COMPARATIVE STUDY AND ANALYSIS OF QUALITY BASED MULTIBIOMETRIC TECHNIQUE USING FUZZY INFERENCE SYSTEM

LYDIA BINTI ABDUL HAMID

UNIVERSITI SAINS MALAYSIA
2013

COMPARATIVE STUDY AND ANALYSIS OF QUALITY BASED MULTIBIOMETRIC TECHNIQUE USING FUZZY INFERENCE SYSTEM

by

LYDIA BINTI ABDUL HAMID

Thesis submitted in fulfillment of the requirements
for the Degree of
Master of Science

July 2013

ACKNOWLEDGEMENTS

At the very most, I am humbly thankful and praise to Allah S.W.T for allowing me to complete this thesis. In the first place I would like to record my gratitude to Dr Dzati Athiar Ramli for her supervision, advice, and guidance from the very early stage of this research as well as giving me extraordinary experiences throughout the work. Above all and the most needed, she provided me unflinching encouragement and support in various ways.

I gratefully thank the Malaysian Ministry of Higher Education and Malaysia Ministry of Science, Technology and Innovations as this research which is headed by Dr Dzati Athiar Ramli was supported under Research University (RU) Grant, Universiti Sains Malaysia, 1001/PELECT/814098 and Incentive Grant, Universiti Sains Malaysia.

I would like to thank my family members, especially my father and mother Abdul Hamid Mohd Amin and Haslina Abdullah for supporting and encouraging me to pursue this study. Without their encouragement, I would not have finished successfully. My grateful thanks also go to my husband Mohd Nor Amin Sukeri for his cooperation during the period of my assignment.

Many thanks to my friend, Farhana Alias for her valuable advice in science discussion, supervision and furthermore, using her precious times to read this thesis and gave their critical comments about it.

Finally, I would like to thank everybody who was important to the successful realization of thesis, as well as expressing my apology that I could not mention personally one by one.

LIST OF TABLES

LIST OF SYMBOLS

w	Direction in space
b	Position in space
w_0	Optimal hyperplane
ρ	Margin
$\emptyset(w)$	Langrange functional
α_i , β_i	Langrange multipliers
ξ_i	Measurement of the misclassification errors
$K\langle x_i , x_j \rangle$	Kernel function
γ , r , d	Kernel parameters
$x_a(t)$	Continuous signal
n	Integer values
T	Sampling period
$x(n)$	Original signal
$z(n)$	Noise drawn from a zero means normal distribution with variance η
$w(n)$	Impulse response of window
N	Length of each frame
$\varepsilon(n)$	Prediction error
$\hat{c}_n$	Complex cepstrums
μ	Mean vector
C	Covariance matrix
H_0 , S_0	Threshold values
$\hat{s}_{audio}$	Audio normalized score

$\hat{s}_{visual}$	Visual normalized score
$\hat{s}_{avi}$	Score distribution
W	Weight
W_{opt}	Weight optimum
ID_{ai}	Current audio
ID_{vi}	Current visual
f_{ai}	Feature audio
f_{vi}	Feature visual
m_{ai}	Model audio
m_{vi}	Model visual

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
ASM	Active Shape Model
ASR	Automatic Speech Recognition
ATM	Auto Teller Machine
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DP	Dynamic Programming
DTW	Dynamic Time Warping
ECP	Equal Correlation Peak
EER	Equal Error Rate
FAR	False Acceptance Rate
FFT	Fast Fourier Transform
FRR	False Rejection Rate
GAR	Genuine Acceptance Rate
GMM	Gaussian Mixture Model
HMM	Hidden Markov Model
LDA	Linear Discriminant Analysis
LPC	Linear Predictive Coding
LSP	Line Spectral Pair
MACE	Minimum Average Correlation Energy
MFCC	Mel Frequency Cepstral Coefficients

PARCOR	Partial Correlation
PCA	Principle Component Analysis
PLP	Perceptual Linear Predictive
PSR	Peak-to-Sidelobe Ratio
RASTA	Relative Spectral Transform
ROC	Receiver Operation Characteristics
ROI	Region of Interest
SDF	Synthetic Discriminant Function
SNR	Signal to Noise Ratio
SVM	Support Vector Machine
UMACE	Unconstrained Minimum Average Correlation Energy
VIV	Verbal Information Verification
VLSI	Very Large Scale Integration
VQ	Vector Quantization

KAJIAN PERBANDINGAN DAN ANALISIS SISTEM MULTIBIOMETRIK BERASASKAN KUALITI DENGAN MENGGUNAKAN KAEDAH GABUNGAN INFEREN KABUR

ABSTRAK

Biometrik adalah sains dan teknologi yang mengukur dan menganalisa data biologi yakni ciri-ciri kelakuan dan fizikal yang mampu membezakan seseorang individu daripada individu yang lain. Kajian terhadap sistem pengesahan biometrik dengan gabungan beberapa sumber biometrik mengesahkan bahawa prestasi sistem adalah lebih baik berbanding prestasi sistem biometrik tunggal. Walau bagaimanapun, pendekatan gabungan yang tidak mengambilkira maklumat kualiti biometrik boleh mempengaruhi prestasi sistem di mana dalam kes tertentu prestasi sistem gabungan berkemungkinan menjadi lebih rendah berbanding prestasi sistem biometrik tunggal. Dalam usaha untuk mengatasi masalah ini, kajian ini mencadangkan kaedah gabungan berasaskan kualiti dengan merekabentuk Sistem Inferen Kabur (FIS) yang mampu menentukan pemberat yang optimum untuk menggabungkan parameter bagi sistem biometrik gabungan dalam persekitaran yang berubah. Bagi tujuan ini, sistem gabungan yang menggabungkan dua modaliti biometrik iaitu pertuturan dan imej bibir telah dilaksanakan. Untuk isyarat pertuturan, Pekali Mel Frekuensi Cepstral (MFCC) telah digunakan sebagai fitur manakala kawasan dikehendaki (ROI) digunakan untuk imej bibir. Mesin Penyokong Vektor (SVM) digunakan sebagai pengelas kepada sistem pengesahan. Untuk pengesahan, kaedah gabungan iaitu peraturan maksimum, peraturan jumlah mudah, peraturan jumlah berpemberat dibandingkan dengan kaedah gabungan berasaskan kualiti yang dicadangkan. Daripada keputusan eksperimen, pada 35dB SNR dan kualiti kepadatan 0.8, peratusan EER sistem pertuturan, bibir, peraturan minimum, peraturan maksimum, peraturan jumlah mudah dan peraturan jumlah berpemberat yang masing-masing mencapai 5.9210%, 37.2157%, 33.2676%, 31.1364%, 4.0112% dan 14.9023% berbanding prestasi sistem inferen kabur sugeno dan mamdani yang masing-masing mencapai 1.9974% dan 1.9745%.

