

2006

Evaluation and performance prediction of multimodal biometric systems

Nevena Samoska
West Virginia University

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Recommended Citation

Samoska, Nevena, "Evaluation and performance prediction of multimodal biometric systems" (2006). *Graduate Theses, Dissertations, and Problem Reports*. 1744.
<https://researchrepository.wvu.edu/etd/1744>

This Thesis is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Thesis in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Thesis has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

Evaluation and Performance Prediction of Multimodal Biometric Systems

Nevena Samoska

Thesis submitted to the College of Engineering and Mineral Resources
at West Virginia University
in partial fulfillment of the requirements for the degree of

Master of Science

in

Computer Science

Approved by:

Bojan Cukic, Ph.D., Committee Chairperson

Arun Ross, Ph.D. Committee Co-Chairperson

Katerina Goseva – Popstojanova, Ph.D., Committee member

Lane Department of
Computer Science and Electrical Engineering

Morgantown, West Virginia

2006

Keywords: biometrics, multimodal biometrics, fusion, normalization,
goodness of fit

Copyright © 2006 Nevena Samoska

Abstract

Evaluation and Performance Prediction of Multimodal Biometric Systems

Nevena Samoska

Multibiometric systems fuse the evidence presented by different biometric sources in order to improve the matching accuracy of a biometric system. In such systems, information fusion can be performed at different levels; however, integration at the matching score level is the most commonly used approach due to the tradeoff between information content and accessibility. This work develops a tool in order to analyze the impact of various normalization schemes on the matching performance of score-level fusion algorithms. The tool permits the systematic evaluation of different fusion rules after employing normalizing and mapping the match scores of different modalities into a common domain. Furthermore, it provides a method to fit various parametric models to the score distribution and analyze the goodness of fit statistic based on the Chi-Squared and Kolmogorov-Smirnov tests. Experimental results on multiple datasets indicate the benefits of normalization, the role of parametric distributions and the variations in matching performance on different databases.

Acknowledgments

Firstly, I take pleasure in expressing my sincere gratefulness to my advisers Dr. Bojan Cukic and Dr. Arun Ross. They have been highly inspiring and prompt when I needed their help. They have been extremely patient with me and have guided this work. It has been a great privilege to be their student.

I would like to thank my committee member, Dr. Katerina Goseva Popstojanova for her continued support throughout my study. I am grateful to her for her constant encouragement and wise suggestions that helped shape not only my thesis but also my approach towards quality research.

I extend my special gratitude to my mother Nada and my husband Damjan for standing by me during this journey that I undertook, for all their love and care. Without their unwitting support, this thesis would not have been possible.

Special thanks to my lab mates and my friend and roommate Dejan. It has been a great experience to share knowledge and work with you.

Table of Contents

ABSTRACT	ii
ACKNOWLEDGMENTS	iii
TABLE OF CONTENTS	iv
LIST OF FIGURES	v
LIST OF TABLES	vii
CHAPTER 1 INTRODUCTION	1
1.1 MOTIVATION	1
1.2 GOAL	2
1.3 ORGANIZATION.....	2
CHAPTER 2 RELATED WORK	3
2.1 OVERVIEW OF BIOMETRICS	3
2.2 PERFORMANCE MEASURES.....	5
2.3 MULTIMODAL BIOMETRICS.....	8
2.3.1 <i>Fusion in Biometrics</i>	10
2.3.2 <i>Combining the modalities at matching score level - Normalization methods</i>	12
2.3.3 <i>Combining the modalities at matching score level – Fusion methods</i>	15
2.2.3 <i>Data Partitioning</i>	17
CHAPTER 3 ANALYSIS OF THE SCORE DISTRIBUTIONS	19
3.1 GOODNESS OF FIT STATISTICS	19
3.1.1 <i>Maximum Likelihood Estimation</i>	19
3.1.2 <i>Chi-Squared statistics</i>	21
3.1.3 <i>Kolmogorov–Smirnov statistic</i>	22
3.1.4 <i>P-values and Critical values</i>	22
CHAPTER 4 IMPLEMENTATION	26
4.1 MUBI OFFLINE ANALYSIS TOOL	26
CHAPTER 5 EXPERIMENTS AND RESULTS	29
5.1. MSU DATASET	29
5.1.1 <i>Experiments with normalization and fusion methods on MSU dataset</i>	30
5.2 NIST DATASET	35
5.2.1 <i>Experiments with normalization and fusion methods on NIST dataset</i>	36
5.3. EXPERIMENTS WITH SYNTHETICALLY GENERATED SCORES USING THE MSU AND NIST DATASETS	41
5.3.1 <i>Case study 1 – MSU dataset</i>	47
5.3.2 <i>Case study 2 – MSU dataset</i>	50
5.3.3 <i>Case study 1 - NIST dataset</i>	55
5.3.4 <i>Case study 2 – NIST dataset</i>	58
CHAPTER 6 DISCUSSION AND CONCLUSION	61
FUTURE WORK	64
APPENDIX A	66
REFERENCES	70

List of Figures

FIGURE 1: BASIC ARCHITECTURE OF A BIOMETRIC SYSTEM.	5
FIGURE 2: FAR VS. FRR ROC CURVE. [11]	6
FIGURE 3: EER EXAMPLE. [10]	7
FIGURE 4: DIFFERENT ROC CURVES FOR TWO MATCHERS WITH SAME D-PRIME. [10]	8
FIGURE 5: ROC CURVE FOR A SYSTEM UTILIZING MULTIPLE BIOMETRIC TRAITS. [4]	9
FIGURE 6: BIOMETRIC FUSION OPTIONS IN THE VERIFICATION MODE OF OPERATION. [15]	10
FIGURE 7: NORMALIZATION METHODS WINDOW IN THE MUBI TOOL.	12
FIGURE 8: FACE NORMALIZED WITH MIN/MAX – PLOT FROM MUBI TOOL.	13
FIGURE 9: FACE NORMALIZED WITH DECIMAL SCALING – PLOT FROM MUBI TOOL.	13
FIGURE 10: FACE NORMALIZED WITH Z-SCORE – PLOT FROM MUBI TOOL.	14
FIGURE 11: FACE NORMALIZED WITH MEDIAN AND MAD – PLOT FROM MUBI TOOL.	14
FIGURE 12: FACE NORMALIZED WITH TANH NORMALIZATION– PLOT FROM MUBI TOOL.	15
FIGURE 13: FUSION METHODS IMPLEMENTED IN THE MUBI TOOL.	16
FIGURE 14: PARTITION DATA METHODS IN THE MUBI TOOL.	18
FIGURE 15: GOODNESS OF FIT TESTS IN THE MUBI TOOL.	23
FIGURE 16: MUBI TOOL – MSU DATASET.	27
FIGURE 17: ADD MODALITY IN THE SYSTEM FROM MUBI TOOL.	28
FIGURE 18: ROC CURVE PLOT FOR ONE BIOMETRIC SYSTEM FROM MUBI TOOL.	28
FIGURE 19: DENSITY PLOT OF FINGER GENUINE AND IMPOSTER SCORES.	31
FIGURE 20: DENSITY PLOT OF FACE GENUINE AND IMPOSTER SCORES.	31
FIGURE 21: DENSITY PLOT OF HAND GENUINE AND IMPOSTER SCORES.	32
FIGURE 22: DIFFERENT NORMALIZATION METHODS FOLLOWED BY THE SIMPLE SUM RULE ON THREE MODALITIES FROM MSU DATASET.	32
FIGURE 23: DIFFERENT FUSION METHODS AFTER MIN-MAX NORMALIZATION ON THREE MODALITIES FROM MSU DATASET.	34
FIGURE 24: OVERALL WINNER FOR THE MSU DATASET – TANH NORMALZITION WITH SIMPLE SUM RULE FUSION METHOD.	34
FIGURE 25: NIST (BSSR1) DATASET [8].	35
FIGURE 26: DENSITY PLOT OF THE GENUINE AND IMPOSTER SCORES OF THE FINGER LEFT INDEX.	36
FIGURE 27: DENSITY PLOT OF GENUINE AND IMPOSTER SCORES OF THE FINGER RIGHT INDEX.	37
FIGURE 28: DENSITY PLOT OF THE GENUINE AND IMPOSTER SCORES OF THE FACE C MATCHER.	37
FIGURE 29: DENSITY PLOT OF THE GENUINE AND IMPOSTER SCORES OF THE FACE G MATCHER.	38
FIGURE 30: FUSION SIMPLE SUM RULE WITH MIN-MAX NORMALIZATION METHOD ON NIST DATASET.	38
FIGURE 31: DIFFERENT FUSION METHODS AFTER MIN-MAX NORMALIZATION ON NIST DATASET.	39
FIGURE 32: DIFFERENT NORMALIZATION METHODS FOLLOWED BY SIMPLE SUM RULE ON NIST DATASET.	39
FIGURE 33: DENSITY PLOTS OF THE FINGER SCORES FOR GENERATED DATA AND ORIGINAL DATA.	47
FIGURE 34: DENSITY PLOTS OF THE FACE SCORES FOR GENERATED DATA AND ORIGINAL DATA.	47
FIGURE 35: DENSITY PLOT OF THE HAND SCORES FOR GENERATED DATA AND ORIGINAL DATA.	48
FIGURE 36: FUSION METHODS WITH DIFFERENT NORMALIZATIONS ON THE GENERATED DATA.	48
FIGURE 37: BGI FUSION METHOD ON GENERATED DATA COMPARED WITH BGI METHOD ON ORIGINAL DATA.	49

FIGURE 38: SIMPLE SUM RULE FUSION METHOD AFTER MIN-MAX NORMALIZATION ON GENERATED DATA COMPARED WITH SAME METHOD ON ORIGINAL DATA.	49
FIGURE 39: DENSITY PLOTS OF THE FINGER SCORES FOR GENERATED DATA AND ORIGINAL DATA.	50
FIGURE 40: DENSITY PLOT OF THE FACE SCORES FOR GENERATED DATA AND ORIGINAL DATA.	50
FIGURE 41: DENSITY PLOT OF THE HAND SCORES FOR GENERATED DATA AND ORIGINAL DATA.	51
FIGURE 42: FUSION METHODS WITH DIFFERENT NORMALIZATIONS ON THE GENERATED DATA.	51
FIGURE 43: BGI FUSION METHOD ON GENERATED DATA COMPARED WITH BGI METHOD ON ORIGINAL DATA.	52
FIGURE 44: PERFORMANCE OF THE SIMPLE SUM RULE FUSION METHOD AFTER Z-SCORE NORMALIZATION ON GENERATED DATA COMPARED WITH THE SAME METHOD ON THE ORIGINAL DATA.	52
FIGURE 45: DENSITY PLOTS OF THE FINGER LEFT INDEX SCORES FOR GENERATED DATA AND ORIGINAL DATA.	55
FIGURE 46: DENSITY PLOTS OF THE FINGER RIGHT INDEX SCORES FOR GENERATED DATA AND ORIGINAL DATA.	55
FIGURE 47: DENSITY PLOTS OF THE FACE C MATCHER SCORES FOR GENERATED DATA AND ORIGINAL DATA.	56
FIGURE 48: DENSITY PLOTS OF THE FACE G MATCHER SCORES FOR GENERATED DATA AND ORIGINAL DATA.	56
FIGURE 49: FUSION METHODS WITH DIFFERENT NORMALIZATIONS ON THE GENERATED DATA.	57
FIGURE 50: SIMPLE SUM RULE FUSION METHOD AFTER MEDIAN NORMALIZATION ON GENERATED DATA COMPARED WITH SAME METHOD ON ORIGINAL DATA.	57
FIGURE 51: DENSITY PLOTS OF THE FINGER LEFT INDEX SCORES FOR GENERATED DATA AND ORIGINAL DATA.	58
FIGURE 52: DENSITY PLOTS OF THE FINGER RIGHT INDEX SCORES FOR GENERATED DATA AND ORIGINAL DATA.	58
FIGURE 53: DENSITY PLOTS OF THE FACE C MATCHER SCORES FOR GENERATED DATA AND ORIGINAL DATA.	59
FIGURE 54: DENSITY PLOTS OF THE FACE C MATCHER SCORES FOR GENERATED DATA AND ORIGINAL DATA.	59
FIGURE 55: FUSION METHODS WITH DIFFERENT NORMALIZATIONS ON THE GENERATED DATA.	60
FIGURE 56: SIMPLE SUM RULE FUSION METHOD AFTER MEDIAN NORMALIZATION ON GENERATED DATA COMPARED WITH SAME METHOD ON ORIGINAL DATA.	60

