Multimodal Biometrics Score Level Fusion Using Non-Confidence Information

Chaw Poh Chia

A thesis submitted in partial fulfillment of the requirements of
Nottingham Trent University for the degree of
Doctor of Philosophy

March 2011

Abstract

Multimodal biometrics refers to automatic authentication methods that depend on multiple modalities of measurable physical characteristics. It alleviates most of the restrictions of single biometrics. To combine the multimodal biometrics scores, three different categories of fusion approaches including rule based, classification based and density based approaches are available. When choosing an approach, one has to consider not only the fusion performance, but also system requirements and other circumstances.

In the context of verification, classification errors arise from samples in the overlapping region (or non- confidence region) between genuine users and impostors. In score space, a further separation of the samples outside the non-confidence region does not result in further verification improvements. Therefore, information contained in the non-confidence region might be useful for improving the fusion process. Up to this point, no attempts are reported in the literature that tries to enhance the fusion process using this additional information. In this work, the use of this information is explored in rule based and density based approaches mentioned above.

The first approach proposes to use the non-confidence region width as a weighting parameter for the Weighted Sum fusion rule. By doing so, the non-confidence region of the multimodal biometrics score space can be minimised. This effectively leads to a better generalisation performance than commonly used Weighted Sum rules. Furthermore, it achieves fusion performances comparable to the more complicated training based approaches. These performances are not only achieved in a wide range of bimodal biometrics experiments, but also in higher dimensional multibiometrics fusion. This method also eliminates the need for score normalization, which is required by other rule based fusion methods.

The second approach proposes a new Gaussian Mixture Model based likelihood ratio fusion method. This approach suggests the application of this density based fusion to the non-confidence region only and directly reject or accept the samples in the confidence region. By applying Gaussian Mixture Model to the non-confidence

region, a smaller and more informative region, the impact of an inaccurately chosen component number on the fusion performance can be reduced. Without tuning or using any component searching algorithm, this proposed approach achieves comparable performance to the one using specific component number searching algorithm. This successful demonstration means less resource is required whilst comparable performance can be achieved and processing time is also significantly reduced.

Declaration

Acknowledgement

First of all, I would like to thank Nottingham Trent University for providing me the Vice-Chancellor Bursary for this PhD study. Without this offer, to pursue for a doctoral degree might be still a dream for me.

My sincere gratitude dedicates to the supervisory team members, Prof. Nasser Sherkat and Dr. Lars Nolle, who are the Director of Studies (DOS) and second supervisor respectively. Thanks for their confidence showing to me to manage the project autonomously. The sharing of their precious research experiences and academic knowledge, the motivation and supervision giving to me are the factors that are indispensable for this research work to be efficiently completed in 3 years time.

I am also very grateful to the authors of NIST-BSSR1 multimodal biometrics database and Xm2vts benchmark database for making these score set databases freely available. It saved me a significant effort in data collection and enabled the core of this research to be conducted.

Thanks to all my colleagues who have always offered to help and advise me. Thanks to all my friends and family for their endless encouragements. Especially my parents and siblings, they all have always offered moral support to empower me to persistently and patiently pursue this research.

Finally and most importantly, my heartfelt thanks go to my wife, Oi Lay. Thanks for the sacrifices you have done and accompanying me in UK during this PhD study period. I appreciate you taking care of yourself on my behalf when I was not beside you. Thank also for your patience, encouragement and compliments.

Chaw Poh Chia

Nottingham, February 2011.

Table of Contents

List of Figures

List of Tables

Page

List of Abbreviations

ANN	Artificial Neural Network
AOI	Area of Interest
ATM	Automatic Teller Machine
BBN	Bayesian Belief Network
B-LLR	Likelihood ratio based fusion using GMM with best component number.
CP	Confidence Partition
CWT	Continuous Wavelet Transformation
DET	Detection Error Trade-Off
DPW	D-Prime Weighted
EER	Equal Error Rate
EERW	Equal Error Rate Weighed
EM	Expectation Maximisation
EW	Equal Weighted
FA	False Acceptance
FAR	False Acceptance Rate
Fc	Facial Matcher C
Fg	Facial Matcher G
Fli	Left Index Fingerprint
Fri	Right Index Fingerprint
FPVP	Feature Points of the Vein Pattern
FSPL	Four-Segments-Piecewise-Linear
FTE	Failure to Enrol
FR	False Rejection
GAR	Genuine Acceptance Rate
GCP	Genuine User Confidence Partition
GMM	Gaussian Mixture Model
IBG	International Biometric Group
ICP	Impostor Confidence Partition
ICPR	International Conference on Pattern Recognition
ID	Identity

IRLS	Iteratively Re-weighted Least Squares
JLLR	Likelihood Ratio Based Fusion Using Joint Density
KCCA	Kernel Canonical Correlation Analysis
KDE	Kernel of Density Estimator
K-NN	K-Nearest Neighbours
LLR	Likelihood Ratio
LP1	Lausanne Protocol 1
LREG	Logistic Regression
LTL	Linear-Tanh-Linear
LVQ	Learning Vector Quantization
MAD	Median Absolute Deviation
MAX	Max rule
MLE	Maximum Likelihood Estimation
MLLR	Likelihood Ratio Based Fusion Using Marginal Density
MLP	Multilayer Perceptron
MP	Merged Pattern
MR	Modality Reliability
NCW	Non-Confidence Width
NCWW	Non-Confidence Width Weighted
ORL	Olivetti Research Laboratory
PIN	Personal Identity Number
PKI	Prior Knowledge Incorporation framework
QLQ	Quadratic-Line-Quadratic
QQ	Two-Quadratics
R-LLR	Likelihood ratio based fusion using GMM with random component number.
R-SBLLR	Likelihood ratio based fusion using GMM with random component number which is applied to sum bounded samples only.
ROC	Receiver Operator Characteristic
SBLLR	Sum Bounded Likelihood Ratio fusion
SMO	Sequential Minimal Optimisation
SPQA	Speech Quality Assurance Algorithm
SUM	Sum rule
SVM	Support Vector Machine

TA	True Acceptance
TER	Total Error Rate
TR	True Rejection
US-VISIT	United States Visitor and Immigrant Status Indicator Technology

List of Publications

The following publications describe parts of the work that have been devised:

[140] C. Chia, N. Sherkat and L. Nolle, “Confidence Partition and Hybrid Fusion in Multimodal Biometric Verification System,” International Conference on Biometrics ID Management and Multimodal Communication, Madrid, Spain, Springer LNCS 5707, pp. 212-219, September 2009.

[141] C. Chia, N. Sherkat and L. Nolle, “Towards a Best Linear Combination for Multimodal Biometric Fusion,” 20th International Conference on Pattern Recognition (ICPR 2010), pp. 1176-1179, Istanbul, Turkey, August 2010.

1. INTRODUCTION

1.1 Biometrics: An Authentication Approach

Biometrics have recently generated a lot of interest to be used as effective methods for identity authentication (identification and verification), particularly after the attacks on the United States on September 11, 2001, and the railway explosion terrorist attack in Madrid, on March 11, 2004. In addition, as the information society increasingly affects every aspect of life, the need to raise security level to ensure the identity of the person accessing information increases [1]. The conventional authentication methods, such as passwords or Automatic Teller Machine (ATM) cards, no longer fulfil the stringent security requirements mentioned above [2]. Some anecdotal references [3] have even predicted the death of passwords, to be replaced by various biometric authentication mechanisms. Biometrics benefit from the fact that they cannot easily to be lost or stolen, they are difficult to copy by others and they require genuine users to be present. Therefore, they appear to be a better option for information security.

The purpose of authentication is to answer the question, “Who is this person?” or “Is this the genuine user?”. Whereas the conventional methods, e.g. Personal Identity Numbers (PIN) or ATM cards are used to authenticate the claimant through answering the questions: “What do you know?” (knowledge-based) and “What do you have?” (token-based) respectively. Biometrics in contrast to these methods are more reliable authentication tools. It is a more intuitive and direct way to provide biometrics to answer the question “Who are you?”

1.1.1 Operation Modes of Biometrics Systems

Based on different application contexts, there are two different operation modes for biometric authentication: verification and identification. Fig. 1-1 presents the operation block diagrams of biometrics systems.

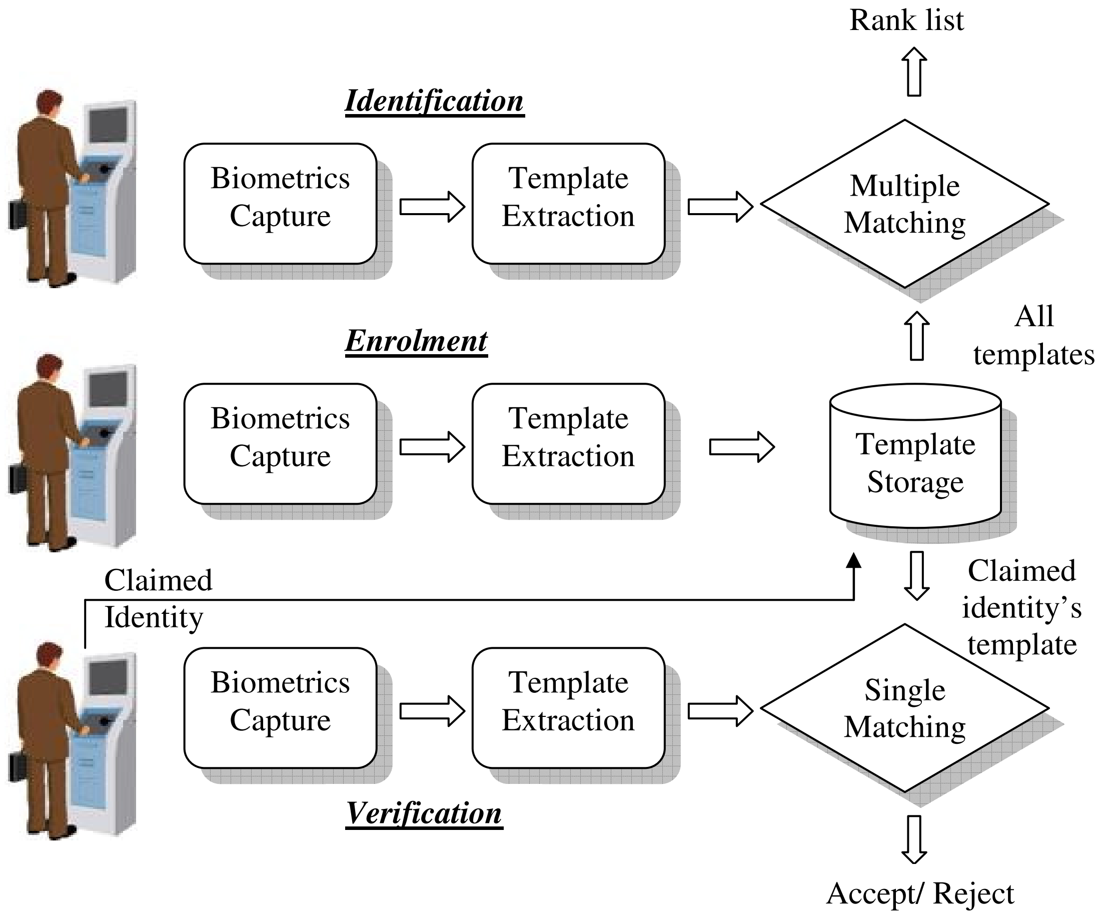


Fig 1-1. Operation modes of biometrics systems.

For both, the verification and identification mode, a user has to enroll to the system before his/her biometrics can be potentially accepted. By providing the required biometrics through the capturing device (fingerprint reader, iris scanner, etc.), the user's biometric template is extracted and stored in the system's central database. A biometric template is a digital reference of distinct characteristics that have been extracted from a biometric sample. Templates are used during the biometric authentication process. It is a main public concern that their biometrics might be compromised because of the central storage architecture of the biometrics system [4]. But in most cases, only the biometrics template will be stored [5], i.e. only certain characteristics of the biometrics will be extracted and stored but not the raw biometrics sample. By doing this, not only the size of the biometric storage can be greatly reduced but the reconstruction of the original biometrics can be avoided [6].

For verification, a claimant tells the system who he/she is and provides biometrics to the system. The pre-stored biometrics template will be retrieved based on the claimed identity and his/her provided biometrics' characteristics will be extracted to form another template. A one-to-one matching is performed in verification mode among these two templates to make the decision: to ACCEPT or to REJECT the claimant as the identity he/she claimed.

In contrast to the verification mode, in identification mode, the user only provides biometrics without telling the system the claimed identity. The new template which is constructed from the provided biometrics is then matched with all the templates in the database to generate a ranked list. The identity on top of the list will be assigned to the claimant. Therefore the identification mode is a one-to-many matching scheme.

1.1.2 Measurement of Biometrics

Fig. 1-2 illustrates how the biometrics is processed after the capturing stage by using the facial biometrics as an example.

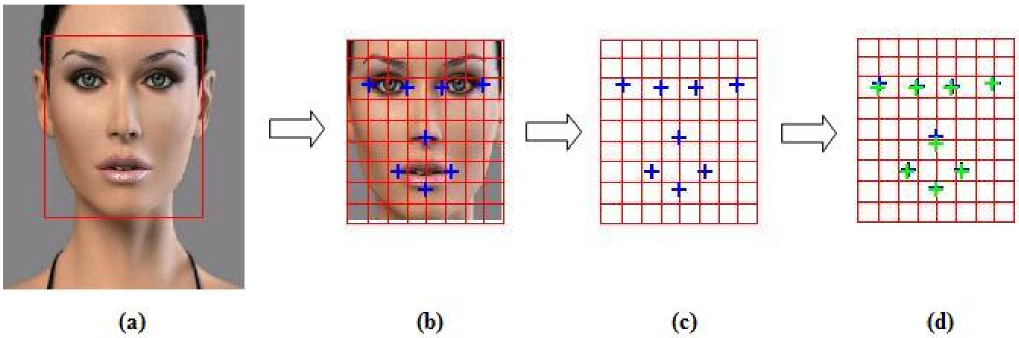


Fig 1-2. An example of biometrics processes.

The user interacts with the sensor to provide a biometrics sample, which is digitised. The digital biometrics will then be enhanced for more efficient feature extraction. For facial biometrics, the Area Of Interest (AOI) will be identified as shown in fig. 1-2(a). In (b), the feature extractor chooses what and how the characteristics are extracted. The template shown in (c) is then generated through integration of all these extracted features. This template is matched in (d) with another template(s) retrieved from the central database to produce the similarity or distance metric as a confidence index to authenticate (to identify or to verify) this person.

1.1.3 Conventional and Novel Biometrics

Some of the conventional biometrics are illustrated in fig. 1-3. The human traits that can be used as biometrics are categorised into physiological and behavioral biometrics. A static or physiological trait, for instances the face, iris, hand geometry or palm print, provides static characteristics. The signature, speech and gait are considered as dynamic or behavioral biometrics. One can extract the recorded dynamic characteristics, e.g. for signature, the writing pressure and inclination over the signing period [7]. Depending on the extracted features, some of the biometrics are in both categories. For example the fingerprint can be used as behavioral or physiological biometrics. When the fingerprint is used as static biometrics, it is easy to spoof by presenting artificial fingerprints. To tackle this issue, the research in [8] uses dynamic features of fingerprints extracted from video sequences that are captured at the image acquisition stage.

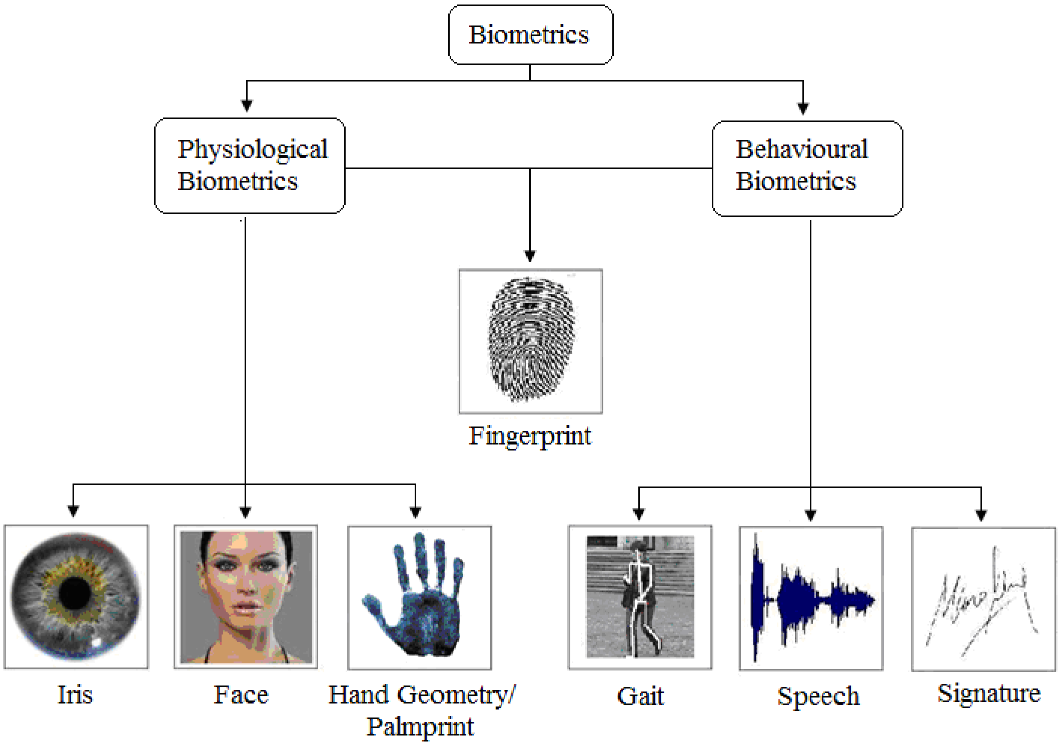


Fig 1-3. The physiological and behavioral biometrics examples.

Whether a human being's biological or behavioral traits can be used to establish identity depends on seven factors listed below [9]:

1. Universality: How common is this biometrics possessed with in the population?
2. Uniqueness: How distinctive this biometrics among the population?
3. Permanence: How invariant is this biometrics over time?
4. Measurability: Is this biometrics collectable and digitisable?
5. Performance: Are the speed, accuracy, robustness and cost of such biometrics acceptable?
6. Acceptability: How willing is the population to present such biometrics?

7. Circumvention: How easy is it to accept a fake biometrics?

Many human traits have been used as biometrics, each of them having limitations. Therefore the search for a new biometric trait has not ended. As long as the human physiological and behavioral traits fulfill the seven conditions listed above, they can be used as biometrics. Some of the state-of-the-art biometric traits are listed in table 1-1.

	Trait	Author	Descriptions
1	Electro-Cardiogram	[10]	Utilises the simple distance measure of heart vector as biometric feature.
2	Eye gaze	[11]	Features are extracted from raw eye-tracking data that preserve the key characteristics of the scan path. The eye gaze is combined with keystroke in this work.
		[12]	Examines the reaction of a human's eyes to visual stimulation. The person to be identified was asked to follow a point on a computer screen.
		[13]	The user sequentially looks at certain parts of a picture to create gaze-based signature.
3	Mouse curve	[14]	Uses the curve's length, curvature, inflection and straightness as features.
4	Gait	[15]	Uses the joint angle trajectories of lower limbs as dynamic information. And uses the Procruste shape analysis to obtain a compact appearance representation as static information.
		[16]	The detected silhouettes are used to build an averaged representations using eigenstance shape models. The similarity measures are based on these averaged representations.
		[17]	Different components of human bodies are shown to have unequal discrimination power. Assigning weights to these components shows improvement to the recognition rate.
5	Finger (Top view)	[18]	Uses the top view image of a finger to create feature map which is called nail code. Nailcode is employed for Euclidean distance computation.
6	Palm vein	[19]	Uses the x and y coordinates, the gray values, temperature gradient and the gradient direction to create

			Feature Points of the Vein-Patterns (FPVPs)
7	Heart sound	[20]	Analyses the heart sound in frequency domain and uses the log cepstral coefficients as the features
8	Ear	[21]	The features are extracted by the convolution of each sub-window with a bank of Gabor Filters. Their dimensionality is reduced by Laplacian Eigen Maps
9	Tongue print	[22]	Uses the geometric features, crack features and texture features of tongue.
10	Eye shape	[23]	Static eye information is obtained by using the Gabor wavelet coefficients of four feature points around black eye area. The dynamic eye shapes (blinking) information is extracted based on the size change of black eye area during blinking.
11	Soft biometrics	[24]	Combines the body weight and fat to aid fingerprint. These characteristics can be used directly without further processing.

Table 1-1. The novel biometrics examples.

1.2 Biometrics History, Development and Its Merits.

To systematically identify a person through referring to individual characteristics measurement had first appeared in 19th century [25]. Alphonse Bertillon, a French police officer had invented anthropometry or Bertillonage. This was a system using the physical measurements as the fig. 1-4 shown for human identification purpose.

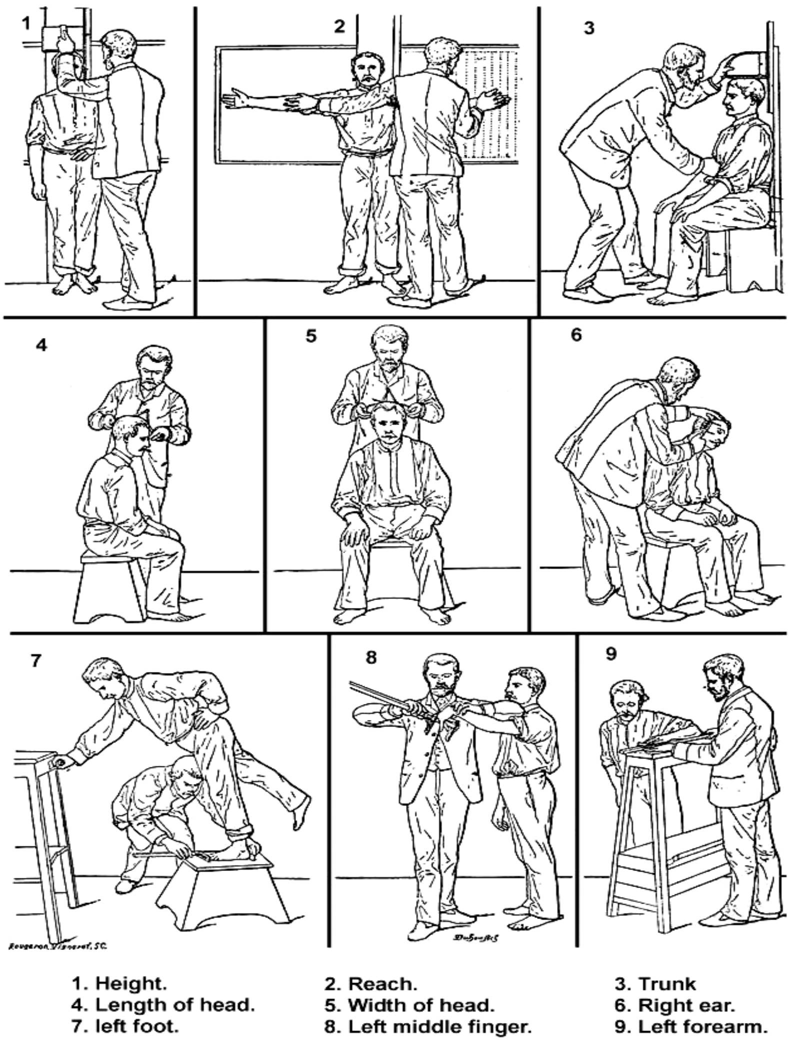


Fig 1-4. Bertillonage or anthopometric measurements (figure obtained from [26]).

With increasing popularity over the 20th century, biometrics have been continuously developed from manual to semi-automated and eventually to fully automated mode like what we see today. Biometrics have been extensively applied in various fields but not just for law-enforcement purpose (i.e. to identify a criminal). According to International Biometric Group (IBG), the biometrics worldwide market was expected to expand to a value between $5.7-$5.8 billion by 2010 [5]. The usage of biometrics not only has been driven by government and the public sector, the private sector has also increasingly shown its interest in such applications. For the public sector, aside from the law-enforcement purpose, biometric applications have been widely implemented for border control and other security purposes. As shown in [27], the United States Visitor and Immigrant Status Indicator Technology (US-VISIT) programme and UK implementation of eBorders are public sector usage examples. Biometric applications were also employed in large scale events such as 2008 Beijing Olympics [28]. Some other private sector's biometric applications examples are: ABN Amro bank introduced the voice verification and recognition technology to their call service centre, fingerprint ticketing was brought in to Disneyland and Mitsubishi Securities used biometrics on their trading floor to cure their too-many-passwords problem [2].

The conventional authentication methods no longer comply with the stringent authentication requirements. A password, for instance, is very easy to forget. People find password management very annoying. For the token based method, an ATM card can be easily used by somebody else or lost.

In contrast to both of these methods, biometrics appears to be a solution to overcome restrictions of conventional authentication methods. Biometric authentication cannot be forgotten or lost. Furthermore, in the authentication process, providing biometrics to the system is the proof of the claimant's presence. Unlike the password or ATM card, a biometrics is more difficult to copy or to falsify. Additionally, a biometrics can be combined with password or/and an ATM card to form two or more authentication factors. By doing so, the authentication rate can be further enhanced without having to replace these existing systems [29].

1.3 Evaluation of Biometrics Verification Performance

The matching between the centrally stored template and the template constructed from a claimant generates a confidence score to verify whether they are an impostor or a genuine user. There is always overlap region between the score distributions of the genuine user and impostor for a practical biometric system as shown in fig. 1-5. It causes the difficulty in classifying the claimant into the correct categories. The reasons of this overlap region formation are discussed in the next section. As the figure shown, there are two types of errors present in biometric verification: False Acceptance (FA) and False Rejection (FR). A verification threshold, Δ is needed in the overlap region as a reference to do the classification. Varying this threshold affects these two error rates.

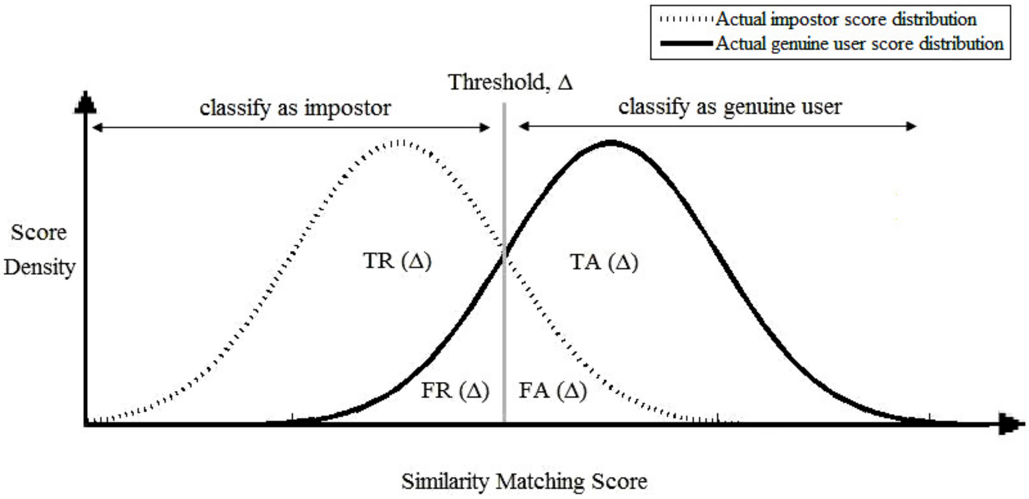


Fig 1-5. Decision making of biometrics based on the threshold (Δ).

Δ is used to establish the security level of a biometrics verification system. It can be seen that for those who obtain a similarity matching score less than Δ will be classified as an impostor. If one is verified with the similarity matching score higher or equals to the threshold, his/her claimed identity will be accepted. A higher Δ represents a higher security level. Undoubtedly, less impostors will get through verification because of the higher security level. But a genuine user with score less than Δ will also be rejected at the same time. Conversely, by adjusting the threshold to a lower level will reduce the number of the genuine users being falsely rejected. However, this will also cause an increase of falsely accepted impostors. In brief, there is a trade-off between these two types of errors.

There are four possible verification outcomes that a claimant can obtain: the FR, FA, which are negative results, or the positive results, True Acceptance (TA) or True Rejection (TR). Four of these outcomes are a function of Δ. Their relationship can be clearly represented through the confusion matrix given in fig. 1-6.

	Predicted Genuine User	Predicted Impostor
Actual Genuine User	TA	FR
Actual Impostor	FA	TR

Fig 1-6. Confusion matrix of biometric verification.

A Receiver Operator Characteristic (ROC) graph is commonly used to visualise the performance of biometrics verification. It is constructed by a series of False Acceptance Rate (FAR) and its associated Genuine Acceptance Rate (GAR) under different operating thresholds. From rule (1.1) and (1.2), FAR is the ratio of total FA cases to the total impostor attempts, *NI.* FRR is the ratio of total FR to total genuine user trials, *NC.* Both FAR and FRR are also the functions of Δ. Rule (1.3) shows the Genuinely Acceptance Rate (GAR) in term of FAR. The Detection Error Trade-Off

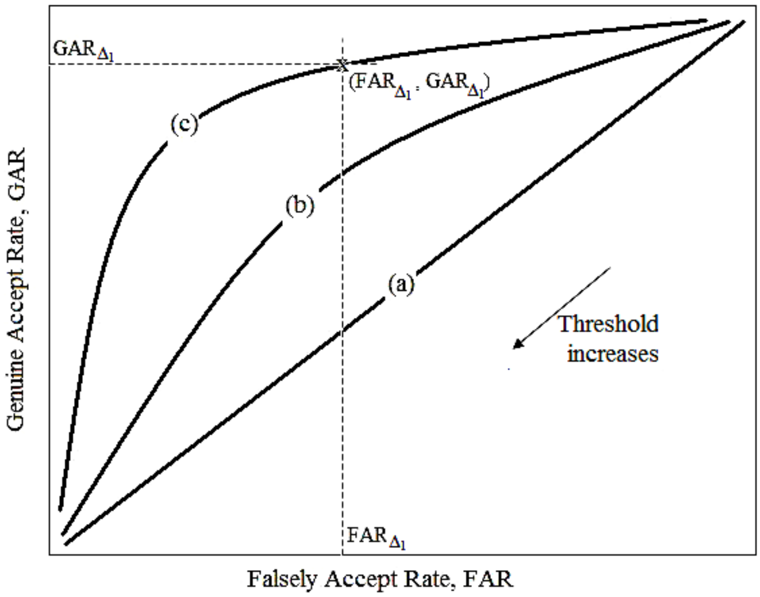
(DET) curve is another commonly used graph to visualise performance. This is a plot of FAR versus FRR that emphasises both types of errors [30], [31].

$$\text{FAR}(\Delta) = \frac{FA(\Delta)}{NI} \times 100\% \tag{1.1}$$

$$\text{FRR}(\Delta) = \frac{FR(\Delta)}{NC} \times 100\% \tag{1.2}$$

$$\text{GAR}(\Delta) = 1\text{- FAR}(\Delta) \tag{1.3}$$

For biometrics fusion research, ROC is more commonly used. Choosing a specific Δ will generate a (FAR, GAR) pair. The manipulation of Δ from minimum to maximum within an appropriate interval will generate a series of (FAR, GAR) pairs. All these pairs are plotted and the connection of these points constructs the ROC curve. It is as shown in fig. 1-7. In the biometrics research community, the GAR is always plotted against FAR in a semi-logarithmic scale in biometrics fusion research field. This is because the value of FAR is much smaller than the GAR. Plotting in a semi-logarithmic scale visualises the verification performance over a series of operating points in a better way. However for numerical result comparison, the logarithmic scale for FAR is not used. Instead, the GAR is frequently reported under certain FAR, e.g. 0.001%, 0.01%, 0.1% and EER. This is because the cost of accepting an impostor may be very different from the cost of rejecting a genuine user (depending on the biometric application).



** This is not a semi-logarithmic ROC plot.*

Fig 1-7. ROC curve examples for biometrics verification with similarity metric.

Fig. 1-7 depicts three ROC curves illustrating three different verification performances. As the curves show, increasing the threshold results in lower GAR and FAR. Curve (a) shows a linear relationship between the FAR and GAR. When increasing the threshold, a FAR decrement is followed by a proportional decrement of GAR. Curve (b) shows a slightly better verification ability. This is because when FAR is decreased, GAR is decreased at a smaller rate. Curve (c) shows the best verification ability amongst the curves. A biometrics verification system always aims to achieve 0% FAR and 100% GAR. Therefore the closer the ROC curve to this operating point, as curve (c) demonstrates, the better.

The work in [32] demonstrates that fingerprint identification rate achieves up to 95% accuracy for a database size of 500 samples. However, it drops to 90% and 86% for database size 10,000 and 100,000 respectively. Therefore it is necessary to evaluate biometrics using appropriate size of databases that depends on different application context[33]. Testing results for large databases corresponds to relatively low FAR. The scales of GAR and FAR will be very different. So for better visualisation of the biometrics performance, it is common for a ROC graph to use semi-log plotting as the logarithmic scale is used for FAR. Semi-log plotted ROC curves can be easier interpreted and compared.

Equal Error Rate (EER) and D-Prime (*d'*) are two other parameters used to report the verification ability. It is mentioned previously that there is a trade-off between FAR and FRR. By varying the threshold, there is a trade off point where the FAR equals FRR and it is termed as EER. *D'* is a statistical measurement of the separation between the impostor and genuine user score distributions. This is depicted in equation (1.4) where μ^G and μ^I are the genuine user and impostor score distributions' mean and σ^G and σ^I are their respective standard deviations.

$$d' = \frac{\mu^G - \mu^I}{\sqrt{(\sigma^G)^2 + (\sigma^I)^2}} \tag{1.4}$$

1.4 Single Biometrics Limitations and Multibiometrics Fusion

Biometrics as the authentication tool have been extensively accepted and employed for practical use. Nevertheless, further development and applicability of single biometrics has come to saturation.

Single biometrics performance in term of enrolment rate is not sufficient for larger population coverage. For example, 2% of the population as reported in [32] failed to enrol to the fingerprint system due to their fingerprints' friction ridges being too damaged to be matched. This can happen to those who work in beauty salons whose fingerprints can be damaged by chemicals or bank cashiers who have to flip over countless banknotes using their fingers. Therefore it is possible that they possess damaged fingerprints as well [1]. Some other reasons of enrolment failures can be due to the fact that they are born with less discriminative biometrics or because of physical body changes. For example, Asian women normally have flatter fingerprint and people with eye illness or pregnant women's irises can change [5].

A single biometrics authentication rate is limited. In the case of verification, there is an overlap region between the similarity score distributions for true matching (genuine user) and false matching (impostor) as shown in fig. 1-8. This region is where the errors arise e.g. a genuine user may have a lower similarity matching score or an impostor may obtain a higher similarity score. Some of the reasons that cause the formation of this overlap region are listed below:

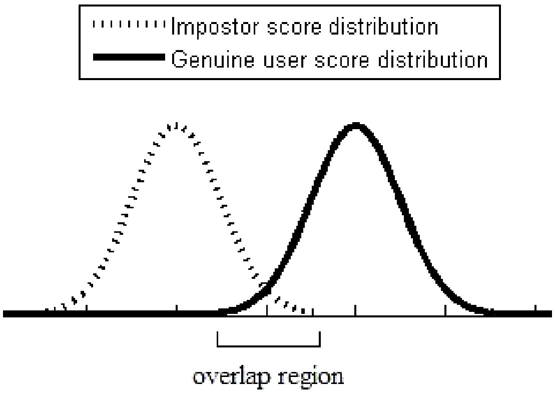


Fig 1-8. Overlap region of the genuine user and impostor score distributions

1. Incorrect interaction with the capturing device: The template quality deteriorates if the biometrics is not properly provided. For instance, the work in [34] suggests pre-alignment of the fingerprint otherwise a non-universal frame of the fingerprint will probably affect the success of the features extraction. The face pose variation is

one of the main concerns in the facial recognition research field [35]. It causes difficulties to facial biometrics authentication as well. Other than the user's improper way of providing biometrics, the system operator may inappropriately manipulate the system settings causing the sensor to less effectively capture the biometrics.

2. Capturing ambient noise: A reliable biometrics acquisition highly relies on the ambient condition. For example the camera or camcorder, which is used as the biometrics capturing device, relies on the ambient light condition for a satisfying capture. The speech/voice biometrics uses the voice print or sound wave to extract the biometrics features. Ambient noise might be integrated into the acoustic signal and cause significant negative impact on the authentication results.

3. Interclass similarity: Identical twins may have distinctive iris or fingerprints but a facial identification system may have difficulties to differentiate them. A biometrics with lower uniqueness will have higher interclass similarity. In [36], the distinctiveness of the fingerprint is classified as "high" whereas the hand geometry is classified as "low". This results in smaller overlap regions for the fingerprint biometrics comparing to the hand geometry biometrics. It can be seen from [37] the ROC graph of hand geometry has a lower ROC curve than fingerprints due to lower distinctness.

Since biometrics applications are widely spread, spoof attacks have attracted great interest from researchers [38], [39], [40]. A single biometrics system can be spoofed easily. The authors in [41] successfully deceive a fingerprint system by using artificial silicone and gelatine fingers. One of the fingerprint systems is even accepted by the ink-printed fingerprint. There are more spoofing examples given in [42] where commercial biometrics applications, using fingerprints, facial and iris recognition systems, are spoofed. By playing back a video of a person's face, the facial biometrics is counterfeited whereas the iris system is fooled by using high resolution digital iris images.

The above mentioned single biometrics restrictions result in errors such as the rejection of a legitimate user or the false acceptance of an impostor. These limitations of single biometrics cannot be lessen by simply improving the individual biometrics. Since insufficient information utilisation in the system might lead to a failure of biometrics [44], the integration of more evidence from the claimant is a feasible way to enhance the biometrics performance and usability [43]. This use of more than one biometric factor in establishing and verifying the identity of a given person to improve the accuracy, reliability and usability of the biometrics system is termed as biometrics fusion.

Multibiometrics systems which are requesting more biometrics evidence tend to reduce the authentication errors and other limitations. The fusion of redundant information from different sources enhances the overall system certainty whilst the fusion of complementary information results in information gain to reduce the system's uncertainty. Therefore to further enhance single biometrics usability and performance, multibiometrics is one of the suitable solutions.

1.5 Research Aim and Objectives

Multibiometrics authentication benefits from additional evidence that are provided by the users (multisample, multiunit and multimodal), or are generated by different capturing devices (multisensor) or different matching algorithms (multialgorithm). All of these sources provide information gain so the error to authenticate a person is reduced. Biometrics information gain can also be achieved by exploring the biometrics sample quality or other underlying information. Biometrics fusion plays a key role to effectively combine all these information.

Therefore, the aim of this research is to improve biometrics verification through effective fusion of multibiometrics and other useful information. To achieve this aim, the objectives of the research more specifically are:

- To establish a baseline for individual biometric verification performance.
- To develop a multibiometrics verification testing framework.
- To investigate different methods of biometrics fusion.
- To explore additional information that aids the biometrics fusion.

1.6 Outline of The Thesis

This chapter sets the scene by providing the foundation knowledge in biometrics and the related performance assessment. Single biometrics performance and usability are limited. Chapter 2 presents the literature review of how the usability and performance of single biometrics are improved in the context of biometrics fusion. These works are discussed from two broad perspectives, information gain and information fusion. Chapter 3 compares different fusion approaches that are reported in the literature with top performance by using two publicly available truly multimodal biometrics databases and 19 cross validated bimodal biometrics experiments. Not only the fusion performances are compared, the limitations, resource requirements and processing/training time are also included in this evaluation. The details of two databases which are used throughout this research work are also given in this chapter. A hybrid fusion method is proposed in chapter 4 to improve the conventional biometrics fusion performance. The fusion improvement is achieved with manual operation of this proposed method. However, from this work, it can be seen non-confidence region sample plays a key role in fusion. This finding supports the idea to incorporate the non-confidence region related information to the conventional state-of-the-art fusion approaches to further improve their performance and usability. Chapter 5 shows how the non-confidence region width can be used in a rule based fusion method, the Weighted Sum rule, to achieve better weighting than the conventional schemes. Through using this parameter for weighting, the generalisation error can be reduced to minimum. In chapter 6, the benefits of employing the non-confidence samples in the state-of-the-art density based fusion method are investigated. It shows the comparable fusion results to the state-of-the-art density based fusion method with significant reduction in training time and requires less resource. Chapter 7 concludes this work and suggests the improvement and extension of this research.

2. LITERATURE REVIEW

A literature review from the perspective of biometrics information gain and information fusion is presented and its structure is depicted in fig. 2-1. Section 2.1 gives the details of additional biometrics information sources that have been explored to enhance authentication. In this section, multibiometrics information and biometrics quality that is directly measured or indirectly derived are reviewed in section 2.1.1 and 2.1.2 respectively. Other reported sources include soft-biometrics, probabilistic reliability of biometrics, failure prediction of ROC and prior knowledge of classifier space. Their details are presented in section 2.1.3.

Section 2.2 discusses the state-of-the-art biometrics fusion algorithm research. Section 2.2.1 further categorises biometrics fusion methods into serial and parallel modes and describes work on biometrics fusion from three different structural levels in section 2.2.2. They are included in section 2.2.2.1~2.2.2.3 for feature level, measurement level and decision level fusion respectively. A significant amount of measurement level fusion has been reported in the literature, mainly because this is the most appropriate fusion level. Therefore measurement level fusion is reviewed separately in section 2.2.3. These measurement level fusion methods are further separated into rule based, classification based and density based approaches according to [45].

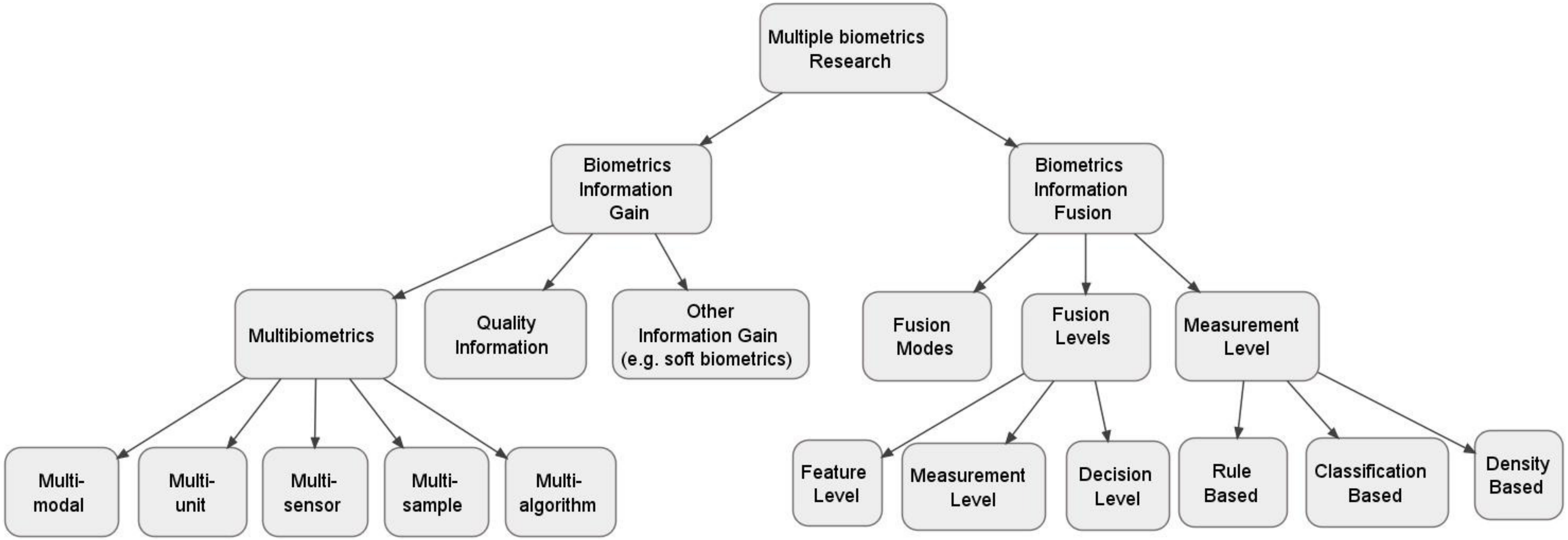


Fig 2-1. Structure of the literature review.

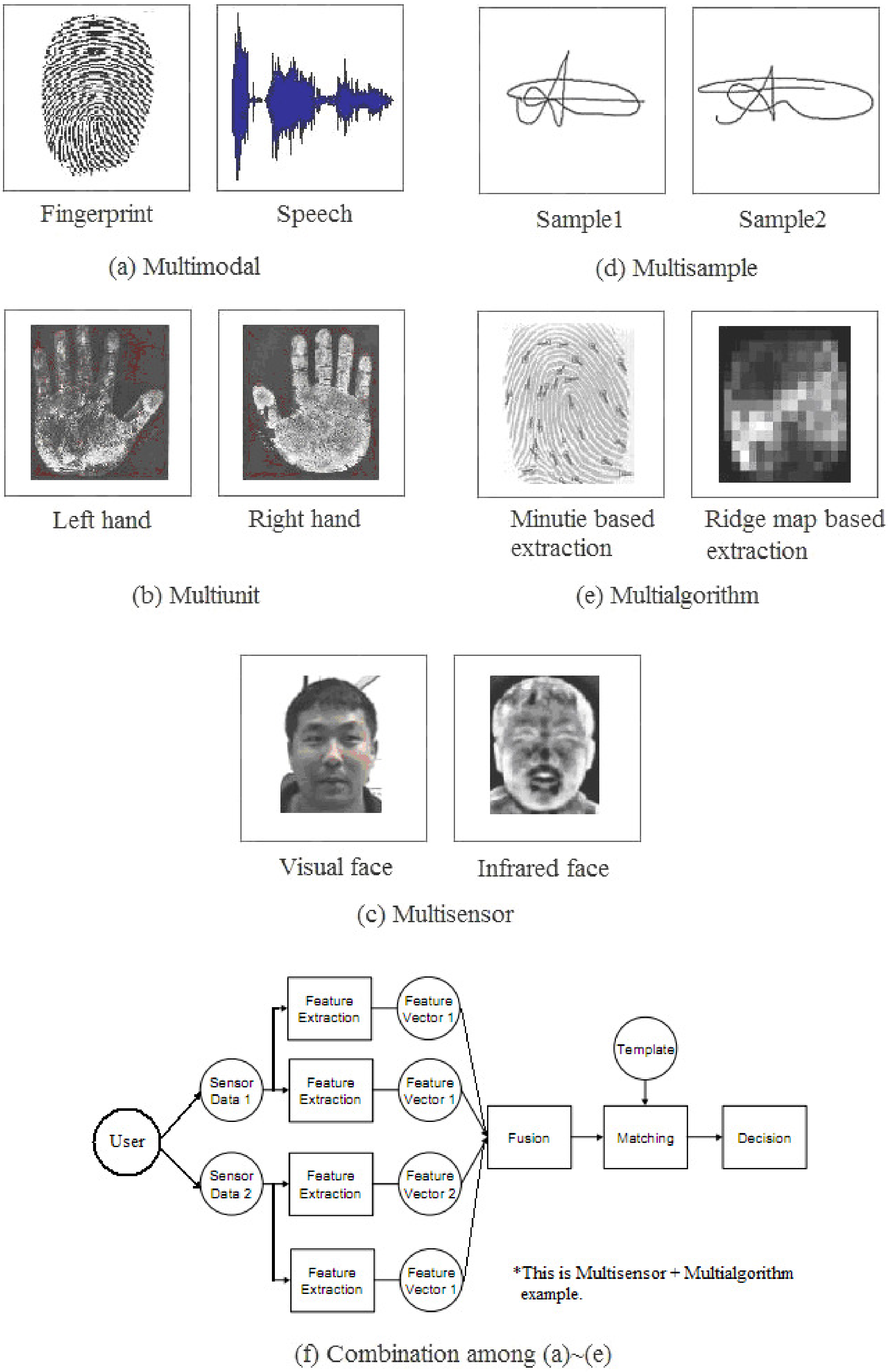
2.1 Biometrics Information Gain

Biometrics information gain mentioned in this thesis generally refers to the additional information source, other than the single modality biometric matching score, that aids fusion performance. Biometrics information gain reduces uncertainty to authenticate a person. In the following section, different sources used to increase the biometrics information gain are presented.

2.1.1 Multibiometrics

Fig. 2-2 illustrates different scenarios of multibiometrics used to increase the information gain. The choice of the best fusion scenario is not based on performance of individual sources but depends on the correlation or statistical independence between the different fused sources [46]. According to experiments published in [47], a positive correlation degrades the fusion performance whereas a negative correlation improves the fusion. Multiple modalities of biometrics are inherently different so this scenario has the lowest correlation and therefore can achieve a better information gain compared to the other scenarios.

