CHAPTER 3
AN APPROACH FOR FACE AND EAR MULTIMODAL SYSTEM

The successful installation of biometric systems in different civilian applications does not entail that biometrics is a completely solved problem. Each biometric has its strengths and weaknesses, the match between a specific biometric and an application is determined depending upon the operational mode of the application and the biometric properties.

3.1 ISSUES IN SINGLE MODAL BIOMETRIC

Single modal biometric characteristics have plenty of error rates and it may not achieve the desired performance requirements of all the applications. There are many factors which degrades the recognition performance. Researchers are addressing to enhance the usability of biometric system. Some of the issues imposed by single modal biometric are given below:

Noise in sensed data

Biometric system has different sensed data. The sensed data might be noisy or deformed. A voice altered by cold, iris recognition with wearing glasses. Finger print with a scar, might be too oily, dry, wet or damaged temporally or permanently. Face sensed weaknesses due to variations in light, pose or illumination. Gait sensed with fluctuation in body weights. Noisy sensed biometric data may be false matched with templates in database resulting is a false rejection (Ross et al., 2011).

Distinctiveness

Biometric trait is expected to vary significantly across two individuals. The characteristics of the individuals are represents with the large inter-class similarity in the feature sets. The information content (number of distinguishable patterns) in two of the most commonly used representations of hand geometry and face are only the order of feature (Golfarelli et al., 1997).
Non-Universality

Problems regarding the quality or consistency of the capture of biometric data may not necessarily due to a fault or error in the sensor (Mane et al., 2006). About 4% of the population may have poor quality fingerprints, due to scars or cuts and it shows erroneous result. Intra-class variations, the biometric data acquired from an person during testing may be different from template data during enrollment. The users are incorrectly interacting with the sensor. This may affect the matching result.

Spoof attacks

Biometric traits of the legitimate user are enrolled in the template database; an imposter may attempt to spoof the sensed data of the biometric system when the traits such as signature (Harrison et al., 1958) and voice (Eriksson et al., 1997) are used. The fingerprint traits can also be spoof with the artificial fingers/fingerprint to thwart a fingerprint verification system (Matsumoto et al., 2002).

Replay attacks

A valid biometric data is maliciously repeated. This is carried out either by the originator or by an adversary who intercepts or acquires a digital copy of the stored biometric sample and retransmits this bypassing the biometric sensor (Sowkarthika et al., 2012).

Substitution attack

The biometric template must be stored to allow user verification. If an attacker gets an access to the storage, either local or remote, anyone can overwrite the legitimate user’s template with his/her own – in essence, stealing their identity.

Tampering

Feature sets on verification or in the templates can be modified in order to obtain a high verification score, no matter which image is presented to the system.

Masquerade attack

A digital “artefact” image can be created from a fingerprint template, so that this artefact, if submitted to the system, will produce a match. The artefact may not even resemble the original image. This attack poses a real threat to the remote authentication
systems (e.g. via the Web), since an attacker does not even have to bother to acquire a
genuine biometric sample. All he needs is just to gain an access to the templates stored on a
remote server (this perfectly fits a description of a typical hacker operating from a rat hole).

**Trojan horse attacks**

Some parts of the system, e.g. a matcher, can be replaced by a Trojan horse
program that always outputs high verification scores.

**Overriding Yes/No response**

An inherent flaw of existing biometric systems is due to the fact that the output of
the system is always a binary Yes/No (i.e., match/no match) response.

High FRR causes inconvenience for legitimate users and prompts the system administrator to lower a verification threshold. This inevitably gives rise to FAR, which, in turn, lowers the security level of the system.