List of Tables

TABLE 1: GENUINE ACCEPTANCE RATE (GAR) (%) OF DIFFERENT NORMALIZATION TECHNIQUES FOLLOWED BY THE SIMPLE SUM RULE FUSION METHOD AND THE BGI METHOD AT 0.1% FALSE ACCEPTANCE RATE (FAR).....	33
TABLE 2: GENUINE ACCEPTANCE RATE (GAR) (%) OF DIFFERENT NORMALIZATION TECHNIQUES FOLLOWED BY THE SIMPLE SUM RULE FUSION METHOD AND THE BGI METHOD AT 0.1% FALSE ACCEPTANCE RATE (FAR).....	40
TABLE 3: RESULTS FROM THE CHI-SQUARE AND KOLOMOGOROV-SMIRNOV TEST ON MSU DATASET.....	41
TABLE 4 RESULTS FROM THE CHI-SQUARE AND KOLOMOGOROV-SMIRNOV TEST ON NIST DATASET.....	42
TABLE 5: SUMMARY FOR THE SELECTED DISTRIBUTIONS FOR MSU DATASET.....	43
TABLE 6: SUMMARY FOR THE SELECTED DISTRIBUTIONS FOR NIST DATASET.....	44
TABLE 7: GENUINE ACCEPTANCE RATE (GAR) (%) OF DIFFERENT NORMALIZATION TECHNIQUES FOLLOWED BY THE SIMPLE SUM RULE FUSION METHOD AND THE BGI METHOD AT 0.1% FALSE ACCEPTANCE RATE (FAR).....	46
TABLE 8 GENUINE ACCEPTANCE RATE (GAR) (%) OF DIFFERENT NORMALIZATION TECHNIQUES FOLLOWED BY THE SIMPLE SUM RULE FUSION METHOD AND BGI AT 0.1% FALSE ACCEPTANCE RATE (FAR).	54
TABLE 9: SUMMARY OF THE EXPERIMENTS ON GENERATED DATA.....	63

Chapter 1 Introduction

Chapter 1

Introduction

1.1 Motivation

There are several requirements in current authentication applications, which drive the need for multimodal biometrics: universal enrollment requirements, accuracy and integrity requirements, risk and viability of spoofing, suitability in usage environment. It has been proven [9] that multimodal biometric systems give better performance than unimodal biometric systems. As multimodal systems are more reliable, a lot of research is conducted to find the best combination of biometric modalities and fusion methods. Currently, there are no tools developed that will allow testing as well as comparison of different schemes on multimodal biometric systems. There is a need to develop standardized biometric testing protocols to provide and ensure valid results. To make the evaluation of biometric systems easier, it is necessary to develop a user-friendly off-line analysis tool, which will aid in evaluating the performance of matching score level fusion methods on multimodal biometric systems.

1.2 Goal

The goal of this study is to implement an off-line analysis tool, to examine the effect of different score normalizations and different fusions on the performance of a multimodal biometric system.

This work consists of four parts:

- i) a verification testing methodology for multimodal biometric systems,
- ii) an evaluation of normalization and fusion algorithms,
- iii) recommendations for designing multimodal biometric systems and
- iv) modular modeling of the software tool that will permit the addition of new methods and new approaches for testing.

1.3 Organization

The rest of this thesis is organized as follows. Chapter 2 includes an introduction to biometrics, unimodal biometrics, different normalization methods and fusion techniques, as well as performance measurements. Chapter 3 examines distribution fitting approaches using two-statistic tests, Chi-Square and Kolmogorov-Smirnov tests. P values and Critical values are explained as a measure for the hypothesis testing. Chapter 4 focuses on the implementation of the MultiBiometric (MUBI) tool. In Chapter 5, a description of MUBI towards conducting various experiments for performance evaluation of a multibiometric system is provided. Also the results for the MSU and NIST datasets utilizing the implemented normalization and fusion methods are presented. In addition, results from the experiments for the generated data are compared using goodness of fit statistic tests. Finally, Chapter 6 describes the conclusions derived from this study and the applications of this tool to the related fields.

Chapter 2 Related work

Chapter 2

Related Work

2.1 Overview of Biometrics

Biometrics refers to the automatic identification or verification of an individual based on his/her measurable physical characteristics (e.g. fingerprint, face, etc.) or personal behavioral traits (e.g. gait, signature, etc). Increasingly large number of applications require user authentication leading to a significant demand for biometric solutions.

A potential biometric trait is expected to meet the following listed requirements:

1. Universality - the trait should be possessed by each individual in the given population. [6].
2. Distinctiveness - the trait should be unique for each person within that population. [6]
3. Permanence - the trait should not change over a period of time with respect to the matching algorithm. [6].
4. Collectability - the trait should be easy to collect automatically in modern biometric systems and must be measurable quantitatively. [6].
5. Performance - the biometric trait should lend itself to fast and accurate identification. [6].

6. Acceptability - people should be able to accept the use of a certain biometric trait. [6].
7. Circumvention - reflects how easily the biometric trait can be spoofed using fraudulent methods [6].

Biometric traits can be physical or behavioral. The traits belonging to the first group are:

- DNA
- Iris
- Face
- Fingerprint
- Hand Geometry
- Palmprint
- Vein Pattern
- Ear
- Retina
- Footprint
- Odor

The second group consists of:

- Gait
- Signature dynamics
- Keystrokes dynamics
- Speech / Voice

Each of these biometric characteristics has its own weaknesses and strengths; no single biometric is expected to effectively meet all the requirements for varying applications. It is obvious that one trait cannot satisfy all the above requirements to the greatest degree. Accordingly, the choice of the biometric trait depends on the application.

The general architecture of a traditional biometric system is shown in Figure 1.

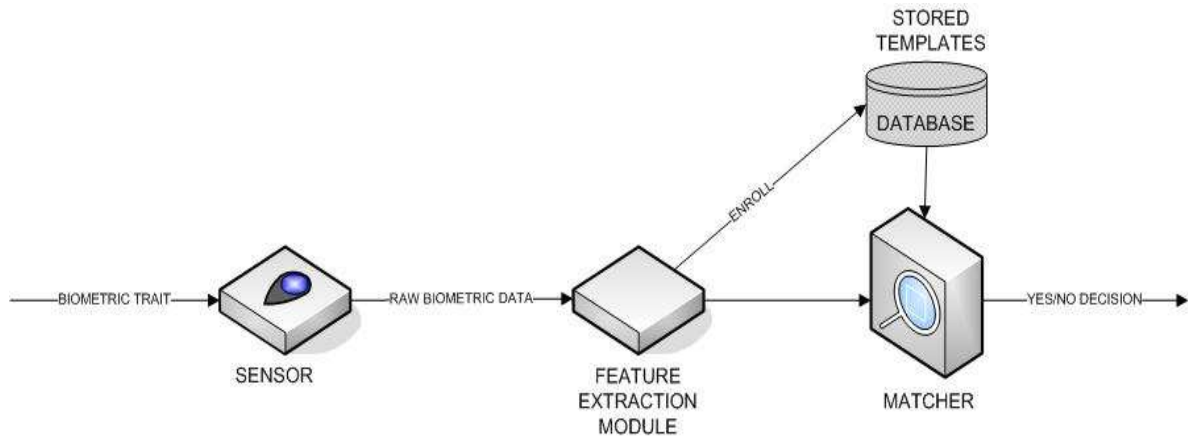


Figure 1: Basic architecture of a biometric system.

In this architecture, there are three main parts:

- the sensor captures the data from the biometric trait and sends raw biometric data to the feature extraction module
- the feature extraction module extracts selected information from the raw data and creates a unique feature vector for the biometric sample. In the enrollment phase, these features are stored in the database as templates.
- The matcher receives the extracted feature vector from the feature extraction module and compares it with the feature vector of one specific template (for verification) or multiple templates (for identification) which are already stored in the database.

2.2 Performance Measures

There are a number of biometric error types, expressed in error rates or error percentages. Usually the performance of biometric systems is measured in terms of two error rates, False Accept Rate (FAR) and False Reject Rate (FRR). The FAR refers to the errors that the biometric system makes by incorrectly declaring a match

between two different individuals and it is called a Type II error. The FRR refers to the errors that the biometric system makes when by mistake it does declare a match between biometric samples from the same individual and is called a Type I error. Some groups also call FAR as False Non-Match Rate (FNMR) and FRR as False Match Rate (FMR). To evaluate the performance of a biometric system usually a graph of FAR vs. FRR is plotted, recognized as a Receiver Operating Characteristic (ROC) curve. ROC curve presents a non-dimensional, basic technical performance measure for comparing two or more biometric systems.

Figure 2 indicates that for different biometric applications different operating points should be chosen which would result in different FRR and FAR. Sometimes the ROC curve is plotted as GAR (Genuine Accept Rate) vs. FAR. The complement of FRR is GAR ($GAR = 1 - FRR$). The ROC curve is a precise and complete specification of the performance of a biometric system, and it is very beneficial in comparing the performance of two or more biometric systems. It also displays the trade-offs between FAR and FRR over a wide range of thresholds.

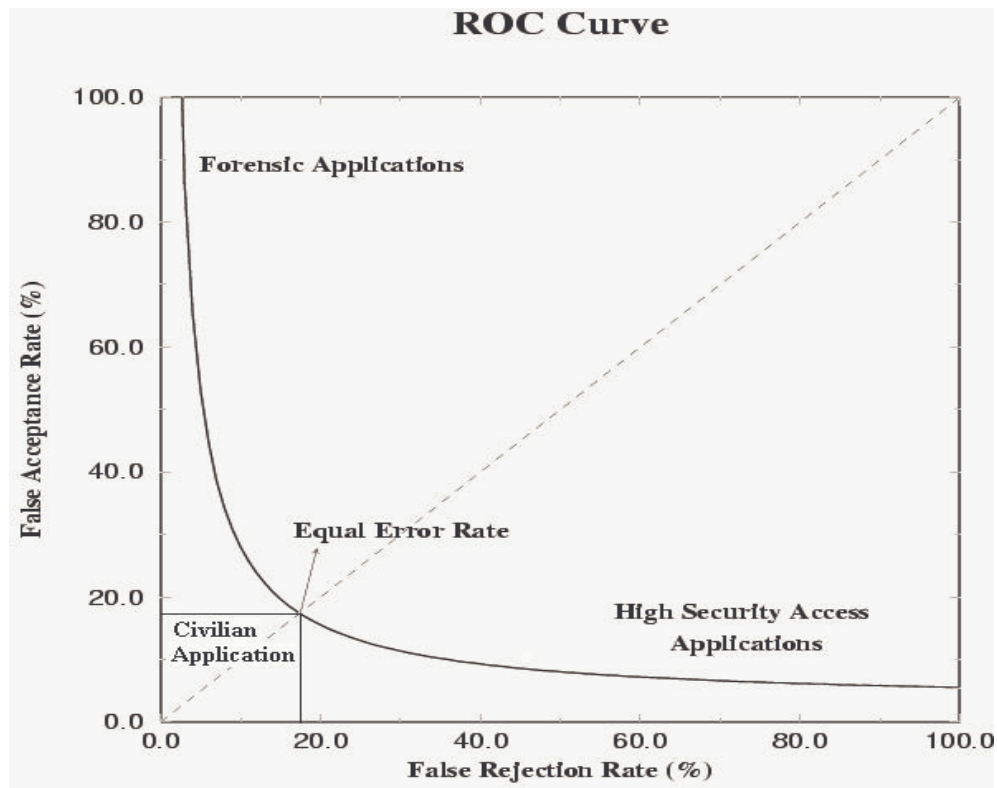


Figure 2: FAR vs. FRR ROC Curve. [11]

Besides the above-mentioned errors, Equal Error Rate (EER) and d-prime are also used to summarize the accuracy of a biometric system. EER is the point on the ROC curve where the FAR equals the FRR or it is the error rate occurring when the decision threshold of a system is set so that the proportion of false rejections are approximately equal to the proportion of false acceptances. For example in Figure 2, the EER is around 18%. The EER can inform us if one system performs better than other but only in narrow range of points $FAR = (EE_a, EE_b)$ and $FRR = (EE_a, EE_b)$. Beyond that range, the ROC curves may cross over each other and the EER would be invalid as displayed on Figure 3. That is why the EER is an unreliable summary of system accuracy.

