Person Recognition through Profile Faces Using Ear Biometrics

Esraa Ratib Alqaralleh

Submitted to the
Institute of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Computer Engineering

Eastern Mediterranean University
March 2018
Gazimağusa, North Cyprus

Approval of the Institute of Graduate Studies and Research

Assoc. Prof. Dr. Ali Hakan Ulusoy
Acting Director

I certify that this thesis satisfies the requirements as a thesis for the degree of Doctor of Philosophy in Computer Engineering.

Prof. Dr. Işık Aybay
Chair, Department of Computer Engineering

We certify that we have read this thesis and that in our opinion it is fully adequate in scope and quality as a thesis for the degree of Doctor of Philosophy in Computer Engineering.

Assoc. Prof. Dr. Önsen Toygar
Supervisor

Examining Committee

1. Prof. Dr. Aytül Erçil
2. Prof. Dr. Fikret S. Gürgen
3. Assoc. Prof. Dr. Önsen Toygar
4. Asst. Prof. Dr. Yıltan Bitirim
5. Asst. Prof. Dr. Ahmet Ünveren

ABSTRACT

Recent studies in biometric systems have shown that the ear biometric is a reliable biometric for human recognition and among a lot of biometric traits it has achieved satisfying results for human recognition. In this thesis, 2D ear recognition approach based on the fusion of ear and tragus (small outer part of ear) using score-level fusion strategy is proposed. An attempt to overcome the effect of challenges such as partial occlusion, pose variation and weak illumination is done since the accuracy of ear recognition may be reduced if one or more of these challenges are available. In this thesis, the effect of the aforementioned challenges is estimated separately, and many samples of ear that are affected by two different challenges at the same time are also considered. The tragus is used as a biometric trait because it is often free from occlusion; it also provides discriminative features even in different poses and illuminations.

The features are extracted using Local Binary Patterns (LBP) and the evaluation has been done on four datasets, namely USTB-1, USTB-2, USTB-3 and UBEAR. It has been observed that the fusion of ear and tragus can improve the recognition performance compared to the use of ear or tragus systems individually. Experimental results show that the proposed approach 1 enhances the recognition rates by fusion of parts that are non-occluded such as tragus in cases of partial occlusion, pose variation and weak illumination. It is observed that the proposed approach 1 that uses score-level fusion strategy performs better than feature-level fusion methods. Additionally, the proposed approach 1 performs better than most of the state-of-the-art ear recognition systems. Experimental results on three datasets show that the proposed

approach 1 is robust and effective since it gives better results than the other matching algorithms under different ear challenges. The maximum accuracies achieved are 100% (under partial occlusion), 97.4% (under weak illumination), 100% (under pose variation), 97.5% (under real occlusion) for USTB-set1, USTB-set2, USTB-set3, UBEAR database, respectively.

On the other hand, this study aims to measure the efficiency of ear and profile face modalities in human recognition under identification and verification modes. In order to obtain a robust multimodal recognition system using different feature extraction methods, we propose to fuse these traits with all possible binary combinations of left ear, left profile face, right ear and right profile face. Fusion is implemented by score-level fusion and decision-level fusion techniques in the proposed approach 2. Additionally, feature-level fusion is used for comparison. All experiments in this approach are implemented on the UBEAR database.

Local Binary Patterns, Local Phase Quantization and Binarized Statistical Image Features approaches are used for feature extraction process in proposed approach 2. Images under different challenge such as illumination variation, pose variation and blurring are tested. Ear and profile face images from UBEAR database are used in the experiments. The experimental results show that the proposed approach 2 is more accurate and reliable than using ear or profile face images separately. The performance of the proposed approach 2 in terms of recognition rate is 100%, and in terms of Equal Error Rates is 1.9%.

Keywords: biometrics, ear recognition, profile face recognition, tragus recognition, score-level fusion, feature-level fusion, decision-level fusion, occlusion, pose variations, illumination variations, multimodal biometrics.

ÖZ

Biyometri sistemleriyle ilgili son zamanlardaki çalışmalar, kulak biyometrisinin insan tanıma problemi için güvenilir bir biyometri olduğunu göstermiş ve diğer birçok biyometrik özellikler arasında, insan tanıma için tatmin edici sonuçlar elde edilmiştir. Bu tezde, kulak ve tragus olarak adlandırılan dışkulak kıkırdağı (kulağın küçük dış kısmı), skor-seviyesi kaynaşım stratejisi ile kaynaştırılıp 2 boyutlu yeni bir kulak tanıma yaklaşımı önerilmiştir. Kulak tanıma sistemlerinin doğruluğunu azaltan kısmi kapatma, poz değişimleri ve zayıf aydınlatma gibi zorlukların varolduğu durumlarda, bunların etkilerinin üstesinden gelmek için girişimler yapılmıştır. Bu tezde, bahsi geçen zorlukların etkileri ayrı ayrı tahmin edilmiş ve ayrıca aynı anda iki farklı zorluk barındıran kulak resimleri de ele alınmıştır. Tragus ise sıklıkla kapatma etkisinden uzak olduğu ve farklı poz ve aydınlatmalar için bile ayırt edici öznitelikler sağladığı için ayrı bir biyometrik özellik olarak kullanılmıştır.

Öznitelikler, Yerel İkili Örüntü yöntemi ile çıkarılmış ve değerlendirmeler USTB-1, USTB-2 ve USTB-3 verisetleri üzerinde yapılmıştır. Deneyler sonucunda, kulak ve tragusun bireysel olarak kullanıldığı tanıma sistemlerine göre, kulak ve tragus kaynaşımı kullanılan sistemin tanıma performansını iyileştirdiği görülmektedir. Kısmi kapatma, poz değişimleri ve zayıf aydınlatmanın bulunduğu durumlarda, birinci önerilen yaklaşımın kapatılmış olmayan tragus ile kaynaşım yapıldığı için deney sonuçlarında tanıma oranlarının arttığı gözlemlenmiştir. Ayrıca, skor-seviyesi kaynaşım stratejisi kullanan birinci önerilen yaklaşım, öznitelik-seviyesi kaynaşım kullanan yöntemlere göre daha iyi performans göstermiştir. Buna ek olarak, birinci önerilen yaklaşım, literatürdeki diğer kulak tanıma sistemlerine göre daha iyi

sonuçlar vermektedir. Üç veriseti üzerinde yapılan deney sonuçlarına göre, güçlü ve etkili olan birinci önerilen yaklaşım, değişik zorluklar altında, diğer eşleştirme algoritmalarından daha iyi sonuç vermiştir. Maksimum doğruluk oranları; kısmi kapatma durumunda %100 (USTB-set1 veriseti üzerinde), zayıf aydınlatma durumunda %97.4 (USTB-set2 veriseti üzerinde), poz değişimleri durumunda %100 (USTB-set3 veriseti üzerinde) ve gerçek kapatma durumunda ise %97.5 (UBEAR veriseti üzerinde) olarak bulunmuştur.

Diğer yandan, bu çalışmanın bir diğer amacı da insan tanıma ve doğrulama işlemleri için kulak ve profil yüz resimlerinin etkisini ölçmektir. Güçlü bir insan tanıma sistemi elde etmek için; sol kulak, sol profil yüz, sağ kulak ve sağ profil yüz resimlerini içeren ve mümkün olan bütün ikili kombinasyonların kaynaşımını kullanan ve değişik öznitelik çıkarma yöntemleri uygulayan bir sistem önerilmiştir. Bu ikinci önerilen yaklaşımda, skor-seviyesi kaynaşım ve karar-seviyesi kaynaşım teknikleri uygulanmıştır. Buna ek olarak, karşılaştırma yapmak için öznitelik-seviyesi kaynaşımı da kullanılmıştır. Bu yaklaşımla ilgili bütün deneyler UBEAR veritabanı üzerinde yapılmıştır.

İkinci önerilen yaklaşımda, öznitelik çıkarma işlemleri için Yerel İkili Örüntü, Yerel Faz Nicemleme ve İkili İstatistiksel Görüntü Öznitelikleri yaklaşımları kullanılmıştır. Aydınlatma değişimleri, poz değişimleri ve bulanıklık gibi farklı zorluklar içeren resimler test edilmiştir. UBEAR veritabanındaki kulak ve profil yüz resimleri deneylerde kullanılmıştır. Deney sonuçları, ikinci önerilen yaklaşımın, kulak veya profil yüz resimlerinin ayrı ayrı kullanıldığı sistemlere göre daha doğru ve güvenilir sonuçlar verdiğini göstermiştir. İkinci önerilen yaklaşım, tanıma oranı açısından %100 ve Eşit Hata Oranları açısından da %1.9 performans elde etmiştir.

Anahtar kelimeler: Biyometri, kulak tanıma, profil yüz tanıma, tragus tanıma, skor-seviyesi kaynaşım, öznitelik-seviyesi kaynaşım, karar-seviyesi kaynaşım, kapatma, poz değişimleri, aydınlatma değişimleri, çoklu biyometri.

ACKNOWLEDGMENT

All thanks and appreciation goes to my supervisor Assoc. Prof. Dr. Önsen Toygar. This work would not have been possible without her supervision and excellent guidance.

Nobody has been more important to me in the pursuit of this thesis than the members of my family. I would like to thank my parents, sisters and brothers, whose love and guidance are with me in whatever I pursue.

Most importantly, I wish to thank my loving and supportive husband Ayman Afaneh and his endless support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BSIF	Binarized Statistical Image Features
EER	Equal Error Rate
EHL	Ear Height Line
FAR	False Accept Rate
FDA	Fisher Discriminant Analysis
FMR	False Match Rate
FNMR	False Non-Match Rate
FRR	False Reject Rate
GAR	Genuine Accept Rate
HE	Histogram Equalization
HOG	Histogram of Oriented Gradients
LBP	Local Binary Patterns
LPQ	Local Phase Quantization
MVN	Mean-Variance Normalization
NLM	Non Local Mean
NPE	Neighborhood Preserving Embedding
PCA	Principal Component Analysis
ROC	Receiver Operating Characteristic
SF	Steerable Filter
SIFT	Scale Invariant Feature Transform
SRC	Sparse Coding Coefficients
SURF	Speed-Up Robust Feature
SVM	Support Vector Machine

Chapter 1

INTRODUCTION

1.1 Biometrics Systems

A biometrics system is basically a pattern recognition system that recognizes a person based on features derived from a specific physiological or behavioral characteristic of the person. Physiological characteristics involve innate human body traits such as fingerprint, iris, face, vein, DNA, hand geometry, ears and many more, whereas behavioral characteristics are related to the measure of uniquely identifying and measurable patterns in human activity such as gait, signature, odor. Figure 1 depicts some examples of several biometric traits.

Biometrics systems are more secure, reliable and provide much higher security solution compared with the traditional systems that depend on magnetic cards, passwords or secret codes that can be stolen, faked or difficult to remember and can be forgotten. Hence, biometric systems are inherently and more reliable and comfortable than traditional authentication methods.

Ideal biometric characteristics have five qualities that are needed for successful authentication [1]:

- **Robustness**: means unchanging on an individual's biometric trait over time (invariant biometric trait).

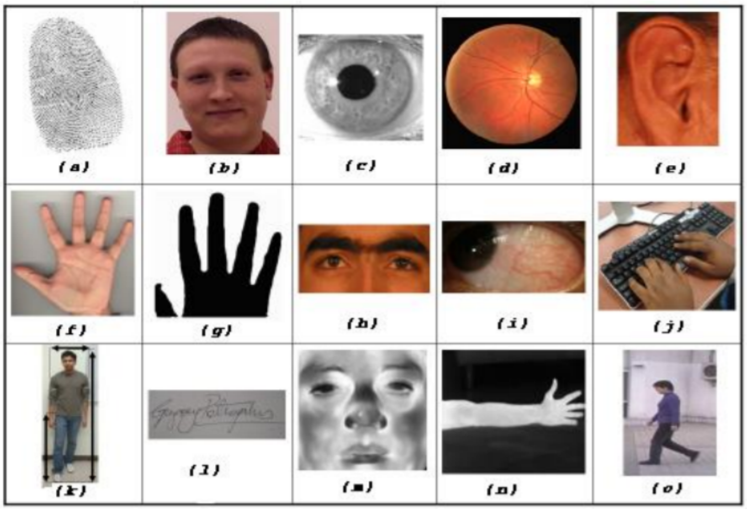


Figure 1: Different Examples of Biometric Traits. a) Fingerprint b) Frontal Face c) Iris d) Retina e) Ear f) Palmprint g) Hand Geometry h) Periocular i) Conjunctival Vasculature j) Keyboard Striking k) Anthropometry l) Signature m) Thermogram of The Face, n) Thermogram of the Hand o) Gait

- **Distinctiveness**: means the trait showing great variation over the population.

- **Availability**: means that all the individuals should ideally have this biometric trait.

- **Accessibility**: means easy to acquire biometric trait using suitable devices such as electronic sensors.

- **Acceptability**: means that individuals do not object having this biometric trait to be taken from them and to be presented to the system.

1.2 Biometric Functionality

The biometrics system commonly operates in one of two functionalities such as verification and identification [1]:

1- Verification or authentication system seeks to answer the question ”Are you who you say you are?”. Under verification system, an individual presents himself/herself as a specific person. The system checks his/her biometric trait against biometric data that exist in the database. A verification system is described as one-to-one matching system because the system matches the biometric traits of the individuals against specific biometric data in the database. Then the decision is obtained as genuine or impostor.

The main stages of verification system pass through the following steps [2, 3]:

- Data acquisition phase is the stage that biometric systems provide the raw data of the individual.
- Pre-processor is mainly used for enhancing the image, eliminating the noise and detecting the Region of Interest (ROI) of the image.
- Feature extractor computes a set of salient and distinguished features of the input biometric data. Feature extraction is defined as the process in which the discriminatory information (feature vector) is obtained.
- Matcher is used to compare identities of two biometrics using the extracted feature vectors and produces the match score that indicates the degree of the similarity/dissimilarity between the sample and the reference template. In other words, if the matcher produces low and insufficient similarity that is not enough to recognize the client, the identity will be rejected; otherwise it will be accepted.
- Decision is the final stage that is based on the generated match score by the matcher of the previous stage. In the verification operation, the output is an ”accept” indicating a genuine match or a ”reject” indicating an impostor. Steps of verification system are presented in Figure 2.

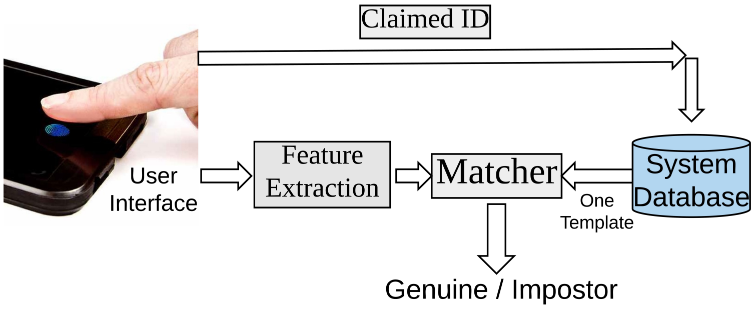


Figure 2: Block Diagram of a Verification System

2- Identification system aims to identify a specific individual (one-to-many matching) where the identity is compared with the all enrolled samples in the database. In this case, the system outputs either the identity of person template or the decision in which the input is not an enrolled user. Identification process is classified based on the cooperation of the user into positive and negative identification.

Generally, positive identification systems prevent multiple individuals from using a single identity and reject an individual's claim to their identity, if no match is found between the acquired data sample and enrolled template. The template in positive identification systems can be stored in decentralized or centralized databases. An attempt to access a restricted area by unauthorized person using his face as biometric trait represents a positive identification.

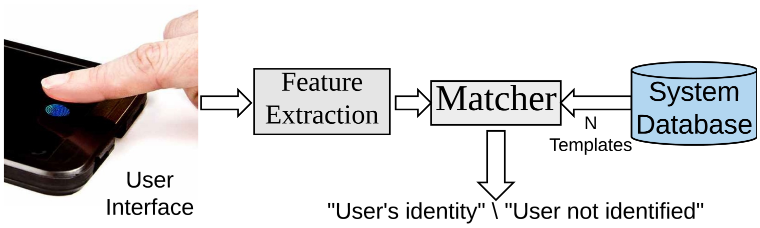


Figure 3: Block Diagram of an Identification System

Negative identification systems prevent multiple identities of a single user, this means that negative identification systems reject a user's claim to no identity if a match is found. Negative identification system can be found in driver licensing and social service systems where multiple enrollments are illegal.

Additionally, the same aforementioned phases of verification are also considered as the main phases of identification system except the difference in the matching phase where the matching in identification is conducted as one-to-many matching, and in the decision step, the output is a list of potential matching identities sorted in terms of their match score. Figure 3 shows the block diagram of an identification system.

1.3 Performance Measures

Two types of errors are produced if the biometric system has large inter-user similarity and large intra-user variations [4], namely false non-match rate (FNMR) and false match rate (FMR). A false non-match error occurs when the two samples of the same trait of an individual may not be matched. False match occurs when two samples from different individuals are incorrectly recognized as a match.