**3.2 LEVEL OF FUSIONS**

The Biometric fusion is the technique to integrate the classification results from
each biometric channel. Multimodal biometric fusion combines the aspect from various
biometric features to improve the strengths and reduce the limitations of the individual aspects. Fig. 3.1 shows the levels of fusions. The efficiency of the fusion scheme greatly influences the accuracy of a multimodal biometric system. The various levels of fusions are:

**Sensor level fusion**

The raw data obtained from multiple sensors can be practiced and merged to
generate new biometric data from which trait can be extracted. Biometric traits from
different sensors like fingerprint, video camera, iris scanner, digital signature etc, are fused to form biometric trait to process. Sensing a speech signal concurrently with two various microphones may be fused and then be subjected to feature extraction and matching. Sensor level is projected to improve the recognition accuracy; it remains possible problems related with unimodal biometric system because of incompatibility of data from various modalities (Karthik *et al.*, 2008).
Fig. 3.1 Level of Fusions (a) Sensor Level fusion, (b) Feature Level fusion, (c) Match Score Level fusion, (d) Rank Level fusion, and (e) Decision Level fusion.

**Feature level fusion**

The feature sets are extracted from different biometric channels can be fused using specific fusion algorithm to form a composite feature set. The feature collections of different modalities agree to extract a minimal feature set from the high-dimensional feature vector.

The feature vectors extracted from the face and ear modalities can be fused is an example of multimodal system. The feature level fusion is the extraction of correlated feature from the different modalities and in course identifies a prominent set of features that can improve recognition accuracy (Maple *et al.* 2006; Thiran *et al.*, 2009). The feature level fusion is likely to achieve superior result in comparison with score level and decision level fusion.

**Match score level fusion**

Feature vectors are generated separately for each modality. Extracted feature vectors compared with the templates residing in the database individually for each biometric trait to generate match scores. Depending on the accuracy of each biometric channel, output set of match scores which are fusion to create composite matching score. As an example, face and hand modalities match score may be combined by the use of simple sum rule in order to obtain a new match score which is then pass to the decision module (Ross *et al.*, 2003).

**Rank level fusion**

Rank level fusion is a new fusion approach where each classifier associates a rank with every enrolled identity. Fusion involves consolidating the rank output by individual biometric subsystems and determining a new rank that would support in establishing the final decision. However, these fusions have one weakness. In multimodal biometric, more different identities output from two or three matching modules which are designed to appear some identities of only one matcher. In this case, the rank level fusion shows the risk of wrong results (Monwar *et al.*, 2009).
**Decision level fusion**

In multimodal biometric system, the final decision is based on the separate decision of different modalities using techniques such as majority voting, behavior knowledge space, weighted voting, AND rule, OR rule (Yi et al., 2010). Decision level fusion is least powerful due to availability of inadequate information and limits the basis for enhancing the system accuracy.

**3.3 BIOMETRIC SYSTEM SECURITY**

Biometric system use physiological or behavioral traits are being increasingly utilized in many applications to enhance the security of physical and logical access systems. The factors to be used for biometric security system are given below:

**Universality:** Universality is that every person using a system should have the characteristics. It is really difficult to get 100% overage. There are mute people, people without fingers or with injured eyes. All these cases must be handled.

**Uniqueness:** The biometric trait should be sufficiently different for individuals in the relevant population such that they can be distinguished from one another. Fingerprints have a high discrimination rate and the probability of two persons with the same iris is estimated as low as 1: 1052. Identical twins, on the other side, cannot be easily distinguished by face recognition and DNA-analysis systems.

**Permanence:** It relates to the manner in which the characteristics should be invariant with time. While the iris usually remains stable over decades, a person’s face changes significantly with time (Valarmathy et al., 2013). The signature and its dynamics may change as well and the finger is a frequent subject to injuries.

**Collectability:** The characteristics must be measured quantitatively and obtaining the characteristics should be easy. Face recognition systems are not intrusive and obtaining of a face image is easy. In the contrast the DNA analysis requires a blood or other bodily sample. The retina scan is rather intrusive as well.

**Performance:** It refers to the achievable identification/verification accuracy and the re-sources and working or environmental conditions needed to achieve an acceptable accuracy. The crossover accuracy of iris-based systems is under 1% and the system is able
to compare over $4 \times 10^6$ iris codes in one second. The crossover accuracy of some signature dynamics systems is as high as 25% and the verification decision takes over one second.