D-Prime is a statistical measure of how well a biometric system can discriminate between different individuals. The larger the d-prime value, the better a biometric system can discriminate between individuals. This measure is most relevant if there is a significant difference in performance between the two biometric systems.

Figure 4 shows that the relative performance of the biometric system with identical d-prime depends on the chosen operating point.

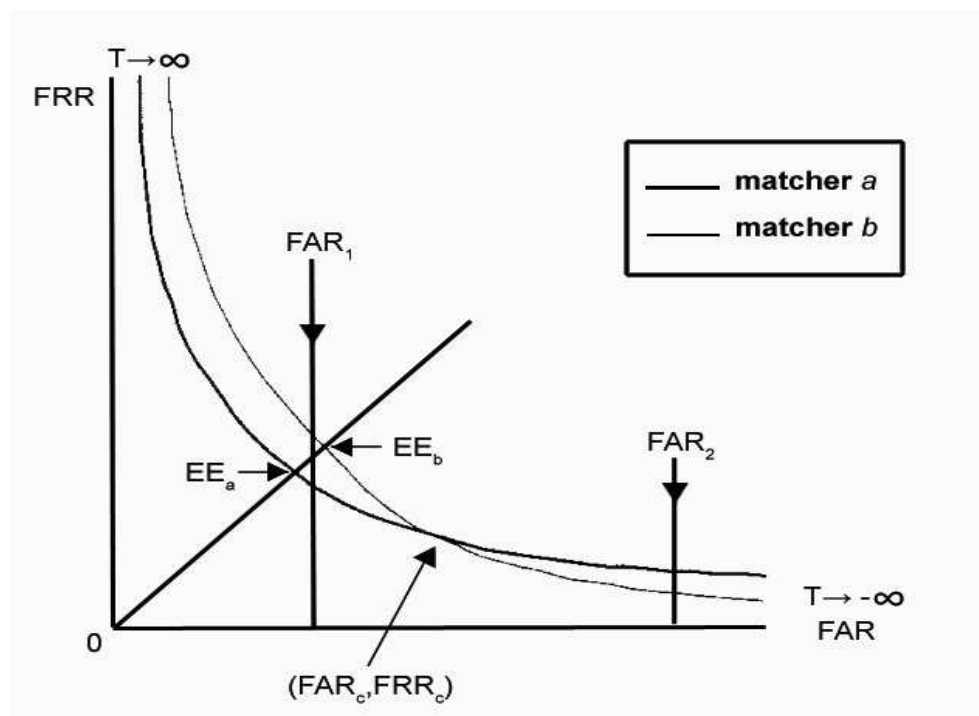


Figure 3: EER example. [10]

There are two more types of errors, namely Failure to Enroll (FTE) and Failure to Acquire (FTA) rates. The FTE rate measures the percentage of individuals who cannot be enrolled in the system. The FTA rate measures the percentage of time in which the biometric system cannot obtain the raw biometric data during enrollment or authentication.

It is obvious that every biometric system makes mistakes, and the true value of these errors cannot be computed or theoretically established. It is only possible to calculate statistical estimates of the errors once we have the database of biometric samples.

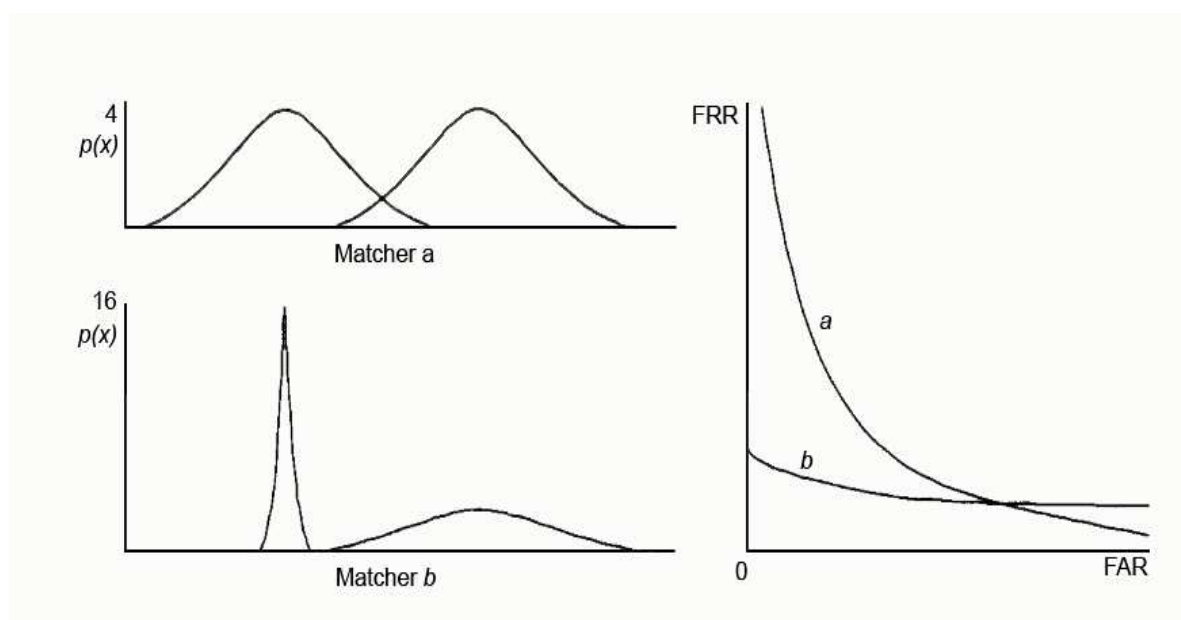


Figure 4: Different ROC curves for two matchers with same d' -prime. [10]

2.3 Multimodal biometrics

Multimodal biometric systems combine two or more modalities (biometric traits). These traits include fingerprints, hand-geometry, face, voice, iris, retina, gait, signature, palm-print, ear, etc. Some of the limitations of unimodal system are overcome by using multiple biometric modalities. A multimodal system offers increased performance, anti-spoofing, acceptable error rates, and the ability to operate on a large user population. Multimodal biometric systems offer an increase in anti-

spoofing by making it more difficult for an imposter to fool the system. It is more difficult because one has to forge multiple biometric traits and probably within a certain time limit. No biometric system can properly work for every user because of various factors, e.g. a mute person cannot use a voice recognition system; a person with missing fingers cannot use the fingerprint identification system. In addition, people may have similar traits that can be difficult to differentiate, e.g., twins will have the same DNA profile.

Another very important advantage in using multimodal biometric systems over unimodal biometric systems is the increase in performance. These systems are more reliable and provide higher verification rates and improved accuracy due to the presence of multiple and independent pieces of evidence. Figure 5 displays the results of a study conducted at Michigan State University evaluating the ROC curves of fingerprint, facial, and hand geometry systems. None of the three individual modalities are as accurate as compared to the combination of all three modalities. Many factors need to be considered when designing a multi-biometric system. These include the choice and number of biometric traits; the level in the biometric system at which information provided by multiple traits should be integrated; the methodology adopted to integrate the information; and the cost versus matching performance trade-off.

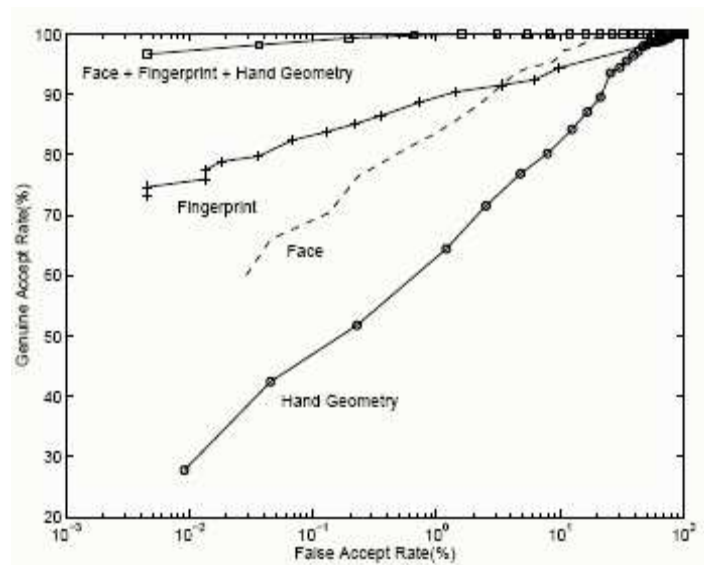


Figure 5: ROC Curve for a system utilizing multiple biometric traits. [4]

2.3.1 Fusion in Biometrics

Information fusion is the key component in multimodal biometrics [1]. Information fusion in multimodal biometrics is the integration of data pertaining to multiple independent biometric devices. This can take place at various levels: feature extraction level, matching score level, decision level, sensor level, and rank level. Figure 6 shows three types of fusion methods.

Fusion in multi-modal biometrics systems can be implemented in five ways:

- Multiple sensors may be used to capture the same biometric;
- Multiple biometrics may be captured;
- Multiple readings of the same biometric may be combined to achieve an optimal reading;
- Readings of two or more units of the same biometric may be taken (*e.g.*, two different fingerprints or both irises) or
- Different matching and/or feature extraction algorithms may be used on the same biometric reading to give separate results.

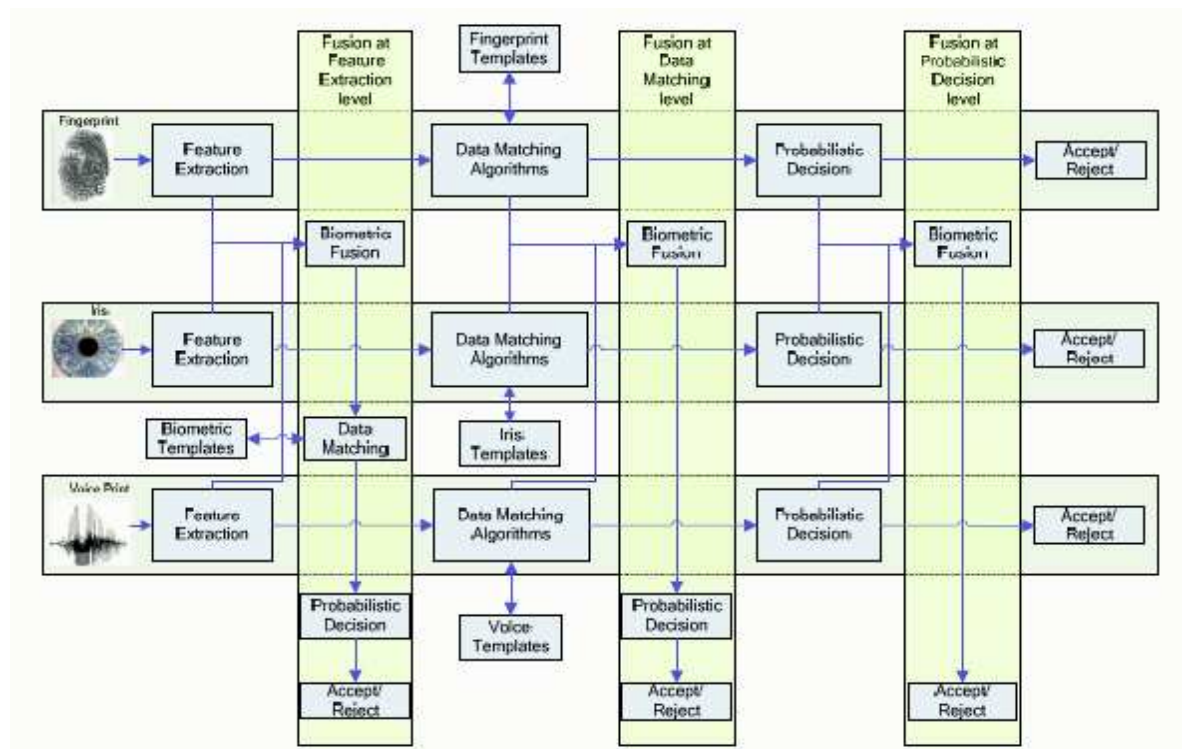


Figure 6: Biometric fusion options in the verification mode of operation. [15]

Fusion of uncorrelated modalities (face and fingerprint) is expected to give better performance than the fusion of correlated modalities (multiple matching algorithms) [6]. Fusion can take place at various levels:

Feature Extraction level: The raw data captured from each sensor is used to build a feature vector, which uniquely identifies a given person in the feature space. Combining more feature vectors results in one vector with higher dimensionality and may increase the probability of correctly identifying a person. However, fusion at this level is difficult due to the following listed problems: curse of dimensionality, removing features that are highly correlated between different biometric modalities, inaccessible feature vectors for most of the biometric systems.