1.3.1 Verification Accuracy

In a verification system, FNMR corresponds to False Reject Rate (FRR), and consequently FMR corresponds to False Accept Rate (FAR). A verification system makes a decision based on comparing the output match score related to genuine and impostor with a threshold value. The proportion of genuine scores that are less than the threshold value is FRR, and the fraction of impostor scores that are greater than or equal to threshold value is FAR, as shown in Figure 4. The value of threshold is chosen based on the purpose and importance of the biometric system. In this study we computed

FAR and FRR as:

$$FAR = \frac{Number\ of\ Accepted\ Imposters}{Total\ Number\ of\ Imposter\ Comparisons} \times 100\%, \tag{1.1}$$

$$FRR = \frac{Number\ of\ Rejected\ Genuine\ Persons}{Total\ Number\ of\ Genuine\ Comparisons} \times 100\%, \tag{1.2}$$

Receiver Operating Characteristic (ROC) curve is a linear or logarithmic scale, which compares the performance of different recognition systems by computing Genuine Acceptance Rate (GAR) and FRR values, where GAR = (100 - FRR)% without using threshold value in the graph. The point in ROC curve where FAR equals FRR is called Equal Error Rate (EER).

The EER of a system can be used to give a threshold independent performance measure, and it is an indicator how accurate the system is, where a lower EER value indicates better performance.

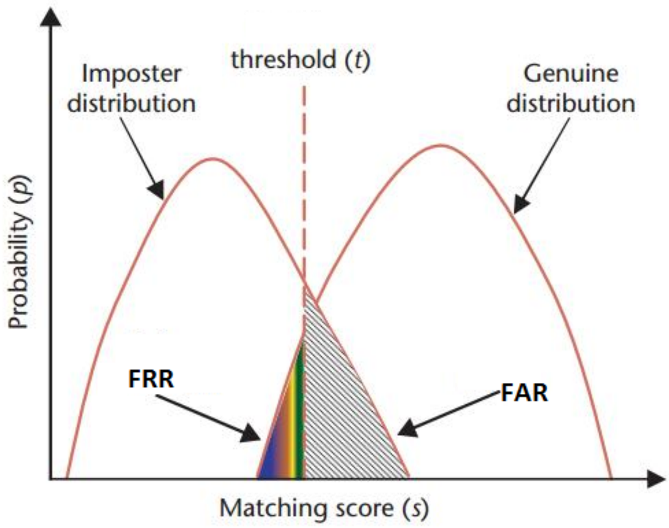


Figure 4: The Relationship between FAR, FRR and Threshold Value [1]

1.3.2 Identification Accuracy

Recognition rate is the most known performance measurement to evaluate a biometric system. It is computed by comparing the enrollment samples with the test samples and after determining their matching scores, ranking them. Recognition rate, which shows how often the template of genuine is found in rank-1 match, is called Rank-1 recognition rate and it is calculated as follows:

$$Recognition\ Rate = \frac{Number\ of\ Genuine\ Matches\ (Top\ k\ Match)}{Total\ Number\ of\ Test\ Matches\ Performed} \times 100\%, \quad (1.3)$$

1.4 Unimodal Biometric Systems

Biometrics systems that use one biometric trait of the individual for recognition is called unimodal system. The major issue with unimodal biometric system is that no one technology can be suitable for all applications. Within a large population, unimodal biometrics is prone to inter-class similarities. For example facial recognition may not work correctly for similar people as the system might not be able to distinguish between the two subjects leading to inaccurate matching.

On the other hand, unimodal biometric systems have to contend with variety of problems such as intra-class variation, noisy data and spoof attacks on stored data. For instance, ear recognition performance decreases due to changes in illumination, pose variation and occlusions [5,6]. Several of these problems can be addressed by deploying multimodal biometric systems by combining multiple source of information.

In this study, we deal with important modalities namely ear and profile face which are widely used for identification systems. Profile face images are recently used to recognize a person in security and surveillance applications. On the other hand, ear contains additional distinctive features. More details about ear biometrics are given in the next subsection.

1.4.1 Ear Biometrics

Recently, ear biometrics is gaining high acceptance for human recognition in high security areas. Ear has many properties that make it strongly desirable in biometric systems. These properties are presented as follows:

1- The structures of the ear's anatomical parts such as outer helix, tragus, antitragus and lobe are discriminate. Anatomical parts of ear are shown in Figure 5.

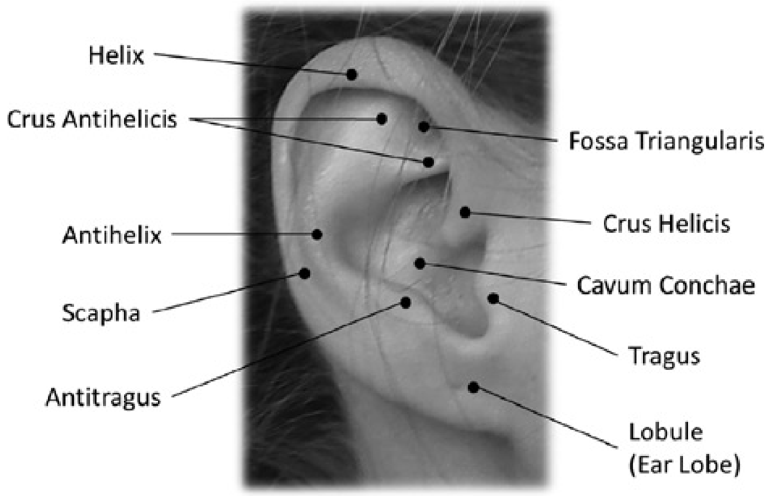


Figure 5: Outer Anatomical Parts of the Ear [2]

2- The ear is not effected by facial expression and aging where one of the first studies that concerned the time effect on the ear recognition [7] proved that the shape of the ear is very stable with the age.

3- The convenient size and the location of the ear, as a part of the profile face, make it easier to be captured compared to other traits such as retina and fingerprint.

4- Capturing of ear image does not require user cooperation. Consequently, it is a candidate solution for passive environment and applications such as forensic image analysis and automated surveillance tasks.

5- Ear can be a supplement for other biometric traits such as face, which may suffer from profile face in some of the surveillance applications that use face recognition. In that case, ear, which is part of the profile face, can provide additional information of the individual.

There are several challenges that can significantly reduce the efficiency of ear recognition performance as well as degrading the extraction of robust and discriminate features. Some of these challenges are described as follows:

- Pose variation: different viewpoints of the camera cause pose variance in the ear image as shown in Figure 6. In this condition, pose variation introduces projective deformations and self-occlusion hence it has more influence than other challenges on ear or profile face recognition process [8, 9]. Additionally, pose variation causes intra-user variation which means that samples of two different individuals taken from single pose may appear more similar than samples that are captured from the same person under different poses (inter-user variation). Many studies that solve the pose variation challenge of ear exist in the literature such as in [10, 11].

Figure 6: Samples of the Same Person at Different Poses

- Illumination variations: many factors affect the appearance of the human ear and face in the images such as nonuniform lighting which may also affect the appearance of the samples because of the internal camera control and the reflected light from skin [12]. The variation in lighting is considered as one of the main technical challenges in recognition systems, especially in ear and face biometric traits, where the trait of a person, which is captured under different illumination conditions, may appear different to a high extent [13] as shown in Figure 7. There are many studies that take into account the illumination variation challenge in ear recognition such as [6].

Figure 7: Ear under Illumination Variation Challenge

- Occlusion: one of the main drawbacks of ear recognition system is occlusion. The ear can be fully or partially occluded by hair, accessories, head dress or headphone as shown in Figure 8. This is not a critical problem in the active identification system where the subject can cooperate with the system to remove them at the capturing process, but it is a problem in passive systems as no assistant on the part of the subject can be assumed. In order to overcome the occlusion problem, a segmentation method and a classifier for each segment can be used [14, 15]. This approach may give a chance to the segments that are not occluded to give a correct classification and successfully identify the individual. Different methods have been proposed to overcome the occlusion problem of ear recognition [16, 17].

Figure 8: Occluded Ears by Hair, Accessories and Headphone [18]

- Ear surgery: recently, the appearance of the ear may be deformed due to the increase in the beauty surgery of the ear [19] as shown in Figure 9. Some parts of ear such as lobe may be stretched or split as shown in Figure 10.

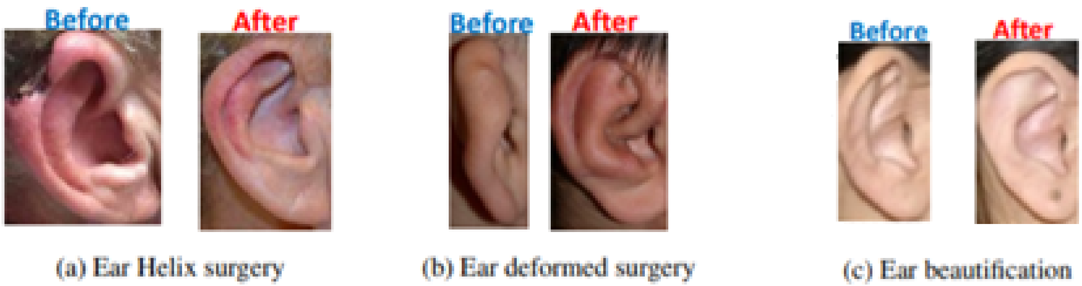
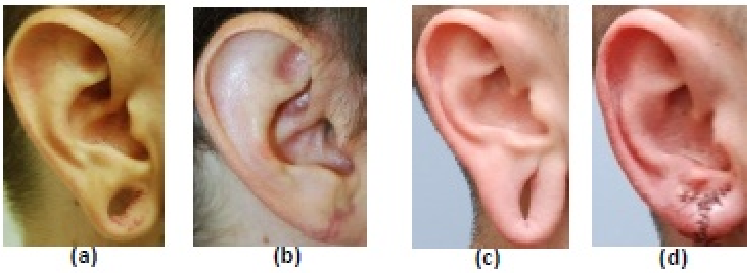


Figure 9: Examples of Different Kinds of Ear Surgery Images [19]



(a) (b) (c) (d)

Figure 10: Ear Before (a , c) and After (b, d) Ear Lobe Surgery

1.5 Multimodal Biometrics

Multimodal biometric systems use combination of two or more biometric modalities. Supplementary information among different modalities are provided in order to increase the recognition performance in terms of accuracy and reliability, and achieve the robustness of a biometric system. Additionally, multimodal biometric systems exploit more than one modality as an alternative to overcome different challenges such as illumination variations, various occlusion and pose variations. Multimodal biometric systems are difficult to spoof as compared to unimodal systems. This makes the multimodal system more appropriate for different applications and customer preference [20–22].

Multimodal biometric systems are classified into different levels according to the level of data fusion. Feature-level, score-level and decision-level fusion are considered as the most commonly used fusion techniques in the literature [23, 24].

Feature-level fusion, in which multiple features acquired from feature extraction processes are fused into new feature vector [25].

Score-level fusion has lesser complexity than the other fusion levels and hence it is widely used. In this fusion level, fusion of scores are obtained from each matching process. These scores are fused to verify the claimed identity and the final decision can be obtained by combining those scores as a new match score using different fusion techniques such as product rule and sum rule [1]. The matching scores that are generated during score-level fusion may be in different numerical range. In this case, normalization techniques are needed to make all the scores under the same domain before the fusion of the individual scores related to different modalities. In this study, min-max and tanh-normalization methods are applied and then tanh-normalization is adopted because it is reported to be robust and highly efficient than other normalization methods [26].

On the other hand, decision-level fusion can be performed after matching. This method can be performed when only the decision outputs by biometric system are available. ”AND” and ”OR” rules and majority voting approach are the most commonly used methods in decision-level fusion [27]. In this study, we used majority voting for the fusion of decision outputs.

The aforementioned fusion techniques on ear-tragus and profile face-ear biometrics are discussed and implemented in this thesis in order to improve the performance of unimodal biometric systems.

1.6 Research Contributions

Our first contribution is to use tragus and ear in order to overcome the limitations of biometric systems under different ear challenges such as weak illumination, occlusion and pose variation. Some of the limitations of unimodal biometric systems may be overcome by using multiple sources of information for recognizing the identity, trying to improve matching performance, and minimizing error rates. In this study, based on tragus which is a small pointed eminence of the external ear, fusion strategies are used to enhance the recognition rate, even with different challenges such as pose variation, variation in illumination and occlusion. Features of tragus are extracted by Local Binary Patterns (LBP) and used because it is almost free from occlusion and it is clearly seen in left and right ears rotation. Features of tragus are fused with other features that are extracted from other segments of the ear of an individual. This fusion has the advantage of capturing the raw data in the same shot. Additionally, we fused tragus and ear biometrics using score-level fusion approach in the Proposed Approach 1 which is not applied for tragus and ear in other studies. In score-level fusion scenario, feature sets from the tragus and non-occluded part of the same ear image are extracted individually by LBP. Then matching process between training and test samples is performed to obtain the match scores and furthermore, the match scores from both traits are fused to get a fused score. Finally, K-Nearest Neighbor (k-NN) classifier is used for classifying the fused scores.

Our second contribution is to use ear and profile face modalities for person identification by multimodal biometric approach. Fusion of these traits with all possible binary combinations of left ear, left profile face, right ear and right profile face are implemented under different challenges. Fusion is implemented by score-level fusion and decision-level fusion techniques in the proposed approach 2. Binarized Statistical Image Features (BSIF), Local Phase Quantization (LPQ) and Local Binary Patterns (LBP) approaches are used for feature extraction process.

1.7 Outline of the Dissertation

The rest of this thesis is organized as follows. Related work of our study is explained in Chapter 2. In Chapter 3, feature extraction methods that are applied in this study are presented in detail. Different databases that are used to evaluate the proposed systems are described in Chapter 4. In proposed scheme 1, fusion approaches for human identification using ear and tragus are described in Chapter 5. Multimodal biometric systems of ear and profile face of proposed scheme 2 are explained in Chapter 6. Finally, Chapter 7 concludes the thesis and states the future work.

Chapter 2

LITERATURE REVIEW

The oldest and the most famous work about ear recognition is done by Alfred Iannarelli [28]. Using large number of ear images manually, ”Iannarelli System”, which includes 12 measurement as shown in Figure 11, found that ear images were different to a high extent. Burge and Burger [29] (1998) noted that detecting the anatomical points is difficult which makes Iannarelli System not applicable. If the first point is not assigned accurately, the rest of the points are not useful. Many different techniques are widely

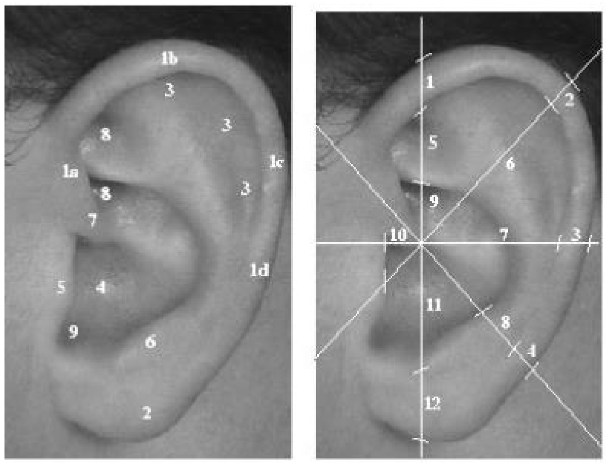


Figure 11: The Measurements of the Iannarelli System [28]

used for ear recognition such as force field transformation, geometric and appearance-based approaches. Ear recognition based on force field transformation method was developed in 2000 by David et al. [30,31]. The functionality and the role of force field energy in extracting features from the images is used to find an energy line, while the pixels are treated as sources of Gaussian source field.

In [7], after constructing adjacency graph from Voronoi diagram of ear edges for each ear, the isomorphism of the constructed graphs can be used to compare the templates of ear.

On the other hand, Choras [32] has used geometrical method of feature extraction in 2006. Contours from ear images were extracted by using methods based on geometrical parameters, GPM-triangle ratio method, GPM- shape ratio method and ABM-angle based contour representation method. A feature vector for each extracted contour is constructed. The results using rank-1 recognition of the GPM and ABM were 100% and 90.4%, respectively.

Another local approach of 2D ear authentication was proposed in 2007 by Nanni and Lumini [33] which was based on multi-matcher approach for ear recognition based on the convolution of a segmented windows with a bank of Gabor filter to extract local features. Laplacian Eigen Maps were used to reduce the dimensionality of the feature vectors. Sub-windows were selected using Sequential Forward Floating Selection (SFFS) algorithm to represent the ear. By combining the decisions of multiple nearest neighbor classifiers for each segmented window, matching process was conducted.

Prakash and Gupta in 2013 [6] proposed an ear recognition approach by applying three different preprocessing techniques as Non Local Mean filter (NLM), Adaptive Histogram Equalization (ADHist) and Steerable Filter (SF). These techniques were used in parallel to overcome the varying illumination, poor contrast and non-registered image. Speed-Up Robust Feature (SURF) was used for feature extraction because it provides high distinctive features in which each point is associated with a descriptor vector of 128 feature elements. The proposed technique enhanced the recognition performance over the existing techniques for UND-E database and the reported accuracy is 96.75% while FAR is 2.58% and FRR is 3.92%.

On the other hand, an approach based on Haar wavelets was proposed to extract features after preprocessing methods such as adaptive histogram equalization and size normalization [34]. Matching is done by fast normalized cross correlation. The proposed method is applied on USTB-set 2 ear image database and IIT Delhi database. An average accuracy on USTB-set 2 for 137 subjects is 97.2% and on IIT Delhi for 125 subjects is 95.2%.

