# We are IntechOpen,
# the world's leading publisher of Open Access books
# Built by scientists, for scientists

## 4,800
Open access books available

## 122,000
International authors and editors

## 135M
Downloads

## 154
Countries delivered to

Our authors are among the

## TOP 1%
most cited scientists

## 12.2%
Contributors from top 500 universities

**BOOK CITATION INDEX**
CLARIVATE ANALYTICS
INDEXED

**WEB OF SCIENCE**™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

## Interested in publishing with us?
## Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

# Person Identification Using Multimodal Biometrics under Different Challenges

Önsen Toygar, Esraa Alqaralleh and Ayman Afaneh

Additional information is available at the end of the chapter

## Abstract

The main aims of this chapter are to show the importance and role of human identification and recognition in the field of human-robot interaction, discuss the methods of person identification systems, namely traditional and biometrics systems, and compare the most commonly used biometric traits that are used in recognition systems such as face, ear, palmprint, iris, and speech. Then, by showing and comparing the requirements, advantages, disadvantages, recognition algorithms, challenges, and experimental results for each trait, the most suitable and efficient biometric trait for human-robot interaction will be discussed. The cases of human-robot interaction that require to use the unimodal biometric system and why the multimodal biometric system is also required will be discussed. Finally, two fusion methods for the multimodal biometric system will be presented and compared.

**Keywords:** person identification, biometrics, multimodal biometrics, face recognition, iris recognition, palmprint recognition, ear recognition, speech recognition

## 1. Introduction

Human identification is one of the oldest behaviors that were done by people to distinguish each other. In the old ages, it was unusual to wrongly identify a person because the number of people was not much in each community. Consequently, memorizing all the persons that you deal within that time was possible. Additionally, it was enough to see the face of any person or to hear his voice to recognize him; therefore, human identification was not considered as a hard issue. The increase of the number of people and the occurrence of commercial and financial transactions forced people to find new reliable methods for human identification in order to prevent the unauthorized person to access authorized information. The new methods of

human identification were classified into two main approaches as traditional and biometrics approaches. Matching process of these methods is conducted not only by humans but also by automated systems, which speed up the matching process in addition to the capability of the large size of memory.

## 2. Person identification approaches (traditional vs. biometrics)

The traditional human identification approaches depend on changeable parameters such as passwords or magnetic/ID cards. These parameters can be easily used by illegal persons, if they know the password or have the card. Losing, forgetting, or stealing are common disadvantages for all the traditional identification methods which make it unreliable and inaccurate especially in the high precise system such as forensics, financial, bank, and border ports systems. The need for more robust systems of person identification in addition to the development of the sensors and automated systems was incentive to construct the systems that depend on the unique features of each person. These features are extracted from a human trait such as fingerprint, face, and speech. Human recognition using features that are extracted from inherent physical or behavioral traits of the individuals is defined as biometrics. In addition to the enhancement of the efficiency and capability of recognition systems, biometrics facilitates identifying, and claiming process, where it is not required to memorize any passwords or to carry any ID cards such as passports or driving license.

Biometrics is the science of establishing the identity of an individual based on a vector of features derived from a behavioral characteristics or specific physical attribute that the person holds. The behavioral characteristic includes how the person interacts and moves, such as their speaking style, hand gestures, signature, etc. The physiological category includes the physical human traits such as fingerprints, iris, face, veins, eyes, hand shape, palmprint, and many more. Evaluating these traits assists the recognition process using the biometric systems [1].

A biometric system includes two main phases as enrollment and recognition. Biometric data (image, video, or speech) are captured and stored in a database in enrollment phase. The recognition phase mainly includes extraction of the salient features and generation of the matching scores in order to compare query features against the stored templates. The biometric system will report an identity at the end of the decision process after performing matching, and this will be the identity of the most resembling person in the database.

## 3. Common biometric traits

In this section, a brief overview, requirements, advantages, and disadvantages of the most commonly used unimodal biometric traits are presented and explained.

### 3.1. Face

Face recognition is one of the most important abilities that we use in our daily lives. Face recognition has been an active research area over the last 40 years, and the first automated face recognition system was developed by Takeo Kanade in 1973 [2]. The increasing interest in the face recognition research is caused by the satisfactory performance in many widely used applications such as the public security, commercial, and multimedia data management applications that use face as biometric trait. Face recognition has several advantages over other biometrics such as fingerprint and iris besides being natural and nonintrusive. First, the most important advantage of face is that it can be captured at a distance and in covert manner. Second, in addition to the identity, the face can also show the expression and emotion of the individual such as sadness, wonder, or scaring. Moreover, it provides a biographic data such as gender and age. Third, large databases of face images are already available, where the users should provide their face image in order to acquire driver's license or ID card. Finally, people are generally more willing to share their face images in the public domain as evinced by the increasing interest in social media applications (e.g., Facebook) with functionalities like face tagging.

A face recognition system generally consists of four modules namely face detection, preprocessing, feature extraction, and matching as shown in **Figure 1**. An original face image and its preprocessed variant are also shown in **Figure 2**.

### 3.2. Iris

Iris recognition is one of the most reliable methods for personal identification. The use of iris texture analysis for biometric identification is clearly well established with the advantages of uniqueness and stability. Iris recognition has been successfully applied in access control systems managing large databases. The United Arab Emirates has been using iris biometrics for border control and expellees tracking purposes for the past decade [3].

Iris is one of the most valuable traits for automatic identification of human being. A number of reasons justify this interest. First of all, the iris is a protected internal organ of the eye that is visible from the exterior. The iris is an annular structure and planar shape that turns easily, and it has a rich texture. Furthermore, iris texture is predominantly a phenotypic with limited genetic penetrance. The appearance is stable over lifetime, which holds tremendous promise
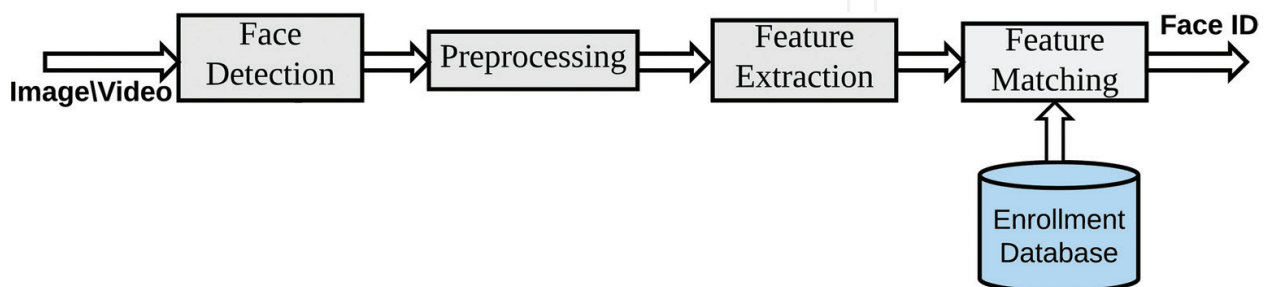