Matching score level: Fusion at the match score level is much more effective than fusion at the decision level. Each unimodal biometric system measures and calculates its own matching score. Matching score is a measure of the similarity between features derived from a presented sample and a stored template. A match / non-match decision is made based on a certain decision threshold.

There are two approaches for consolidating the scores obtained from different matchers. One approach is to pose it as a classification problem where for each biometric modality a feature vector is constructed using the matching scores. A trained classifier will decide one of two classes: "Accept" (genuine user) or "Reject" (imposter user), based on the feature vector. The second approach is a combination problem where the individual matching scores are combined to generate a single scalar score, which is then used to make the final decision (see Figure 6). Since the matching scores are heterogeneous, normalization is required to transform these scores into a common domain.

Decision level: Fusion on this level is the least informative, each biometric system makes a decision and then those decisions are combined usually based on majority voting scheme. Methods to weight the decisions from each biometric system are also used, for example iris over hand geometry.

2.3.2 Combining the modalities at matching score level - Normalization methods

Each of the biometric modalities will output a matching score. Currently, there is not a standard as to what this number represents. It could represent a similarity or a dissimilarity (distance) score. The matching scores at the output of the matcher may be distinguished as genuine or imposter scores. The sets of the genuine and imposter scores may follow different statistical distributions. There is no guarantee that the score belongs to specific probability distribution. The scores could also be on different numerical scales. Various scores need to be converted into a common domain so that they can be combined. It is highly desirable to use a score normalization technique that has high robustness and is efficient. Robustness refers to insensitivity to outliers in the data and efficiency refers to the proximity of the obtained estimate to the optimal estimate when the distribution of the data is known.

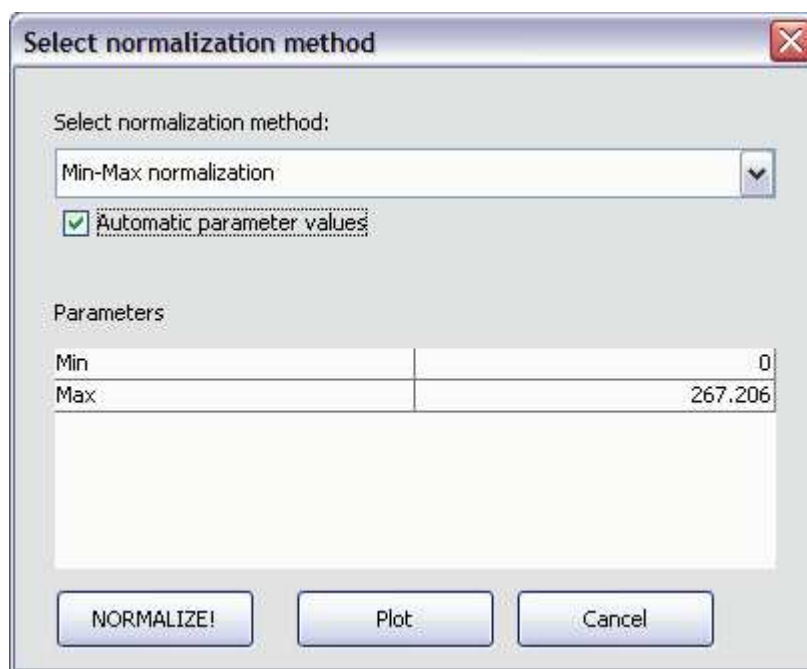


Figure 7: Normalization methods window in the MUBI tool.

The following is a list of normalization techniques that are available and implemented in the MUBI. If the matching score set is $\{s_k\}$, normalized scores are denoted as $\{s'_k\}$:

- Min max normalization is best suited for the case where the bounds (maximum and minimum values) of the scores produced by the matcher are known. This method is not robust; therefore, it is highly sensitive to outliers. [5]

$$s'_k = \frac{(s_k - \min)}{(\max - \min)}$$

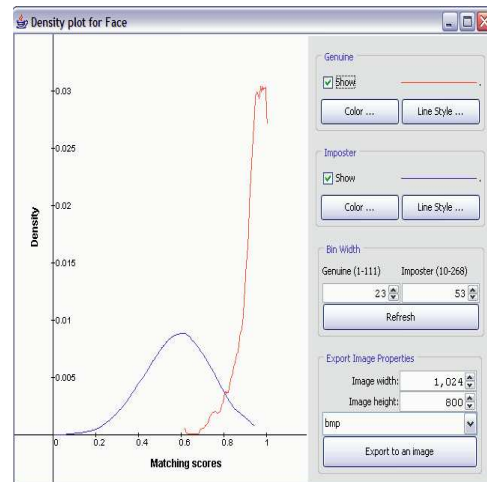


Figure 8: Face normalized with Min/Max – plot from MUBI tool.

- Decimal scaling can be applied when the scores of different matchers are on a logarithmic scale (lack of robustness and assumption that the scores of different matchers vary by a logarithmic factor) [5]

$$s'_k = \frac{s_k}{10^n}, n = \log_{10} \max(s_i)$$

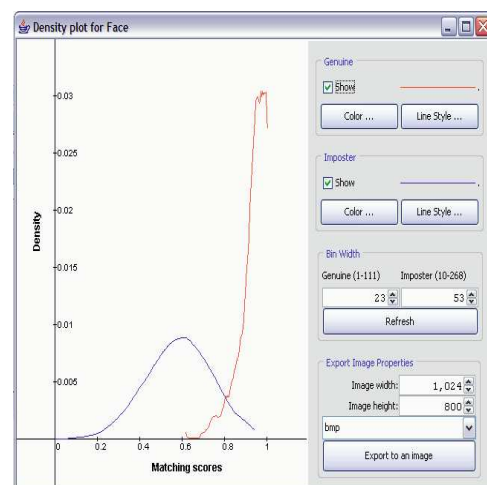


Figure 9: Face normalized with Decimal Scaling – plot from MUBI tool.

- Z-score normalization – is calculated using the arithmetic mean and standard deviation of the given data. Sensitive to outliers and does not guarantee a common numerical range for the normalized scores from different matchers. [5]

$$s'_k = \frac{s_k - \mu}{\sigma}$$

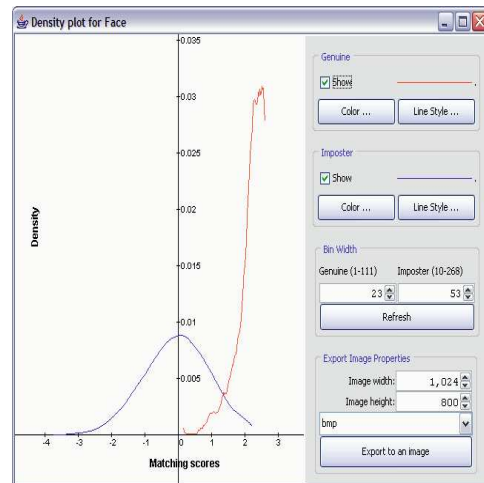
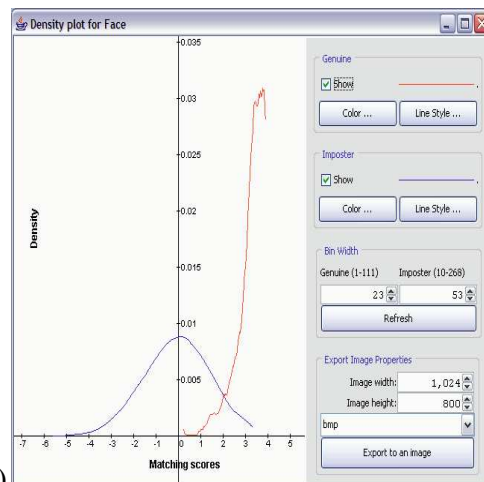


Figure 10: Face normalized with Z-score – plot from MUBI tool.

- Median and median absolute deviation (MAD) method is insensitive to outliers and the points in the extreme tails of the distribution, but has low efficiency compared with Z-scores. [5]

$$s'_k = \frac{s_k - \text{median}}{MAD}$$



$$MAD = \text{median}(|s_k - \text{median}|)$$

Figure 11: Face normalized with Median and MAD – plot from MUBI tool.

- TanH estimators are robust and highly efficient. They use the Hampel estimators, based on the influence ψ - function below, to reduce the influence of outliers in the distribution. The mean and standard deviation estimates of genuine scores are represented with μ_{GH} and σ_{GH} , respectively[5]

$$s'_k = \frac{1}{2} \left\{ \tanh\left(0.01 \frac{s_k - \mu_{GH}}{\sigma_{GH}}\right) + 1 \right\}$$

$$\psi(u) = \begin{cases} u, & 0 \leq |u| < a, \\ a * \text{sign}(u), & a \leq |u| < b, \\ a * \text{sign}(u) * \left(\frac{c - |u|}{c - b}\right), & b \leq |u| < c, \\ 0, & |u| \geq c. \end{cases}$$

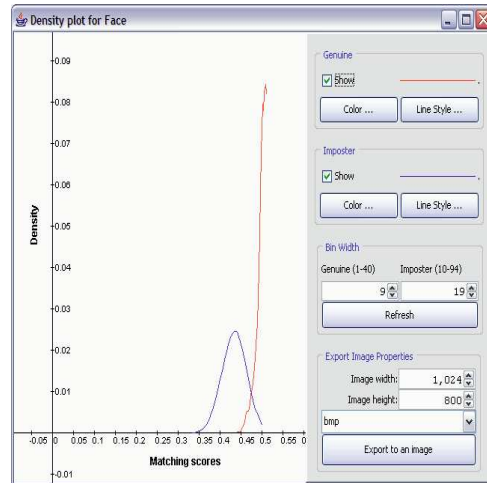


Figure 12: Face normalized with Tanh normalization– plot from MUBI tool.

2.3.3 Combining the modalities at matching score level – Fusion methods

A list of fusion techniques that can be used to combine multiple normalized scores into a single score are provided in this section.

If s_i is the matching score from i^{th} modality, s represents the resulting fused score.

- The *Simple Product Rule* combines the scores by multiplying all of the individual scores.

$$s = s_1 * s_2 * \dots * s_n$$

- The *Simple Sum Rule* combines the scores as a linear transformation.

$$s = (a_1 s_1 - b_1) + \dots + (a_n s_n - b_n)$$

a_i and b_i represents the weights and biases, respectively, which can be entered by the user.

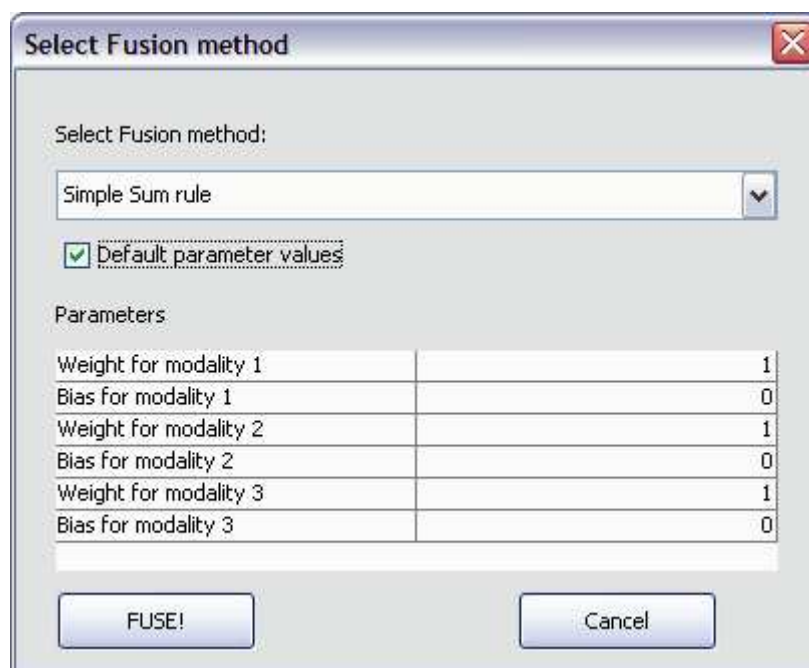


Figure 13: Fusion methods implemented in the MUBI tool.