Some of the multimodal biometrics research examples are the combination of hand geometry, fingerprint and facial biometrics in [43] and BioID identification system in [48] that uses lip movement, facial image and speech biometrics. BioID extracts the multimodal biometrics through a recorded video of a speaking claimant. It is a preferred method because this single section acquisition process is more user friendly than the one in [43], i.e. it does not need to interact with multiple capturing devices. Furthermore lip movement and speech are dynamic biometrics. Such biometrics aid to identify the living state of the claimant to prevent spoof attack [38], [39]. Moreover a dynamic biometrics can also be combined with knowledge based authentication to further enhance security. For instance, a system can require the claimant to provide his password utterance as speech biometrics or to instruct the user to sequentially look at certain points on a screen (gaze based biometrics) can be used to create password entries [13], [49]. Further multimodal biometrics combination studies are available in [50], [51], [52]. From these works, it can be summarised that multimodal biometrics is a preferred approach than the other scenarios to tackle single biometrics limitations because of the following reasons:



(a) Multimodal

(d) Multisample

(b) Multiunit

(e) Multialgorithm

(c) Multisensor

(f) Combination among (a)~(e)

Fig 2-2. Multibiometrics combination scenarios.

1. Multimodal biometrics have the lowest correlation among the sources therefore higher information gain can be achieved. The uncertainties that are caused by low quality biometrics, interclass similarities, etc. can be reduced.

2. It provides alternative biometric options for a claimant who is unable to provide a specific biometrics. Consequently it increases the universality of the system to cover a larger population. By doing this, the Failure to Enrol (FTE) rate can be significantly reduced.

3. The usability of the multimodal biometrics system is also better than the single biometrics. For example, a speech biometrics can be used instead of facial biometrics under faint light condition or a facial biometrics can be given higher weight than the speech biometrics in a noisy ambiance.

4. Multimodal biometrics is more difficult to spoof because multiple modalities have to be presented at the same time especially for the system that combines static and dynamic biometrics which involves temporal analysis. The liveness of the claimant can be detected to prevent the impostor to spoof the system with artificial biometrics.

Although the development in the field of multimodal biometrics has received considerable attention, there are some disadvantages in this approach:

- Additional hardware costs.
- Additional enrolment time and system processing times.

- More complicated software design and data management, e.g. elective control and management of access to the biometric data and systems where privacy, confidentiality and trust are of primary concern [156].

The fingerprint from different fingers, the iris from left and right eyes, the left and right side profile faces and the hand geometry or the palm print from left and right hands are the examples of multiunit biometrics. Some of the multiunit biometrics usage examples in the literature are given as followed. In [65], three different views of the face with different head poses are used for gaining additional information. The NIST-BSSR1 database detailed in [53] contains the verification matching scores from left index and right index fingerprints. The multiunit biometrics fusion works employing these multiunit fingerprints' matching scores are available in [54], [55], [56], [57], [58], [59], [60]. In [61], the authors combine the fingerprints from little finger, ring finger, middle finger and index finger. This kind of combination has the benefit of biometrics acquisition can be done at the same time.

In the multisensor scenario, multiple sensors are used to complement the shortcomings of a specific sensor. For example, the use of a capacitive and an optical fingerprint sensor can be found in [62]. The capacitive sensor is used to eliminate the need of optical sensor for clean, undamaged epidermal skin and a clean sensing surface. Another example is shown in [63]. It uses an infrared camera, which is robust against ambient lighting and other variations such as facial hair, wrinkles and expression. It overcomes some limitations of the conventional visual camera used for facial recognition.

The same biometric is used more than once to achieve the best "scan" possible in multisample scenario. The dataset used in [62] comprises 10 impressions of different is a multisample scenario example. As in [64], such a scenario is also named as Single Source Multiple Sample fusion (SSMS). Since it uses only a single sensor, its implementation costs will not be as high as a multimodal biometrics system. Conventional SSMS uses the samples with the same view, but this work discloses that different views of the same source contain more disjointed features to increase the authentication performance.

Biometrics internal processing components include a preprocessing module to enhance the raw data, a feature extraction and a matching module. Different algorithms are capable of extracting and matching different discriminative information. Processing the biometrics with more than one algorithm is termed a multialgorithm scenario and it helps in gaining complementary information from a single source. For instance in [66], facial biometrics authentication performance is improved through applying Principal Component Analysis, Independent Component Analysis and Linear Discriminant Analysis facial classifiers. Both the minutiae and ridge flow fingerprint features are used in a hybrid fingerprint system in [67]. Instead of using multiple features, [68] uses a single feature but combining it with three different feature matching algorithms. Another approach, using multiple feature extractors and also multiple matchers, is presented in [69]. This approach not only uses minutiae and texture extractors, the minutiae extractor is further combined with two feature matchers, a string matcher and a dynamic matcher. All these approaches show the feasibility of gaining complementary information through different feature extractors and matchers from a single biometrics source to enhance the authentication performance.

Other scenarios aside from the multimodal biometrics are to obtain the information from a single modality of biometrics. These scenarios can be used to increase single modality biometrics usability and performance through complementary information obtained from the single biometrics source. However, the improvement is limited compared to a multimodal biometrics sources which are inherently different. Furthermore, these scenarios do not provide alternative biometrics option to cover larger population and are easier to be spoofed. Therefore, these scenarios are more frequently used to improve individual biometrics performance. Nevertheless, these scenarios can be implemented in a multimodal biometrics system to further enhance the usability and performance. In [70], the authors empirically show by averaging *m* modalities of biometrics and *n* samples per modality, the errors can be reduced by a factor from the [1, *nm*]. The NIST-BSSR1 database contains multimodal and multiunit biometrics score. All these scores (multimodal and multiunit samples scores) are combined in [55], [56], [57], [59]. In [71], the authors combine infrared iris with infrared and visual facial biometrics (multimodal and multisensor combination). By doing this, the risk of a spoof attack (e.g. using a high resolution picture) against the infrared iris is reduced.

2.1.2 Incorporation of Quality Measure

A quality measure of the biometrics is indicative of the classification errors (e.g. the systematic errors, presentation-dependant errors and user-dependant errors [72]) in biometrics authentication. In this part, quality related fusion research is reported. It provides details about how these biometrics qualities can be obtained and used to enhance the biometrics authentication process.

In [73], the quality of the fingerprint is assessed using the NIST Fingerprint Image Software 2 [74] and the speech quality by using the NIST Speech Quality Assurance Algorithm (SPQA) [75]. These quality measures are further combined into a global quality which is used as a weighing parameter for the Weighted Sum fusion rule. In [76], fingerprint images are divided into sub-windows and transformed by Pet Hat's Continuous Wavelet Transformation (CWT) [77]. Then the wavelet coefficients are used as an image quality reference to weight the fingerprint features' Euclidean distance. A fingerprints database with quality labeled by a human expert [78] in the range between 0~2 is used in [79]. It then uses a modified Support Vector Machine (SVM) to do the fusion with inclusion of these quality measures. Biometrics quality can also be calculated by the relative biometrics information entropy between the population's and individual's feature distribution [80].

Probability based quality measures can easily be incorporated into the density based fusion algorithm (e.g. using the product of density or a jointly modeled density). The works in [56] and [81] use coherence-based local quality estimation from [82] and wavelet-based algorithm [83] for fingerprint and iris quality estimations respectively. Two of these quality measures along with relative biometrics scores are then jointly modeled and combined using likelihood ratio fusion.

Cross device matching significantly degrades authentication performance. The work in [84] proposes a score normalisation approach that includes the qualitative device information to solve this problem. The joint density is modeled using each modality's match score, quality measure and the quality cluster information. It is then combined using the Naive Bayes principal. The quality of facial images and fingerprints are assessed using Omniperception SDK [85] and a fingerprint quality assessment algorithm in [82] correspondingly. All these quality measurements and its relative biometrics scores are available in a recent developed database [86], [87].

Quality of the biometrics does not contain information on whether the claimant is an impostor or a genuine user. However it is an indicator for the reliability of the biometrics measurements. Compared to the scores that are obtained from poorer quality biometrics, a score from a higher quality biometrics is more reliable therefore should be given a higher weighting.

The Quality information and multibiometrics are the main information sources to improve the biometrics performance. Aside from this information, some other sources in the literature are further presented in the following section.

2.1.3 Other Information Gain Aids in Biometrics Improvement

Soft biometrics are empirically shown to be capable of improving the performance of conventional hard biometrics [24], [88]. Some of the soft biometrics examples are human height, weight, colour of skin, colour of iris, body fat, gender, age, etc. By using the colour of the iris, the author in [88] achieves an improvement by using soft biometrics that does not request any additional capturing devices. Using body fat and weight which are considered soft biometrics, is a low-cost and easy to understand method to enhance the fingerprint performance and make it more difficult for circumvention [24].

The research work in [89] uses Bayesian Networks to estimate the probability for verification errors. They derive Modality Reliability (MR) using the speech verifier outputs (classified identity and score) and the acoustic environment condition (Signal-to-Noise Ratio). This MR is used to indicate whether the verifier make a reliable decision. Authors in [60] suggest a failure prediction model. It is based on the construction of a learning system using several features extracted from the biometrics scores. By exploiting the classifier space for class-specific information, the face and fingerprint fusion can be enhanced [90]. If the output of the biometrics is well clustered and distinct from other clusters, this information can be used to further extend the separation of different classes.

Norman et al. proposed a Prior Knowledge Incorporation framework (PKI) in [91] which to incorporate additional information sources into the biometrics score. In their work, they utilise Client-dependent F-ratio normalised scores (as proposed in [92]) and margin derived confidence (as proposed in [93]) as additional sources of information. The outputs from the different modalities of biometrics after the PKI are then further combined using a second classifier.

From the works presented in section 2.1 it can be seen that information gain reduces the authentication uncertainties and errors. To increase the information gain through multibiometrics source is the most popular and direct way. Among the multibiometrics scenarios, multimodal biometrics having the lowest correlation between the different sources and hence provides the greatest information gain compared to the rest. Not only the authentication rate can be improved, but most of the limitations of single biometrics can be overcome. Biometrics scores may not be suitable for further processing because of poor quality. Therefore a quality measure can be used to tune the scores to more effectively authenticate a person. Aside from the above, it is possible that there is other useful knowledge underlying among the training biometrics samples. This knowledge can also be exploited and used as additional information to enhance the authentication performance.

With the availability of these biometric information, how to effectively combine or fuse them becomes another great challenge in the biometrics research field. This is reviewed in section 2.2.

2.2 Biometrics Information Fusion

Biometrics fusion approaches reported in the field are firstly reviewed in the operation mode which are presented in session 2.2.1. This includes serial and parallel fusion modes. Thereafter parallel fusion which is a preferred mode is classified into three different levels fusion according to biometrics internal processing stage. This is as shown in fig. 1-2 and detailed in section 2.2.2. Amongst these three levels, measurement level fusion is more commonly applied therefore it is further discussed in section 2.3.

2.2.1 Serial and Parallel Fusion Mode

Fusion can be done in serial or parallel. These modes are also referred to as hierarchical fusion and holistic fusion respectively in [94]. Fig 2-3 illustrates these fusion modes in schematic diagrams. Serial fusion authenticates a person through sequentially assessing the claimant's biometrics as shown in fig. 2-3(a). Individual biometrics module generates a decision and passes it to the next biometrics module or simply terminates the process if a reliable decision is obtained. In this mode, the information is not really "fused" but each biometrics acting as a filter. As depicted in fig. 2-3(b), in contrast to the serial mode, all biometrics outputs are combined simultaneously using a fusion algorithm in parallel mode.

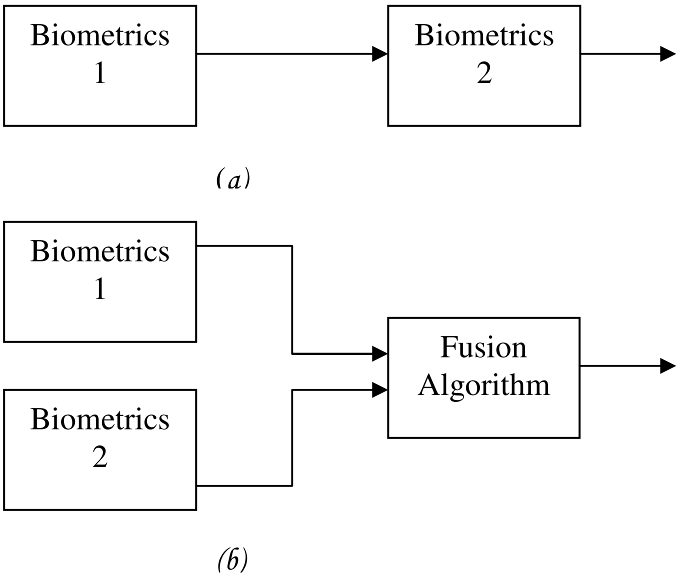


Fig 2-3. Biometrics fusion modes: (a) Serial fusion mode (b) Parallel fusion mode.

An authentication system named "Sequential Selection Multimodal Authentication System" is devised in [1]. The user is requested to present his/her first preference of biometrics. When the biometrics score is sufficiently high to clear the security level, the authentication is accepted. Otherwise, the user has to provide the next preferred biometrics. The author claimed that a higher security level without having to request extraneous information from the user. The research work in [23] defines the pupil and iris parts as "blackeye". They use the dynamic blackeye shape (blackeye shape changes during blinking) as the first biometrics subsystem and the static blackeye shape as a second subsystem. Prior of employment of this sequential combining method, the optimal thresholds for each of the modalities to achieve the best fusion result have to be determined.

In the work presented in [95], the single biometrics matcher with better authentication performance is used in the first step and the less performed matcher is as the next step for serial combination. However, in contrast to this work, [94] uses the less performed matcher prior of the one with better authentication performance. There is no reason given for these arrangements in these literature. In [94], the authors show serial fusion is less effective than the Sum and the Product rules which are parallel fusion rules. Different experimental result is obtained in [95], the serial fusion method outperforms the simple Sum rule (parallel fusion). However, this might be caused by the fact that unnormalised scores are used in the experiments in [95].

The serial mode may request less evidence from the user to reduce the authentication time and increases the user friendliness of the system. When reliable decision can be made at the first stage, it is simply a single source biometrics. If not, the system forfeits previous biometrics and seeks for the next evidence. Although reliable decision cannot be made, there is still useful information contained in these forfeited biometrics. The serial mode behaves likes a filter rather than a fusion engine and hence results in information wastage. In parallel fusion mode, gathering all the biometrics information and combining them simultaneously is a more efficient way of using the biometrics information. Therefore the parallel fusion approach is used more often in the literature.

According to the biometrics processing stage as illustrated in fig.1-2, parallel fusion methods are broadly classified into feature level, measurement level and decision level fusion [96]. These methods are reviewed in the section 2.2.2.

2.2.2 Three Different Levels of Fusion

As shown in fig. 2-4, in the context of verification, to accept or to reject a claimant referring to the biometrics is a process of information reduction.

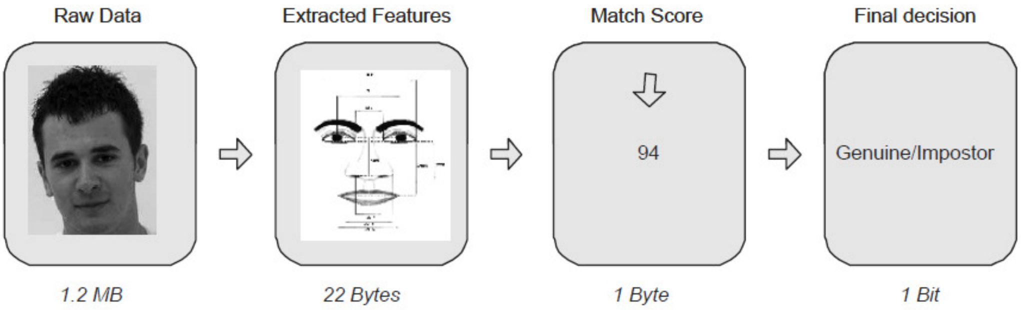


Fig 2-4. The contents of processing biometrics in verification mode (adapted from [97])

Combination at earlier stage is desired because of the richer available information. However to combine the biometrics at the feature level is difficult especially when the features are different (e.g. to combine the minutiae of the fingerprint to the eigenface coefficient.) Although it is much easier to do the fusion at the decision level, because only one bit information is involved, this information is too limited for a significant fusion improvement. All biometrics output matching score containing more useful information than a binary decision. As a result, the score level biometrics fusion is more popular than the two others in the field of fusion research.

Some of the feature level and decision level fusion research works are reviewed in session 2.2.2.1 and 2.2.2.2 respectively. The score level fusion researches are more widely explored than the feature level and decision level fusion. Therefore it is reviewed and discussed separately in session 2.2.3.

2.2.2.1 Feature Level Biometrics Fusion

Feature level fusion is to concatenate the extracted features from multiple biometrics sources. In some of the work, the feature combination is done at sensor level or image level.

Khuwaja uses compact Learning Vector Quantization (LVQ) neural networks to combine face and fingerprint images and get a blurred image called Merged Pattern (MP) [98]. It contains both biometrics features and is considered to be more discriminative. The MP features are then extracted by an adaptive artificial neural network. In this work, the achieved 100% identification rate using the Olivetti Research Laboratory (ORL) database is considered as the major achievement. Concatenation of ear and face features in [99] achieves rank one identification rate at 90.9% where ear and face provide 71.6% and 70.5% respectively. Because of physiological relationship, ear and profile face are combined at feature level in [100]. Kernel Canonical Correlation Analysis (KCCA) is a feature fusion method. It is to extract the non-linear associated features of ear and face and classify them using minimum distance classifiers. In contrast to 90.8% and 77.6% recognition rate for ear and profile face respectively, the fusion result is 98.7%.

2.2.2.2 Decision Level Biometrics Fusion

Since the biometrics verification decision is only one bit of information (i.e. to accept or to reject), very limited information is available for fusion at this level. Therefore its performance is normally not comparable to the feature and score level fusion. However such fusion method is commonly applied in identification mode to reduce the processing time.

The AND, the OR rule and majority voting [101] are commonly used for decision level fusion. AND and OR rules are applied in [24] to combine fingerprint and body weight. From their reported results, it is surprising that AND rule, which is one bit information fusion, achieves 1.45% Total Error Rate (TER, i.e. the summation of FAR and FRR) which outperforms the more complicated methods: Sum rule, Multi Layer Perceptron and Support Vector Machine, which obtained 2.51%, 1.69% and 3.28% TER respectively.

Majority voting is, for example, implemented in BioID [48], a commercial biometrics application. This system integrates speech, face and lip movement biometrics. The system assigns identity to a claimant if 2 or 3 (under ‘2 out of 3’ or ‘3 out of 3’ scheme respectively) of his/her biometrics passed the relative thresholds set in advance. The fusion result is not reported. However, to achieve the best fusion result using this method, the preset thresholds have to be chosen carefully. The work in [102] finds the optimal thresholds using individuals biometrics’ ROC and then applies these thresholds prior of using AND or OR rule for decision fusion. This fusion method is claimed to be more robust to outliers and insensitive to the deviation between the training and testing scores.

A biometrics identification system generates a rank list as decision. Lin and Anil apply decision level fusion in an identification system [103]. In their work, the top *n* possible identities are established by the face module and then passed to the fingerprint module to create the final list. For identification, the system has to compare the biometrics template with all the templates in the database, which is very time consuming. The amount of template comparison and processing time is drastically reduced using such decision level fusion. Unlike this fusion method, Djamel and Abbes simultaneously combine the top five rank determined by different biometrics using Borda count [65]. Borda count assigns a specific score to the possible identities according to their obtained rank and determines the identity based on the accumulated score. In [104], the ranks assigned to survivors (users) who pass the thresholds are directly summed. By doing this, 99% authentication rate is achieved.

2.2.3 Measurement Level Fusion Approaches

Measurement level is the most popular biometrics fusion level. Biometrics generates confidence value or score to authenticate a person. Such information is homogenous and accessible, therefore majority of biometrics fusion research concentrates on this type of fusion. It can be further subdivided into three different types: (a) rule based fusion (b) classification based fusion (c) density based fusion.

2.2.3.1 Rule Based Fusion

Rule based fusion combines the biometrics score by using a fixed rule, e.g. Sum, Min or Max rules. The main advantages of such combination are that there is no training session required, and the method is very efficient in processing time and conceptually simple. However each biometrics module might have different measurement scale. For the biometrics to be effectively combined using a fusion rule, score normalisation presents the greatest challenge. Different score normalisation techniques are proposed and compared. These proposed methods and comparisons are available in [44], [58], [61], [105], [106].

Table 2.1 presents a summary of score normalisation techniques from the literatures. An effective score normalisation algorithm should be less sensitive to outliers. Whether the normalised scores have to be in a common range or retaining its original distribution depends on the applied fusion rule.

In the table, S'_1 is the normalised score and S_i is the original score. Min-max normalisation is the simplest algorithm that only involves finding the minimum (*min*) and maximum (*max*) scores generated by a specific biometrics matcher. These parameters are sometimes directly available from the biometric application vendor so no training session is required to find these parameters. To use Z-score normalisation, the score distribution's mean (μ) and standard deviation (σ) must be found in advance. Such prior knowledge can only be estimated from a training set and it is sensitive to outlier. Tanh normalisation is also sensitive to outlier. However Jain shows that using the Hampel influence function [107] able to greatly reduce this problem [44]. Median and Median Absolute Deviation (Median and MAD), which uses the median of the biometrics score distribution is less affected by outliers than the Z-score and Tanh method. However the risk of this normalisation is that once the normalised score is a Gaussian distribution, it cannot be normalised effectively. Double Sigmoid normalisation requires careful tuning of the t, r_1 and r_2 to choose the region with linear mapping characteristic. t is the reference point, r_1 and r_2 denote the left and right edges of the linear mapping region. The scores outside this linear mapping region are transformed non-linearly to increase the separation of genuine user and impostor score distributions.

		Retain Distribution ?	Sensitive to outliers ?	Map into common range?
1	Min-max $S'_i = \frac{S_i - \min}{\max - \min}$	Yes	Yes	Yes
2	Z-score $S'_i = \frac{S_i - \mu}{\sigma}$	No	Yes	No
3	Tanh $S'_i = \frac{1}{2}\{\tanh\left(0.01\left(\frac{S_k - \mu}{\sigma}\right)\right) + 1\}$	No	No	Yes
4	Median and Median Absolute Deviation (Median and MAD) $S'_i = \frac{S_i - median}{MAD}$ $MAD=median(\vert S_i\text{-}median\vert)$	No	No	No
5	Double Sigmoid $S'_i = \begin{cases} \frac{1}{1+\exp(-2((S_k - t)/r_1))}, & \text{if } S_k < t, \\ \frac{1}{1+\exp(-2((S_k - t)/r_2))}, & \text{otherwise} \end{cases}$	No	No	Yes

Table 2-1. Commonly used score normalisation algorithms.

[105] proposes to use different normalisation techniques for different regions. In this so called Two-Quadratics method (QQ), two quadratic segments are setup in the mapping function. This is further modified into three segments normalisation in Quadratic-Line-Quadratic method (QLQ). It also includes two quadratic segments but leaves certain region unnormalised. In [106], Four-Segments-Piecewise-Linear (FSPL) and Linear-Tanh-Linear (LTL) are proposed. The authors use the linear function (for FSPL) and non-linear function (for LTL) in the overlap region for score normalisation. The comparison between these methods with other commonly used normalisation techniques, by applying the Weighted Sum rule, indicates that QLQ and LTL perform better than the rest.

In the following, different rules proposed for fusion are presented. A common theoretical framework for combining classifiers is developed in [108]. Commonly used combination schemes, e.g. Product, Sum, Min, Max, Median rules and Majority Voting are compared. The result shows that the Sum rule outperforms other schemes. Through sensitivity analysis, Kittler concludes that the superior performance of the Sum rule is due its resilience to the estimation errors. The effectiveness of the Sum rule is further justified by Ross and Jain’s research in [43]. This simple fusion rule outperforms the complicated Decision Trees and the Linear Discriminant Analysis fusion methods. However, the Max rule is reported in [65] and [109] to outperform the Sum rule. Nevertheless, the Sum rule has been widely employed in the literature. Different Weighted Sum rules have been proposed and they are summarised in the table 2-2.

	Weighting Scheme	Applied in
1	Equal Weighted Same weights are assigned to all biometrics. This weighting scheme does not use any parameter.	[16], [67], [109], [110], [111]
2	Equal Error Rate Weighted Each biometrics Equal Error Rate, EER, is used to weight their contributions. Biometrics with higher EER is assigned with lower weight.	[106]
3	D-Prime Weighted Each biometrics genuine and impostor scores separation, *d’,* is used to weight their contributions. Biometrics with higher *d’ is* assigned with higher weight.	[106]
4	Quality Weighted The biometrics with better quality is assigned with higher weight.	[73], [76]
5	FAR/FRR Weighted False Acceptance Rate (FAR) and False Rejection Rate (FRR) are threshold-dependent, therefore a training section is required for different operating point to find these parameters. The biometrics with lower FAR/FRR is assigned with higher weight.	[106]
6	Rank Weighted This scheme is only applicable in the identification mode. A score with higher rank is assigned with higher weight for combination.	[16], [59], [64]

7	Exhaustively Searching The best weights used to achieve optimal fusion performance are exhaustively searched. However this searching has to be repeated for different operating points.	[112], [113], [114]

Table 2-2. Commonly used weighting schemes for Sum rule.

2.2.3.2 Classification Based Fusion

The scores from different biometrics sources can be treated as feature vectors. The fusion therefore is viewed as a classification problem. A classifier is used to construct a separation boundary between the genuine user and impostor in a verification system. The classifier used for this purpose includes K Nearest Neighbours, Decision Trees, Neural Networks, Support Vector Machine and Logistic Regression [115].

No advance training is required for K Nearest Neighbours. By referring to the distances from the tested biometric sample to *k* nearest reference points, the sample is then assigned to the category that has the majority of nearest neighbours. Although no training section being required, the distances from the tested sample to all the reference points have to be found. This leads to a very time-consuming fusion process. To achieve a better verification rate, this method was modified in [116], [117], [118], [119].

A Decision Tree categorises the biometric samples according to a series of tests on a specific attribute of the data. These hierarchical tests lead to a particular class. Each of the tested attributes is found based on maximising the information gain at the particular node. This method has the advantage that it provides direct insight into the predictive structure [120]. However, it is very sensitive to small changes in the dataset

[121]. The well-known C4.5 classifier is devised by Quinlan [122]. This is the most widely employed Decision Trees algorithm where the related fusion works are available in [123], [124], [125].

Another classification method that can be used for biometrics fusion is Artificial Neural Network (ANN). An ANN is composed of many artificial neurons that are interlinked by synaptic connections. Each of these connections is associated with a specific weight. To train an ANN the weights have to be adjusted according to the error between the predicted and actual outputs. This process is performed mostly by a back-propagation algorithm. These weights and the relative biometric scores are then used by a function to transform this information into a meaningful output. The Multilayer Perceptron (MLP) [124] and Radial Basis Function [66], [126] are two commonly used transform function in the literature. MLP uses a linear transform function whereas the RBF uses a non-linear one. In [66], it is commented that RBF is preferred because their experiment shows better fusion performance than MLP. Also because of the RBF kernel is able to learn from both the positive and negative samples (genuine user and impostor samples).

K Nearest Neighbours, Decision Tree and ANN operating thresholds are not adjustable because their output is not a score but a class label, which is threshold independent. Although Support Vector Machines and Discriminant Analysis operating thresholds are also non-adjustable, these algorithms can be modified to generate a confidence value but not a class label. So a threshold can be used to classify these biometrics samples associated with confidence value.

In the biometrics fusion problem for verification (two class classification problem), given a set of training samples, a Support Vector Machine constructs a separation boundary so the distance from it to the nearest data points which are termed as support vector on each side is maximised. Such a classifier is a linear classifier. A non-linear Support Vector Machine can be built by applying this algorithm in a transformed feature space [127]. This feature space can be created through a kernel function to project the samples to higher dimensional space. Polynomial and Radial Basis Function kernels are employed in [124] for multimodal biometrics fusion problems. In [124], significant fusion performance difference is obtained by using the Polynomial and Gaussian kernel. Therefore it can be said to choose a suitable kernel function is the main challenge of this fusion approach. The SVM has been reported to have the best fusion performance in [128], [129], [130], [131] compared to the methods including decision level fusion approach, Sum rule, K Nearest Neighbours, Decision Trees and ANN. Instead of using the output class label by SVM, the signed distance from the tested sample to the Support Vector Machine's separating surface can be used as output score [132].

Logistic Regression uses the logistic function to transform the weighted biometric scores into a value between 0 and 1. The input of the logistic function, the variable *z*, is a measure of the total contribution of all biometrics sources. The weights of biometrics or the regression coefficients are usually found using Maximum Likelihood Estimation (MLE). This is an iterative process which is similar to the back-propagation in ANN. Logistic regression is applied in [43], [123], [124], [126], [133] to solve the biometrics fusion problem. In the comparative study in [134], Logistic Regression is evaluated as one of the most effective score level fusion techniques among three different rule based fusion categories.

2.2.3.3 Density Based Fusion

Density based fusion first transforms the scores of biometrics into probability densities. These probabilities can then easily be combined using the product rule. Unlike the scores used in rule based fusion, these densities can be applied directly without normalisation. Furthermore, provided that the underlying densities are known, the optimal fusion performance is directly achieved. Since this method is probability based, additional information (e.g. the probability based quality) that aids the fusion process can also be incorporated without having to modify the fusion algorithm. Some individuals might not possess certain biometrics or its measurements are not reliable. This causes the non-density based fusion algorithm cannot be applied because of not sufficient input is provided. This missing data problem can also be easily solved in this fusion method.

Different attempts have been tried to improve the authentication rate using the density based fusion. Dass et. al. consider the biometrics score distributions associated with discrete and continuous components [55]. Their algorithm detects and removes discrete components before modeling the biometrics' marginal continuous density. The mixture of continuous density and discrete components are used in a product rule to do the fusion. Aside from this method, they also utilise the copula function for joint densities estimation. The likelihood ratio is then used to categorise the user. They successfully demonstrate that the proposed approaches outperform the single biometric.

The work in [73] shows that biometrics quality can be easily incorporated for density based fusion. They directly use the joint density modeling conditioned on the identity (genuine user and impostor) and biometrics quality. This is modeled by using Gaussian, Gamma, Log-normal or beta distribution. These joint densities are applied in their developed Bayesian Belief Network (BBN) which is shown to outperform the Sum fusion rule.

To achieve a higher authentication performance, the work in [135] models the genuine and impostor score distributions by adaptively using both the user dependent and user independent parameters. The model parameters which are estimated from the entire samples (global estimation as the user independent parameter) and a specific user samples (local estimation as the user dependant parameter) are adaptively used in their model. They achieve a biometrics fusion improvement of 80% and 55% compared to the non-adapted density fusion method for small and large training set respectively.

Nandakumar demonstrates in [57] that the problem of missing biometrics can be solved without having to modify the density based fusion method. They use the likelihood ratio as the input to the product rule. By assigning unity value as the likelihood ratio to the missing biometrics, this problem is catered for.

The Gaussian Mixture Model (GMM) is demonstrated to be an effective model to estimate the genuine user and impostor score densities and is easy to implement in [56]. Their work consistently achieves good fusion performance comparable to the Sum rule and Support Vector Machine. This performance is further enhanced by incorporating the probability based quality information into the density [81]. The component number is the only parameter required for the model. Due to its effectiveness and requiring less parameter tuning, this modeling tool has been widely employed, e.g. in [129], [131], [136], [137]. However, choosing an appropriate component number for the model is challenging. The authors of [131] tune this parameter manually on the training samples. In [56], this parameter is searched automatically by using the state-of-the-art GMM fitting algorithm in [138].

2.2.3.4 Selecting A Fusion Approach

Aside from the fusion performance, there are other considerations in making a choice amongst the rules based, classification based and density based fusion approaches. These concerns are presented in the following

(a) Availability of resource:

Training based fusion methods, either in classification or density based categories, normally produce better authentication rates than the non-training rule based fusion. But such training set may not be available or large training sets have to be collected for a reliable prior knowledge exploration. Furthermore one has to consider the availability of these complicated training based algorithms.

(b) Advantages of the approach:

Although rule based fusion is less efficient, such a fusion method is the conceptually simplest, fast and does not use a specific training algorithm. The density based fusion is preferred because of its ability to cope with the missing value problem and to incorporate additional information without having to modify its fusion algorithm.

Furthermore, it does not require score normalisation and it is able to achieve the optimal performance at any operating points directly, provided the underlying densities are known.

(c) System requirements:

A biometrics system has to be threshold-adjustable to accommodate different security levels. Therefore, some of the classification based algorithms that generate class label are not suitable for biometrics fusion or have to be modified. A training based method, especially when used in large scale application, the training process has to be efficient. Such efficiency is necessary for the fusion algorithm to accommodate new enrolments and the variation of the biometrics of the population instantly.

2.3 Summary

Single biometrics after many years of development has come to saturation to meet the desired performance requirements for larger population [46]. Generally, this saturation is governed by the limitations of authentication ability and system usability. The related works done to overcome these limitations are broadly addressing two issues: to provide additional information and to more effectively combine this information. The above review can be summarised as follows.

1. Using multibiometrics is the most effective and direct way to increase the single biometrics authentication performance. For example, a multimodal biometrics system can easily outperform a single biometric. Amongst the different multibiometrics scenarios, multimodal biometrics is preferred. This is due to the fact that multimodal biometrics is inherently different, so more information gain

can be obtained compared to other scenarios. At the same time, multimodal biometrics enhances the system usability. It further provides alternative biometrics entry option. As a result, the system is more robust to adapt to different operating ambiance (e.g. dim light or noisy conditions) and to cover larger population. Providing different modalities of biometrics also makes the spoof attack more difficult.

2. Multibiometrics information are combined with other information such as biometrics quality, soft biometrics and other biometrics score underlying knowledge, to further improve the authentication performance. Considering that the verification errors arise from the overlap region, this region related information might be informative to further aid fusion. However, using this region information for biometrics score level fusion is not found in the literature.

3. Effective combination of the biometrics information also plays a key role for biometrics authentication improvement. Parallel fusion mode is preferred. This is because the information can be "fully utilised" before making a reliable decision. Biometrics fusion is easier to be dealt with under verification mode (two class problem) rather than the identification mode (multi class problem). As a result, it is common the fusion method to be firstly developed for verification purpose. It is then extended to the more complicated identification problem, e.g. the work presented in [57].

4. The biometrics information fusion attempts have been tried on feature level, measurement level and decision level. Vast majority of works focus on

measurement level fusion. This is because of the balance between the complexity and richness of information. Furthermore, the biometrics measurement sample for the fusion method evaluation is easier to obtain, e.g. from vendor matcher systems [46].

5. Three different types of measurement level fusion, including rule based, classification based and density based fusion are available. Each of these methods has different features to accommodate in different biometrics fusion requirements and circumstances. The rule based fusion, e.g. the Sum rule, regardless of being simple, has been reported to outperform the complicated training based algorithms (Decision Trees and Linear Discriminant Analysis) [43]. It is efficient, effective and does not require training session and additional resource. Most of the classification based methods' performances are reported on a single operating point. This is due to the fact that a classification based method outputs the class label but not a measurement. Therefore, it is not threshold adjustable to accommodate different security levels. Further modification is needed to make it threshold adjustable. Both the classification based and density based fusion methods require sufficient training sets to find the reliable parameter value or to fit in the density model. These methods' performances always rely on the training sample size and quality. A density based method also has the feature of directly achieving the optimal fusion performance at any operating point. This is achieved without parameter tuning. However, this optimal fusion performance can only be achieved provided that the density is estimated accurately. Additional density based information can be incorporated in this fusion method and the missing of

biometrics problem can be solved using this method, without any ad-hoc modification.

2.4 Research Questions Arising

The overlap region contained in the biometrics verification score as described in section 1.4 might be informative in further improving the biometrics authentication performance and/or usability. Using such information has not been found in the biometrics score level fusion research community. Therefore, this research is directed in exploration of the overlap region information and the benefits of implementing such information to the existing fusion approaches. From the summaries in section 2.3, the measurement level fusion methods from different categories have different features. Therefore, this exploration is conducted to approaches from different categories. The single biometrics limitations can be more comprehensively solved by combining multiple modalities of biometrics. Thus this research is conducted to combine the multimodal biometrics at score level, under the parallel fusion and the verification mode. From this research direction, several gaps in the literature are identified and this work is shaped to address these gaps. These gaps are listed below.

1. Extensive evaluation of the conventional multimodal biometrics fusion approaches from different categories has to be established as a baseline for this research work. As well as, to compare them over a wide range of experiments, since different fusion strategies have been claimed outperforming the others. Such comparison is available in the literature. However, they are not extensive enough. These works either do not cover all three different approaches or not tested over a wide range of multimodal biometrics experiments. For example, the works in [43],

[108], [123], [129] use only rule based and classification based methods for comparison. In [56], even though all three different categories' approaches are covered, their experiments are tested on multibiometrics combination only (mixing of multimodal and multiunit biometrics). The work in [134] includes the multimodal biometrics evaluation and various approaches from three different categories. Nevertheless, there are only three multimodal biometrics experiments tested (the rest are multiunit biometrics). Furthermore, their work is only tested using a single database that includes two modalities of biometrics. Therefore, it is clear that there is no extensive score level fusion algorithms evaluation specifically for multimodal biometrics. Amongst the approaches claimed with top performance, whether there is a specific method outperforming all extensively, has remained unanswered. Furthermore, whether a fusion strategy with top performance is the most appropriate one for practical implementation also has not been discussed previously.

2. The rule based fusion research has been focused on score normalisation techniques and Weighted Sum fusion rule in the literature (refer to section 2.2.3.1). These works successfully demonstrate that the authentication rate can be improved by these techniques. Different fusion rules (e.g. Sum and Max rules) have been reported outperforming other compared methods. Aside from using these combination strategies separately, the improvement can also be achieved by incorporating a selection scheme on these combination strategies. Such a fusion approach selection mechanism is not common in the biometrics score level fusion research. The only work can be found in [62] where the authors formulate a selection mechanism aimed to further separate the genuine and impostor scores.

Their mechanism selects the Max and Min rules based on the estimated error rate. This is further modified by introducing classifiers (K-NN, Quadratic Bayes and Parzen Windows) to make the selection between Max and Min rule [139]. The rule based fusion is well known for its simplicity. However, further complexity is introduced in their work by using the mechanism. Different score regions (overlap and non-overlap regions), which can be easily identified, exhibit different confidence level in discriminating a claimant. Therefore, the selection mechanism might be conducted from this perspective. The question of whether applying the fusion method selection mechanism based on the score region is feasible has not been answered in the literature.

3. The Weighted Sum rule is the most effective rule based fusion method. The Equal Weighted Sum rule has been commonly used in the literatures [16], [67], [109], [110], [111]. However, due to the fact that different biometrics always has different authentication ability, whether equal weighting is a good practice is not answered in the literature. Different weighting helps to further improve the Sum rule fusion performance. One of the attempts is to use the threshold-dependant parameter. For example, using the FAR and FRR for weighting [106] or exhaustive searching for the optimal weights [112], [113], [114]. However, these Weighted Sum rules require searching for the new weighting values whenever different operating threshold is required. This is to say that the performance is not maximised for all operating thresholds. Some of the threshold-independent parameters, e.g. the *d'* and EER, are used in [106]. It is usual to use *d'* and EER to evaluate individual biometrics performance. Nevertheless, whether these parameters can be used to produce consistent Weighted Sum rule fusion

performance (for most of the operating points) is also remained unanswered. Furthermore, since the overlap region is where the errors arise, if reducing such a region is a way to enable maximisation on performance has to be addressed.

4. In the density based fusion method, the authentication performance depends on the accurate modeling of the underlying biometrics scores density distribution. The Gaussian Mixture Model has been widely employed for this purpose (e.g. the works in [56], [129], [131], [136], [137]) because of its effectiveness in modeling and less parameter to tune. However, choosing the component number is challenging because this causes direct impact to the fusion performance. This parameter is manually tuned on the training set [131] or is searched by specific density fitting algorithm [56] in the literature. In another word, to achieve optimal density based fusion performance, these works focus in the question of "How the exact component number can be obtained to boost the biometrics fusion performance?". Considering the resource availability, the inaccuracy of manual tuning and the additional searching times is required, the focus can be transferred. A new research question can be asked: "How the impact of inaccurate assignment of component number to the biometrics fusion performance can be reduced?". There are no relevant works found in the biometrics score level fusion research community. The attempt to answer this question aims to address this gap.

3 COMPARATIVE EVALUATION OF MULTIMODAL BIOMETRICS SCORE LEVEL FUSION APPROACHES

The aim of this chapter is to provide a comparison of different categories of fusion techniques on large scale databases. This comparison is based on the fusion accuracy that is obtained through bimodal biometrics fusion. Other factors associated with this performance, such as its limitations, training and processing time and availability of resources are also jointly considered in the work. Such comparison provides comprehensive guidance for selecting an appropriate strategy for a particular application. Moreover, it provides a baseline for the proposed fusion strategies presented in chapters 5 and 6.

3.1 Introduction

A significant amount of score level fusion techniques have been proposed in recent research. Most of these studies are focusing on increasing the fusion accuracy. However, the reported performances from different attempts are not directly comparable. This is because the databases used for evaluations are different in size and quality, which are factors having a direct impact on performance [142], [143]. Therefore, it is difficult for one to choose the best fusion method. Furthermore, some assumptions have to be made before the reported performance can be achieved. Also, for one to choose the most appropriate fusion algorithm, the fusion accuracy is not the only criterion [144]. Some other relevant factors that are listed below are associated with the achieved performance and have to be considered as well:

1) Processing or training time of the fusion strategy,
2) Ease of implementation of the algorithm,
3) Resources availability, such as training data and specific algorithm,
4) Fusion algorithm's robustness against the training data variation,
5) Designing and calibrating of the algorithm.

Therefore, there is a need to evaluate and compare the conventional state-of-the-art fusion strategies from a comprehensive perspective on a common database. Some comparisons can be found in the following literatures:

- Kittler develops a common framework to combine different classifiers via a wide range of strategies: Sum, Max, Min, Median and Majority Voting [108]. Experimental results of combining frontal face, profile face and voice reveal that Sum rule outperforms other combination scheme. Kittler concludes the robustness of the Sum rule is due to its sensitivity to estimation error.

- Ross and Jain combine face, hand and fingerprint biometrics using Sum rule, linear discriminant classifier and decision tree [43]. They "surprisingly" find Sum rule outperforming the complicated linear discriminant classifier and decision tree that are learning based.

- Fierrez-Aguilar et. al. find the linear SVM and logistic regression yield the same best results over a wide range of parametric and non-parametric fusion methods in combining face, signature and fingerprint biometrics [129].

- Verlinde compares fusion performance by combining profile face, frontal face and speech biometrics [123]. Logistic regression is reported to outperform K-NN and decision tree based methods.

- Nandakumar et. al. transform the multibiometrics scores into joint densities by using a GMM for density estimation [56]. They use the likelihood ratio as the fusion score. Their method is reported to outperform or comparable to SVM for two different databases.

- Eight fusion techniques from the literature are chosen for comparison based on the reported performance in [134]. The comparison is established on the fingerprint and face biometrics fusion. The work concludes that the product of likelihood ratio and logistic regressions is highly effective.

In the following session, several fusion strategies that have been commonly reported effective for fusion are introduced. They are fusion algorithms chosen from the comparative works mentioned above.

3.2 Compared Fusion Schemes from Three Different Categories

The commonly reported best fusion strategies from three different categories: rule based, classification based and density based fusion methods are compared. Based on the literature review in the previous section, the compared methods are Sum rule from the rule based category, Support Vector Machine and Logistic Regression from the classification based category and Likelihood Ratio based fusion from the density based category. Marginal and joint densities are used in Likelihood Ratio based fusion.

These densities are estimated by a Gaussian Mixture Model (GMM) and its relevant component number is determined by a state-of-the-art algorithm. This algorithm automatically estimates the number of component and the component parameters. Since Max rule is reported to perform better than the Sum rule [65], [109], it is included in the comparison. Min-max score normalisation is chosen to retain the original score distribution. Brief details of these fusion methods are given in sections (a) ~ (f). Here, S_{fi} refers to the fused score and $S'_{I,k}$ is the user *i*'s normalised score that is generated by matchers *k*, $S_{i,k}$ is the raw biometrics score (without normalisation) and the *K* is the total number of matchers.

(a) Sum rule (SUM)

K modalities' biometrics scores are added after the scores are normalised. The Equal Weighted Sum rule is the most popular Weighted Sum rule. It is as shown below:

$$S_{fi} = \sum_{k=1}^{K} \frac{1}{K} \times S'_{i,k} \ , \forall i \qquad (3.1)$$

(b) Max rule (MAX)

The maximum score among the multimodal biometrics scores is chosen as the fusion score. For effective comparison, raw biometrics score has to be normalised in advance.

$$S_{fi} = \max(S'_{i,1}, S'_{i,2}, ..., S'_{i,K}) \, , \forall i \qquad (3.2)$$

(c) Support Vector Machine (SVM)

As shown in fig. 3-1, by viewing multimodal biometric scores as a set of vectors in n-dimensional score space, SVM constructs a separating hyperplane. Such hyperplane is constructed so that the distance from this hyperplane to the nearest data points (the support vectors) on both sides is maximised. Vapnik proved that maximising of this distance minimises the generalised classification error [145]. For non-separable samples, a kernel function can be used to project the samples to a higher dimensional score space and to construct the separation hyperplane in that space. Some commonly used kernel functions are Polynomial function, Radial Basis Function and Hyperbolic Tangent Function. Vapnik further suggests a soft margin to allow the existence of mislabeled samples. By doing this the samples can be classified as less error as possible while the maximum distance between the separating hyperplane and nearest support vectors (parallel hyperplanes) can be maintained. In this work, a linear kernel function is used and the separating hyperplane is found using the Sequential Minimal Optimisation method (SMO). To make this method to be threshold adjustable, it was modified to produce a fusion score based on the proximity of the test sample to the separating hyperplane [130].

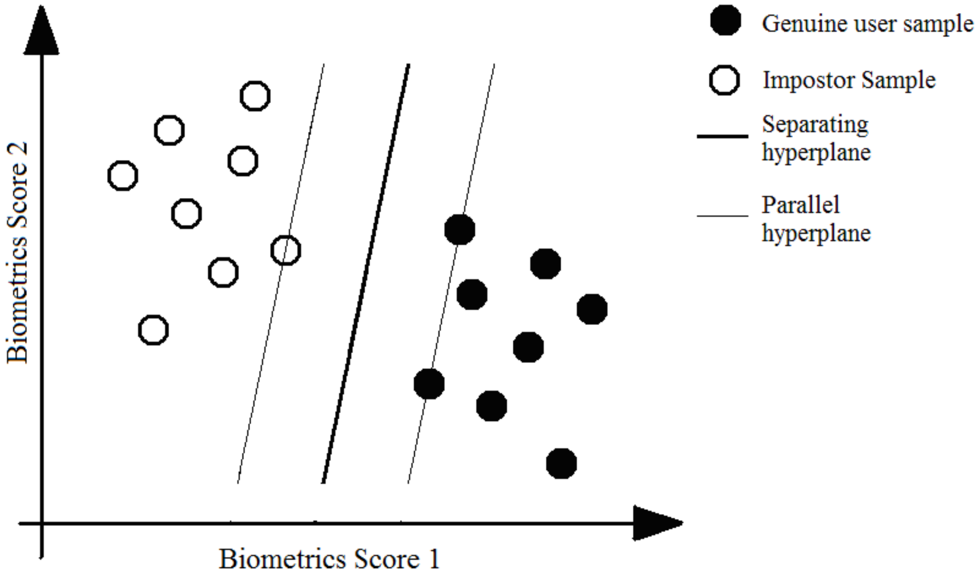


Fig 3-1. Support Vectors Machine schematic diagram.

(d) Logistic Regression (LREG)

Fig. 3-2 illustrates the logistic function *f(z)*. The variable *z* is called logit. It is a measure of the total contribution of different biometrics scores based on the training samples. Their contributions are weighted by the regression coefficients as shown in (3.4). *f(z)* in (3.3) transforms these combined score into the probability values between 0 and 1.

$$f(z) = \frac{1}{1+e^{-z}} \tag{3.3}$$

$$z = \beta_0 + \beta_1 S_{i,1} + \beta_2 S_{i,2} + ... + \beta_K S_{i,K} \tag{3.4}$$

For the expression of *z*, β_0 is the intercept and β_1, β_2, …, β_K are the regression coefficients. These parameters are estimated from the training samples by using the Maximum Likelihood Estimation algorithm (MLE), which is based on the Iteratively Re-weighted Least Squares method (IRLS).

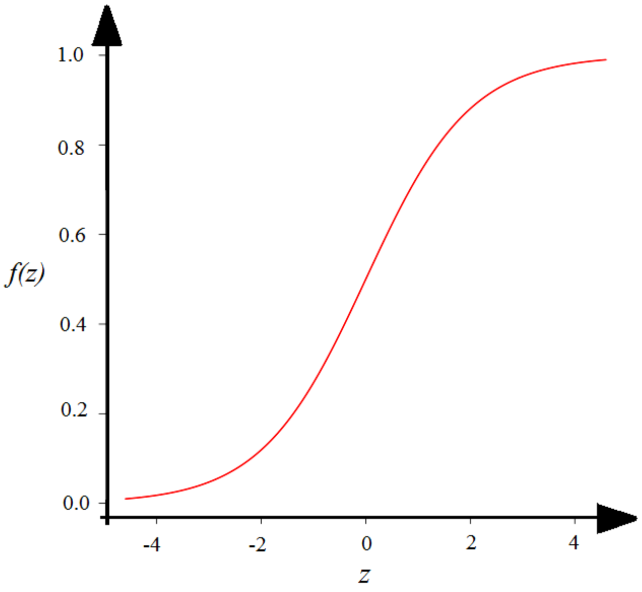


Fig 3-2. Logistic function used in Logistic Regression Analysis.