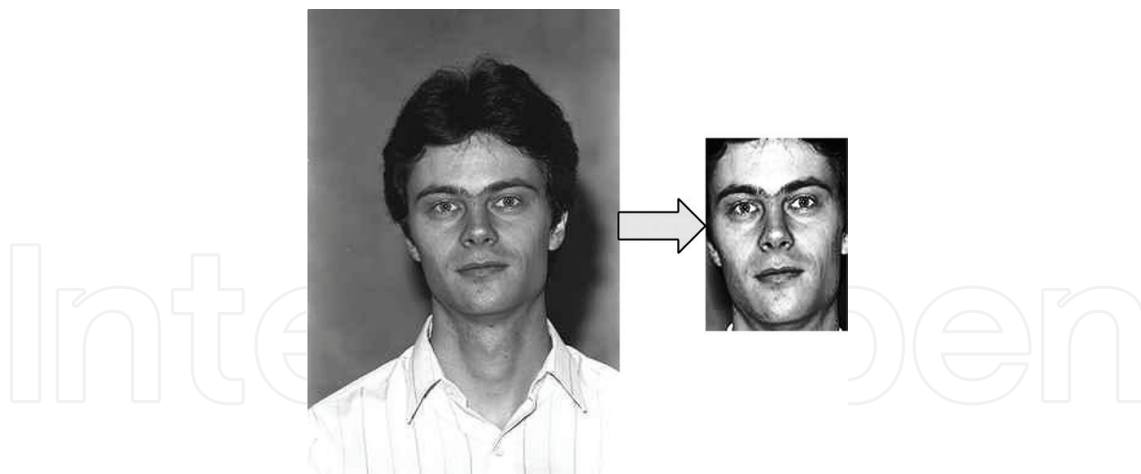**Figure 1.** Block diagram of a face recognition system.

**Figure 2.** An original and a preprocessed face image.

for leveraging iris recognition in diverse application scenarios such as border control, forensic investigations, and cryptosystems.

There are also some drawbacks with it. It needs much user cooperation for data acquisition, and it is often sensitive to occlusion. Iris data acquisition needs a controlled environment. Additionally, data acquisition devices are quite costly. Iris recognition cannot be used in a covert situation.

A typical iris recognition system has four different modules such as acquisition, segmentation, normalization, and matching. These modules are shown in **Figure 3** for a general iris recognition system.

### 3.3. Palmprint

The palmprint recognition system is considered as one of the most successful biometric systems that are reliable and effective. This system identifies the person based on the principal lines, wrinkles, and ridges on the surface of the palm. Studies and research over 10 years have proven that the interesting feature of palmprint is fixed and invariant, and a palmprint acquired from any person is unique, so it can be reliable as a biometric trait.

Some of the advantages of the palmprint recognition compared with other biometric trait systems are invariant line structure, low intrusiveness, and the low cost of capturing device. Palmprint identification requires either high (refers to 400 dpi or more) or low (refers to 150 dpi or less) resolution images in which high-resolution images are suitable for forensic applications such as criminal detection [4] and low-resolution images are more suitable for civil and commercial applications such as access control. High-resolution and low-resolution palmprint images are demonstrated in **Figure 4**. Additionally, the area of palmprint is larger than fingerprint; consequently, there is a possibility of capturing more distinctive features in it.
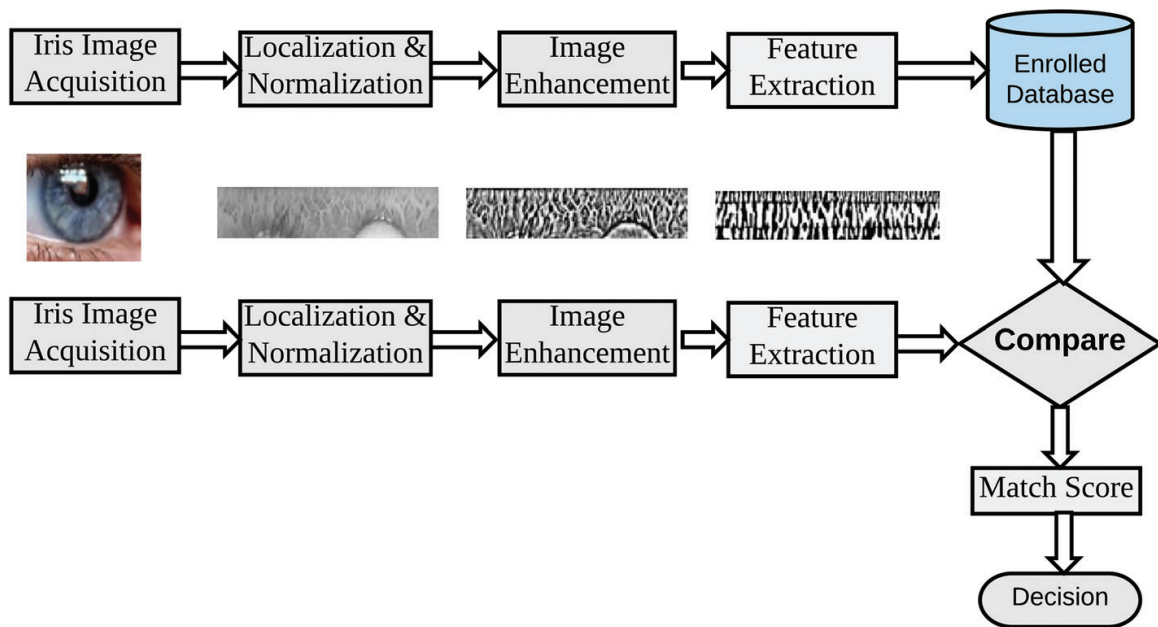
**Figure 3.** Block diagram of an iris recognition system [1].

Due to its low cost, user friendly system, high speed, and high accuracy of palmprint recognition, it can be considered as one of the most reliable and suitable biometric recognition system. A lot of work has already been done about palmprint recognition, since it is a very interesting research area. However, more research is needed to obtain efficient palmprint system [4].

There are three groups of marks which are used in palmprint identification [5] as geometric features, line features (e.g., principle lines, wrinkles), and point features (e.g., minutiae points). A typical palmprint recognition system consists of palmprint acquisition, preprocessing, feature extraction, and matching phases [6].
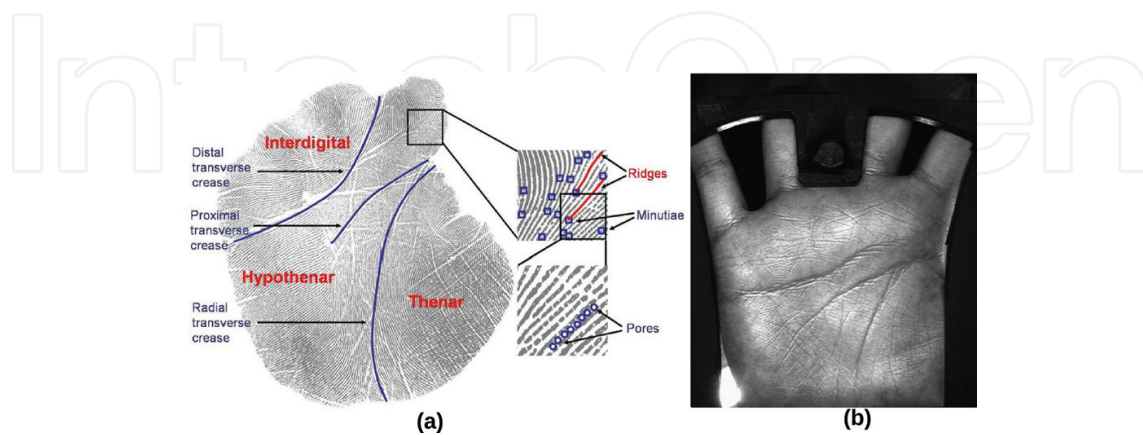


**Figure 4.** Palmprint features (a) a high-resolution image and (b) a low-resolution image.

### 3.4. Fingerprint

The modern history of fingerprint identification begins in the 19th century with the development of identification bureaus charged with keeping accurate records about indexed individuals. The acquisition of fingerprint was performed firstly by using ink technique [7].