Recently, an approach using geometric information of the ear has been presented in [35]. The proposed method depends on the shape of the ear which involves image pre-processing using Gaussian filter, ear helix detection by Canny edge operator which consists of both the outer and inner helixes. Geometric feature extraction based on maximum and minimum EHL (Ear Height Line) uses Euclidean distance for feature matching. The algorithm was applied on USTB-1 and IIT Delhi databases and the recognition rates are 99.6 % and 98 .3 % for IIT Delhi and USTB-set1, respectively.

Benzaouiet et al. in [36] have proposed a technique using local texture-based techniques such as LBP, LPQ, and BSIF in addition to k-NN and SVM classifiers. The images were segmented using horizontal-vertical projection. The experiments were applied on three databases namely IIT Delhi-1, IIT Delhi-2 and USTB-1. The best recognition rate was 98.46% by using k-NN classifier.

Some researches were done on ear recognition in order to solve occlusion challenge. In 2006, Yuan et al. [37] improved non-negative Matrix Factorization with Sparseness Constraint (INMFSC) for ear recognition with occlusion. The ear image was divided into three parts without overlapping and INMFSC was applied for feature extraction. The final classification was based on a Gaussian model-based classifier. The results of USTB dataset3 with 79 subjects were reported and the best rank-1 accuracy for 10% occlusion from above of the ear was nearly 91%.

On the other hand, a new model-based approach for ear recognition was proposed in [38] that fuses the model-based and outer ear metrics. Profile faces of 63 subjects from XM2VTS dataset were tested. The rank-1 accuracy for 30% occlusion from above of the ear was 89.4%.

Many other researches identify people based on occluded ear images. Bustard and Nixon [39] tried to solve the occlusion problem using an ear registration and recognition method by treating the ear as a planar surface and creating a homography transform using SIFT feature matches. Ear recognition under partial occlusions was discussed in that paper. The relationship between occlusion percentage and recognition rate was presented. The rank-1 accuracy on XM2VTS dataset was 92% (for 30% occlusion from above and left side).

Yuan and Mu [40] proposed a 2D ear recognition approach based on local information fusion and under partial occlusion challenge. In this approach, images are separated to sub-windows and features from each sub-window were extracted by Neighborhood Preserving Embedding (NPE). Sub-window regions form sub-classifiers, and the partial occlusion was removed on different levels and different locations. Rank-1 recognition rate depends on 28 sub-classes that are extracted from 28 sub-windows of each image. For 24th sub-class of USTB database with 50% occlusion, 80% recognition rate is achieved in rank-1; for 19th subclass 72% recognition rate is achieved with 50% occlusion in UND database.

A recent approach named Sparse Coding Coefficients (SRC) [5] is applied to represent a test image with occlusion as the combination of sparse linear combination of training samples and sparse error incurred by image noise. To develop the SRC model under partial ear occlusion, Yuan et al. [41] have used non-negative descriptors extracted by Gabor feature descriptors and non-negative occlusion descriptors. Experimental results on USTB database subset-3 with occlusion are 93.8%, 85.4% and 79.2% for 15%, 25% and 35% occlusion, respectively.

A few works have been done in the field of fusion of ear with other modalities. Chang et al. [42] have aimed to compare ear and face image recognition rates using a Principal Component Analysis (PCA) technique on faces and ear images. The normalized masked ear and face images of a subject are concatenated to combine face and ear images which provided better performance than using each one separately. The recognition rate for face and ear separately were 70.5% and 71.6% respectively, where the multimodal recognition rate was 90.9%.

A fusion method between left and right ears using shape feature for recognition was proposed by Zhang et al. [43] to increase the recognition rate. They achieved recognition rate of 93.3% by using left or right ear image and 95.1% by fusing both sides of ear images.

In [44], a combination between palmprint and ear was done based on features extracted from palm and ear images. HMAX model with Gabor filter for palmprint and Gaussian filter for ear were implemented. SVM and k-NN were used for classification. The recognition performance reached to 100%.

A recent work by Hezil and Boukrouche [45] in 2017 have fused two biometric modalities such as ear and palmprint. The authors used BSIF texture descriptor at canonical correlation analysis and feature-level fusion, attaining recognition rate of 100% using IIT Delhi-2 ear and IIT Delhi palmprint database.

One of the most important modality that is fused with ear is profile face. In Pan et al. in 2008 [46], they fused ear and profile face at feature-level fusion. Fisher Discriminant Analysis (FDA) technique was used and achieved a recognition rate of 96.84% using the USTB database.

Additionally, Xu and Mu [47] used Kernel Canonical Correlation Analysis (KCCA) for the fusion of profile face and ear. They performed decision fusion using the

Weighted-Sum rule. USTB database is used and they achieved a recognition rate of 98.68%.

On the other hand, a local feature extraction technique called Speed-Up Robust Feature (SURF) was used in [48]. The recognition performance was improved by the fusion of ear and profile face. It was noted that the score-level fusion was better than the feature-level fusion. The recognition rates were 98.02%, 96.02% and 99.36% on UND-E, UND-J2 and IITK datasets, respectively.

Recently, a new method has been proposed by Annapurani et al. [49] to fuse the shape of the ear and tragus. An enhanced edge detection method was used to extract the features from tragus. The shape of the ear was also extracted and a fused template was formed by combining the tragus and shape of the ear by feature-level fusion. IIT Delhi ear database which has no occlusions and AMI ear database that includes mild occlusions [50] were used in the experiments. The accuracies were 99.2% and 100% for AMI and IIT Delhi databases, respectively.

Compared to the recent work of Annapurani et al. [49] and by exploiting the advantages of the tragus, our Proposed Approach 1 describes a new technique for 2D ear image recognition and ear is recognized from different poses, under different illumination conditions and with different ratios and locations of occlusion. The score-level fusion of tragus and ear is applied to demonstrate that the accuracy of the ear recognition system in the proposed method with score-level fusion is better than the accuracy in the systems that uses ear biometric traits individually.

Chapter 3

FEATURE EXTRACTION METHODS REVIEW

3.1 Overview

A general biometric system can be divided into two basic activities: feature extraction and classification. In feature extraction, there are two main classes: global and local feature extraction approaches [51,52]. Global approaches are based on the pixel information; all the pixels of the image are treated as a single vector, and the total number of pixels represents the size of the vector. Most methods in this approach use another representation subspace to reduce the number of pixels and to eliminate redundancies. The aim of using global feature is to utilize more specific and less frequent features to represent more discriminative knowledge of a class domain.

Principal Component Analysis (PCA) [53], Linear Discriminate Analysis (LDA) [54], and Independent Component Analysis (ICA) are the most popular methods used for dimensionality reduction and the extraction of useful information. In this study we used Principal Component Analysis (PCA) for comparison purposes.

Local approaches depend on the description of the local neighborhood of specific points in the image. The aim of these approaches is to find local features that are robust against illumination and pose variations since they do not depend on the location or relations between those points.

Recently, researchers have focused on local approaches that are considered more robust than global approaches; they are mainly based on geometric information such as distances, landmark points, angles and spatial relationships between the components of the biometric modality. However, neither global nor local approaches are efficient in uncontrolled conditions. In this study, we used feature extraction approaches essentially based on local texture descriptors in order to identify people from their 2D ear images which are described below.

3.2 Local Texture Descriptors

One of the main characteristic that played a critical role in the field of pattern recognition is the texture. Texture is an important characteristic of many kinds of images that range from multispectral remotely sensed data to microscopic images. Image texture may provide information about physical properties of objects like smoothness, roughness or differences in surface reflectance like color [55].

Local texture descriptor methods can easily derive an effective feature model that combines the global form of the analyzed object and the local texture of its appearance in a single feature vector. With this type of descriptor, the entire image is scanned pixel by pixel, providing local information, and the co-occurrences of the texture descriptor are accumulated in a discrete histogram, providing global information. In addition, these approaches codify and collect the co-occurrence of the micro features as a histogram. They are characterized by a very high discriminative power, simplicity of calculation, and invariance to any monotonic changes in gray level.

Texture-based methods have achieved satisfying results on different biometrics such as face, iris and ear [56]. In this study, we tested and compared three recent local texture descriptors, namely Local Binary Patterns (LBP), Local Phase Quantization (LPQ) and Binarized Statistical Image Features (BSIF). The details related to these methods are given below.

3.2.1 Local Binary Patterns (LBP)

Local Binary Patterns (LBP) is one of the widely-used texture-based schemes that is firstly proposed by Ojala et al. [57, 58] due to the high calculation speed, low computational complexity, simplicity, effectiveness and insensitivity for gray scale change and illumination variation [59]. Additionally, LBP achieves high performance on face recognition [60–62] and ear recognition compared with other texture descriptors [36, 63].

The original LBP operator was founded on the assumption that texture has locally two complementary aspects: a pattern and its force. The operator works in a neighborhood of (3×3), using the central value as a threshold [62]. An LBP code describing the local texture pattern is generated as follows: all neighbors take the value 1 if their value is higher or equal to the current pixel and 0 otherwise. The pixels of this binary code are multiplied by corresponding weights and summed in order to get the LBP code of the current pixel. As the neighborhood is composed of 8 pixels, a total of 2^8 different labels can be obtained depending on the gray values relating to the center and its neighborhood.

The LBP value of the center pixel in the P neighborhood on a circle of radius R is calculated by:

$$LBP_{P,R}(x_c, y_c) = \sum_{n=0}^{7} \delta(g_n - g_c)2^n, \tag{3.1}$$

where g_c is a center pixel value positioned at (x_c , y_c), g_n is one of the eight surrounding center pixel values with the radius R, P is the whole neighborhood number, and a sign function δ is defined such that:

$$\delta(x) = \begin{cases} 1, & x > 0 \\ 0, & otherwise \end{cases}$$

A basic implementation of the original LBP operator is shown in Figure 12, and Figure 13 shows some normalized samples of ear images and their local binary pattern representation.

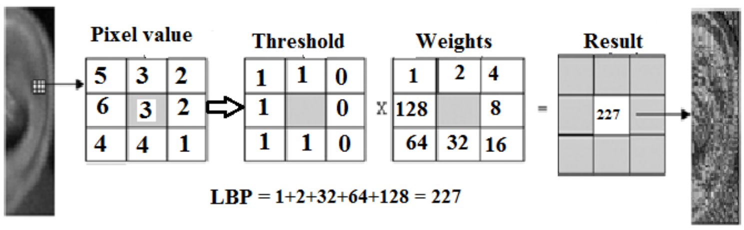


Figure 12: Local Binary Pattern Operator Applied on Normalized Ear Image

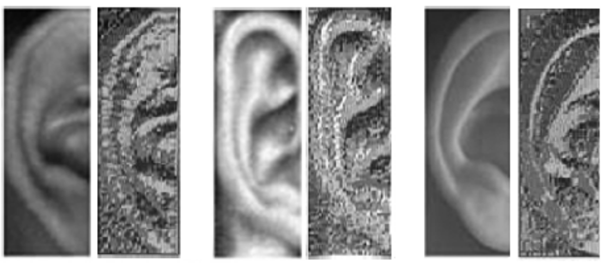


Figure 13: Normalized Ear Images and their Local Binary Pattern Representation

In order to handle many challenges that appear in image processing applications such as illumination variation, blur, different scales and rotation, new texture descriptors are developed. The most important and the latest texture descriptors are LPQ and BSIF in which the details are given below.

3.2.2 Local Phase Quantization

Ojansivu et al. [64] proposed a new descriptor for texture classification that is robust to image blurring and invariant to uniform illumination changes called Local Phase Quantization (LPQ), based on quantizing the Fourier transform phase in local neighborhoods. LPQ has proven to be a very efficient descriptor in face recognition [61] and ear recognition [36].

A convolution between the Point Spread Function (PSF) and the image intensity represents LPQ spatial blurring method. After applying LPQ operator at each pixel location, the results are presented as histogram codes which are insensitive to centrally symmetric blur such as out of focus and motion [63, 65].

3.2.2.1 LPQ Blur Invariant Using Fourier Transform Phase

Assuming that a blurred original image is *f(x)*, and an observed image is *g(x)*, then, the discrete model for spatially invariant blurring of *f(x)* can be expressed by a convolution [64]:

$$g(u) = f(u) \otimes h(u), \tag{3.2}$$

where *h(x)* is the PSF of the blur, $\otimes$ denotes 2-D convolution and x is a vector of coordinates. In the Fourier domain, this corresponds to:

$$G(u) = F(u) \otimes H(u), \tag{3.3}$$

where *G(u)*, *F(u)* and *H(u)* are the discrete Fourier transforms (DFT) of the blurred image *g(x)*, the original image *f(x)*, and the PSF *h(x)*, respectively, and u is a vector of coordinates $[u,v]^T$. The magnitude and phase can be separated from:

$$|G(u)| = |F(u)| \otimes |H(u)|, \tag{3.4}$$

$$\angle G = \angle F + \angle H \tag{3.5}$$

If we assume that blurred PSF *h(x)* is centrally symmetric, namely *h(x)=h(-x)*, its Fourier transform is always real-valued, and as a consequence its phase is only a two-valued function, given by:

$$\angle H(u) = \begin{cases} 0, & H(u) \geq 0 \\ \pi, & H(u) < 0 \end{cases}$$

The phase for each pixel is computed, then, the image is quantized by considering the sign of the local phase which includes imaginary and real part as shown in Figure 14. The quantized neighborhood of each pixel is reported as an eight digit binary string. Next, local histograms with 256 bins dimensional feature vector are computed. Then, for different window sizes and radii, the concatenated histogram descriptor is computed. In this study, radii 5, 7, 9 and 11 (different window sizes) of LPQ are implemented and compared. The normalized ear images and their corresponding LPQ codes are shown in Figure15.

3.2.3 Binarized Statistical Image Features (BSIF)

Binarized Statistical Image Features was recently proposed by Kannala and Rahtu [66]. It has been used for face and ear recognition and texture classification. The BSIF descriptor has two parameters: the size of the filter *(l)* and the length of the binary string *(n)*. Each bit in the binary string code is related to different filters and the number of these filters are assigned by required number of the bit string [67].

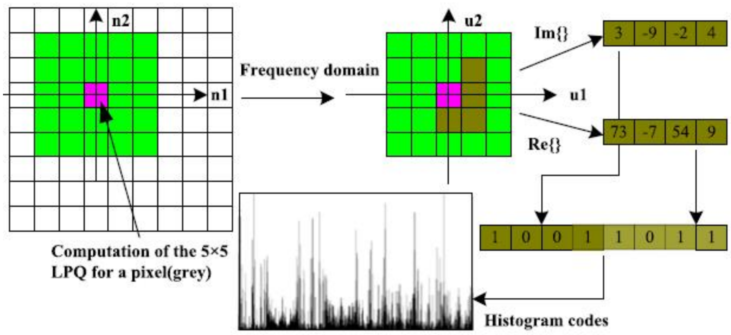


Figure 14: Procedure of Computing LPQ

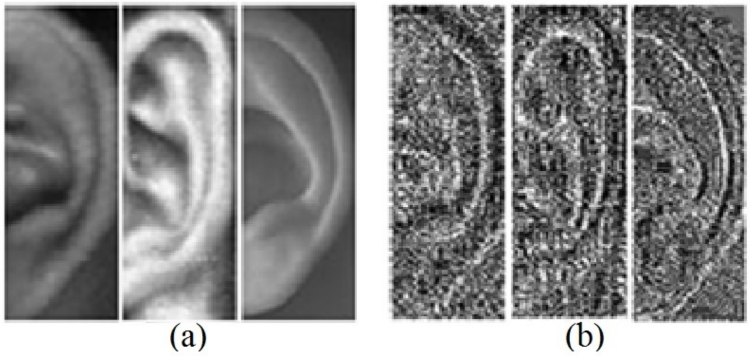


Figure 15: (a) Normalized Ear Images (b) Local Phase Quantization Codes

The descriptor in BSIF is determined based on the statistical properties of image patches therefore it is called Binarized Statistical Image Feature (BSIF). In BSIF, an image patch X of size $l \times l$ pixels and a linear filter Wi of the same size are used to obtain the filter response *Si* [68]:

$$s_i = \sum_{u,v} W_i(u,v)X(u,v) = w_i^T x, \tag{3.6}$$

Vectors *w* and *x* contain the pixels of *Wi* and *X*. The binarized feature bi is obtained by setting *bi* = *1 if si* > *0* and *bi* = *0* otherwise.

After mapping the binary code to real value (0 to 2^x), the final histogram is constructed by theses values. To achieve a good performance in ear recognition using BSIF, filter size and filter length should be considered. In this study, we use the standard filters that extract a local descriptor for different window sizes, overlap between neighboring windows and different filter sizes and concatenate each local histogram to a global histogram representation. The original filters are proposed by Kannala and Rahtu, which are available online for use and test. These filters were learned from 50,000 image patches. Figure16 shows an example of filters with factors l =*7*, n = *8*. Figure 17 shows samples of normalized ear images and BSIF code representation.

In this study, for all experiments, we use 8-bit, 9-bit, 10-bit, 11-bit, 12-bit code words and 5×5, 11×11, 13×13, 15×15, and 17×17 filters.

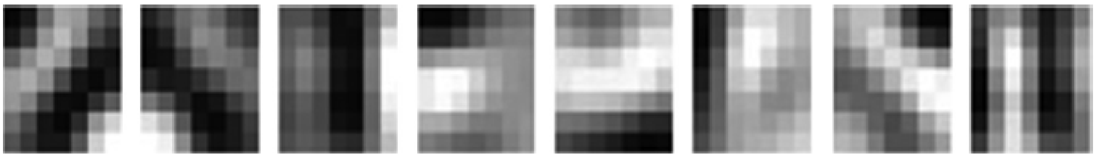


Figure 16: An Example of BSIF Filters with 7×7 Pixels

On the other hand, other local approaches, namely HOG and SIFT, are used for comparison purposes with the Proposed Approach 1 as presented below.