- The *Simple Max Rule* is the maximum score from the different modalities.

$$s = \max(s_1, s_2, \dots, s_n)$$

- The *Simple Min Rule* is the minimum score from the different modalities.

$$s = \min(s_1, s_2, \dots, s_n)$$

- *Biometric Gain against Imposter (BGI)* is the likelihood ratio of genuine to imposter scores.

$$\text{BGI} = \frac{\text{Probability of being an imposter, given the biometric evidence too}}{\text{Probability of being an imposter, given only prior knowledge}}$$

A very good approximation of BGI is Likelihood Ratio of Genuine to Imposter (LRGI)

$$\text{BGI} \sim \text{LRGI} = \frac{\text{Probability of seeing the evidence from an imposter}}{\text{Probability of seeing it from expected genuine subject}}$$

Each score is normalized by transforming it into the LRGI scale, which represents a normalization process. Then the various scores are combined by multiplication or by addition of the log likelihood ratios [20].

Information fusion at the matching score level is advantageous as it allows the possibility to adjust user specific parameters. For example, to compensate for a genuine user who has dry skin, a user specific weighting scheme can be implemented so that more weight will be placed on the user's other biometric traits. It has been shown [2] that the use of user-specific weights and thresholds can improve the performance by approximately 3% and 2%, respectively.

2.2.3 Data Partitioning

When testing the performance of the selected combination of normalization and fusion method, the input data can be partitioned. In this way, we can test the chosen combination of normalization and fusion methods on a new data that has not been used while choosing the methods. One way to do this is not to use the entire dataset during the training phase. Some of the data is removed before training begins. Then when the training is over, the data that was removed can be used to test the performance of the normalization and fusion methods on the “new” data. This is the basic idea for a whole class of model evaluation methods called *cross validation*.

The datasets can be divided into training and test set in three different ways in MUBI:

- a) re-substitution method where all the available data is used for training as well as testing, training and test sets are the same (special case of hold-out method)
- b) hold-out method – input value for the percentage can be entered and data will be divided into independent training and test sets according to the specified percentages.

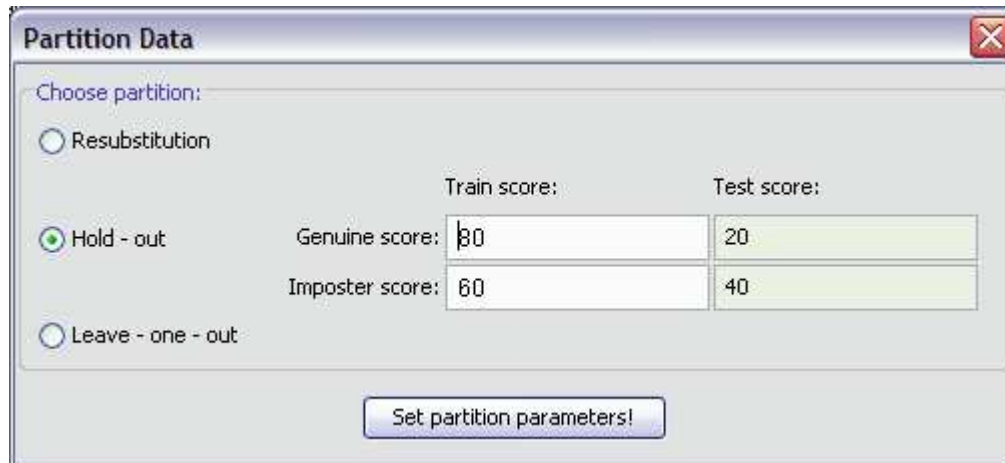


Figure 14: Partition Data methods in the MUBI tool.

The holdout method is the simplest kind of cross validation. The dataset is separated into two sets, called the training set and the testing set. Normalization parameters are estimated using only the training set. Then the test set representing the part of dataset that has never been seen before in the system is normalized using these parameters and the fusion method is executed. The advantage of this method is that it takes less time to perform fusion on a large dataset. However, its evaluation can have a high variance. The evaluation depends heavily on the selected data points in the training set and the test set, and thus the evaluation may be significantly different depending on the partitioning the dataset.

- c) leave one out method - In this method the dataset is divided to $n-1$ different training samples and 1 test sample, N different times. This means that N separate times, normalization parameters are estimated from all data points except one score and then normalization and fusion are performed to test scores from each modality.

By using the training set we can compare parameters for normalization and then test the performance of the fusion method on the test set.

Chapter 3 Analysis of the score distributions

Chapter 3

Analysis of the score distributions

3.1 Goodness of Fit Statistics

By analyzing the match scores, conclusions about the performance of the biometric system can be derived. The distributions of the scores are very important for describing the data. One way to examine the data is to fit a parametric probability distribution. There are statistics, which measure how good the distribution fits the input data and how confident we can be that the data was produced by the distribution function. For each of these statistics the smaller the value, the better the fit. We used two different fit statistics: Chi-squared and Kolmogorov-Smirnov. When more than one fit statistic is available, there is no hard rule to decide which test will give you the "best" result. Each test has its own strengths and weaknesses. The decision is up to the user and the user must decide which information is more important when considering which test to use.

3.1.1 Maximum Likelihood Estimation

The Maximum likelihood Estimates (MLE) of a distribution are the parameters of the distribution function that maximize the probability of obtaining the given dataset.

This method is robust and generates estimators with good statistical properties. The implementation can be mathematically intense.

For any density distribution $f(x)$ with one parameter α , and a corresponding set of n sampled values X_i , an expression called the likelihood is defined as:

$$L = \prod_{i=1}^n f(X_i, \alpha)$$

To find the MLE, simply maximize L with respect to α :

$$\frac{\partial L}{\partial \alpha} = 0,$$

and solve for α . The method described above can be easily generalized to distributions with more than one parameter.

For example, an exponential function with a fixed lower bound of zero has only one adjustable parameter, and its MLE is easily calculated. The distribution's density function is:

$$f(x, \beta) = \frac{1}{\beta} e^{-x/\beta}$$

and the likelihood function is:

$$L(\beta) = \prod_{i=1}^n \frac{1}{\beta} e^{-x_i/\beta} = \beta^{-n} \exp\left(-\frac{1}{\beta} \sum_{i=1}^n X_i\right)$$

To simplify, we can use the natural log of the likelihood function:

$$l(\beta) = \ln L(\beta) = -n \ln(\beta) - \frac{1}{\beta} \sum_{i=1}^n X_i$$

To maximize the log of the likelihood, simply set its derivative with respect to β to zero:

$$\frac{\partial l}{\partial \beta} = -\frac{n}{\beta} + \frac{1}{\beta^2} \sum_{i=1}^n X_i,$$

$$\beta = \sum_{i=1}^n \frac{X_i}{n}$$

Therefore, MUBI tries to fit the data to the best Exponential function with a fixed lower bound of zero. It first finds the mean of the input data and uses it as the MLE for β .

3.1.2 Chi-Squared statistics

The chi-squared statistic is the best-known goodness-of-fit statistic. It can be used with both continuous and discrete sample data. To calculate the chi-squared statistic first the x-axis domain must be broken into several “bins”. The chi-squared statistic is defined as:

$$\chi^2 = \sum_{i=1}^K \frac{(N_i - E_i)^2}{E_i}$$

where:

K is the number of bins;

N_i is the observed number of samples in the i^{th} bin;

E_i is the expected number of samples in the i^{th} bin.

The chi-squared statistic is calculated directly from the difference between the observed and expected values and a characteristic called the degrees of freedom. Whether the observed and expected values are similar enough to be able to claim no association is measured by the P-value, which is calculated from the chi-squared statistic and its degrees of freedom. A commonly adopted convention is to reject the hypothesis of no association between the two variables if the P-value is less than 0.05 ($P < 0.05$).

The chi-squared statistic has a weakness, as there are no clear rules regarding the selection criteria for the number and location of the bins. Different conclusions for the same data can be derived depending on the specified bins. Some unpredictability can be removed by choosing equiprobable bins.

3.1.3 Kolmogorov–Smirnov statistic

This test is implemented in MUBI and is used to decide if the match score dataset comes from a population with a specific distribution.

Kolmogorv-Smirnov statistic is described by:

$$D_n = \max \left[\left| F_n(x) - \hat{F}(x) \right| \right]$$

where

n = total number of data points

$\hat{F}(x)$ = the fitted cumulative distribution function

$$F_n(x) = \frac{N_x}{n}$$

N_x = the number of X_i 's less than x .

Since the Kolmogorov-Smirnov statistic does not require binning, it is less arbitrary than the chi-squared statistic. A weakness of the Kolmogorov-Smirnov statistic is its inability to detect tail discrepancies.

3.1.4 P-values and Critical values

The goodness-of-fit statistic reports a measure of the deviation of the fitted distribution from the input data. As mentioned earlier, the smaller the fit statistic is the better is the fit. However, how small a value is needed for a “good” fit? This section explains how P-values and critical values can be used to analyze the “goodness” of a fit.

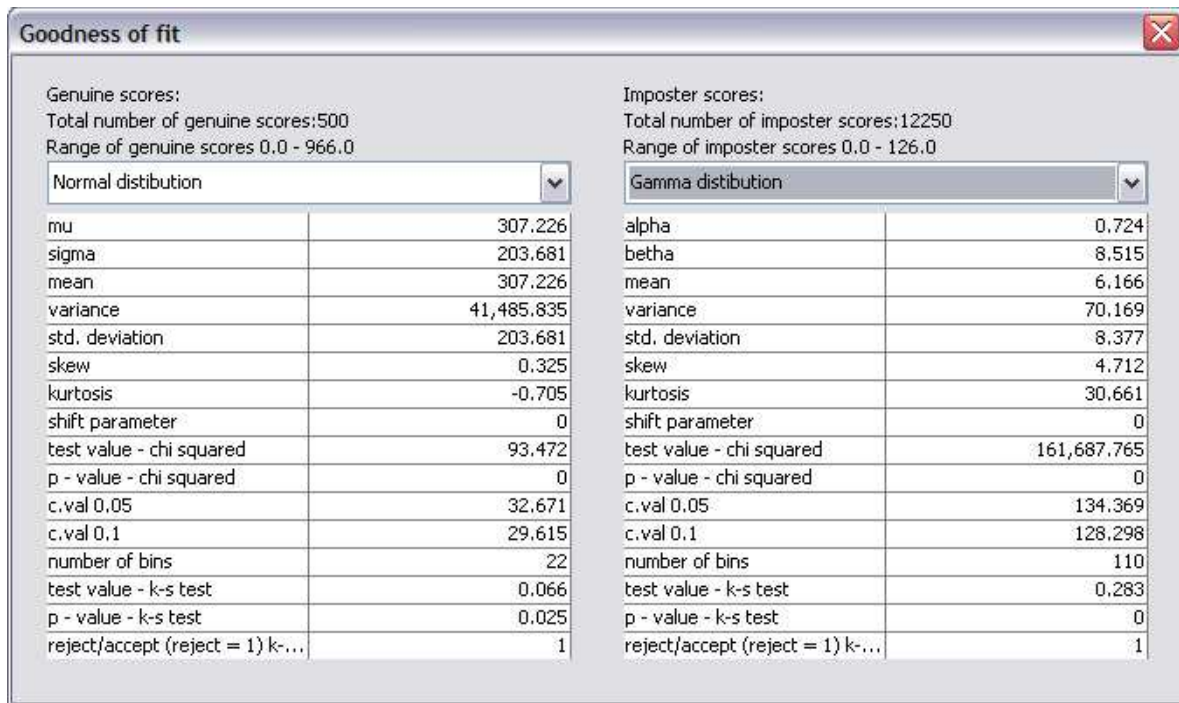


Figure 15: Goodness of fit tests in the MUBI tool.

Suppose we have a distribution fitted to a set of N sampled values, and a corresponding fit statistic - t . How likely is it that a new set N samples drawn from the fitted distribution would generate a fit statistic greater than or equal to t ? This probability is referred to as the P -value. As the P -value decreases to zero, we are less and less confident that the fitted distribution could possibly have generated our original data set. On the other hand, as the P -value approaches one, we have no basis to reject the hypothesis that the fitted distribution actually generated our data set.

Often a particular level of significance needs to be specified, usually denoted by α . The value t of the fit statistic at α level of significance is known as the “critical value”. This value of α is the probability that we will incorrectly reject a distribution because it generated, due to statistical fluctuations, a value of t that was very large. Then we need to know, given this significance level, what is the largest value of t that would be accepted as a valid fit. Any fit that has a value of t above the critical value is rejected, while fits with values of t below the critical value are accepted. Typically, critical values depend on the type of distribution fit, the particular fit statistic being used, the number of data points, and the significance level.