(e) Likelihood Ratio Based Fusion (JLLR)

The term $f_g(x)/f_i(x)$ in (3.5) is referred to as the likelihood ratio. The logarithm of this likelihood ratio is taken as the fusion score S_{fi}. In this method, the multimodal biometrics scores joint densities, $f(S_{i,1}, S_{i,2}, \ldots, S_{i,K})$ of the impostor and genuine user are estimated using GMM whereas the component numbers are determined by a fitting algorithm in [138].

$$S_{fi} = \log\left[\frac{f_g\left(S_{i,1}, S_{i,2,}\ldots, S_{i,K}\right)}{f_i\left(S_{i,1}, S_{i,2}, \ldots, S_{i,K}\right)}\right] \tag{3.5}$$

(f) Product of Likelihood Ratio Fusion (MLLR)

In contrast to the JLLR algorithm (e) which uses the joint densities, here the marginal density $f(S_{i,k})$ of each biometrics is modeled. The likelihood ratio of each matcher is then multiplied and the logarithm of the product is used as the fusion score. Again, GMM and the fitting algorithm mentioned in (e) are used to estimate the marginal densities.

$$S_{fi} = \log \prod_{k=1}^{K} \frac{f_g(S_{i,k})}{f_i(S_{i,k})} \tag{3.6}$$

The comparison is conducted on two publicly available truly multimodal databases: NIST-BSSR1 [53] and Xm2vts databases [146]. Cross validation over the matchers with different modality in these databases is carried out. 4 and 15 bimodal biometrics fusion experiments are conducted using NIST-BSSR1 and Xm2vts correspondingly.

3.3 Multimodal Biometrics Score Set Databases

Two publicly available databases are used throughout the experiments in this thesis. These are the NIST-BSSR1 multimodal database [53] and the Xm2vts benchmark database [146]. A large number of samples is needed for an evaluation of fusion methods, but it is difficult and time consuming to collect biometrics samples from large populations. Therefore research is often based on the chimerical assumption, i.e. to use chimeric users whose biometrics are constructed by combining multimodal biometrics from different individuals [59], [61], [73], [105], [116]. However, according to Norman's experimental results [147], such practice is questionable. Therefore both databases chosen for this work are truly multimodal. A genuine user score can only be obtained through true sample matching whilst an impostor score can be obtained through cross sample matching. This results unbalanced training set and poses challenge when the biometrics score fusion is considered as a classification problem [45]. However in this work, the proposed methods are in density based and rule based fusion categories. The following section provides the details about these databases.

3.3.1 NIST-BSSR1 Multimodal Biometrics Score Database

The NIST-BSSR1 multimodal biometrics database comprises three matching score datasets. Only Set 1 is used because this is the only truly multimodal database. Set 1 is based on faces and fingerprints from 517 individuals, collected using two commercial facial matchers and one freely available fingerprint recognition system. Unfortunately, the details of the matchers are not provided by the authors. Each of the 517 individuals' left index (Fli) and right index (Fri) fingerprint is verified by the fingerprint matcher whereas their facial images are verified by the facial matchers referred to as Fc and Fg. Through cross validation, all the enrolled 517 users are verified using their own templates to generate genuine user score and using the rest 516 users' templates to generate impostor scores). Therefore there are 517 (517*1) genuine user scores and 266,772 (517*516) impostor scores in total. Fig. 3.3 depicts the matching score distributions. The ROC curve indicating the performance of the matching constructed by using all available scores is shown in fig. 3.4.

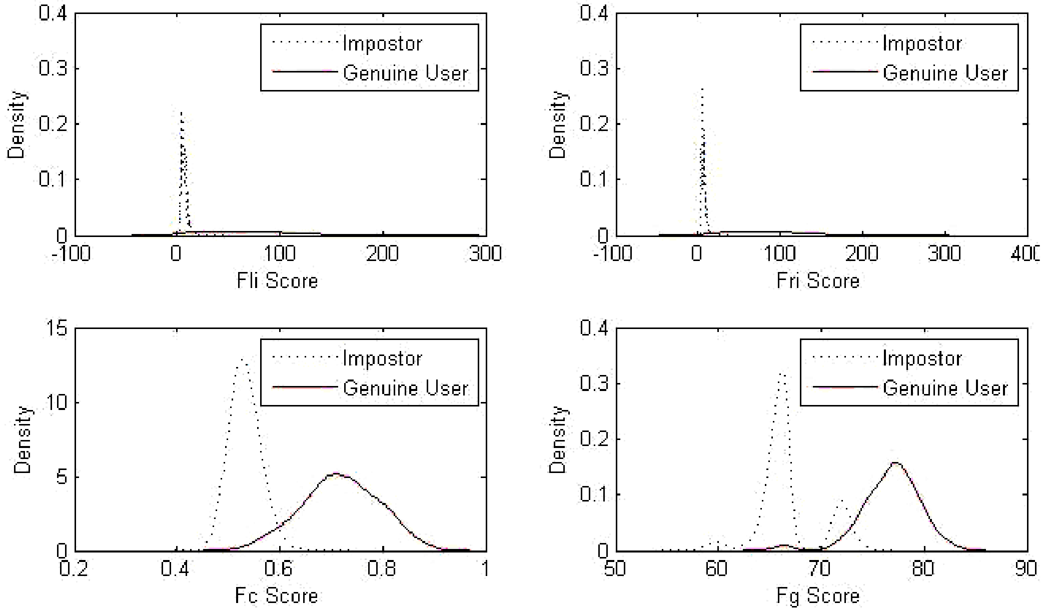


Fig 3-3. NIST-BSSR1 all matchers' score distributions.

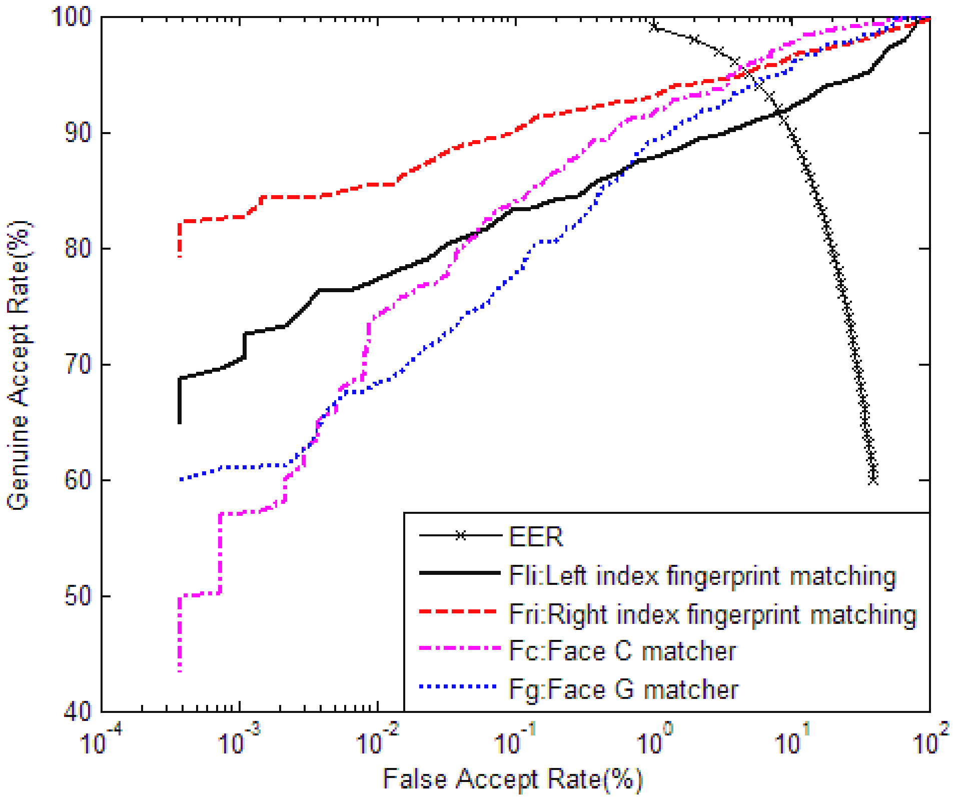


Fig 3-4. ROC curves for the matching in NIST-BSSR1 multimodal database.

3.3.2 Xm2vts Benchmark Score Database

There are five facial matchers (F1~F5) and three speech matchers (S6~S8) in Xm2vts benchmark database. Each matcher is constructed using different feature and classifier [146]. The facial and speech matchers are based on the following features:

1. **FH**: Normalised face image concatenated with its RGB Histogram.
2. **DCTs**: Discrete Cosine Transform coefficients that are calculated from face image (with size of 40x32 pixels) features.
3. **DCTb**: Discrete Cosine Transform coefficients that are calculated from face image (with size of 80x64 pixels) features.
4. **LFCC**: The Linear Filter-bank Cepstral Coefficient speech features.

5. **PAC**: The Phase Auto-Correlation Mel Filter-bank Cepstral Coefficient speech features.
6. **SSC**: Spectral Subband speech features.

Two different classifiers were used for these experiments: MLP and a Bayes Classifiers using GMM. So these feature and classifier combinations form five facial matchers and three speech matchers as the baseline systems. These eight combinations are as listed:

1. F1: (**FH, MLP**)
2. F2: (**DCTs, GMM**)
3. F3: (**DCTb, GMM**)
4. F4: (**DCTs, MLP**)
5. F5: (**DCTb, MLP**)
6. S6: (**LFCC, GMM**)
7. S7: (**PAC, GMM**)
8. S8: (**SSC, GMM**)

The Xm2vts database contains of 295 individuals' speech and facial score samples. Each individual contributes eight samples per modality which are taken within four sessions and one month interval with two samples for each session. Within the score sets, there are 1000 genuine scores and 151,800 impostor scores from the development and evaluation sets. 200 out of 295 users enroll on the systems and 5 sample images are acquired from them to create 1000 genuine scores (5 x 200). The remaining users (95 users) acted as external impostors that are not enrolled on the systems. Eight samples from these external impostors are used to generate 152,000 impostor scores (95 x 8 x 200) against the enrolled users. By elimination of the samples that failed to be compared, the actual available impostor scores are 151,800. The author divides this database into training and testing sets. The training set under the Lausanne Protocol 1 (LP1) [146] consists of 600 (3 x 200) genuine user scores generated using 3 genuine samples and 40,000 (25 x 8 x 200) impostor scores from 25 external impostors. The testing set includes 400 (2 x 200) genuine user scores from the rest of the genuine samples and 111,800 (≈ 70 x 8 x 200) impostor scores from the rest of the external impostors. These matchers' score distributions are given in fig. 3-5 and their verification ROC curves are depicted in fig. 3-6.

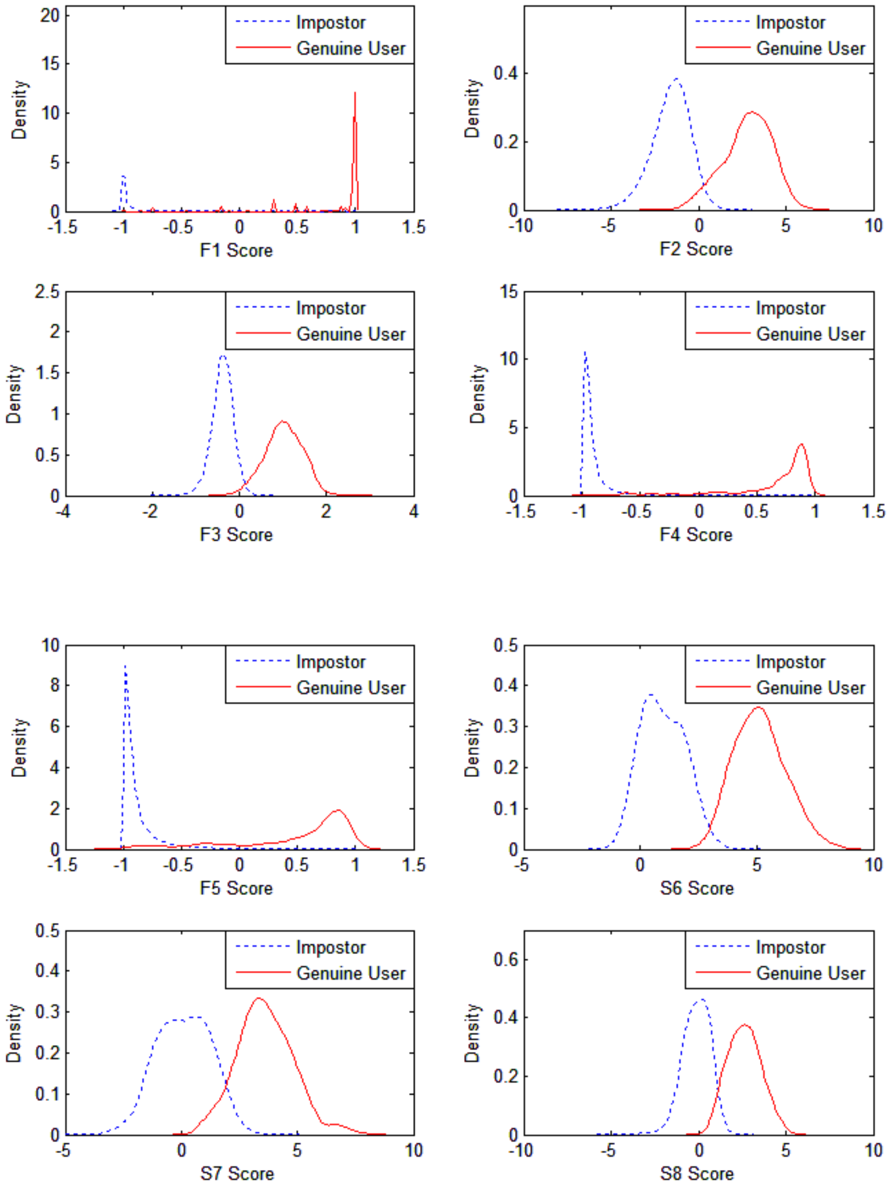


Fig 3-5. Xm2vts benchmark database all matchers' score distribution.

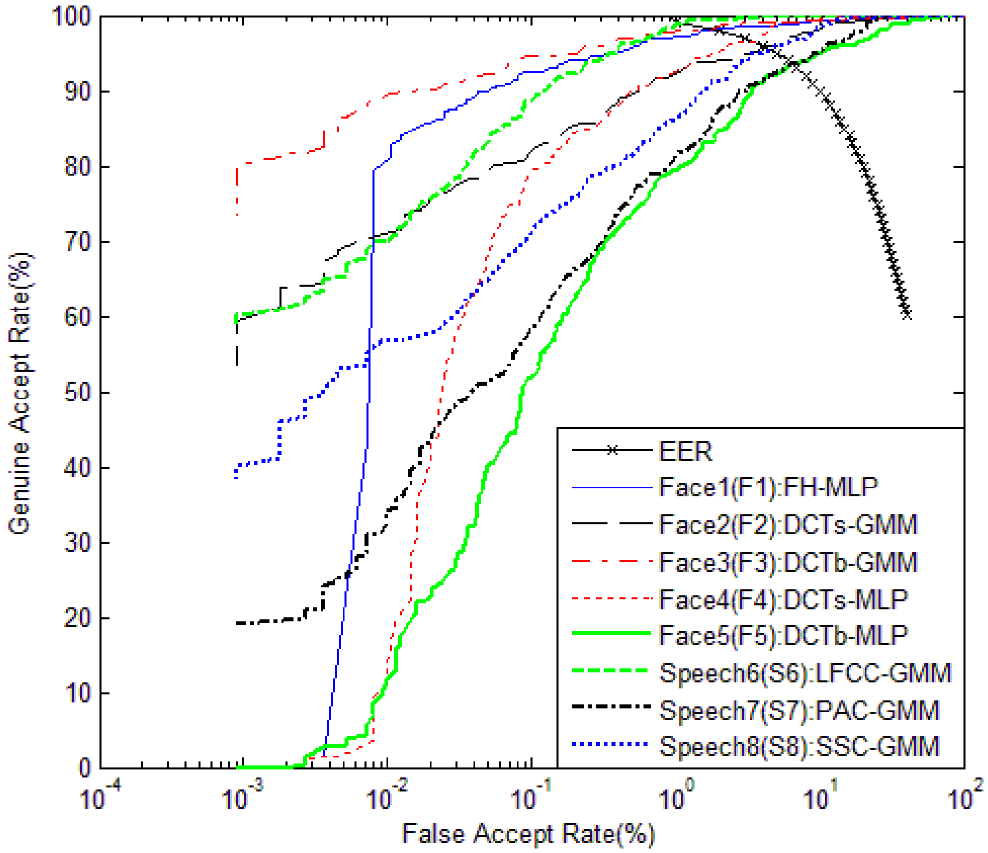


Fig 3-6. ROC curves for the matchers in Xm2vts benchmark database.

Both NIST-BSSR1 and Xm2vts benchmark database use the impostors that are created using cross sample matching. Such impostors are never intended to defeat the system therefore are termed as unskilled forgeries. Those created by a user who is instructed to make such an attempt given information about the targeted user are termed as skilled forgeries. The biometrics evaluation that depends on only unskilled forgeries can be insufficient. However, there is no strong means by which one can define a good forger and prove his/her existence (or non-existence) that such analysis is theoretically impossible [157]. Therefore most studies in the biometric community to date only incorporate unskilled forgeries, and very rarely skilled forgeries [158].

3.4 Experiment Set Up, Comparisons and Result Analysis

Using the NIST-BSSR1 multimodal biometrics database, 4 bimodal biometrics score level fusion experiments are conducted. This database involves fingerprint and frontal face biometrics. There are 15 bimodal biometrics fusion experiments from the Xm2vts benchmark database that use frontal face and speech biometrics. To evaluate the robustness of the fusion algorithms towards the sample variation, it is preferred to use several random partitions of the testing and training set rather than just use the single defined partition. Therefore, the Xm2vts defined testing and training sets are mixed and equally separated into testing and training score sets. Such partitions are repeated 30 times in both databases in all experiments. For the density based methods, GMM component numbers are searched in the range of 1~5 and 1~10 respectively for genuine user and impostor. In the following section, the comparisons between different fusion strategies are presented. The comparisons are in terms of average verification performance, relative performance variation against different partitions of the training and testing sets. Such comparisons also include the required training and processing times and their implementation details.

3.4.1 Verification Performance and Its Consistency

Four key operating points are extracted from the ROC curves. Table 3-1~3-4 present the average performance of the single biometrics. The standard deviations of the average results over 30 trials are given in brackets. In table 3-2 and 3-4, the shaded results are the best one within a category whereas the bolded figures (*) are the best fusion results amongst all categories in that particular fusion experiment. M1 and M2 are the first and second matchers used in that particular bimodal biometrics fusion experiment. The lowest operating point performances are also shown graphically in fig. 3-7.

Exp. No	Matcher		**GAR at FAR equals to 0.001%**	
	M1	M2	M1	M2
1	Fli	Fc	72.38(1.58)	52.84(3.25)
2	Fli	Fg	73.50(1.93)	60.35(1.47)
3	Fri	Fc	83.02(2.27)	53.79(4.25)
4	Fri	Fg	83.13(2.50)	62.62(2.55)

(a)

Exp. No	Matcher		**GAR at FAR equals to 0.01%**	
	M1	M2	M1	M2
1	Fli	Fc	77.94(0.83)	72.01(4.70)
2	Fli	Fg	77.85(1.14)	67.61(1.57)
3	Fri	Fc	85.95(1.87)	73.40(2.85)
4	Fri	Fg	85.16(2.00)	70.04(2.55)

(b)

Exp. No	Matcher		**GAR at FAR equals to 0.1%**	
	M1	M2	M1	M2
1	Fli	Fc	82.73(0.75)	84.40(1.51)
2	Fli	Fg	83.21(1.57)	77.19(1.28)
3	Fri	Fc	90.74(1.98)	84.17(1.05)
4	Fri	Fg	90.08(1.86)	79.07(0.85)

(c)

Exp. No	Matcher		**EER**	
	M1	M2	M1	M2
1	Fli	Fc	8.13(0.70)	4.48(0.22)
2	Fli	Fg	8.01(0.73)	5.57(0.47)
3	Fri	Fc	4.64(1.11)	4.31(0.73)
4	Fri	Fg	4.50(0.89)	5.88(0.50)

(d)

** The average performances are reported based on 30 trials of 50% testing and training sets partitions of the score dataset.*

Table 3-1. Single matchers' performances (average GAR) in four NIST-BSSR1 bimodal biometrics fusion experiments under (a) FAR=0.001% (b) FAR=0.01% (c) FAR=0.1% and (d) EER.

Exp. No.	GAR at FAR equals to 0.001%					
	SUM	MAX	LREG	SVM	JLLR	MLLR
1	91.03(0.83)	56.37(3.72)	**93.32(0.41)***	93.14(0.50)	92.70(0.88)	92.05(1.17)
2	92.45(0.91)	61.21(1.43)	91.53(1.51)	91.74(1.52)	**92.88(0.71)***	91.93(2.03)
3	93.08(1.59)	58.81(5.19)	**95.55(0.96)***	95.30(1.01)	95.33(1.17)	95.30(1.72)
4	**94.89(0.98)***	64.04(2.62)	94.71(2.07)	94.65(1.62)	94.12(2.12)	93.75(3.54)

(a)

Exp. No.	GAR at FAR equals to 0.01%					
	SUM	MAX	LREG	SVM	JLLR	MLLR
1	94.82(0.61)	75.35(4.60)	95.38(0.76)	95.39(1.01)	95.78(0.43)	**95.91(0.78)***
2	94.52(0.76)	68.79(1.68)	94.97(0.64)	95.09(0.68)	95.10(0.69)	**95.36(0.65)***
3	95.95(1.18)	77.27(3.46)	97.43(0.64)	97.54(0.50)	**97.75(0.78)***	97.44(1.14)
4	96.37(1.09)	71.48(2.44)	97.03(0.95)	96.81(0.65)	97.69(0.80)	**97.82(1.28)***

(b)

Exp. No.	GAR at FAR equals to 0.1%					
	SUM	MAX	LREG	SVM	JLLR	MLLR
1	97.21(0.51)	87.97(1.34)	98.31(0.73)	98.18(0.61)	**98.46(0.57)***	98.30(0.67)
2	96.18(0.42)	78.71(1.00)	96.24(0.68)	96.47(0.66)	**97.39(0.74)***	97.17(0.43)
3	97.25(0.63)	87.37(1.00)	**98.73(0.37)***	**98.73(0.37)***	98.59(0.51)	98.64(0.45)
4	97.79(0.60)	80.25(0.69)	99.08(0.41)	98.84(0.54)	99.02(0.39)	**99.12(0.38)***

(c)

Exp. No.	EER					
	SUM	MAX	LREG	SVM	JLLR	MLLR
1	1.25(0.38)	3.48(0.46)	1.20(0.46)	1.10(0.32)	**1.04(0.38)***	1.09(0.32)
2	1.63(0.33)	4.69(0.46)	1.52(0.36)	1.38(0.19)	1.20(0.30)	**1.17(0.24)***
3	0.69(0.16)	3.08(0.30)	1.02(0.45)	0.64(0.20)	**0.47(0.13)***	0.49(0.14)
4	1.49(0.45)	5.29(0.47)	0.62(0.28)	0.60(0.20)	0.39(0.10)	**0.38(0.17)***

(d)

** The shaded figures are the best result in those particular fusion categories and the bolded figure with '*' is the best fusion result achieved in that particular experiment.*

*** The average performances are reported based on 30 trials of 50% testing and training sets partitions of the score dataset which are used to obtain the results in Table 3-1.*

Table 3-2. Conventional fusion strategies' performances (average GAR) in four NIST-BSSR1 bimodal biometrics fusion experiments under (a) FAR=0.001% (b) FAR=0.01% (c) FAR=0.1% and (d) EER.

	Matcher		**GAR at FAR equals to 0.002%**	
Exp. No.	M1	M2	M1	M2
1	F1	S6	1.87(4.01)	55.53(1.72)
2	F1	S7	3.51(7.00)	15.94(0.94)
3	F1	S8	4.68(13.77)	41.88(5.13)
4	F2	S6	54.97(4.23)	56.96(3.16)
5	F2	S7	58.04(4.09)	15.78(1.54)
6	F2	S8	62.68(5.00)	39.14(4.72)
7	F3	S6	76.89(3.73)	57.02(2.38)
8	F3	S7	76.19(4.52)	18.21(3.44)
9	F3	S8	79.63(4.83)	38.71(4.60)
10	F4	S6	1.09(0.82)	55.65(2.77)
11	F4	S7	0.50(0.63)	16.86(2.50)
12	F4	S8	0.42(0.38)	37.75(3.39)
13	F5	S6	0.10(0.16)	56.13(2.43)
14	F5	S7	0.18(0.36)	17.44(1.73)
15	F5	S8	0.06(0.09)	38.28(4.23)

(a)

	Matcher		**GAR at FAR equals to 0.01%**	
Exp. No.	M1	M2	M1	M2
1	F1	S6	80.59(4.35)	67.55(3.96)
2	F1	S7	79.72(4.24)	25.39(1.93)
3	F1	S8	78.00(6.95)	50.36(1.69)
4	F2	S6	67.59(2.28)	70.02(2.18)
5	F2	S7	69.25(1.88)	25.48(1.78)
6	F2	S8	70.68(2.47)	50.26(1.66)
7	F3	S6	88.56(1.24)	67.17(3.75)
8	F3	S7	88.53(1.58)	26.91(2.44)
9	F3	S8	88.43(2.08)	49.64(2.15)
10	F4	S6	29.35(11.45)	67.04(3.38)
11	F4	S7	24.06(7.35)	26.08(1.98)
12	F4	S8	25.60(11.13)	50.67(1.80)
13	F5	S6	5.55(3.20)	68.26(3.20)
14	F5	S7	4.52(2.29)	26.70(1.95)
15	F5	S8	5.44(3.32)	50.15(1.71)

(b)

	Matcher		**GAR at FAR equals to 0.1%**	
Exp. No.	M1	M2	M1	M2
1	F1	S6	93.12(0.73)	88.32(1.33)
2	F1	S7	93.46(0.66)	54.98(1.41)
3	F1	S8	93.23(0.88)	67.19(1.35)
4	F2	S6	81.22(1.53)	88.84(0.72)
5	F2	S7	82.63(1.04)	54.10(1.48)
6	F2	S8	81.61(1.29)	68.47(1.18)
7	F3	S6	94.39(0.50)	88.32(0.99)
8	F3	S7	94.15(0.63)	55.38(1.65)
9	F3	S8	94.30(0.62)	67.64(1.19)
10	F4	S6	80.00(1.74)	89.00(1.18)
11	F4	S7	79.65(1.13)	54.74(1.56)
12	F4	S8	80.03(1.13)	68.04(1.40)
13	F5	S6	53.12(2.52)	88.95(0.96)
14	F5	S7	52.28(3.31)	54.90(1.75)
15	F5	S8	51.60(2.42)	68.00(1.58)

(c)

	Matcher		**EER**	
Exp. No.	M1	M2	M1	M2
1	F1	S6	1.74(0.29)	1.01(0.08)
2	F1	S7	1.62(0.33)	5.84(0.27)
3	F1	S8	1.86(0.26)	4.63(0.28)
4	F2	S6	4.61(0.45)	0.93(0.11)
5	F2	S7	4.20(0.40)	5.87(0.25)
6	F2	S8	4.30(0.36)	4.33(0.37)
7	F3	S6	1.52(0.21)	1.00(0.06)
8	F3	S7	1.79(0.21)	5.90(0.22)
9	F3	S8	1.78(0.22)	4.59(0.28)
10	F4	S6	3.10(0.36)	1.02(0.10)
11	F4	S7	3.24(0.18)	5.91(0.28)
12	F4	S8	3.10(0.30)	4.48(0.41)
13	F5	S6	5.63(0.61)	1.01(0.09)
14	F5	S7	5.44(0.47)	5.84(0.41)
15	F5	S8	5.54(0.46)	4.23(0.35)

(d)

** The average performances are reported based on 30 trials of 50% testing and training sets partitions of the score dataset.*

Table 3-3. Single matchers' performances (average GAR) in fifteen Xm2vts bimodal biometrics fusion experiments under (a) FAR=0.002% (b) FAR=0.01% (c) FAR=0.1% and (d) EER.

Exp. No.	**GAR at FAR equals to 0.002%**					
	SUM	MAX	LREG	SVM	JLLR	MLLR
1	93.47(0.88)	1.28(2.68)	96.59(1.04)	**96.65(1.03)***	95.91(1.21)	96.14(1.00)
2	**90.00(1.06)***	2.98(5.02)	88.07(0.83)	88.22(0.85)	88.93(2.57)	89.84(2.19)
3	90.26(0.95)	4.58(14.79)	83.95(3.27)	83.89(3.37)	89.01(3.72)	**91.75(1.06)***
4	**93.18(1.43)***	59.80(6.90)	92.12(1.26)	92.41(1.21)	93.05(1.18)	93.13(1.10)
5	79.01(1.44)	52.90(7.80)	80.40(0.96)	**80.88(1.48)***	78.85(3.59)	80.35(1.61)
6	81.59(3.12)	60.94(10.20)	**82.10(2.73)***	**82.10(2.87)***	79.17(4.94)	81.50(3.38)
7	97.46(0.79)	80.01(3.26)	97.30(1.13)	97.47(0.85)	96.94(1.02)	**97.65(0.78)***
8	83.13(3.38)	31.30(5.47)	87.28(2.60)	87.19(2.60)	**87.86(2.38)***	87.11(2.79)
9	91.95(0.81)	44.11(5.09)	92.45(1.30)	**92.87(1.78)***	92.18(2.12)	92.53(1.43)
10	80.09(2.78)	1.18(0.76)	89.08(12.85)	92.08(6.71)	92.20(3.00)	**92.76(4.96)***
11	66.77(1.57)	0.79(0.67)	76.19(5.32)	**76.21(5.26)***	72.86(2.94)	73.59(3.36)
12	66.56(3.83)	0.60(0.46)	**83.58(2.46)***	83.56(2.49)	79.18(3.63)	81.58(2.19)
13	68.01(4.56)	0.29(0.25)	86.63(3.10)	86.73(3.15)	86.34(2.83)	**87.98(1.82)***
14	50.45(2.44)	0.33(0.46)	62.52(3.92)	62.40(3.93)	57.96(3.78)	**62.94(2.43)***
15	53.77(4.12)	0.30(0.20)	**81.08(3.43)***	80.99(3.36)	77.32(1.72)	79.67(2.10)

(a)

Exp. No.	**GAR at FAR equals to 0.01%**					
	SUM	MAX	LREG	SVM	JLLR	MLLR
1	93.85(0.81)	80.23(5.08)	98.12(0.48)	98.13(0.49)	97.67(0.61)	**98.18(0.54)***
2	92.24(1.01)	79.96(4.30)	91.78(1.21)	91.68(1.28)	94.48(0.63)	**94.49(0.80)***
3	92.89(0.83)	77.18(8.11)	93.73(1.42)	93.72(1.33)	94.25(0.82)	**94.81(1.06)***
4	95.19(0.67)	72.55(4.39)	**96.43(0.56)***	96.36(0.53)	96.32(0.66)	96.37(0.53)
5	85.76(1.17)	64.70(5.01)	85.76(1.38)	**85.89(1.24)***	85.82(1.35)	85.84(1.10)
6	**89.68(1.25)***	73.50(5.88)	89.02(1.22)	89.31(1.26)	89.15(1.38)	89.49(1.47)
7	98.47(0.25)	88.74(3.27)	98.47(0.29)	98.45(0.27)	**98.51(0.38)***	98.45(0.25)
8	92.84(1.50)	43.99(4.75)	94.71(0.82)	**94.73(0.84)***	94.43(0.92)	94.69(0.72)
9	95.59(0.87)	56.20(2.48)	95.70(0.67)	95.48(0.62)	95.22(0.68)	**95.71(0.65)***
10	85.18(1.96)	29.15(11.47)	96.97(0.57)	97.02(0.55)	96.98(0.50)	**97.56(0.49)***
11	75.23(2.37)	24.78(7.27)	86.80(1.72)	**86.84(1.68)***	81.19(1.96)	83.56(1.62)
12	77.46(1.53)	25.65(11.03)	89.01(1.42)	**89.22(1.68)***	87.19(1.48)	89.07(1.14)
13	74.74(1.51)	5.75(3.00)	93.02(1.22)	93.01(1.26)	92.54(1.07)	**93.25(1.56)***
14	58.99(3.11)	4.84(2.45)	75.52(2.14)	**75.89(2.06)***	67.65(2.94)	72.28(3.18)
15	63.20(1.44)	5.80(3.31)	86.03(1.07)	**86.10(0.98)***	81.45(1.39)	85.01(1.09)

(b)

Exp. No.	GAR at FAR equals to 0.1%					
	SUM	MAX	LREG	SVM	JLLR	MLLR
1	95.53(0.82)	93.16(0.74)	**99.35(0.32)***	99.33(0.34)	99.26(0.33)	**99.35(0.24)***
2	95.16(0.52)	93.50(0.60)	97.40(0.44)	97.38(0.42)	**97.58(0.43)***	**97.58(0.43)***
3	94.95(0.83)	93.86(0.97)	98.22(0.31)	98.20(0.32)	97.99(0.35)	**98.23(0.33)***
4	98.66(0.38)	86.54(2.82)	99.05(0.28)	99.00(0.30)	**99.11(0.20)***	99.07(0.22)
5	93.54(0.49)	87.14(1.44)	93.76(0.57)	**93.85(0.58)***	93.82(0.51)	**93.85(0.58)***
6	95.97(0.91)	86.57(3.41)	96.10(0.87)	**96.16(0.94)***	95.76(0.86)	96.00(0.85)
7	**99.49(0.19)***	98.04(0.54)	**99.49(0.19)***	**99.49(0.19)***	**99.49(0.19)***	**99.49(0.19)***
8	97.04(0.41)	74.64(2.86)	97.56(0.38)	97.56(0.35)	**97.58(0.36)***	**97.58(0.40)***
9	97.88(0.55)	74.61(1.75)	97.86(0.50)	97.74(0.48)	**98.05(0.38)***	97.92(0.50)
10	92.45(0.93)	81.16(1.47)	99.30(0.22)	99.29(0.23)	99.37(0.21)	**99.42(0.21)***
11	90.17(0.97)	81.13(1.26)	95.42(0.70)	95.40(0.65)	95.20(0.62)	**95.49(0.74)***
12	89.30(0.83)	85.02(1.03)	**96.86(0.50)***	96.80(0.49)	96.18(0.66)	96.45(0.59)
13	82.37(1.50)	53.70(2.73)	**98.43(0.42)***	98.42(0.44)	97.83(0.47)	98.24(0.35)
14	78.04(0.95)	53.13(3.32)	89.07(0.77)	88.97(0.83)	89.24(0.72)	**89.66(0.71)***
15	76.93(1.45)	54.62(2.87)	93.77(0.90)	**93.79(0.81)***	91.91(0.78)	92.94(0.89)

(c)

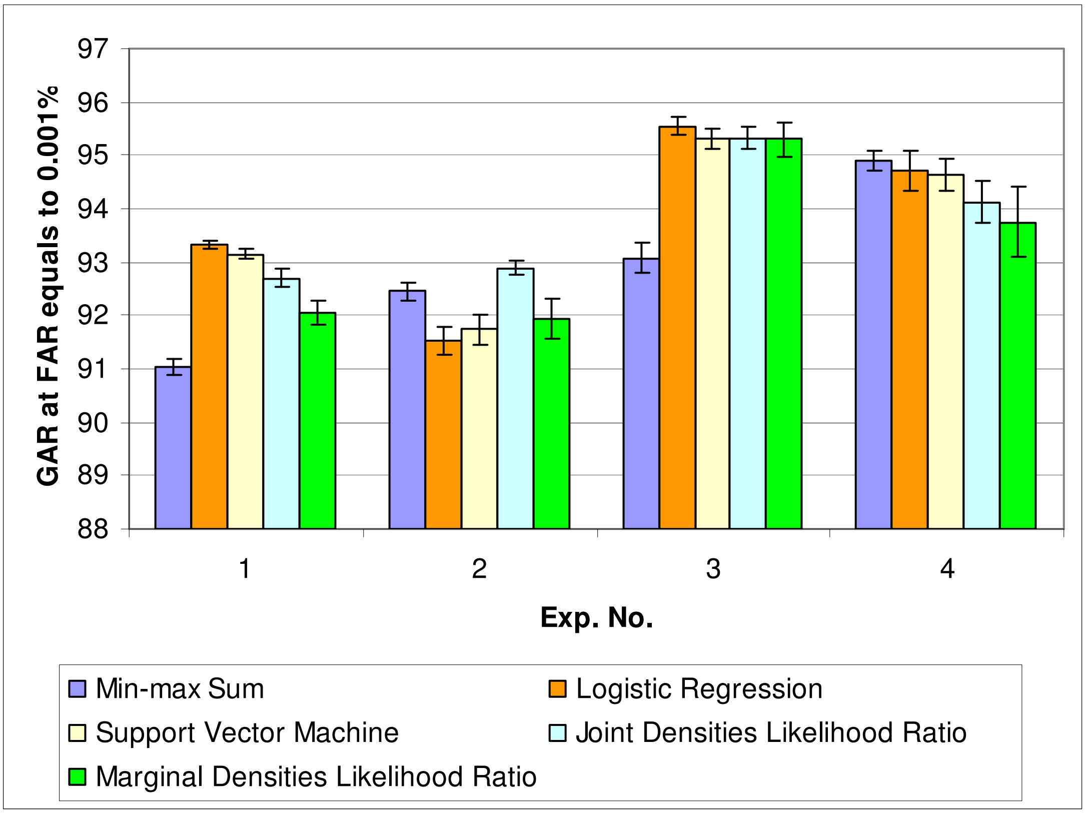
Exp. No.	EER					
	SUM	MAX	LREG	SVM	JLLR	MLLR
1	0.82(0.10)	0.59(0.11)	0.46(0.21)	0.35(0.11)	0.29(0.08)	**0.25(0.05)***
2	1.03(0.21)	1.03(0.08)	0.75(0.08)	**0.72(0.07)***	0.86(0.12)	0.75(0.14)
3	1.15(0.14)	0.99(0.20)	0.71(0.24)	**0.63(0.11)***	0.82(0.14)	0.67(0.17)
4	0.40(0.09)	2.52(0.62)	0.38(0.14)	**0.37(0.14)***	0.42(0.15)	0.41(0.15)
5	**1.30(0.21)***	1.99(0.29)	1.31(0.20)	1.31(0.18)	1.32(0.20)	1.32(0.20)
6	1.08(0.23)	2.19(0.36)	**1.02(0.19)***	1.03(0.20)	1.07(0.20)	1.03(0.23)
7	0.43(0.13)	0.40(0.18)	0.47(0.18)	0.43(0.13)	**0.33(0.10)***	0.38(0.09)
8	1.09(0.17)	2.31(0.29)	1.00(0.21)	0.99(0.21)	**0.89(0.15)***	0.92(0.15)
9	0.75(0.31)	3.36(0.25)	0.76(0.32)	**0.69(0.24)***	0.73(0.22)	0.72(0.23)
10	0.99(0.09)	0.84(0.10)	0.39(0.16)	0.33(0.08)	**0.28(0.09)***	0.29(0.10)
11	1.31(0.10)	1.68(0.18)	0.68(0.05)	0.68(0.04)	0.62(0.08)	**0.58(0.12)***
12	1.50(0.14)	1.73(0.13)	0.83(0.21)	0.83(0.20)	0.78(0.13)	**0.71(0.12)***
13	2.28(0.22)	1.52(0.16)	0.50(0.12)	**0.49(0.09)***	0.56(0.10)	0.53(0.11)
14	3.10(0.24)	2.69(0.32)	1.65(0.24)	**1.62(0.26)***	1.68(0.26)	1.69(0.26)
15	3.33(0.23)	2.54(0.21)	1.47(0.18)	1.46(0.18)	1.33(0.22)	**1.31(0.23)***

(d)

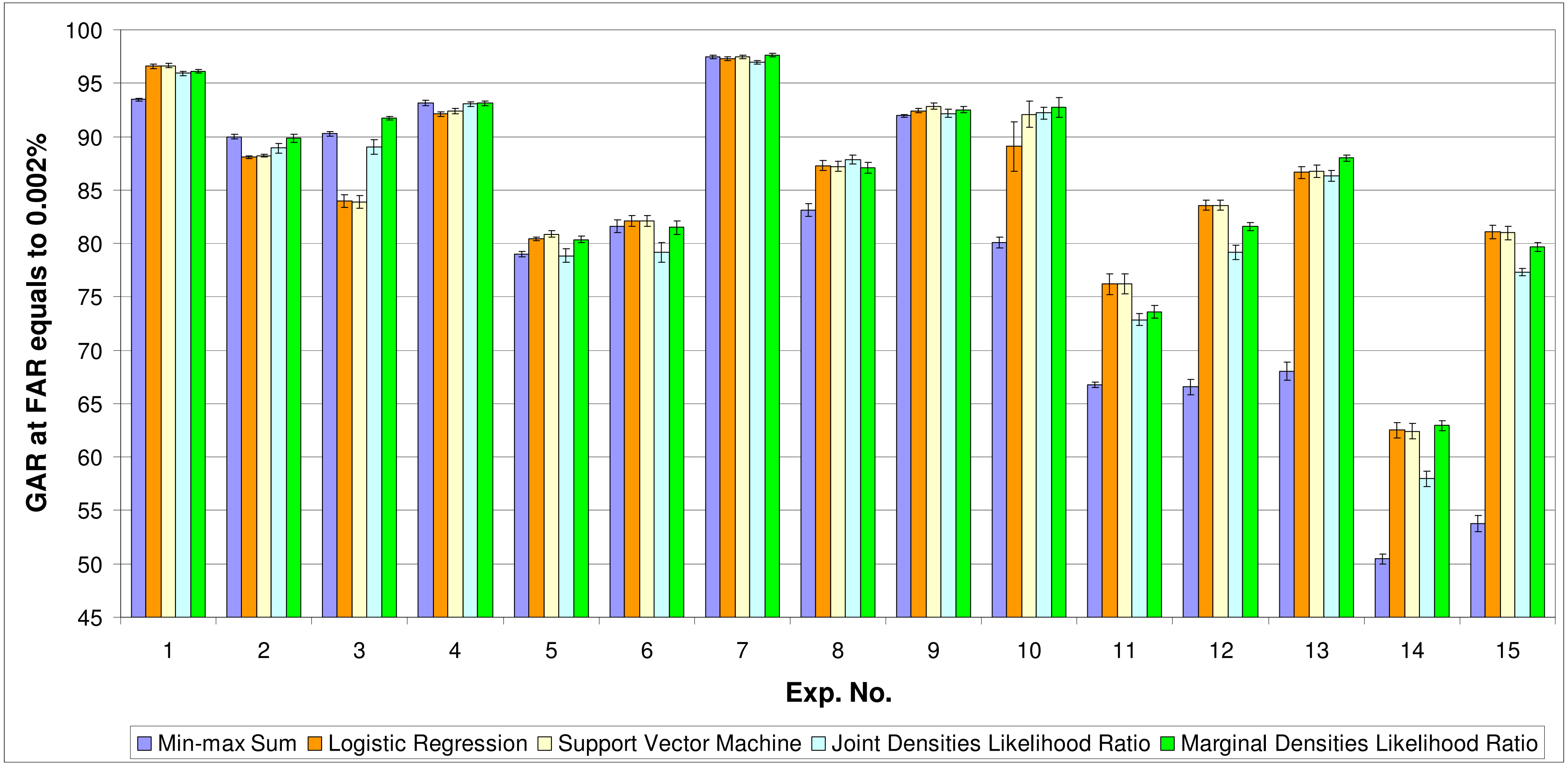
** The shaded figures are the best result in those particular fusion categories and the bolded figure with '*' is the best fusion result achieved in that particular experiment.*

*** The average performances are reported based on 30 trials of 50% testing and training sets partitions of the score dataset which are used to obtain the results in Table 3-3.*

Table 3-4. Conventional fusion strategies' performances (average GAR) in fifteen Xm2vts bimodal biometrics fusion experiments under (a) FAR=0.002% (b) FAR=0.01% (c) FAR=0.1% and (d) EER.


GAR at FAR equals to 0.001%
97
96
95
94
93
92
91
90
89
88
1
2
3
4
Exp. No.
Min-max Sum
Logistic Regression
Support Vector Machine
Joint Densities Likelihood Ratio
Marginal Densities Likelihood Ratio

(a)



** The error bars are plotted at +/- 1 standard error (sample no. = 30).*

(b)

Fig 3-7. Conventional fusion strategies average performance at lowest operating point in (a) NIST-BSSR1 (b) Xm2vts.

From table 3-2 and 3-4, it can be seen that there are no single fusion algorithm that outperforms all the others in the majority of the experiments. However it is obvious that the training based methods (classification and density based method) perform better than the non-training based rule based methods, because they utilise prior knowledge. However, both the density based and the classification based methods generally achieve comparable fusion performances.

An overall comparison result, including the single biometrics performances that are given in table 3-1 and 3-3, shows that all fusion approaches, except the Max rule outperform single biometrics at four operating points. For instance, M1 almost outperforms Max in all experiments and operating points in NIST-BSSR1. This supports the argument that Max rule using only a single source is not efficient. Sum rule, as another rule based method, performs much better than Max rule. For example at the lowest operating point, in 3 experiments out of 19, it achieves the best results over all of compared methods. Nevertheless, for the remaining experiments, the differences between the Sum rule and the best achieved results vary from 0.18%~27.3% at the lowest operating point. This shows significant fusion performance variation, probably because no reliable weighting reference is used in the algorithm. This performance inconsistency of the Sum rule fusion at the lowest operating point can be seen from fig. 3-7.

In classification based fusion, both the SVM and LREG perform very similarly to each other. For the Xm2vts database, the difference between the two methods is around 3% for Exp. no.10 (at lowest operating point). For the rest of the experiments over all the operating points, the differences are just within 0.4%. This 3% difference achieved by LREG however has a high standard deviation. In this case, some of the lower genuine scores are projected to an even lower score region by the logistic function, therefore a very low threshold is needed to pass these genuine users, which will significantly increase the FAR. Therefore LREG is not able to operate on a very low FAR in some fusion cases. Classification based methods do not perform well at some operating points compared to the density based methods. For instance, in Exp. no. 3 using the Xm2vts database at lowest operating point, both the classification based methods' performances are around 8% less than the one for the density based methods and around 3% less in Exp. no.2, i.e. for at the second lowest operating point. This is because the classification based methods inherently find the single best separation boundary, i.e. the performance is not optimised for all operating points.

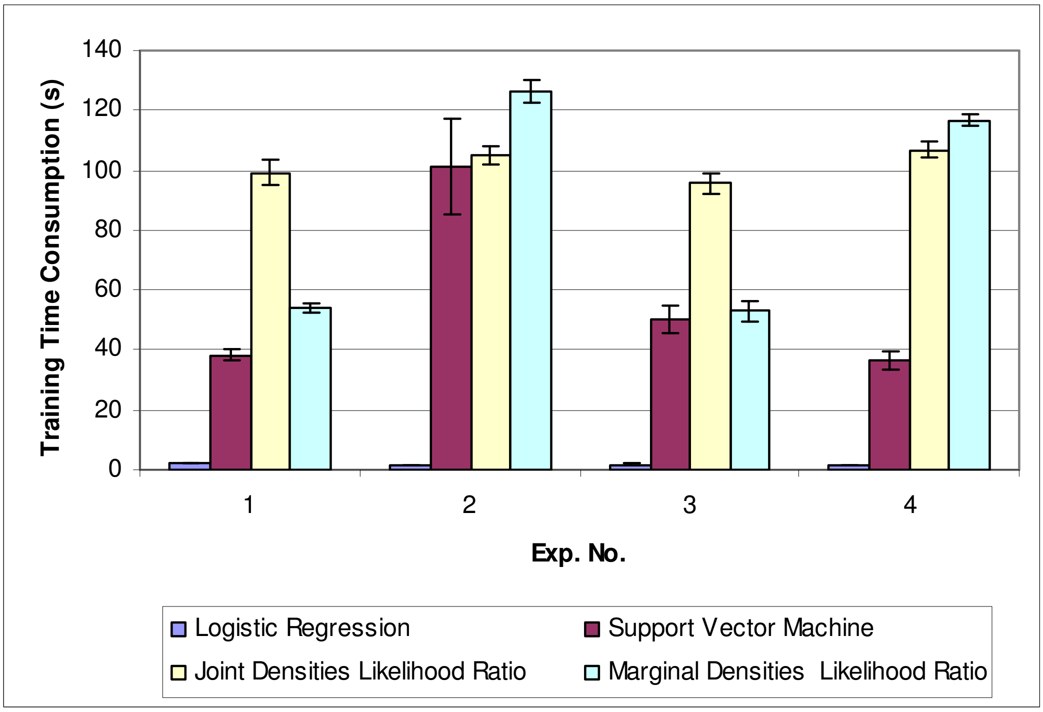
For NIST-BSSR1 dataset, both density based methods perform comparable to each other. Their performance differences over all the operating points are just within 1%. However, MLLR using the marginal density generally outperforms the other algorithm in the Xm2vts database. At FAR equals to 0.002%, there are 5 experiments are with performance differences more than 2%. Whereas at FAR equals 0.01%, there are 3 experiments are with performance differences more than 2%. These performances are accompanied by a high standard deviation. For instance, for Xm2vts Exp. no. 6, there is a 2.33% GAR difference with a 4.94 standard deviation. In this comparison work, GMM uses the component numbers from the range 1~10 and 1~5 respectively for impostor and genuine user density estimation. This might be not enough to build the joint density model accurately and therefore might causes the degradation of MLLR.