The main application of fingerprint identification is forensic investigation of crimes. John Maloy performed a forensic identification in the late 1850s [8] by designing a high-security identification system that has always been the main goal in the security business.

The main reasons for the popularity of fingerprint recognition are as follows:

- The pattern of fingerprint is unique to each individual and immutable throughout life from infancy to old age and the patterns of no two hands resemble each other,

- Its success in various applications in the forensics, government, and civilian domains,

- The fact that criminals often leave their fingerprints at crime scenes,

- The existence of large legacy databases such as National Institute of Standards and Technology (NIST), Fingerprint Verification Competition (FVC) evaluation databases from 2000, 2002, and 2004.

- The availability of compact and relatively inexpensive fingerprint readers.

A typical fingerprint feature called minutiae is extracted from fingerprint images, as shown in **Figure 5**, and used for matching process for a fingerprint recognition system.

### 3.5. Ear

Recognizing people by their ear has recently received significant attention in the literature. There are many factors that made ear a widely used biometrics. First, the shape of the ear and the structure of cartilaginous tissue of the pinna are very discriminate. It is formed by the outer helix, the antihelix, the lobe, the tragus, the antitragus, and the concha. The ear recognition approaches are based on matching the distance of salient points on the pinna from a landmark location. Second, ear has a structure which does not vary with facial expressions or time, and it is very stable for the end of life. It has been shown
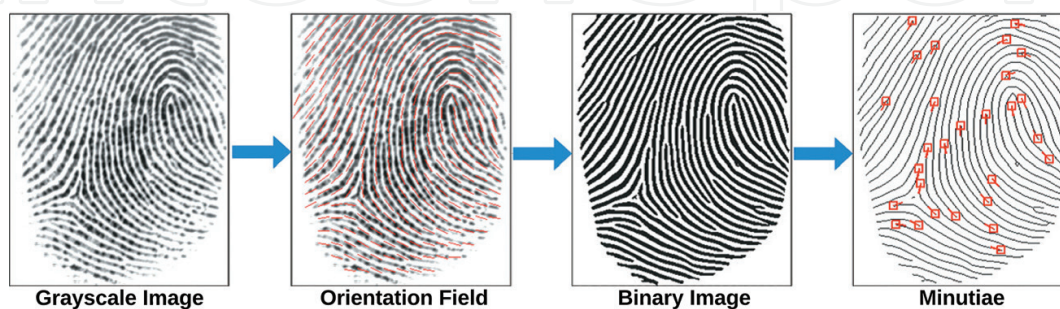


| Grayscale Image | Orientation Field | Binary Image | Minutiae |

**Figure 5.** A typical minutiae feature extraction algorithm [9].

that the recognition rate is not affected by aging [10]. Third, ear biometric is convenient as its acquisition is easy because the size of the ear is larger than fingerprint, iris, and retina and smaller than face. Ear data can also be captured even without the knowledge or cooperation of the user from far distance [11]; therefore, it can be used in passive environment. This makes ear recognition especially interesting for smart surveillance tasks and for forensic image analysis, because ear images can typically be extracted from profile head shots or video footage.

The main drawback of ear biometric is occlusion, where the ear can be partially or fully covered by hair or by other items such as head dress, hearing aids, jewelry, or headphone. In an active identification system, it is not a critical point as the subject can pull his or her hair back, but in a passive identification, it is a problem as there will be nobody informing the subject. Other challenges on ears are different poses (angles), left and right rotation, and different lighting conditions.

### 3.6. Speech

The activities of automatic speaker verification and identification have a long history going back to the early 1960s [12]. Dragon systems were the early applications that were used as speech recognizer [13], which focused on the ability of recognition system to provide acoustic knowledge about speaker. Baum-Welch HMM procedures were employed by these systems to train models.

Speech or voice is one of the behavioral traits that can be used in biometric systems to identify the user based on the stored voice in the enrollment phase, where the voice characteristics such as pronunciation style and voice texture are unique and distinctive for each person. On the other hand, voice can also be considered physiological in addition to behavioral feature based on the shape of the vocal track.

#### 3.6.1. Advantages and disadvantages of voice recognition

Generally, voice recognition is nonintrusive, and people are willing to accept a speech-based biometric system with as little inconvenience as possible. It also offers a cheap recognition technology, because general purpose voice recorders can be used to acquire the data. However, a person's voice can be easily recorded and can be used for authorized access, and the noise can be canceled by specific software. As a result, these make speech recognition to be used in many applications such as financial applications, security, retail, crime investigation, entertainment, etc.

Speech-based features are sensitive to a number of factors such as background noise, room reverberation, the channel through which the speech is acquired (such as cellular, land-line, and VoIP), overlapping speech, and Lombard or hyper-articulated speech. Additionally, the emotional and physical state of the speaker are important. An illness such as flu can change a person's voice, and it makes voice recognition difficult. Speech-based authentication is currently restricted to low-security applications because of high variability in an

individual's voice and poor accuracy performance of a typical speech-based authentication system. Existing techniques are able to reduce variability caused by additive noise or linear distortions, as well as compensating slowly varying linear channels [14].

### 3.6.2. Speech recognition

Speech recognition process starts by acquiring the sound from a user using microphone, and then, the series of acoustic signals are converted to a set of identifying words. The speech recognition depends on many factors such as language model, vocabulary size, speaking style, speaker enrollment, and transducer [15]. Speech recognition system is classified to "speaker dependent system," if the user should train the system before using it, and to "speaker independent system," if the system can recognize any speaker's speech without the need to train phase. Speech recognition systems can also be divided into "isolated word speech" or "continuous speech" based on the number of the used vocabularies for identification process.

Speaker models [16, 17] enable us to generate the scores from which we will make decisions. As in any pattern recognition problem, the choices are numerous, and the most popular and dominated technique in last two decade is Hidden Markov Models. There are also other techniques used for speech recognition systems such as Artificial Neural Networks (ANN), Back Propagation Algorithm (BPA), Fast Fourier Transform (FFT), Learn Vector Quantization (LVQ), and Neural Networks (NN). A typical speech recognition system is shown in **Figure 6**.