Skewness and Kurtosis are measures of the shape of the distribution. Skewness characterizes the degree of asymmetry of a distribution around its mean. A positive value of skewness signifies a distribution with asymmetric tail extending out towards more positive x . A negative value signifies a distribution whose tail extends out towards more negative x . If the skew value is then it suggests that the distribution may be symmetric.

$$Skew(x_1, \dots, x_n) = \frac{1}{N} \sum_{j=1}^N \left(\frac{x_j - \mu}{\sigma} \right)^3$$

where:

(x_1, \dots, x_N) represents the scores;

N – is the total number of scores;

μ -is the mean of the scores;

σ - is the standard deviation of the scores.

Kurtosis measures the relative peakness or flatness of a distribution (relative to normal distribution). If the value of the kurtosis is positive then the distribution is leptokurtic (leptokurtic distribution has a more sharp "peak" around the mean- it means higher probability than a normally distributed scores near the mean) and negative value suggests platykurtic (a platykurtic distribution has a smaller "peak" around the mean -it means a lower probability than a normally distributed scores near the mean).

$$Kurt(x_1, \dots, x_n) = \frac{1}{N} \sum_{j=1}^N \left(\frac{x_j - \mu}{\sigma} \right)^4 - 3$$

where:

(x_1, \dots, x_N) represents the scores

N – is the total number of scores

μ -is the mean of the scores

σ - is the standard deviation of the scores

The shift parameter is calculated for each distribution. It indicates any x-value that is applied if the input data exceeds the range of the fitted distribution. A shifted distribution is expressed in the form $\text{Distribution}(\text{arg}_1, \text{arg}_2 \dots \text{arg}_n) \pm \text{Shift}$, where a sample drawn from the shifted distribution will have the shift amount added to or subtracted from it.

Chapter 4 Implementation

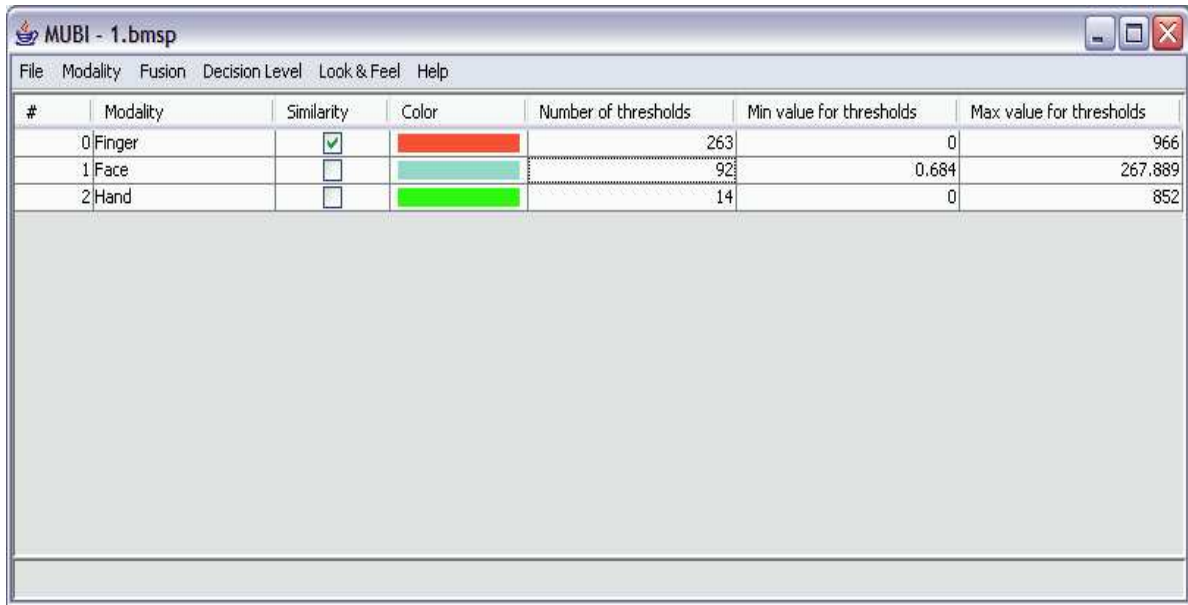
Chapter 4

Implementation

4.1 MUBI offline analysis tool

With the MUBI off-line analysis tool we can empower biometric system designers to chose “*the best*” integration techniques in the context of their application. It allows the evaluation of different normalization and fusion methods as multimodal biometric system design alternatives. The tool is implemented in Java 5 and it is portable to any operating system containing JAVA installation. For the goodness of fit evaluations, a Matlab installation is also required on the operating system.

The MUBI is an application for analyzing biometric systems. Only one biometric system can be analyzed at a time. Each system contains a number of biometric modalities. The inputs for each modality are the genuine and the imposter scores. The tool combines similarity as well as distance scores. When a modality to the project (biometric system), needs to be added, two text files or comma-delimited files containing the genuine and imposter scores are required.



The screenshot shows a window titled "MUBI - 1.bmsp" with a menu bar containing "File", "Modality", "Fusion", "Decision Level", "Look & Feel", and "Help". Below the menu bar is a table with the following data:




#	Modality	Similarity	Color	Number of thresholds	Min value for thresholds	Max value for thresholds
0	Finger	<input checked="" type="checkbox"/>		263	0	966
1	Face	<input type="checkbox"/>		92	0.684	267.889
2	Hand	<input type="checkbox"/>		14	0	852

Figure 16: MUBI tool – MSU Dataset.

After modalities are added to the project (Figure 17), the project, including information about the modalities can be saved into a single binary file. For each modality, we can plot the estimated density curves for its genuine and imposter scores, create partitioning with some of the methods, normalize and then fuse all the modalities with one of the fusion methods. A ROC curve can be plotted to see the performance of the selected combination and it can be exported into jpg or bmp image file (Figure 18). When moving the mouse above the graph the FAR and GAR for that specific point on the graph are indicated. In addition, all normalized and fused scores can be exported into text files for performing extra analysis.

Testing different strategy rules can be developed for a specific database. The distribution fit method helps in guessing the underlying distribution. Partitioning data aids in evaluating the performance of the specific model on the test set which is completely separate from the training set. Results of this tool can be used to identify the combinations of biometric traits that give significantly improved performance.



Figure 17: Add modality in the system from MUBI tool.

The performance of the multimodal biometric system with MUBI can be studied with fusion techniques like: the simple sum of scores, the simple max score, the simple min-score, simple product score and BGI fusion methods can be applied on the normalized scores. The scores are transformed into similarity scores with simple distance-to-similarity transformation and with no scale change. Then one of the following normalization techniques can be used: simple Min-Max, Z-score, Median-MAD, and Tanh. For BGI fusion normalization methods are not required since the method itself normalizes the scores.

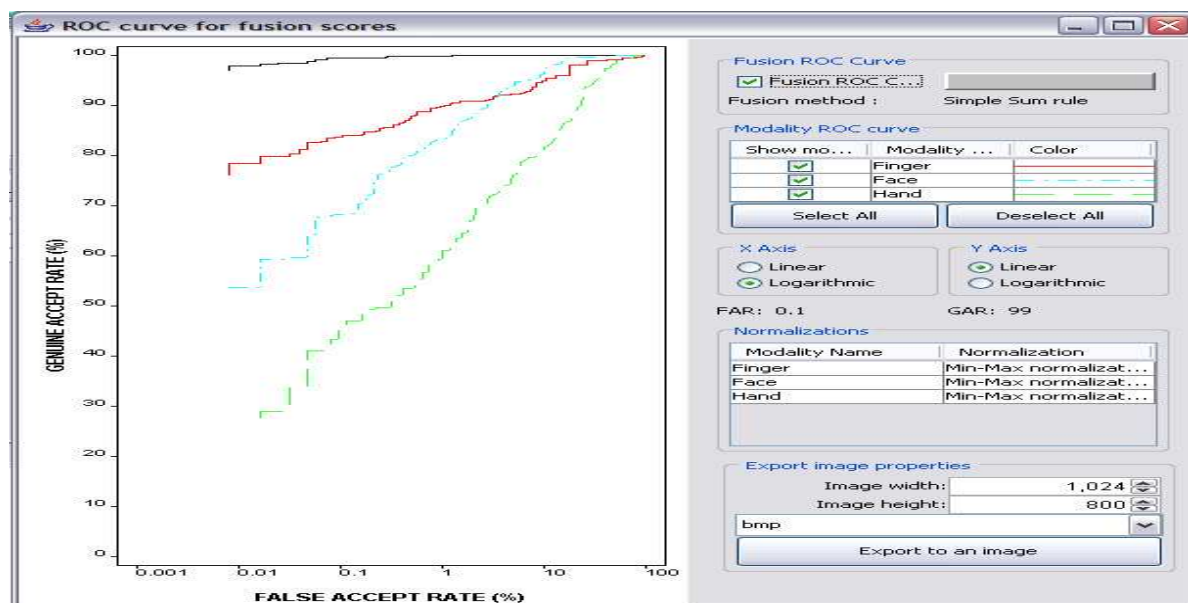


Figure 18: ROC curve plot for one biometric system from MUBI tool.

Chapter 5 Experiments and results

Chapter 5

Experiments and results

5.1. MSU dataset

The MSU multimodal database used in our experiments contains genuine scores and the other with 12,250 imposter scores. The data was collected from 50 users using different sensors, over different time periods. Five face images and five fingerprints of the same finger were obtained from each user. Five hand-geometry images were taken from 50 different users (some of the users were present in both sets). For the biometric traits, it is assumed mutual independence. This assumption permit to randomly pair users from the two sets. Thus, a database of 50 users was constructed, each user having five biometric templates for each modality. Fingerprint matching using minutiae features was used and the scores are similarity scores. For face, eigenface coefficients were used as a representation of the face image. Euclidean distance was calculated between the eigenface coefficients of the template and that of the input face image. For the hand geometry, also Euclidean distance was used to calculate the matching score between the 14-dimensional feature vectors representing hand images [5].

5.1.1 Experiments with normalization and fusion methods on MSU dataset

First, the distributions of the score density plots of genuine and imposter scores for each modality are presented below. Analyzing the shape of the genuine and imposter scores, one can see overlapping region between the two distributions. Accordingly, we can predict the impact of that modality on the performance of the system. In Figure 19, the genuine and imposter density functions of the fingerprint system are plotted using MUBI. We can observe that the imposter scores are very close to zero, and the genuine scores are spread over a wide range of values [5]. Density plots of the face genuine and imposter scores are displayed in Figure 20 and the density plots of hand genuine and imposter scores in Figure 21.

The density plots of genuine and imposter scores indicate that the overlapping area for finger is the smallest and for hand geometry is the largest.

Before discussing the fusion methods, it should be noted that if there are modalities with different kind of scores, for example, similarity scores from finger and distance scores from face and hand, MUBI transforms all the scores to similarity scores before performing normalization and fusion. Thus the raw scores from all three modalities are made comparable. The scores produced in these experiments are unbounded and can, in theory, produce any value.

Figure 22 displays the ROC curves generated by utilizing different normalization methods followed by the Simple Sum Rule. In addition, the ROC curves of each modality are shown for comparison purposes. Tanh normalization followed by the Simple Sum Rule gave the best ROC curve for these three modalities from the MSU dataset.

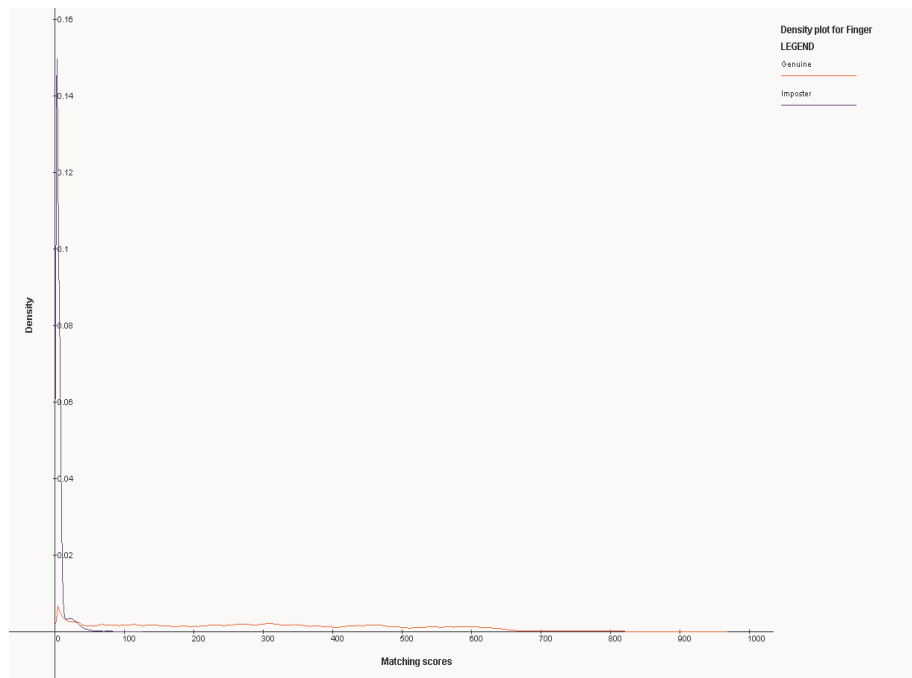


Figure 19: Density plot of finger genuine and imposter scores.