In rule based fusion, by choosing only a single source score, the Max rule is not as efficient as the other compared schemes. The Sum rule can perform better. However, using equal weights for the Sum rule fusion cannot perform consistently. Classification based methods and density based methods perform consistently and comparable to each other, however these methods have certain limitations. LREG might fail at very low operating points, and the same goes for SVMs, Their performance is not optimised for all operating points. SVMs can achieve better performances by replacing the linear kernel function with a more complex one such as Radial Basis Function, Polynomial Function or Hyperbolic Tangent Function. But such a kernel function and its parameters need to be chosen carefully on a case-by-case basis. The performance of MLLR and JLLR that use GMMs for density modeling highly relies on the modeling accuracy. It heavily depends on the selected component number. Inaccurate component numbers cause low and inconsistent fusion performance. Searching for the accurate component numbers requires specific algorithms and will on the other hand increase the training and processing time.

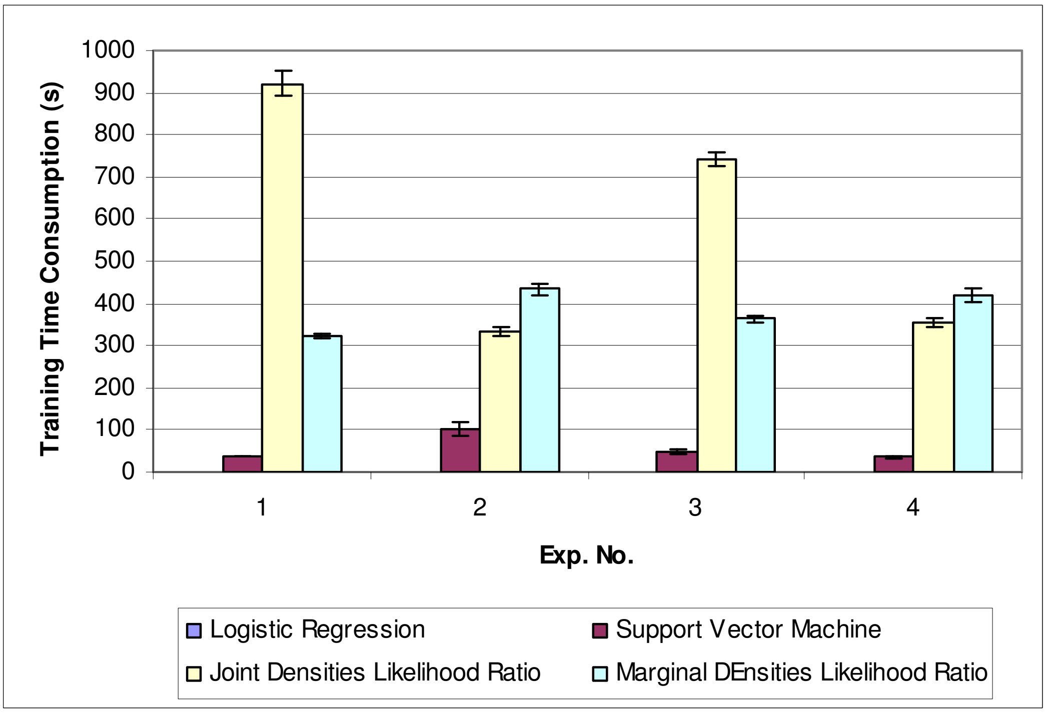
3.4.2 Required Training Times

Biometrics performance is sensitive to factors such as ambient condition, aging effects, matcher setting, user interaction with matcher and etc. Such effects become more obvious when dealing with a larger population. For a training based fusion algorithm, to maintain its performance against the mentioned factors, online retraining is needed [148]. Therefore it is important to consider the training time when assessing the fusion performance.

Fig. 3-7 shows that the five compared methods, except the Sum rule, can perform consistently well near top of the verification rate at lowest operating point. Despite of this, Sum rule requires only an addition operation and can almost instantly do the fusion whereas training will be required for the four other approaches. Relative training times used to achieve the reported performances are shown in fig. 3-8 and 3-9. These experiments are carried out using Matlab testing platform under Microsoft Windows Environment with 1.6GHz CPU speed and 2GB RAM.



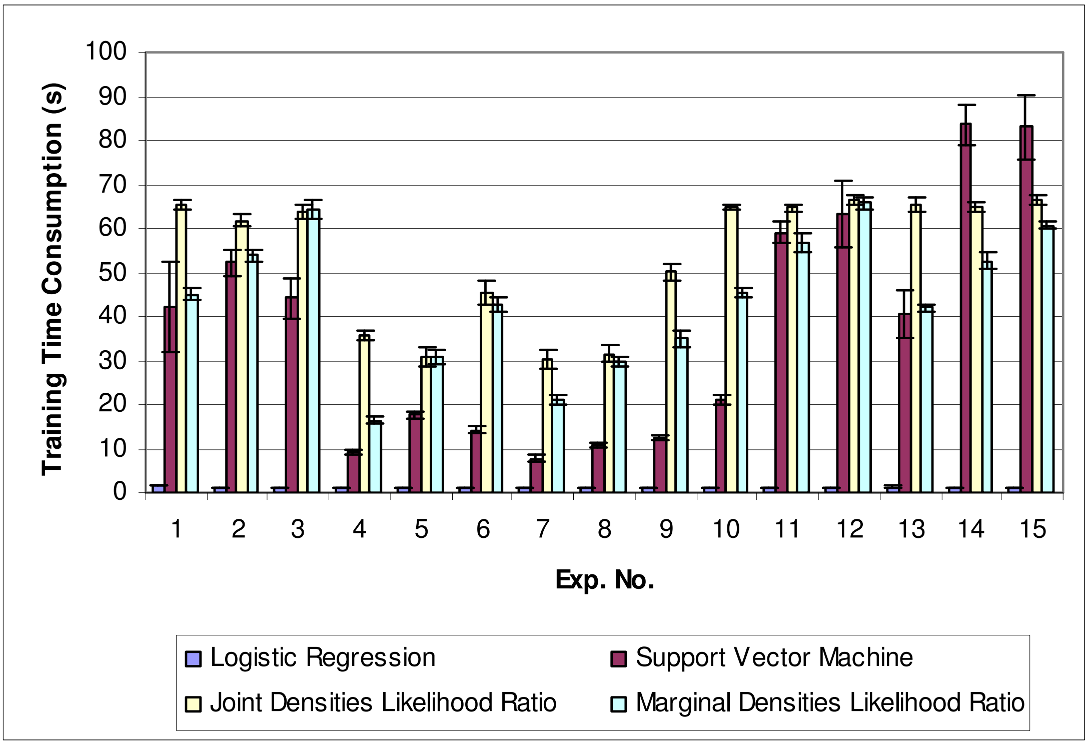
(a)



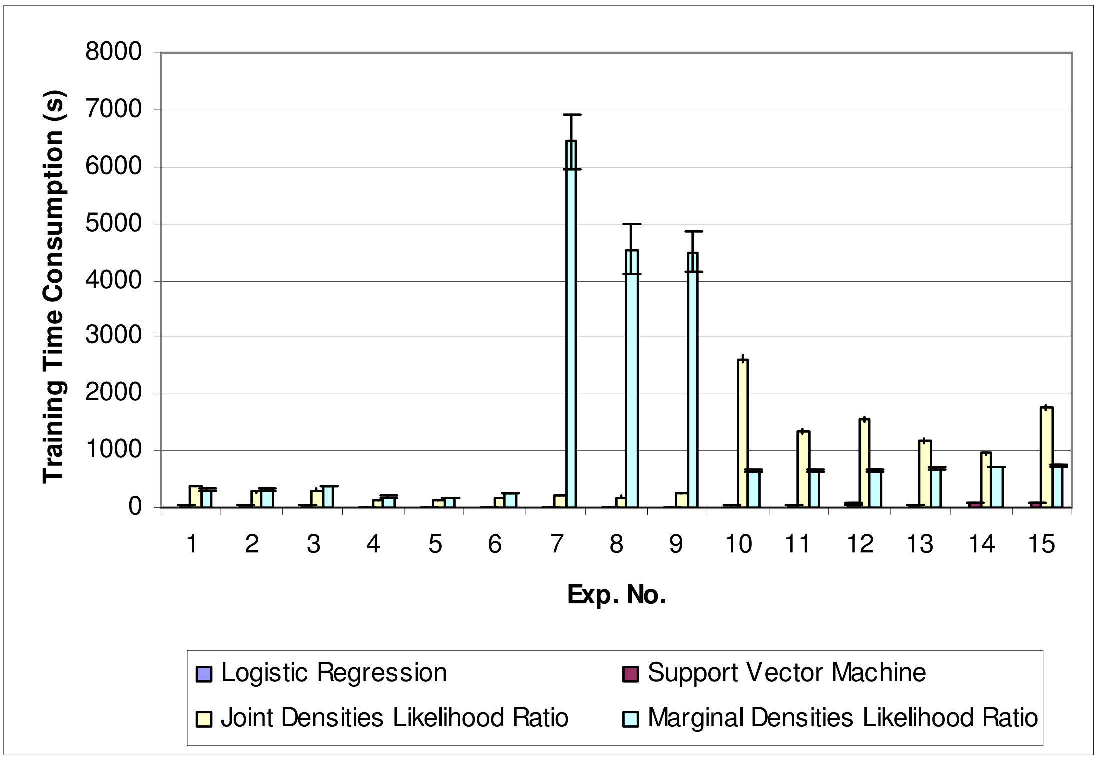
(b)

* *The error bars are plotted at +/- 1 standard error (sample no. = 30).*

Fig 3-8. NIST-BSSR1 fusion required training times (a) without GMM component numbers searching algorithm (b) with GMM component numbers searching algorithm.



(a)



(b)

** The error bars are plotted at +/- 1 standard error (sample no. = 30).*

Fig 3-9. Xm2vts fusion required training times (a) without GMM component numbers searching algorithm (b) with GMM component numbers searching algorithm.

The only difference between (a) and (b) in fig. 3-8 and 3-9 is the training time for density based methods. (a) illustrates the training time only for GMM density modelling whereas (b) includes the components searching time by using the algorithm from [138]. It can be seen that although the fact that density based methods guarantee optimum performance at all operating points provided the underlying densities are modelled accurately, they require significant modelling time and long periods to search for appropriate component numbers. SVM uses SMO to search for the hyper plane with maximum margin, which requires less training time than density based method. However it consume more times than LREG. LREG, using the MLE algorithm for searching for the logit parameters, requires the least training times among the algorithms.

3.5 Conclusions

Most of the score level fusion comparison works in the literature only consider the fusion performance. However for choosing the most appropriate fusion strategy, one cannot solely used fusion performance as a criterion but also has to consider some other key characteristics of the fusion approach.

SUM, LREG, SVM and GMM-LLR (JLLR and MLLR) were found to be the most effective score level fusion approaches. Through a wider range of bimodal biometrics fusion experiments, their superior performances are again confirmed. However, it is empirically shown that there is no one single approach that guarantees the best performance at all times. Although all these approaches are comparable in achieved verification performance, they have individual limitations which are identified in this comparative work.

For practical implementation, not only the verification performance and limitation have to be considered, the required processing and training times and availability of the resources have to be taken into account as well. All these factors influence the selection of the most appropriate strategy. Table 3-5 summarises all these factors.

		SUM	LREG	SVM	GMM-LLR
1	Verification Performance	Occasionally outperforms training based method. Top performance is not guaranteed in different fusion.	Consistent top performance is only guaranteed at higher operating point (e.g. >0.001% FAR).	Top performance at most of the operating points. But might not be the optimal performance.	Optimal performance is guaranteed at all operating points provided the underlying density distributions are known.
2	Limitation	Inconsistent performance. Weighting scheme can be applied to improve the generalisation performance.	Not suitable for application operates at very low FAR. Optimal result is not guaranteed for all operating points.	Kernel function and its parameters have to be chosen carefully for further improvement. Optimal result is not guaranteed for all operating points.	Greatly relies on density model accuracy. Performance and robustness against samples variation are sensitive to the chosen component numbers.
3	Processing and Training Time	Fastest, no training session. Fusion only involves addition arithmetic.	Fast, MLE for logit parameters' searching is very efficient.	Longer training time to search for the hyperplane.	Long modeling time and very long component numbers searching time.
4	Resources Availability	No resource is required. To improve the consistency, weighting might require certain information.	Sufficient training samples and standard statistical package are needed.	Sufficient training samples and advance statistical package is needed to construct the hyperplane with maximum separation margin.	Sufficient training samples and advance statistical package are needed. Accurate component numbers searching require special algorithm .

Table 3-5. Conventional fusion strategies comprehensive comparisons.

4 CONFIDENCE PARTITION AND HYBRID FUSION IN MULTIMODAL BIOMETRICS VERIFICATION

The Equal Weighted Sum rule is a very promising biometric fusion algorithm. However, it might be of benefit to not applying it to the entire score space. By examining the score distributions of each biometric, it can be seen that confidence regions exist, which enable the introduction of the Confidence Partition in biometrics score space. Here, it is proposed that the Sum rule can be replaced by the Min or the Max rule in the Confidence Partitions to further enhance the verification performance. It is empirically shown that this novel Hybrid Fusion method is able to improve the Sum rule. The performance depends on the careful manual assignment of the Confidence Partition where prior knowledge of the sample distributions will be required. Nevertheless, the results and analysis presented in this chapter suggest that the non-confidence samples play a key role in improving the fusion performance. The results and analysis lead to the concept of using the non-confidence related information to aid multimodal biometrics fusion, which is presented in chapter 5 and 6.

4.1 Introduction

The Equal Weighted Sum rule (EW Sum) is one of the well known score level fusion approaches. This method simply uses the average value of multiple biometrics scores as the fusion result. Surprisingly, this simple and non-training based method appears to be outperforming many complicated training based fusion algorithms [43] and is widely studied in biometric researches [16], [67], [109], [110], [111]]. Through sensitivity analysis, Kittler concludes that the superior performance of the Sum rule is due to its resilience against estimate error [108].

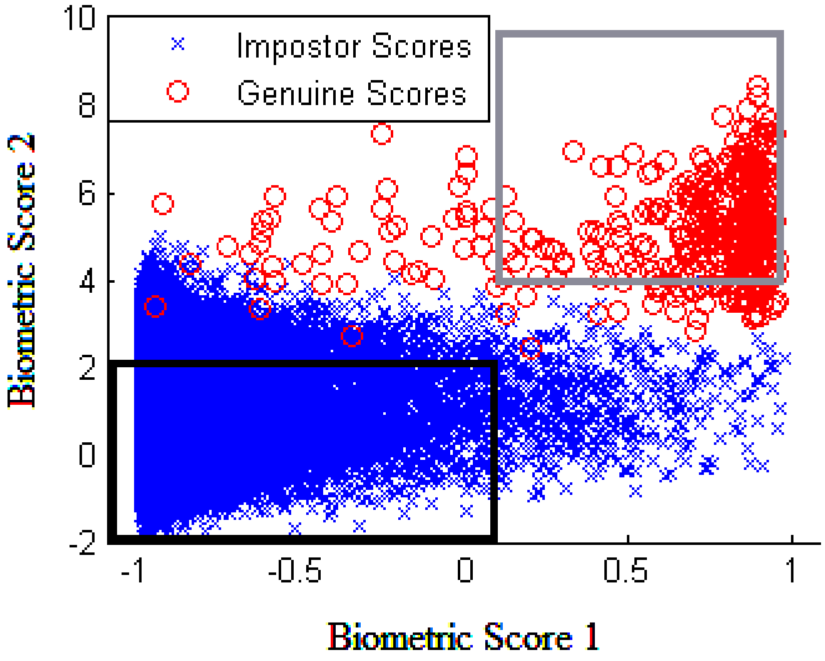


Fig 4-1. Bimodal biometrics score space and the confidence regions.

Fig. 4-1 shows the score space constructed by two biometric matchers that are in similarity measurement (i.e. a claimant is more likely be verified with a higher biometrics score). In the figure, the samples within the black square have a lower bimodal biometrics score. Because there are no appearances of genuine user's samples, it indicates that the testing samples located in this partition are more likely from the impostor group. Testing samples appearing in the grey square are more likely from the genuine user group. Intuitively, due to the higher confidence in these regions, instead of applying the EW Sum rule over the entire score space, the samples residing in the black square can be fused by the Min rule (assigning lower scores for samples which are more likely to be from an impostor) and the samples within the grey square can be fused using the Max rule (assigning higher score to samples which are more likely to be from a genuine user). By doing this, the fusion score from two different groups can be further separated to achieve better verification accuracy. In this chapter, the assignment of Confidence Partitions (CP) to multimodal biometrics score spaces is introduced. Replacing the Sum rule with more appropriate rules in these CPs to increase the fusion verification performance is evaluated. This new approach enables the fusion of multimodal biometrics in a hybrid manner, including the EW Sum, Min and Max rule. This fusion scheme is referred to as Dynamic Score Selection in [139], [149].

The rest of this chapter is organised as follows: Section 4.2 provides details of the proposed method. Section 4.3 presents the experiment set up, results and their analysis. Finally section 4.4 gives the conclusion for this investigation's finding and suggestions for future development.

4.2 Confidence Partition and Hybrid Fusion

In general the proposed method is applicable to higher dimensional score space, it is being tested on bimodal biometrics samples in this section. The reason for choosing only two biometrics sources to fuse at this stage is to investigate the feasibility of this proposed method prior of introducing further complexity.

Referring to fig. 4-1, the Genuine User Confidence Partition (GCP) in the score space is assembled by setting up higher thresholds for two of the biometric matchers. A user with scores higher than these thresholds will be considered as more likely to be a genuine user. The Impostor Confidence Partition (ICP), the score space is formed by two lower thresholds. If the user's multimodal biometric scores are smaller than these thresholds, he/she is more likely to be an impostor.

Score normalisation is required in rule based fusion for effective combination [44]. The simplest normalisation technique is the Min-max normalisation which is shown in table 2-1. It maps the biometric scores into the interval between 0 and 1. This normalisation equation is shown in (4.1). The notation shown in the equation represents the following: S_i is the biometric score of user i, S'_i represents the normalised score. The minimum value (min) and the maximum value (max) of the biometric scores can be estimated from a set of training scores or available from the commercial biometric matcher vendor.

$$S'_i = \frac{S_i - \min}{\max - \min} \tag{4.1}$$

By introducing CP, different rules are applied to different regions. Here, rule (4.2) ~ (4.4) are applied. They are integrated into a hybrid fusion method as shown in (4.5). K is the total number of matchers and S_{fi} denotes the fused score.

1. EW Sum Rule:

$$S_{fi} = \sum_{k=1}^{K} \frac{1}{K} S'_{i,k} \ , \forall i \tag{4.2}$$

2. Min Rule:

$$S_{fi} = \min(S'_{i,1}, S'_{i,2}, \ldots, S'_{K}) \ , \forall i \tag{4.3}$$

3. Max Rule:

$$S_{fi} = \max(S'_{i,1}, S'_{i,2}, ..., S'_{i,K}), \forall i \qquad (4.4)$$

4. Hybrid Rule:

$$S_{fi} = \begin{cases} \text{Apply Min Rule, when } < S'_{i,1}, S'_{i,2}, \ldots, S'_{i,K} > \text{ falls in ICP.} \\ \text{Apply Max Rule, when } < S'_{i,1}, S'_{i,2}, \ldots, S'_{i,K} > \text{ falls in GCP.} \\ \text{Apply Sum Rule, elsewhere.} \end{cases} \qquad (4.5)$$

As shown in equation (4.5), for the confidence partitions, Min or Max rule (instead of Equal Weighted Sum) is applied. Applying Min or Max rule in confidence partition instead of Equal Weighted Sum rule further separates the impostor and genuine user score distributions. The Non-Confidence Partition (NCP) is the complement region of the CPs. It denotes the part where the sample can be easily misclassified. Sum rule is applied in this part due to its good performance in dealing with the estimation error.

4.3 Experiment Set Up and Results Analysis

The proposed methods are tested on two publicly available databases, which are the NIST-BSSR1 multimodal database detailed in section 3.3.1 and the Xm2vts benchmark database detailed in session 3.3.2.

Only the best and the worst biometrics from each modality are chosen for the fusion experiments. In the NIST-BSSR1 multimodal database, the right index fingerprint (Fri) is paired with the facial matcher C (Fc) and the left index fingerprint (Fli) is paired with the facial matcher G (Fg) to develop the best and the worst bimodal biometrics fusion respectively. Fc is chosen as a better facial matcher because it has better performance than Fg at most of the operating points. All the scores in NIST-BSSR1 are used for testing. For the XM2VTS database, the best facial matcher DCTb-GMM (F3) is paired with the best speech matcher LFCC-GMM (S6) whereas the worst DCTb-MLP (F5) facial matcher is paired with the worst speech matcher PAC-GMM (S7) in the experiments. Only the scores from the evaluation set are used for testing for this database.

The aim of this investigation is to find out if the proposed approach is able to enhance the verification accuracy. Therefore, the GCP and ICP are assigned manually at this stage. The chosen confidence partitions' thresholds are listed in table 4.1. These partitions are optimised using the testing samples. Four fusion results that are based on the best and worst biometrics are graphically shown in fig. 4-2 ~ 4-5. Their numerical results are presented in table 4.2 and 4.3. The reported GAR is at FAR equal to 0.001%.

	ICP	GCP
NIST-BSSR1 Best Matchers Fusion	$S_{face} < 0.55$ $S_{finger} < 0.15$	$S_{face} > 0.34$ $S_{finger} > 0.20$
NIST-BSSR1 Worst Matchers Fusion	$S_{face} < 0.35$ $S_{finger} < 0.09$	$S_{face} > 0.20$ $S_{finger} > 0.20$
Xm2vts Best Matchers Fusion	$S_{speech} < 0.48$ $S_{face} < 0.44$	$S_{speech} > 0.41$ $S_{face} > 0.60$
Xm2vts Worst Matchers Fusion	$S_{speech} < 0.43$ $S_{face} < 1.00$	$S_{speech} > 0.67$ $S_{face} > 0.79$

Table 4-1. Manually assignment of confidence partitions for Hybrid Fusion.

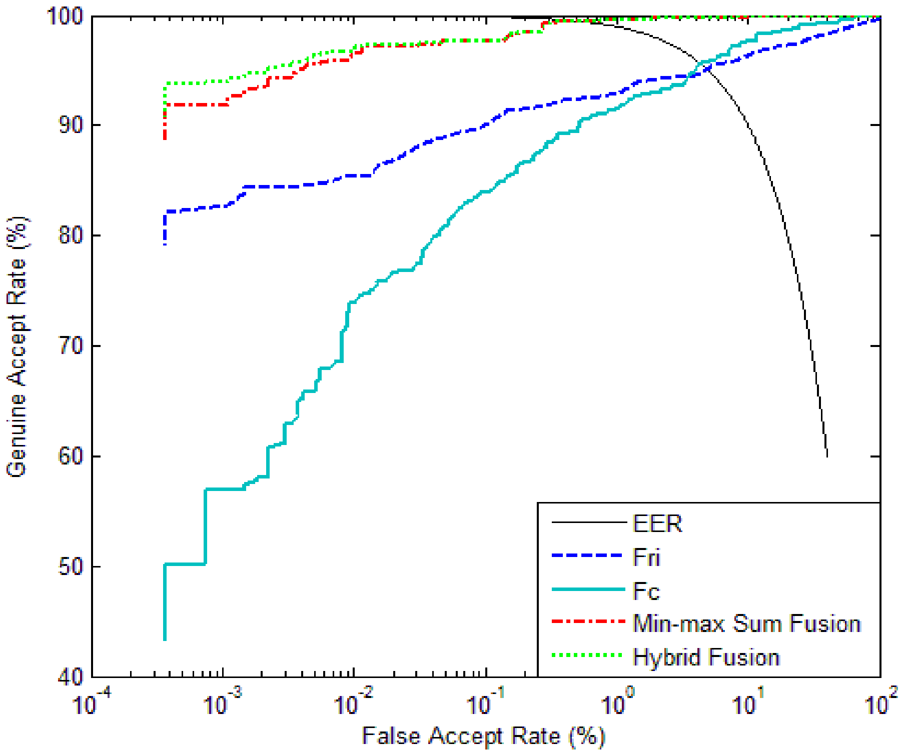


Fig 4-2. NIST-BSSR1 best matchers Hybrid Fusion and its baselines ROC curves.

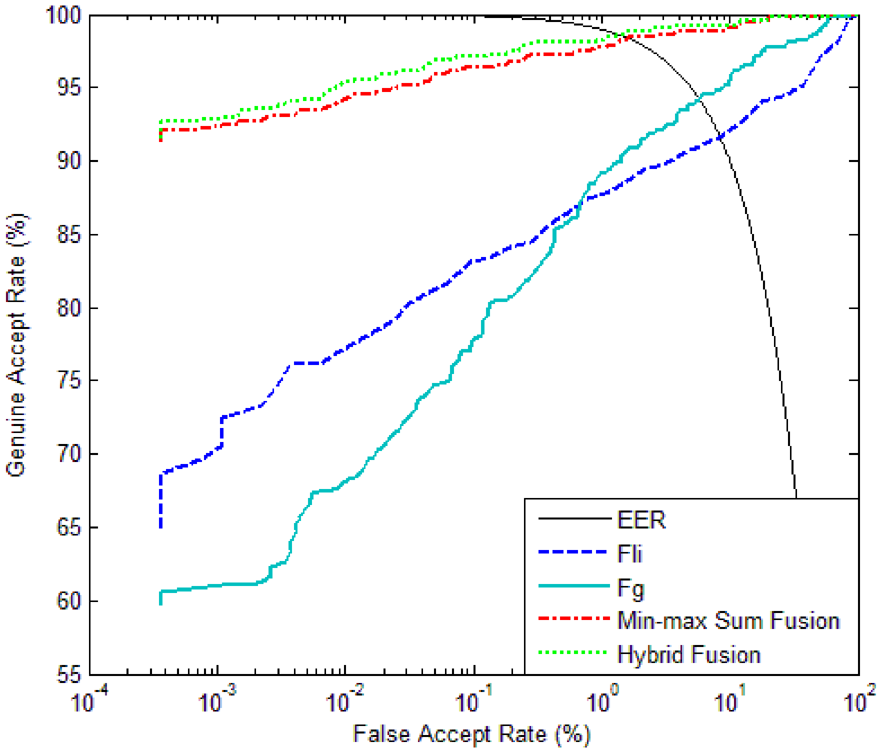


Fig 4-3. NIST-BSSR1 worst matchers Hybrid Fusion and its baselines ROC curves.

	Fingerprint		Face		Min-max Sum		Hybrid Fusion	
	EER (%)	GAR (%)	EER (%)	GAR (%)	EER (%)	GAR (%)	EER (%)	GAR (%)
Best Matchers Fusion	4.5	82.7	4.3	56.9	0.6	91.9	1.0	93.6
Worst Matchers Fusion	8.6	70.0	5.8	61.1	1.6	92.3	1.3	93.0

Table 4-2. Genuine Accept Rate and Equal Error Rate of Hybrid Fusion and its baselines in NIST-BSSR1 multimodal database.

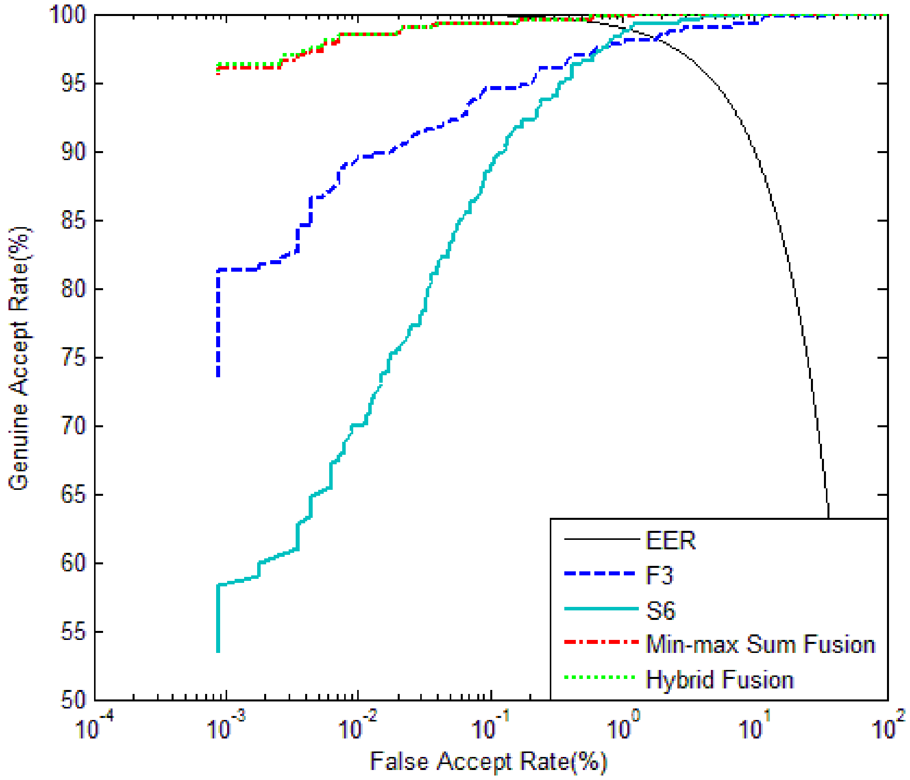


Fig 4-4. Xm2vts best matchers Hybrid Fusion and its baselines ROC curves.

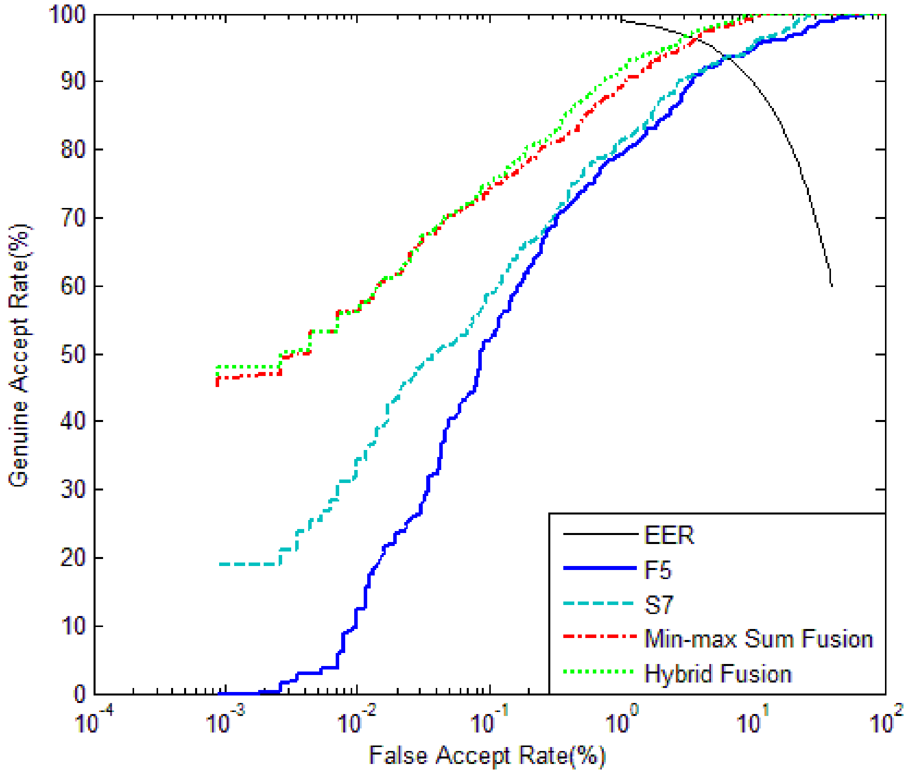


Fig 4-5. Xm2vts worst matchers Hybrid Fusion and its baselines ROC curves.

	Face		Speech		Min-max Sum		Hybrid Fusion	
	EER (%)	GAR (%)	EER (%)	GAR (%)	EER (%)	GAR (%)	EER (%)	GAR (%)
Best Matchers Fusion	1.8	81.3	1.1	58.3	0.5	96.0	0.5	96.3
Worst Matchers Fusion	6.5	0.0	6.5	19.0	3.7	46.3	3.2	48.0

Table 4-3 Genuine Accept Rate and Equal Error Rate of Hybrid Fusion and its baselines in Xm2vts benchmark database

From the graphical and numerical results shown in fig. 4-2 ~ 4-5 and table 4-2 and 4-3, it can be seen that both the Hybrid Fusion and Min-max Sum fusion outperform their single best biometrics. It also can be seen that the proposed Hybrid Fusion is able to further improve the results achieved using the Min-max Sum rule fusion. The GAR improvement at FAR equal to 0.001% over the Sum rule in all the experiments is between 0.3% ~ 3.7%. The lower the EER is the better is the performance. EER improvements for worst matcher fusion in both databases are 18.8% and 13.5% and remained unchanged for the Xm2vts best matcher fusion. Due to inappropriate assignment of ICP in NIST-BSSR1 best matcher fusion, its EER rises up 66.7%. The reason is discussed further in the following section (4.4).

In conclusion, from the reported results have demonstrated that the proposed approach is able to improve the EW Sum rule based fusion. However, such improvement depends on careful manual assignment of CP. Inappropriate CP assignment on the other hand reduces the accuracy. In the next section, an in-depth investigation into the achieved improvements is presented.

4.4 Further Analysis of Proposed Approach

Fig. 4-6 depicts the separation boundary of the EW Sum rule in bimodal biometrics score space. By changing the verification threshold, position of the separation boundary can be adjusted whilst retaining its gradient. The accept and reject regions' size can be controlled by this adjustment to adapt to different security levels. Fig. 4-7 ~ 4-10 illustrate the bimodal biometrics score spaces for four of the bimodal biometrics fusion experiments. By showing the EW Sum rule's separation boundary in the score space, the achieved verification improvements can be visualised.

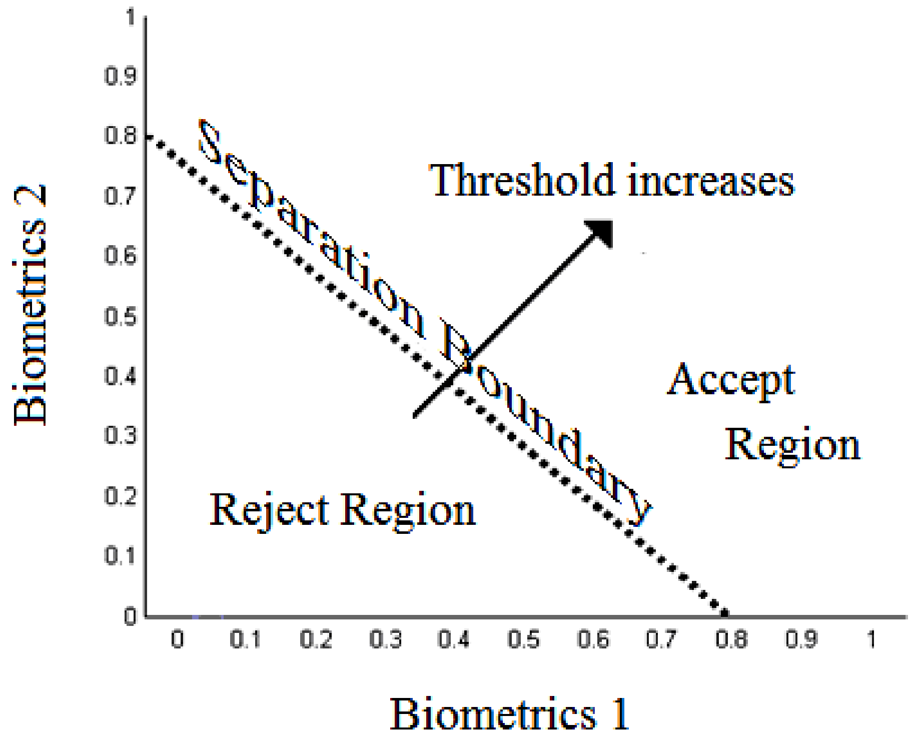


Fig 4-6. EW Sum rule separation boundary in bimodal biometrics score space.

Fig. 4-7 shows a scatter plot of the NIST-BSSR1 best matchers' fusion samples. A non-confidence region is defined by the Min-max Equal Weighted Sum rule separation boundary operated with maximum and minimum thresholds. Only single class sample available out of this region, i.e. the samples located out of the non-confidence region can be safely rejected or accepted.

To apply the Max or the Min rule to the assigned CP equivalent to project the samples onto the *x*=*y* function line. For example, a claimant is initially assigned with the normalised bimodal biometrics scores of 0.3 and 0.5. If this score vector is in the ICP, the Min rule will be applied to his/her score vector. This can be seen as his/her score vector is transformed from (0.3, 0.5) to (0.3, 0.3) on the biometrics score space when the EW Sum separation boundary is used[1]. In ICP, the samples above the separation boundary are vertically projected onto the *x*=*y* function line whereas the samples below are horizontally projected. Similarly for the GCP, if a score vector e.g. (0.7, 0.9) is in GCP, the Max rule is applied. This is equivalent to transform the score vector from (0.7, 0.9) to (0.9, 0.9)[2]. On the score space showing EW Sum separation boundary, this can be seen as the samples above the boundary are horizontally projected whereas the samples below are vertically projected. The figure on the right in fig. 4-7 shows the equivalent scatter plot when the Hybrid Fusion rule (replace EW Sum to Min or Max rule in CP) is applied. The verification improvement over the EW Sum is due to the samples in the green regions. It can be seen the green regions' samples are projected away from the non-confidence region to the confidence region.

As a result of the inappropriate assignment of ICP in the experiment in fig. 4-7 (some of the genuine users are included in the ICP), Hybrid Fusion achieves a higher EER than EW Sum and rises the EER of the EW Sum from 0.6% to 1.0%. However there is still an improvement when the separation boundary operates at FAR equals to 0.001%, the GAR rises from 91.9% to 93.6%. This is contributed by the non-confidence samples that are projected to the GCP. The fusion score distributions of the EW Sum and Hybrid Fusion are shown in fig. 4-11. It can be seen that even further separation between two classes score is achieved by Hybrid Fusion compared to EW Sum, the Hybrid Fusion's improvement is not proportional to this separation.

[1] Applying the Min rule to score vector (0.3, 0.5) has the same fusion result as applying EW Sum rule to (0.3, 0.3), which is 0.3.

[2] Applying the Max rule to score vector (0.7, 0.9) has the same fusion result as applying EW Sum rule to (0.9, 0.9), which is 0.9.

From fig. 4-8, the improvement is mainly contributed by the samples projected to GCP. Since most of the ICP samples are already in the confidence region of the EW Sum rule, further separation (projection of the ICP samples to the x=y function line) of these samples as shown in fig. 4-12 does not result in further improvement. Very moderate improvement is achieved in the experiment in fig. 4-9. This is because of the very small non-confidence region of the EW Sum involved in the projection. This result is also justified by their fusion score distributions as shown in fig. 4-13. It shows that the overlap region of the two fusion strategies does not have a significant difference. EER and GAR improvement of the experiment in fig. 4-10 is due to the large number of ICP samples that are projected. Also from fig. 4-14, the Hybrid Fusion score distribution's overlap region is reduced compared to the EW Sum rule.

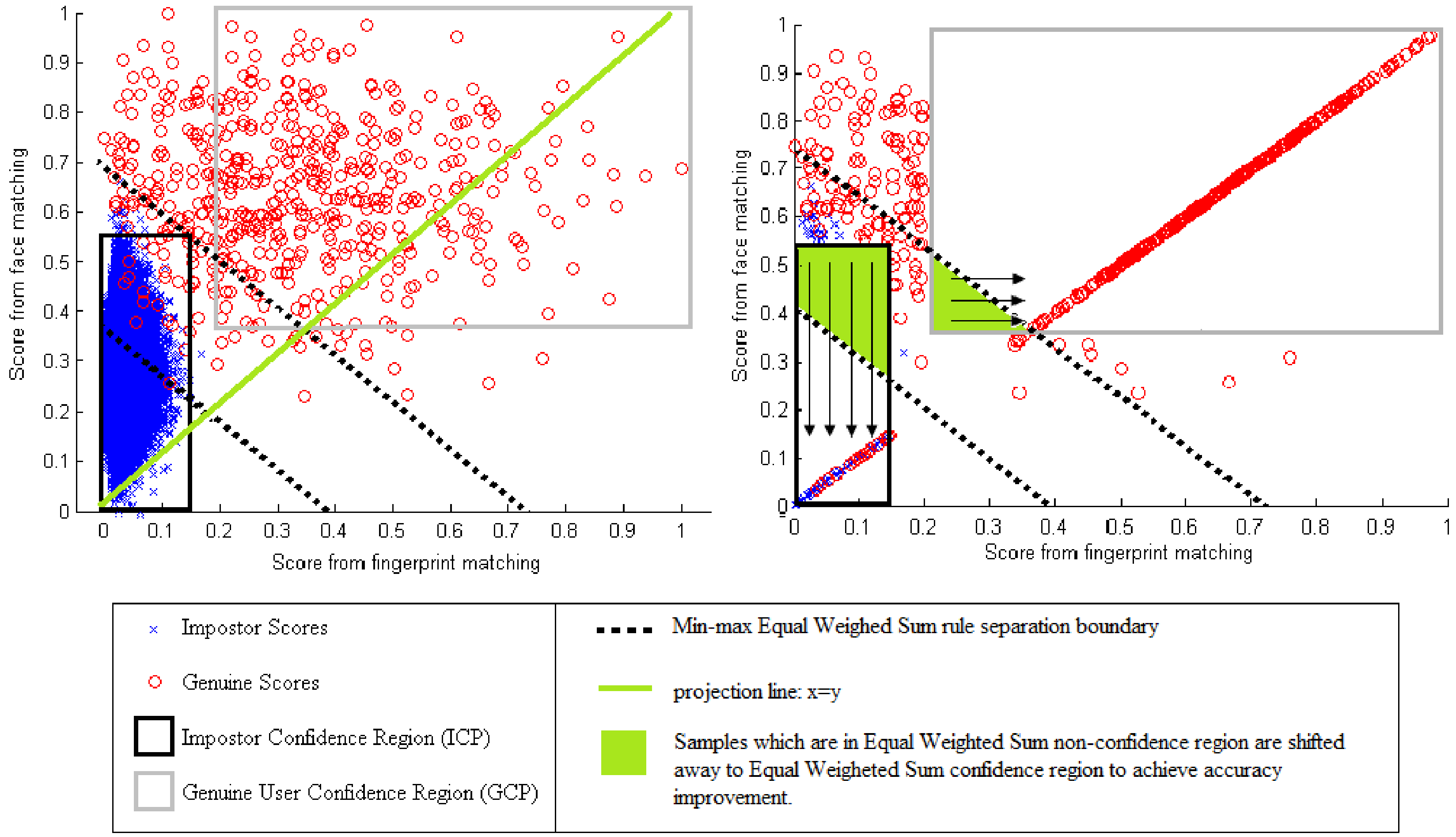


Fig 4-7. NIST-BSSR1 best matchers fusion's testing score space with confidence partition assignments (left) and equivalent confidence sample projections (right).

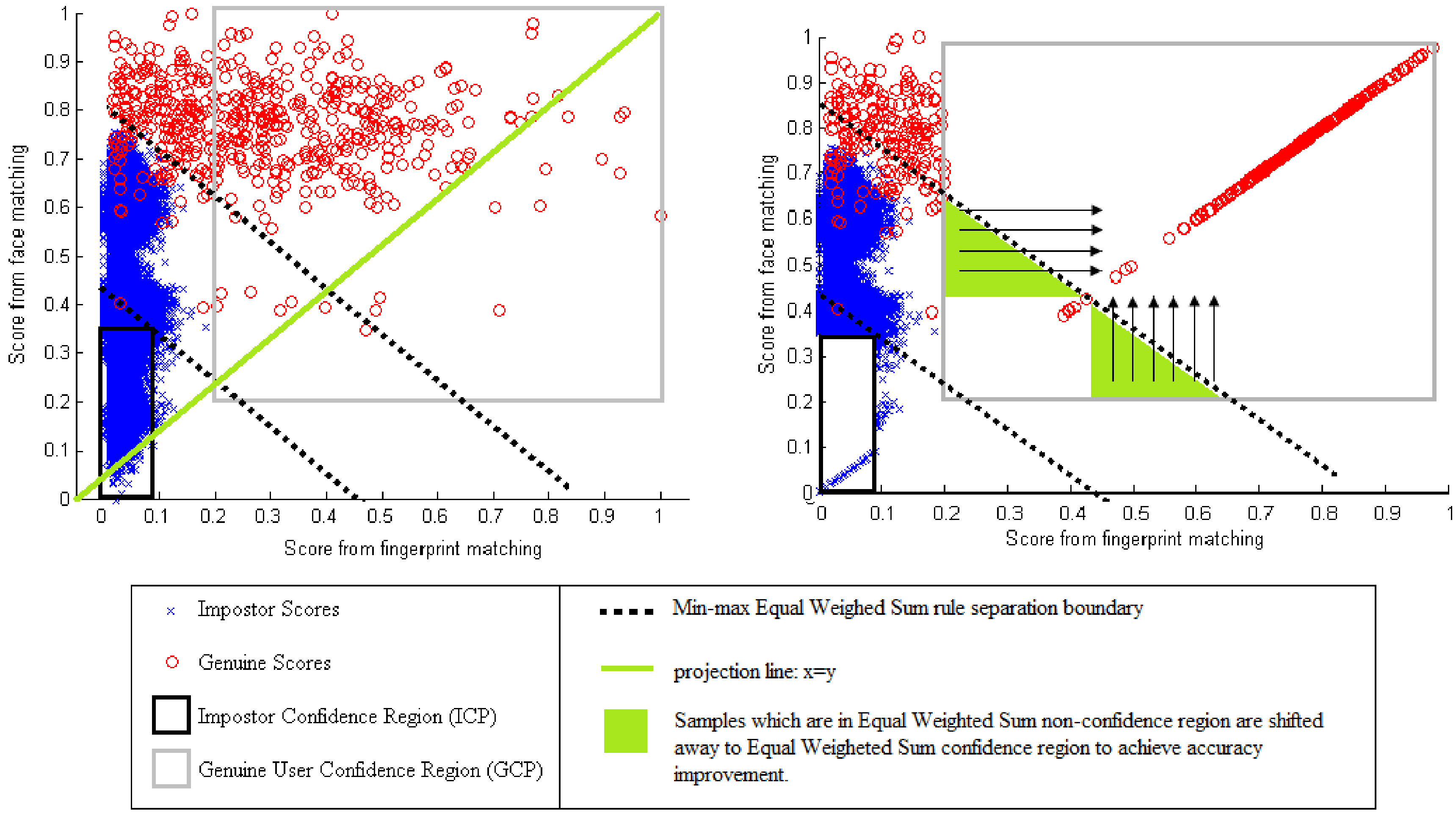


Fig 4-8. NIST-BSSR1 worst matchers fusion's testing score space with confidence partition assignments (left) and equivalent confidence sample projections (right).