3.3 Histogram of Oriented Gradients (HOG)

Histogram of Oriented Gradients feature descriptor is used in image processing field, and it is one of the texture descriptors used for ear recognition [63, 69]. Computation of the global HOG descriptor [70] includes five steps as described in the following subsection.

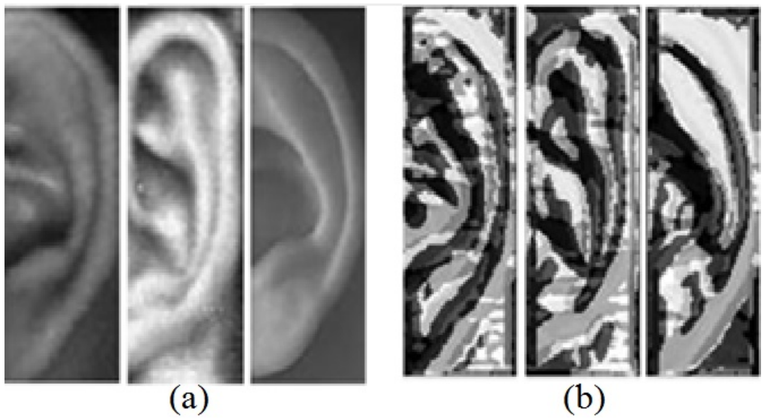


Figure 17: (a) Normalized Ear Images (b) Binarized Statistical Image Features Codes

3.3.1 HOG Algorithm

Step 1: Gradient Computation

A pair of filters, [-1 0 1] and [-1 0 1 $]^T$, are convolved wiht 3×3 HOG cells to compute the local gradient values. For each cell, the local orientation is obtained by the weighted sum of the responses of filter for each pixel.

Step 2: Orientation Binning

Quantizing the local orientations within blocks, which includes group of cells, into bins in the [0, π] interval or [0, 2π] interval.

Step 3: Histogram Computation

Group the cells together into larger blocks of equal size as shown in Figure 18, and a local histogram of quantized orientations is extracted.

Step 4: Histogram Normalization

Normalization process is performed on local histogram using one of the normalization factors such as L1-norm, L2-norm or L1-sqrt to overcome the variation in lighting and contrast as given below:

$$L2\!-\!norm : f = \frac{v}{\sqrt{||v||_2^2 + e^2}}, \tag{3.7}$$

$$L1\!-\!norm : f = \frac{v}{(||v||_1 + e^2)}, \tag{3.8}$$

$$L1\!-\!sqrt : f = \sqrt{\frac{v}{(||v||_1 + e^2)}}, \tag{3.9}$$

where v is the non-normalized vector containing all histograms in a given block, $\|v_k\|$ is its K- norm for K= 1,2 and e is a small constant.

Step 5: Concatenation Of Local Histograms

The final HOG descriptor is obtained by concatenating all the local histograms in image and the global descriptor is used to compare train and test images.

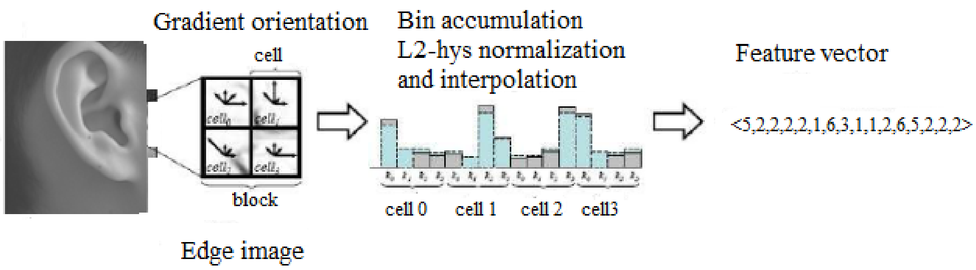


Figure 18: The Steps of HOG Algorithm

3.4 Scale-Invariant Feature Transformation (SIFT)

Scale-Invariant Feature Transformation (SIFT) was proposed by David Lowe in 1999. Detecting keypoint locations in the image and computing separate descriptors for each of the keypoints are the main steps of SIFT approach. There are many studies that used SIFT technique in ear recognition and achieved satisfying result [39, 71]. The major steps used to generate the set of stable features in both location and scale are described below [72]:

1- Scale- space extrema detection: in this stage, the location and scale of the interest points are recorded using Difference of Gaussian function in order to identify potential interest points that are invariant to scale and orientation and remove low contrast and unstable edge points.

2- Orientation assignment: orientations, location, and scale are assigned to each selected feature. By quantizing the orientations into 36 bins, a histogram is formed. The result of this step determines multiple keypoints with different orientations for the same scale and location.

3- Keypoint descriptor: the image gradients are measured around each keypoint, and the gradient strength and direction of neighborhood are computed. Based on 4*4 subregions, 8 bins exist in each subregion. The total number of subregions is 4*4*8=128 dimensions.

4- Matching: finally, the ear image is matched by individually comparing each feature from the ear image to the database and finding candidate matching features based on Euclidean Distance of their feature vectors. Figure 19 shows the matching result of ears of the same individuals.

Additionally, one of the global approaches which is also used for comparison purposes is Principal Component Analysis (PCA).

Figure 19: Comparison of Two Ear Image by Using SIFT Keypoint Matching

3.5 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) was proposed by Karl Pearson. It is a way for expressing the data and highlighting the similarities and differences for classification purposes. PCA is one of the earliest statistical methods proposed for face and ear recognition [42, 73, 74], The main goal of Principal Component Analysis (PCA) is to reduce the dimensionality of a data set consisting of many variables correlated with each other, the challenge here is to project a high-dimensional data onto a smaller dimensional subspace while retaining most of the discriminatory information. In order to reduce the information loss during dimensionality reduction process, the best low-dimensional space can be determined by the best principle component [53].

A Summary of PCA technique [75] is as follows:

- The data is standardized and the mean of the stored data is calculated.
- Covariance matrix is calculated.
- Eigenvectors and Eigenvalues are obtained from the covariance matrix.
- Eigenvalues are sorted (highest to lowest) and the k eigenvectors are chosen that correspond to the k largest eigenvalues where k is the number of dimensions of the new

feature subspace ($k \leq d$).

- The projection matrix W is constructed from the selected k eigenvectors.

- The original dataset X is transformed via W to obtain a k-dimensional feature subspace Y.

Chapter 4

DESCRIPTION OF DATABASES

Different subsets of ear databases are employed to perform a set of experiments in order to investigate the performance of our proposed systems. In this thesis, we used USTB ear database [76] including USTB- set1, USTB-set2, USTB-set3 datasets and UBEAR database [77].

4.1 USTB Ear Datasets

USTB database contains ear images that were captured by University of Science and Technology, Beijing under different conditions of illumination and pose. The following subsections have a brief overview on each ear database separately.

4.1.1 USTB Database-Set1

USTB database-set1 contains 60 users in which each has at least 3 samples. There exist totally 180 images in the database. This dataset contains three right ear samples for each subject and all the samples are captured under standard conditions. Some samples from this dataset are shown in Figure 20.

4.1.2 USTB Database-Set2

USTB database-set2 includes 308 images of 77 users in which each has 4 samples for each user. The first sample is captured under standard conditions, the other samples are captured with different angle (30° or -30°) or under weak illumination. Sample images from this dataset are demonstrated in Figure 21.

Figure 20: Ear Samples of USTB Database-set1 from Different Subject

Figure 21: Ear Samples of USTB Database-set2 from Different Subject

4.1.3 USTB Database-Set3

USTB database-set3 consists of images of full right profile faces including ear for 79 users. For each user, 10 images were captured under specific angles, with images turned to left by 0°, 5°, 10 °, 15° and 20 °. These images have partial, trivial, and regular occlusions as shown in Figure 22. Two samples from this dataset are presented in Figure 23.

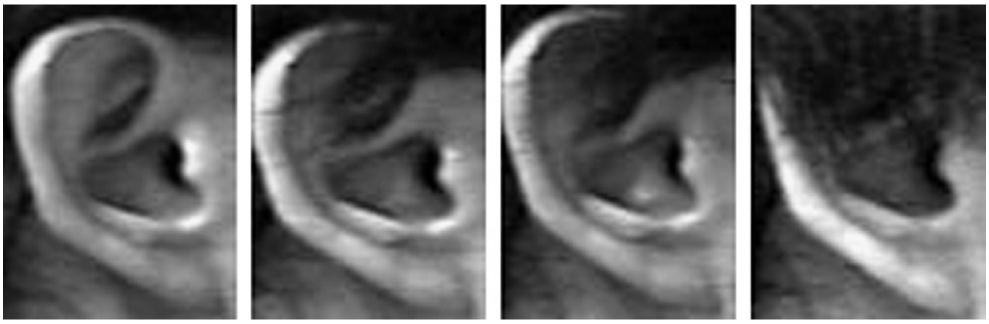


Figure 22: Samples of Occluded Ear Images of USTB Database-set3

Figure 23: Samples of USTB Database-set3

Figure 24: Samples of UBEAR Database (Different Illumination)

4.2 UBEAR Dataset

The UBEAR dataset images [77] were captured under different challenges namely partial occlusions and illumination in addition to blurring. The right and left profile face images were acquired from recorded video while the individuals were moving. It includes 126 subjects and 4429 samples of profile face (left and right) as shown in Figures 24 and 25.

Figure 25: Samples of UBEAR Database (Different Poses)

Chapter 5

EAR RECOGNITION BASED ON FUSION OF EAR AND TRAGUS UNDER DIFFERENT CHALLENGES

5.1 Preparatory Work

The contribution of this study is to use ear and tragus for person recognition by applying different fusion techniques on ear and tragus. In order to obtain high accuracy in recognition system, different global and local techniques of feature extraction were examined to find the most convenient method to recognize a human being using ear and tragus.

In the Proposed Approach 1, four different feature extractors, namely Local Binary Patterns (LBP) [78], Histogram of Oriented Gradients (HOG) [70], Scale-Invariant Feature Transform (SIFT) [79] and Principal Component Analysis (PCA) [3] are implemented and compared as shown in Table 1. According to the comparison results, LBP is selected for feature extraction due to the highest recognition rates compared with other algorithms. The comparison experiments are conducted under standard conditions using USTB-1 dataset. Additionally, 5$\times$5 segments are empirically selected to be used for LBP algorithm in this study.

Table 1: Comparison of feature extraction algorithms under standard conditions applied on USTB-1 dataset

		Algorithms			
		LBP	HOG	SIFT	PCA
Recognition Rate (%) with	Tragus	93.3	91.6	91.6	83.3
	Ear	100	98.3	96.6	90

On the other hand, two widely used classifiers, namely Support Vector Machine (SVM) [80] with linear kernel and k-Nearest Neighbor (k-NN) [81] classifier, are used to compare tragus and ear recognition performance as presented in Table 2. The performance using SVM and k-NN classifiers are equal in most of the cases and slightly different in two cases (shown in bold on Table 2). On the other hand, the execution time of SVM is high compared to k-NN which is very significant as it requires less time for identification. Because of the simplicity and less computation time of k-NN compared to SVM classifier, k-NN classifier is used in further experiments using other datasets under occlusion, pose and illumination challenges.

5.2 Description of the Proposed Technique

Ear biometrics have many limitations such as illumination variation, occlusion, and pose variation as described earlier. The presence of these problems in the ear image prevents taking the advantage of discriminate features of ear. Using the fusion stage, the aforementioned challenges, which are decreasing the system performance, can be solved for ear recognition system.

The Proposed Approach 1 is mainly based on a fusion of two parts of the same sample of ear, namely tragus and any other part of the ear that is not suffering from occlusion.

Tragus is often not affected by occlusion because its location is away from hair and accessories compared with other parts of the ear and it is the most apparent part in case of rotation. Fusion is conducted by score-level fusion technique implemented in the same way as in [24, 82, 83].

Table 2: Comparison of classifiers under horizontal and vertical occlusions applied on USTB-1 dataset

Occlusion Type	Biometric Trait	Classifier	Recognition Rate Under Occlusion Ratio(%) 0	10	20	30	40	50
Horizontal	Tragus	k-NN	93.3	93.3	93.3	93.3	93.3	93.3
		SVM	93.3	93.3	93.3	93.3	93.3	93.3
	Ear	k-NN	100	100	98.3	98.3	86.6	**65**
		SVM	100	100	98.3	98.3	86.6	**66.6**
Vertical	Tragus	k-NN	93.3	93.3	93.3	93.3	93.3	93.3
		SVM	93.3	93.3	93.3	93.3	93.3	93.3
	Ear	k-NN	100	100	100	96.6	**86**	75
		SVM	100	100	100	96.6	**85**	75

The following is the explanation steps that are applied in the Proposed Approach 1:

Step 1: Preprocessing is important step before starting recognition. All images (ear and tragus) are histogram equalized (HE) for adjusting image intensities to improve the contrast of the image, and Mean Variance Normalization (MVN) is used to spread the energy of all images and minimizes the noise of the image and the variation in illumination.

Step 2: All the entire ear and tragus images are divided into several blocks and Local Binary Patterns method is performed. Each image of ear and tragus are divided into 5×5 windows to obtain 25 windows for each trait.

Step 3: A global LBP feature vector is generated by concatenating histogram of all divided windows for ear and tragus, separately.

Step 4: In the matching step, Manhattan distance, as represented in equation (5.1), is used to determine the match score between test and train feature vectors.

$$d_{x,y} = \sum_{i=1}^{n} |x_i - y_i|, \tag{5.1}$$

where x and y denote the feature vectors of length n.

Step 5: For normalization process, the individual score of the ear and tragus are normalized using most efficient scheme which is tanh normalization and represented as:

$$S'_k = \frac{1}{2}\left\{tanh\left(0.01\frac{(S_k - \mu_{GH})}{\sigma_{GH}}\right) + 1\right\}, \tag{5.2}$$

where S'_k represents the normalized score for $k=1,2,...,n$; μ_{GH} and σ_{GH} are the mean and standard deviation, respectively.

Step 6: An efficient and simple fusion technique (Sum Rule) is applied to combine the normalized scores of ear and tragus.

Step 7: k-Nearest Neighbor (k-NN) classifier is used for classification stage. The block diagram of the first proposed fusion scheme is shown in Figure 26.

The information fusion of ear and tragus can be performed at feature-level fusion and score-level fusion. Fusion between tragus and the uncovered segment of ear at score level is done by constructing two separated templates for both tragus and uncovered ear's segment.

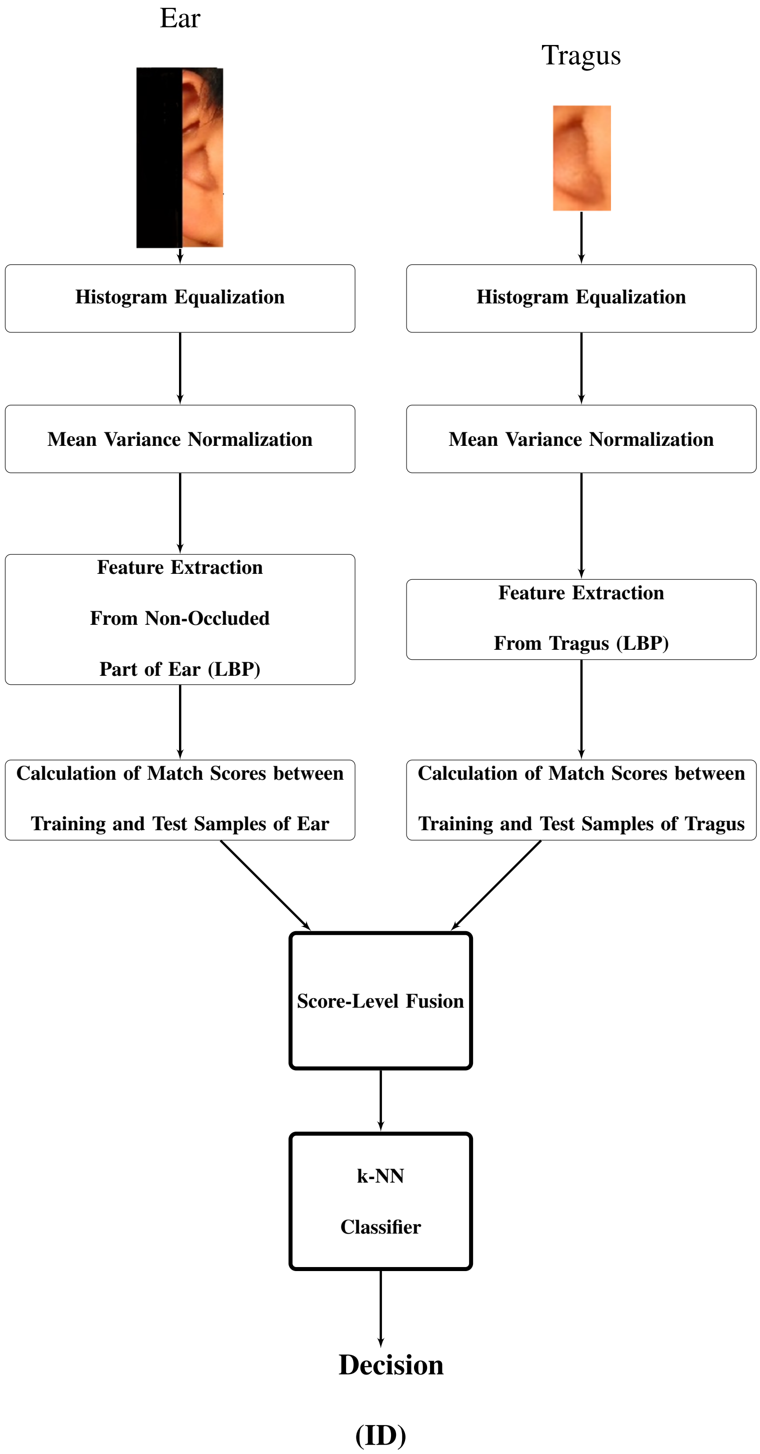


Figure 26: Block Diagram of The Proposed Approach 1

Then, the matching score between enrolled and test templates for each trait is calculated. Finally, the matching scores of tragus (S_{tragus}) and ear's segment ($S_{\text{ear seg}}$) are fused together to give the fused score (S_{fused}). Score fusion is calculated by transformation-based fusion using Weighted Sum Rule as follows:

$$S_{\text{fused}} = (W_{\text{ear seg}} \times S_{\text{ear seg}}) + (W_{\text{tragus}} \times S_{\text{tragus}}) \quad (5.1)$$

where $W_{\text{ear seg}}$ and W_{tragus} are the weights of ear and tragus, respectively. The values of weight are selected after many experiments. The best recognition rates under standard conditions are acquired using the values 0.75 and 0.25 for the weights of ear and tragus, respectively.