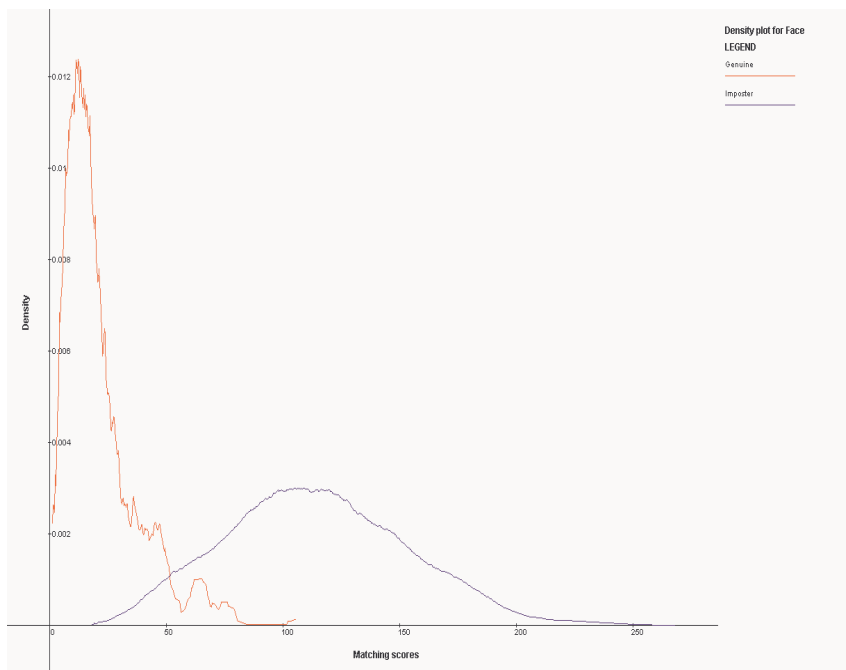


Figure 20: Density plot of face genuine and imposter scores.

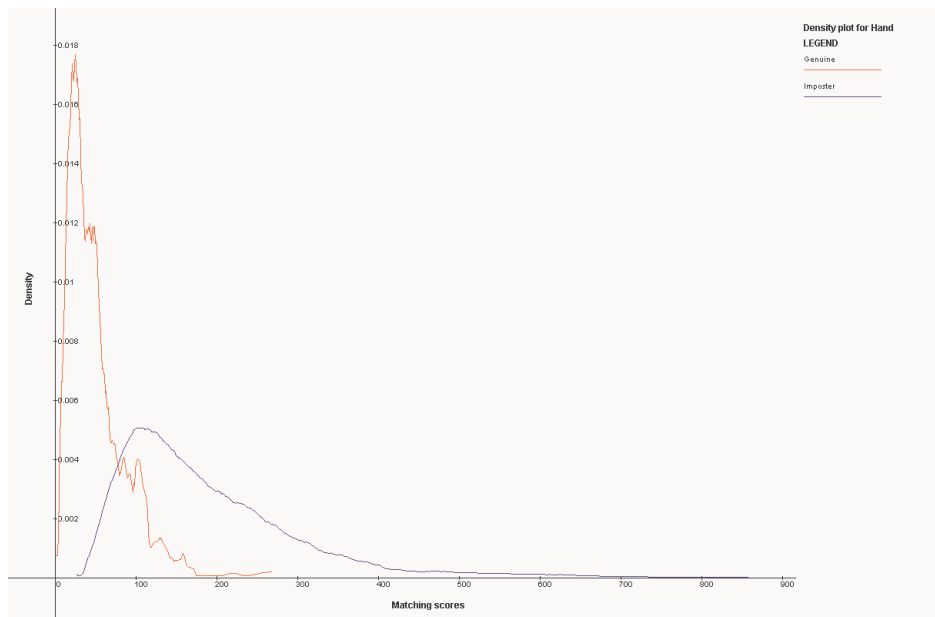


Figure 21: Density plot of hand genuine and imposter scores.

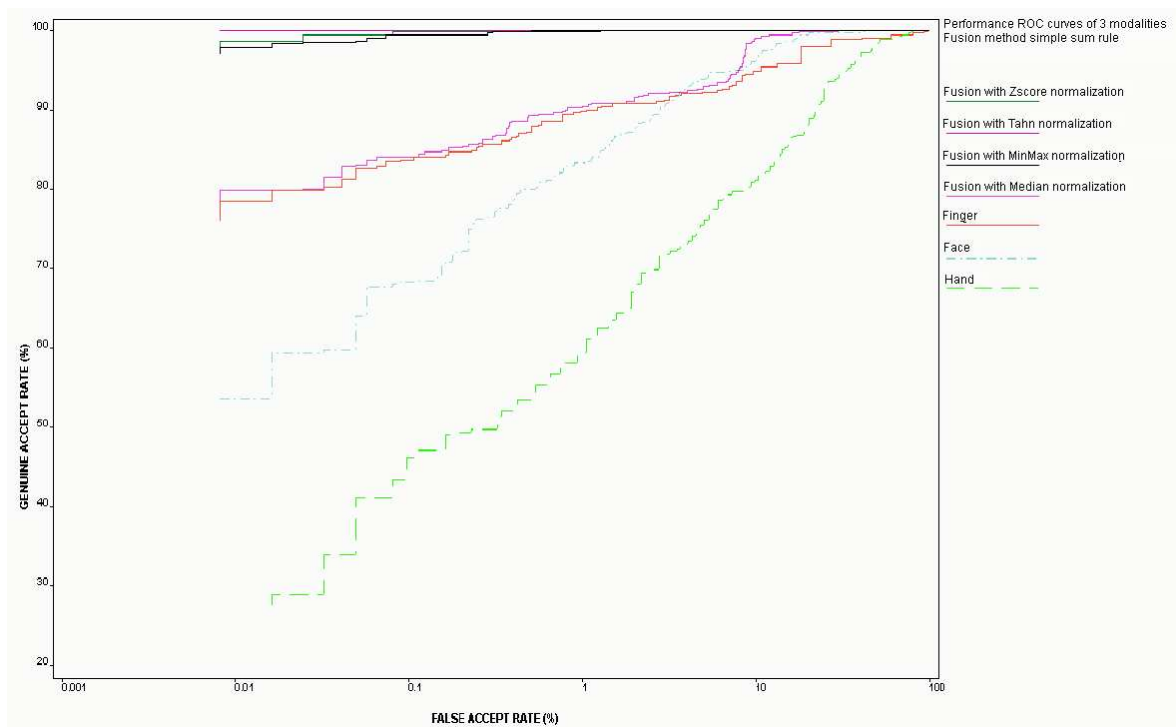


Figure 22: Different normalization methods followed by the Simple Sum Rule on three modalities from MSU dataset.

Figure 23 shows the ROC curves of the three modalities from the MSU dataset. Scores were normalized with the Min-Max normalization method and then combined with the Simple Sum Rule, Simple Product Rule, Minimum Rule, Maximum Rule and BGI fusion method. The Simple Sum Rule after Min-Max normalization gives the best performance in this experiment.

For this dataset, it can be observed from the experiments that Min-Max, Z-Score and Tanh normalization methods outperform the other normalization methods. It should be noted that if the scores from the modalities are significantly different there is no guarantee that these methods will work.

Figure 24 shows the overall winner for the MSU dataset which is the Tanh normalization method followed by the Simple Sum Rule fusion.

Table 1 summarizes the Genuine Acceptance Rate (GAR) of the multimodal system for different normalization techniques followed by the Simple Sum Rule fusion method and BGI fusion at a False Acceptance Rate (FAR) of 0.1%. For the BGI method half of the scores are used for density estimation based on the Parzen Window method [24]. The user can specify the kernel function and the window size. It should be noted that the training set is randomly chosen and hence the results even for the same fusion rule may not be similar. In Table 1 the average GAR value is presented for five BGI fusions.

Normalization techniques	Fusion techniques	
	Simple Sum Rule	BGI (average value)
Min-Max	99.548	98.3726
Z-Score	99.699	
Tanh	99.825	
Median	84.036	
Decimal	99.328	

Table 1: Genuine Acceptance Rate (GAR) (%) of different normalization techniques followed by the Simple Sum Rule fusion method and the BGI method at 0.1% False Acceptance Rate (FAR).

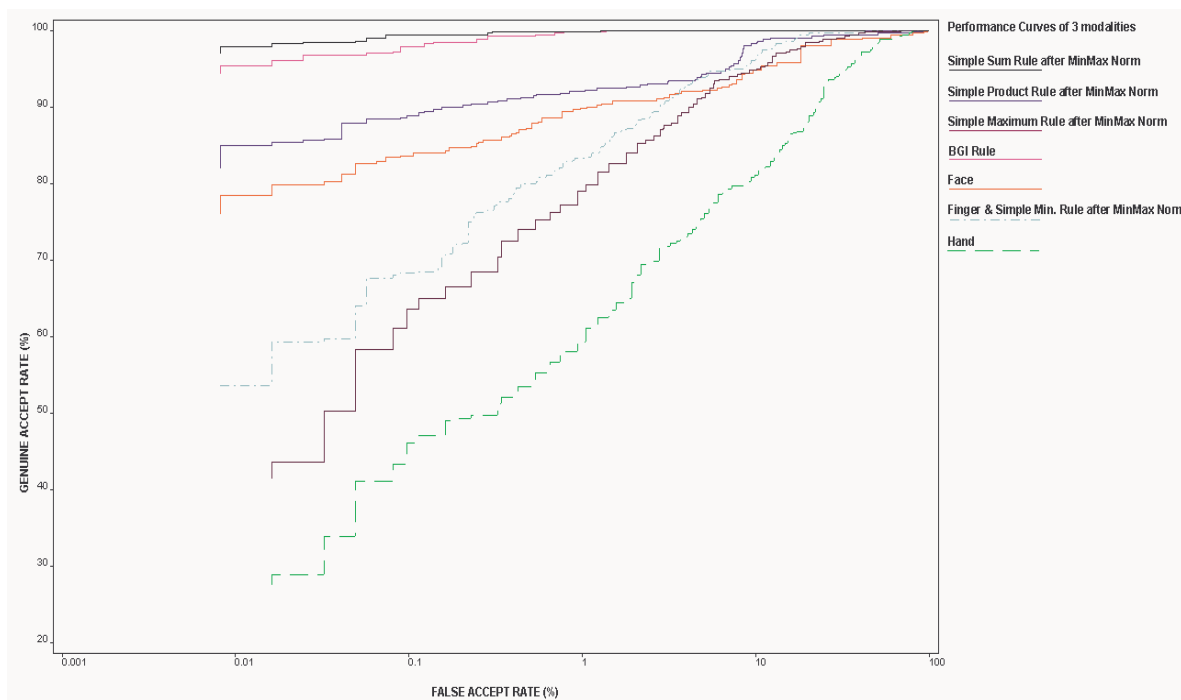


Figure 23: Different fusion methods after Min-Max normalization on three modalities from MSU dataset.