### 3.7. Performance evaluation of biometrics systems

Different measurements can be used to evaluate the performance of biometric systems. The most famous measurement is the recognition rate, which is defined as the percentage of the samples that are correctly matched samples to the total tested samples. Another popular measurement is False Reject Rate (FRR) versus False Accept Rate (FAR) at various threshold values, where FRR refers to the expected probability for two mate samples which
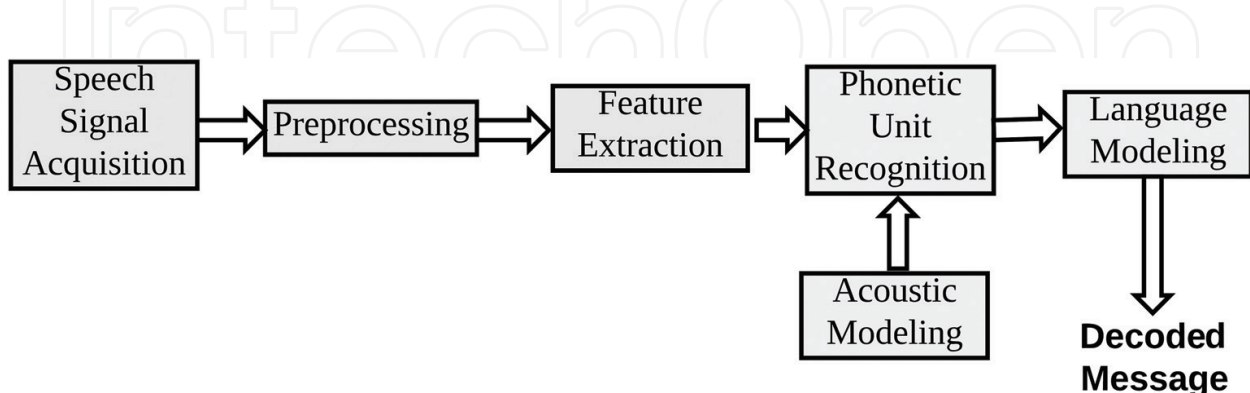


**Figure 6.** Block diagram of a speech recognition system.

are wrongly mismatched and FAR refers to the expected probability that two non-mate samples are incorrectly matched.

Single-valued measure "Equal Error Rate (EER)," that is threshold independent, can also be used to evaluate the performance of recognition systems. EER is the value, where FRR and FAR are equal.

Detection Error Trade-off (DET) or Receiver Operating Characteristic (ROC) curves are also used to compare the performance of biometric systems in which both curves plot FRR against FAR in the normal deviate and linear scale, respectively.

## 4. Biometric challenges

There are several challenges and key factors that can significantly affect the recognition performance as well as degrading the extraction of robust and discriminant features. Some of these challenges such as pose, illumination, aging, facial expression variations, and occlusions are briefly described below, and these challenges are illustrated in **Figure 7**.
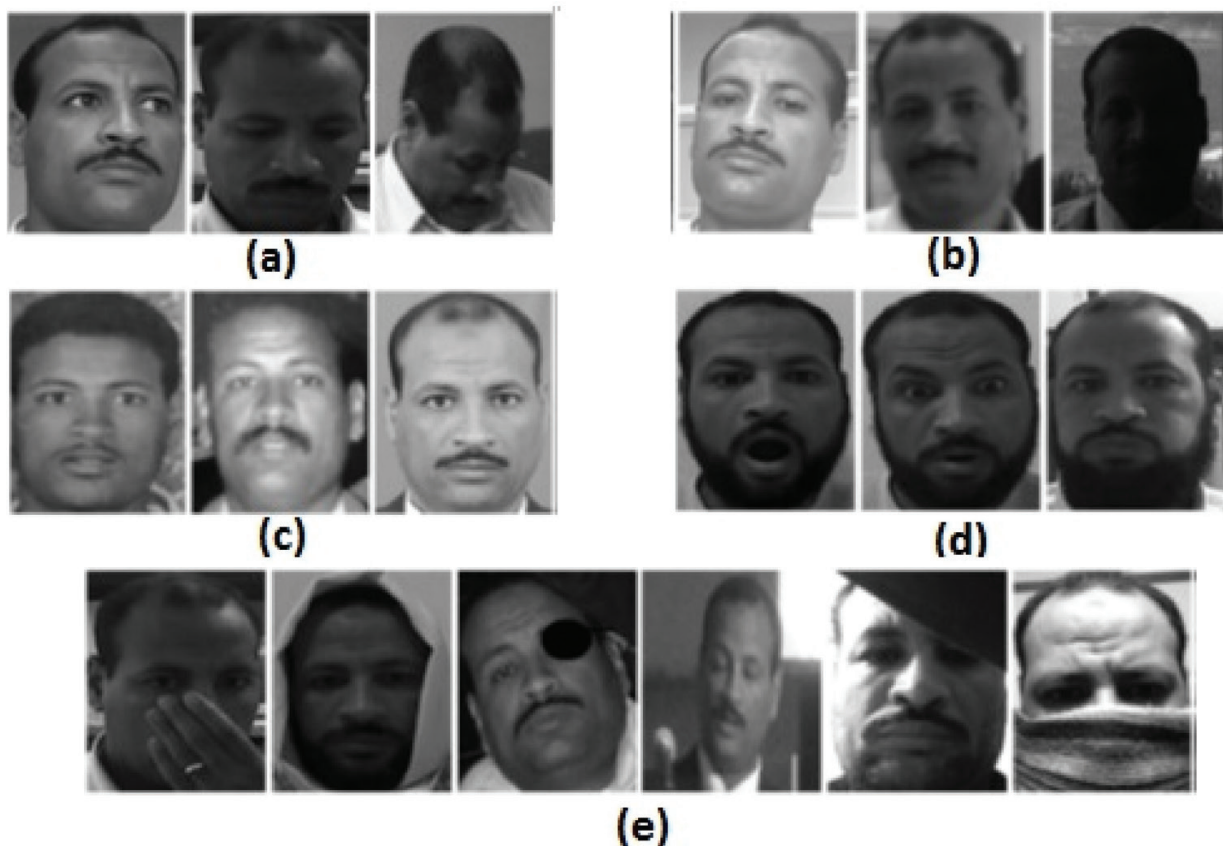


**Figure 7.** The challenges in the context of face recognition: (a) pose variations, (b) illumination variations, (c) aging variations, (d) facial expressions, (e) occlusions.

1. Pose variation: the images of a face or ear vary because of the camera pose (different viewpoints) as shown in **Figure 7a**. In this condition, some facial parts such as the eyes or nose may become partially or fully occluded. Pose variation has more influence on recognition process because of introducing projective deformations and self-occlusion. Thus, it is possible that images of the same person taken from two different poses may appear more different (intra-user variation) than images of two different people taken with the same poses (inter-user variation). There are many studies that deal with pose variation challenges in [18–20].

2. Illumination variation: when the image is captured, it may be affected by many factors to some degree. The appearance of the human face or ear is affected by factors such as lighting that includes spectra, source distribution, and intensity and also camera characteristics such as sensor response and lenses. Illumination variations can also have an effect on the appearance because of skin reflectance properties and the internal camera control [21]. The problem of illumination variation is considered to be one of the main technical challenges in biometric systems especially for face and ear traits, where the face of a person can appear dramatically different as shown in **Figure 7b**. In order to handle variations in lighting conditions or pose, an image relighting technique based on pose-robust albedo estimation [22] can be used to generate multiple frontal images of the same person with variable lighting.