COMPARATIVE STUDY AND ANALYSIS OF QUALITY BASED MULTIBIOMETRIC TECHNIQUE USING FUZZY INFERENCE SYSTEM

ABSTRACT

Biometric is a science and technology of measuring and analyzing biological data i.e. physical or behavioral traits which is able to uniquely recognize a person from others. Prior studies of biometric verification systems with fusion of several biometric sources have been proved to be outstanding over single biometric system. However, fusion approach without considering the quality information of the data used will affect the system performance where in some cases the performances of the fusion system may become worse compared to the performances of either one of the single systems. In order to overcome this limitation, this study proposes a quality based fusion scheme by designing a fuzzy inference system (FIS) which is able to determine the optimum weight to combine the parameter for fusion systems in changing conditions. For this purpose, fusion systems which combine two modalities i.e. speech and lip traits are experimented. For speech signal, Mel Frequency Cepstral Coefficient (MFCC) is used as features while region of interest (ROI) of lip image is employed as lip features. Support vector machine (SVM) is then executed as classifier to the verification system. For validation, common fusion schemes i.e. minimum rule, maximum rule, simple sum rule, weighted sum rule are compared to the proposed quality based fusion scheme. From the experimental results at 35dB SNR of speech and 0.8 quality density of lip, the EER percentages for speech, lip, minimum rule, maximum rule, simple sum rule, weighted sum rule systems are observed as 5.9210%, 37.2157%, 33.2676%, 31.1364%, 4.0112% and 14.9023%, respectively compared to the performances of sugeno-type FIS and mamdani-type FIS i.e. 1.9974% and 1.9745%.

CHAPTER 1

INTRODUCTION

1.1 Introduction

Section 1.2 discusses the problem statement involved in this study. Objectives of this study are given in section 1.3. Section 1.4 explains the scope of research. Section 1.5 discusses the performance evaluation. Finally, the thesis outline is presented in section 1.6.

1.2 Overview of Biometrics

Previously, the traditional verification uses passwords, keys or smart cards which are less secure since few problems may occur due to forgotten password, duplicated keys or stolen smart cards. Nowadays, biometric data for verification systems are commercially used in data security, internet access, ATMs, network logins, credit cards and government records. More studies on biometric system have been done by researchers due to the increase of requirement of automatic information processing in many industrial fields (Chia et al. 2011). Biometrics is defined as the development of statistical and mathematical methods applicable to data analysis problems in the biological sciences. Biometrics is also a technology, which uses various individual attributes of a person to verify his or her identity. Biometric characteristics can be divided into two main classes i.e. physiological and behavioral characteristics. Physiological characteristics refers to the human body such as face, fingerprints, palm print, iris, DNA, hand geometry and finger vein structure while behavioural characteristics are related to the actions of a person such as voice, keystroke dynamics, gait, typing rhythm and signature (Jain et al. 2004). This study implements biometric system for speaker verification systems. Speaker verification system is used to verify a person's claim from the enrolment database by using speech signal as the input data.

Single biometric systems have to face few limitations such as non-universality, noisy sensor data, large intra-user variations and susceptibility to spoof attacks. For example, a single biometric system uses voice patterns to identify the individuals may fail to operate because of a noisy data signal captured by the system. Limitations faced by single biometric system can be overcome by applying the multibiometric system. Multibiometric system enhanced the matching accuracy of a biometric system in noisy condition as well as increases the population coverage with multiple traits (i.e. lip, iris, voice and face). Studies on multibiometrics are further discussed in Ben-Yacoub et al. (1999) and Pan et al. (2000). Besides that, multibiometric system may continuously operate even though a certain trait is unreliable due to user manipulation, sensor or software malfunctions.

1.3 Problem Statement

Single biometric systems that use voice features to verify individuals may fail to operate due to noisy voice signal captured by the systems. However, this limitation can be overcome by applying multibiometric systems which are capable to enhance the individual matching accuracy of the single biometric systems in noisy condition as reported in Ross and Jain (2004).

These systems which consist of more than one modalities fused together can continuously maintain their function even though one of the modalities is faulty. This is because these systems can rely on the perfect modality for the correct verification result.

However, this is only true when fusion scheme is done at the decision level where hard decision fusion for example OR operator is executed. For the score level decision fusion, the multibiometric systems are at optimum performance when all traits operate in clean condition. In noisy condition, the systems are likely to give false fusion scores because the authentic and imposter scores are no more reliable.

This study proposes the use of quality based score fusion approach to improve the performances of multibiometric system. This approach measures the degree of quality of the biometric sample hence incorporating the measurement to the fusion algorithm. This method is very useful to ensure the speaker verification systems are at optimum performance especially in noisy condition. For this purpose, a Fuzzy Inferred System (FIS) is designed and the weight inferred from FIS is used as fusion weight in the multibiometric systems.

1.4 Objectives

The objectives of this study are:

1. to develop single and multimodal systems based on speech and lip traits.
2. to integrate the proposed quality based fusion algorithm to the multibiometric systems.
3. to evaluate the performances of proposed fusion systems hence to compare their effectiveness with the baseline methods.

1.5 Scope of Research

The database used in this study is the Audio-Visual Digit Database (Sanderson and Paliwal 2001). The database consists of audio and visual of people reciting zero. The video of each person is stored as JPEG images with 512 x 384 pixels while corresponding audio provided 16bit, 32 kHz, WAV format. For the noisy signals, this study uses the generic signals by corrupting the clean audio with simulated additive white Gaussian noise. Whereas, salt and pepper noise are imposed to the visual data.

For audio feature extraction module, this study implements two types of parameter analysis i.e. Mel Frequency Cepstrum Coefficient (MFCC) and Linear Prediction Coding (LPC). For visual feature extraction module, Region of Interest (ROI) is used as features. The Support Vector Machine (SVM) classifier is used in pattern matching module while min-max normalization is used for score normalization. Particularly, for audio biometric system, two types of systems i.e. MFCC-SVM and LPC-SVM systems are developed while for visual biometric system, one type of system i.e. ROI-SVM system is developed.