In feature-level fusion, multiple feature sets of the same individual are consolidated together to obtain fused template (T_{fused}). The aforementioned strategy is performed in this study for comparison purposes whenever the fusion of the feature templates of the tragus (T_{tragus}) and uncovered ear segment ($T_{\text{ear seg}}$) are implemented as follows:

$$T_{\text{fused}} = (\, T_{\text{tragus}} \cup T_{\text{ear seg}}) \quad (5.2)$$

In the final step of the Proposed Approach 1, the fused scores are used in the classification process to reach the final decision.

5.3 Experiments of the Proposed Approach 1

The validity of the Proposed Approach 1 is estimated by conducting many experiments over the three sets of USTB and UBEAR database as explained in the following four subsections.

5.3.1 Experiments on USTB Dataset 1

In USTB database-set1 (USTB-1), the ear images and tragus were considered separately. The ear image was taken under partial occlusion challenge by creating a mask that covers test images of the ear with 10%, 20%, 30%, 40% and 50% occlusion ratio.

The position of occlusion is not fixed and predictable, so two different ways of occlusion are simulated as horizontal occlusion (as shown in Fig.27), and vertical occlusion (as shown in Fig.28). In both occlusion strategies, tragus is not covered in any case.

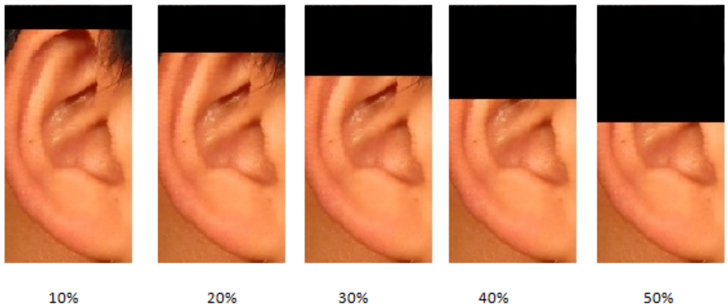


Figure 27: Percentage of Horizontal Occlusion

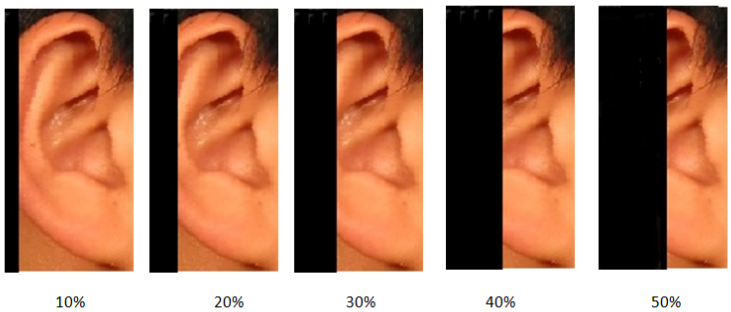


Figure 28: Percentage of Vertical Occlusion

The features of occluded test samples are extracted by LBP algorithm. Matching score is obtained between test image (occluded) and training images (not occluded) using k-NN classifier. Accuracy for various occlusion ratio for each ear and tragus separately are shown in Table 3 and Table 4. It is clear that the accuracy decreases when occlusion ratio increases. By USTB-1, efficiency of the Proposed Approach 1 is evaluated under the challenge of occlusion. Test image of the ear is masked by different occlusion ratios as 0%, 10%, 20%, 30%, 40% and 50% and is fused with tragus images using feature-level fusion and the Proposed Approach 1 with score-level fusion. It is demonstrated that both of these fusion techniques record the same very high performance up to 50% vertical and horizontal occlusion scenarios.

Table 3: Recognition rates (%) on USTB- set1 (horizontal occlusion)

Occlusion Ratio(%)		0	10	20	30	40	50
Recognition Rates with	Tragus	93.3	93.3	93.3	93.3	93.3	93.3
	Ear	100	100	98.3	98.3	86.6	65
Fusion of Ear and Tragus							
	Fusion Strategy \ Occlusion Ratio(%)	0	10	20	30	40	50
Recognition Rates with	Feature-level fusion	100	100	100	100	100	98.3
	Proposed Method	100	100	100	100	100	98.3

Table 4: Recognition rates (%) on USTB- set1 (vertical occlusion)

Occlusion Ratio(%)		0	10	20	30	40	50
Recognition Rates with	Tragus	93.3	93.3	93.3	93.3	93.3	93.3
	Ear	100	100	100	96.6	86	75
Fusion of Ear and Tragus							
	Fusion Strategy \ Occlusion Ratio(%)	0	10	20	30	40	50
Recognition Rates with	Feature-level fusion	100	100	100	100	100	98.3
	Proposed Method	100	100	100	100	100	98.3

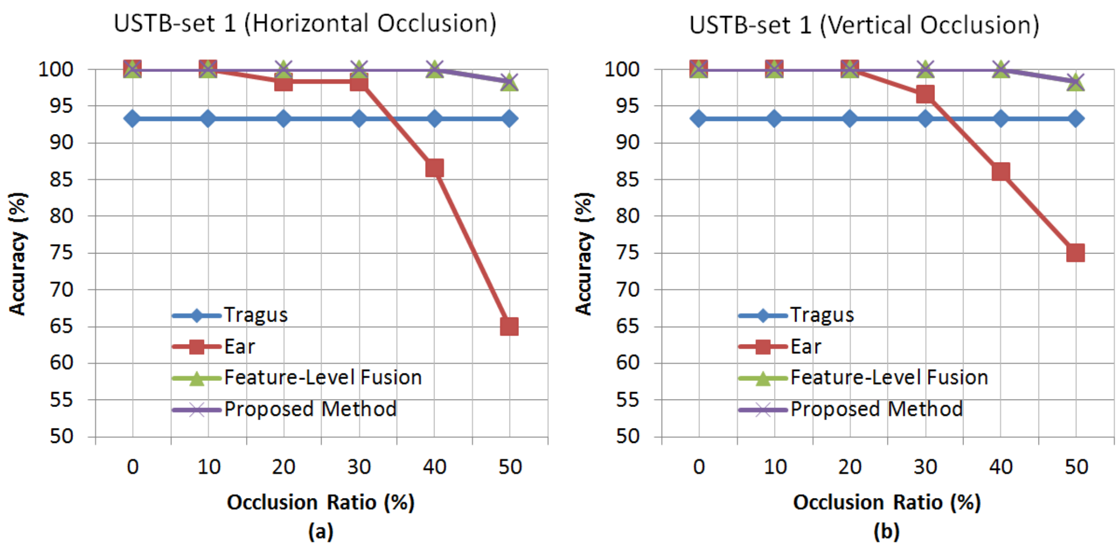


Figure 29: Recognition Rates for the Proposed Method on USTB Ear Dataset-1 with Various Levels of Occlusion

5.3.2 Experiments on USTB Dataset 2

The images of ear in USTB dataset 2 are captured under different challenges namely pose and illumination variation. The dataset includes 77 users, each having 4 samples. The first sample is frontal (with angle 0°) ear image under standard illumination condition. The second and the third samples are captured under -30° and +30° angles (different poses), respectively; and the fourth sample is captured under weak illumination condition. In order to demonstrate the effectiveness of the Proposed Approach 1, experiments are conducted in this section under occlusion, pose variation and weak illumination challenges. The experiments in this section are divided into three different cases. For the first case, in dataset 2, the ear images and tragus were taken separately. The ear image was taken under partial occlusion challenge by covering the test image of the ear by different horizontal occlusion ratios as 10%, 20%, 30% and 40%. Test samples in this case also suffer from weak illumination in addition to occlusion challenge as shown in Table 5. For the second case, test images of the ear are captured in two different poses in -30° and +30° angles. From Table 5, it can be observed that the pose variation has more influence on ear recognition performance than other factors such as occlusion and illumination. Tragus helps to enhance the performance in both left and right rotation cases when using score-level fusion of ear and tragus. For the third case, test images are captured under the challenge of weak illumination. This challenge negatively affects the accuracy as shown in Table 5. It is observed that the best results are obtained from fused (ear and tragus) recognition system at score-level fusion under occlusion problem, pose varying and weak illumination are 91.1%, 88.31% and 97.4%, respectively. The reason for this enhancement is that although the ears suffer from different challenges, tragus modality is able to provide discriminative features for identification because it is almost free from occlusion and it is clearly seen in left and right ears rotation.

Table 5: Recognition rates (%) on USTB-set2 (horizontal occlusion, pose variation and weak illumination)

		Occlusion Ratio(%) & Weak Illumination				Pose Variation (°)		Illumination
	Segment \ challenge	10	20	30	40	30	-30	Weak
Recognition Rate with	Tragus	74.02	74.02	74.02	74.02	38.9	40.25	74.02
	Ear	78.8	73.1	71.25	66.66	79.22	79.22	88.31
Fusion of Ear and Tragus								
Recognition Rate with	Feature-Level Fusion	90.6	89.5	88.95	85.2	72.72	83.11	96.10
	Proposed Method	91.1	90	89.82	85.5	81.81	88.31	97.4

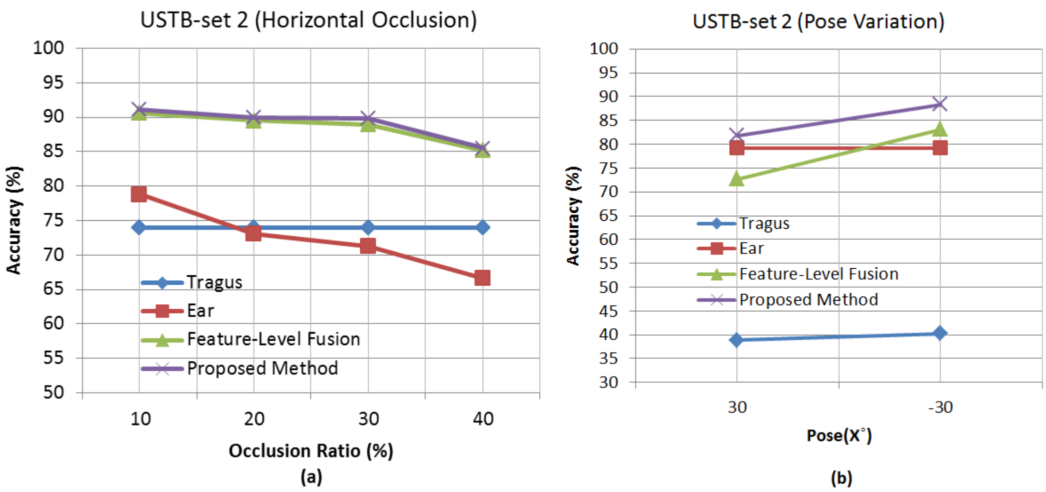


Figure 30: Recognition Rates for the Proposed Method on USTB Ear Dataset-2 with Various Levels of Occlusion and Pose Angles

5.3.3 Experiments on USTB Dataset 3

Samples in USTB-3 are captured under pose variation (with angles of 0°, 5°, 10°, 15° and 20°), therefore two different cases of experiments are conducted. In the first case, test images are horizontally and vertically covered with occlusion ratios as 10%, 20%, 30%, 40%, 50% and 60 %. In this case, test images suffered from two different challenges as partial occlusion and pose variation. Table 6 and Table 7 show the accuracy with horizontal and vertical occlusions, respectively. In the second case, as shown in Table 8, test images only suffer from pose variation (with angles of 5°, 10°, 15° and 20°) with 0% occlusion. It can be observed from Table 8 that the pose variation has more influence on ear recognition performance than other factors such as occlusion and illumination. We have noticed that the enhancement ratio increases by increasing the occlusion ratio. A slight improvement can also be noticed on accuracy in the case of the pose varying problem as presented in Table 8.

Table 6: Recognition rates (%) on USTB-set 3 (horizontal occlusion & pose variation 5°,10°,15° 20°)

	Occlusion Ratio(%)	10	20	30	40	50	60
Recognition Rate with	Tragus	74.1	74.1	74.1	74.1	74.1	74.1
	Ear	96.6	96.3	95.6	93.2	86.2	80.1
Fusion of Ear and Tragus							
Recognition Rate with	Feature-Level Fusion	95.9	95.2	94	92.2	89.9	87.1
	Proposed Method	97.2	96.8	95.9	94.6	91.5	88.8

Table 7: Recognition rates (%) on USTB-set 3 (vertical occlusion & pose variation 5°,10°,15°, 20°)

	Occlusion Ratio(%)	10	20	30	40	50	60
Recognition Rate with	Tragus	74.1	74.1	74.1	74.1	74.1	74.1
	Ear	95.8	95.8	95.4	92.4	85.9	78.8
Fusion of Ear and Tragus							
Recognition Rate with	Feature-Level Fusion	95.7	94.8	94.5	92.3	91.1	89.4
	Proposed Method	96.5	96.3	96.3	94.3	92.3	89.2

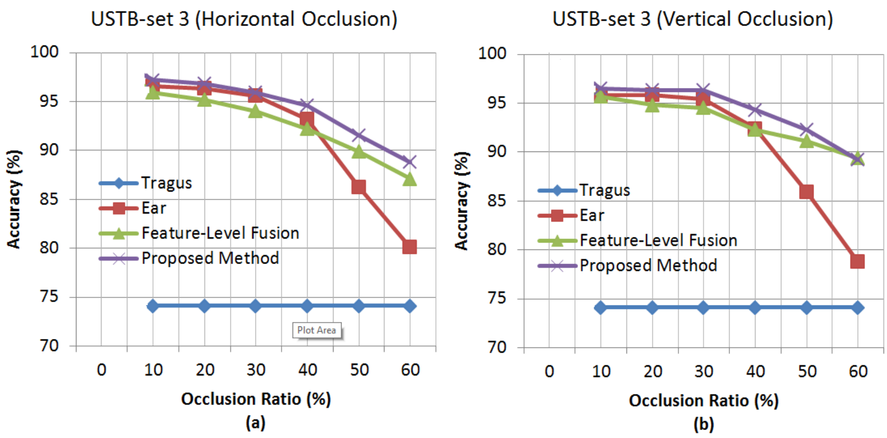


Figure 31: Recognition Rates for the Proposed Method on USTB Ear Dataset-3 with Various Levels of Occlusion

Table 8: Recognition rates (%) on USTB-set3 (pose variation)

	Pose Variation (°)	5	10	15	20
Recognition	Tragus	91.02	89.1	71.15	44.8
Rate with	Ear	100	98.01	96.1	92.3
Fusion of Ear and Tragus					
Recognition	Feature-Level Fusion	100	98.01	97.4	87.2
Rate with	Proposed Method	100	98.8	98.1	93.6

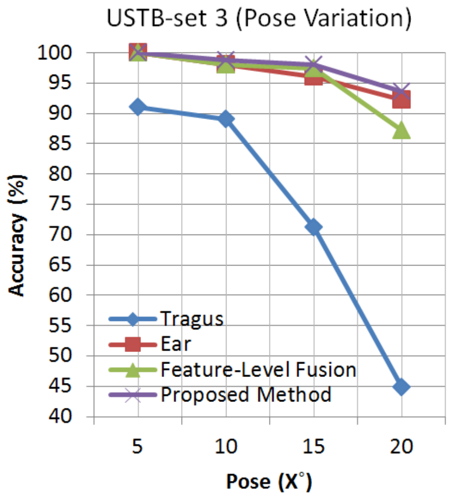


Figure 32: Recognition Rates for the Proposed Method on USTB Ear Dataset-3 with Various Pose Angles

5.3.4 Experiments on Real Occluded Ear Images

In order to test the proposed system under real occlusion challenge, the experiments in this part use ear images that are naturally occluded by hair. Eighty samples of 20 users are selected from UBEAR database. The selected samples suffer from non-uniform occlusion by hair as shown in Figure 33. Table 9 shows the accuracy of ear and tragus as unimodal systems and the accuracies of multimodal systems using feature-level and score-level fusion are also demonstrated. It is clearly shown that the proposed method is better than the unimodal methods using ear and tragus. Additionally, the proposed method has improved performance compared to the multimodal method using feature-level fusion of ear and tragus on real occlusion conditions.