**Acceptability:** It indicates to what extent people are willing to accept the biometric system. Face recognition systems are personally not intrusive, but there are a country where taking pictures of persons is not viable (Lifang et al., 2011). The retina scanner requires an infrared laser beam directed through the cornea of the eye. This is rather invasive and only few users accept this technology.

**Circumvention:** It relates to how difficult it is to fool the system by fraudulent techniques. An automated access control system that can be easily fooled with a fingerprint model or a picture of a user’s face does not provide much security.

### 3.4 BIOMETRIC MODALITIES

A biometric system is essentially a pattern-recognition system that recognizes a person based on a feature vector derived from a specific physiological or behavioral characteristic that the person possesses (Hao et al., 2006; Jerome et al., 2010). The commonly used biometric traits are shown in the fig. 3.2.

**Face:** Humans have a remarkable ability to recognize fellow beings based on facial appearance. So, face is a natural human trait for automated biometric recognition. Face recognition systems typically utilize the spatial relationship among the locations of facial features such as eyes, nose, lips, chin, and the global appearance of a face (Brendan et al., 2012).

The forensic and civilian applications of face recognition technologies pose a number of technical challenges both for static mug-shot photograph matching (e.g., for ensuring that the same person is not requesting multiple passports) to unconstrained video streams acquired in visible or near-infrared illumination (e.g., in surveillance) (Ichino et al., 2006). The problems associated with illumination, gesture, facial makeup, occlusion, and pose variations adversely affect the face recognition performance (Nageshkumar et al., 2009). While face recognition is non-intrusive, has high user acceptance, and provides acceptable levels of recognition performance in controlled environments, robust face recognition in non-ideal situations continues to pose
challenges. Facial recognition tools can be improved by training on a set of synthetic facial expressions and appearance/environment variations generated from real facial images (Turk et al., 1991).

**Fingerprint:** Fingerprint-based recognition has been the longest serving, most successful and popular method for person identification. Fingerprints consist of a regular texture pattern composed of ridges and valleys. These ridges are characterized by several landmark points, known as minutiae, which are mostly in the form of ridge endings and ridge bifurcations. The spatial distribution of these minutiae points is claimed to be unique to each finger; it is the collection of minutiae points in a fingerprint that is primarily employed for matching two fingerprints (Youn et al., 2008). In addition to minutiae points, there are sweat pores and other details which can be acquired in high resolution (1000 ppi) fingerprint images. These extended features are receiving increased attention since forensics experts seem to utilize them particularly for latent and poor quality fingerprint images. Nearly all forensics and law enforcement agencies worldwide utilize Automatic Fingerprint Identification Systems (AFIS). In a recent development, the vitality of the appendage pressed onto the sensor can be determined by detecting the perspiration between the ridges, this helps combat faked latex fingerprints and the use of high resolution printed images of fingers, which still affects some cheaper systems (Karaboga et al., 2008).

**Iris Recognition:** The visual texture of the human iris is determined by the chaotic morphogenetic processes during embryonic development and is supposed to be unique for each person and each eye. The iris is a protected internal organ whose texture is stable and distinctive, even among identical twins, and extremely difficult to surgically spoof (Ross et al., 2006). An excellent survey on the current iris recognition technologies and future research challenges is available (Ross et al., 2004). An iris image is typically acquired using a non-contact imaging process: capturing an iris image involves cooperation from the user, both to register the image of iris in the central imaging area and to ensure that the iris is at a predetermined distance from the focal plane of the camera first invented by John Daugman, both the accuracy and matching speed of currently available iris recognition systems is very high. Primarily used for authentication, Iris recognition involves a high resolution picture being taken of a subject’s eye and comparing it to a
data set (Romain et al., 2012). The blood vessels in an Iris have complete uniqueness across the population, as they are determined randomly during gestation (Yazhuo et al., 2013). However, relatively high sensor cost, along with relatively large Failure to Enroll (FTE) rate reported in some studies, and lack of legacy iris databases may limit its usage in some large-scale government applications.