3. Aging: aging can be a natural cause of age progression and an artificial cause of using make-up tools. Facial appearance changes more drastically at younger ages less than 18 years due to the change in subject's weight or stiffness of skin. All aging related variations such as wrinkles, speckles, skin tone, and shape degrade face recognition performance. One of the main reasons for the small number of studies concerning face recognition in the context of age factor was the absence of a public domain database for studying the effect of aging [23], since it was very difficult to collect a dataset for face images that contains images for the same subject taken at different ages along his/her life. An example set of images for different ages of the same person is presented in **Figure 7c**.

4. Occlusion: faces may be partially occluded by other objects such as scarf, hat, spectacles, beard, and mustache as shown in **Figure 7e**. This makes the face detection process a difficult task and the recognition itself might be difficult because of some hidden facial parts making features hard to be recognized. For these reasons, in surveillance and commercial applications, face recognition engines reject the images when some part of it is not detected. In the literature, local-feature based methods have been proposed to overcome these occlusion problems [24]. On the other hand, the iris could potentially be occluded due to the eyelashes, eyelids, shadows, or specular reflections, and these occlusions can lead to higher false non-match rates.

5. Facial expression: the appearance of faces is directly affected by a person's facial expression such as anger, surprise, and disgust as shown in **Figure 7d**. Additionally, facial hair such as beard and mustache can change facial appearance specifically near the mouth and

chin regions. Moreover, facial expression causes large intra-class variations. In order to handle these facial expression problems, local-feature-based approaches and 3D-model-based approaches are designed [25].

# 5. Human robot interaction (HRI)

Human-robot interaction (HRI) is the study of how people can interact with robots and to what extent robots are exploited and used for successful interaction with human beings. It could also be defined as a field of study dedicated to understanding, designing, and evaluating robotic systems for use by or with humans. In general, the interaction is based on the communication with or reaction to each other, either people or things as shown in **Figure 8**.

### 5.1. The importance and the role of person identification in human-robot interaction

Person identification is a very important function for robots, which work with humans in the real world [26]. Human identification by robot may enhance the extent of interaction and
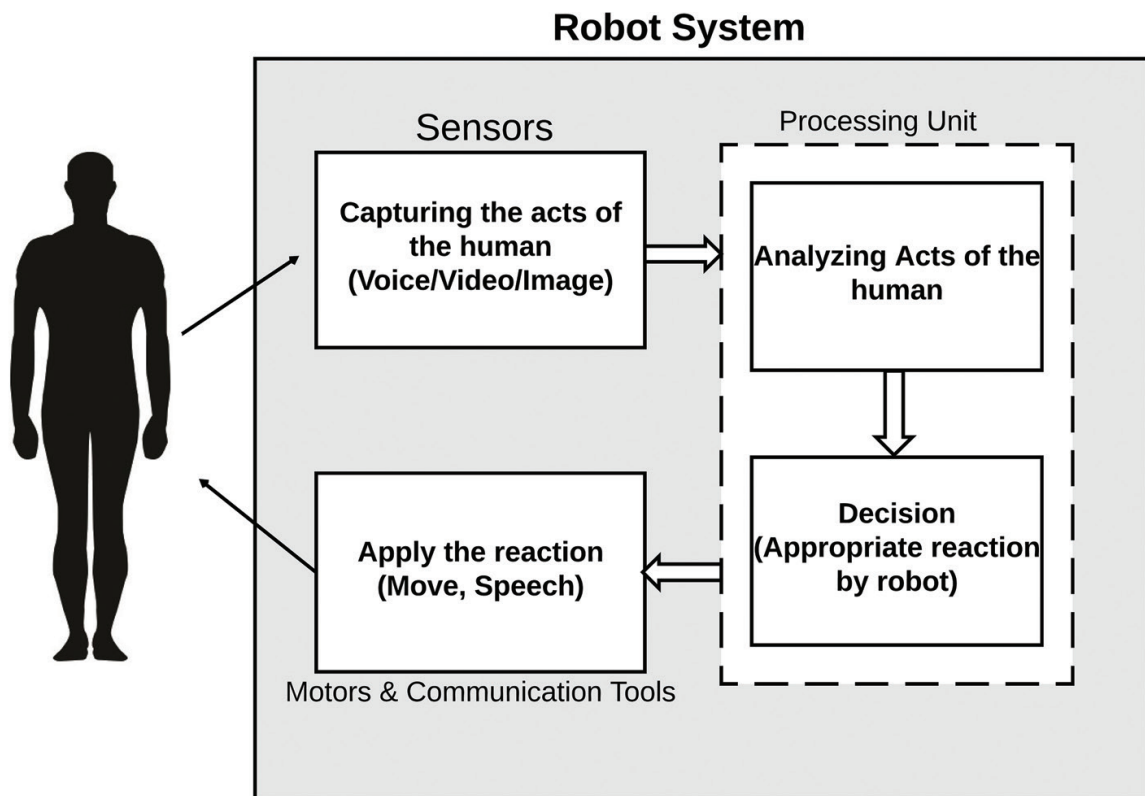


**Figure 8.** Block diagram of a human-robot interaction system.

communication with each other, where identifying the user does not only require ID but also many other information such as age, gender, interests/hobbies, and language of each user. Knowing the age of the user will help the robot to choose the tone of voice, where child may prefer childish voice tone instead of the manly voice and vice versa. Calling "Mr, Ms, Sir, Madam" when communicating with a person is based on gender, which is also important. Additionally, identifying the interest/hobby of the user will highly enhance the interaction, since it is not acceptable to discuss boxing with a person whose interest is ballet. In addition, communicating with a person using his/her original language ensures promotion of the interaction.

### 5.2. The most appropriate biometric traits of a person that can easily be identified by robot

Interaction depends on the extent of communication between robots and humans. Human and a robot can construct a communication between each other using several forms. Proximity to each other is the main factor that impacts the communication forms between human and robot. Thus, communication and interaction can be classified into two general categories [27]:

- Remote interaction: the human and the robot are not at the same place and are separated spatially or even temporally (different rooms, countries, or planets)

- Proximate interaction: the humans and the robots are collocated (same room)

Choosing biometric traits that robot should use to identify the user should be compatible with the aforementioned interaction categories. For the remote interaction, the biometric traits whose raw features are images such as face, ear, and iris are not convenient choices, since the majority of remote interaction is conducted by voice communication. Therefore, speech recognition may be the best choice, since it is suitable for direct (different room) and mobile calling. For proximate interaction (face-to-face interaction) and in order to create more real interaction, identification process should use a biometric trait that does not require direct contact with the user in order to capture the biometric traits such as face, ear, and voice, which are captured from a far distance.

