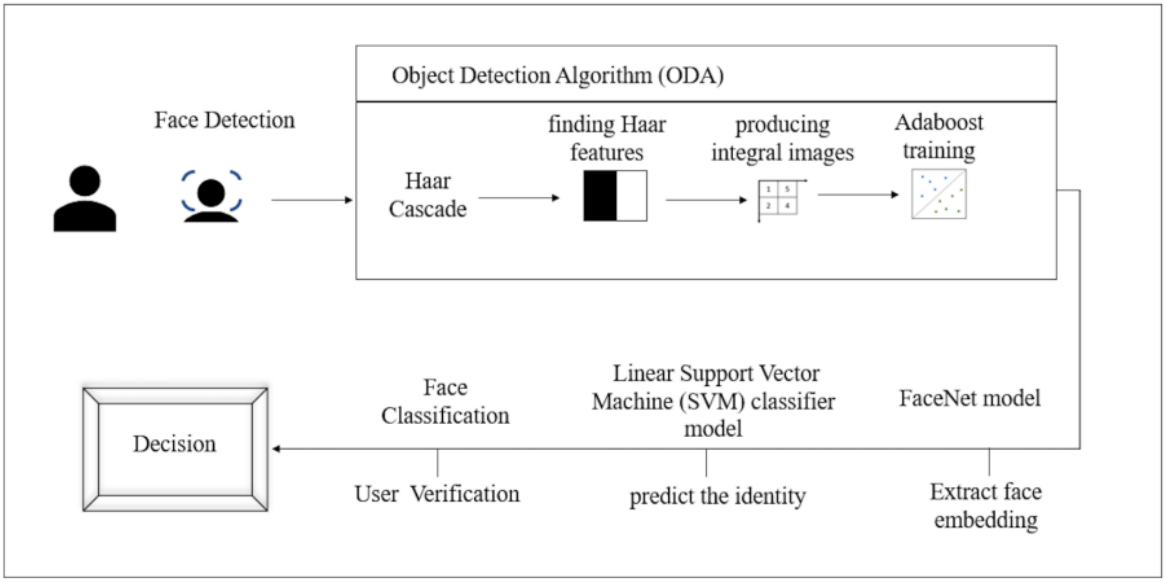

Figure 5: Proposed face recognition process.

Author	Database		Results (EER	%)
		Face	Voice	Fusion
Chetty & Wagner (2008)	AVOZES	3.2	4.2	0.73
Palanivel & Yegnanarayana (2008)	nEWSPAPERS	2.9	9.2	0.45
Soltane (2015)	eNTERFACE	0.399	0.0054	0.281
Kasban (2017)	PROPOSED	1.95	2.24	0.64
Abozaid et al. (2019)	PROPOSED	2.98	1.43	0.62
Proposed scheme	AVSpeech	0.21	0.12	0.01

Table 2: Comparison of the proposed scheme's obtained EER with other published results.

3. Results

Chetty & Wagner (2008): The fusion method had the lowest EER at 0.73%, with face and voice methods having higher EERs at 3.2% and 4.2%, respectively. Palanivel & Yegnanarayana (2008): The fusion method had an EER of 0.45%, which is lower than both face (2.9%) and voice (9.2%) methods. Raghavendra, Rao & Kumar (2010): The face method had the lowest EER at 2.1%, with voice and fusion methods at 2.7% and 1.2%, respectively.Elmir, Elberrichi & Adjoudj (2014): The face method had the lowest EER at 1.02%, with the voice method significantly higher at 22.37%, and fusion at 0.39%.Soltane (2015): The voice method had an extremely low EER at 0.0054%, with face and fusion methods at 0.399% and 0.281%, respectively.Kasban (2017): The fusion method had the lowest EER at 0.64%, with face and voice methods at 1.95% and 2.24%, respectively.Abozaid et al. (2019): The face method had the highest EER at 2.98%, with voice and fusion methods at 1.43% and 0.62%, respectively.Proposed scheme: The fusion method showed the best performance with an EER of 0.01%, followed by face (0.21%) and voice (0.12%).model, which is a leading model for face recognition.

4. Discussion

A multifunctional biometric authentication system using FaceNet for facial recognition and Gaussian Mixture Models (GMM) for voice recognition combines the strengths of both modalities to enhance security and accuracy. FaceNet provides high accuracy and robustness in facial recognition, while GMMs effectively model voice characteristics for speaker identification. Integrating these technologies allows for more secure authentication by cross-verifying identities through both face and voice, reducing the risk of spoofing. However, this approach also involves increased system complexity and resource demands, requiring careful design to ensure a seamless and efficient user experience.

References

[1] Abozaid, A., Haggag, A., Kasban, H., & Eltokhy, M. (2019). Multimodal biometric scheme for human authentication technique based on voice and face recognition fusion. Multimedia Tools and Applications, 78(12), 16345-16361. https://doi.org/10.1007/s11042-018-7012-3

[2] Alan, V., Ronald, W., & Buck, J. (1989). Discrete-time signal processing. United Kingdom: Prentice Hall Inc.

[3] Ammour, B., Bouden, T., & Boubchir, L. (2018). Face-iris multimodal biometric system based on hybrid level fusion. In 2018 41st International Conference on Telecommunications and Signal Processing (TSP) (pp. 1-5). Piscataway: IEEE.

[4] Batool, A., & Tariq, A. (2011). Computerized system for fingerprint identification for biometric security. In 2011 IEEE 14th International Multitopic Conference (pp. 102-106). Piscataway: IEEE.

[5] Bracewell, R. N. (2000). The Fourier transform and its applications. McGraw-Hill Series in Electrical and Computer Engineering: Circuits and Systems.

[6] Brunelli, R., & Falavigna, D. (1995). Person identification using multiple cues. IEEE Transactions on Pattern Analysis and Machine Intelligence, 17(10), 955-966. https://doi.org/10.1109/34.464560

[7] Chetty, G., & Wagner, M. (2008). Robust face-voice based speaker identity verification using multilevel fusion. Image and Vision Computing, 26(9), 1249-1260. https://doi.org/10.1016/j.imavis.2008.02.009

[8] Dodangeh, P., & Jahangir, A. H. (2018). A biometric security scheme for wireless body area networks. Journal of Information Security and Applications, 41, 62-74. https://doi.org/10.1016/j.jisa.2018.06.001