**Palmprint Recognition:** The palms of the human hands contain pattern of ridges and valleys much like the fingerprints. The area of the palm is much larger than the area of a finger and, as a result, palm prints are expected to be even more distinctive than the fingerprints. Since palm print scanners need to capture a large area, they are bulkier and more expensive than the fingerprint sensors. Human palms also contain additional distinctive features such as principal lines and wrinkles that can be captured even with a lower resolution scanner, which would be cheaper (Fahd et al., 2010; Yaoa et al., 2007). While law enforcement and forensics agencies have always collected fingerprints, it is only in recent years that large palmprint databases are becoming available (Raghavendra et al., 2011). Based on the success of fingerprints in civilian applications, some attempts have been made to utilize low resolution palmprint images (about 75 dpi) for access control applications (Mingxing et al., 2010; Rahman et al., 2007; Steven et al., 2009).

**Hand Geometry:** It is claimed that individuals can be discriminated based on the shape of their hands. Person identification using hand geometry utilizes low resolution (~20 ppi) hand images to extract a number of geometrical features such as finger length, width, thickness, perimeter, and finger area. The discriminatory power of these features is quite limited, and therefore hand geometry systems are employed only for verification applications (1:1 matching) in low security access control and time-and-attendance applications (Gang et al., 2007). The hand geometry systems have large physical size, so they cannot be easily embedded in existing security systems. Hand geometry bio-metrics fall into two main categories: geometric measurements and contour description (Charles et al., 2010). The automatic extraction of geometric measurements from a hand geometry image is a rather difficult error pruned task. Contour description methods have in general lower accuracy but they are more robust in automatic authentication processes (Choi et al., 2009).
**Voice:** Speech or voice-based recognition systems identify a person based on their spoken words. The generation of human voice involves a combination of behavioral and physiological features (Li *et al.*, 2013). The physiological component of voice generation depends on the shape and size of vocal tracts, lips, nasal cavities, and mouth. The movement of lips, jaws, tongue, velum, and larynx constitute the behavioral component of voice which can vary over time due to person’s age and medical condition (e.g., common cold). The spectral content of the voice is analyzed to extract its intensity, duration, quality, and pitch information, which is used to build a model (typically the Hidden Markov Model) for speaker recognition. Matching result is obtained by measuring the minimum error distance when the observation is aligned to the model. The matching techniques popularly used for model templates include Dynamic Time Warping algorithm, Vector Quantization and Nearest Neighbors algorithm (Harrison *et al.*, 1958). Speaker recognition is highly suitable for applications like tele banking but it is quite sensitive to background noise and playback spoofing.

**Signature:** Signature is a behavioral biometric modality that is used in daily business transactions (e.g., credit card purchase). However, attempts to develop highly accurate signature recognition systems have not been successful. This is primarily due to the large *intra-class variations* in a person’s signature over time. Attempts have been made to improve the signature recognition performance by capturing dynamic or online signatures that require pressure-sensitive, pen-pad. Dynamic signatures help in acquiring the shape, speed, acceleration, pen pressure, sharpness of loops, order and speed of strokes, during the actual act of signing (Maiorana *et al.*, 2010; Rua *et al.*, 2012). This additional information seems to improve the verification performance (over static signatures) as well as circumvent signature forgeries. Still, very few automatic signature verification systems have been deployed (Umut *et al.*, 2005).

**DNA:** DNA (deoxyribonucleic acid) is the well-known double helix structure present in every human cell. A DNA sample is used to produce either a DNA fingerprint or a DNA profile (Eberhart *et al.*, 1995). The molecular structure of DNA can be imagined as a zipper with each tooth represented by one of the letters: A (Adeline), C (Cytosine), G (Guanine), T (Thymine) and with opposite teeth forming one of two pairs, either A-T or G-C. The information in DNA is determined by the sequence of
letters along the zipper (Rabia et al., 2009). This method takes advantage of the different biological pattern of the DNA molecule between individuals. Unique differences in the banding pattern of the DNA fragments occur. DNA prints were first used in 1983 in United Kingdom (Bui et al., 2010). DNA fingerprinting is unpopular for authentication; it is only commonly used to compare two samples to check if they are from the same person (Yao et al., 2010). The reason it is not more widespread is many see it as a violation of their privacy. It is also computationally complex and thus time intensive to perform. DNA is unique among the majority of the population; however, DNA is not always unique between monozygotic twins (Khalil-Hani et al., 2010).