Subsequently, MFCC-ROI-SVM and LPC-ROI-SVM systems are developed for multibiometric systems. Four fusion schemes i.e. minimum rule, maximum rule, simple sum rule and weighted sum rule are used as baseline techniques. This study consists of software implementation only and no hardware integration is implemented during the entire process.

1.6 Performance Evaluation

According to Kung et al. (2004), the biometric system performances are evaluated using False Rejection Rate (FRR) and False Acceptance Rate (FAR). FRR is the percentage of authorized persons rejected by the system. In term of sensitivity or Genuine Acceptance Rate (GAR), it can be explained as the percentage of authorized individuals is admitted by the system. FRR and GAR are computed as in equation (1.1) and (1.2), respectively where t_o is the threshold value. The FRR is achieved when the number of genuine is less then t_o. For GAR, this condition is achieved when the number of genuine is equal or greater then t_o.

$$\mathrm{FRR} = \left[\frac{\text{number of genuine} < t_o}{\text{number of genuine}}\right] 100\% \tag{1.1}$$

and

$$\mathrm{GAR} = 1 - \mathrm{FRR} = \left[\frac{\text{number of genuine} \geq t_o}{\text{number of genuine}}\right] 100\% \tag{1.2}$$

Besides that, FAR is the percentage of unauthorized individuals which accepted by the system. The FAR is given as in equation (1.3) where t_o is the threshold value. This condition is achieved when the number of imposter is greater then t_o.

$$\mathrm{FAR} = \left[\frac{\text{number of imposter} > t_o}{\text{number of imposter}}\right] 100\% \tag{1.3}$$

The Receiver Operating Characteristic (ROC) plot is a visual characterization of the trade-off between FAR versus FRR or GAR versus FAR. This study implements the FAR against GAR graph to describe the system performances as shown in Figure 1.2. The performances of the biometric systems based on FAR and GAR can be verified based on several threshold values which need to be adjusted according to the desired security standards.

Consequently, Equal Error Rate (EER) is another method used to measure the system performances where the error rates for both accept and reject are equal. The value of the EER can be easily obtained from the ROC curve. The EER is computed in order to identify the accuracy of the systems. In general, the system with the lowest EER is the most accurate system.

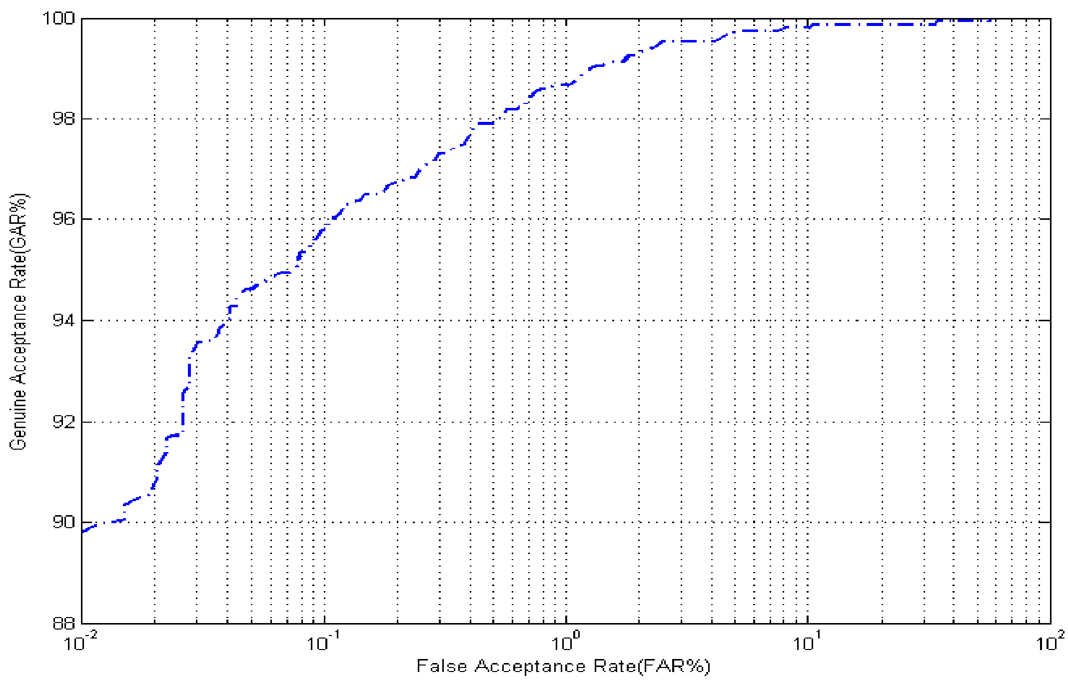

Figure 1.2: Receiver Operation Characteristics (ROC) GAR versus FAR

1.7 Thesis Outline

This thesis consists of 5 chapters. Chapter 1 discusses the concept and definition, problem statement, objectives and scope of research. Chapter 2 covers the previous and current researches on multibiometric systems, speech signal and image biometric system. It also explains the theory of SVM and fuzzy logic which involved in this study.

Methodology is explained in Chapter 3. Further details on steps involved throughout this study are discussed in this chapter. Results and discussion obtained from this study are stated in Chapter 4. Finally, conclusion and suggestion are presented in Chapter 5.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Section 2.2 discusses the concept and definition involved in this study. Recent studies on multibiometric systems are given in section 2.3. Section 2.4 explains the researches on single biometric systems using speech signal and lip image traits. Section 2.5 discusses the theory of support vector machine. Reviews on Fuzzy Inference System (FIS) are explained in Section 2.6. Finally, the summary is presented in section 2.7.

2.2 Concept and Definition

Originally, the word "biometric" is derived from the Greek words 'bios' and 'metric' which means life and measurement (Bohm and Testor, 2007). Biometrics in general can be defined as a technology that employs person's physiological and behavioural traits for verification and identification purposes (Kung et al., 2004). Previously, the traditional way to verify and identify people uses passwords, keys or smart cards. These approaches are less secure compared to the biometric approach since some problems may occur due to passwords can be forgotten and, duplicated keys or smart cards can be misplaced or lost.