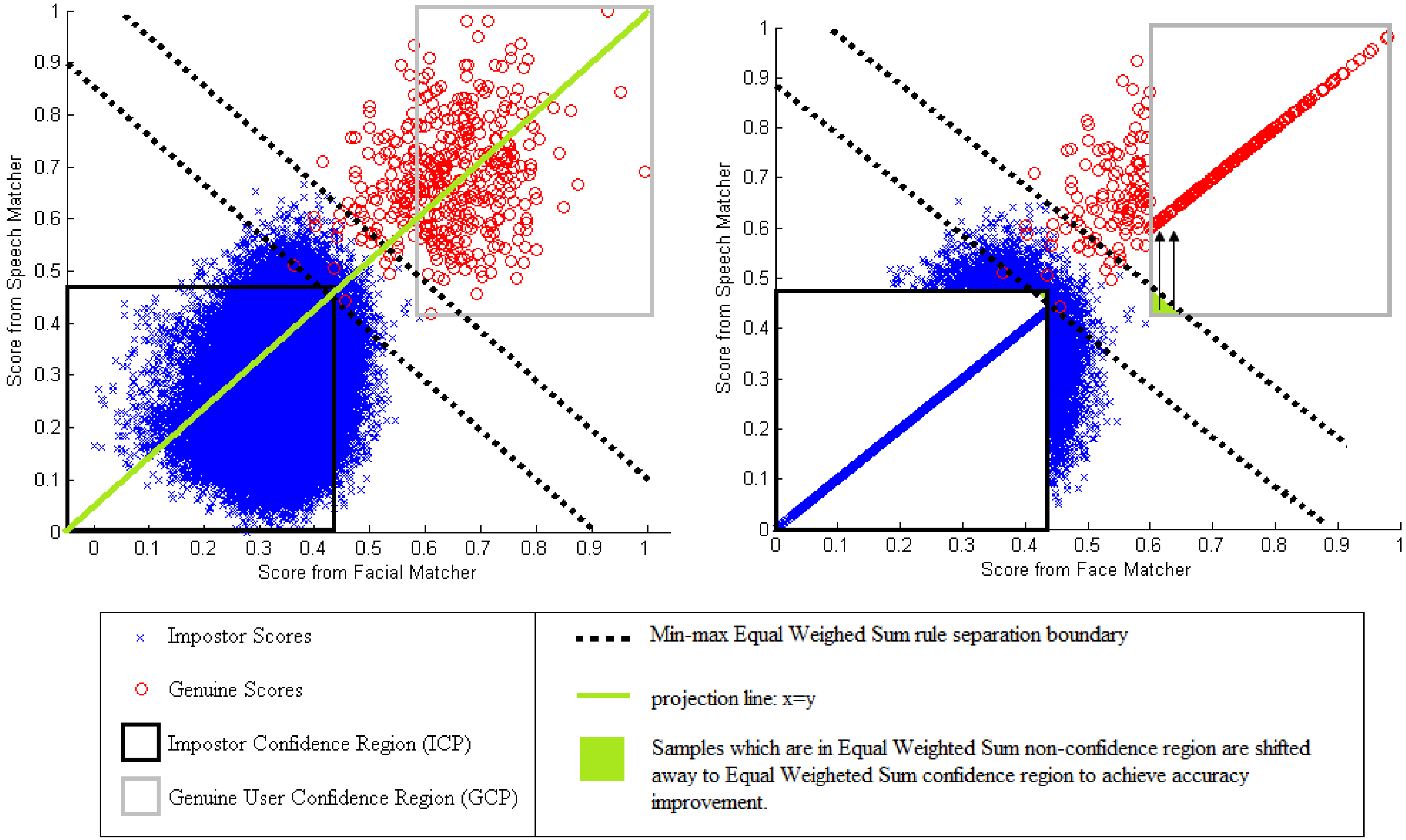


Fig 4-9. Xm2vts best matchers fusion's testing score space with confidence partitions assignments (left) and equivalent confidence sample projections (right).

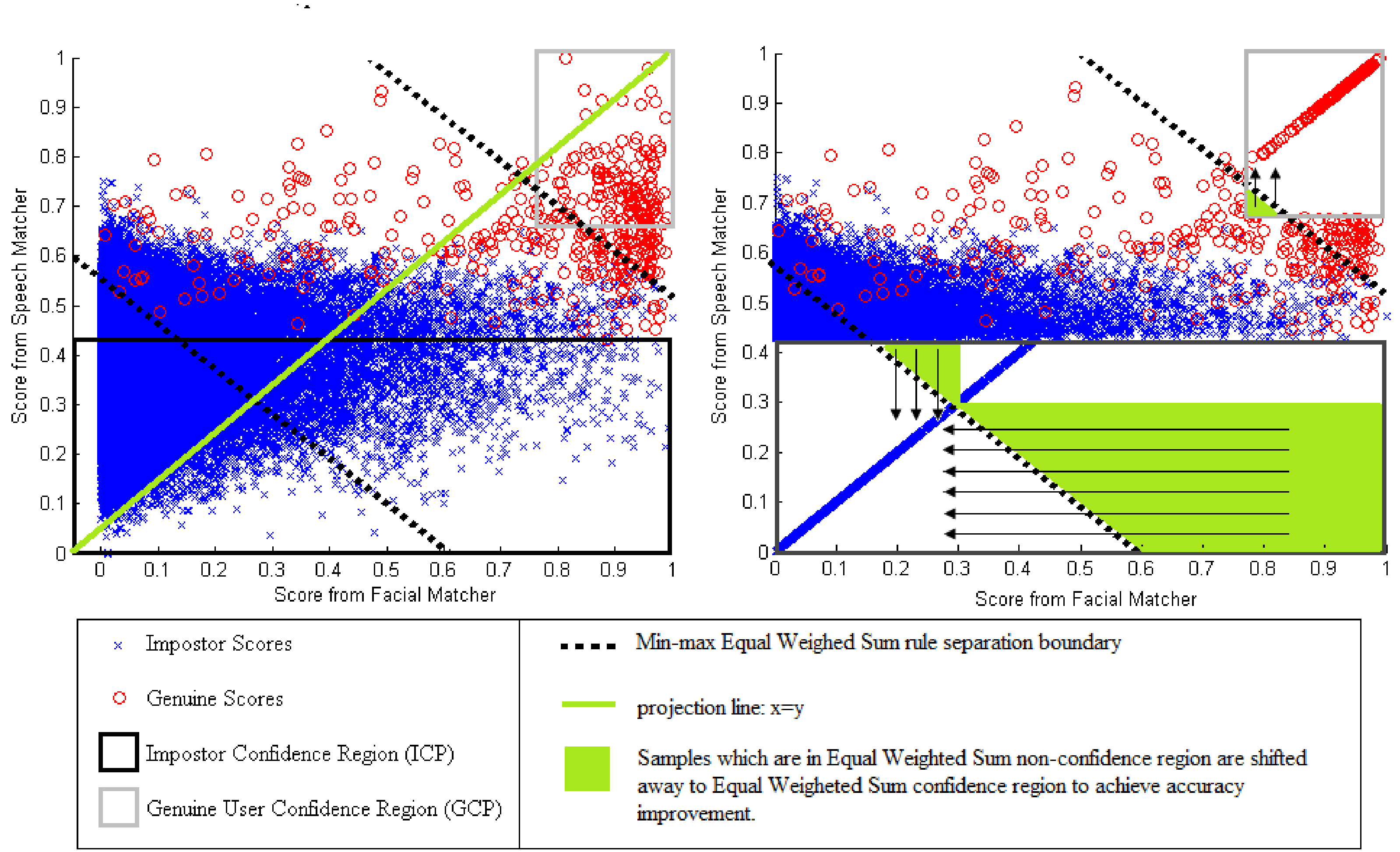


Fig 4-10. Xm2vts worst matchers testing fusion's score space with confidence partition assignments (left) and equivalent confidence sample projections (right).

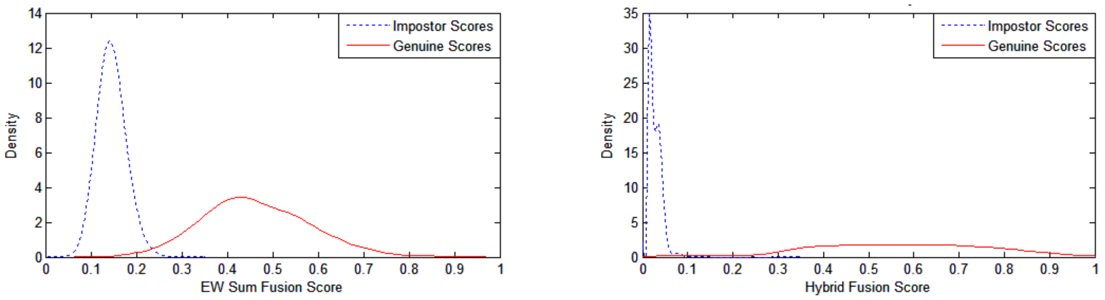


Fig 4-11. NIST-BSSR1 best matchers fusion score distribution densities: Min-max Sum (Left) and Hybrid fusion (right).

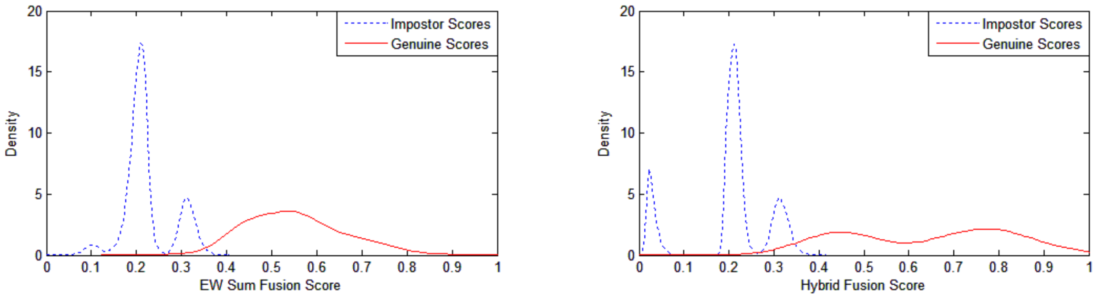


Fig 4-12. NIST-BSSR1 worst matchers fusion score distribution densities: Min-max Sum (Left) and Hybrid fusion (right).

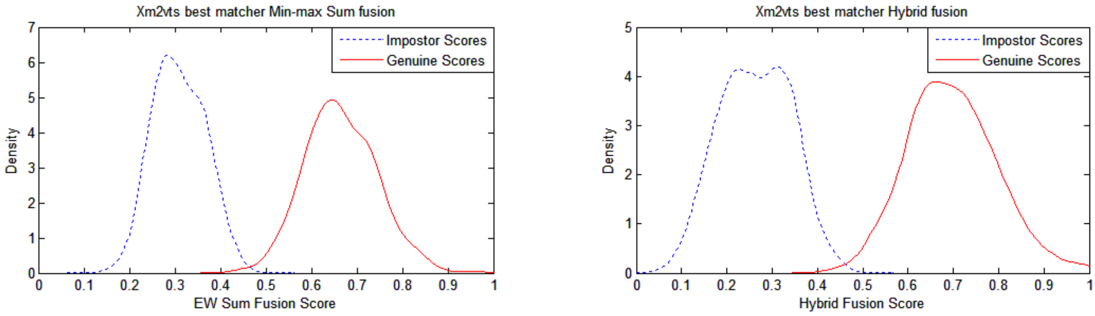


Fig 4-13. Xm2vts best matchers fusion score distribution densities: Min-max Sum (Left) and Hybrid fusion (right).

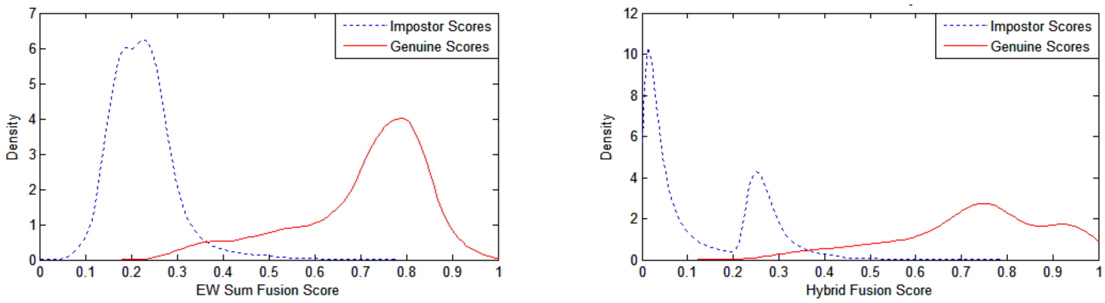


Fig 4-14. Xm2vts worst matchers fusion score distribution densities: Min-max Sum (Left) and Hybrid fusion (right).

4.5 Conclusions

Three conclusions can be made from the experimental results and further analysis in section 4.4:

1) Hybrid fusion is a feasible approach to improve the EW Sum rule. However, it depends on the accurate assignment of the CPs in order to reduce the non-confidence samples by projecting the samples to the confidence partitions so the verification improvement can be achieved. Prior knowledge of the score distributions is required for such an accurate assignment. Inaccurate assignment of CPs results in the degradation of the fusion approach.

2) The samples in the confidence region can be safely rejected or accepted whereas the sample acceptance or rejection in the non-confidence region has to depend on the security threshold. The smaller non-confidence region means less non-confidence samples therefore better fusion performance.

3) Non-confidence samples carry more information than the confidence samples for improving the fusion algorithm's performance.

Based on conclusions (2) and (3), the non-confidence related information is able to aid effective fusion. In the following chapters, the exploration of this information incorporating in the conventional fusion approaches is investigated. The first method uses the non-confidence region width as weighting parameter in the Weighted Sum rule. The second approach employs the GMM likelihood based fusion to the non-confidence region only. Both of these approaches are to further enhance the fusion algorithm's performance and usability.

5 TOWARDS A BEST LINEAR COMBINATION FOR MULTIMODAL BIOMETRICS FUSION

Owing to effectiveness and ease of implementation, the Sum rule has been widely applied in the biometric fusion research. Different matcher information has been used as weighting parameters in the Weighted Sum rule. In this work, a new parameter is devised to reduce the genuine/imposter distribution overlap. It is shown that the overlap region width can be used as effective weighting parameter to achieve the best generalisation performance compared to other commonly used matcher information. Furthermore, this conceptually simple and fast method is demonstrated achieving comparable performance to other conventional training based methods. This proposed method is tested using the 19 bimodal biometrics experiments conducted in chapter 2.

5.1 Introduction

The Sum Rule is one of the effective score level fusion approaches for biometric score level fusion research. Although it is a very simple algorithm, it outperforms some of the more complex fusion methods [43]. Weighting is used in the Sum rule to indicate the importance of each modality in the fusion. There are generally two different weighting schemes. The first one is to apply a same weight to all the scores generated by the same biometric matcher. This is equivalent to adjusting the separation boundary's gradient (e.g. using the biometric matcher performance measure as weighting parameter [106]). Another is to apply different weight to different users accordingly even the scores are generated by the same biometric matcher. This is equivalent to adjusting the score vector's position (e.g. using individual biometrics quality measure as weighting parameter [73], [76]). In fig. 5-1, P represents the bimodal biometrics score vector of a specific user. $P_{original}$ is relocated to $P_{user\ weight}$ after applying 0.5 and 1.5 weighting scalars to its bimodal biometrics. Such weighting scheme highly depends on the reliability of the user specific information. However, the training data that underpins such information is usually not sufficient or adequately representative [132].

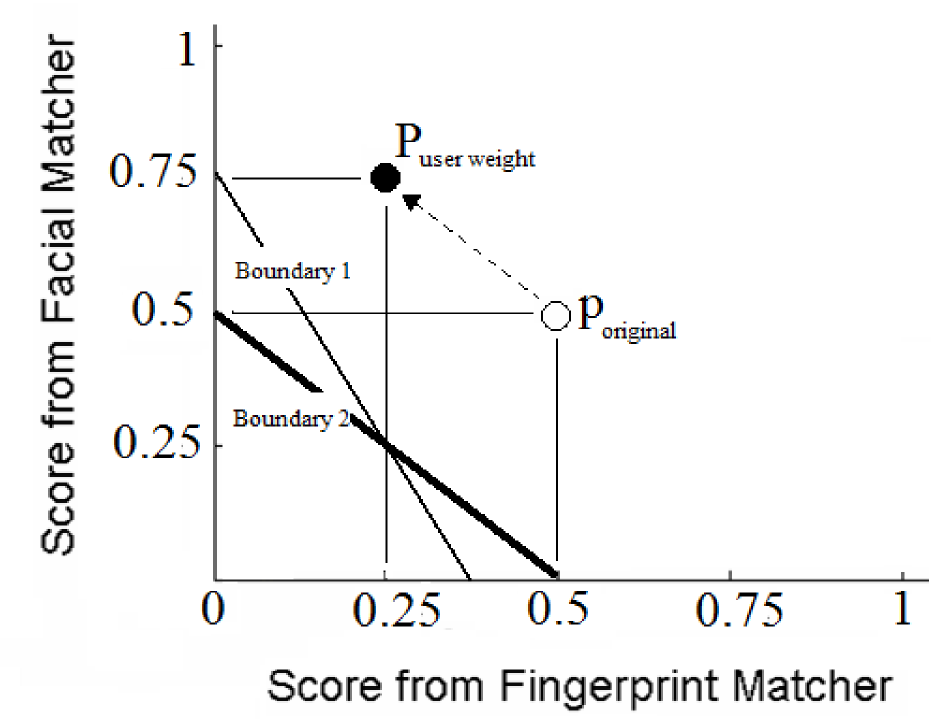


Fig 5-1. The difference between adjusting samples position and separation boundary's gradient in Weighted Sum rule fusion.

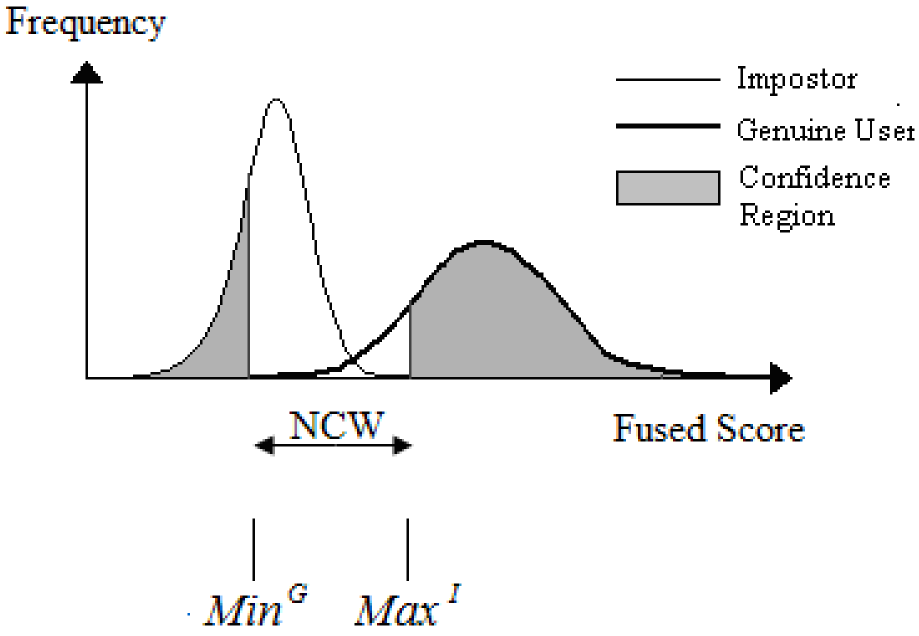
In contrast to the above, the gradient of the decision boundary is adjusted by using individual matcher information. Boundary 2's gradient in fig. 5-1 is adjusted to become Boundary 1's by assigning higher weight to the x-axis fingerprint matcher but a lower weight to the y-axis facial matcher. The weighting parameter under this weighting scheme can be easily obtained from a training set. The commonly used matcher information are EER and D-Prime [106]. For a given FAR, the best gradient that is used to achieve the best GAR can be found through exhaustive search [40], [112], [113]. Although exhaustive search guarantees a high verification rate, an optimal weighting depends on a special purpose algorithm that might be complex and or time consuming. In addition, for determining the optimal weights, training is required for every single operating point.

From chapter 3, it is concluded that the biometrics verification errors arise from the overlap region. By projecting away the overlap region's samples to the correct confidence region, the verification rate is improved. A smaller overlap region generally contains less non-confidence samples thus it produces less errors and results in a better generalisation performance over the entire operating points. Aside from projecting away the overlap region's samples, adjusting the gradient of the separation boundary will also reduces the overlap region. Therefore the aim of this work is to achieve the best linear combination by reducing the overlap region in a novel way by adjusting the gradient of the separation boundary. The Equal Weighted (EW), Equal Error Rate Weighed (EERW) and D-Prime Weighted (DPW) Sum rules are the commonly used methods. They are used as the baselines to evaluate this proposed method. The method is also further compared with the conventional best performing training based method to evaluate its effectiveness.

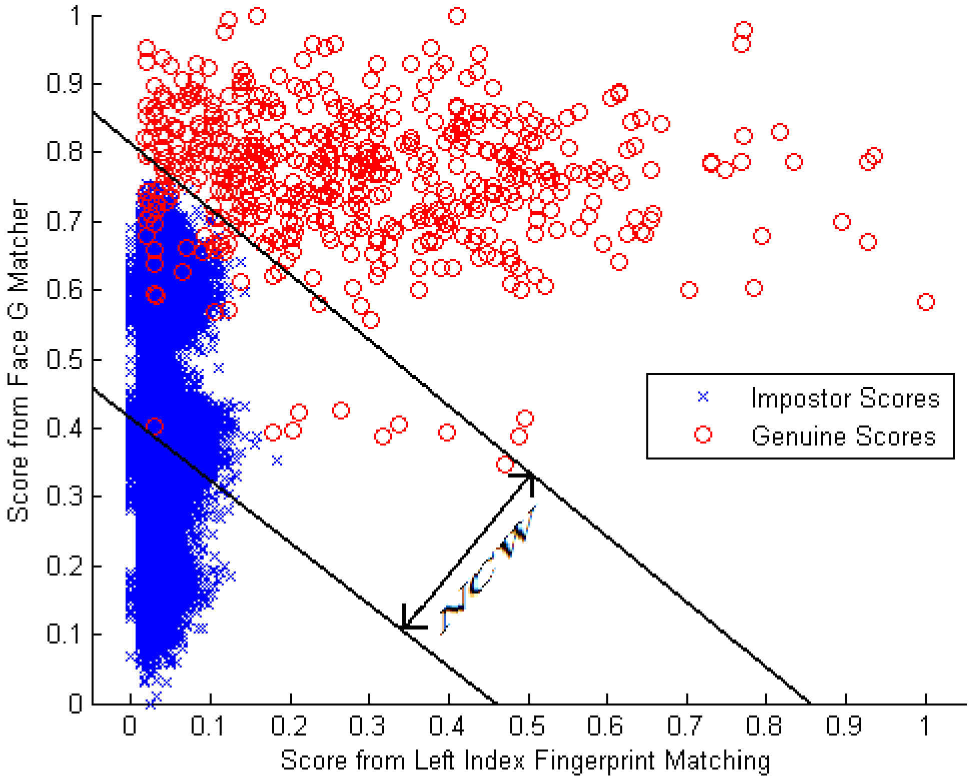
The details of the new method are given in the following section. Section 5-3 describes the experimental setup and the results analysis. Section 5-4 shows the comparisons of the proposed approach to other categories' best performing training based fusion approaches and to the state-of-the-art fusion algorithm in a higher dimension. This section is then followed by the conclusion.

5.2 Non-Confidence Width Weighted Sum Rule and Its Baselines

Fig. 5-2 illustrates a bimodal biometrics fusion that is viewed from one-dimensional and two-dimensional score spaces. There is a significant overlap region causing the difficulty to classify the claimant into genuine user or impostor groups. The reasons of this overlap region formation are discussed in section 1.4. The grey regions in (a), which are located outside the overlap part, are confidence regions where only a single class of users can be found. Therefore the confidence regions' samples can be safely rejected or accepted whereas the samples in the overlap region can only be classified with referring to the threshold boundary.



(a)



(b)

Fig 5-2. Bimodal biometrics linear fusion views: (a) one dimensional view and (b) two dimensional view.

The width of the overlap region is termed Non-Confidence Width (NCW). NCW can be determined from the difference between the maximum impostor score and the minimum genuine user score in a similarity based measurement biometrics system. It is shown in (5.1).

$$NCW = Max^{I} - Min^{G} \tag{5.1}$$

From fig. 5-2(b), the NCW, which is manipulated by the maximum impostor fused score, Max^{I}, and the minimum genuine user fused score, Min^{G}, can be adjusted by manipulating the separation boundary's gradient. It is depicted in fig. 5-3:

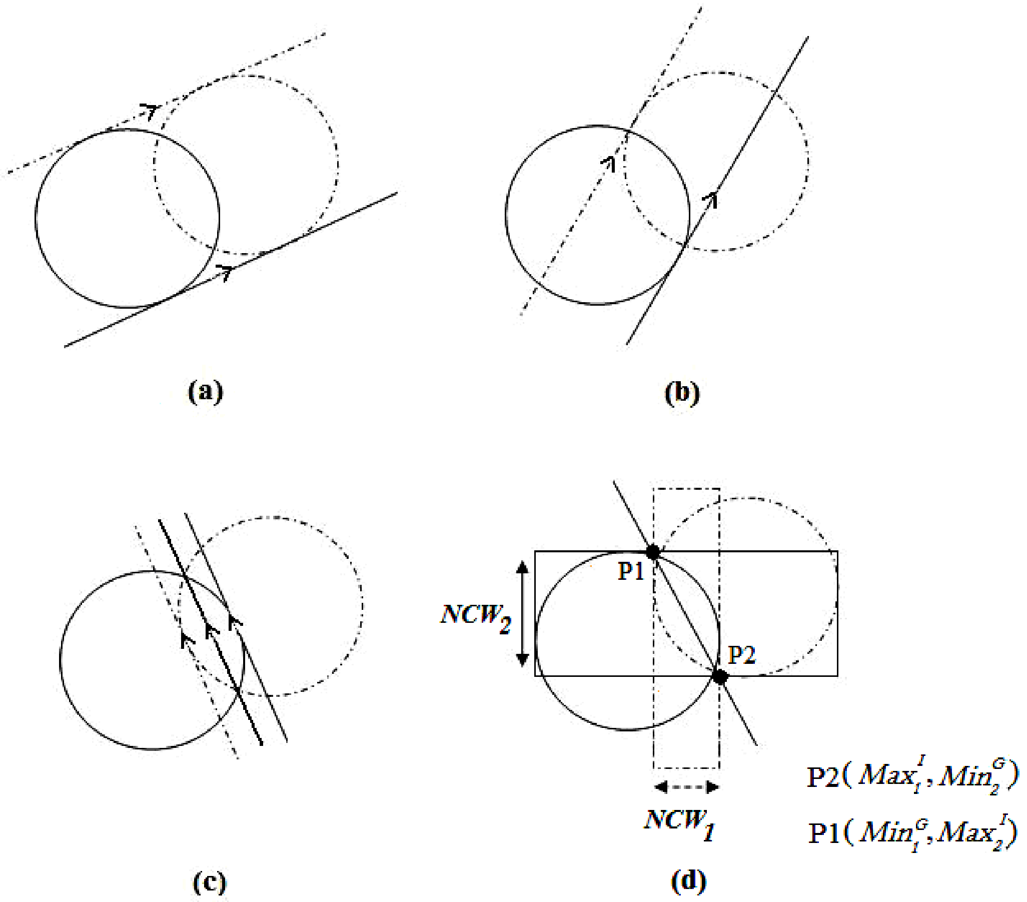


Fig 5-3. Reducing the NCW by adjusting the gradient of the linear boundary.

The two overlap circles in fig. 5-3 represent the approximated scatter of the genuine user and impostor scores. The straight lines are separation boundaries with different verification thresholds. Varying the decision threshold is a process of moving the boundary while preserving its gradient. As shown by (a), the circles' area between the separation boundaries is the non-confidence region and is at maximum. The non-confidence region is the area where the samples cannot be clearly classified by the separation boundary. However, by adjusting the gradient of the boundary, the non-confidence region can be reduced as shown in (b). When the boundary is parallel to the line connecting the intersection points of the two circles (P_1 and P_2), the non-confidence region is restricted to a minimum as shown in (c). Such adjustment enables the samples to be classified with minimum error so a better ROC can be obtained. Therefore it is desired that the separation boundary has the same gradient as the line connecting the circles' intersection points. As (d) depicts, this specific gradient *m* can be approximated by the NCW of the two matchers in bimodal biometrics fusion, where *m* in (d) equals to (5.2).

$$m = \frac{Min_2^G - Max_2^I}{Max_1^I - Min_1^G} \tag{5.2}$$

$$\frac{1}{NCW_1}x + \frac{1}{NCW_2}y = c \tag{5.3}$$

By using the common form of a linear equation, *y=mx+c*, (5.3) can be derived. c is an adjustable threshold for controlling of the boundary position. In this Weighted Sum rule, biometrics scores are inversely proportional weighted by their NCW. Their respective weights *W_k* can be obtained by applying (5.4) so that $\sum_{k=1}^{K} W_k = 1$, where K is the total matcher number. Therefore (5.3) can be rewritten as (5.5), S_i is the fused score for user i and $S'_{i,k}$ is the biometric score that is generated by matcher k for user i.

$$W_k = \frac{\frac{1}{NCW_k}}{\sum_{k=1}^{K} \frac{1}{NCW_k}} \tag{5.4}$$

$$S_i = \sum_{k=1}^{K} W_k S'_{i,k} \tag{5.5}$$

This method is referred to as the Non-Confidence Width Weighted Sum rule (NCWW). In contrast to conventional rule based fusion methods, the advantages of this method are that it is more robust in obtaining better fusion result and it does not need score normalisation. Moreover, the NCW information is very easy to obtain. Three commonly used weighting schemes are applied as baselines in the experiments to evaluate the effectiveness of the proposed method:

1) Equal Weighted: This weighting scheme assumes that the different modalities of biometrics have the same performance and therefore the scores are combined by using the same weight, W_k as shown in (5.6). However, this is not practical when the different biometrics are having different discrimination abilities.

$$W_{k=1\ldots K} = \frac{1}{K} \tag{5.6}$$

2) EER Weighted: EER is the error rate where false acceptance rate is equal to the false rejection rate. It is used to evaluate the performance of a biometrics matcher. Nonetheless, this single operating point measurement is not the only factor that determines the discrimination ability. For instance, a biometric matcher may have a better GAR at lowest FAR but with a poorer EER than another biometric matcher. Moreover, the EER of a matcher varies for different testing populations. Therefore EER is not a reliable parameter to weight different biometrics' matcher contribution. As shown in (5.7), the weight W_k assigning to the biometrics matcher k is inversely proportional to its EER.

$$W_k = \frac{\frac{1}{EER_k}}{\sum_{k=1}^{K} \frac{1}{EER_k}} \tag{5.7}$$

3) D-Prime Weighted: *d'* statistically measures the separation of impostor and genuine user biometrics scores. As depicted in (5.8), μ_k^G and μ_k^I are the mean genuine user and impostor scores for biometric matcher k where σ_k^G and σ_k^I are the standard deviations. The further separation of the two classes is desired. Therefore the associated matcher weight is directly proportional to its *d'* as shown in (5.9).

$$d'_k = \frac{\mu_k^G - \mu_k^I}{\sqrt{(\sigma_k^G)^2 + (\sigma_k^I)^2}} \tag{5.8}$$

$$W_k = \frac{d_k^{'}}{\sum_{k=1}^{K} d_k^{'}} \tag{5.9}$$

d' is a measure of two classes' score separation, which includes confidence samples in its computation. A statistical distance measurement without considering the confidence and non-confidence regions is not robust if used as weighting parameter. For instance, a biometrics with greater *d'* might have lower discrimination ability because of having a greater overlap region which causes the worse performance. Moreover, the sensitivity of the mean and the standard deviation to outliers might further degrade the robustness and performance.

5.3 Experiment Set Up and Results Analysis

It is desirable to examine the effectiveness of the proposed method before introducing further complexities. Therefore, although the proposed method can be generalised to higher dimensions, this investigation focuses on the performance of bimodal biometric fusion. The 19 bimodal biometrics fusion experiments that were introduced in chapter 3 are once again used to test the proposed method and its baselines. All the score sets are from the truly multimodal databases: the NIST-BSSR1 multimodal database and the Xm2vts benchmark database. Since no matcher information is given, each of the databases is evenly divided into two sets. Using the Xm2vts database's training and testing set defined by the author restricts the number of experiments, i.e. it can be carried out only once. Therefore, to examine the robustness of the proposed method in this database, the testing and training scores are randomly chosen from this database to form equal training and testing sets. The required weighting parameters are obtained from the first set and the remaining samples are used for testing. Such equal size partition for testing and training enables one to evaluate the proposed approach without having a bias for the testing and training sets size.

For the weighting schemes acting as baselines, before the Weighted Sum rule is applied, the biometrics scores are normalised using Min-max normalisation which is detailed in table 2-1. Each of the experiments were repeated 30 times with different random partitions of the databases for statistical reason. For numerical comparison, four operating points captured over the ROC are listed in the tables below. The average results are reported and their standard deviations are shown in brackets in table 5-1 and 5-2. Best results are reported solely based on average performance (shown as shaded figure) and using the t-test, a statistical significance hypothesis test at 95% confidence interval (including the shaded and bordered figures). M1 and M2 is the first and second biometrics matcher involving in the experiments. These matchers' details can be found in section 3.3.

Exp. No.	GAR(%) at FAR equals to 0.001%					
	M1	M2	EW	EERW	DPW	NCWW
1	70.99(2.94)	56.37(3.56)	90.66(1.20)	84.78(3.21)	83.34(2.46)	92.36(1.62)
2	72.33(2.67)	61.07(2.62)	92.76(1.23)	91.19(1.63)	87.49(1.76)	89.07(1.90)
3	82.65(1.35)	55.17(4.82)	92.50(1.25)	91.48(2.12)	83.49(2.78)	95.40(1.18)
4	82.73(1.78)	60.88(2.30)	95.00(0.80)	94.94(1.04)	90.27(1.67)	92.34(1.94)
Total Best Result	0	0	2	0(1)	0	2

(a)

Exp. No.	GAR(%) at FAR equals to 0.01%					
	M1	M2	EW	EERW	DPW	NCWW
1	76.82(1.61)	72.86(3.31)	94.27(1.21)	90.71(2.51)	89.59(2.43)	94.83(1.33)
2	77.76(1.96)	68.36(2.18)	94.39(1.05)	92.76(1.38)	90.45(1.55)	93.51(1.50)
3	84.79(1.24)	73.01(2.50)	96.63(0.99)	96.19(1.43)	92.13(1.46)	97.52(0.83)
4	84.66(1.82)	68.08(1.70)	96.80(0.73)	96.95(0.91)	93.26(1.38)	96.26(0.98)
Total Best Result	0	0	1(2)	1	0	2

(b)

Exp. No.	GAR(%) at FAR equals to 0.1%					
	M1	M2	EW	EERW	DPW	NCWW
1	82.95(1.55)	83.97(1.78)	97.31(0.74)	95.67(1.34)	95.45(1.20)	97.66(0.86)
2	83.54(1.49)	77.84(1.92)	96.49(0.89)	95.58(1.08)	93.47(1.38)	96.41(0.84)
3	89.82(1.11)	83.74(1.77)	97.73(0.81)	97.68(0.88)	96.60(1.14)	99.07(0.57)
4	89.94(1.50)	77.82(2.17)	97.94(0.65)	98.24(0.66)	95.96(0.91)	98.51(0.56)
Total Best Result	0	0	1	0	0	3(4)

(c)

Exp. No.	EER					
	M1	M2	EW	EERW	DPW	NCWW
1	8.37(1.04)	4.45(0.45)	1.26(0.36)	1.74(0.47)	1.86(0.41)	1.04(0.39)
2	8.39(0.84)	5.75(0.77)	1.76(0.44)	2.44(0.74)	2.99(0.67)	1.39(0.37)
3	4.82(0.57)	4.45(0.48)	0.51(0.20)	0.57(0.24)	1.05(0.30)	0.38(0.19)
4	4.94(0.70)	5.77(0.73)	1.34(0.48)	0.99(0.49)	2.52(0.61)	0.49(0.19)
Total Best Result	0	0	0	0	0	4

(d)

** The shaded figures represent the best average fusion results. The figures with border represent the average fusion results that are not significantly different from the best results tested using t-test at 95% confidence interval.*

Table 5-1. Weighted Sum fusion performance (average GAR) in four NIST-BSSR1 bimodal biometrics fusion experiments under (a) FAR=0.001% (b) FAR=0.01% (c) FAR=0.1% and (d) EER.

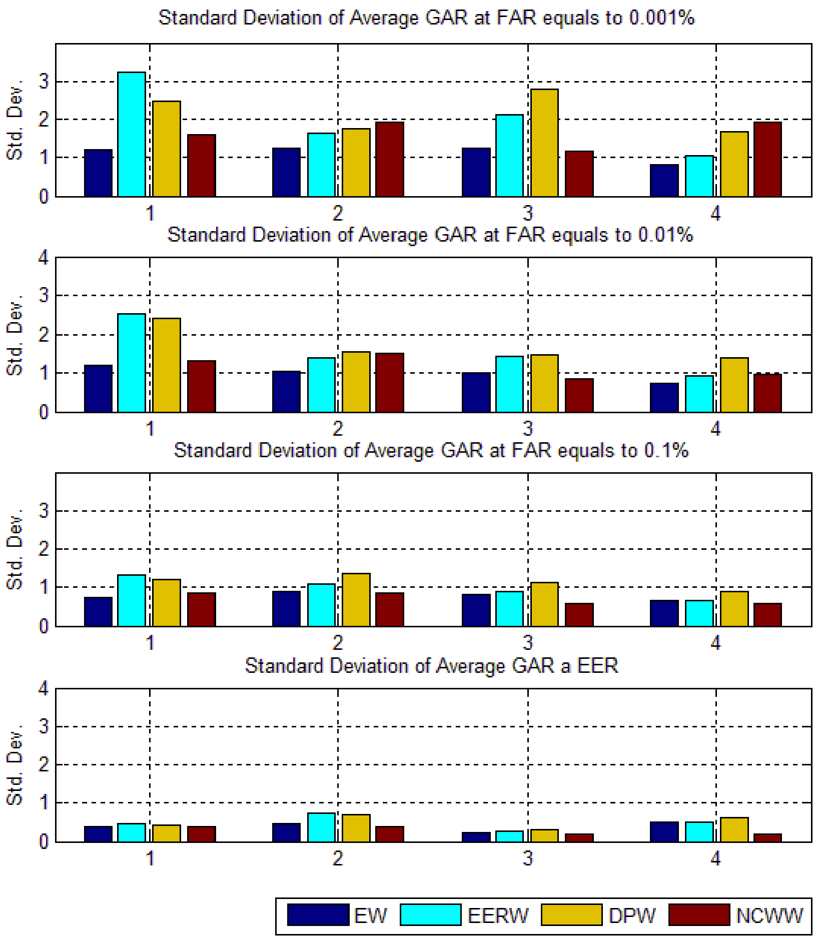


Fig 5-4. NIST-BSSR1 bimodal biometrics fusion experiments: 30 partitions average result's standard deviation to show the fusion performance consistency.

In NIST-BSSR1 experiments, all the fusion strategies outperform their respective single best biometrics. It is obvious that the NCWW has a better generalisation performance than the compared schemes. It generally achieves more best fusion results (GAR) over all compared operating points. Under the lowest operating point, NCWW outperforms baselines in the range of 1.7%~11.9% with standard deviation equal to or less than 1.62. However it does not perform well in the experiments involved with Fg (Exp. no. 2 and 4) under this lowest operating point. Comparing the results outperforming the NCWW solely based on the average value shows that the performance differences are in the range of 2.1% ~ 3.7%. As seen in fig. 3-2, Fg's score distribution has a long tail and multiple components of Gaussian within the distribution. Therefore the two circles model assumed in fig. 5-3 may not suitably fit experiments involving Fg because there might be several clusters available within the score space instead of a single one. This causes the high variation of NCW to be used in the Weighted Sum rule to degrade this fusion approach. This unreliability is reflected in the performance of the standard deviation in fig. 5-4. In this figure, this approach has the highest standard deviations under the lowest operating point in Exp. 2 and 4. However, for the consecutive operating points, the NCWW performance differences to the other weighting scheme that achieves best fusion result are just less than 0.7%. NCWW outperforms the other methods in most of the experiments and operating points and hence demonstrates the robustness of this fusion approach.

Exp. No.	GAR(%) at FAR equals to 0.002%					
	M1	M2	EW	EERW	DPW	NCWW
1	5.14(13.93)	57.02(2.86)	93.74(0.88)	94.64(0.90)	92.80(1.15)	96.56(0.93)
2	6.08(14.52)	15.80(1.63)	90.05(0.93)	87.20(1.74)	88.67(1.44)	88.58(1.05)
3	2.69(4.61)	38.34(5.49)	90.43(1.51)	88.87(1.16)	89.94(1.23)	85.87(2.51)
4	60.01(6.07)	56.13(2.21)	93.70(1.07)	76.88(4.94)	93.52(1.01)	91.76(1.77)
5	59.77(4.89)	15.98(1.64)	79.22(2.11)	80.25(2.77)	80.62(2.05)	79.11(2.20)
6	59.39(5.23)	38.89(5.55)	82.31(2.93)	82.22(3.28)	82.19(3.02)	83.32(2.90)
7	77.54(4.86)	56.77(3.07)	97.22(0.67)	94.37(2.35)	97.33(0.68)	96.27(1.30)
8	78.87(3.39)	16.41(1.77)	83.04(3.37)	88.99(2.32)	87.18(2.62)	84.44(3.19)
9	79.58(3.33)	39.43(6.11)	92.60(1.54)	90.26(2.56)	93.11(1.79)	91.71(1.95)
10	1.22(1.03)	57.12(2.63)	78.34(2.39)	92.37(3.04)	81.29(2.38)	94.37(4.32)
11	1.60(2.62)	16.07(2.05)	67.05(1.69)	54.91(3.81)	65.58(2.33)	77.12(4.05)
12	1.12(1.76)	38.05(5.85)	66.18(4.31)	58.68(4.44)	66.74(4.20)	83.92(2.58)
13	0.28(1.02)	56.80(3.29)	67.12(3.25)	86.39(3.63)	78.4(2.98)	86.85(3.17)
14	0.06(0.10)	16.06(1.64)	50.14(2.56)	49.35(3.06)	57.56(2.37)	63.16(3.52)
15	0.17(0.49)	35.81(4.60)	52.43(4.04)	56.99(5.79)	66.17(4.13)	78.70(2.40)
Total Best Result	0	0	3(6)	1(4)	3(6)	8

(a)

Exp. No.	GAR(%) at FAR equals to 0.01%					
	M1	M2	EW	EERW	DPW	NCWW
1	78.06(11.4)	68.41(2.63)	94.23(0.82)	95.22(1.10)	93.49(1.10)	98.04(0.52)
2	79.25(7.14)	25.77(2.05)	92.20(0.96)	90.41(1.42)	91.66(1.26)	91.78(1.40)
3	75.99(12.01)	49.75(2.64)	92.91(0.97)	89.87(1.08)	91.19(1.30)	93.91(1.46)
4	68.80(2.69)	67.45(2.28)	95.60(0.80)	87.52(2.91)	96.02(0.77)	96.22(0.59)
5	69.09(2.04)	25.45(2.96)	85.70(1.72)	86.34(1.44)	86.28(1.61)	85.73(1.49)
6	69.02(2.13)	50.19(1.98)	89.48(1.29)	89.24(1.38)	89.52(1.29)	88.91(1.18)
7	88.61(1.69)	68.88(2.66)	98.38(0.42)	98.00(0.59)	98.36(0.43)	98.38(0.42)
8	88.92(1.02)	25.19(1.83)	92.61(1.62)	94.02(0.71)	94.81(0.73)	93.33(1.77)
9	89.01(1.55)	50.18(2.56)	95.30(1.12)	94.82(1.14)	95.64(0.83)	94.79(1.19)
10	25.32(7.98)	68.83(2.64)	85.07(1.71)	95.92(1.46)	86.92(2.28)	97.34(0.52)
11	25.63(10.66)	25.44(1.93)	75.09(2.72)	66.40(3.25)	73.35(2.81)	86.31(1.46)
12	26.68(11.09)	49.60(1.65)	78.10(2.00)	73.04(2.51)	78.29(2.13)	88.86(1.91)
13	5.24(3.82)	68.68(2.49)	73.63(1.98)	92.58(1.11)	85.39(1.50)	92.64(1.10)
14	4.43(1.83)	25.68(2.01)	58.56(2.43)	57.72(2.73)	65.33(2.47)	73.53(2.54)
15	4.27(2.49)	49.40(1.78)	62.84(2.17)	65.99(3.43)	72.26(2.54)	83.52(1.80)
Total Best Result	0	0	2(5)	1(3)	3(6)	10(12)

(b)

Exp. No.	GAR(%) at FAR equals to 0.1%					
	M1	M2	EW	EERW	DPW	NCWW
1	93.25(0.92)	88.71(1.05)	95.74(0.85)	97.29(1.25)	95.13(0.96)	99.15(0.30)
2	93.54(0.85)	55.25(1.83)	95.20(0.69)	94.15(0.90)	94.73(0.79)	97.47(0.59)
3	93.08(0.76)	67.25(2.30)	94.97(0.69)	93.83(0.80)	94.36(0.93)	98.11(0.43)
4	82.14(1.27)	88.90(1.03)	98.57(0.39)	98.20(0.65)	98.78(0.37)	99.00(0.33)
5	82.20(1.32)	54.59(1.83)	93.80(0.76)	93.82(0.81)	93.95(0.72)	93.72(0.76)
6	81.97(1.13)	67.66(1.29)	95.49(0.71)	95.41(0.76)	95.52(0.70)	95.46(0.78)
7	94.31(0.76)	88.96(1.56)	99.38(0.22)	99.32(0.22)	99.38(0.22)	99.38(0.22)
8	94.49(0.61)	54.84(1.87)	97.05(0.62)	97.51(0.63)	97.59(0.55)	97.25(0.63)
9	94.29(0.63)	67.64(1.71)	98.11(0.49)	97.55(0.63)	97.97(0.53)	97.87(0.51)
10	80.16(1.42)	88.76(1.01)	92.61(1.09)	98.61(0.70)	93.53(1.31)	99.38(0.25)
11	80.83(1.35)	55.44(1.83)	90.66(0.83)	87.30(1.22)	89.96(1.05)	95.42(0.54)
12	80.29(1.56)	67.61(1.65)	89.43(0.97)	87.60(1.30)	89.63(1.32)	96.88(0.63)
13	50.78(2.90)	88.87(1.11)	82.25(1.39)	98.25(0.36)	91.09(1.63)	97.95(0.40)
14	51.45(2.85)	55.74(1.58)	77.24(1.55)	76.39(2.32)	81.20(1.91)	88.32(1.03)
15	51.28(3.23)	67.51(1.44)	76.41(1.75)	78.09(2.48)	83.72(2.31)	92.41(1.18)
Total Best Result	0	0	2(4)	1(5)	4(5)	10(12)

(c)

Exp. No.	EER					
	M1	M2	EW	EERW	DPW	NCWW
1	1.73(0.28)	0.99(0.09)	0.84(0.13)	0.59(0.18)	1.06(0.22)	0.32(0.09)
2	1.67(0.30)	5.83(0.42)	1.07(0.21)	1.42(0.30)	1.29(0.28)	0.72(0.15)
3	1.78(0.24)	4.42(0.38)	1.10(0.15)	1.49(0.24)	1.31(0.19)	0.60(0.18)
4	4.44(0.45)	0.98(0.09)	0.41(0.09)	0.50(0.13)	0.40(0.11)	0.42(0.15)
5	4.39(0.32)	5.78(0.31)	1.38(0.22)	1.34(0.16)	1.35(0.19)	1.38(0.21)
6	4.39(0.33)	4.42(0.33)	1.16(0.17)	1.19(0.20)	1.18(0.17)	1.01(0.21)
7	1.68(0.24)	0.96(0.10)	0.45(0.13)	0.40(0.12)	0.46(0.15)	0.44(0.14)
8	1.61(0.23)	5.76(0.44)	1.02(0.20)	1.01(0.22)	0.95(0.15)	0.99(0.19)
9	1.68(0.22)	4.43(0.45)	0.69(0.17)	0.94(0.29)	0.68(0.15)	0.74(0.15)
10	3.14(0.38)	0.97(0.10)	1.04(0.13)	0.36(0.10)	0.87(0.15)	0.32(0.08)
11	3.03(0.36)	5.81(0.37)	1.26(0.13)	1.78(0.25)	1.37(0.17)	0.65(0.04)
12	3.15(0.32)	4.41(0.37)	1.49(0.14)	1.78(0.22)	1.47(0.18)	0.81(0.19)
13	5.48(0.33)	0.97(0.12)	2.34(0.24)	0.48(0.09)	1.02(0.12)	0.52(0.09)
14	5.37(0.42)	5.72(0.38)	3.12(0.26)	3.20(0.35)	2.29(0.24)	1.57(0.23)
15	5.58(0.47)	4.36(0.37)	3.35(0.23)	2.99(0.39)	2.13(0.24)	1.53(0.24)
Total Best Result	0	0	0(5)	3(4)	3(4)	9(14)

(d))

** The shaded figures represent the best average fusion results. The figures with border represent the average fusion results that are not significantly different from the best results tested using t-test at 95% confidence interval.*

Table 5-2. Weighted Sum fusion performance (average GAR) in fifteen Xm2vts bimodal biometrics fusion experiments under (a) FAR=0.002% (b) FAR=0.01% (c) FAR=0.1% and (d) EER.