## 6. Multibiometric systems

Some of the limitations imposed by unimodal biometric systems (that is, biometric systems that rely on the evidence of a single biometric trait) can be overcome by using multiple biometric modalities. Increasing the discriminant information and constraints leads to decrease the error in recognition process. More information can be acquired when using different sources of information simultaneously, and the sources of information may be on several types such as multiple biometric traits, algorithms, instances, samples, and sensors. Various scenarios in a multimodal biometric system are demonstrated on **Figure 9**.
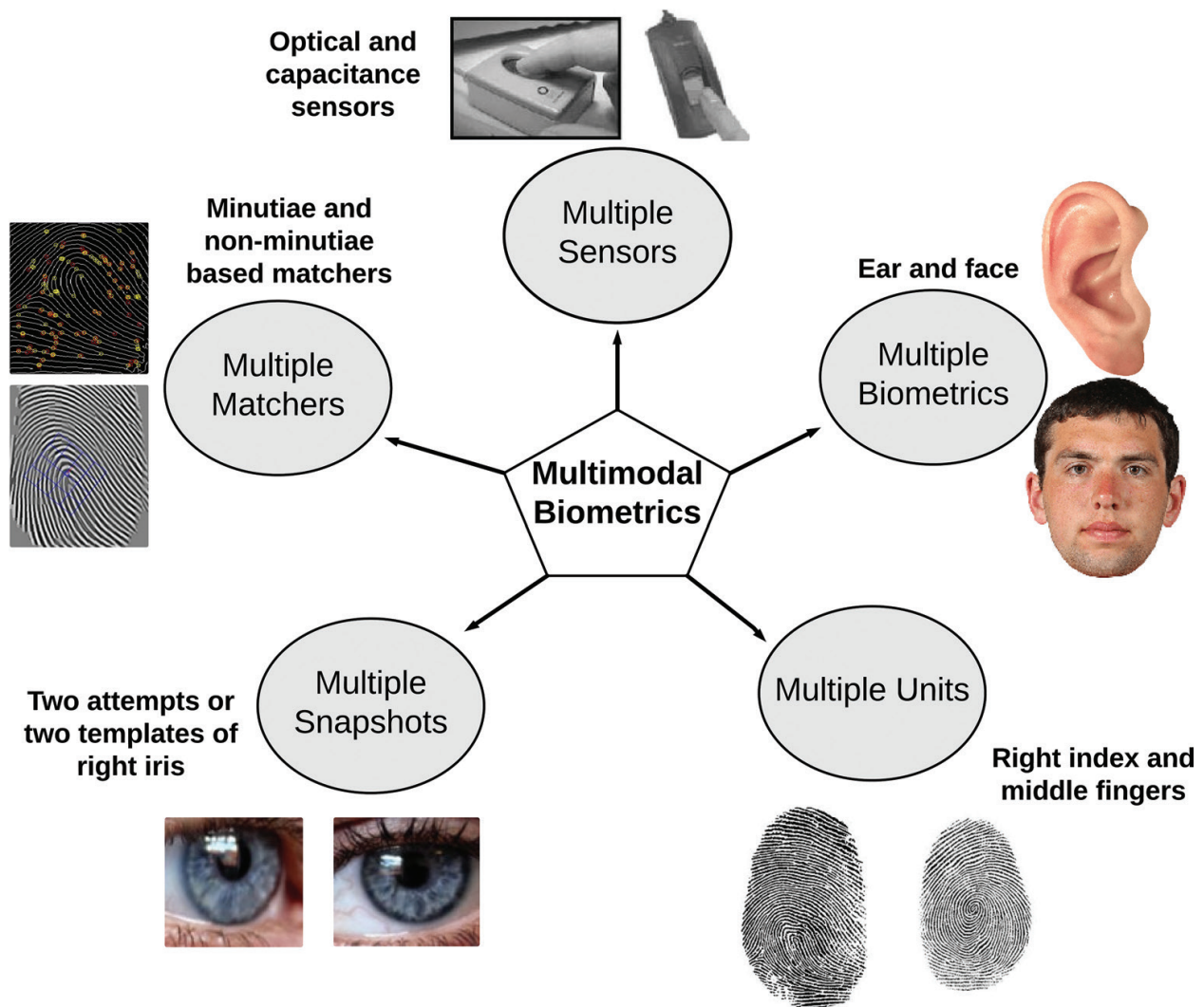
**Figure 9.** Various scenarios in a multimodal biometric system.

Consolidating multiple features that are acquired from different biometric sources in order to construct a person recognition system is defined as multibiometric systems. For example, fingerprint and palmprint traits, or right and left iris of an individual, or two different samples of the same ear trait may be fused together to recognize the person more accurate and reliable than unimodal biometric systems. Due to the use of more than one biometric source, many of the limitations of unimodal systems can be overcome by the multimodal biometric systems [28].

Multibiometric systems are able to compensate a shortage of any source using the other source of information. In addition, the difficulty of circumvention of multiple biometric sources simultaneously creates more reliable systems than unimodal systems. On the other hand, the unimodal biometric systems are low cost and require less enrollment and recognition time compared to multimodal systems. Hence, it is essential to carefully analyze the tradeoff between the added cost and the benefits earned when making a business case for

the use of multibiometrics in a specific application such as commercial, forensics, and the biometric systems that include large population.

The information used in recognition process can be fused in five different levels [29]:

1. Sensor level fusion: information of the individual is captured by multiple sensors in order to generate new data that is afterward subjected to feature extraction phase. For instance, in the case of iris biometrics, samples from "Panasonic BM-ET 330" and "LG IrisAccess 4000" sensors may be fused to obtain one sample.

2. Feature level fusion: in this level, the extracted features from multiple biometric sources are fused to obtain a single feature vector that contains rich biometric information about a client. Integration at feature level is expected to offer good recognition accuracy because it detects the correlated feature values generated by different biometric algorithms, thereby identifying a set of distinguished features.

3. Score level fusion: it is the most commonly used fusion technique due to the ease of performing a fusion of the match scores in multibiometric systems. Match scores of multiple classifiers are integrated in score-level fusion to produce a single match score, which is used to get a final decision. Score level fusion requires performing score normalization, which converts the scores into common scale. The fused match score is then calculated by three categories, namely likelyhood ratio–based score fusion, transformation-based score fusion, and classifier-based score fusion.

4. Rank level fusion: it is defined as consolidating associated ranks of multiple classifiers in order to derive consensus rank of each identity to establish the final decision. Rank-level fusion provides less information compared to score level fusion, and it is relevant in iden-tification mode. The final decision of rank-level fusion is obtained by three well-known methods namely Highest Rank, Borda Count, and Logistic Regression methods.

5. Decision level fusion: the outputs (decisions) of different matchers may be fused to obtain a single/final decision (genuine or imposter in a verification system or the identity of the client in an identification system). A single class label can be obtained by employing tech-niques like majority voting, behavior knowledge space, etc.

Among the aforementioned fusion techniques, the most popular ones are score-level fusion and feature-level fusion. Most of the person identification systems use these fusion techniques because of their simplicity and high performance. These systems are compared in **Table 1** by demonstrating many details of the state-of-the-art multibiometric systems.