Each person has different personal characteristics which can be distinguished from the others. According to Virk and Maini (2012), biometric characteristics can be divided into i.e. physiological and behavioural types. Physiological characteristics refers to the human body such as face, fingerprint, palm print, iris, DNA, hand geometric and finger vein structure while behavioural characteristics is related to the behaviour of a person such as keystroke dynamic, gait, typing rhythm and signature. Consequently, voice can be considered as behavioural or physiological characteristics since both physical information such as nasality and pitch; and behavioural information i.e. conversational style and dialect are contained in the signal (Jain et al., 1999a).

As discussed above, biometric systems operate as two different approaches i.e. verification and identification. According to Reynolds (2002), verification system is a task of determining whether a person is who he/she claims to be while the identification system is used to determine who is talking from a set of known voices or speakers.

According to Campbell (1997), the architecture of biometric speaker verification systems consists of four different components i.e. data acquisition, feature extraction, pattern matching and decision as shown in Figure 2.1. Speaker verification systems can be grouped into text-dependent and text-independent applications. For text-dependent applications, the systems have prior knowledge of the text to be spoken while in text-independent application, the systems have no prior knowledge of the text to be spoken (Rydin, 2001).

The data acquisition consists of two sections which are enrolment data and current data. Enrolment data is a group of data from people who have authorization to use the system while current data is data to be verified by the system which are obtained from both authorized and non-authorized users. During the data acquisition process, the speech signal which is an analogue signal will be converted to digital signal using the microphone.

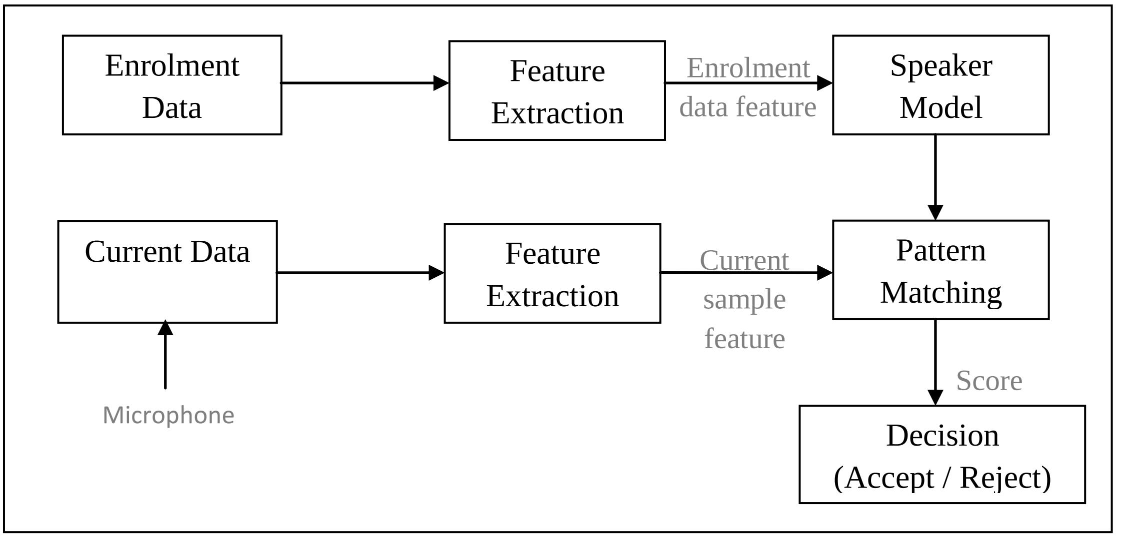

Figure 2.1: Architecture of biometric speaker verification systems

Consequently, feature extraction consists of few sub-processes. Firstly, the non-speech portion is removed from the speech signal. Next, only the informative part will be extracted from the speech signal as features to the system. The speakers are differentiated according to their vocal tract and glottal source. Commonly, the spectral based features using Linear Prediction Coding (LPC) and Fast Fourier Transform (FFT) are executed (Reynolds, 2002).

For the pattern matching process, the score between the enrolment and current speech features are measured. Here, the extracted features from the enrolment speech data are first used to construct speech model. The current speech data is then compared with the model during the verification process. Typical pattern matching methods for speaker biometric system are Artificial Neural Network (ANN), Dynamic Time Warping (DTW) and Hidden Markov Model (HMM) (Anusuya and Katti, 2009). Finally, for the decision process, the current sample's score is compared to a threshold that earlier specified by the systems and a decision is made whether to accept or reject the current speaker.

2.3 Literature Reviews on Multibiometric Systems

Initially, researches on biometric systems have been focused on single modal verification which only consider single trait as biometric data. The systems only use single modality to find the genuine person from the given database (Jain et al., 1999a). However, single biometric systems tend to obtain low performances when the data is corrupted due to noisy condition. Other than that, physical appearance and behavioral characteristics of a person tend to vary with time which also can affect biometric system performances (Kittler et al., 1997b). One of the solutions to overcome this problem is by implementing multibiometric systems. Multibiometric systems combine multiple traits (i.e. speech, iris and fingerprint) in order to improve the systems recognition accuracy when one of the traits is corrupted (Rowe et al, 2007). Multibiometric refers as an extension of a single biometrics in which information from multiple sources such as sensors, units, samples, algorithms and traits are combined. Further reports on multibiometric systems have been reviewed in (Ross et al., 2004) and (Ross et al., 2007).

2.3.1 Advantages and Disadvantages of Multibiometric Systems

According to Rowe et al. (2007), multibiometric systems are capable to solve the non-universality problem faced by single biometric systems. For example, if a mute person who is unable to provide information required by the speaker verification systems, the person can aid the problem by using other biometric traits such as fingerprint, iris or face. Besides, multibiometric systems can avoid problems caused by noisy data where the information obtained is not sufficient for decision-making. The systems can implement the data from other traits which provide sufficient information in order to enable the decision-making. Another advantage of multibiometric systems is where the spoof attacks can be avoided since the systems require more than one trait which is harder for the imposter to mimic the enrollment speakers.

However, one of the disadvantages faced by multibiometric systems is that it requires more sensors which contribute to higher implementation cost compared to single biometric systems. In addition, multibiometric systems require user to interact with more than one sensor. For example, multibiometric system using microphone and fingerprint scanner may increase the user inconvenience since a user needs to provide information for microphone as well as to touch the fingerprint scanner. Hence, more computation, memory and storage are required for this purpose. Furthermore, the operating times during enrolment and verification process are also increased (Rowe et al., 2007).