Figure 33: Real Samples of Occluded Ear

Table 9: Recognition rates (%) on UBEAR database (real occlusion)

	Biometric Trait	Recognition rates (%)
Recognition Rates with	Tragus	82.5
	Ear	90
Fusion of Ear and Tragus		
	Fusion Strategy	Recognition rates (%)
Recognition Rates with	Feature-level fusion	95
	Proposed Method	97.5

5.4 Comparison of the Proposed System with the State-of-the-Art Systems

Finally, we compare the Proposed Approach 1 with several state-of-the-art methods involving 2D ear identification. Table 10 lists the recognition rates of the state-of-the-art methods on USTB-1, USTB-2 and USTB-3 datasets under different pose and occlusion conditions. Since there is no study on tragus and ear recognition with various challenges, the approaches used in the table include only ear identification results.

The comparison results show that the Proposed Approach 1 is better than most of the state-of-the-art ear recognition systems under different conditions.

Table 10: Comparison of recognition performance of different 2D ear identification methods on USTB-1,USTB-2,USTB-3 datasets

		Recognition Rate (%)		
Identification Approach	Feature Extraction Method	USTB-1	USTB-2	USTB-3
Yuan et al. (2006) [37]	Improved Non-Negative Matrix Factorization with Sparseness Constraints (INMFSC)	N/A	N/A	91 (10% Occlusion)
Wang et al. (2008) [10]	Uniform local binary patterns(ULBPs)and Haar wavelet transform	N/A	N/A	92.4 (Pose 20°)
Zhichun (2009) [11]	Independent Component Analysis (ICA)	N/A	N/A	90 (Pose 15°)
Guiterrez et al. (2010) [84]	Wavelet Transform & Neural Network	N/A	**97.5**	N/A
Wang and Yan (2011) [16]	Local Binary Pattern	N/A	92.2	N/A
Yuan and Mu (2012) [40]	Neighborhood Preserving Embedding	N/A	N/A	90 (50% Occlusion)
Tariq and Akram (2012) [34]	Haar wavelets	98.3	96.1	N/A
Zhang et al. (2013) [17]	Sparse Representation Classification SRC	N/A	N/A	**96.96** (20% Occlusion)
Yuan and Mu (2014) [85]	Gabor filter	N/A	N/A	96.46(0% Occlusion)
Omara et al. (2016) [35]	Shape of the ear	98.3	N/A	N/A
Benzaoui et al. (2017) [86]	BSIF descriptor/ Anatomical and Embryological information	98.97	N/A	N/A
Our Proposed approach	Fusion of Ear and Tragus Using Local Binary Pattern	**100**	97.4	**100**(0%Occlusion) **97.2**(10%Occlusion) 96.8(20% Occlusion) **92.3**(50%Occlusion) **98.1** (Pose 15°) **93.6** (Pose 20°)

5.4.1 Discussion on Experimental Results

The enrollment and testing phases that are used for ear are used in the same way for tragus. Due to the large size of the ear compared to the tragus, the recognition process of ear requires longer execution time than the process required for the tragus. All the previous tables show the performance in terms of recognition rate of ear and tragus individually in addition to the recognition rates of the Proposed Approach 1 which uses score-level fusion and for comparison purposes, we used feature-level fusion recognition rates of ear and tragus under different challenges using three sets of USTB database in addition to UBEAR database. In most cases, score-level and feature-level fusion systems are better than individual systems for the three datasets used. Additionally, the Proposed Approach 1 that uses score-level fusion outperforms feature-level fusion. The best recognition rate for USTB-1 with 40% occlusion and less is 100%, and with 50% occlusion is 98.3% as shown in Table 3 and 4. For USTB-2, the best recognition rates under 10% occlusion with weak illumination are 91.1%, 88.3% under -30° pose and 97.4% under weak illumination as shown in Table 5. For USTB-3, the best recognition rates are 100% under 5° rotated pose and 97.2% under 10% occlusion with pose variation as shown in Tables 6 to 8. Figures from 29 to 32 compare the individual systems with the proposed fused method under different challenges, and show the enhancement of the proposed fused system. In general, the proposed system is robust to pose and occlusion challenges and it is performing better than most of the other state-of-the-art ear recognition systems.

5.5 Conclusion of Proposed Approach 1

In this chapter, a novel recognition system based on the fusion between ear and tragus in a single captured image is proposed to overcome the effect of challenges of ear such as partial occlusion, pose variation and weak illumination. Features from ear and tragus are extracted using LBP algorithm with score-level fusion in the Proposed Approach 1. It has been observed that score-level fusion is superior to feature-level fusion in all experiments. Experimental results on three datasets show that our proposed technique is robust and effective since it gives better results than the other matching algorithms under different ear challenges. The maximum accuracies achieved are 100% (under partial occlusion), 97.4% (under weak illumination), 100% (under pose variation), 97.5% (under real occlusion) for USTB-set1, USTB-set2, USTB-set3, UBEAR respectively. The results obtained are better than most of the state-of-the-art ear recognition systems. Further work will focus on the improvement of the performance of the proposed system under other various ear challenges.

Chapter 6

MULTIMODAL BIOMETRICS FOR PERSON IDENTIFICATION USING EAR AND PROFILE FACE

6.1 Introduction

Biometrics systems aim to construct recognition system with minimum error rate by choosing any trait whose features are discriminant and not duplicated for different individuals [87, 88]. In this context, a fusion between the ear and other traits should be considered in order to construct a reliable and accurate system in all cases even when biometric samples suffer from different challenges such as occlusion.

In this study, the effect of profile face and ear traits in recognition of individuals is independently assessed. Both face and ear are passive in nature and the active part of the authenticator is not needed [49]. In some cases, ear trait is preferred instead of face traits due to some characteristics. For example, the variation of expression may change the appearance of face and it is found that it is strongly affected by ageing [49, 89]. Additionally, the background of ear is predictable and its color distribution is almost uniform.

Recent studies proved that some features of the ear are unique enough to recognize the similar persons such as identical twins [90]. This fact has important effects for security applications and makes the recognition performance of ear on the par with other biometric traits such as the fingerprint.

Ear as biometric trait is less commonly used compared to face biometrics because of the high discrimination of face biometrics. Consequently, the amount of features that can be extracted from face is more than the extracted features of ear. The serious challenges for both are lighting variation, pose variation, and occlusion. Profile face and ear are fused in this study because they can be easily captured in a single device and shot, which makes the time and cost of collecting the biometric data low compared to other fusion possibilities. Therefore, presenting the biometric trait to the system by individuals will be easier and more acceptable.

In the Proposed Approach 2, variation in lighting is considered since both face and ear are affected by variation in illumination, consequently, the sample images used in the experiments were captured under different lighting conditions (controlled and uncontrolled), additionally, many samples suffer from slight blurring and occlusion problem.

The Proposed Approach 2 is mainly based on Binarized Statistical Image Features (BSIF) algorithm in addition to two different types of fusion, namely score-level fusion and decision-level fusion. Fusion is conducted between ear and profile face. All possible binary combinations of fusion are considered using the following pairs of biometric traits: right ear - left ear (Figure 34-a), left profile - right profile (Figure 34-b), right profile - right ear (Figure 35-a), left Profile - left ear (Figure 35-b), left profile - right ear (Figure 35-c), and right profile - left ear (Figure 35-d).

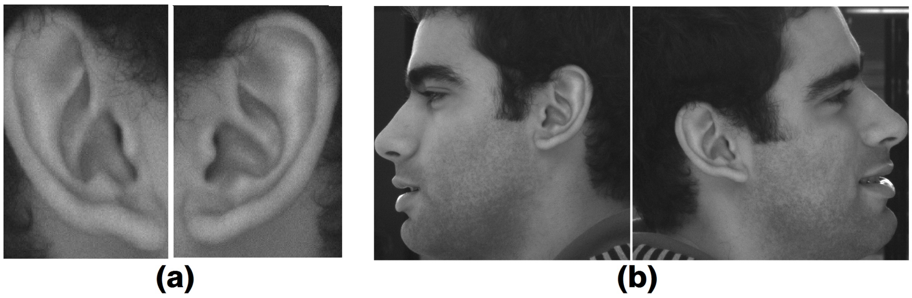


Figure 34: Examples of Different Sides of the Same Trait Used in Fusion (a) Right Ear-Left Ear (b) Left Profile-Right Profile

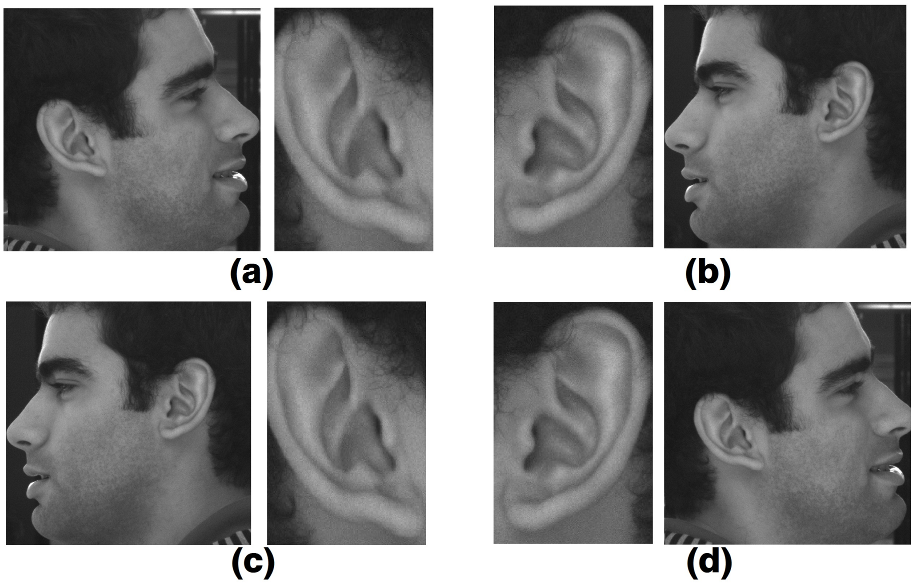


Figure 35: Examples of Different Traits Used in Fusion

6.2 Proposed Approach 2

In this study, we used multimodal biometrics for the recognition process. The Proposed Approach 2 mainly depends on applying fusion between two different modalities in order to enhance the recognition performance of biometric system. As previously described, fusion often increases the reliability and accuracy of the system. Therefore, it can be a good choice in the case of non uniform lighting condition. Profile face and ear are fused in two different levels of fusion, namely score-level and decision-level fusion. Six different combinations of fusion were implemented: right ear-right profile, right ear-left ear, left ear-left profile, right ear-left profile, left ear-right profile and right profile-left profile. Figure 36 shows the flowchart of one of the fusion combinations using left ear and left profile face with score-level fusion using BSIF feature extraction algorithm.

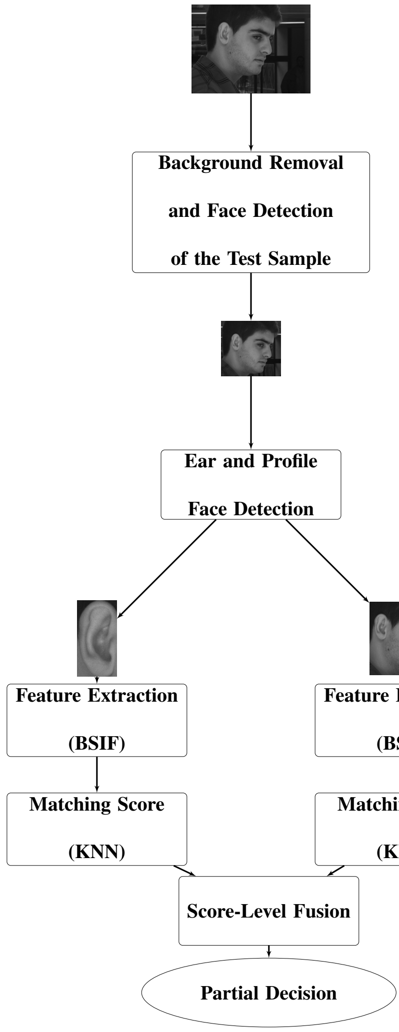


Figure 36: Flowchart of Score-Level Fusion of Right Ear and Right Profile Face

The final decision of the Proposed Approach 2 that uses score-level fusion of two different traits (ear and profile face) or two different samples of the same trait (right ear and left ear) can be found by applying majority voting of all the aforementioned combinations of fusion. Six different experimental outputs are used as partial decisions in order to make the final recognition decision as shown in Figure 37.

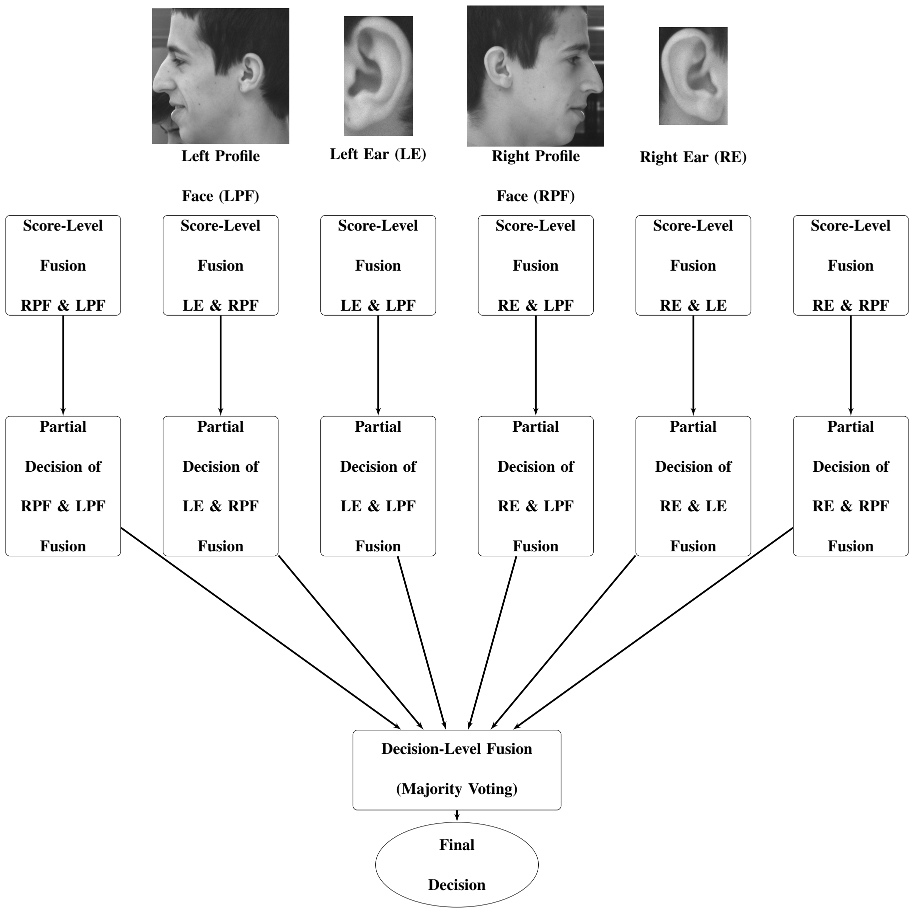


Figure 37: Block Diagram of the Proposed Approach 2

6.3 Experiments and Results

Seven different experiments are conducted on UBEAR database in order to demonstrate the efficiency of the Proposed Approach 2 for the recognition of individuals under different illumination conditions.

6.3.1 Fusion of Facial and Ear Data in Different Levels

In this study, three different levels of fusion are used: feature-level fusion which is used for comparison purposes, score-level fusion and decision-level fusion which are used in the Proposed Approach 2. Many biometric systems employ fusion of different levels in order to fuse more than one biometric system [82, 91].

Feature or representation-level fusion can be defined as the concatenation of multiple sets of feature of the same individual in one template to form a single feature set. In this study, a heterogeneous fusion technique is implemented by fusing multiple feature sets that are extracted from different traits (profile face and ear) using three feature extraction algorithms.

In score-level fusion approach, more than one match score of different traits or different biometric matchers can be fused in order to acquire a final recognition decision using the fused matching scores. In this study, transformation-based fusion, which is one of types of score-level fusion, with Sum Rule technique is employed as in [92].

On the other hand, fusion of multiple modalities may be performed after matching process using decision-level fusion methods. This is performed when only the decisions output by biometric systems are available. In this study, we used Majority Voting approach for the fusion of decision outputs obtained from the score-level fusion of different combinations of ear and profile faces.

6.3.2 Experimental Setup

In the Proposed Approach 2, 92 users of UBEAR dataset are used to evaluate the Proposed Approach 2 and each user has 16 samples. The description of UBEAR can be seen in Chapter 4. After detecting right and left profile face in addition to ear images, each trait is tested individually using three different feature extractors namely LBP with 5×5 segments, LPQ with radii 7 and BSIF with 8-bit code words and 17×17 filter. Then KNN classifier is used for classification. Score-level fusion and feature-level fusion of traits are implemented using all possible binary combinations of fusion such as right ear - right profile, right ear - left ear, left ear - left profile, right ear - left profile, left ear - right profile, and right profile - left profile. Feature extraction and matching are performed with the same experimental setup using LBP, LPQ and BSIF. In all experiments, accuracies for controlled and uncontrolled illumination conditions are calculated and presented in Tables 11 to 14.