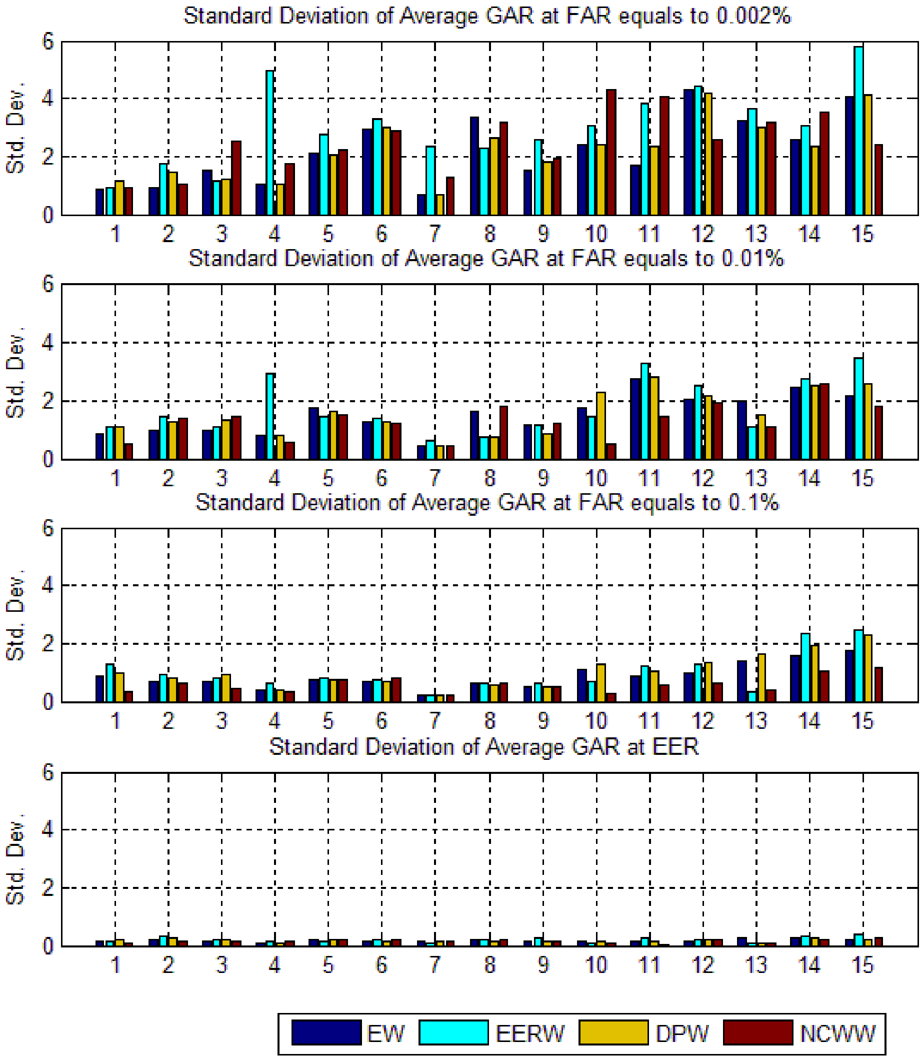


Fig 5-5. Xm2vts benchmark database bimodal biometrics fusion experiments: 30 partitions average result's standard deviation to show the fusion performance consistency.

Most of the fusion results using Xm2vts outperform the single best biometrics. Again, as the total best results shown in table 5-2, NCWW significantly provides more best GAR than the other methods under all operating points. Amongst the best results based on average value, NCWW outperforms the other fusion results with differences up to 26.27%. At lowest operating point, for NCWW's results that are not outperforming others, the performance differences are in the range of 0.09%~4.56% compared to the best achieved fusion results. For the consecutive operating points, this is greatly reduced. Except Exp. no. 8, which has a performance difference of

1.48% at second lowest operating point, the performance differences of the other experiments and operating points are equal to or less than 0.85%.

NCWW outperforms with significant difference when the experiments involve face matchers, F4 and F5. Referring to fig. 3-5, it can be seen that two different classes' scores are around -1 to 1. By applying an inverse tangent to these scores prior to the Min-max normalisation, the author in [56] demonstrates a fusion improvement for the EW Sum rule. This trial once again depicts the importance of choosing the appropriate normalisation algorithm on a case by case basis for rule based fusion, whereas the proposed method does not has to make this choice to achieve the best result. Fig. 5-5 depicts the average performance standard deviation. All the Weighed Sum rules' results show a relative high variation in the experiments involving F4 and F5 (Exp. no. 10 ~15). This is due to the significant spread but low density of the non-confidence region that causes the significant difference of non-confidence score scatters through different partition trials.

5.4 NCWW Comparison to Other Conventional Approaches and in Higher Dimension.

Based on the same experimental set-up, NCWW results obtained in this chapter are directly comparable to the results obtained by conventional fusion approaches in chapter 3. Table 5-3 and fig. 5-6 show comparisons at the lowest operating point. In the table, the shaded figures represent the best achieved fusion results and the figures with border are the best results that are in 95% confidence interval.

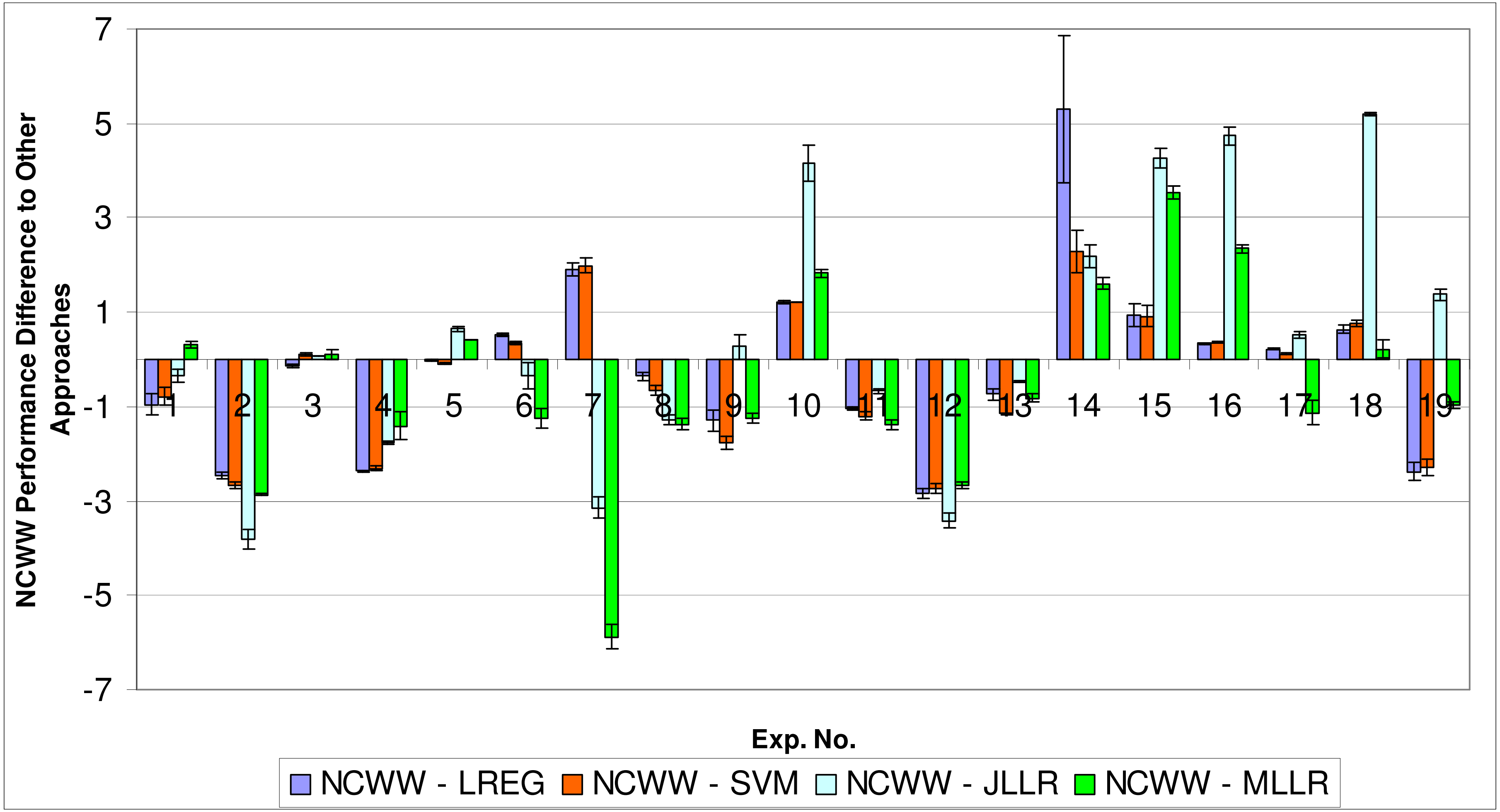
Fig. 5-6 shows the performance differences of NCWW and the compared approaches. In this figure, the positive value represents that NCWW outperforms the compared approach. Table 5-3 illustrates that there are 7 out of 19 experiments (Exp. no. 3, 5, 10, 14, 15, 16 and 19 in fig. 5-6.) where NCWW performs comparable or better, with statistical significance over the rest. Such comparable or outperforming result's differences are in the range of -0.15% ~ 5.29%. Compared to the best result achieved by the other approaches, there are 6 experiments (Exp. no. 1, 6, 8, 11, 13 and 17 in fig. 5-6.) where NCWW performs less effective in the moderate range of -0.96% ~ -1.38%. For the rest of the experiments, NCWW is significantly underperforming. From fig. 5-6, it can be seen that these are experiments 2, 4, 7, 9, 12 and 19. The best approaches outperform NCWW in the range of -1.77% ~ -5.88%.

Based on the analysis above, and considering that NCWW is a conceptual simple and fast, parameters can be easily obtained from the training samples without using a specific algorithm and does not require careful modelling of the score distributions, NCWW is an alternative fusion approaches to the other conventional state-of-the-art approaches.

NIST-BSSR1 Exp. No.	NCWW	LREG	SVM	JLLR	MLLR
1	92.36(1.62)	93.32(0.41)	93.14(0.50)	92.70(0.88)	92.05(1.17)
2	89.07(1.90)	91.53(1.51)	91.74(1.52)	92.88(0.71)	91.93(2.03)
3	95.40(1.18)	95.55(0.96)	95.30(1.01)	95.33(1.17)	95.30(1.72)
4	92.34(1.94)	94.71(2.07)	94.65(1.62)	94.12(2.12)	93.75(3.54)
Xm2vts Exp. No.	**NCWW**	**LREG**	**SVM**	**JLLR**	**MLLR**
1	96.56(0.93)	96.59(1.04)	96.65(1.03)	95.91(1.21)	96.14(1.00)
2	88.58(1.05)	88.07(0.83)	88.22(0.85)	88.93(2.57)	89.84(2.19)
3	85.87(2.51)	83.95(3.27)	83.89(3.37)	89.01(3.72)	91.75(1.06)
4	91.76(1.77)	92.12(1.26)	92.41(1.21)	93.05(1.18)	93.13(1.10)
5	79.11(2.20)	80.40(0.96)	80.88(1.48)	78.85(3.59)	80.35(1.61)
6	83.32(2.90)	82.10(2.73)	82.10(2.87)	79.17(4.94)	81.50(3.38)
7	96.27(1.30)	97.30(1.13)	97.47(0.85)	96.94(1.02)	97.65(0.78)
8	84.44(3.19)	87.28(2.60)	87.19(2.60)	87.86(2.38)	87.11(2.79)
9	91.71(1.95)	92.45(1.30)	92.87(1.78)	92.18(2.12)	92.53(1.43)
10	94.37(4.32)	89.08(12.85)	92.08(6.71)	92.20(3.00)	92.76(4.96)
11	77.12(4.05)	76.19(5.32)	76.21(5.26)	72.86(2.94)	73.59(3.36)
12	83.92(2.58)	83.58(2.46)	83.56(2.49)	79.18(3.63)	81.58(2.19)
13	86.85(3.17)	86.63(3.10)	86.73(3.15)	86.34(2.83)	87.98(1.82)
14	63.16(3.52)	62.52(3.92)	62.40(3.93)	57.96(3.78)	62.94(2.43)
15	78.70(2.40)	81.08(3.43)	80.99(3.36)	77.32(1.72)	79.67(2.10)

** The shaded figures represent the best average fusion results. The figures with border represent the average fusion results that are not significantly different from the best results tested using t-test at 95% confidence interval.*

Table 5-3. NCWW comparisons with conventional fusion methods at lowest operating points.



* *The error bars are plotted at +/- 1 standard error (sample no. = 30).*

Fig 5-6. NCWW performance difference (GAR at lowest FAR) to other conventional approaches in 19 experiments.

Subsequently, the approach is applied to higher dimensional score space to further examine its effectiveness and compared to the state-of-the-art method. As an example, the Gaussian Mixture Modelling likelihood ratio based fusion in [56] is used for comparison. In this work, all the matchers in both databases are fused respectively in a multi-biometric fusion experiment. In NIST-BSSR1, four biometrics sources (Fli, Fri, Fg and Fc) are fused whereas eight sources (F1~F5 and S6~S8) are combined in the Xm2vts benchmark database. Table 5-4 shows the comparison results.

The compared fusion method in [56] requires the fitting algorithm presented in [138] to search for a suitable component numbers for GMM. Both the modelling and fitting processes are complicated and time consuming. This is especially the case when dealing with large scale biometrics data evaluation. NCWW only requires the minimum genuine user score and maximum impostor score as input parameters. Moreover, it involves only simple addition, therefore it is conceptually simple and very easy to implement. The results show that this simple approach outperforms or performs comparable to the complicated fusion method in [56] in the higher dimensional fusion experiments.

		Likelihood Ratio Based Fusion	NCWW
NIST-BSSR1	Mean GAR at 0.01% FAR	99.1%	99.2%
	95% Confidence Interval on increase in GAR at 0.01% FAR	[13.5%, 14.0%]*	[14.0%, 14.4%]**
Xm2vts Benchmark Database	Mean GAR at 0.01% FAR	98.7%	99.0%
	95% Confidence Interval on increase in GAR at 0.01% FAR	N/A	N/A

* *With refer to best single matcher's performance at GAR equals to 85.3%.*

** *With refer to best single matcher's performance at GAR equals to 85.0%.*

Table 5-4. Higher dimensional fusion comparisons of NCWW to [56].

5.5 Conclusions

It was demonstrated that NCWW consistently outperforms the conventional Weighted Sum rule in most of the experiments. While it is possible to further enhance the conventional rule based fusion methods, by carefully choosing the normalisation algorithm on a case by case basis, NCWW can achieve the best result without having to make this choice or using any normalisation algorithm.

NCWW performs comparably well to the conventional state-of-the-art fusion approaches not only in bimodal biometrics fusion but also at higher dimensional biometrics fusion. However such performance can only be assured provided that the biometrics scores' distributions are close to the assumed model in fig. 5-3. In this case, NCW, as defined in (5.1), can be effectively used as weighting reference.

While the conventional approaches provide comparable fusion results, NCWW has the advantage that it is conceptually simple and easy to implement and to understand. Its parameters can be easily obtained from the training samples without the need of using specific algorithms and does not need to model the biometrics score distributions. Furthermore, the fusion can be done almost instantly as it only involves simple arithmetic and in fact training is not required.

Basing the NCW on Max_k^I and Min_k^G difference alone is sensitive to outliers and may lead to unreliability and degradation of this fusion approach. However, NCW can be extended to include the corresponding density and other overlap region's information. By doing this, unreliability and degradation can be reduced and a further improvement can be achieved.

6 A NEW APPROACH TO LIKELIHOOD RATIO BASED MULTIMODAL BIOMETRICS SCORE FUSION

Density based score fusion for biometrics has the advantage of not needing score normalisation or tuning of the parameters and additional biometric information can be easily incorporated. It is able to consistently achieve a high verification rate at any operating point, provided the score densities are estimated accurately. The Gaussian Mixture Model has been successfully used to estimate the biometrics score density of impostor and genuine user. However, the estimation accuracy highly relies on the selected component numbers and this selection can be a very time consuming process, especially when encountering a significant amount of training samples. This restricts the usability of this fusion approach. In this chapter, only the non-confidence samples are used to train the Gaussians Mixture Model by applying random component numbers. By doing this, not only a comparable verification rate can be achieved in most of the experiments without having to use the component number searching algorithm, but the method also demonstrates a considerable reduction in training time.

6.1 Introduction

Likelihood Ratio (LLR) based fusion is one of the density based score fusions. It transforms the score into a density before it employs the LLR to make a decision. One of the main advantages of this method is that it enables additional probability based biometrics information to be incorporated into the algorithm directly without having to modify it [56]. Furthermore, a multimodal biometric application covering a large scale population has a better chance of overcoming the missing data problem as mentioned in section 2.2.3.3. By assigning unity probability to the missing biometrics value, the missing data does not influence the use by a claimant who does not possess sufficient biometrics traits [57].

Unlike the linear separation boundary, e.g. in the rule based approach and some of the classification based fusion methods, density based fusion achieves optimal performance at any desired operating point directly, provided the density estimation is accurate. By transforming the score into density, the score normalisation which has to be carefully chosen by the rule based fusion method, is not needed.

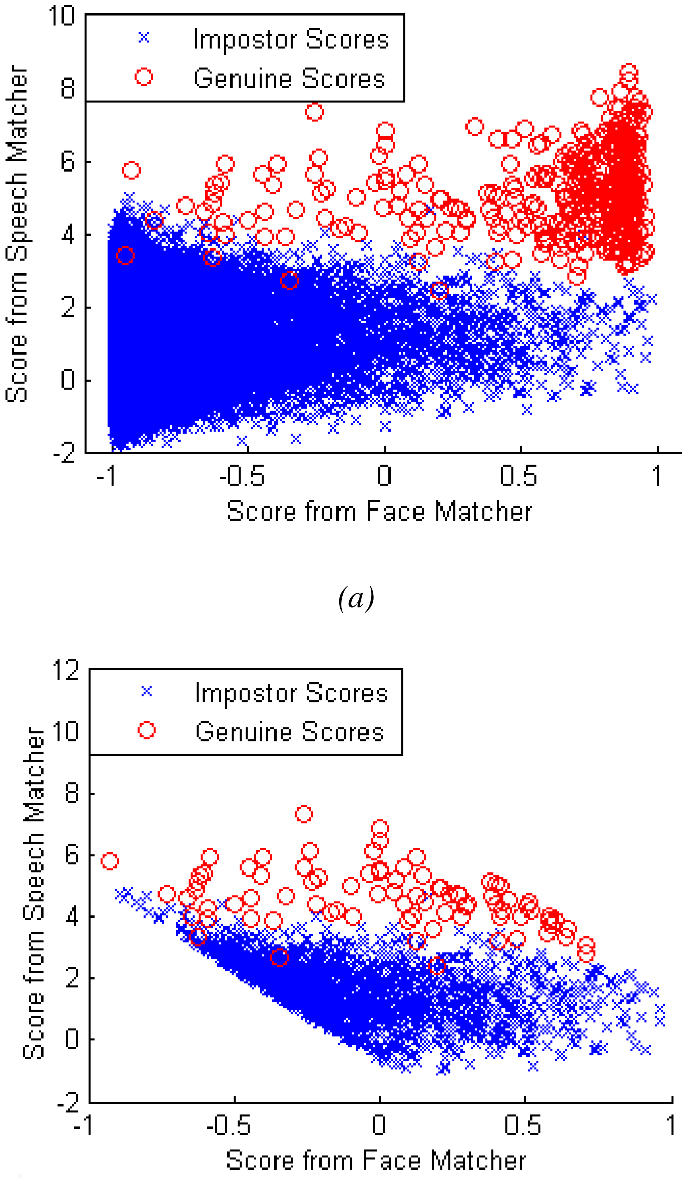
Eight state-of-the-art fusion strategies are chosen in [133] for comparison. They are from three different fusion categories and are selected based on their reported performance. Their work confirms that the product of LLR fusion is the most accurate method. However, it is very important to estimate the genuine user and impostor's densities reliably and accurately, because the LLR fusion performance highly depends on these estimations. This method is therefore complicated to implement because accurate curve fitting and density estimation is needed. In this work, Kernel of Density Estimator (KDE) is used. Gaussian Mixture Model (GMM) is successful in modelling density in [150] and [151] and Nandakumar et. al. demonstrate that it is not only effective in modeling the score densities, but it is also easier to implement [56].

Regardless of the fact that the parameter for tuning is not required for GMM, choosing the precise component numbers is a critical issue. The modeling accuracy is very sensitive to this parameter. Higher component number causes over fitting but lower component number results in a less accurate estimation. The work in [56] determines the component numbers automatically by using the state-of-the-art fitting algorithm available from [138]. However, it is a very time consuming searching process.

For practical implementation, one not only has to consider the availability of this searching algorithm, the required searching time restricts the usability of this method. For example, the real world biometrics systems are easily affected by the environmental factors, template aging, incorrect interaction with the sensors etc. The pre-trained models might not account for all these variations. Furthermore, enrolments of new users might affect these density models' accuracy [148]. Recall that LLR fusion relies on the accuracy of the modeling. The factors mentioned above are able to degrade the performance especially in large scale applications. Therefore to ensure the density models are always reliable, online learning or training is necessary, whereas such retraining has to be time-efficient.

This work presents a new approach to LLR based fusion that uses GMM. The proposed method is able to achieve comparable verification rates compared to conventional LLR based fusion. It is less sensitive to the selected component numbers, does not require a specific component searching algorithm and it requires significantly less in training time.

This novel method of improving the LLR based score fusion is inspired by the experiments described in chapter 4, i.e. on separating the score space into confidence and non-confidence regions. Fig. 6-1 shows an example of the entire score space including the confidence and non-confidence partitions. The non-confidence samples shown in fig. 6(b) are partitioned using the Min-max Sum rule fusion. The rest of the samples are regarded as confidence region samples. According to the findings of chapter 4, without considering the verification threshold, samples in the confidence region can be verified directly as impostor or genuine user. Therefore it can be say the verification performance is mainly affected by the samples in the non-confidence region.



(a)

(b)

Fig 6-1. Bimodal biometrics score distribution: (a) entire sample (b) Min-max Sum rule separated non-confidence samples.

As a result, it is suggested to use only the non-confidence samples for building the GMM. Elimination of the confidence region samples greatly reduces the training set and hence the training time can be reduced significantly. Also because the modeling is restricted to a smaller and more important region (the non-confidence region), the fusion performance is less affected by the component numbers, i.e. approximate component numbers can be used so the component number searching algorithm is not needed.

In the following sections, the proposed method is introduced. The conventional state-of-the-art GMM likelihood ratio based fusion and the concept of confidence and non-confidence regions are outlined. Section 6.3 presents the experimental results using 19 bimodal biometric fusion experiments which are also used in the previous chapters. The likelihood ratio fusion uses three GMMs in parallel. The first is a conventional model built by using the entire training set and the component numbers searched by the state-of-the-art algorithm. The second and third models are built by applying random component numbers to the entire training samples and non confidence samples. Likelihood ratio based fusion using three different models are compared and it is further compared using conventional fusion methods presented in chapter 3.

6.2 New Approach of Likelihood Ratio Based Score Fusion

The basic idea of the new fusion approach is to directly accept or reject the users when their biometrics vectors are in the confidence region or to apply the GMM-LLR fusion when the vectors are in non-confidence region. So in this method, the density models are built specifically for non-confidence regions only for impostor and genuine users. The non-confidence region is where the impostor and the genuine user scores co-exist.

6.2.1 Gaussian Mixture Model and Likelihood Ratio Fusion

GMM is formed by combining multiple Gaussian distribution density with a single component. The following shows the original model and the estimation of the densities:

Gaussian Mixture Model:

$$P(x \mid \theta_1 ... \theta_k, w_1 ... w_k) = \sum_{k=1}^{K} w_k p^k (x \mid \theta_k) \tag{6.1}$$

Where K is the number of mixture components, w_k is the weight assigned to the k^{th} mixture component $p^k(x \mid \theta_k)$ with mean vector μ_k and covariance matrix, Σ_k and $\Sigma_{k=1}^{K} w_k = 1$. x is the vector of J matcher scores $x=[x_1, x_2, ..., x_J]$. The equation for the individual component is shown in (6.2)

$$\begin{aligned} p^k(x \mid \theta_k) &= p^k(x \mid \mu_k \Sigma_k) \\ &= \frac{1}{(2\pi)^{K/2} \mid \Sigma_k \mid^{1/2}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma_k{}^{-1}(x-\mu_k)} \end{aligned} \tag{6.2}$$

Genuine user and the impostor score density estimation models are expressed in (6.3) and (6.4):

$$f_{gn}(x) = \sum_{k=1}^{K_{gn}} w_{gn,k} p^k (x; \mu_{gn,k}, \Sigma_{gn,k}) \tag{6.3}$$

$$f_{im}(x) = \sum_{k=1}^{K_{im}} w_{im,k} p^k (x; \mu_{im,k}, \Sigma_{im,k}) \tag{6.4}$$

From (6.3) and (6.4), it is necessary to choose the appropriate component numbers, K_{gn} and K_{im} to avoid over fitting or inaccurate modelling. Here the GMM fits to the training samples by using the expectation maximisation (EM) algorithm to achieve a maximum likelihood estimation of the parameters. This fitting method uses an iterative algorithm that converges to a local optimum. x represents the vector constructed by first and second modality's biometric scores in this work.

Bimodal biometrics score transformed densities of claimant whose biometrics score vector falls in non-confidence region are applied to find the log likelihood ratio, *LLR(x)* which is shown in (6.5). *LLR(x)* is used to classify the claimant into the impostor or genuine user categories based on the verification threshold.

$$LLR(x) = \log \frac{f_{gn}(x)}{f_{im}(x)} \tag{6.5}$$

6.2.2 Non-Confidence Samples and Sum Bounded Likelihood Ratio Fusion

Fig. 6-2 shows that the confidence region is a region where only a single class of users is available. The overlap part, referred to as non-confidence region, lies between the confidence regions. It represents the part where both impostor and genuine user scores co-exist.

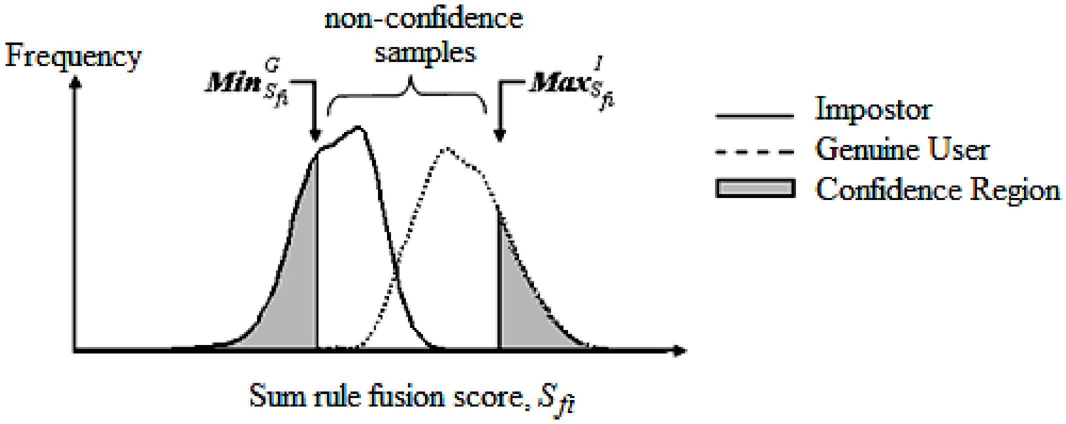


Fig 6-2. Non-confidence samples bounded by Sum rule fusion.

$$Max^I_{S_{fi}} \leq S_{fi} \leq Min^G_{S_{fi}} \tag{6.6}$$

To choose the non-confidence samples for GMM training, the samples which lie in non-confidence region of the Min-max Equal Weighted Sum fusion score, S_{fi} are used. For the biometrics, which uses the similarity measure, a sample that fulfils the rule (6.6) is treated as a non-confidence sample. $Max^I_{S_{fi}}$ and $Min^G_{S_{fi}}$ represent the maximum impostor Sum fusion score and the minimum genuine user Sum fusion score respectively. The Sum rule, as mentioned in section 2.2.3.1 is well known for its simplicity and outperformance to many complicated fusion methods. As a result, it is deemed appropriate to be used to identify the confidence and non-confidence region samples. The Min-max score normalisation is used for the Sum fusion so the score distribution's shape can be retained. By doing this the proposed approach can be evaluated without any side effects of the normalisation algorithm. Details of Sum rule fusion and the score normalisation can be found in chapter 2 and 3.

In the testing or verification phase, if a claimant's score vector is located in a confidence region, the claimed identity is accepted or rejected directly. Otherwise, the vector is evaluated by the likelihood ratio fusion as shown in (6.7). The score vector is transformed into densities using $f_{gn}(x)$ and $f_{im}(x)$. These are the density models created using the non-confidence or bounded training samples. Varying the verification threshold, η can be used to produce the ROC curve. This innovative density based score level fusion is referred to as Sum Bounded Likelihood Ratio fusion (SBLLR).

$$LLR(x) = \begin{cases} \text{Genuine User, when } \log \dfrac{f_{gn}(x)}{f_{im}(x)} \geq \eta, \\ \\ \text{Impostor, when } \log \dfrac{f_{gn}(x)}{f_{im}(x)} < \eta \end{cases} \tag{6.7}$$

6.3 Experiment Set Up and Results Analysis

To evaluate this proposed approach feasibility before introducing further complexity, it is again tested using bimodal biometrics fusion experiments. 19 bimodal biometrics fusion experiments that are used in chapter 3 and 5 are used again to test the proposed method. Each of the experiments has two stages: training and testing. Because the training and testing sets are not defined in NIST-BSSR1, half of the impostor and genuine user matching scores are randomly chosen to form the training set. The rest of the matching scores are used as testing set. Such partitions and the random assignments of the component numbers both in NIST-BSSR1 and Xm2vts are repeated 30 times for a statistical significant result. For the Xm2vts, the training and testing sets defined in [146] are used. All the parameters and models are strictly obtained in the training stage. The experiments are carried out on Matlab which is under Microsoft Windows environment with 1.6GHz CPU speed and 2GB RAM.

To investigate the SBLLR, three joint density models are built in parallel. The first model is built with the entire training sample by applying the best component numbers estimated using the effective algorithm given in [138], which was successfully applied in [56]. This is denoted B-LLR in the following discussion. The second and the third model are built by applying the same component numbers, which are randomly chosen. The only difference between R-LLR and R-SBLLR is that the GMM likelihood ratio based fusion is applied to the entire training samples and only to the non-confidence samples respectively. They are denoted R-LLR and R-SBLLR respectively. The best component numbers are estimated from the range of 1-20 for both impostor and genuine user joint density models. The genuine user training sample size is much smaller than the impostor training sample size, therefore less components might be sufficient to cover the genuine user score density distribution. So the random component numbers are chosen from the range of 1-20 for the impostor models and 1-5 for the genuine user model.

The experimental results of the three models together with the Min-max Equal Weighted Sum fusion (denoted by SUM) are presented and discussed. Their comparison is carried out over four operating points that include the lowest operating point and the point of EER. It is then compared with the conventional state-of-the-art approaches which are detailed in chapter 3.

6.3.1 Results and Discussions

The experimental results are presented numerically in tables 6-1, 6-2 and 6-3. M1 and M2 represent the two matchers used in the experiments. The results presented are the average GAR over 30 trials of the 19 experiments and its standard deviation is shown in the bracket. For the total better results over R-SBLLR in the tables, it adds up all the better results solely based on average performance. The number shown in the bracket is the total better result with statistical significance. These statistical significances are tested using t-test in the 95% confidence interval.

From NIST-BSSR1 fusion results in table 6-1, it can be seen that all the fusion strategies outperform individual single matchers. R-SBLLR outperforms most of the SUM rules except the experiments involving Fg (Exp. no. 2 and 4) at the lowest operating point. NCWW, which is the approach proposed in chapter 5, also struggles in these experiments. As shown in fig. 3-3, this is because Fg's score distributions contain long tails and multiple components resulting in an inability to effectively extract the non-confidence regions by just using the $Max^{I}_{S_{fi}}$ and $Min^{G}_{S_{fi}}$. Such ineffectiveness causes the R-SBLLR, in the experiments involving Fg at the lowest operating point with relative high performance variation (standard deviation >1.9), to degrade. However, only Exp. no. 2 is more statistically significant than R-SBLLR under this operating point.

R-LLR has more or equal better results than R-SBLLR at three higher operating points. However most of these results exhibit a higher performance variation because of the random component numbers. Among these results, R-SBLLR has performance differences of not more than 0.65% but performs more consistently (standard deviation < 1%). Furthermore, R-SBLLR at the lowest operating point, as depicted in fig. 6-3, significantly outperforms R-LLR in the range of 1.06%~3.59%. At the same time, it greatly enhances the performance consistency as shown in fig. 6-4. The statistical significance test indicates that the R-LLR has no performances significantly better than R-SBLLR at two lower operating points. However, it is comparable to or outperforms the R-SBLLR at two higher operating points. Therefore, it can be said that R-SBLLR's performance is better or comparable to R-LLR at three lower operating points but is performing more consistently over all experiments in this database.

A comparison with B-LLR, that employs state-of-the-art component numbers searching algorithm, shows that the proposed method is comparable without having to use the searching algorithm. Except the 1.37% performance difference at the lowest operating points of the Exp. no. 2, all the rest of the operating points and experiments' performance differences are approximately 1% or less. The performance standard deviations presented by B-LLR and R-SBLLR are shown in fig. 6-4. R-SBLLR is as robust as B-LLR when producing these comparable results.

Fig 6-5 and 6-8 plot the fusion processing time reduction against the verification rate changes at the lowest operating point for R-SBLLR over B-LLR. This is to show how much the fusion processing time is required for the proposed method to achieve the comparable performance. It can be seen that R-SBLLR greatly reduces B-LLR training and fusion time because of the reduction of training samples and without having to use any component number searching algorithm. At least 95% of the training and fusion time is saved for this database. For example, the average training time for Exp. no. 1 requires 1456s to build the B-LLR model and do the fusion whereas the proposed approach just requires 9.05s.

Exp. No.	GAR(%) at FAR equals to 0.001%					
	M1	M2	SUM	B-LLR	R-LLR	R-SBLLR
1	71.25(3.07)	54.81(3.61)	90.48(0.61)	92.94(0.91)	88.96(6.95)	92.27(0.99)
2	73.18(2.64)	59.91(1.29)	92.21(0.99)	92.20(2.82)	87.23(7.55)	90.82(2.24)
3	82.62(1.71)	56.92(5.19)	92.06(1.13)	95.46(1.17)	95.54(2.24)	95.32(0.93)
4	82.69(1.63)	60.55(2.10)	94.91(1.13)	95.77(1.57)	93.69(3.08)	94.76(1.97)
Total Better Results Over R-SBLLR (with statistical significance)			2(1)	4(3)	1(0)	

(a)

Exp. No.	GAR(%) at FAR equals to 0.01%					
	M1	M2	SUM	B-LLR	R-LLR	R-SBLLR
1	77.36(1.27)	74.29(2.67)	94.02(0.81)	96.13(0.65)	94.58(2.61)	95.98(0.52)
2	77.81(1.68)	67.37(1.10)	94.47(1.11)	95.80(0.66)	94.24(3.55)	94.82(0.80)
3	85.42(2.06)	75.20(2.57)	96.29(0.82)	97.74(0.74)	97.71(1.85)	97.28(0.74)
4	84.63(1.25)	68.66(2.23)	96.58(0.84)	98.39(0.81)	97.96(1.55)	97.68(0.91)
Total Better Results Over R-SBLLR (with statistical significance)			0(0)	4(3)	2(0)	

(b)

Exp. No.	GAR(%) at FAR equals to 0.1%					
	M1	M2	SUM	B-LLR	R-LLR	R-SBLLR
1	82.56(1.52)	84.86(1.16)	96.97(0.45)	98.39(0.49)	98.12(1.18)	97.98(0.66)
)2	83.76(1.24)	76.29(1.71)	96.49(0.55)	97.91(0.41)	97.38(2.57)	97.51(0.77)
3	89.81(1.51)	84.84(1.23)	97.47(0.78)	98.87(0.44)	98.81(0.60)	98.21(0.51)
4	89.64(0.87)	78.02(1.63)	97.47(0.69)	99.55(0.52)	99.46(0.75)	98.97(0.86)
Total Better Results Over R-SBLLR (with statistical significance)			0(0)	4(4)	3(2)	

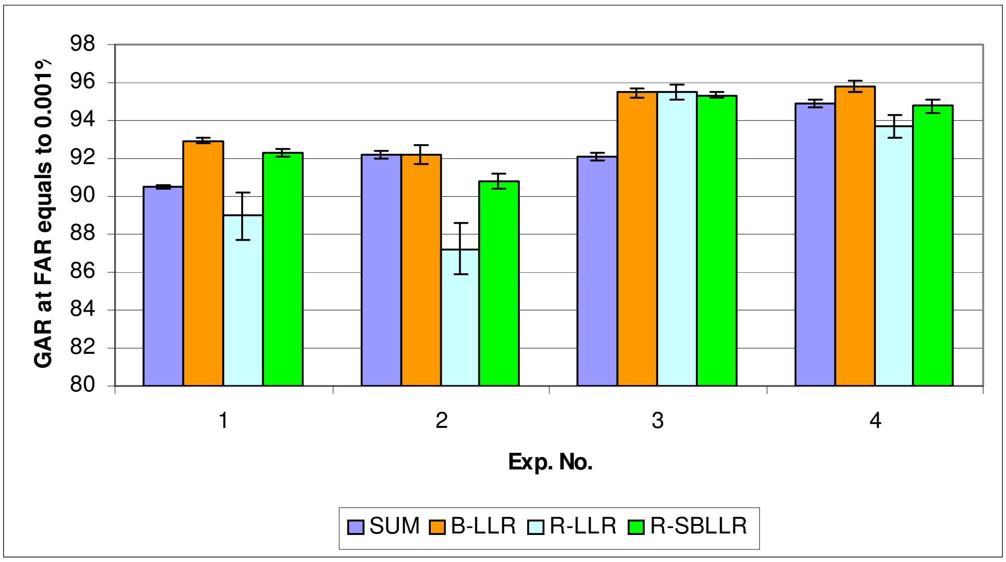
(c)

Exp. No.	EER					
	M1	M2	SUM	B-LLR	R-LLR	R-SBLLR
1	8.12(0.94)	4.36(0.29)	1.19(0.35)	0.90(0.25)	0.93(0.34)	1.32(0.27)
2	7.92(0.66)	6.31(0.76)	1.63(0.21)	1.01(0.38)	0.95(0.25)	1.38(0.15)
3	4.68(0.45)	4.29(0.26)	0.76(0.17)	0.41(0.10)	0.43(0.14)	0.73(0.24)
4	5.34(0.48)	5.67(0.76)	1.81(0.71)	0.27(0.25)	0.28(0.29)	0.74(0.54)
Total Better Results Over R-SBLLR (with statistical significance)			1(0)	4(4)	4(4)	

(d)

* *The shaded results are the average results (without consider statistical significance) that outperform R-SBLLR.*

Table 6-1. GMM based likelihood ratio fusion performance (average GAR) in four NIST-BSSR1 bimodal biometrics fusion experiments under (a) FAR=0.001% (b) FAR=0.01% (c) FAR=0.1% and (d) EER.



** The error bars are plotted at +/- 1 standard error (sample no.= 30)*

Fig 6-3. GMM based likelihood ratio fusion lowest operating point performance in NIST-BSSR1.

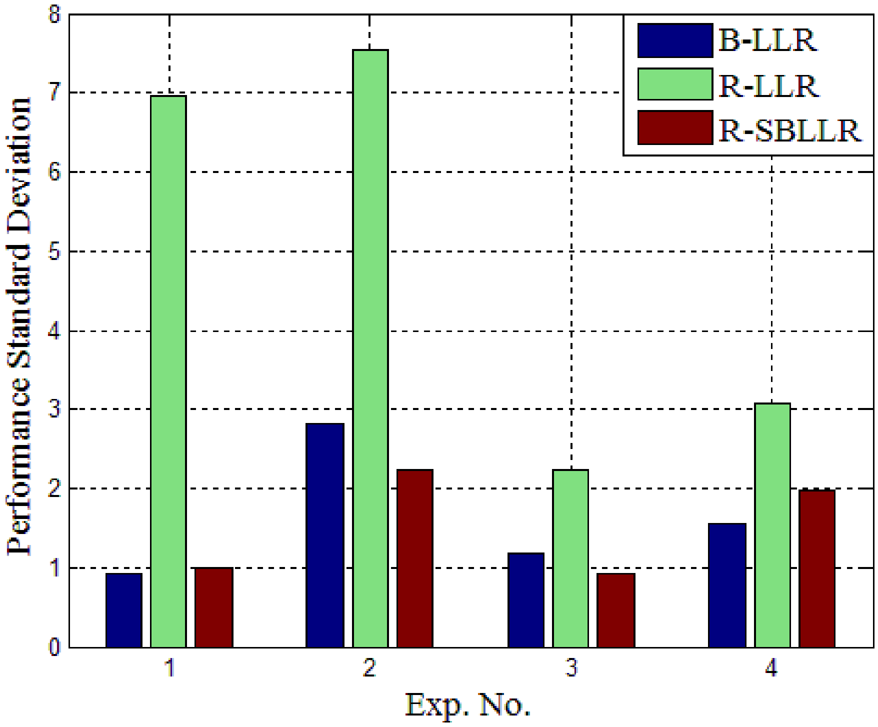
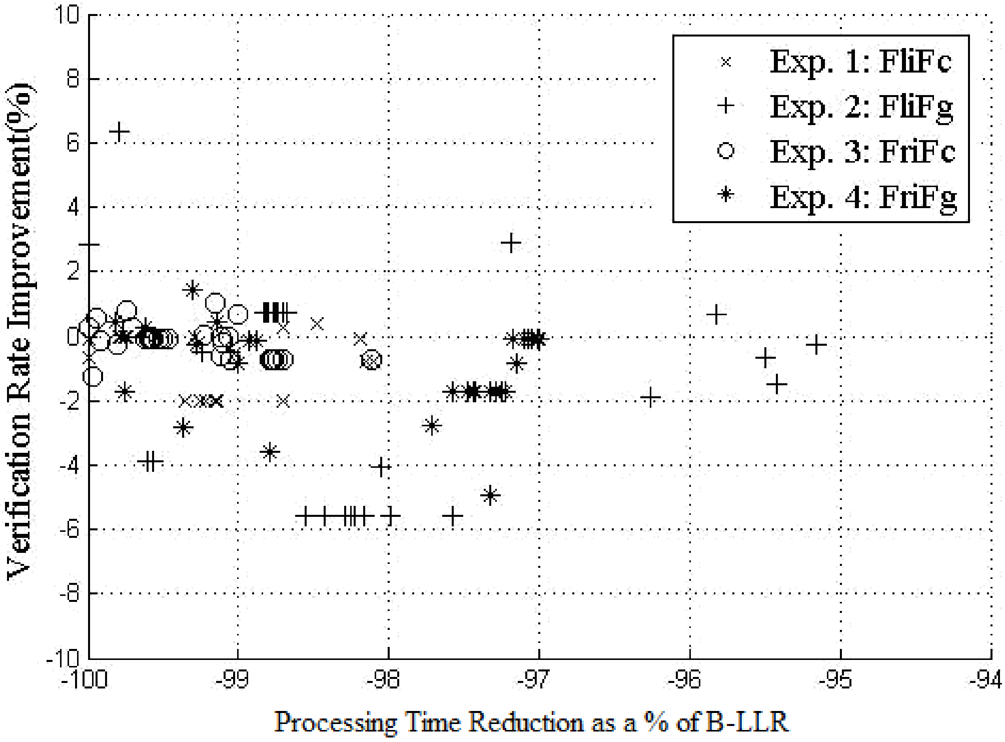


Fig 6-4. GMM based likelihood ratio fusion lowest operating point performance's standard deviation in NIST-BSSR1 to show fusion performance consistency.



** The sample points show all 30 trials of the four bimodal biometrics experiments.*

Fig 6-5. Performance variation of R-SBLLR over B-LLR in terms of verification rate at lowest operating points and processing time in NIST-BSSR1.

It is clear from the Xm2vts experiments that the proposed method outperforms SUM and R-LLR significantly. The SUM fusion achieves 2.01% and 1.62% performance differences that are better than R-SBLLR in Exp. no. 2 and 4 respectively under the lowest operating point. Aside from these performance differences, the SUM rule outperformance differences are just within the range up to 0.53% for other better results under all operating points and experiments. For the better results of R-SBLLR over SUM, the performance differences are up to 19.28%. Also, there are just 2 ~ 4 experiments outperforming with statistical significance using SUM in all operating points and experiments.

R-LLR outperforms R-SBLLR with 4.85% difference in Exp. no. 10 under lowest operating point. However, the rest of the outperformance is just within the range up to 0.57% for all experiments and operating points. R-SBLLR on the other hand is able to produce much better results that are up to 8.18% better compared to R-LLR. Furthermore, by considering the better average results' statistical significance, there are just 1 ~ 4 out of 15 R-LLR's fusion results outperforming with statistical significance in all operating points and experiments.

As shown in fig. 6-6, B-LLR obtains 5 better fusion results over R-SBLLR under the lowest operating points. These differences are in the range of 0.72% ~5.91%. Despite of this, R-SBLLR obtains 10 better fusion results over B-LLR with performance differences in the range of 0.11% ~ 3.12%. At the second lowest operating point, there are 7 experiments in which B-LLR outperforms R-SBLLR. Aside from 2.57% and 1.80% differences presented by Exp. no. 2 and 13 correspondingly, the other outperformance differences are in the range just up to 0.57%. For the other two operating points, the better performance's differences remain in the range up to 0.76%. There are just 4 and 6 experiments in which B-LLR that have better statistical significances than R-SBLLR at two lower operating points and 8 and 7 experiments at two higher operating points. This demonstrates that the proposed method is outperforming or comparable to the conventional state-of-the-art approach for this database.

Fig. 6-7 shows the lowest operating point average results' standard deviation. It can be seen that the experiments involving F4 and F5 (Exp. no. 10~15) are with significant performance variation for R-LLR. From table 6-2 it can be seen that these matchers (F4 and F5), which are denoted as M1 in the table, exhibit a verification rate of 0 ~ 0.31% verification rate at the lowest operating point. The matchers with near-zero performance are the reason for such high performance variation. However, by using R-SBLLR, these variations can be significantly reduced.

The performance improvement for R-SBLLR over B-LLR in terms of verification rate at lowest operating point and corresponding training time reduction for all Xm2vts experiment trials are shown in fig. 6-8. At least 97% of the fusion and training time is improved. For example, the average training time for Exp. no. 10 requires 1126s to do the fusion and to build the B-LLR model, whereas the proposed approach just requires 0.77s.