2.3.2 Taxanomy of Multibiometric Systems

Multibiometric systems can be classified into five systems i.e. multimodal, multi-sample, multi-algorithm, multi-sensor and multi-instance systems as shown in Figure 2.1.

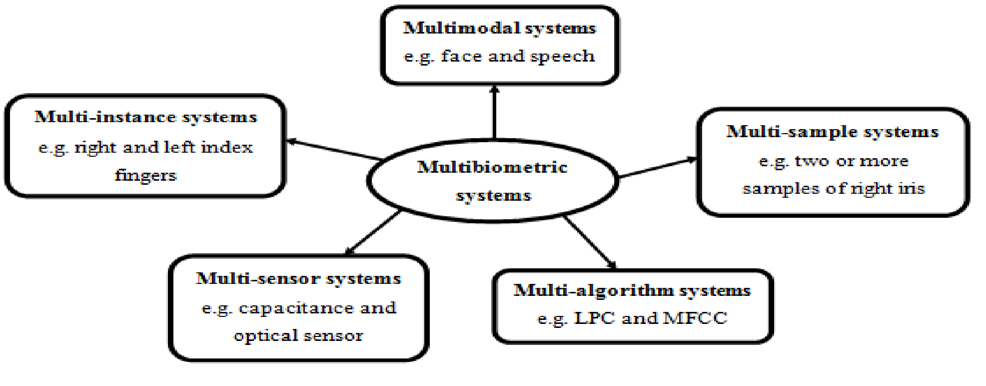

Figure 2.2: Scenarios in multibiometric systems

a. Multimodal systems: multimodal systems extract biometric information from multiple modalities such as speech, lip and face for verification of individuals. The multimodal systems are highly reliable especially if one of the traits has insufficient information**.** However, cost of developing these systems are higher due to more sensors needed to extract the traits information. Study on multimodal systems using face and speech modalities has been reported by Brunelli and Falagivna (1995). According to Kittler et al., (1998), the multimodal systems performances have been improved by combining three biometrics i.e. frontal face, face profile and voice using sum rule combination scheme. In another research, a multimodal system implementing three different traits i.e. fingerprint, face and finger vein has been discussed in Hong et al. (1999).

b. Multi-sample systems: Multi-sample systems use multiple samples extracted from a single trait which obtained from a single sensor**.** The scores are extracted from each sample by applying the same algorithm in order to obtain an overall recognition results. The benefit of using multiple samples is to avoid poor performance due to the bad properties of sample if only one sample is used during the process. Research by Samad et al. (2007) proposed multi-sample fusion schemes in order to increase the system performances. This fusion scheme computes the score from each sample using maximum, minimum, median, average and majority vote operator. According to Suutala and Roning (2005), the combinations of multi-samples have improved the performances of footstep profile-based person identification. This study employed multiple footsteps from each person and the scores were fused using simple product and sum rules fusion schemes.

c. Multi-algorithm systems: multi-algorithm systems are a combination of output obtained from multiple methods such as classification algorithms or/and feature extraction for the same biometric data (Ross and Jain, 2007). The outputs are combined to obtain an overall recognition result**.** The advantage of multi-algorithm is that no multiple sensors are required which is very cost effective. However, the system computation will be complicated due to many feature extractions and matching modules are verified during the process. Studies on multi-filter bank approach for speaker verification based on genetic algorithm have been discussed by Charbuillet et al. (2007b). This study proposes a feature extraction system based on the combination of three feature extractors (i.e. MFCC, LPCC and LFCC) adapted to the speaker verification task. Results proved that the proposed method improves the system performances. In addition, subspace algorithms such as PCA, Fisher Linear Discriminant (FLD) and ICA have been applied for palm print and face separately in order to determine the best algorithm performance. This study can be found in Imran et al. (2010).

d. Multi-sensor systems: Multi-sensor systems implement multiple sensors to capture single biometric trait of an individual. Marcialis and Roli (2004) reported that the multi-sensor systems can perform better than traditional fingerprint matchers using a single sensor. According to Lee et al. (2004), images of a subject are captured using multiple 2D cameras. Next, extraction of a person’s face using an infrared sensor and visible-light sensor has been illustrated in Kong et al. (2005). Subsequently, multi spectral cameras were used to extract the images of iris, face and finger have been explained in Rowe and Nixon (2006).

e. Multi-instance systems: For multi-instance systems, the biometric information is extracted from the multiple instances of a single biometric trait. As an example, Prabhakar and Jain (2000) proposed the used of the left and right index finger and iris of an individual. A study on multi-instance speech signal data fusion by evaluating the multi-instance of speech signal (i.e. zero, seven and eight) has been discussed in Ramli et al. (2010). According to Ramli et al. (2011), combination of three speech modality subsystems from different verbal zero, seven and eight multi-instance were proposed to overcome the limitations faced by single modal system.

2.3.3 Levels of Fusion of Multibiometric Systems

The levels of fusion of multibiometric systems can be classified into fusion before matching and fusion after matching as illustrated in Figure 2.2. Fusion before matching is known as pre-classification while fusion after matching is known as post-classification. For fusion before matching, the fusion process is computed at sensor and feature levels while fusion after matching, the fusion process is done at match score, rank and decision levels (Ross and Jain 2007).

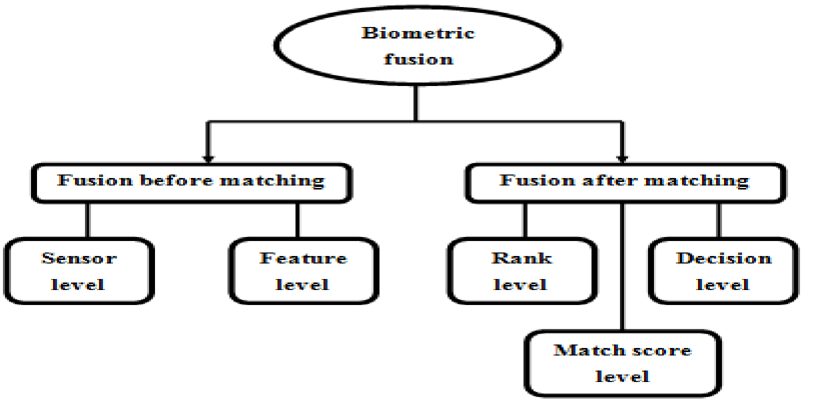

Figure 2.3: Level of fusions

a. Fusion before matching

i. Sensor level fusion: Sensor level fusion is a combination of raw data extracted from the sensor as displayed in Figure 2.3. According to Iyengar et al. (1995), the system performances may be affected due to contaminate raw data which results from many factors such as background clutter and non-uniform illumination. As discussed by Singh et al. (2005), two types of conditions i.e. data from a single trait obtained from multiple sensors and data from multiple snapshots of a single biometric trait extracted from a single sensor can be performed during sensor level fusion.