**Hand Veins:** The pattern of blood vessels hidden underneath the skin is quite distinct in individuals, even among identical twins and stable over long period of time. The primary function of veins is to carry blood from one part of the body to another and therefore vascular pattern is spread throughout the body (Kumar et al., 2013). The veins that are present in hands, *i.e.* palm, finger and palm dorsal surface, are easy to acquire (using near infrared illumination) and have been employed for the biometric identification (Yang et al., 2010). The vein patterns are generally stable for adults (age of 20-50 years) but begin to shrink later due to decline in strength of bones and muscles. There are several diseases, like diabetes, atherosclerosis, or tumors, which can influence the vein patterns and make them thick or thin. Biometric authentication devices using finger and palm vein imaging are now available for some commercial applications (Khan et al., 2008). To the best of the knowledge, there is no known large scale vascular biometric system. This could be primarily due to concerns about the system cost and lack of large scale studies on vein individuality and stability. On the plus side, these vascular systems are touchless which often appeals to the user.

**Gait Recognition:** Gait is the peculiar way one walks and is a complex spatiotemporal biometric. Gait is not supposed to be very distinctive, but is sufficiently discriminatory to allow verification in some low-security applications. Gait is a behavioral biometric and may not remain invariant, especially over a long period of time, due to fluctuations in body weight, major injuries involving joints or brain, or due to inebriety (Zhang et al., 2007). Acquisition of gait is similar to acquiring a facial picture and, hence, may be an acceptable biometric. Since gait based systems use the video sequence footage
of a walking person to measure several different movements of each articulate joint, it is input intensive and computationally expensive (Hanmandlu et al., 2011).

**Ear Recognition**: Ear biometric has become an admired biometric modality that facilitates enhanced biometric performance because of the distinctiveness, permanence and unchanged by aging. Ears have been shown to be one of the most unique physiological features on the human body (Dahel et al., 2003; Qinghai, 2010). Ear biometrics is that they retain their shape throughout life and are quite static. However, it is a topic of extensive current research.

![Commonly used biometric traits](image)

Fig. 3.2 Commonly used biometric traits (a) face, (b) fingerprint, (c) iris, (d) palmprint, (e) Hand geometry, (f) voice, (g) signature, (h) DNA, (i) veins, (j) gait, and (k) ear.
The recognition accuracy of individual biometric traits outlined above may not be adequate to meet the requirements of some high security applications. The low individuality or uniqueness and lack of adequate quality of individual biometric traits for some users in the target population can also pose problems in large scale applications (Ajay et al., 2012; Chan et al. 2012). The biometric modality employed for large-scale deployments demands high universality among the user population (Huang et al., 2011).

3.5 MULTI BIOMETRIC FUSION SCENARIOS

Depending on the number of traits, sensors, and feature sets used, a variety of fusion scenarios are possible in a multimodal biometric system and it is shown in the fig. 3.3.

**Multiple sensors:** A single bio-metric modality is acquired by using a number of sensors. One example is multiple face cameras for creating a 3D input face or for combining the output scores of the different baseline face images.

![Fig. 3.3 Various biometric fusion scenarios (a) multiple sensors, (b) multiple algorithms, (c) multiple units, (d) multiple instances, and (e) multiple modalities.](image)
Multiple algorithms: A single biometric input is processed with different feature extraction algorithms in order to create templates with different information content (Debnath et al., 2009). One example is processing fingerprint images according to minutiae and texture-based representations.

Multiple units: A single biometric modality but multiple parts of the human body are used. One example is the use of multiple fingers in fingerprint verification.

Multiple Instances: The same biometric modality and instance is acquired with the same sensor multiple times. One example is the sequential use of multiple impressions of the same finger in fingerprint verification.

Multiple modalities: Multiple biometric modalities are combined. This is also known as multimodal biometrics.