Exp. No.	GAR(%) at FAR equals to 0.001%					
	M1	M2	SUM	B-LLR	R-LLR	R-SBLLR
1	0.78(0)	60.12(0)	91.81(0)	96.55(0.60)	94.46(2.42)	94.32(1.88)
2	0.78(0)	19.04(0)	88.25(0)	86.13(0.59)	84.82(1.55)	86.24(1.64)
3	0.78(0)	40.37(0)	88.12(0)	89.98(0.26)	88.49(2.72)	88.86(1.89)
4	61.72(0)	59.05(0)	93.43(0)	90.63(1.65)	91.22(1.36)	91.81(1.38)
5	61.72(0)	19.46(0)	78.62(0)	78.79(0.69)	78.94(0.95)	79.12(1.44)
6	61.72(0)	41.49(0)	79.56(0)	79.61(1.32)	79.46(1.47)	80.30(0.72)
7	81.49(0)	59.05(0)	96.25(0)	95.98(2.08)	96.05(1.91)	96.06(1.82)
8	81.49(0)	19.46(0)	80.87(0)	86.11(2.01)	85.67(2.32)	86.77(1.89)
9	81.49(0)	41.49(0)	91.75(0)	90.45(0.67)	90.44(1.48)	91.61(0.31)
10	0.31(0)	59.05(0)	75.56(0)	87.99(0.74)	86.93(2.99)	82.08(3.00)
11	0.31(0)	19.46(0)	65.81(0)	69.23(1.14)	68.86(4.18)	68.51(2.25)
12	0.31(0)	41.49(0)	67.31(0)	77.32(1.24)	75.11(8.88)	80.44(1.83)
13	0(0)	60.12(0)	63.87(0)	83.40(0.50)	79.28(4.41)	81.80(1.51)
14	0(0)	19.04(0)	47.55(0)	60.92(2.32)	54.71(11.54)	62.89(5.61)
15	0(0)	40.37(0)	57.05(0)	78.53(1.87)	73.65(6.57)	78.65(2.29)
Total Better Results Over R-SBLLR (with statistical significance)			4(3)	5(4)	3(1)	

(a)

Exp. No.	GAR(%) at FAR equals to 0.01%					
	M1	M2	SUM	B-LLR	R-LLR	R-SBLLR
1	80.7(0)	70.18(0)	92.50(0)	97.52(0.48)	97.09(1.93)	97.94(0.50)
2	80.7(0)	34.04(0)	89.34(0)	93.63(0.32)	90.61(1.51)	91.06(1.91)
3	80.7(0)	56.75(0)	91.39(0)	93.58(0.47)	93.28(1.39)	94.64(0.81)
4	71.00(0)	70.10(0)	95.77(0)	96.16(0.16)	96.14(0.20)	95.91(0.31)
5	71.00(0)	34.04(0)	84.84(0)	84.77(0.36)	85.08(0.48)	84.51(0.63)
6	71.00(0)	56.75(0)	88.53(0)	88.26(0.66)	88.11(0.53)	88.51(0.49)
7	89.30(0)	70.10(0)	98.50(0)	98.20(0.84)	98.25(0.85)	97.97(0.92)
8	89.30(0)	34.04(0)	91.65(0)	93.74(0.96)	93.83(0.93)	93.98(0.80)
9	89.30(0)	56.75(0)	94.52(0)	95.12(0.19)	94.97(0.25)	94.72(0.28)
10	14.43(0)	70.10(0)	81.75(0)	96.59(0.39)	95.57(2.92)	96.02(1.30)
11	14.43(0)	34.04(0)	71.30(0)	77.30(0.72)	76.33(4.41)	81.15(2.13)
12	14.43(0)	56.75(0)	77.25(0)	84.25(0.78)	81.58(6.58)	85.94(0.82)
13	11.86(0)	70.18(0)	70.00(0)	91.08(0.34)	87.83(3.59)	89.28(1.52)
14	11.86(0)	34.04(0)	57.05(0)	69.93(2.97)	62.69(9.79)	70.33(4.34)
15	11.86(0)	56.75(0)	64.48(0)	82.10(0.83)	78.62(4.15)	83.29(0.87)
Total Better Results Over R-SBLLR (with statistical significance)			3(2)	7(6)	4(3)	

(b)

Exp. No.	GAR(%) at FAR equals to 0.1%					
	M1	M2	SUM	B-LLR	R-LLR	R-SBLLR
1	92.5(0)	88.50(0)	94.57(0)	99.31(0.11)	98.82(1.21)	98.99(0.05)
2	92.5(0)	58.35(0)	94.25(0)	96.88(0.12)	96.10(1.15)	96.46(0.23)
3	92.5(0)	71.22(0)	94.50(0)	98.49(0.11)	98.17(1.17)	98.57(0.12)
4	81.98(0)	88.50(0)	98.32(0)	98.76(0.04)	98.75(0.06)	98.76(0.04)
5	81.98(0)	58.22(0)	92.66(0)	94.09(0.16)	93.91(0.22)	93.47(0.31)
6	81.98(0)	71.29(0)	95.40(0)	95.35(0.20)	95.40(0.21)	95.23(0.22)
7	94.50(0)	88.50(0)	99.25(0)	99.18(0.38)	99.18(0.38)	99.03(0.40)
8	94.50(0)	58.22(0)	96.00(0)	97.29(0.38)	97.23(0.40)	96.82(0.54)
9	94.50(0)	71.29(0)	97.53(0)	97.62(0.12)	97.60(0.15)	97.35(0.21)
10	78.75(0)	88.50(0)	90.25(0)	99.00(0.00)	98.58(1.09)	98.75(0.02)
11	78.75(0)	58.22(0)	88.97(0)	93.72(0.34)	92.78(1.29)	93.82(0.34)
12	78.75(0)	71.29(0)	87.29(0)	95.30(0.30)	94.06(2.33)	95.74(0.19)
13	52.00(0)	88.50(0)	78.50(0)	97.62(0.17)	95.25(2.88)	97.56(0.33)
14	52.00(0)	58.35(0)	74.20(0)	85.44(0.43)	81.11(6.09)	84.68(1.07)
15	52.00(0)	71.22(0)	74.71(0)	90.69(0.64)	89.14(1.28)	91.07(0.78)
Total Better Results Over R-SBLLR (with statistical significance)			3(3)	10(8)	5(4)	

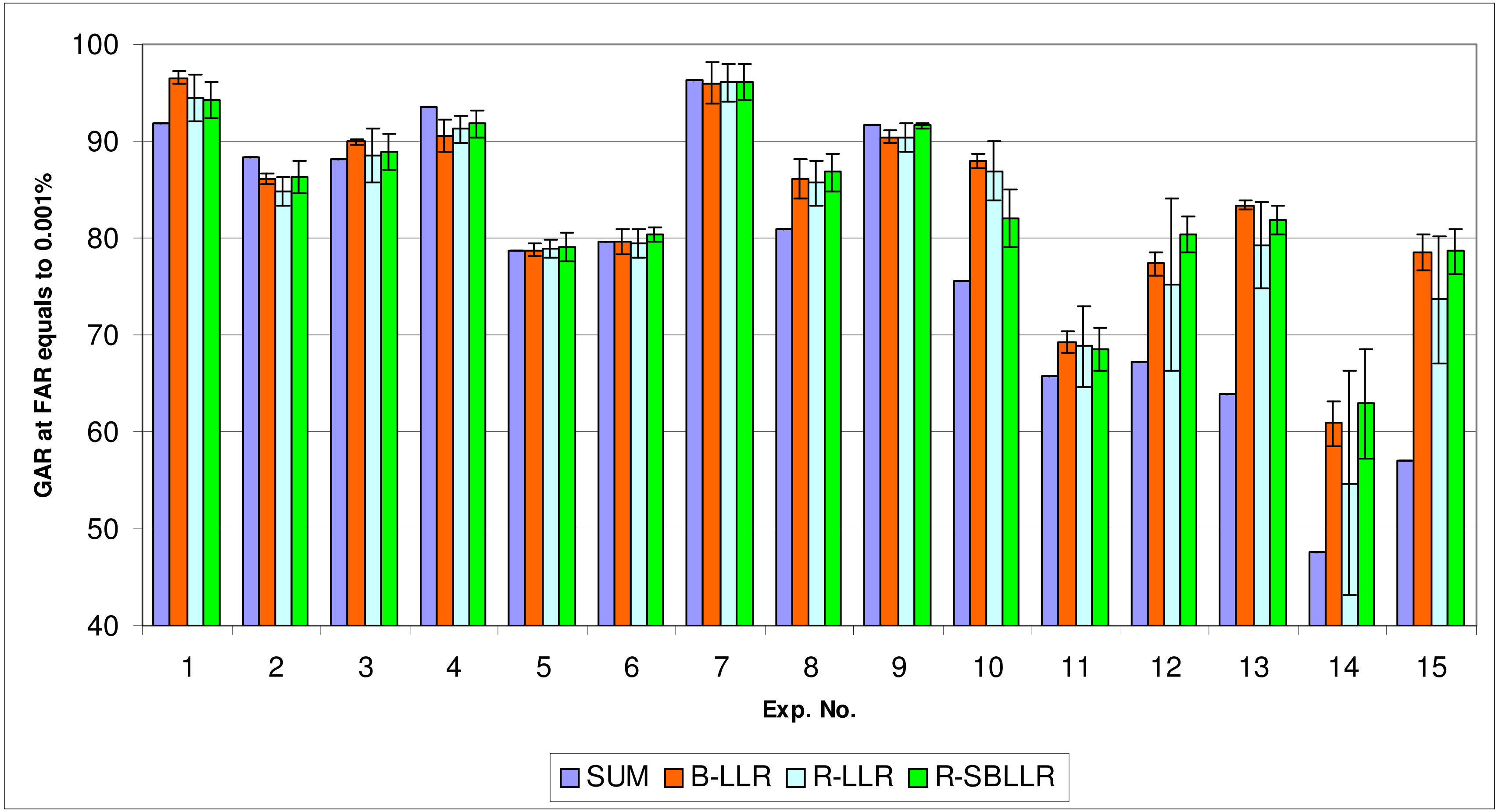
(c)

Exp. No.	EER					
	M1	M2	SUM	B-LLR	R-LLR	R-SBLLR
1	1.82(0)	1.11(0)	0.91(0)	0.31(0.04)	0.47(0.17)	0.57(0.11)
2	1.82(0)	6.50(0)	1.25(0)	1.07(0.06)	1.26(0.25)	0.95(0.06)
3	1.82(0)	4.51(0)	1.10(0)	0.75(0.00)	1.00(0.11)	0.75(0.00)
4	4.12(0)	1.11(0)	0.50(0)	0.59(0.02)	0.57(0.06)	0.68(0.10)
5	4.12(0)	6.50(0)	1.63(0)	1.50(0.00)	1.65(0.11)	1.71(0.07)
6	4.12(0)	4.50(0)	1.19(0)	1.25(0.00)	1.24(0.02)	1.21(0.05)
7	1.82(0)	1.11(0)	0.46(0)	0.51(0.08)	0.53(0.17)	0.50(0.14)
8	1.82(0)	6.50(0)	1.25(0)	0.97(0.09)	1.17(0.21)	1.25(0.19)
9	1.82(0)	4.50(0)	0.75(0)	0.85(0.03)	0.85(0.08)	0.88(0.08)
10	3.50(0)	1.11(0)	1.16(0)	0.38(0.03)	0.41(0.11)	0.50(0.03)
11	3.50(0)	6.50(0)	1.38(0)	0.78(0.05)	0.81(0.10)	0.75(0.01)
12	3.50(0)	4.50(0)	1.50(0)	0.85(0.06)	0.98(0.21)	0.85(0.11)
13	6.50(0)	1.11(0)	2.60(0)	0.56(0.05)	0.70(0.18)	0.58(0.05)
14	6.50(0)	6.50(0)	3.53(0)	2.44(0.09)	2.45(0.22)	2.44(0.08)
15	6.50(0)	4.51(0)	3.79(0)	1.69(0.08)	1.79(0.18)	1.78(0.05)
Total Better Results Over R-SBLLR (with statistical significance)			5(4)	8(7)	6(4)	

(d)

* *The shaded results are the average results (without consider statistical significance) that outperform R-SBLLR.*

Table 6-2. GMM based likelihood ratio fusion performance (average GAR) in fifteen Xm2vts bimodal biometrics fusion experiments under (a) FAR=0.001% (b) FAR=0.01% (c) FAR=0.1% and (d) EER.



* *The error bars are plotted at +/- 1 standard error (sample no.= 30)*

Fig 6-6. GMM based likelihood ratio fusion lowest operating point performance in Xm2vts.

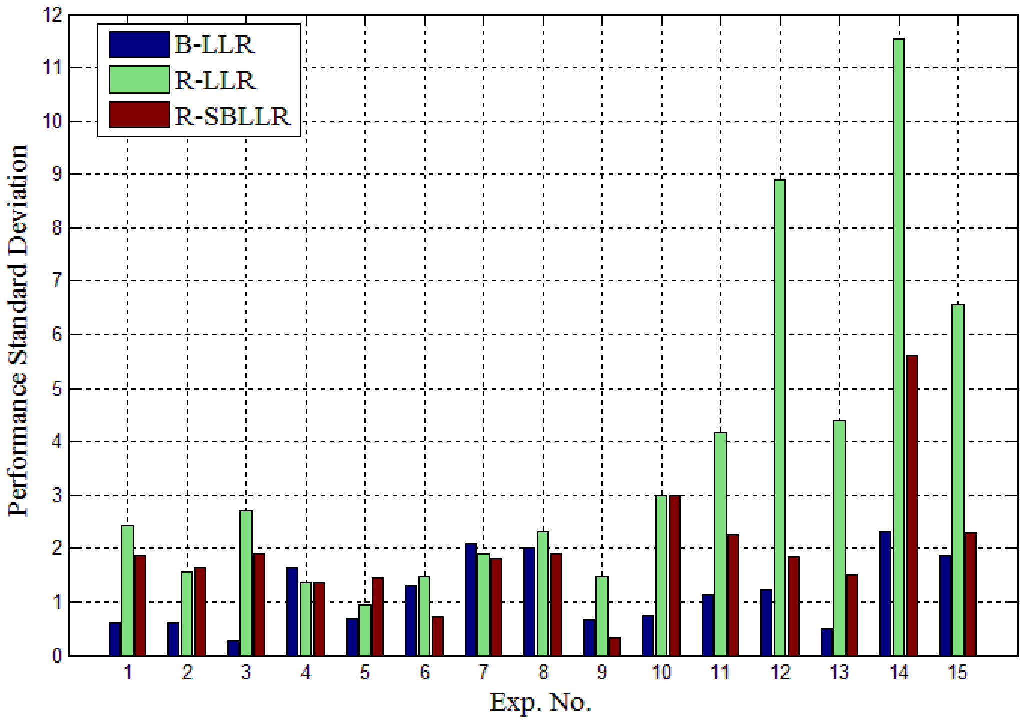
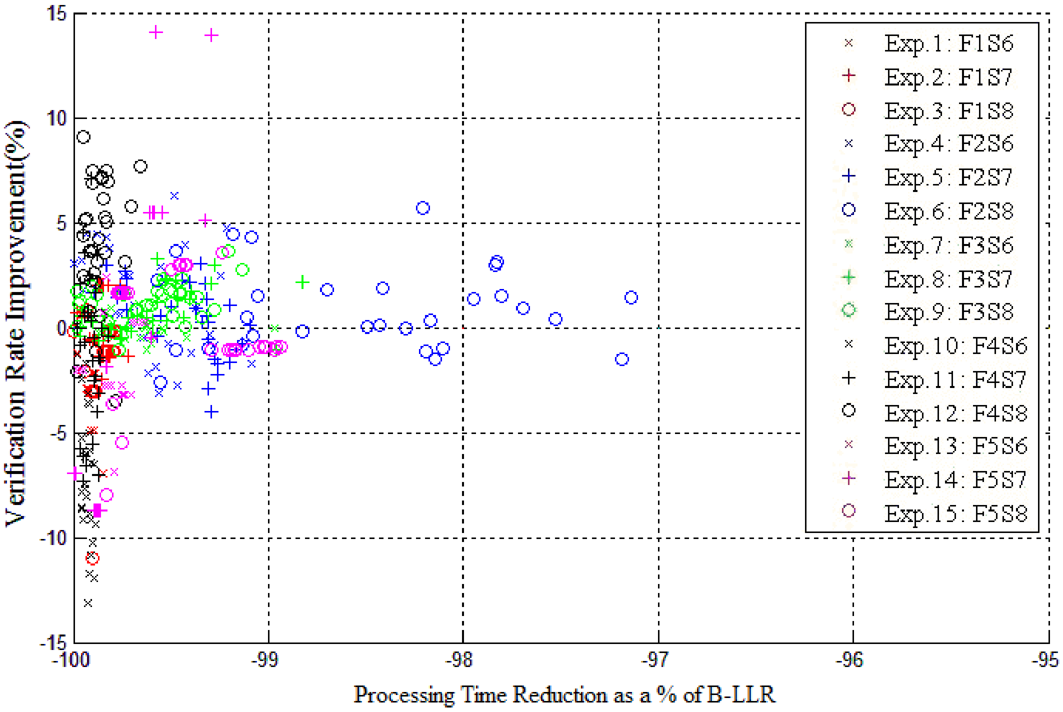


Fig 6-7. GMM based likelihood ratio fusion lowest operating point performance's standard deviation in. Xm2vts to show fusion performance consistency.



* *The sample points show all 30 trials of the fifteen bimodal biometrics experiments.*

Fig 6-8 Performance variation of R-SBLLR over B-LLR in terms of verification rate at lowest operating points and processing time in Xm2vts.

6.3.2 SBLLR Comparison to Other Conventional Approaches

Table 6-3 shows the fusion results achieved by the proposed method and the conventional state-of-the-art approaches at the lowest operating point, which are detailed in chapter 3. The shaded results in the table are the results that are statistically significant better than the proposed method. In both databases, it is clear that R-SBLLR is a better choice than SUM, SVM and MLLR. This is because R-SBLLR obtains more comparable or better results than the rest.

For the LREG, there are 10 experiments out of 19 that perform better than the proposed method. Aside from 5.17% performance difference in Exp. no. 14 in Xm2vts, the other outperformances are just in the range of 0.14% ~ 2.10%. In contrast to this, LREG achieves the performances which are 23.59%, 38.69%, 29.11% and 20.11% less comparing to R-SBLLR in Xm2vts Exp. no. 4, 7, 10 and 13 respectively. These large performance variations are caused by the inconsistent performance of LREG at very low operating points as mentioned in chapter 3.

NIST-BSSR1	GAR(%) at FAR equals to 0.001%					
Exp. No.	M1M2	SUM	LREG	SVM	MLLR	R-SBLLR
1	FliFc	90.48(0.61)	92.85(0.59)	93.24(0.52)	92.30(1.32)	92.27(0.99)
2	FliFg	92.21(0.99)	91.89(1.16)	91.53(1.20)	88.41(7.29)	90.82(2.24)
3	FriFc	92.06(1.13)	96.08(0.94)	95.26(1.11)	94.96(1.72)	95.32(0.93)
4	FriFg	94.91(1.13)	94.44(0.95)	95.11(1.20)	95.02(1.53)	94.76(1.97)

Xm2vts	GAR(%) at FAR equals to 0.001%					
Exp. No.	M1M2	SUM	LREG	SVM	MLLR	R-SBLLR
1	F1S6	91.81(0)	95.25(0)	95.78(0)	96.39(0.52)	94.32(1.88)
2	F1S7	88.25(0)	86.25(0)	86.31(0)	83.63(0.59)	86.24(1.64)
3	F1S8	88.12(0)	82.66(0)	81.84(0)	90.46(0.35)	88.86(1.89)
4	F2S6	93.43(0)	68.22(0)	90.56(0)	93.26(0.24)	91.81(1.38)
5	F2S7	78.62(0)	80.31(0)	79.28(0)	78.06(0.50)	79.12(1.44)
6	F2S8	79.56(0)	81.05(0)	80.28(0)	78.61(1.03)	80.30(0.72)
7	F3S6	96.25(0)	57.37(5.44)	94.76(0)	96.25(0.04)	96.06(1.82)
8	F3S7	80.87(0)	85.18(5.44)	85.31(0)	85.22(0.43)	86.77(1.89)
9	F3S8	91.75(0)	91.75(0)	91.78(0)	92.15(0.17)	91.61(0.31)
10	F4S6	75.56(0)	52.97(9.26)	80.31(0)	83.86(1.25)	82.08(3.00)
11	F4S7	65.81(0)	68.55(0)	69.65(0)	71.04(0.63)	68.51(2.25)
12	F4S8	67.31(0)	81.02(0)	79.56(0)	79.70(0.50)	80.44(1.83)
13	F5S6	63.87(0)	61.69(0)	78.81(0)	80.42(0.57)	81.80(1.51)
14	F5S7	47.55(0)	68.06(0)	60.96(0)	62.29(0.33)	62.89(5.61)
15	F5S8	57.05(0)	80.75(0)	78.87(0)	76.29(0.84)	78.65(2.29)

** The shaded results are the average results that statistical significantly outperform R-SBLLR.*

Table 6-3. R-SBLLR performance comparisons with conventional state-of-the-art approaches.

6.4 Conclusions

From the analysis of the fusion results in both databases, the proposed method is demonstrated to significantly outperform R-LLR in most of the experiments at most operating points. It successfully reduces the performance variations of R-LLR caused by using random component numbers. This proposed method achieves comparable performance to the conventional state-of-the-art B-LLR especially at lower operating points in most of the experiments without having to use component number searching algorithm. Because of the reduction in the number of samples and not needing to search for the component number, at least 95% of the fusion and training time can be saved. Also, this method is very easy to understand and to implement because of its simplicity. R-SBLLR achieves comparable performance to B-LLR and outperforms the conventional state-of-the-art fusion approaches in terms of performance and consistency.

Regardless of the effectiveness and the benefits offered by the proposed method, there are also some limitations on this approach. Firstly, choosing the non-confidence samples just based on the $Max_{S_{fi}}^{I}$ and $Min_{S_{fi}}^{G}$ is not reliable. Such a choice can be affected by the outlier. Furthermore, when the non-confidence region contains multiple clusters, it requires a more accurate density model to achieve optimum performance. R-SBLLR, which uses an approximate model, performs less effectively than B-LLR in these cases. In future work, a more accurate separation of the confidence and non-confidence samples should be investigated to further enhance the SBLLR performance. Also, a criterion to choose this fusion approach as an effective fusion approach has to be investigated.

7 CONCLUSIONS AND FUTURE WORK

7.1 Conclusions

Multimodal biometrics alleviates many restrictions of single biometrics. Based on the literature, to improve the biometrics authentication performance, one can increase the information gain or design a more effective fusion algorithm. Due to the lower correlation between the sources, multimodal biometrics provides maximum information gain for authentication.

The work presented in this thesis demonstrates that information from the non-confidence region can be used as additional information to further improve conventional state-of-the-art fusion approaches for combining the multimodal biometrics. In the literature, additional information such as biometrics quality, soft biometrics, modality reliability measure, failure prediction model, etc., are embedded into the conventional fusion approaches. The use of non-confidence information for biometrics score level fusion, to the best of author's knowledge, has not been reported yet.

In the literature, different fusion approaches from three different categories (rule based, classification based and density based fusion) have been utilised to combine the biometrics information. Some of the approaches, such as Sum Rule, Logistic Regression, Support Vector Machine and Likelihood Ratio based fusion, are reported in the literatures to achieve top performance and outperforming others. However, which fusion approach (amongst the above mentioned) achieves the best results when combining multimodal biometrics score remained unknown in this research community. In this work, these approaches are simultaneously compared and extensively evaluated by using 19 bimodal biometrics experiments conducted from two large scale databases. It is found that these approaches have comparable fusion performance. Even though these conventional state-of-the-art approaches have been claimed to achieve the best performance, there is no single fusion method which outperformed others in all experiments and operating thresholds. Furthermore, aside from the fusion performance, there are other factors have to be considered to choose an appropriate strategy. This suggests to the community that the strategy with the best performance might not necessarily be an appropriate one. The selection of an appropriate strategy has to comprehensively consider the prerequisites and other factors. For instance, the required fusion and training time, implementation requirements, resource availability as well as their advantages and disadvantages. This explains the existence of three different approaches.

Applying a fusion strategy selection mechanism is a way of improving fusion performance. However, this is not common in the biometrics fusion research community. The only work reported uses the estimated errors and classifier to make the selection [62]. Different biometrics score spaces exhibit different confidence in discriminating a claimant. This region information might be employed for the selection scheme. However, whether it is feasible for further fusion improvement is unknown. The proposed Hybrid Fusion in chapter 4 answers this question. This method manually assigns the confidence partitions and replaces the Sum rule with more confidence rules (Min and Max) in these partitions. This hybrid method achieves increases in the range of 0.3% ~ 1.7% compared to the Min-max Equal Weighted Sum rule. However, the results rely on careful manual assignment of the confidence and non-confidence partitions. It is found that such assignment of the partitions has to successfully increase the separation of the non-confidence sample (but not the entire sample) to achieve an improvement. Therefore, the answer to the research question, whether selecting different rules for different confidence partitions will further improve the authentication rate, is no. This is to say that there is no fusion improvement even though a further separation of the samples in the confidence region is achieved. The fusion improvement depends on the non-confidence region samples only but not on the samples from the entire score space. Nevertheless, this finding confirms the importance of non-confidence region related information for fusion improvement.

The Weighted Sum rule has been widely used for biometrics score level fusion in the literature. The commonly used Weighted Sum rules use *d'* Weighting, EER Weighting and Equal Weighing. However, through the experimental results obtained from a wide range of experiments, it can be concluded that these commonly used weighting scheme are not able to achieve generalisation performance. In the literature, the optimal fusion performance can be achieved by using a specific optimal weight searching algorithm. Nevertheless, this searching has to be repeated whenever the operating point is changed. Therefore, there is a need to explore for a new weighting parameter that will enable consistent fusion performance. From chapter 4, it is found that the non-confidence region is directly related to the fusion improvement. Therefore, it is proposed to minimise the non-confidence region to achieve generalisation fusion performance, by using the non-confidence region width as the weighting parameter. By doing this, the gradient of the linear separation boundary can be adjusted to enable the minimisation of the non-confidence region.

The proposed NCWW Sum rule in chapter 5 successfully demonstrates that it has better generalisation ability than the rival weighting methods. In 19 experiments, it outperforms the rest (EER, *d'* and Equal Weighted Sum rule) with obtaining 10, 14, 16 and 18 best results at the lowest operating point to the EER operating point respectively. At the lowest operating point, the NCWW Sum rule performs up to 26.27% better than the compared schemes. This experimental result also gives the answer that the commonly used Weighted Sum rules (EER, *d'* and Equal Weighted Sum rule) are not optimal enough although of these approaches are regularly used in the literature. Not only the NCWW Sum rule achieves the generalisation performance, it is also comparable to the conventional state-of-the-art fusion approaches. From the lowest operating point fusion results, it can be seen that NCWW is comparable or outperforms the other methods with statistical significance, e.g. the LREG, SVM and LLR based fusions with 7 out of 19 experiments. Such comparable or outperforming differences are in the range of -0.15% ~ 5.29%. There are another 6 experiments where NCWW performs less effective just in the moderate range of -0.96% ~ -1.38%. In a higher dimensional score space, it achieves comparable fusion results to the state-of-the-art fusion presented in [56] with verification results of more than 99%. Considering these results and the simplicity of this approach, it is considered an alternative option to the conventional state-of-the-art approaches. Furthermore, this approach also does not require the score normalisation which is required by other rule based fusion approaches in the literature.

SBLLR (or referred to as R-SBLLR) proposed in chapter 6 is an attempt to use the GMM based likelihood ratio fusion without having to assign accurate component number. Other works in the literature focus in looking for the optimal component number to boost the fusion performance. This work is novel in the way that it tries to reduce the impact of an inaccurate assignment of the components number to the fusion performance. By doing this, fewer resources are needed whilst the fusion performance can be maintained. Furthermore, the component number searching time can be saved to improve this method's usability. Without using any component number searching algorithm, it is found that the proposed method is comparable to the one that uses the state-of-the-art component number searching algorithm. The component numbers used by the non-confidence sample model have been randomly chosen and this likelihood ratio fusion is only applied in the non-confidence region. At the lowest operating point, the proposed method outperforms the other algorithms with statistical significance or is comparable in 12 experiments (out of 19). This performance differences are in the range of -0.14% ~ 3.12%. For the experiments in which R-SBLLR did not outperform the others, except 2 experiments with 2.23% and 5.91% differences, the differences are just equal or less than 1.38%. At least 95% and 97% of the modelling and fusion time can be saved for both databases. For example, the average training time for Exp. no. 10 in Xm2vts was 1126s for doing the fusion to build the B-LLR model, whereas the new approach required just 0.77s. From these results, it can be concluded that this work successfully addresses the identified gap. It presents a new GMM based likelihood ratio fusion method that eliminates the need of component searching algorithm which is required by other algorithms reported in the literature. Furthermore, the fusion time required by the conventional GMM based likelihood ratio fusion is also significantly reduced.

Two of the proposed methods use non-confidence information to improve the conventional biometrics score level fusion approaches. While this information is not explored before, the experimental results presented in this work suggest that this region is informative for multimodal biometrics score level fusion.

The remaining sections give the suggestions based on the experiences and findings gained whilst this research work was conducted. These suggestions include the extension of the conducted work on the partially working or non-working cases, and how the findings of this work are applicable to other applications.

7.2 Future Work

This work has achieved significant progress to improve the conventional state-of-the-art fusion methods' performance and usability by using non-confidence information. However, it is believed that these devised methods can be further refined, expanded and applied in other applications. In this final part of thesis, a non-exhaustive list for potential extension of this work is given:

- **Non-confidence region width redefinition.** Using the difference between the maximum impostor score, Max^I and minimum genuine user score, Min^G alone to define the non-confidence region is not accurate enough. Such a definition can be easily affected by outliers, which Grubbs defines as: "An outlying observation, or outlier, is one that appears to deviate markedly from other members of the sample in which it occurs" [152]. Density related information from this region should be included in the definition. This redefined non-confidence region is expected to be more representative for specific biometrics authentication ability. It therefore can be used more reliable as a weighting reference for Weighted Sum rule to further enhance the proposed approach.

- **NCWW extension.** NCW defined in this work relies on searching through a large scale training sample set. However, the parameters used by the NCW definition, maximum impostor score, Max^I and minimum genuine user score, Min^G can be searched through the worst impostor and genuine user authentication cases for a specific biometrics. Therefore, the large training set is not needed to save sample collection times. NCWW demonstrates the importance of testing the biometrics over a series of critical cases. However, it leads to the research question of how these cases can be identified for a specific biometrics. For instance, a fingerprint biometrics' low genuine user score might be caused by a damaged fingerprint or contamination of the fingerprint capturing device. A high impostor score might be the result of interclass similarity or spoof attack (e.g. using rubber fingerprints). NCWW work extension can be directed to collection of a series of problematic cases to critically evaluate the specific biometrics. NCW therefore can be used as a parameter to evaluate the specific biometrics performance, as well as using it as a reliable weighting reference for biometrics score level fusion, without requiring a large training sample set.

- **SBLLR extension.** A density based method relies on the availability of sufficient training samples to create an accurate density model. For SBLLR, there might be an insufficient number of training samples available for modeling, because only the non-confidence samples are used. A training sample size assessment has to be included to ensure the appropriateness of applying this method. To enhance the proposed approach, finding the appropriate range for the random component numbers needs to be studied further as well. Using only the non-confidence samples for training is tested on joint density models in this work. It is interesting to investigate the benefits of applying it to the marginal density modeling as well as other classification based fusion methods.

- **Intelligent switcher.** No single fusion strategy, including the proposed methods, has achieved the best results in all biometrics fusion experiments and at all operating points. Switching between different fusion algorithms enables a more robust performance. The effectiveness for such switching can be explained using the consistent fusion concept presented in [128]. This kind of algorithm is also termed a multiple classifier system or dynamic score selection in the literature. This is demonstrated in several applications in [153], [154]. The comparison of the fusion result using this switching scheme and the optimised individual classifiers in [149] further shows the feasibility of this idea.

- **Higher dimensional fusion and sources selection.** Only bimodal biometrics is explored in this work. More biometrics sources can be included to further evaluate the proposed method's effectiveness. With the availability of multi biometrics sources, whether the inclusion of as many sources as possible in the fusion is of benefit has to be answered. If not, the new research question will be how to find the best combination amongst different sources to achieve an optimum performance. Since the non-confidence region is the key region to determine the fusion performance, the non-confidence region formed by multiple sources can potentially be used as a guide for searching for such best combination.

- **Biometrics Identification System.** The central concept of the work is to increase the authentication performance by combining multimodal biometrics and using the non-confidence information. In this work, this is implemented in verification mode only. Since the proposed method is effective in verification mode, it probably helps to improve the fusion in identification mode as well. In contrast to the two class problem presented by the verification mode, the identification mode is a more complicated multiclass problem. While identification is another important authentication task, it is desired to investigate the modification of the proposed fusion method to deal with the identification problems. By doing this, wider applications can be covered.

- **Bioinformatics Application.** Fusion of different evidences or information enhances the classification performance. Machine learning algorithms proposed or mentioned in this work are not only applicable to biometrics, they may be applied, for example, to the fusion of genomic, proteomic and transcriptomic data in bioinformatics community [155].

References

[1] Yoshiaki Isobe, "Personal Authentication Infrastructure for the Ubiquitous Information Age", http://www.hitachi.com/rd/sdl/people/bio/index.html [Last checked on 14-04-2010]

[2] W. Sturgeon, "Biometrics Curing Password Headaches", http://software.silicon.com/security/0,39024655,39152802,00.htm, [Last checked on 14-04-2010]

[3] S. Granneman, "Bill Gates Is Right?"[Online], Available at http://www.securityfocus.com/columnists/277, 2004, [Last checked on 14-04-2010]

[4] S. Prabhakar, S. Pankanti and A. Jain, "Biometric Recognition: Security and Privacy Concerns", IEEE Security and Privacy, Vol. 1, No.2, pp. 33-42, March-April 2003.

[5] C. Everett, "Biometrics–Ready To Be Used in Anger?", Infosecurity Today, Vol. 3, Issue 5, pp. 30-32, September-October 2006.

[6] I. Maghiros, Y. Punie, S. Delaitre and et. al., "Biometrics at the Frontiers: Assessing the Impact on Society", Institute for Prospective Technological Studies Technical Report Series (EC- DG JRC – IPTS), http://ec.europa.eu/justice_home/doc_centre/freetravel/doc/biometrics_eur21585en.pdf, [Last checked on 15-04-2010]

[7] M. Bashir and J. Kempf, "Bio-Inspired Reference Level Assigned DTW for Person Identification Using Handwritten Signatures ", International Conference on Biometrics ID Management and Multimodal Commu-nication, Madrid, Spain, Springer LNCS 5707, pp. 200-206, September 2009.

[8] C. Dorai, N. Ratha and R. Bolle, "Dynamic Behavior Analysis in Compressed Fingerprint Videos", IEEE Trans. Circuits Syst. Video Techn. Vol. 4(1), pp. 58-73, 2004.

[9] A. Jain, P. Flynn and A. Ross, "Handbook of Biometrics", Springer, 2008.

[10] G. Wubbeler, M. Stavridis, D. Kreiseler, R. Bousseljot and C. Elster, "Verification of Humans Using Electrocardiogram", Pattern Recognition Letters, Vol. 28, pp. 1172-1175, 2007.

[11] D. Silver and A. Biggs, "Keystroke and Eye-Tracking Biometrics for User Identification", Proceedings of the International Conference on Artificial Intelligence (ICAI '06), Vol.2, pp. 344-348, Nevada, USA, June 2006.

[12] P. Kasprowski and J. Ober, "With The Flick of An Eye", Biometric Technology Today, Vol. 12, Issue 3, pp. 7-8, March 2004.

[13] A. Maeder, C. Fookes and S. Sridharan, "Gaze Based User Authentication for Personal Computer Applications", IEEE Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, pp. 727-730, October 2004.

[14] D. Schulz, "Mouse Curve Biometrics", Biometric Consortium Conference, 2006 Biometrics Symposium: Special Session on Research at the, pp. 1-6, September 2006.

[15] L. Wang, H. Ning, T. Tan and W. Hu, “Fusion of Static and Dynamic Body Biometrics for Gait Recognition”, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 14, No. 2, pp.149-158, February 2004.

[16] Z. Liu and S. Sarkar, “Outdoor Recognition at A Distance by Using Gait and Face”, Image and Vision Computing, Vol. 25, pp. 817-832, 2007A

[17] N. Boulgouris and Z. Chi, “Human Gait Recognition Based on Matching of Body Components”, Pattern Recognition, Vol. 40, pp. 1763-1770, 2007.

[18] P. Chaikan and M. Karnjanadecha, “The Use of Top-view Finger Image for Personal Identification”, Proceedings of 5th International Symposium on Image and Signal Processing and Analysis (ISPA ‘07), pp.343-346, Istanbul, Turkey, November 2007.

[19] K. Fan C. Lin, “The Using of Thermal Images of Palm-dorsa Vein-patterns for Biometric Verification”, IEEE Computer Society Proceedings of the 17th International Conference on Pattern Recognition (ICPR ‘04), Vol. 4, pp. 450-453, 2004.

[20] K. Phua, J. Chen, T. Dat and L. Shue, “Heart Sound as A Biometric”, Pattern Recognition, Vol. 41, Issue, 3, pp. 906-919, March 2008.

[21] L. Nanni and A. Lumini, “A Multi-matcher for Ear Authentication”, Pattern Recognition Letters, Vol. 28, pp. 2219-2226, 2007.

[22] D. Zhang, Z. liu and J. Yan, “Dynamic Tongueprint: A Novel Biometric Identifier”, Pattern Recognition, Vol. 43, Issue 3, pp. 1071-1082., March 2010.

[23] K. Ishikawa, K. Fujita and T. Hamamoto, “Biometrics Verification Using Dynamic and Static Eye Shapes”, IEEE International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS 06), pp. 235-238, Tottori, Japan., 2006.

[24] H. Ailisto, E. Vildjiounaite, K. Lindholm, S. Makela and J. Peltola, “Soft Biometrics-Combining Body Weight and Fat Measurements with Fingerprint Biometrics”, Pattern Recognition Letters 27, pp. 325-334, 2005.

[25] National Science and Technology Council (NSTC), “Biometric History”, http://www.biometrics.gov/Documents/BioHistory.pdf, [Last checked on 18-04-2010]

[26] E. German, “The History of Fingerprints” [Online], Available at http://www.onin.com/fp/fphistory.html, [Last checked on 18-04-2010]

[27] A. Wang, “Biometrics market: where are we now?”, Biometric Technology Today, Vol. 14, Issue 9, pp. 7-9, September 2006.

[28] Y. W and P. Li, “Biometric Identification: Is it an arduous task and long road?”, China Auto-ID, Vol. 3, pp. 36-44, June 2007.

[29] T. Best, “Expanding the Fusion Concept”, Biometric Technology Today, Vol. 14, Issue 6, pp. 7-8, June 2006.

[30] A. Martin , G. Doddington , T. Kamm , M. Ordowski and M. Przybocki, “The DET Curve in Assessment of Detection Task Performance”, In Proc. Eurospeech '97, pp. 1895—1898, 1997A

[31] A. Clark and C. Clark, “Performance Characterization in Computer Vision – A Tutorial”, http://peipa.essex.ac.uk/benchmark/tutorials/essex/tutorial.pdf, [Last checked on 18-04-2010]

[32] NIST Report to the United States Congress, “Summary of NIST Standards for Biometric Accuracy, Tamper Resistance, and Interoperability”, Available at http://www.itl.nist.gov/iad/894.03/NISTAPP_Nov02.pdf, [Last checked on 20-04-2010]

[33] R. Snelick, U. Uludag, A. Mink, M. Indovina, and A. Jain. Large Scale Evaluation of Multimodal Biometric Authentication Using State-of-the-Art Systems”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 3, pp. 450-455, March 2005.

[34] J. Jeffers and A. Arakala, “Fingerprint Alignment for A Minutiae-Based Fuzzy Vault”, IEEE Proceedings of the Biometrics Symposium, pp. 1-6, September 2007.

[35] M. Sarfraz and O. Hellwich, “Head Pose Estimation in Face Recognition Across Pose Scenarios”, in Proceedings of the Third International Conference on Computer Vision Theory and Applications (VISAPP 2008), Vol. 1, pp. 235-242, January 2008,

[36] A. Jain, “An Introduction to Biometric recognition”, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 14, No.1, January 2004.

[37] A. Ross, A. K. Jain, “Multimodal Biometrics: An Overview”, 12th European Signal Processing Conference (EUSIPCO), pp. 1221-1224, September 2004.

[38] C. Roberts, “Biometric Attack Vectors and Defences”, Computer and Security, Vol. 26, pp. 14-25, 2007.

[39] J. Galbally, J. Fierrez and J. Ortega-Garcia, “Vulnerabilities in Biometric Systems: Attacks and Recent Advances in Liveness Detection”, Proc. Spanish Workshop on Biometrics, SWB-07, Girona, Spain, June 2007.

[40] R. Rodriguesa, L. Lee and V. Govindarajua, “Robustness of Multimodal Biometric Fusion Methods Against Spoof Attack”, Journal of Visual Languages and Computing, Vol. 20, Issue 3, pp. 169-179, June 2009.

[41] T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino, “Impact of Artificial Gummy Fingers on Fingerprint Systems”, in Proceedings of SPIE, Optical Security and Counterfeit Deterrence Techniques IV, Vol. 4677, pp. 275-289, San Jose, USA, January 2002.

[42] L. Thalheim, J. Krissler and P. Ziegler, “Body Check: Biometrics Defeated”, c't Magazine article, June 3, 2002, English translation is available at http://www.extremetech.com/article2/0,2845,13919,00.asp [Last checked on 20-04-2010]

[43] A. Ross, A. Jain, “Information Fusion in Biometrics”, Pattern Recognition Letters 24, pp. 2115-2125, 2003.

[44] A. Jain, K. Nandakumar, A. Ross, “Score Normalization in Multimodal Biometric Systems”, Pattern Recognition, Vol. 38, No.12, pp. 2270-2285, December 2005.

[45] K. Nandakumar, “Multibiometric Systems: Fusion Strategies and Template Security,” PhD Thesis, Michigan State University, 2008.

[46] T. Ko, "Multimodal Biometric Identification for Large User Population Using Fingerprint, Face and Iris Recognition", IEEE Proceedings of the 34th Applied Imagery and Pattern Recognition Workshop (AIPR05), pp. 218 – 223, 2005.

[47] N. Poh and S. Bengio, "How Do Correlation and Variance of Base Classifiers Affect Fusion in Biometric Authentication Tasks?", IEEE Transactions on Signal Processing, Vol. 53, No. 11, pp. 4384-4396, November 2005.

[48] R. Frischholz and U. Dieckmann, "BioID: A Multimodal Biometric Identication System", IEEE Computer, Vol. 33, No. 2, pp. 64-68, February 2000.

[49] M. Kumar, T. Garfinkel, D. Boneh and T. Winograd, "Reducing Shoulder-surfing by Using Gaze-based Password Entry", ACM International Conference Proceeding Series, Proceedings of the 3rd symposium on Usable privacy and security, Vol. 229, pp. 13-19, 2007.

[50] C. Sanderson, K. Paliwal, "Information Fusion and Person Verification Using Speech and Face Information", Research Paper IDIAP-RR 02-33, IDIAP, September 2002.

[51] S. Lucey and T. Chen, "Improved Audio-Visual Speaker Recognition Via the Use of a Hybrid Combination Strategy" AVBPA, Vol. 2688, pp. 929-936, 2003.

[52] K. Toh and W. Yau, "Fingerprint and Speaker Verification Decisions Fusion Using a Functional Link Network", IEEE Transactions on Systems, Man and Cybernetics - Part A: Applications and Reviews, Vol. 35, No. 3, pp. 357-370, August 2005.

[53] A. Ross, A. Jain, "Information Fusion in Biometrics", Pattern Recognition Letters 24, pp. 2115-2125, 2003.

[54] S. Tulyakov, C. Wu and V. Govindaraju, "Iterative Methods for Searching Optimal Classifier Combination Function", [1]First IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS '07), pp. 1-5, December 2007.

[55] S. Dass, K. Nandakumar and A. Jain, "A Principled Approach to Score Level Fusion in Multimodal Biometric Systems," Proceedings of Audio and Video-Based Biometric Person Authentication (AVBPA '05), Vol. 2546, pp. 1049-1058, 2005.

[56] K. Nandakumar, Y. Chen, S. Dass and A. Jain, "Likelihood Ratio Based Biometric Score Fusion", IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), Vol. 30(2), pp. 342-347, 2008.

[57] K. Nandakumar, A. Jain and A. Ross, "Fusion in Multibiometric Identification Systems: What about Missing Data?", Proceedings of the 3rd International Conference on Advances in Biometrics, Springer LNCS, Vol. 5558, pp. 743-752, Alghero, Italy, 2009.

[58] F. Tsalakanidou, S. Malassiotis and M. Strintzis, "A 3D Face and Hand Biometric system for Robust User-friendly Authentication", Pattern Recognition Letters, Vol. 28, pp. 2238-2249, 2007.

[59] S. Tulyakov, Z. Zhang and V. Govindaraju, "Comparison of Combination Methods Utilizing T-normalization and Second Best Score Model", IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW '08), pp. 1-5, June 2008.

[60] W. Scheirer and T. Boult, "A Fusion Based Approach to Enhance Multi-modal Biometric Recognition system Failure", 2nd IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS '08), pp. 1-7, December 2008.

[61] S. Ribaric and I. Fratric, "Experimental Evaluation of Matching-Score Normalization Techniques on Different Multimodal Biometrics Systems", IEEE Mediterranean Electrotechnical Conference, pp. 498–501, Malaga, Spain, May 2006.

[62] G. Giacinto, F. Roli, and R. Tronci, "Score Selection Techniques for Fingerprint Multi-Modal Biometric Authentication", IEEE Computer Society Conference Proceedings of 13th International Conference on Image Analysis and Processing (ICIAP '05), Vol. 3617, pp. 1018-1025, 2005.

[63] R. Singh, M. Vatsa and A. Noore, "Integrated Multilevel Image Fusion and Match Score Fusion of Visible and Infrared Face Images for Robust Face Recognition", Pattern Recognition, Vol. 41, Issue 3, pp. 880-893, March 2008.

[64] M. Cheung, K. Yiu, M. Mak and S. Kung, "Multi-sample Fusion with Constrained Feature Transformation for Robust Speaker Verification", In 8th International Conference on Spoken Language Processing (ICSLP), pp. 1813-1816, Korea, October 2004.

[65] D. Bouchaffra and A. Amira, "Structural Hidden Markov models for biometrics: Fusion of face and fingerprint", Pattern Recognition, Vol. 41, No. 3, pp. 852-867, June 2007.

[66] X. Lu, Y. Wang and A. Jain, "Combining Classifiers for Face Recognition", IEEE International Conference on Multimedia & Expo (ICME 03), Vol. 3, pp 13-16, July 2003.

[67] A. Ross, A. Jain and J. Reisman, "A Hybrid Fingerprint Matcher", Pattern Recognition, Vol. 36, pp. 1661-1673, 2003.

[68] A. Jain, S. Prabhakar and S. Chen, "Combining Multiple Matchers for a High Security Fingerprint Verification System", Pattern Recognition Letters, Vol. 20, pp. 1371-1379, 1999.

[69] S. Prabhakar and A. Jain, "Decision-Level Fusion in Fingerprint Verification", Pattern Recognition, Vol. 35, pp. 861-874, 2002.

[70] N. Poh, J. Korczak and S. Bengio, "A Multi-Sample Multi-Source Model for Biometric Authentication", International Workshop on Neural Network and Signal Processing (NNSP), pp. 275-284, 2002.

[71] K. Toh, J. Kim and S. Lee, "Biometrics Scores Fusion based on Total Error Minimization" Pattern Recognition, Vol. 41 , Issue 3 , pp. 1066-1082, March 2008.

[72] J. Richiardi, K. Kryszczuk, "Quality Measures in Unimodal and Multimodal Biometric Verification", Proc. 15th European Conference on Signal Processing (EUSIPCO), Poznan, Poland, September 2007.

[73] D. Maurer and J. Baker, "Fusing Multimodal Biometrics with Quality Estimates via a Bayesian Belief Network", Pattern Recognition, Vol. 41, pp. 821-832, 2008.

[74] C. Watson, NIST Fingerprint Image Software 2 (NFIS2), Rel. 28-2.2, CDROM and Documentation, 2004.

[75] National Institute of Standards and Technology (NIST) Speech Quality Assurance (SPQA) Package 2.3 Documentation, available at http://www.nist.gov/speech/tools/spqa_23sphere25tarZ.htm

[76] L. Nanni and A. Lumini, "A Hybrid Wavelet-based Fingerprint Matcher", Pattern Recognition, Vol. 40, No. 11, pp 3146-3151, November 2007.

[77] P. Frick, R Beck and E. Berkhuijsen and I. Patrickeyev, "Scaling and Correlation Analysis of Galactic Images", Monthly Notices of the Royal Astronomical Society, Vol. 327, Issue 4, pp. 1145-1157, 2001.

[78] J. Fierrez-Aguilar, J. Ortega-Garcia, J. Gonzalez-Rodriguez and J. Bigun,, "Kernel-based Multimodal Biometric Verification Using Quality Signals", in Proceedings of SPIE, Vol. 5404, pp. 544-554, 2004.

[79] J. Fierrez, J. Ortega, J. Gonzalez and J. Bigun, "Discriminative Multimodal Biometrics Authentication based on Quality Measures", Pattern Recognition 38, pp 777-779, 2005.

[80] R. Youmaran, A. Adler, "Measuring Biometric Sample Quality in Term of Biometric Information", pp. 1-6, 2006 Biometrics Symposium, September 2006.

[81] K. Nandakumar, Y. Chen and A. Jain, "Quality-based Score Level Fusion in Multibiometric Systems", IEEE Proceedings of 18th International Conference on Pattern Recognition (ICPR '06), Vol. 4, pp. 473-476, 2006.

[82] A. Jain, Y. Chen and S. Dass, "Fingerprint Quality Indices for Predicting Authentication Performance', Proc. of the 5th International Conf. on Audio- and Video-Based Biometric Person Authentication (AVBPA), Rye Brook, NY, pp. 160-170, July 2005.

[83] Y. Chen, S. Dass and A. Jain, "Localized Iris Image Quality. Using 2-D Wavelets," in Proc. IEEE Int. Conf. Biometrics, pp. 373–381, 2006.

[84] N. Poh, T. Bourlai, and J. Kittler, "Quality-based Score Normalisation with Device Qualitative Information for Multimodal Biometric Fusion", IEEE Transactions on Systems, Man, and Cybernetics (part B), 2009, accepted for publication.

[85] available at http://www.omniperception.com/products/affinity

[86] N. Poh, T. Bourlai, J. Kittler and et. al., "Benchmarking Quality-dependent and Cost-sensitive Multimodal Biometric Fusion Algorithms", IEEE Transactions on Information Forensics and Security, Vol. 4, Issue 4, pp. 849-866, December 2009.

[87] N. Poh, T. Bourlai and J. Kittler, "A Multimodal Biometric Test Bed for Quality-dependent, Cost-sensitive and Client-specific Score-level Fusion Algorithms", Pattern Recognition, Vol. 43, Issue 3, pp. 1094-1105 , March 2010.

[88] R. Zewail, A. Elsafi, M. Saeb and N. Hamdy, "Soft and Hard Biometrics Fusion for Improved Identity Verification", in 47th IEEE International Midwest symposium on circuits and systems, Hiroshima, Japan, pp. 225-228, July 2004.

[89] J. Richiardi, P. Prodanov and A. Drygajlo, "A Probabilistic Measure of Modality Reliability In Speaker Verification", IEEE Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05), pp. 709–712, May 2005.

[90] A. Patra and S. Das, “Enhancing Decision Combination of Face and Fingerprint by Exploitation of Individual Classifier Space: An Approach to Multi-modal Biometry”, Pattern Recognition, Vol. 41, pp. 2298-2308, 2008.