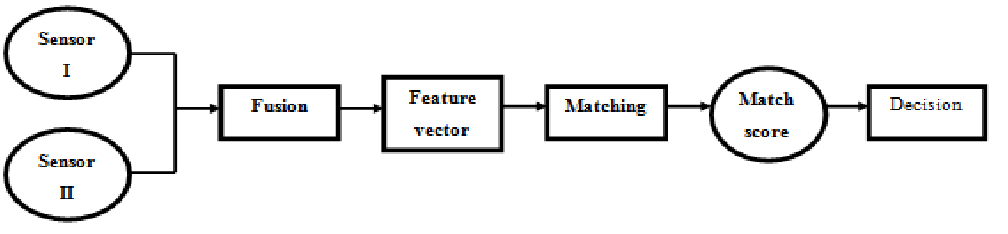

Figure 2.4: Sensor level fusion process flows

ii. Feature level fusion: Feature level fusion is a combination of different feature vectors extracted from multiple biometric sources into a single feature vector as illustrated in Figure 2.4. Feature normalization and feature selection have been executed during the process. According to (Jain et al., 2005), the feature normalization is implemented to modify the location and scale of feature values via a transformation function where can be done by using appropriate normalization schemes. As an example, the min-max technique and median scheming have been used for hand and face biometric traits. Next, feature selection is executed in order to reduce the dimensionality of a feature vector. This process has the ability to improve the matching performance. As stated in Kumar and Zhang (2005), feature selection algorithms such as Sequential Backward Selection (SBS), Sequential Forward Selection (SFS) and Partition About Medoids have been studied.

However, the feature level fusion is hard to perform due to the joint feature set extracted from different biometric sources may not be linear and incompatible (Ross and Jain, 2004). Therefore, more researchers focused on other types of fusion schemes such as score level fusion and decision level fusion.

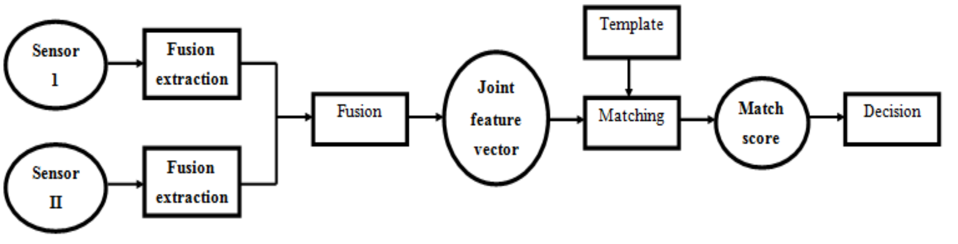

Figure 2.5: Feature level fusion process flows

b. Fusion after matching

i. Rank level fusion: Rank level fusion is a combination of identification ranks obtained from multiple unimodal biometrics as shown in Figure 2.5. A novel approach to improve biometric recognition using rank level fusion has been reported in Bhatnagar et al (2007). This paper used the rank level fusion which improved the system performances. Rank level fusion using fingerprint and iris biometric has been discussed in Radha and Kavitha (2011). The experimental results have revealed better performances of the proposed rank level fusion in multimodal biometrics system.

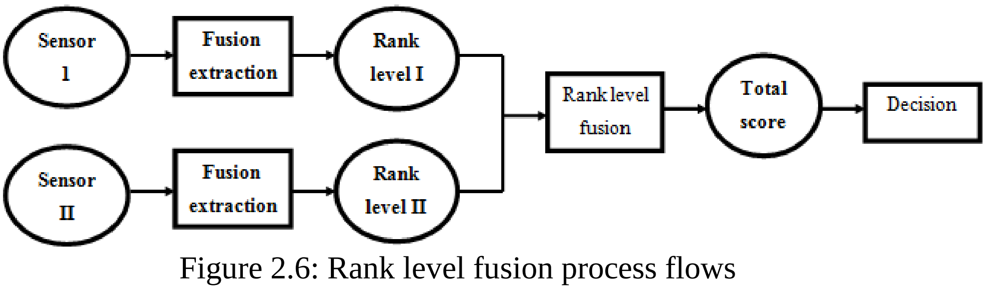

Figure 2.6: Rank level fusion process flows

ii. Score level fusion: Score level fusion is a combination of match outputs from multiple biometrics to improve the matching performances in order to verify or identify individual (Jain and Ross, 2004). This approach is illustrated as in Figure 2.6. The fusion of this level is widely applied in multibiometric systems due to its simplicity. Moreover, the matching scores consist of sufficient information which enables the systems to distinguish the authentic users from the imposter users (He el al., 2010). However, the combination process may be defected due to degraded biometric performance. Score level fusion can be grouped into three schemes i.e. density-based schemes, transformation-based scheme, and classifier-based scheme in order to overcome this limitation (Ross and Jain, 2007). In density-based scheme, a training set of the genuine and imposter match scores is estimated using a joint density function. Next, the posterior probability of observing genuine (or imposter) class are defined from the Bayes formula. However, the density-based fusion scheme requires large training samples in order to develop accurate system performances. Therefore, it is not suitable in most of the multimodal systems because of the time consuming and high cost factors. The transformation-based scheme is applied during the score normalization process which requires simple normalization technique, i.e. min-max normalization, z-score and tanh-estimator in order to transform the score into the same domain. For transformation-based fusion, the match score is directly combined using simple fusion operators such as sum rule, product rule, min rule and max rule techniques (Parviz and Moin, 2011). For the classifier-based scheme, the matched scores extracted from each biometric source are used as inputs to a trained pattern classifier such as SVM, HMM and ANN. Next, the input score is classified by the classifier in order to identify the genuine and imposter class (Nandakumar et al., 2007).