The performances of the Proposed Approach 2 with the other methods employing all feature extraction algorithms that are used in unimodal and multimodal systems and fusion approaches of all possible combinations are measured under identification mode using recognition rate [4]. Additionally, Equal Error Rate (EER) is used under verification mode to evaluate the Proposed Approach 2 with all unimodal systems and the score-level fusion of binary combinations that use BSIF algorithm.

Match scores of all experiments are calculated by Manhattan Distance measure. In theory, client similarity scores of clients whose templates are stored in the database should always be higher than the scores of unknown users whose templates are not stored in the system. If this would be true, a single threshold that separates the scores of authorized people from unauthorized counterparts could be used to distinguish between clients and impostors. In real world of biometric systems, this assumption is not true since in some cases client samples generate scores that are less than the scores of some impostor samples. For that reason, it is a fact that however the classification threshold is chosen, some classification errors occur. Consequently, another performance measure such as EER can be used to employ a threshold independent system.

6.3.3 Experiments on Unimodal Systems

Unimodal systems experiments employ four different biometric traits namely right ear (RE), right profile face (RPF), left ear (LE) and left profile face (LPF). These traits are tested in unimodal mode. In each experiment, the samples are captured and tested under Controlled and Uncontrolled illumination conditions. The recognition rates under identification mode and EER's under verification mode of unimodal systems are shown in Tables 11 and 12, respectively. For each biometric trait, three feature extraction approaches are implemented namely LBP, LPQ and BSIF and the results are presented in the aforementioned tables.

6.3.4 Experiments on Multimodal Systems

All possible fusion combinations under identification mode of right ear, right profile face, left ear and left profile face using feature-level and score-level fusion are presented in Tables 13 and 14, respectively. Illumination challenge is also considered in the multimodal experiments. Fusion is conducted for each combination possibility using LBP, LPQ and BSIF feature extraction approaches. BSIF results are better than the results of LBP and LPQ for both feature-level and score-level approaches under all llumination conditions.

Table 11: Accuracy of unimodal systems using LBP, LPQ and BSIF feature extraction algorithms under identification mode

Recognition Rate (%)						
	Cont Illumination			Uncont Illumination		
Method / Trait	LBP	LPQ	BSIF	LBP	LPQ	BSIF
RE	90.38	90.38	94.23	82.5	80	86.25
RPF	90.38	92.30	96.15	83.5	82.5	88.75
LE	88.46	90.38	92.30	80	78.75	85
LPF	89.42	91.34	95.19	85	81.25	87.5

Table 12: Equal Error Rate (EER) of unimodal systems using LBP, LPQ and BSIF feature extraction algorithms under verification mode

EER (%)						
	Cont Illumination			Uncont Illumination		
Method / Trait	LBP	LPQ	BSIF	LBP	LPQ	BSIF
RE	15.2	14.8	8.5	22.7	24.3	17.6
RPF	13.6	9.3	7.1	17.8	21.2	12.9
LE	14.2	11.3	9.7	25.7	26.3	18.6
LPF	11.7	10.1	6.4	19.6	20.6	14.3

6.3.5 Experiments on the Proposed Approach 2

The Proposed Approach 2 is applied to find the final recognition decision based on the outputs of score-level fusion that uses BSIF algorithm in all possible binary combina-

Table 13: Accuracy of multimodal systems using feature-level fusion with LBP, LPQ and BSIF under identification mode

Recognition Rate (%)						
	Cont Illumination			Uncont Illumination		
Fused Traits \ Method	LBP	LPQ	BSIF	LBP	LPQ	BSIF
RE-RPF	90.38	92.30	96.15	85.9	80.3	91.6
RE-LE	91.34	93.26	97.11	88.73	85.9	93.80
RE-LPF	92.30	94.23	97.11	89.13	86.7	92.16
LE-LPF	89.42	91.34	96.15	86.7	82.6	90.8
LE-RPF	93.26	94.23	97.11	89.13	85.9	92.16
RPF-LPF	94.23	95.19	97.11	90.3	88.73	95.73
LE-RE-LPF	94.23	94.23	96.15	91.6	87.5	95.73
LE-RE-RPF	94.23	95.19	97.11	91.6	87.5	95.73
LE-RPF-LPF	96.15	96.15	98.07	92.16	88.73	97.82
RE-RPF-LPF	96.15	96.15	98.07	91.6	88.73	97.28
RE-LE-RPF-LPF	96.15	97.11	98.07	92.16	89.6	97.82

tions of unimodal systems by conducting Majority Voting fusion technique as shown in Figure 37. Tables 15 and 16 show the recognition rates under identification mode and EER's under verification mode of the Proposed Approach 2 under different illumination conditions. The ROC curves of the Proposed Approach 2 in addition to two binary fusion approaches of profile face and ear under uncontrolled illumination are presented in Figure 38.

On the other hand, the accuracy of some state-of-the-art methods on ear and profile face biometric traits as unimodal and multimodal systems are listed in Table 18. It is

Table 14: Accuracy of multimodal systems using score-level fusion under identification mode

Recognition Rate (%)						
	Cont Illumination			Uncont Illumination		
Fused Traits \ Method	LBP	LPQ	BSIF	LBP	LPQ	BSIF
RE-RPF	91.34	92.30	96.15	85.9	80.3	91.6
RE-LE	91.34	93.26	97.11	89.13	85.9	93.80
RE-LPF	93.26	93.26	97.11	89.13	86.7	92.16
LE-LPF	90.38	92.30	96.15	87.5	82.6	90.8
LE-RPF	92.30	94.23	97.11	90.8	85.9	92.16
RPF-LPF	94.23	96.15	98.07	90.3	88.73	97.1
LE-RE-LPF	94.23	94.23	97.11	91.6	88.73	95.73
LE-RE-RPF	94.23	95.19	97.11	91.6	87.5	96.3
LE-RPF-LPF	96.15	97.11	99.03	92.16	87.5	98.3
RE-RPF-LPF	96.15	97.11	98.07	93.80	88.73	97.28
RE-LE-RPF-LPF	96.15	98.07	100	94.30	89.6	98.3

shown that multimodal biometric results are better than their unimodal counterparts. Since there is no study on UBEAR dataset including both ear and profile face biometrics, we added several state-of-the-art systems using different databases on ear and profile face. Although the table does not give an exact comparison between the presented approaches, it gives an idea on how to improve unimodal ear and profile face biometric systems using different feature extractors.

Table 15: Recognition rate of the proposed method under identification mode

	Traits \ Illumination	Recognition Rate (%) Controlled	Recognition Rate (%) Uncontrolled
Score-Level Fusion using BSIF	Right Ear-Right Profile Face	96.15	91.6
	Right Ear-Left Ear	97.11	93.80
	Right Ear-Left Profile Face	97.11	92.16
	Left Ear-Left Profile Face	96.15	90.8
	Left Ear-Right Profile Face	97.11	92.16
	Right Profile Face-Left Profile Face	98.07	97.1
Proposed Method		100	99.47

6.3.6 Discussion on Experimental Results

All the experimental results demonstrated that the variation in lighting conditions has a negative impact on the recognition rate. Mostly, the fused feature vector of profile face and ear of the same side contains redundant information, consequently, the recognition rate will negatively be affected compared with the recognition rate of the fused traits that are captured from different sides.

Comparing the feature extraction algorithms, it is clearly seen that BSIF approach outperforms LBP and LPQ in all cases of the experiments. On the other hand, LBP and LPQ show close performance under controlled illumination while LBP outperforms the more recent approach LPQ under uncontrolled illumination which means that LBP method still provides acceptable results for nonuniform illumination conditions.

Table 16: Equal Error Rate (EER) of the proposed method under verification mode

		EER (%)	
	Illumination / Traits	Controlled	Uncontrolled
Score-Level Fusion using BSIF	Right Ear-Right Profile Face	7.8	12.6
	Right Ear-Left Ear	5.25	6.1
	Right Ear-Left Profile Face	4.7	7.5
	Left Ear-Left Profile Face	6.5	9.8
	Left Ear-Right Profile Face	5.3	6.9
	Right Profile Face-Left Profile Face	4.2	6.3
Proposed Method		1.9	3.1

Tables 13 and 14 show the results of multimodal systems that fuse two, three and four biometric traits. The results proved the fact that increasing number of information sources may lead to enhance the recognition rates in some cases but they are usually more expensive than unimodal systems. This is because of the need for additional computational and storage resources in addition to larger enrollment and recognition processing times as presented in Table 17. Comparison of computation times of unimodal and multimodal systems under uncontrolled illumination is presented in Table 17. Consequently, the processing time of the multimodal systems that use only two biometric traits is less than the processing time of the systems that use three biometric traits and so on. Hence, the tradeoff between the extra cost and the benefits when making an application that uses multibiometrics approach should be analyzed. Table 17 also compares the computation times of BSIF, LPQ and LBP feature extraction ap-

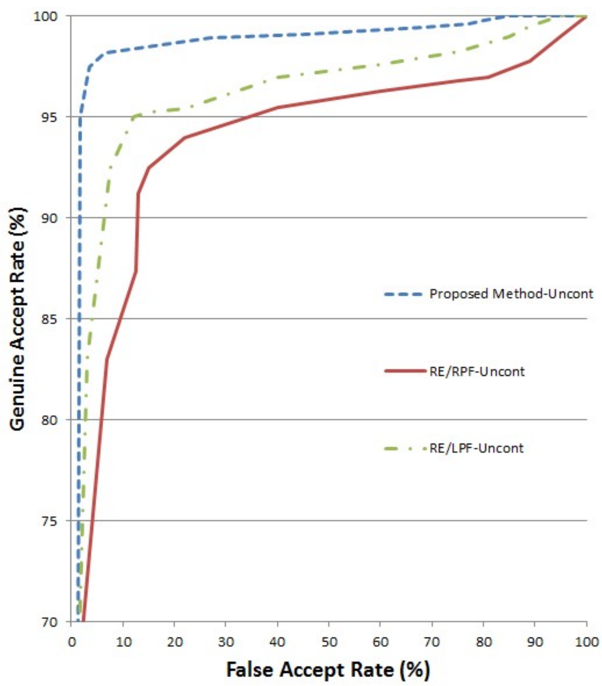


Figure 38: ROC Curve of the Proposed Approach 2 with RE-RPF and RE-LPF Fusion Under Uncontrolled Illumination

proaches. BSIF consumes the longest computation while LBP is the shortest, however BSIF outperforms the other feature extractors in terms of accuracy for all unimodal and multimodal systems presented in the table.

The performance of score-level fusion in multimodal systems is comparable in many cases and better in some of the cases than feature-level fusion, consequently, score-level fusion is used in the Proposed Approach 2. Additionally, the binary combinations of fusion are chosen to be used in the Proposed Approach 2 due to low processing time and the comparable performance compared with other multimodal systems that fuse three or four traits.

The Proposed Approach 2 overcomes the non-uniform illumination challenges and outperforms the unimodal and other multimodal methods fused with feature-level or score-level fusion techniques in all the cases that suffer from non-uniform illumination. In the case of controlled lighting, the recognition rates of the Proposed Approach 2 are better than or comparable to the unimodal biometric systems. Applying the Majority Voting on the results of the multimodal methods increases the best recognition rate that was acquired under uncontrolled illumination from 97.1% that was acquired by the fusion of right and left profile face to 99.47% as shown in Table 15. Therefore, the proposed approach 2 is better or achieves the highest recognition rate than the unimodal or the other multimodal systems under different lighting conditions.

6.4 Conclusion of Proposed Approach 2

The performance of biometric systems can be enhanced when different biometric traits are fused together compared with the recognition rates of unimodal systems. Fusion of two traits in different sides (Left-Right) produce higher accuracy than fusion of two traits on the same sides (Left-Left or Right-Right) since the traits that are captured from the same side may have redundant features. Recognition process is strongly affected by uncontrolled illumination. Our proposed method, which applies the score-level fusion of all possible binary combinations of ear and profile faces, employs Majority Voting technique to fuse partial decisions of all binary combinations of ear and profile face to reach a final decision. The Proposed Approach 2 achieves 100% recognition rate under controlled illumination conditions and 99.47% for uncontrolled illumination conditions which are better than the other multimodal fusion methods. As a future work, more realistic recognition systems constructed by considering further challenges such as pose variation, partial occlusion and beauty surgery of face and ear will be studied. Additionally, other biometric traits should be used and tested as unimodal and multimodal recognition systems under different challenges.

Table 17: Comparison of computation times of unimodal and multimodal recognition approaches under uncontrolled illumination

Trait	Method	Accuracy (%)	Computation Time (s)
RE	LBP	82.5	9.025
	LPQ	80	13.77
	BSIF	86.25	21.04
RPF	LBP	83.5	11.29
	LPQ	82.5	14.53
	BSIF	88.75	16.92
RE+RPF	LBP	85.9	29.32
Score-Level Fusion	LPQ	80.3	31.05
	BSIF	91.6	63.54

Table 18: Unimodal and multimodal state-of-the-art approaches on ear and profile face biometrics

Identification Approach	Feature Extraction Method	Biometric Trait	Recognition rate (%)	Database
Unimodal Systems				
Benzaoui et al [86]	BSIF descriptor LPQ	Ear Ear	98.97 92.82	USTB-1 USTB-1
Anam and Usman [34]	Haar Wavelets	Ear	98.33	USTB-1
Omara et al [35]	LBP Geometric Measurements	Ear Ear	95.8 99.6	USTB-1 IIT Delhi
Nejati et al [90]	Expection Report Model (ERM)	Ear	92.77	Private DB
Xu et al [93]	Kernel Principal Component Analysis (KPCA)	Ear Profile Face	90.08 92.19	USTB-3 USTB-3
Rathore et al [48]	Speed-Up Robust Feature (SURF)	Ear Profile Face	98.1 99.03	IITK-1 IITK-1
Hezil et al [45]	BSIF Descriptor	Ear	98.9	IITDelhi-2
Our Study	BSIF Descriptor	Ear Profile Face	86.25 94.23 88.75 96.15	UBEAR (Uncontrolled Illumination) UBEAR (Controlled Illumination) UBEAR (Uncontrolled Illumination) UBEAR (Controlled Illumination)
Multimodal Systems				
Xu et al [93]	Kernel Principal Component Analysis (KPCA)	Ear + Profile Face	94.52	USTB-3
Rathore et al [48]	Speed-Up Robust Feature (SURF)	Ear + Profile Face	99.36	IITK-1
Xu and Mu [47]	Kernel Conical Correlation Analysis (KCCA)	Ear + Profile Face	98.86	USTB-3
Pan et al. [46]	Fisher Discriminate Analysis (FDA)	Ear + Profile Face	96.84	USTB-3
Our Study	BSIF Descriptor	Ear + Profile Face	98.3 100	UBEAR (Uncontrolled Illumination) UBEAR (Controlled Illumination)
Proposed Method	BSIF Descriptor	Ear + Profile Face	**99.47** **100**	UBEAR (Uncontrolled Illumination) UBEAR (Controlled Illumination)

Chapter 7

CONCLUSION

In this thesis, we have presented a literature review about ear biometrics which involves all cases of using ear biometrics where it contains ear as unimodal biometrics system, ear in multimodal systems and ear recognition under different challenges.

Exploiting the properties of tragus part of the ear, which is almost free from occlusion, and clearly seen with different poses, we have presented the Proposed Approach 1 based on fusion of tragus features and the remaining part of ear biometric. In the Proposed Approach 1, we have taken into account three challenges: firstly occlusion with different ratio and locations, secondly pose variation at specific angles, thirdly weak illumination. LBP algorithm is used for feature extraction process with score-level fusion in the Proposed Approach 1. The Proposed Approach 1 could be validated by conducting extensive experiments on available USTB database which includes three subsets namely USTB-1, USTB-2 and USTB-3. It seems from the results that the Proposed Approach 1 emerges as an efficient tool for ear recognition, the best accuracies achieved are 100% (under partial occlusion), 97.4% (under weak illumination), 100% (under pose variation) for USTB-set1, USTB-set2, USTB-set3, respectively. The score-level fusion outperforms feature-level fusion and the results obtained are better than most of the state-of-the-art ear recognition systems.

The key idea of the Proposed Approach 2 is to show that the ear biometrics can be integrated with other biometric traits in order to improve the recognition performance. Our proposed method and empirical results support the aforementioned fact. Fusion of the ear and profile face traits in different sides (Left-Right) was implemented using score-level fusion and decision-level fusion under lighting, blurring and occlusion challenges.

In the Proposed Approach 2, we show how the precise representation of the ear features plays significant role in enhancing the accuracy of the ear recognition. To achieve this purpose, we use different local texture descriptors namely LBP, LPQ and BSIF which are tested and compared in unimodal and multimodal systems. The results demonstrate that BSIF descriptor is superior to the other feature extractors in terms of accuracy for all unimodal and multimodal systems. However, despite the high accuracy of the BSIF, it must be noted that its processing time is longer compared to LBP and LPQ. The Proposed Approach 2 yielded excellent results compared with the state-of-the-art methods in unimodal and multimodal systems. The recognition rates under controlled and uncontrolled illumination conditions were 100% and 99.47%, respectively.

REFERENCES

[1] A. K. Jain, A. Ross, and S. Prabhakar, “An introduction to biometric recognition,” *IEEE Transactions on circuits and systems for video technology*, vol. 14, no. 1, pp. 4–20, 2004.