The results shown in **Table 1** prove that consolidation of different unimodal biometric sys-tems construct a recognition system that is robust against many challenges such as occlusion, pose, and nonuniform illumination. Additionally, the studies presented in **Table 1** demon-strate that score-level fusion of more than one biometric trait overcomes the limitations of unimodal biometric systems, and in most of the studies, score-level fusion results outperform feature-level fusion results for person identification.

| Identification approach | Biometric traits | Databases and challenges | Fusion strategy | Recognition rate (%) |
|---|---|---|---|---|
| Toygar et al. [30] | Face Voice | *XM2VTS:* (P) *BANCA:* (P, I, E, O, N) | Score-level fusion | *XM2VTS:* Voice: 78.01 Face: 86.53 Face + Voice: 94.24 *BANCA:* Voice: 91.54 Face: 92.07 Face + Voice: 97.43 |
| Eskandari and Toygar [31] | Iris Face | *CASIA-Iris_Distance:* (I, O, N, D) *FERET, ORL, BANCA (used for weight optimization):* (P, I, E, O, N) *UBIRIS (used for weight optimization):* (I, O, N) | Feature-level and Score-level fusion | *CASIA-Iris_Distance:* Face: 92.77 Iris: 77.65 Face + Iris: 98.66 |
| Farmanbar and Toygar [32] | Palmprint Face | *FERET:* (P, I, E) *PolyU:* (P) | Feature-level and Score-level fusion | *FERET ± PolyU:* Palmprint: 94.30 Face: 83.21 Palmprint + Face: 99.17 |
| Hezil and Boukrouche [33] | Ear Palmprint | *IITDelhi-2 Ear IITDelhi Palmprint* | Feature-Level Fusion | *IITDelhi-2 Ear ± IITDelhi Palmprint* Palmprint: 97.73 Ear: 98.9 Palmprint + Ear: 100 |
| Ghoualmi et al. [34] | Iris Ear | *CASIA IrisV1 USTB 2* (P,I) | Feature-Level Fusion | *CASIA IrisV1 ± USTB-2* Iris: 95.8 Ear: 91.36 Iris + Ear: 99.67 |
| Telgad et al. [35] | Face Fingerprint | *FVC 2004* | Score-level fusion | *FVC 2004:* Face-PCA: 92.4 Fingerprint-Minutiae: 93.05 Fingerprint-Gabor Filter: 95 Face + Fingerprint: 97.5 |
| Patil and Bhalke [36] | Fingerprint Palmprint Iris | *FVC IITD CASIA* | Score-level fusion | *FVC ± IITD ± CASIA* Fingerprint: 72.73 Plamprint: 65.57 Iris: 80 Fingerprint + Palmprint + Iris = 95.23 |

P, pose; I, illumination; E, expression; O, occlusion; N, noise; D, distance.

**Table 1.** Comparison of person identification approaches using multimodal biometric traits under different challenges.

## 7. Fusion of face and speech traits

Based on the purpose of the robot, a unimodal or a multimodal recognition system could be selected to be used for human-robot interaction. For example, a military purpose robot should be more accurate than home purpose robot. As mentioned in Section 5.2, the common trait that can be used for human identification by robot in both remote and proximate interaction is voice biometric trait. On the other hand, the face is the most realistic biometric trait in case of proximate interaction.

It will be appropriate to fuse face and voice in human-robot interaction, since both of these traits are noncontacted and the user is unaware that recognition is being performed. Many studies proved that the fusion of face and speech is appropriate for many purposes [37–39], where face and speech are the best choices since both of them do not need physical or direct contact with sensors [40, 41]. Another advantage of speech over face is that speech can be recognized even when a human and robot are not found in the same physical place. This is useful for voice recognition purposes by mobile phone or when a user and robot are in two different rooms in the same place. Consequently, a realistic human-robot interaction system is achieved, either HRI is conducted by face-to-face, blind, or invisible interaction.

## 8. Conclusion

Multimodal biometrics in the context of human-robot interaction is discussed under different challenges. The most commonly used biometric traits namely face, iris, fingerprint, ear, palmprint, and voice are discussed in this chapter. Various challenges such as pose, illumination, expression, aging variations, and occlusion are explained, and many state-of-the-art biometric systems involving these challenges are presented and compared. The comparison of these systems shows that multimodal biometrics overcomes the limitations of unimodal systems and achieves better person identification performance. Additionally, score-level fusion technique applied on more than one biometric trait obtains higher recognition rates for person identification. On the other hand, fusion of face and speech is an appropriate choice for human-robot interaction, since the enrollment phase of face and speech biometric systems does not require physical or direct contact with sensors. The face image or speech of a person can be captured by a robot, even if the person is far away from the robot.

## Author details

Önsen Toygar*, Esraa Alqaralleh and Ayman Afaneh

*Address all correspondence to: onsen.toygar@emu.edu.tr

Computer Engineering Department, Faculty of Engineering, Eastern Mediterranean University, Famagusta, North Cyprus, via Mersin, Turkey

# References

[1] Jain AK, Ross AA, Nandakumar K. Introduction to Biometrics. Springer; 2011. 312 p. DOI: 10.1007/978-0-387-773261

[2] Takeo K. Picture Processing by Computer Complex and Recognition of Human Faces [Thesis]. Kyoto, Japan: Dept. of Science, Kyoto University; 1974. 143 p . DOI: 10.14989/doctor.k1486Available from: http://hdl.handle.net/2433/162079

[3] Bowyer KW, Hollingsworth K, Flynn PJ. Image understanding for iris biometrics: A survey. Computer Vision and Image Understanding. 2008;**110**(2):281-307. DOI: 10.1016/j.cviu.2007.08.005

[4] Zhang D, Kong WK, You J, Wong M. Online palmprint identification. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2003;**25**(9):1041-1050. DOI: 10.1109/TPAMI.2003.1227981

[5] Raut SD, Humbe VT. Biometric palm prints feature matching for person identification. International Journal of Modern Education and Computer Science. 2012;**4**(11):61. DOI: 10.5815/ijmecs.2012.11.06

[6] Kong A, Zhang D, Kamel M. A survey of palmprint recognition. Pattern Recognition. 2009;**42**(7):1408-1418. DOI: 10.1016/j.patcog.2009.01.018

[7] Berry JS, David A. The history and development of fingerprinting. In: Lee HC, Ramotowski R, Gaensslen RE, editors. Advances in fingerprint Technology. 2nd ed. CRC press; 2001:13-52. DOI: 10.1201/9781420041347.ch1

[8] Cole S, Col A. Suspect Identities: A History of Fingerprinting and Criminal Identification. Cambridge, Mass /London: Harvard University Press; 30-10-2009.38 p. DOI: 10.1023/B:MESC.0000005857.89878.a9

[9] Jain AK, Prabhakar S, Hong L, Pankanti S. Filterbank-based fingerprint matching. IEEE Transactions on Image Processing. 2000;**9**(5):846-859. DOI: 10.1109/83.841531

[10] Mina I. Mark N. Sasan M. The effect of time on ear biometrics. In: International Joint Conference on Biometrics (IJCB), 2011; Washington, DC, USA. IEEE; 2011. p. 1-6. DOI: 10.1109/IJCB.2011.6117584

[11] Pflug A, Busch C. Ear biometrics: A survey of detection, feature extraction and recognition methods. IET Biometrics. 2012;**1**(2):114-129. DOI: 10.1049/iet-bmt.2011.0003

[12] Pruzansky S, Mathews MV. Talker-recognition procedure based on analysis of variance. The Journal of the Acoustical Society of America. 1964;**36**(11):2041-2047. DOI: 10.1121/1.1795335. PACS

[13] Peskin B et al. Topic and speaker identification via large vocabulary continuous speech recognition. In: Proceedings of the workshop on Human Language Technology. Association

for Computational Linguistics: Stroudsburg, PA, USA; 1993. p. 119-124. DOI: 10.3115/1075671.1075697