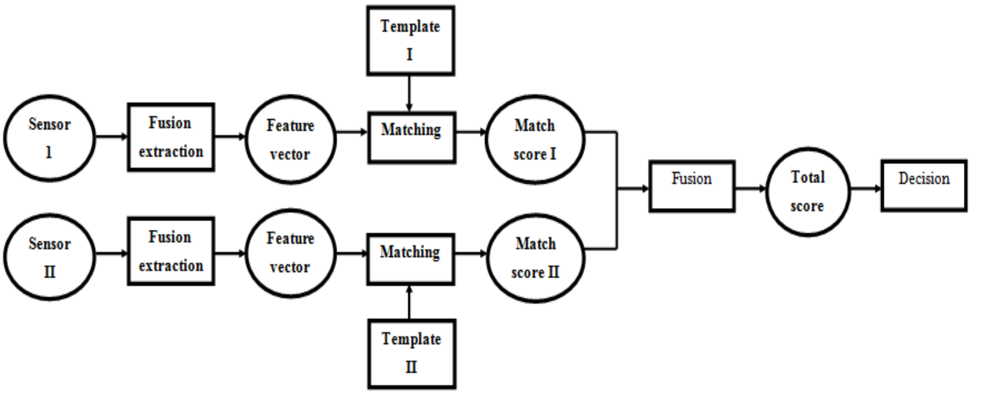

Figure 2.7: Match score level fusion process flows

iii. Decision level fusion: Fusion at the decision level is executed after a match decision has been made by the individual biometric source as shown in Figure 2.7. Few methods such as “AND” and “OR” rules, majority voting, weighted majority voting and Bayesian decision fusion have been implemented in order to combined the distinct decisions to a final decision (Ross and Jain, 2007). Messer and Kittler (2000) discussed the data and decision level fusion of temporal information for automatic target recognition (ATR). An adaptive ATR system which decides how to best distinguish the target from a particular background has been proposed in this study. In decision level fusion, Kahler and Blasch (2011) discussed the used of decision levels fusion in High Range Resolution (HRR) radar and Synthetic

Aperture Radar (SAR). Results proved that the decision level fusion is able to enhance identification performance.

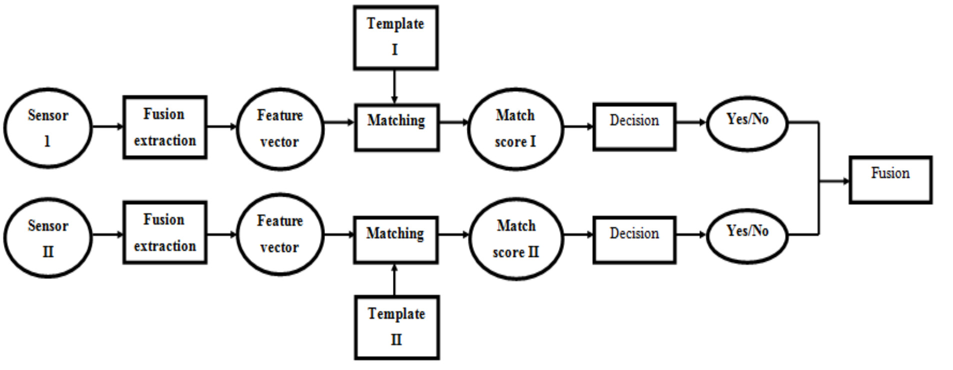

Figure 2.8: Decision level fusion process flows

2.3.4 Quality Based Fusion Systems

Quality measures are measurements to identify the degree of excellence of biometric samples to some predefined criteria which may influence the system performance. This study implements the quality measures into the biometric systems. Figure 2.8 described the general theory of incorporating quality measures in biometric systems.

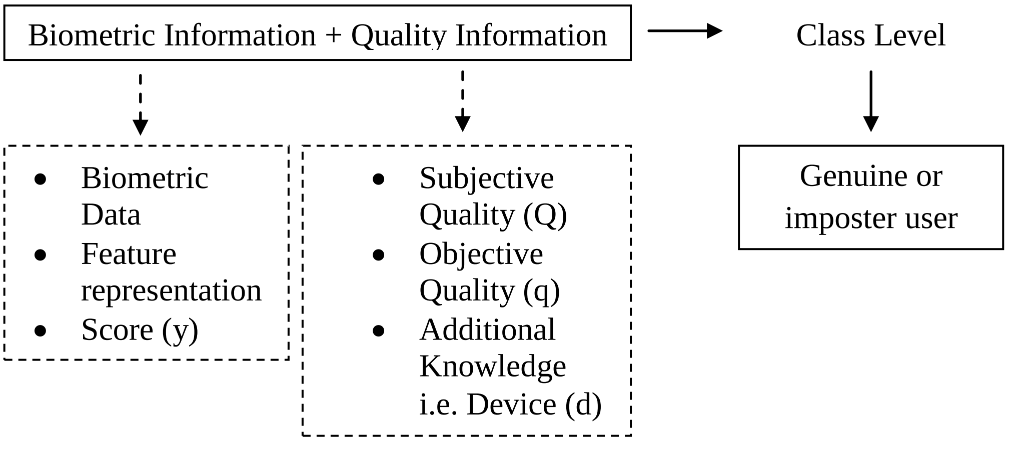

Figure 2.9: General theory of incorporating quality measures

The quality information can be grouped into two classes, i.e. subjective quality (Q) and objective quality (q), respectively. The subjective quality (Q) is derived from human judgement, manual (not computable) and observable only during training while the objective quality (q) is derived from biometric sample, automatic and computable as well as observable during both training and testing. Examples for subjective and objective quality for face and iris are shown in Figure 2.10 and 2.11, respectively.

Good quality or bad quality?

Q

Uniform background
High spatial resolution
High colour resolution

Cluttered background
Low spatial resolution
Low colour resolution

q So, background uniformity, spatial resolution and colour resolution are potential quality measures for face

Figure 2.10: Example 1 (face) for subjective and objective quality (Poh et al., 2010)

Good quality or bad quality?

Q

(i) (ii) (iii) (iv)

i. Good quality, ii. Small iris area, iii. Displaced contact lens and iv. Blurred

q So, iris area, contact lens and image focus are potential quality measures for iris

Figure 2.11: Example 2 (iris) for subjective and objective quality (Poh et al., 2010)

Research on discriminative multimodal biometric authentication based on quality measures has been illustrated in Fierrez et al. (2005). Chen et al. (2005) implemented the fingerprint quality indices for predicting authentication performance in their studies. Subsequently, incorporating image quality multi-algorithm fingerprint verification has been discussed in Fierrez et al. (2006).