[91] N. Poh and S. Bengio, “A Novel Approach to Combining Client-Dependent and Confidence Information in Multimodal Biometric”, 5th International Conference of Audio and Video Based Biometric Person Authentication (AVBPA), Springer LNCS , Vol. 3546, pp. 1120-1129, 2005.

[92] N. Poh and S. Bengio, “An Investigation of F-ratio Client-Dependent Normalisation on Biometric Authentication Tasks”, IEEE International Conference of Acoustics, Speech, and Signal Processing (ICASSP), Vol. 1 , pp. 721-724, 2005.

[93] N. Poh and S. Bengio, “Improving Fusion with Margin-Derived Confidence in Biometric Authentication Tasks”, 5th International Conference of Audio and Video Based Biometric Person Authentication (AVBPA), Springer LNCS 3546, pp. 474-483, 2005.

[94] A. Kale, A. RoyChowdhury and R. Chellappa, “Fusion of Gait and Face for Human Identification”, In IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Vol. 5, pp. 901-904, Montreal, Canada, May 2004.

[95] A. Sharma, “Step Integration based Information Fusion for Multimodal Biometrics”, appears in Systems, Signals and Image Processing, 2007 and 6th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services. pp. 213-216, November 2007.

[96] R. Brunelli and D. Falavigna, “Person Identication Using Multiple Cues”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 17, No.10, pp. 955-966, October 1995.

[97] A. Ross, K. Nandakumar, and A. Jain.” Handbook of Multibiometrics”. Springer, 2006.

[98] A. Khuwaja, “Merging face and finger images for human identification”, Pattern Analysis and Applications, Vol. 8, Issue 1, pp. 188-198, September 2005.

[99] K. Chang, K.. 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, September 2003.

[100] X. Xu and Z. Mu, “Feature Fusion Method Based on KCCA for Ear and Profile Face Based Multimodal Recognition”, Proc. IEEE International Conference on Automation and Logistics, pp. 312-314, November 2007.

[101] L. Lam and C. Suen, “Application of Majority Voting to Pattern Recognition: An Analysis of Its Behaviour and Performance”, IEEE Transactions on Systems, Man and Cybernetics-Part A: Systems and Humans, Vol. 27, No.5, pp. 553-568, September 1997.

[102] Q. Tao and R. Veldhuis, “Hybrid Fusion for Biometrics: Combining Score-Level and Decision-Level Fusion”, 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp. 1-6, 2008.

[103] L. Hong, A. Jain, “Integrating Faces and Fingerprints for Personal Identification”, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), Vol. 20, No. 12, December 1998.

[104] J. Bhatnagar, A. Kumar and N. Saggar, "A Novel Approach to Improve Biometric Recognition Using Rank Level Fusion", IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-6, USA, June 2007.

[105] M. Indovina, U. Uludag, R. Snelick, A. Mink, A. Jain, "Multimodal Biometric Authentication Methods: A COTS Approach", Proc. MMVA, Workshop Multimodal User Authentication, pp. 99-106, December 2003.

[106] Y. Singh and P. Gupta, "Quantitative Evaluation of Normalization Techniques of Matching Scores in Multimodal Biometric Systems", ICB 2007, Springer LCNS 4642, pp. 574-583, 2007.

[107] F. Hampel, E. Ronchetti, P. Rousseeuw and W. Stahel, "Robust Statistics - The Approach Based on Influence Functions", Wiley, 2005.

[108] J. Kittler, "On Combining Classifiers", IEEE Transactions on Pattern Analysis and Machine Intelligence, (TPAMI),Vol. 20, No. 3, pp. 226 – 239, March 1998.

[109] H. Vajaria, T. Islam, P. Mohanty and et. al., "Evaluation and Analysis of a Face and Voice Outdoor Multi-biometric System", Pattern Recognition Letters, Vol. 28, pp. 1572-1580, 2007.

[110] S. Krawczyk1 and A. Jain, "Securing Electronic Medical Records Using Biometric Authentication", AVBPA 2005, Springer LNCS, Vol. 3546, pp. 1110-1119, 2005.

[111] G. Shakhnarovich, L. Lee and T.J. Darrell, "Integrated Face and Gait Recognition from Multiple Views", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 439-446, Hawaii, USA, December 2001.

[112] A. Jain, A. Ross, "Learning User Specific Parameters in A Multibiometric System," International Conference on Image Processing (ICIP), pp. 57-60, 2002.

[113] C. Bergamini, L. Oliveira, A. Koerich and R. Sabourin, "Combining Different Biometric Traits with One-class Classification", Signal Processing, Vol. 89, pp. 2117-2127, 2009.

[114] H. Korves, L. Nadel, B. Ulery and D. Bevilacqua "Multi-biometric Fusion: From Research to Operations", Sigma, Summer 2005.

[115] J. Yu, "Information Fusion in Multimedia Information Retrieval: An Overview" Data and Vision Weekly Seminar, Department of Computer Science, University of Texas, San Antonio, USA, Spring 2007.

[116] A, Teoh, S. Samad and A. Hussein, "Nearest Neighbourhood Classifiers in a Bimodal Biometric Verification System Fusion Decision Scheme", Journal of Research and Practice in Information Technology, Vol. 36, No. 1, February 2004.

[117] H. Liu, Y. Wang, T. Tan, "Multi-Modal Data Fusion for Person Authentication based on Improved ENN", Acta Automatica Sinica, Vol. 30, No. 1, 2004.

[118] J. Wang, P. Neskovic and L. Cooper, "Improving Nearest Neighbor Rule with a Simple Adaptive Distance Measure", Pattern Recognition Letters, Vol. 28, pp. 207-213, 2007.

[119] S. Manocha and M. Girolami, "An Empirical Analysis of the Probabilistic K-nearest Neighbour Classifier", Pattern Recognition Letters, Vol. 28, pp. 1818-1824, 2007.

[120] L. Breiman, J. Friedman, C. Stone and R. Olshen, “Classification and Regression Trees”, Wadsworth, 1984.

[121] Y. Ma, B. Cukic and H. Singh, “A Classification Approach to Multi-biometric Score Fusion”, Springer LNCS, Vol. 3546, pp. 484-493, June 2005.

[122] J. Quinlan, “C4.5: Programs for Machine Learning”, Morgan Kaufmann Publishers, 1993.

[123] P. Verlinde and G. Ghollet, “Comparing Decision Fusion Paradigms Using k-NN Based Classifiers, Decision Trees and Logistic Regression in a Multi-modal Identity Verification Application”, 2nd international Conference on AVBPA, pp. 188-193, Washington, USA, 1999.

[124] S. Ben-Yacoub, Y. Abdeljaoued and E. Mayoraz, “Fusion of Face and Speech Data for Person Identity Verication”, IEEE Transactions on Neural Networks, Vol. 10, No. 5, pp. 1065-1075, September 1999.

[125] S. Gutta, J. Huang, I. Imam and H. Wechsler, “Face and Hand Gesture Recognition Using Hybrid Classifiers”, IEEE Proceedings of the International Conference on Automatic Face and Gesture Recognition (ICAFGR ’96), pp. 164-169, 1996.

[126] Y. Wang, T. Tan, and A. Jain, “Combining Face and Iris Biometrics for Identity Verification”, 4th International Conference on Audio and Video based Biometric Person Authentication (AVBPA), pp. 805-813, Guildford, UK, June 2003.

[127] B. Boser, I. Guyon and V. Vapnik, “A training algorithm for optimal margin classifiers”, in Fifth Annual ACM Workshop on COLT, pp. 144-152, Pittsburgh, PA, 1992.

[128] S. Kung and M. Mak, “On Consistent Fusion of Multimodal Biometrics”, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP ’06), Vol. 5, pp. 1085-1088, May 2006.

[129] J. Fierrez-Aguilar, J.Ortega-Garcia and J. Gonzalez-Rodriguez, “Fusion Strategies in Multimodal Biometric Verification”, IEEE Proceedings of the 2003 International Conference of Multimedia and Expo (ICME ‘03), Vol.. 3, pp. 5-8, 2003.

[130] J. Fierrez, J. Ortega, D. Garcia, J. Gonzalez, “A Comparative Evaluation of Fusion Strategies for Multimodal Biometric Verification,” AVBPA 2003, LCNS 2688, pp. 830-837, 2003.

[131] N. Poh and S. Bengio, “Non-Linear Variance Reduction Techniques in Biometric Authentication”, in Multimodal User Authentication Workshop, Santa Barabara, pp. 123-130, 2003.

[132] J. Fierrez, D. Garcia, J. Ortega and J. Gonzalez, “Adapted User-Dependant Multimodal Biometric Authentication Exploiting General Information,” Pattern Recognition Letters 26, pp. 2628-2639, 2005.

[133] B. Ulery, A. Hicklin, C. Watson and et. al., “Studies of Biometric Fusion”, Technical Report IR 7346, NIST, September 2006.

[134] B. Ulery, W. Fellner, P. Hallinan and et. al., “Studies of Biometric Fusion (Appendix C): Evaluation of Selected Biometric Fusion Techniques”, Technical Report IR 7346, NIST, September 2006.

[135] J. Fierrez, D. Garcia, J. Ortega and J. Gonzalez, “Bayesian Adaptation for User-Dependent Multimodal Biometric Authentication”, Pattern Recognition 38, pp 1317-1319, 2005.

[136] M. Bengherabi, L. Mezai, F. Harizi, A. Guessoum, M. Cheriet “Robust authentication using Likelihood Ratio and GMM for the fusion of voice and face Risk Bounds for Mixture Density Estimation”, in International Conference on Signals, Circuits and Systems, pp. 1-6, 2009.

[137] R. Raghavendra, R. Ashok and G. Kumar, “Multimodal Biometric Score Fusion Using Gaussian Mixture Model and Monte Carlo Method”, Journal of Computer Science and Technology, Vol. 25, No. 4, pp. 771-782, 2010.

[138] M. Figueiredo and A. Jain, “Unsupervised Learning of Finite Mixture Models,” IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), Vol..24, No. 3, pp. 381-396, March, 2002.

[139] R. Tronci, G. Giacinto and F. Roli, “Dynamic Score Selection for Fusion of Multiple Biometric Matchers”, IEEE Computer Society Conference Proceedings of 14th International Conference on Image Analysis and Processing (ICIAP ‘07), pp. 15-22, 2007.

[140] C. Chia, N. Sherkat and L. Nolle, “Confidence Partition and Hybrid Fusion in Multimodal Biometric Verification System,” International Conference on Biometrics ID Management and Multimodal Communication (BioID_MultiComm’ 09), Madrid, Spain, Springer LNCS 5707, pp. 212-219, September 2009.

[141] C. Chia, N. Sherkat and L. Nolle, “Towards a Best Linear Combination for Multimodal Biometric Fusion,” 20th International Conference on Pattern Recognition (ICPR’ 10), pp. 1176-1179, Istanbul, Turkey, August 2010.

[142] National Science and Technology Council (NSTC), “Biometric Testing and Statistics”, Biometrics Reference Room, Available at http://www.biometrics.gov/Documents/BioTestingAndStats.pdf, [Last checked on 23-06-2010]

[143] R. Bolle, S. Pankanti, N. Ratha “Evaluation Techniques for Biometrics-based Authentication Systems (FRR)”, 15th International Conference on Pattern Recognition (ICPR '00) , Vol. 2, pp. 2831- 2837, Barcelona, Spain, September 2008.

[144] J. Xue and D. Titterington, “Comments on @On Discriminative vs. Generative Classifiers : A Comparison of Logistic Regression and Naive Bayes», Neural Processing Letters, Vol. 28, Issue 3, pp. 169-187, December 2008.

[145] V. Vapnik, “The Nature of Statistical Learning Theory”, Springer-Verlag, 1995.

[146] N. Poh and S. Bengio, “Database, Protocol and Tools for Evaluating Score-Level Fusion Algorithms in Biometrics Authentication”, Pattern Recognition, Vol. 39, No. 2, pp. 223-233, February 2006.

[147] N. Poh and S. Bengio, “Can Chimeric Persons Be Used in Multimodal Biometric Authentication Experiments?”, in 2nd International Machine Learning and Multimodal Interaction Workshop (MLMI'05), Springer LNCS 3869, pp. 87-100, 2005.

[148] R. Singh, M. Vatsa, A. Ross and A. Noore, “Biometric Classifier Update Using Online Learning: A Case Study in Near Infrared Face Verification”, Image and Vision Computing, Vol.28, Issue 7, pp. 1098-1105, July 2010.

[149] K. Woods, “Combination of Multiple Classifiers Using Local Accuracy Estimates” IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), Vol. 19, No. 4, pp. 405-410, April 1997.

[150] J. Li, A. Barron, “Mixture Density Estimation” , in Advances in Neural Information Processing Systems 12, pp. 279-285, 1999.

[151] A. Rakhlin1, D. Panchenko and S. Mukherjee, “Risk Bounds for Mixture Density Estimation”, in ESAIM: P&S, Vol. 9, pp. 220-229, June 2005.

[152] F. Grubbs, “Procedures for Detecting Outlying Observations in Samples ”, Technometrics, Vol.. 11, No. 1, pp. 1-21, February 1969.

[153] L. Xu, A. Krzyzak and C. Suen, “Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition”, IEEE Transactions on Systems, Man and Cybernetics, Vol. 22, No. 3, pp. 418-435, May 1992.

[154] T. Ho, J. Hull and S. Srihari, “Decision Combination in Multiple Classifier systems”, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), Vol. 16, No. 1, pp. 66-75, January 1994.

[155] S. Kung, and M. Mak, “Machine Learning for Multi-Modality Genomial Signal Processing”, IEEE Signal Processing Magazine, pp. 117-121, May 2006.

[156] F. Deravi, M Fairhurst, R. Guest, N. Mavity and A. Canuto, “Intelligent agents for the management of complexity in multimodal biometrics”, Universal Access in the Information Society, Vol. 2, No. 4, pp. 293-304, 2003.

[157] R. Plamondon and S.N. Srihari, “On-line and off-line handwriting recognition: a comprehensive survey.”, in IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 22, No.1, pp 63-84, 2000.

[158] L. Ballard, F. Monrose and D. Lopresti, “Biometric authentication revisited: understanding the impact of wolves in sheep's clothing.”, Proceedings of the 15th conference on USENIX Security Symposium, pp. 29-41, 2006.

Confidence Partition and Hybrid Fusion in Multimodal Biometric Verification System

Chaw Chia, Nasser Sherkat and Lars Nolle

Department of Computing and Technology
Nottingham Trent University, Nottingham, UK
{chaw.chia, nasser.sherkat, lars.nolle}@ntu.ac.uk

Abstract. Sum rule fusion is a very promising multimodal biometrics fusion approach. However, it is proposed not to widely applying it across the multimodal biometrics score space. By examining the score distributions of each biometric matcher, it can be seen that there exist confidence regions which enable the introduction of the Confidence Partition in multimodal biometric score space. It is proposed that the Sum rule can be replaced by the Min or the Max rule in the Confidence Partition to further increase the overall verification performance. The proposed idea which is to apply the fusion rules in a hybrid manner has been tested on two publicly available databases and the experimental results shows 0.3% ~ 2.3% genuine accept rate improvement at relatively low false accept rate.

1 Introduction

Multimodal biometrics have attracted great interest in the biometric research field in recent years. Given its potential to out perform single biometrics verification, many researchers have put their efforts in exploration of different integration techniques. However, integration at the score level is the most preferred approach due to the effectiveness and ease in implementation [1]. The Sum rule, one of the well known score level fusion rule is a method that simply utilises the addition of each biometric scores as fusion result. Surprisingly, it appears to be outperforming many complicated fusion algorithms [2] and being widely employed in biometric research [3, 4, 5, 6, 7, 8]. Through sensitivity analysis, Kittler concluded that the superior performance of the Sum rule is due to it resilient ability to estimate error [9].

In this paper, the assignment of Confidence Partitions (CP) in multimodal biometrics score space has been introduced. Instead of applying the Sum rule over the complete region of multimodal biometrics score space, we suggest to replace the Sum rule in the different CPs with more appropriate rules (Min and Max rule in this paper). This scheme enables the fusion of multimodal biometrics in a hybrid manner including the Sum rule.

Figure 1 illustrates a typical biometric matcher score distribution that includes a genuine user and an impostor score distributions. There is a significant overlap region of the curves that causes the main difficulty to classify the claimant into the genuine user or impostor groups. The shaded regions outside the overlap part are confidence regions. They represent the regions where only a single class of users can be found. Although the Sum rule performs well to produce reliable fusion scores, when the biometric scores are located in a confidence region it is suggested to apply a more appropriate rule instead of the Sum rule for a more reliable fusion score, for example the Min, Max rule [9] or the decision fusion rule [10].

The rest of the paper is organised as follows: Section 2 provides details about the proposed integration method. Section 3 presents the databases used, experiments, results and their analysis. Finally section 4 concludes the paper.

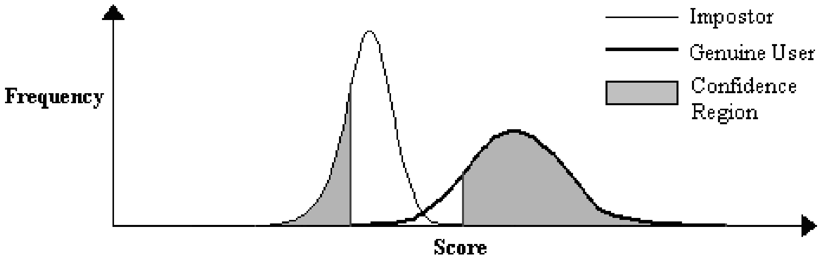


Fig. 1. Biometric matcher score distribution.

2 Confidence Partition and Hybrid Fusion

Even though the proposed idea is feasible in higher dimensional score space, it has only being used for the bimodal biometrics fusion in this paper. First of all, the score distributions of bimodal matchers are constructed (the distributions will be modeled by density estimation algorithm in future research). The regions within the distribution where only one type of user (either genuine user or impostor) is present are marked. Within the genuine user score distribution, the marked region is termed as genuine user confidence region whereas the region within the impostor score distribution is termed as impostor confidence region. Consequently, a two dimensional score space is created. The Genuine User Confidence Partition (GCP) in the score space is assembled from both modalities' genuine user confidence regions. Also the Impostor Confidence Partition (ICP) in the score space is formed by both modalities' impostor confidence regions.

Prior to applying the fusion rule, we need to normalise the scores from different biometric matchers into a common domain before they can be effectively combined [11]. The simplest normalisation technique is the Minmax normalisation [11] which is showed in (1). It is a rule that maps the biometric scores into the interval between 0 and 1. The minimum value (min) and the maximum value (max) of the score distribution can be estimated from a set of matching scores. The notations shown in the equation represent the follows: S_i is the biometric score of user i, S'_i represents the normalised score for user i, S_{fi} is the after fusion score for the particular user, K represents the total number of matchers.

$$S'_i = \frac{S_i - \min}{\max - \min} \tag{1}$$

By introducing the CP, multiple rules can be applied over the multimodal biometric system in a hybrid manner. In this work, the rules (2) ~ (4) have been applied. The hybrid fusion scheme is implemented according to scenario shown in (5).

1. Sum Rule:

$$S_{fi} = \sum_{k=1}^{K} S'_{i,k} \ , \forall i \qquad \textbf{(2)}$$

2. Min Rule:

$$S_{fi} = \min(S'_{i,1}, S'_{i,2}, ..., S'_{K}) \ , \forall i \qquad \textbf{(3)}$$

3. Max Rule:

$$S_{fi} = \max(S'_{i,1}, S'_{i,2}, ..., S'_{i,K}) \ , \forall i \qquad \textbf{(4)}$$

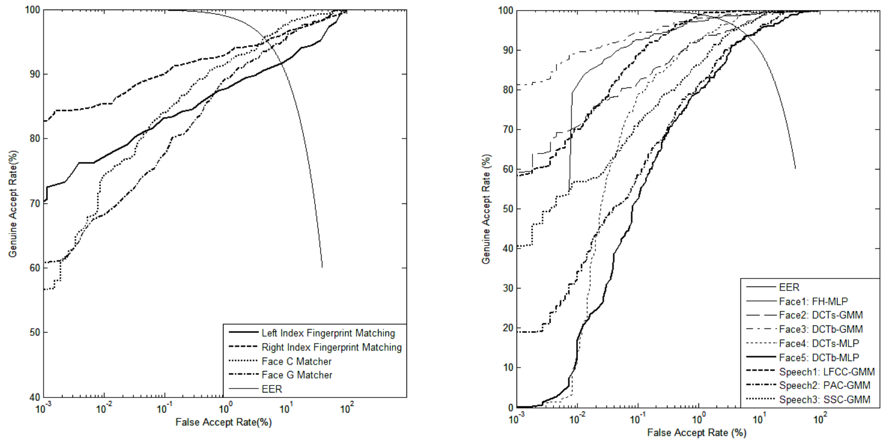
4. Hybrid Rule:

$$S_{fi} = \begin{cases} \text{Apply Min Rule, when} < S'_{i,1} , S'_{i,2} ,\ldots,S'_{i,K} > \text{fall in ICP.} \\ \text{Apply Max Rule, when} < S'_{i,1} , S'_{i,2} ,\ldots,S'_{i,K} > \text{fall in GCP.} \\ \text{Apply Sum Rule, elsewhere.} \end{cases} \qquad \textbf{(5)}$$

As shown in equation (5), for the partitions where we have high confidence from the biometric matchers we can apply the Min or Max rule which is considered as the more appropriate rule than the Sum rule. The non-confidence partition which is the complement region of the CP exhibits the part that can be easily misclassified. Due to the superior performance of Sum rule in dealing with the estimation error mentioned in section 1, we employ this rule to these non-confidence partitions.

3 Experimental Results

The proposed method has been tested on two publicly available databases, which are the NIST-BSSR1 multimodal database [12] and the XM2VTS benchmark database [13]. In the NIST-BSSR1 multimodal database, there are 517 genuine user scores and 266,772 impostor scores, whereas the XM2VTS database (evaluation set) includes 400 genuine user scores and 111,800 impostor scores. Both the databases are truly multimodal (chimeric assumption is not in used [14]). The performance graphs of each matcher in the databases are depicted in figure 2.



(a) **(b)**

Fig. 2. Performance of baseline matchers (a) NIST-BSSR1 Multimodal Matchers and (b) XM2VTS Matchers Performance.

Only the best and the worst biometric matchers from each modality are chosen for the experiments. In the NIST-BSSR1 multimodal database, the right index fingerprint has been paired with the facial matcher C and the left index fingerprint has been paired with the facial matcher G to develop the best and worst multimodal biometrics fusion respectively. For the XM2VTS database, the best facial matcher DCTb-GMM is paired with the best speech matcher LFCC-GMM whereas the worst DCTb-MLP facial matcher is paired with the worst speech matcher PAC-GMM in the experiments.

The GCP and ICP are assigned manually according to the figures in table 1. All the fusion results based on the best and worst multimodal matcher's combination are graphically shown in figure 3 and figure 4. Their numerical results are also presented in table 2 and table 3. This is worth mention that the genuine accept rate (GAR) listed in the tables is reported to be 0.001% of the false accept rate (FAR).

Table 1. Assignment of Confidence Partitions in the experiments.

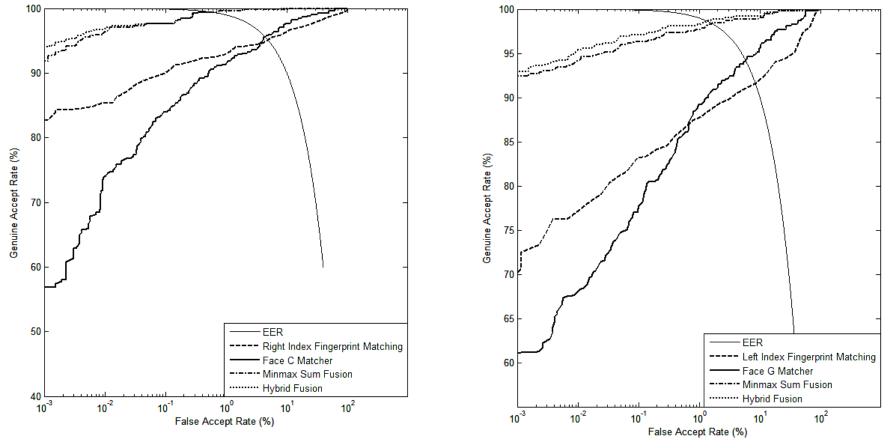
	Impostor	Genuine User	Non-
	Confidence Partition	Confidence Partition	Confidence Partition
NIST-BSSR1 Best Matchers	$S_{face} < 0.55$ $S_{finger} < 0.15$	$S_{face} > 0.34$ $S_{finger} > 0.20$	Other than Confidence Partitions
NIST-BSSR1 Worst Matchers	$S_{face} < 0.35$ $S_{finger} < 0.09$	$S_{face} > 0.20$ $S_{finger} > 0.20$	Other than Confidence Partitions
XM2VTS Best Matchers	$S_{speech} < 0.48$ $S_{face} < 0.44$	$S_{speech} > 0.41$ $S_{face} > 0.60$	Other than Confidence Partitions
XM2VTS Worst Matchers	$S_{speech} < 0.43$ $S_{face} < 1.00$	$S_{speech} > 0.67$ $S_{face} > 0.79$	Other than Confidence Partitions

From the graphical and numerical results shown in figures 3 and 4 and tables 2 and 3, we can conclude that the proposed method outperforms the Sum rule fusion especially at lower FAR even though there are no significant improvements of the equal error rate (EER) which is the rate where FAR is equal to the false reject rate (FRR).

The best matchers hybrid fusion for the NIST-BSSR1 dataset achieved 93% GAR which is 0.7% more than the Sum rule whereas in the XM2VTS the best matchers hybrid fusion achieved 96.3% GAR which is 0.3% better than the Sum rule. The GAR improvement becomes more obvious in the worst matchers hybrid fusion in both databases. The hybrid fusion gains additional 2.1% and 2.3% GAR improvement compared to the Sum rule in NIST-BSSR1 and XM2VTS databases respectively. The relative Sum rule performances are 91.9% and 62.0% in NIST-BSSR1 and XM2VTS. As it can be observed from the scatter plots, the best matchers achieved very good separation between the genuine user and impostor score distribution. Therefore the Sum rule is able to produces a very reliable fusion score. As a result no significant hybrid fusion improvement can be obtained when comparing it with the Sum rule. However, the Sum rule performs poorer to fuse multimodal biometrics with lower authentication rate. In this case, the use of a hybrid fusion rule leads to an improvement over the Sum rule fusion. Like the work shown in [4], our work justifies again that the higher accuracy biometric system leaves less room for improvement.

In a bimodal biometric system, the Sum fusion score can be considered as the average value between the Min fusion score and the Max fusion score. Further, within the confidence partition the difference between minimum score and maximum score will not be significant. As a result, the improvements of the GAR achieved in the experiments are within the range between 0.3%~2.3%. It is assumed that the improvement can be further increased when the Min and Max rules being replaced by a higher degree confidence fusion rule, for example the decision fusion rule.

In fact, the improvement also relies on a more accurate assignment of the CP and depends on the amount of claimants whose multimodal biometric scores are falling in the confidence partitions. The more scores falls in the CP, the more improvement of the hybrid fusion can be obtained.

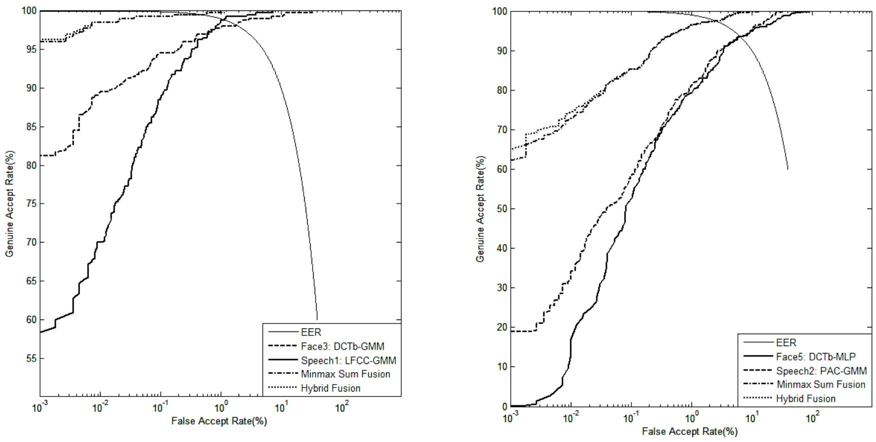


(a) (b)

Fig. 3. Performance of the NIST-BSSR1 bimodal biometrics fusion on (a) the best multimodal matchers and (b) the worst multimodal matchers.

Table 2. Accept rates and error rates of NIST-BSSR1 Multimodal database single biometrics and the combined multimodal biometrics.

	Fingerprint		Face		Sum		Hybrid	
	EER	GAR	EER	GAR	EER	GAR	EER	GAR
Best Matchers	8.6%	70.0 %	5.8%	61.1%	1.6%	92.3%	1.3%	93.0%
Worst Matchers	4.5%	82.7%	4.3%	56.9%	0.5%	91.9%	0.5%	94.0%



(a) **(b)**

Fig. 4. Performance of the XM2VTS bimodal biometrics fusion on (a) the best multimodal matchers and (b) the worst multimodal matchers.

Table 3. Accept rates and error rates of XM2VTS single biometrics and their combined multimodal biometrics.

	Face		Speech		Sum		Hybrid	
	EER	GAR	EER	GAR	EER	GAR	EER	GAR
Best Matchers	1.8%	81.3%	1.1%	58.3%	0.5%	96.0%	0.5%	96.3%
Worst Matchers	6.4%	0.0%	6.4%	19.0%	2.5%	62.0%	2.5%	64.3%

4 Conclusions

After the introduction of the confidence partition, we have proposed to use more appropriate fusion rules (Min and Max rule in this paper) in the confidence partitions instead of Sum rule. This approach enables the rule based fusion to be applied in a hybrid manner that includes Sum, Min and Max rules. In the preliminary experiments, we showed that the manually operated hybrid rule performed better than the Sum rule. The future exploration will be focusing on automatic assignment of confidence partitions across the biometric score space. An investigation into integration of decision rule in the developed hybrid fusion framework will also be conducted.

References

1. A. Ross, A. K. Jain, "*Multimodal Biometrics: An Overview*", 12th European Signal Processing Conference (EUSIPCO), pp. 1221-1224, September, 2004.
2. A. Ross, A. K. Jain, "*Information Fusion in Biometrics*", Pattern Recognition Letters 24, pp. 2115-2125, 2003.
3. A. K. Jain, A. Ross, "*Learning User Specific Parameters in A MultiBiometric System*", IEEE ICIP, pp. 57-60, 2002.
4. M.Indovina, U. Uludag, R. Snelick, A. Mink, A. K. Jain, "*Multimodal Biometric Authentication Mehods: A COTS Approach*", Proc. MMVA, Workshop Multimodal User Authentication, pp. 99-106, December, 2003.
5. H. Ailisto, E. Vildjiounaite, K. Lindholm, S. Makela. J. Peltola, "*Soft Biometrics- Combining Body Weight and Fat Measurements with Fingerprint Biometrics*", Pattern Recognition Letters 27, pp. 325-334, 2006.
6. Xiaoguang Lu Yunhong Wang , A. K. Jain, "*Combining Classifiers for Face Recognition*", ICME '03. vol.3 , pp 13-16, July 2003.
7. D. Bouchaffra, A. Amira, "*Structural Hidden Markov models for biometrics: Fusion of face and fingerprint*", Pattern Recognition , vol. 41 , no. 3 ,pp. 852-867, March 2008.
8. L. Nanni, A. Lumini, "*A Hybrid Wavelet-based Fingerprint Matcher*", Pattern Recognition, vol. 40 , no. 11, pp 3146-3151 ,November 2007.
9. Josef Kittler, "*On Combining Classifiers*", IEEE Transactions On Pattern Analysis and Machine Intelligence, vol. 20, no. 3, pp. 226 - 239 March 1998.
10. L. Lam, C. Y. Suen, "*Application of Majority Voting to Pattern Recognition: An Analysis of Its Behaviour and Performance*", IEEE Trans. Systems Man Cybernet. Part A: Systems Humans 27(5) pp. 553-568, 1997.
11. A. K. Jain, K. Nandakumar, A. Ross, "*Score Normalization in Multimodal Biometric Systems*" Pattern Recognition, vol. 38, no.12, pp. 2270-2285, December 2005.
12. National Institute of Standards and Technology: NIST Biometric Scores Set, http://www.itl.nist.gov/iad/894.03/biomtricscores
13. N. Poh, S. Bengio, "*Database, Protocol and Tools for Evaluating Score-Level Fusion Algorithms in Biometrics Authentication*", Pattern Recognition, vol.39, no. 2, pp.223-233, February 2006.
14. N. Poh, S. Bengio, "*Can Chimeric Persons Be Used in Multimodal Biometric Authentication Experiments?*", 2nd Int'l Machine Learning and Multimodal Interaction Workshop 2005 (MLMI'05), LNCS 3869, pp. 87-100, 2005.

Towards a Best Linear Combination for Multimodal Biometric Fusion

Chaw Chia, Nasser Sherkat, Lars Nolle
Computer and Science Department,
Nottingham Trent University, UK
{chaw.chia, nasser.sherkat,
lars.nolle}@ntu.ac.uk

Abstract

Owing to effectiveness and ease of implementation Sum rule has been widely applied in the biometric research field. Different matcher information has been used as weighting parameters in the weighted Sum rule. In this work, a new parameter has been devised in reducing the genuine/imposter distribution overlap. It is shown that the overlap region width has the best generalization performance as the weighting parameter amongst other commonly used matcher information. Furthermore, it is illustrated that the equal weighted Sum rule can generally perform better than the Equal Error Rate and d-prime weighted Sum rule. The publicly available databases: the NIST-BSSR1 multimodal biometric and Xm2vts score sets have been used.

1. Introduction

Combining several modalities of biometrics is a promising approach to achieve high verification rate. The Sum Rule is one of the effective score level fusion approaches for multiple biometric score combination. Although it is a very simple algorithm, it out performs some of the complex fusion methods [1] and has been extensively applied in various biometric fusion attempts. However, different biometrics tend to perform differently. A weighted Sum rule is preferred since it can be used to indicate the importance of each biometric modality in the fusion.

The Weighted Sum Rule is a linear boundary in bimodal biometric score space where the weighting can be viewed as a means to adjust its gradient. There exists a best linear boundary for every single operating point. An exhaustive search for this best linear boundary through searching for the optimal weights has been conducted and other similar works have been reported in [2, 3]. Although exhaustive searching promises high verification rate, a training session that might be complex or time consuming is requested for every single operating point.

For the biometric verification problem the classification errors arise from the overlap region. A smaller overlap region tends to produce less classification error. Therefore the aim of this work is to achieve the best linear combination by reducing the overlap region through adjusting the gradient of the linear boundary. This proposed method is described in the following section.

The Equal Weighted (EW), Equal Error Rate Weighed (EERW) and D-Prime Weighted (DW) Sum rules are commonly used methods. A further contribution of this work is to carry out the comparison between these methods and the proposed work. To the best of our knowledge, similar comparative work has not been reported. The third section describes the experimental setup and the results analysis and the fourth section concludes the paper

2. Non-Confidence Width Weighted Sum Rule

Fig. 1 illustrates a typical biometric matcher score distribution that includes the genuine user and impostor score distributions. There is a significant overlap region of the curves that causes the difficulty to classify the claimant into genuine or impostor groups. The grey regions outside the overlap part are confidence regions where only a single class of users can be found and therefore the samples can be safely rejected or accepted. Whereas the samples in the overlap region can only be classified with referring to the threshold boundary.

The width of the overlap region is termed Non-Confidence Width (NCW). NCW can be determined from the difference between the matcher *k's* maximum impostor score Max_k^I and the minimum genuine user score Min_k^G as shown in (1).

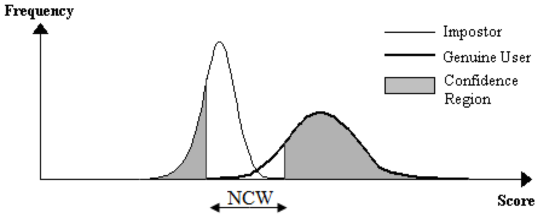


Figure 1. Biometric matcher score distribution.

$$NCW_k = Max_k^I - Min_k^G \qquad (1)$$

In a practical biometric matcher, the NCW will always exist. Some of the reasons for formation of non-confidence region are the noise in the sensed data, the interclass similarities in the feature space of multiple users and intraclass variations that are typically caused by users who incorrectly interacting with the sensor [4].

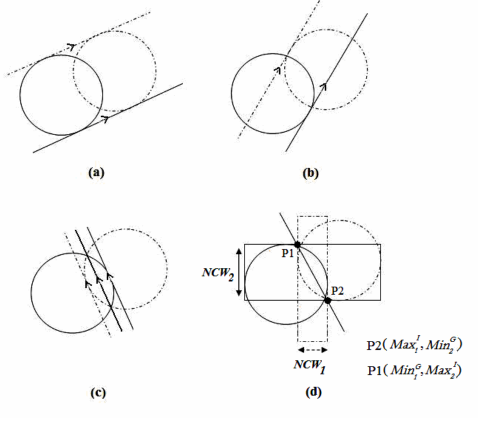


Figure 2. Reducing the overlap region by adjusting the gradient of the linear boundary.

Fig. 2 illustrates how the NCW weighted Sum rule can aid in reducing the overlap region in bimodal biometric score space to maximise the verification rate. Two of the overlap circles in this figure represent the approximated scatter of the genuine user and impostor scores. The straight lines are separation boundaries with different verification thresholds. It denotes varying the decision threshold is a process of moving the boundary while preserving its gradient. As shown by 2(a), the circles' area between the separation boundaries is at maximum. The bounded area is the area where the samples cannot be clearly classified by the boundary. However, by adjusting the gradient of the boundary, the bounded area can be reduced as shown in 2(b). When the boundary is parallel to the line connecting the intersection points of the two circles, the bounded area is restricted to a minimum as shown in 2(c). A smaller bounded area contains less non-confidence samples so a better ROC can be obtained. Therefore it is desired that the boundary has the same gradient with the line connecting the circle's points of intersection. As 2(d) depicts, this specific gradient m can be approximated by the NCW of the two matchers where m in 2(d) equals to (2).

$$m = \frac{Min_2^G - Max_2^I}{Max_1^I - Min_1^G} \qquad (2)$$

$$\frac{1}{NCW_1}x + \frac{1}{NCW_2}y = c \qquad (3)$$

By using the common form of a linear equation (3) can be derived. *c* is an adjustable threshold for controlling the boundary position. In this weighted Sum rule, biometric matcher's scores *x* and *y* are inverse proportionally weighted by their NCW. Their respective weights W_k can be obtained by applying (4) so that $\sum_{k=1}^{K} W_k = 1$ where the K is the total matcher number. Therefore (3) can be rewritten as (5), S_i is the fused score for user *i and* $S'_{i,k}$ is his biometric score that is generated by matcher k.

$$W_k = \frac{\frac{1}{NCW_k}}{\sum_{k=1}^{K} \frac{1}{NCW_k}} \tag{4}$$

$$S_i = \sum_{k=1}^{K} W_k S'_{i,k} \tag{5}$$

This method is termed as Non-Confidence Width Weighted Sum rule (NCWW). Three of the following commonly used weighted schemes are carried out in the experiments for comparison and used to evaluate the effectiveness of the proposed method.

a) Equal Weighted: Multiple biometrics will be assigned the same weight, W_k:

$$W_{k=1...K} = \frac{1}{K} \tag{6}$$

b) EER Weighted: Equal Error Rate (EER) is where the Falsely Accept Rate (FAR) equals to Falsely Reject Rate (FRR). It is inverse proportionally used as weighting parameter in (7).

$$W_k = \frac{\frac{1}{EER_k}}{\sum_{k=1}^{K} \frac{1}{EER_k}} \tag{7}$$

c) D-Prime Weighted: D-prime has been used to statistically measure the separation of impostor and genuine user biometric scores as depicted in (8). The μ_k^G and μ_k^I are the genuine score and impostor score mean where σ_k^G and σ_k^I are their standard deviations. The associated matcher weight is directly proportional to its d-prime as shown in (9).

$$d'_k = \frac{\mu_k^G - \mu_k^I}{\sqrt{(\sigma_k^G)^2 + (\sigma_k^I)^2}} \tag{8}$$

$$W_k = \frac{d_k^{'}}{\sum_{k=1}^{K} d_k^{'}} \tag{9}$$

3. Experimental Setup and Results

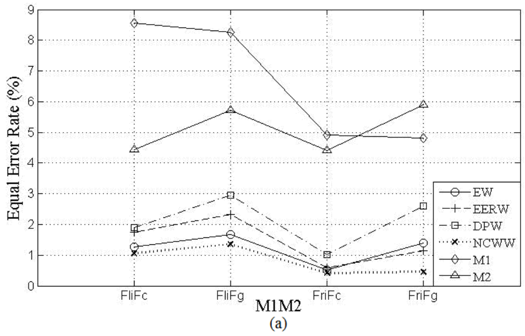
Although the proposed method can be generalized to higher dimensions, it was decided to focus on investigating the performance of bimodal biometric fusion. This is because it is desirable to examine the effectiveness of the proposed method before introducing further complexities. The NIST-BSSR1 multimodal database [5] and the Xm2vts benchmark database [6] are used for the experiments. These databases are truly multimodal. Since no matcher information is given, each of the databases has been evenly separated into two parts. The weighting parameters are obtained through the first half part and the rest for the testing purpose.

The BSSR1 multimodal database consists of 517 genuine user scores and 266,772 impostor scores from the user's left and right fingerprints (Fli and Fri) and facial scores from two matchers (Fc and Fg). Fli and Fri are paired with Fc and Fg to form four different bimodal biometric fusion experiments.

For the Xm2vts score database, there are five facial matchers (F1~F5) and three speech matchers (S6~S8). It contains 295 individuals' speech and facial scores. There are 1000 genuine scores and 151,800 impostor scores from both the development set and evaluation set. Even though the training and testing partitions have been defined by the author, in our experiments this partitioning has not been done. To examine the statistical significance of the proposed method, the testing and training scores are mixed and randomly chosen to form equal training and testing sets. Different permutations between five facial matchers and three speech matchers create 15 bimodal biometric fusion experiments.

Before the weighted Sum rule is applied, the biometric scores are normalised through Min-Max normalisation [7]. Each of the experiments has been repeated 100 times through different partitions of the databases. Due to the page constraint, only the average EER are reported graphically in fig. 3. In the figure, the M1 and M2 represent the matchers involve in the experiments

The superior performance of the NCWW can be justified from the figure that NCWW's EER seem to be always the lowest in both of the databases. In NIST-BSSR1, the NCWW obtains the best EER over all the experiments where the obtained EERs are in the range of 0.41%~1.35%. Whereas in the Xm2vts database, the NCWW obtains the best EER from 9 experiments out of 15. The rest of the results are very comparable to the best EER where the differences from the best EER is just in the range of 0.01%~0.08%. However, the NCWW EERs vary in the range of 0.29%~~1.72%. At lowest operating points, NCWW and EW obtain 2 best GAR respectively at FAR equals to 0.001% in BSSR1 experiments. EW, EERW and DPW obtain 3 best GAR correspondingly at FAR equals to 0.002% in Xm2vts whereas NCWW obtains 6 best GAR out of 15.



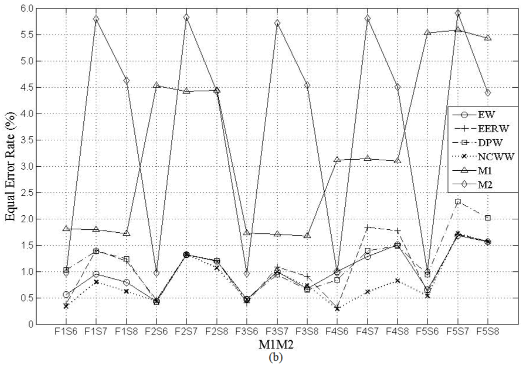


Figure 3. NIST BSSR1 (a) and Xm2vts (b) bimodal biometric fusion average results.

It can be seen from fig. 4, that NCWW not only obtains the best EERs in all the NIST-BSSR1 experiments, their performance standard deviations are the lowest also. The standard deviations are no more than 0.35 for the NIST-BSSR1 where the Xm2vts standard deviations are no more than 0.23.

It is clear that NCWW has the best generalisation capability in producing the best EER. From the comparison of the other three methods, surprisingly, the EW which is independent of any parameters performs broadly better than the other two parametric methods. From fig. 4, in contrast to the EW and NCWW, the DPW and EERW generally perform more inconsistently and worse in several experiments. This is because of the EER and d-prime is very sensitive against the sample variation. Furthermore EER cannot be a key factor in weighting the discrimination power of a matcher. For example a matcher with a lower EER may have higher lowest FAR than the other one. As for the d-prime, it includes the samples outside the overlap region for its calculation. However the errors of verification arise from the overlap region. Therefore it cannot be appropriately used as a weighting parameter. Due to the nature of the biometric matcher that produce similarity score tends to give a high score for a genuine user and a low score for an impostor (i.e. like the circles in fig. 2 shown, the two circles will normally align from lower left to upper right direction), EW which has an gradient of -1 for its linear boundary will generally out perform the other two methods.

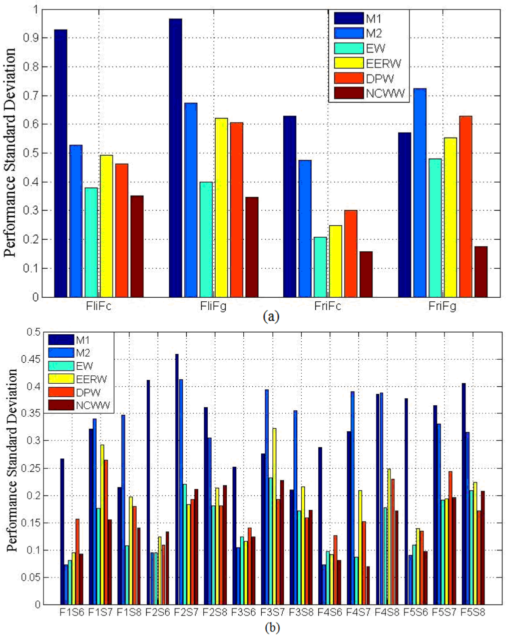


Figure 4. NIST BSSR1 (a) and Xm2vts (b) 100 bimodal biometric fusion averaged result standard deviations.

4. Conclusion

Although EER and d-prime have been applied in the weighted Sum Rule in biometric research, it is shown that they cannot be reliably used. EW can generally perform better than both of them but NCWW will be preferred. This is because it is aimed at reducing the overlap region and in fact the score normalization process can be eliminated by using NCWW.

Furthermore, NCWW is preferred not only due its simplicity, parameter accessibility and no need for parameter tuning, but because it out performs other non-linear methods. As an example the Gaussian Mixture Modeling likelihood ratios test in [8] is considered. The multi-biometric fusion research conducted through NIST-BSSR1 obtains mean GAR at 99.1% (with FAR equals to 0.01%). The 95% confidence interval on increase in GAR with respect to the best single matcher performance is [13.5%, 14%]. By using the same experimental setting, the NCWW can generate mean GAR at 99.2% with the 95% confidence interval on increase in GAR is [13.7%, 14.1%]. The Xm2vts multibiometric experiment that follows the partition in [6] obtains 98.7% GAR in (9) whereas NCWW performs better again at 99.1% GAR.

In this work, we have demonstrated how to maximize the verification rate by reducing the overlap region in a bimodal biometric linear fusion problem. However, basing the detection of the overlap width on Max_k^I and Min_k^G difference alone is sensitive to outliers and may lead to unreliability. Therefore for a more consistent performance, the NCW's definition needs to be extended to include the corresponding density and other overlap region information.

References

[1]A. Ross and A. Jain, "Information Fusion in Biometrics", Pattern Recognition Letters 24, pp. 2115-2125, 2003.

[2] A. Jain and A. Ross, "Learning User Specific Parameters in A MultiBiometric System", ICIP, pp. 57-60, 2002.

[3]C. Bergamini, L.

Olivieira, A. Koerich and R. Sabourin, "Combining different biometric traits with one-class classification", Signal

Processing Journal, vol. 89, pp. 2117-2137, 2009.

[4]A. Jain, "An Introduction to Biometric recognition", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 14, No.1, 2004.

[5]National Institute of Standards and Technology: NIST Biometric Scores Set, http://www.itl.nist.gov/ iad/ 894.03/ biomtric scores

[6]N. Poh and S. Bengio, "Database, Protocol and Tools for Evaluating Score-Level Fusion Algorithms in Biometrics Authentication," Pattern Recognition, vol.39, no. 2, pp.223-233, February 2006.

[7]A. Jain, K. Nandakumar and A. Ross, "Score Normalization in Multimodal Biometric Systems" Pattern Recognition, vol. 38, no.12, pp. 2270-2285, December.

[8]K. Nandakumar, Y. Chen and A..Jain, "Likelihood Ratio-Based Biometric Score Fusion," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, pp.342-347, 2008.