3.6 MULTIMODAL BIOMETRIC USING FACE AND EAR

Multi-modal biometrics increase accuracy by considering other highly specific biological traits to limit the number of applicants for an identity. This system is expected to be more reliable due to the presence of multiple, independent traits and not easy to forge multiple biometric traits. Variety of biometric scenarios is depending on the traits, feature sets and sensors applied. Some of the scenarios are multiple sensors, multiple algorithms, repeated instances, multiple modalities (Wang et al., 2009). Multimodal system functions are in three different modes. In Serial mode, the output of the one modality is used to reduce the number of possible identities before the next modality is used. In Parallel mode, sensed data from multiple modalities are used concurrently. In Hierarchical mode, individual modality is combined in a hierarchy structure.

Multimodal biometrics based on the fusion of two different biometric modalities face and ear; provide a new approach of non-invasive biometric authentication. There are several inspirations to choose face and ear for a multi-modal biometric recognition. Face and ear traits are the passive and non-intrusive unlike fingerprint and signatures (Tjark et al., 2010). During image acquisition, ear and face data can be captured using conventional cameras. The data collection for face and ear does not require participation or cooperation from the user (Ee-Chien et al., 2006). The traits of face and ear are in close physical proximity to
each other. Both biometric features are jointly present in an image or video captured of a user’s head and are both available to a biometric system.

### 3.7 PERFORMANCE EVALUATION METRICS

The performance of the multimodal system is expressed in terms of matching errors and image acquisition errors. Matching errors consist of false match rate (FMR), in which an impostor’s sample matches a legitimate user’s template, and False Non Match Rate (FNMR), in which a legitimate user’s sample does not match the own template. Genuine Acceptance Rate (GAR), is the assessment of the presentation that properly classified the genuine model as authentic. Dice Coefficient (DC) was intended to be applied to presence/absence data, Jaccard coefficient (JC) measures similarity between sample sets, and is defined as the size of the intersection divided by the size of the union of the sample sets.

**False Matching Rate (FMR)**

False Matching Rate is the possibility that the system inaccurately matches the input model to a non-matching template in the database. It calculates the percent of illogical inputs which are imperfectly acknowledged.

\[
FMR = \frac{\text{Invalid inputs which are incorrectly accepted}}{\text{Total number of inputs}}
\]  

(3.1)

**False Non Matching Rate (FNMR)**

False Non Matching Rate is the possibility that the system not succeeds to identify a match among the input model and a matching template in the database. It calculates the percent of valid inputs which are imperfectly discarded.

\[
FNMR = \frac{\text{Invalid inputs which are incorrectly rejected}}{\text{Total number of inputs}}
\]  

(3.2)

**Genuine Acceptance Rate (GAR)**

GAR is the assessment of the presentation that properly classified the genuine model as authentic

\[
GAR = 1 - FNMR
\]  

(3.3)
**Dice Coefficient**

Dice Coefficient (DC) was intended to be applied to presence/absence data, and is given by

\[
DC = \frac{2|A \cap B|}{|A| + |B|}
\]

(3.4)

where A and B are the number of species in samples A and B.

**Jaccard Coefficient**

The Jaccard index known as the Jaccard similarity coefficient is a statistic used for comparing the similarity and diversity of sample sets. The Jaccard coefficient (JC) measures similarity between sample sets, and defined as the size of the intersection divided by the size of the union of the sample sets:

\[
J(A, B) = \frac{|A \cap B|}{|A \cup B|}
\]

(3.5)

where A and B are the number of species in samples A and B.

**3.8 SUMMARY**

Multimodal biometric systems address numerous problems observed in single modal biometric systems. The complex methods employed to find a good combination of multiple biometric modalities and various level of fusion applied to get the best possible recognition result has been discussed. Performance of the system has been evaluated with the GAR, FMR, FNMR, JC and DC metrics. A multimodal approach is also investigated where face and ear images are combined to enhance the identification process of the individuals. The next chapter discusses a multi biometric fuzzy vault cryptosystem for face and ear recognition and shows the experimental results with the database.