[2] A. Pflug and C. Busch, “Ear biometrics: a survey of detection, feature extraction and recognition methods,” *IET biometrics*, vol. 1, no. 2, pp. 114–129, 2012.

[3] A. K. Jain, A. A. Ross, and K. Nandakumar, “Introduction,” in *Introduction to Biometrics*, pp. 1–49, Springer, 2011.

[4] S. Prakash and P. Gupta, “Introduction,” in *Ear Biometrics in 2D and 3D*, pp. 1–20, Springer, 2015.

[5] B. Zhang, Z. Mu, C. Li, and H. Zeng, “Robust classification for occluded ear via gabor scale feature-based non-negative sparse representation,” *Optical Engineering*, vol. 53, no. 6, pp. 061702–061702, 2014.

[6] S. Prakash and P. Gupta, “An efficient ear recognition technique invariant to illumination and pose,” *Telecommunication Systems*, pp. 1–14, 2013.

[7] M. Burge and W. Burger, “The effect of time on ear biometrics,” in *In Proc of Intl Conf on Pattern Recognition(ICPR 00).*, pp. 822–826, 2000.

[8] S. J. Prince, J. H. Elder, J. Warrell, and F. M. Felisberti, “Tied factor analysis for face recognition across large pose differences,” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 30, no. 6, pp. 970–984, 2008.

[9] A. Asthana, T. K. Marks, M. J. Jones, K. H. Tieu, and M. Rohith, “Fully automatic pose-invariant face recognition via 3d pose normalization,” in *Computer Vision (ICCV), 2011 IEEE International Conference on*, pp. 937–944, IEEE, 2011.

[10] Y. Wang, Z.-c. Mu, and H. Zeng, “Block-based and multi-resolution methods for ear recognition using wavelet transform and uniform local binary patterns,” in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pp. 1–4, IEEE, 2008.

[11] D. W. M. Zhichun, “An ica-based ear recognition method through nonlinear adaptive feature fusion [j],” *Journal of Computer-Aided Design & Computer Graphics*, vol. 3, p. 017, 2009.

[12] D.-H. Liu, K.-M. Lam, and L.-S. Shen, “Illumination invariant face recognition,” *Pattern Recognition*, vol. 38, no. 10, pp. 1705–1716, 2005.

[13] V. M. Patel, T. Wu, S. Biswas, P. J. Phillips, and R. Chellappa, “Dictionary-based face recognition under variable lighting and pose,” *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 954–965, 2012.

[14] B. El-Desoky, M. El-Kady, M. Rashad, M. M. Eid, and A. Tharwat, "Ear recognition and occlusion," *International Journal of Computer Science & Information Technology*, vol. 4, no. 6, p. 97, 2012.

[15] X. Tan, S. Chen, Z.-H. Zhou, and J. Liu, "Face recognition under occlusions and variant expressions with partial similarity," *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 2, pp. 217–230, 2009.

[16] Z.-q. Wang and X.-d. Yan, "Multi-scale feature extraction algorithm of ear image," in *Electric Information and Control Engineering (ICEICE), 2011 International Conference on*, pp. 528–531, IEEE, 2011.

[17] B. Zhang, Z. Mu, C. Jiang, and J. Dong, "A robust algorithm for ear recognition under partial occlusion," in *Control Conference (CCC), 2013 32nd Chinese*, pp. 3800–3804, IEEE, 2013.

[18] L. Nanni and A. Lumini, "Fusion of color spaces for ear authentication," *Pattern Recognition*, vol. 42, no. 9, pp. 1906–1913, 2009.

[19] R. Raghavendra, K. B. Raja, and C. Busch, "Ear recognition after ear lobe surgery: A preliminary study," in *Identity, Security and Behavior Analysis (ISBA), 2016 IEEE International Conference on*, pp. 1–6, IEEE, 2016.

[20] A. A. Ross, K. Nandakumar, and A. K. Jain, *Handbook of multibiometrics*, vol. 6.

Springer Science & Business Media, 2006.

[21] N. Saini and A. Sinha, "Face and palmprint multimodal biometric systems using gabor–wigner transform as feature extraction," *Pattern Analysis and Applications*, vol. 18, no. 4, pp. 921–932, 2015.

[22] L. Shen, L. Bai, and Z. Ji, "Fpcode: An efficient approach for multi-modal biometrics," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 25, no. 02, pp. 273–286, 2011.

[23] R. Raghavendra, B. Dorizzi, A. Rao, and G. H. Kumar, "Designing efficient fusion schemes for multimodal biometric systems using face and palmprint," *Pattern Recognition*, vol. 44, no. 5, pp. 1076–1088, 2011.

[24] M. Eskandari and Ö. Toygar, "Selection of optimized features and weights on face-iris fusion using distance images," *Computer Vision and Image Understanding*, vol. 137, pp. 63–75, 2015.

[25] A. B. Mansoor, H. Masood, M. Mumtaz, *et al.*, "Personal identification using feature and score level fusion of palm-and fingerprints," *Signal, Image and Video Processing*, vol. 5, no. 4, p. 477, 2011.

[26] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern recognition*, vol. 38, no. 12, pp. 2270–2285, 2005.

[27] A. Afaneh, F. Noroozi, and Ö. Toygar, "Recognition of identical twins using fusion of various facial feature extractors," *EURASIP Journal on Image and Video Processing*, vol. 2017, no. 1, p. 81, 2017.

[28] A. V. Iannarelli, *Ear identification*. Paramont Publishing Company, 1989.

[29] M. Burge and W. Burger, "Ear biometrics in computer vision," in *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, pp. 222–226, IEEE, 1998.

[30] D. J. Hurley, M. S. Nixon, and J. N. Carter, "Automatic ear recognition by force field transformations," 2000.

[31] D. J. Hurley, M. S. Nixon, and J. N. Carter, "Force field energy functionals for image feature extraction," *Image and Vision computing*, vol. 20, no. 5, pp. 311–317, 2002.

[32] M. Choras, "Further developments in geometric algorithms for ear biometrics," in *In Proc of Intl Conf on Articulated Motion and Deformable Objects (AMDO06).*, pp. 58–67, 2006.

[33] L. Nanni and A. Lumini, "A multi-matcher for ear authentication," *Pattern Recognition Letters*, vol. 28, no. 16, pp. 2219–2226, 2007.

[34] A. Tariq and M. U. Akram, "Personal identification using ear recognition," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 10, no. 2, pp. 321–326, 2012.

[35] I. Omara, F. Li, H. Zhang, and W. Zuo, "A novel geometric feature extraction method for ear recognition," *Expert Systems with Applications*, vol. 65, pp. 127–135, 2016.

[36] A. Benzaoui, A. Hadid, and A. Boukrouche, "Ear biometric recognition using local texture descriptors," *Journal of Electronic Imaging*, vol. 23, no. 5, pp. 053008–053008, 2014.

[37] L. Yuan, Z.-c. Mu, Y. Zhang, and K. Liu, "Ear recognition using improved non-negative matrix factorization," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 4, pp. 501–504, IEEE, 2006.

[38] B. Arbab-Zavar and M. S. Nixon, "On guided model-based analysis for ear biometrics," *Computer Vision and Image Understanding*, vol. 115, no. 4, pp. 487–502, 2011.

[39] J. D. Bustard and M. S. Nixon, "Toward unconstrained ear recognition from two-dimensional images," *IEEE transactions on systems, man, and cybernetics-Part A: Systems and Humans*, vol. 40, no. 3, pp. 486–494, 2010.

[40] L. Yuan and Z. chun Mu, "Ear recognition based on local information fusion," *Pattern Recognition Letters*, vol. 33, no. 2, pp. 182–190, 2012.

[41] L. Yuan, W. Liu, and Y. Li, "Non-negative dictionary based sparse representation classification for ear recognition with occlusion," *Neurocomputing*, vol. 171, pp. 540–550, 2016.

[42] K. Chang, K. W. Bowyer, S. Sarkar, and B. Victor, "Comparison and combination of ear and face images in appearance-based biometrics," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 25, no. 9, pp. 1160–1165, 2003.

[43] L. Lu, X. Zhang, Y. Zhao, and Y. Jia, "Ear recognition based on statistical shape model," in *Innovative Computing, Information and Control, 2006. ICICIC'06. First International Conference on*, vol. 3, pp. 353–356, IEEE, 2006.

[44] K. Faez, S. Motamed, and M. Yaqubi, "Personal verification using ear and palmprint biometrics," in *Systems, Man and Cybernetics, 2008. SMC 2008. IEEE International Conference on*, pp. 3727–3731, IEEE, 2008.

[45] N. Hezil and A. Boukrouche, "Multimodal biometric recognition using human ear and palmprint," *IET Biometrics*, 2017.

[46] X. Pan, Y. Cao, X. Xu, Y. Lu, and Y. Zhao, "Ear and face based multimodal recognition based on kfda," in *Audio, Language and Image Processing, 2008.*

ICALIP 2008. International Conference on, pp. 965–969, IEEE, 2008.

[47] X. Xu and Z. Mu, "Feature fusion method based on kcca for ear and profile face based multimodal recognition," in *Automation and Logistics, 2007 IEEE International Conference on*, pp. 620–623, IEEE, 2007.

[48] R. Rathore, S. Prakash, and P. Gupta, "Efficient human recognition system using ear and profile face," in *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*, pp. 1–6, IEEE, 2013.

[49] K. Annapurani, M. Sadiq, and C. Malathy, "Fusion of shape of the ear and tragus–a unique feature extraction method for ear authentication system," *Expert Systems with Applications*, vol. 42, no. 1, pp. 649–656, 2015.

[50] Ž. Emeršič, V. Štruc, and P. Peer, "Ear recognition: More than a survey," *Neurocomputing*, 2017.

[51] R. S. Choras, "Image feature extraction techniques and their applications for cbir and biometrics systems," *International journal of biology and biomedical engineering*, vol. 1, no. 1, pp. 6–16, 2007.

[52] K.-L. Lim and H. K. Galoogahi, "Shape classification using local and global features," in *Image and Video Technology (PSIVT), 2010 Fourth Pacific-Rim Symposium on*, pp. 115–120, IEEE, 2010.

[53] I. T. Jolliffe, “Principal component analysis and factor analysis,” *Principal component analysis*, pp. 150–166, 2002.

[54] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces vs. fisherfaces: Recognition using class specific linear projection,” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[55] S. Brahnam, L. C. Jain, L. Nanni, A. Lumini, *et al.*, *Local binary patterns: new variants and applications*. Springer, 2014.

[56] L. Nanni, A. Lumini, and S. Brahnam, “Survey on lbp based texture descriptors for image classification,” *Expert Systems with Applications*, vol. 39, no. 3, pp. 3634–3641, 2012.

[57] T. Ojala, M. Pietikäinen, and D. Harwood, “A comparative study of texture measures with classification based on featured distributions,” *Pattern recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[58] T. Ojala, M. Pietikainen, and D. Harwood, “Performance evaluation of texture measures with classification based on kullback discrimination of distributions,” in *Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference on*, vol. 1, pp. 582–585, IEEE, 1994.

[59] L. Nanni, A. Lumini, and S. Brahnam, "Local binary patterns variants as texture descriptors for medical image analysis," *Artificial intelligence in medicine*, vol. 49, no. 2, pp. 117–125, 2010.

[60] W. Yu, L. Gan, S. Yang, Y. Ding, P. Jiang, J. Wang, and S. Li, "An improved lbp algorithm for texture and face classification," *Signal, Image and Video Processing*, vol. 8, no. 1, pp. 155–161, 2014.

[61] S.-R. Zhou, J.-P. Yin, and J.-M. Zhang, "Local binary pattern (lbp) and local phase quantization (lbq) based on gabor filter for face representation," *Neurocomputing*, vol. 116, pp. 260–264, 2013.

[62] G. Anbarjafari, "Face recognition using color local binary pattern from mutually independent color channels," *EURASIP Journal on Image and Video Processing*, vol. 2013, no. 1, p. 6, 2013.

[63] A. Pflug, P. N. Paul, and C. Busch, "A comparative study on texture and surface descriptors for ear biometrics," in *Security Technology (ICCST), 2014 International Carnahan Conference on*, pp. 1–6, IEEE, 2014.

[64] V. Ojansivu and J. Heikkilä, "Blur insensitive texture classification using local phase quantization," in *International conference on image and signal processing*, pp. 236–243, Springer, 2008.

[65] V. Ojansivu and J. Heikkila, "A method for blur and similarity transform invariant object recognition," in *Image Analysis and Processing, 2007. ICIAP 2007. 14th International Conference on*, pp. 583–588, IEEE, 2007.

[66] J. Kannala and E. Rahtu, "Bsif: Binarized statistical image features," in *Pattern Recognition (ICPR), 2012 21st International Conference on*, pp. 1363–1366, IEEE, 2012.

[67] A. Hyvärinen, J. Hurri, and P. O. Hoyer, *Natural Image Statistics: A Probabilistic Approach to Early Computational Vision.*, vol. 39. Springer Science & Business Media, 2009.

[68] R. Raghavendra and C. Busch, "Texture based features for robust palmprint recognition: a comparative study," *EURASIP Journal on Information Security*, vol. 2015, no. 1, p. 5, 2015.

[69] N. Damer and B. Führer, "Ear recognition using multi-scale histogram of oriented gradients," in *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2012 Eighth International Conference on*, pp. 21–24, IEEE, 2012.

[70] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, pp. 886–893, IEEE, 2005.

[71] K. Dewi and T. Yahagi, "Ear photo recognition using scale invariant keypoints.," in *Computational Intelligence*, pp. 253–258, 2006.

[72] R. Szeliski, *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010.

[73] B. Victor, K. Bowyer, and S. Sarkar, "An evaluation of face and ear biometrics," in *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, vol. 1, pp. 429–432, IEEE, 2002.

[74] M. Morchid, R. Dufour, P.-M. Bousquet, G. Linarès, and J.-M. Torres-Moreno, "Feature selection using principal component analysis for massive retweet detection," *Pattern Recognition Letters*, vol. 49, pp. 33–39, 2014.

[75] W. S. Yambor, "Analysis of pca-based and fisher discriminant-based image recognition algorithms," Master's thesis, Colorado State University, 2000.

[76] "Ustb ear image database [online] available:," *http://www1.ustb.edu.cn/resb/en/index.htm*.

[77] R. Raposo, E. Hoyle, A. Peixinho, and H. Proença, "Ubear: A dataset of ear images captured on-the-move in uncontrolled conditions," in *Computational Intelligence in Biometrics and Identity Management (CIBIM), 2011 IEEE Workshop on*, pp. 84–90, IEEE, 2011.

[78] T. Ojala, M. Pietikainen, and T. Maenpaa, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[79] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[80] V. Vapnik, *The nature of statistical learning theory*. Springer science & business media, 2013.

[81] M. Khan, Q. Ding, and W. Perrizo, “K-nearest neighbor classification on spatial data streams using p-trees,” in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 517–528, Springer, 2002.

[82] M. Eskandari and Ö. Toygar, “Fusion of face and iris biometrics using local and global feature extraction methods,” *Signal, image and video processing*, vol. 8, no. 6, pp. 995–1006, 2014.

[83] M. Eskandari, Ö. Toygar, and H. Demirel, “A new approach for face-iris multimodal biometric recognition using score fusion,” *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 27, no. 03, p. 1356004, 2013.

[84] L. Gutierrez, P. Melin, and M. Lopez, “Modular neural network integrator for

human recognition from ear images," in *Neural Networks (IJCNN), The 2010 International Joint Conference on*, pp. 1–5, IEEE, 2010.

[85] L. Yuan and Z. Mu, "Ear recognition based on gabor features and kfda," *The Scientific World Journal*, vol. 2014, 2014.

[86] A. Benzaoui, I. Adjabi, and A. Boukrouche, "Experiments and improvements of ear recognition based on local texture descriptors," *Optical Engineering*, vol. 56, no. 4, pp. 043109–043109, 2017.

[87] M. Eskandari, Ö. Toygar, and H. Demirel, "Feature extractor selection for face–iris multimodal recognition," *Signal, image and video processing*, vol. 8, no. 6, pp. 1189–1198, 2014.

[88] A. Ross and A. K. Jain, "Multimodal biometrics: An overview," in *Signal Processing Conference, 2004 12th European*, pp. 1221–1224, IEEE, 2004.

[89] C. Shan and R. Braspenning, "Recognizing facial expressions automatically from video," *Handbook of ambient intelligence and smart environments*, pp. 479–509, 2010.

[90] H. Nejati, L. Zhang, T. Sim, E. Martinez-Marroquin, and G. Dong, "Wonder ears: Identification of identical twins from ear images," in *Pattern recognition (ICPR), 2012 21st international conference on*, pp. 1201–1204, IEEE, 2012.

[91] M. Farmanbar and Ö. Toygar, “Feature selection for the fusion of face and palmprint biometrics,” *Signal, Image and Video Processing*, vol. 10, no. 5, pp. 951–958, 2016.

[92] S. Soltanpour and Q. J. Wu, “Multimodal 2d–3d face recognition using local descriptors: pyramidal shape map and structural context,” *IET Biometrics*, vol. 6, no. 1, pp. 27–35, 2016.

[93] X. Xiaona, Z. Yue, and P. Xiuqin, “Multimodal recognition fusing ear and profile face based on kpca,” in *Systems and Control in Aerospace and Astronautics, 2008. ISSCAA 2008. 2nd International Symposium on*, pp. 1–5, IEEE, 2008.