[14] Yu D, Deng L, Droppo J, Wu J, Gong Y, Acero A. Robust speech recognition using a cepstral minimum-mean-square-error-motivated noise suppressor. IEEE Transactions on Audio, Speech and Language Processing. 2008;**16**(5):1061-1070. DOI: 10.1109/TASL.2008.921761

[15] Varile GB, Zampolli A. Survey of the State of the Art in Human Language Technology. Linguistica Computazionale. Cambridge University Press; 1997. 413 p. DOI: 10.1.1.366.9300

[16] Cowling M, Sitte R. Comparison of techniques for environmental sound recognition. Pattern Recognition Letters. 2003;**24**(15):2895-2907. DOI: 10.1016/S0167-8655(03)00147-8

[17] Furui S. Recent advances in speaker recognition. Pattern Recognition Letters. 1997;**18**(9):859-872. DOI: 10.1016/S0167-8655(97)00073-1

[18] Blanz V, Grother P, Phillips PJ, Vetter T. Face recognition based on frontal views generated from non-frontal images. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR; 2005. p. 454-461

[19] Prince SJ, Elder JH, Warrell J, Felisberti FM. Tied factor analysis for face recognition across large pose differences. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2008;**30**(6):970-984. DOI: 10.1109/TPAMI.2008.48

[20] Asthana A, Marks TK, Jones MJ, Tieu KH, Rohith MV. Fully automatic pose-invariant face recognition via 3D pose normalization. In: IEEE International Conference on Computer Vision (ICCV), 2011; Barcelona, Spain. IEEE; 2011. p. 937-944. DOI: 10.1109/ICCV.2011.6126336

[21] Liu DH, Lam KM, Shen LS. Illumination invariant face recognition. Pattern Recognition. 2005;**38**(10):1705-1716. DOI: https://doi.org/10.1016/j.patcog.2005.03.009

[22] Patel VM, Wu T, Biswas S, Phillips PJ, Chellappa R. Dictionary-based face recognition under variable lighting and pose. IEEE Transactions on Information Forensics and Security. 2012;**7**(3):954-965. DOI: 10.1109/TIFS.2012.2189205

[23] Park U, Tong Y, Jain AK. Age-invariant face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2010;**32**(5):947-954. DOI: 10.1109/TPAMI.2010.14

[24] Tan X, Chen S, Zhou ZH, Liu J. Face recognition under occlusions and variant expressions with partial similarity. IEEE Transactions on Information Forensics and Security. 2009;**4**(2):217-230. DOI: 10.1109/TIFS.2009.2020772

[25] Levine MD, Yu Y. Face recognition subject to variations in facial expression, illumination and pose using correlation filters. Computer Vision and Image Understanding. 2006;**104**(1):1-15. DOI: 10.1016/j.cviu.2006.06.004

[26] Fukui K, Yamaguchi O. Face recognition using multi-viewpoint patterns for robot vision. Robotics Research. 2005;**15**:192-201. DOI: 10.1.1.474.7008

[27] Goodrich MA, Schultz AC. Human-robot interaction: A survey. Foundations and trends (r) in human-computer interaction. 2007;**1**(3):203-275. DOI: 10.1561/1100000005

[28] Jain AK, Ross A. Multibiometric systems. Communications of the ACM. 2004;**47**(1):34-40. DOI: 10.1145/962081.962102

[29] Ross A, Govindarajan R. Feature level fusion using hand and face biometrics. In: Jain AK, Ratha NK, editors. Proceedings of SPIE Conference on Biometric Technology for Human Identification II. Orlando, USA; 2005. p. 196-204. DOI: org/10.1117/12.606093

[30] Toygar Ö, Ergün C, Altinçay H. Using local features based face experts in multimodal biometric identification systems. In: Fifth International Conference on Soft Computing, Computing with Words and Perceptions in System Analysis, Decision and Control, 2009. ICSCCW 2009. 2009. p. 1-4

[31] Eskandari M, Toygar Ö. Selection of optimized features and weights on face-iris fusion using distance images. Computer Vision and Image Understanding. 2015;**137**:63-75. DOI: 10.1016/j.cviu.2015.02.011

[32] Farmanbar M, Toygar Ö. Feature selection for the fusion of face and palmprint biometrics. Signal, Image and Video Processing. 2016;**10**(5):951-958. DOI: 10.1007/s11760-015-0845-6

[33] Hezil N, Boukrouche A. Multimodal biometric recognition using human ear and palmprint. IET Biometrics. 2017:9. DOI: 10.1049/iet-bmt.2016.0072

[34] Ghoualmi L, Chikhi S, Draa A. A SIFT-based feature level fusion of iris and ear biometrics. In: Schwenker F, Scherer S, Morency LP, editors. Multimodal Pattern Recognition of Social Signals in Human-Computer-Interaction. MPRSS 2014. Lecture Notes in Computer Science, vol 8869. Springer, Cham; 2014:102-112. DOI: 10.1007/978-3-319-14899-1_10

[35] Telgad RL, Deshmukh PD, Siddiqui AM. Combination approach to score level fusion for Multimodal Biometric system by using face and fingerprint. In IEEE: International Conference on Recent Advances and Innovations in Engineering (ICRAIE), 9-11 May 2014; Jaipur, India. IEEE; p. 1-8. DOI: 10.1109/ICRAIE.2014.6909320

[36] Patil AP, Bhalke DG. Fusion of fingerprint, palmprint and iris for person identification. In IEEE: Automatic Control and Dynamic Optimization Techniques (ICACDOT), International Conference on; 9-10 Sep 2016; Pune, India. IEEE; p. 960-963. DOI: 10.1109/ICACDOT.2016.7877730

[37] Soltane M, Doghmane N, Guersi N. Face and speech based multi-modal biometric authentication. International Journal of Advanced Science and Technology. 2010;**21**(6):41-56

[38] Ben-Yacoub S, Abdeljaoued Y, Mayoraz E. Fusion of face and speech data for person identity verification. IEEE Transactions on Neural Networks. 1999;**10**(5):1065-1074. DOI: 10.1109/72.788647

[39] Jain AK, Hong L, Kulkarni Y. A multimodal biometric system using fingerprint, face and speech. In: Proceedings of 2nd Int'l Conference on Audio- and Video-Based Biometric Person Authentication, Washington DC. 1999. p. 182-187

[40] Demirel H, Anbarjafari G. Probability Distribution Functions Based Face Recognition System Using Discrete Wavelet Subbands. In: Olkkonen JT, editor. Discrete Wavelet Transforms-Theory and Applications. ISBN: 978-953-307-185-5, InTech, Available from: http://www.intechopen.com/books/discrete-wavelet-transforms-theory-and-applications/probability-distribution-functions-based-face-recognition-system-using-discrete-wavelet-subbands; 2011

[41] Maucec MS, Zgank A. Speech recognition system of Slovenian broadcast news. In: Ipsic I, editor. Speech Technologies. InTech, DOI: 10.5772/17161. Available from: https://www.intechopen.com/books/speech-technologies/speech-recognition-system-of-slovenian-broadcast-news; 2011