2.4 Literature Reviews on Single Biometric Systems

2.4.1 Speech Signal Trait

The technology of Automatic Speech Recognition (ASR) has progressed greatly over the past few years. During 1939, studies on automatic speech recognition and transcription have started by the ATT&T's Bell Labs. In 1960, Gunnar Fant created a model which can describe the physiological components of speech signal using the x-rays analysis. Each speech signal produces different phonic sounds which can be used to identify the speakers. In 1970, the Fand model has been expanded by Dr. Joseph Perkell by adding the tongue and jaw to the model (John et al, 2003). As mentioned by Haberman and Fejfar (1976), the National Institute of Standard and Technology (NIST) developed the NIST Speech Group in the mid 1980s to study the uses of speech processing techniques. Studies on ASR have been further discussed in Campbell (1997) and Reynolds (2002). Features and techniques for speaker recognition can be found in Singh (2003).

Research on speaker recognition using Mel Frequency Cepstrum Coefficient (MFCC) features has been implemented in Hazen et al. (2004) and Hazen (2006). Linear Prediction Coding (LPC) features has been discussed in Atal (1974) and Furui (1981a). Furui (1981b) has used the fixed Japanese utterences features for the speaker recognition system. Other features that have been suggested are Perceptual Linear Predictive (PLP) coefficient which has been explained in Xu (1989) and Line Spectral Pair (LSP) frequencies which has been studied in Liu (1990). The LPC feature is modified to PLP coefficient which is based on human perception and physiological effect sound while LSP coefficient which is based on formant bandwidths and locations. Partial Correlation (PARCOR) is also modified from the LPC feature which has been discussed in Atilli (1988). Researches by Dupont and

Luettin (2000) and Heckmann et al. (2002b) have implemented noise-robust RASTA-PLP features for the speaker recognition system. Features extracted using Linear Discriminant Analysis (LDA) technique is a combination of LPC coefficient and dynamic LPC coefficient in order to reduce the dimension of speech vector space has been reported in Hai and Joo (2003). The UBM_GMM cepstral features, prosodic statistics and pronounciation modelling has been explained in Campbell (2003). Study by Jiang et al. (2003) has proposed the wavelet packet strategy during the feature extraction. In another research, filter bank features has been discussed in Ravindran et al. (2003).

For the pattern matching process, few techniques such as Euclidean, Manhattan Distance, Dynamic Time Warping (DTW), Vector Quantization (VQ), Hidden Markov Model (HMM), Gaussion Mixture Model (GMM), Artificial Neural Network (ANN) and Support Vector Machine (SVM) have been proposed (Qiao and Yasuhara, 2006; Ouzounov, 2010). For text independent recognition, speaker specific VQ codebooks or the more advanced GMM have been used regularly. Furthermore, Euclidean and Manhattan Distance are simpler techniques compared to more advanced technique such as DTW, VQ and HMM. DTW is an algorithm for measuring similarity between two sequences which may vary in time or speed. DTW approach may be a better choice for a real-world speaker recognition system if the amount of available training data is not sufficient. DTW has been used in speech signal processing (Rabiner and Juang, 1993), manufacturing (Gollmer and Posten, 1995), gesture recognition (Gavrila and Davis, 1995), medicine (Caiani et. al, 1998) and robotics (Schmill et. al, 1999).

In another research, the vector quantization (VQ) is another technique for pattern matching process. The training data has been used to form a speaker's codebook. During the recognition stage, the test data is compared to the codebook of each reference speaker and a measure of the difference has been used to make the recognition decision. This model used a vector quantized codebook, which is generated for a speaker by using the training data. The VQ technique has been discussed in Linde et al. (1980) and Soong et al. (1985), respectively.

A model that is widely used for modeling of sequences is the Hidden Markov Model (HMM). It provides more flexibility and produce better matching score. In speech recognition, HMMs have been used for modeling observed patterns from 1970s. However, the system performances are quite slow compared to other methods. Many researchers i.e. Fu (1980) and Russell and Moore (1985) have published a large number of papers, which present HMM as tool for use on these practical problems. The HMM algorithm and its implementation has been described by Rabiner and Juang (1986).

Next, GMM has also been used for pattern matching. This method has been the most successful because of many factors such as the high-accuracy recognition, the probabilistic framework, and the remarkable capability to model irregular data. This characteristic makes it very suitable to have a smooth estimation of speaker's acoustic variability Studies on GMM based speaker recognition on readily available databases has been discussed in Wildermoth and Paliwal (2003). Audio signal processing using GMM has been implemented by Kumar et al. (2010).

Speech recognition by self-organizing feature finder using ANN has been stated in Lerner and Deller (1991). ANN based on multi-layer perceptions or radial basic function has been trained to discriminate between enrolled speaker and non-enrolled speakers. Consequently, phonetic speaker recognition using SVM has been explained in Campbell (2003). Study by Solomonof et al (2005) used SVM in order to derive fully non-linear channel compensations. Another method i.e. correlation filters has been explained in Kumar (1992) and Alfalou et al. (2010). Face verification using correlation filters can be found in Savvides et al. (2002) while lower face verification centered and lip using correlation filters have been discussed in Samad et al. (2007) and Ramli et al. (2007). Synthetic Discrimination Function (SDF) and Equal Correlation Peak SDF (ECP SDF) based techniques have been proposed to overcome the problem when matched filters tend to drops due to changes of scale, rotation and pose of reference images. Research on SDF has been implemented in Rehman et al. (2005) while ECP SDF has been explained in Alkanhal (2006).

Advanced correlation filter namely MACE and UMACE have been applied in image processing which mostly used in authentication and identification process. The advantages of MACE and UMACE are that they are easy to be implemented and do not require large number of training images as reported in Tahir et al. (2005), Ramli et al. (2008) and Ghafar et al. (2008).

Pattern matching techniques using probability density function has been discussed in Schwartz et al. (1982). The K nearest neighbours method combines the strengths of the DTW and VQ methods where it keeps all the data obtained from training phase. Higgins et al. (1993) has proposed the used of K nearest neighbours technique during the pattern matching process. Another technique i.e. verbal information verification (VIV) has also been discussed by Li et al. (2000).

2.4.2 Lip Reading Trait

According to Petajan and Brooke (1988), humans use lip reading to enhance speech recognition especially when the signal is degraded by noise or hearing impairment. A number of techniques have been reported to extract mouth features for