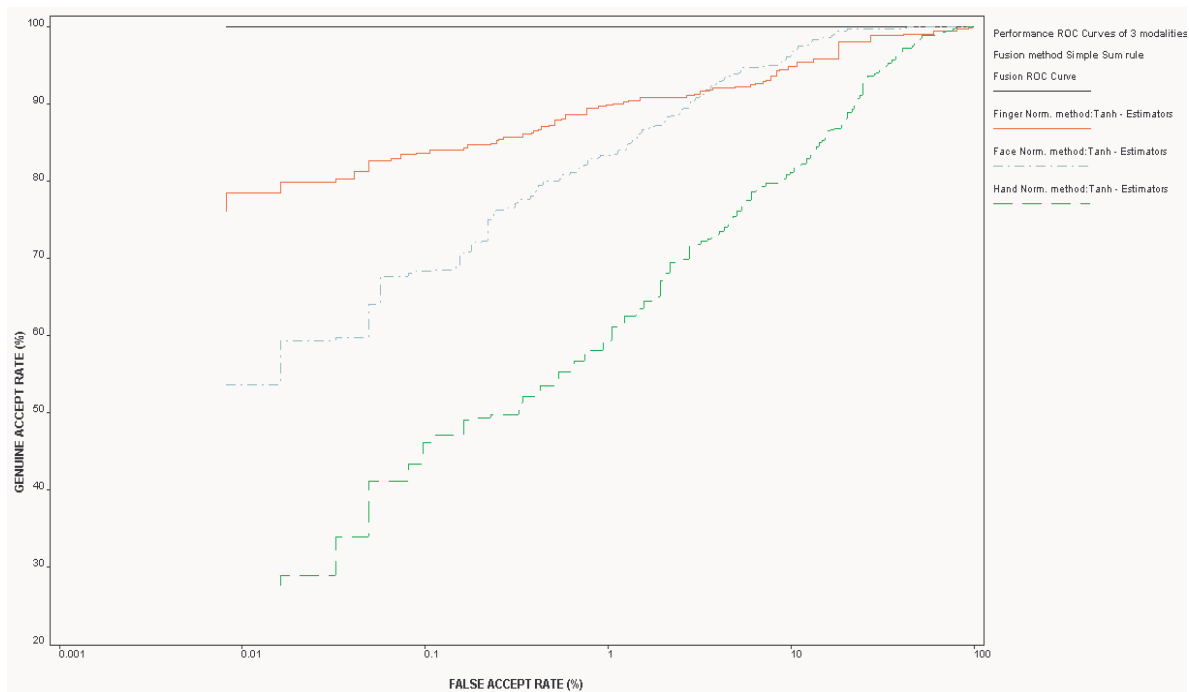


Figure 24: Overall winner for the MSU dataset – Tanh normalization with Simple Sum Rule fusion method.

5.2 NIST dataset

Biometric Scores Set - Release 1 (BSSR1) is a set of raw output similarity scores from two face recognition systems and one fingerprint system, operating on frontal faces, and left and right index live-scan fingerprints, respectively.

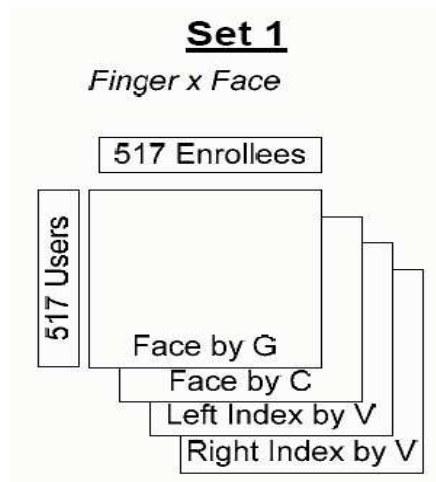


Figure 25: NIST (BSSR1) dataset [8].

The release includes true multimodal score data, i.e. similarity scores from the comparisons of faces and fingerprints of the same people. The dataset is comprised of face and fingerprint scores from the same set of 517 individuals. For each individual, the set contains one score from the comparison of two right index fingerprints, one score from the comparison of two left index fingerprints, and two scores (from two separate matchers) from the comparison of two frontal faces. These scores were computed from the fingerprint images and the face images of the same person collected at the same time. NIST database consists of four modalities Face C Matcher, Face G Matcher, Left Index Finger and Right Index Finger each with 517 genuine scores and 266,772 imposter scores.

5.2.1 Experiments with normalization and fusion methods on NIST dataset

For analyzing the genuine and imposter distributions of the scores, density plots of these scores for the finger left index are displayed in Figure 26. Figure 27 presents density plots of the genuine and imposter scores for the right finger index. Density plots for the genuine and imposter distributions of the face C matcher and face G matcher are displayed in Figure 28 and Figure 29, respectively. The overlap area is the smallest for the right finger index compared to the other modalities. Accordingly, the ROC curve for the right finger index in Figure 30 is the best among the four modalities. Figure 30 shows the performance for each biometric from the NIST dataset. Scores are fused with the Simple Sum Rule using Min-Max normalization method on system with two modalities, three modalities and four modalities. It is obvious that the ROC curve when two modalities are fused is much better than the ROC curve of each individual biometric trait. In addition, it can be observed that adding more modalities to the system improves the performance of the multimodal biometric system (for this dataset).

Figure 31 shows ROC curves from the experiments with different fusion methods. BGI and the combination with MinMax Normalization and Simple Sum Rule fusion perform the best from all combinations.

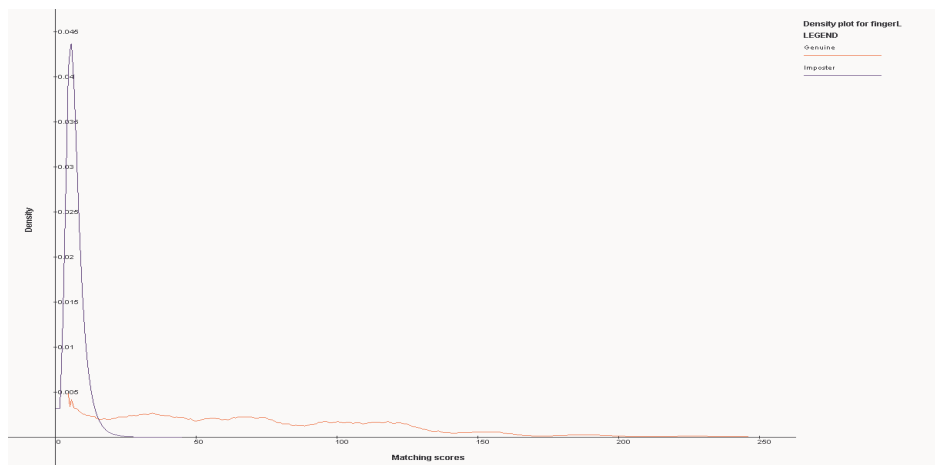


Figure 26: Density plot of the genuine and imposter scores of the finger left index.

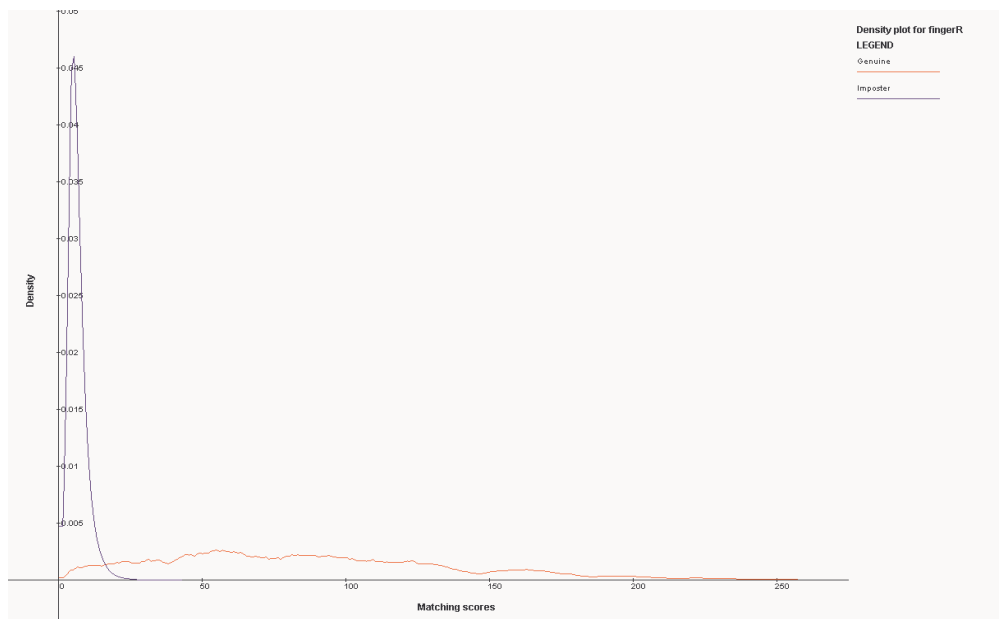


Figure 27: Density plot of genuine and imposter scores of the finger right index.

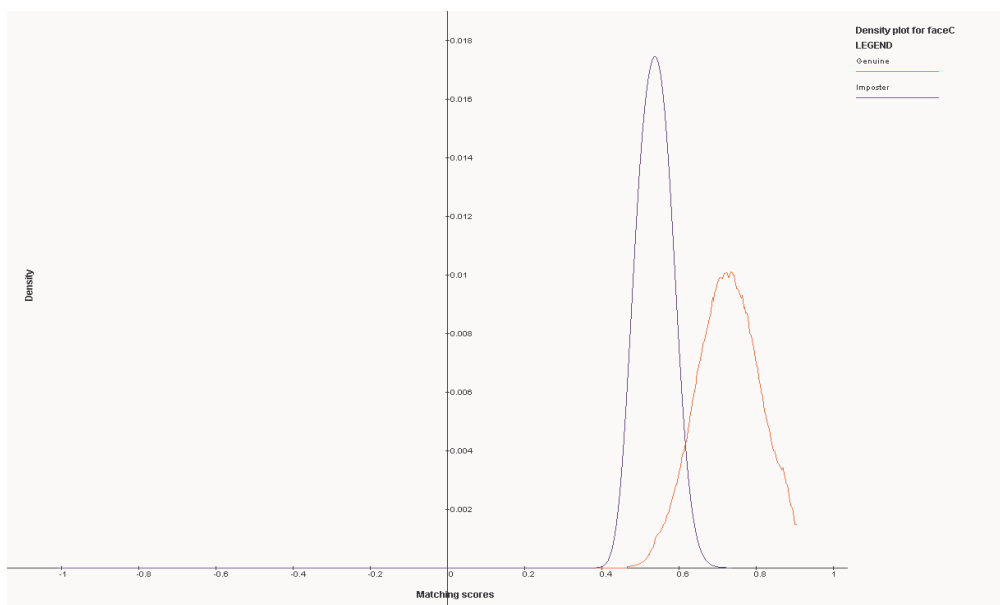


Figure 28: Density plot of the genuine and imposter scores of the face C matcher.

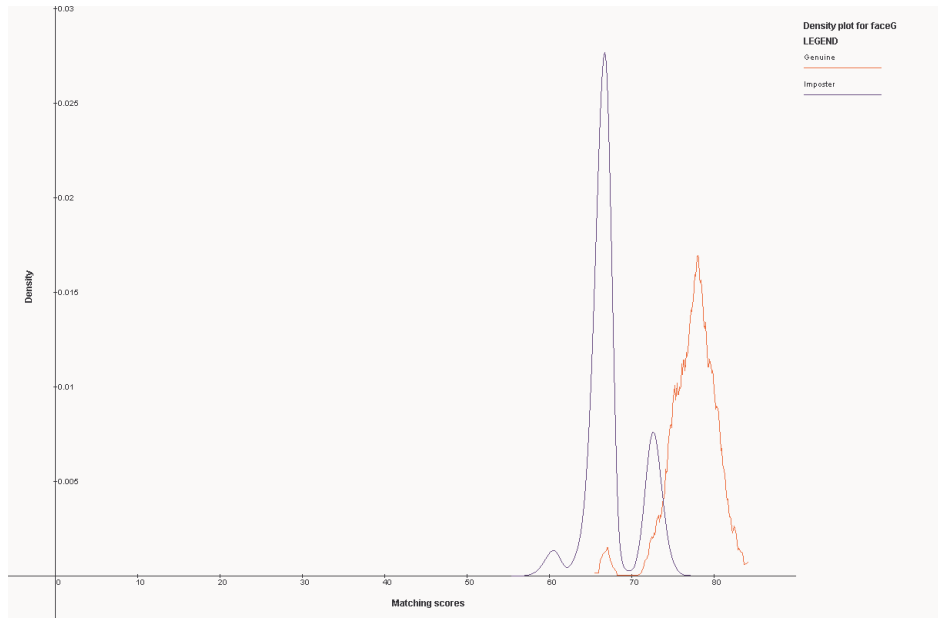


Figure 29: Density plot of the genuine and imposter scores of the face G matcher.

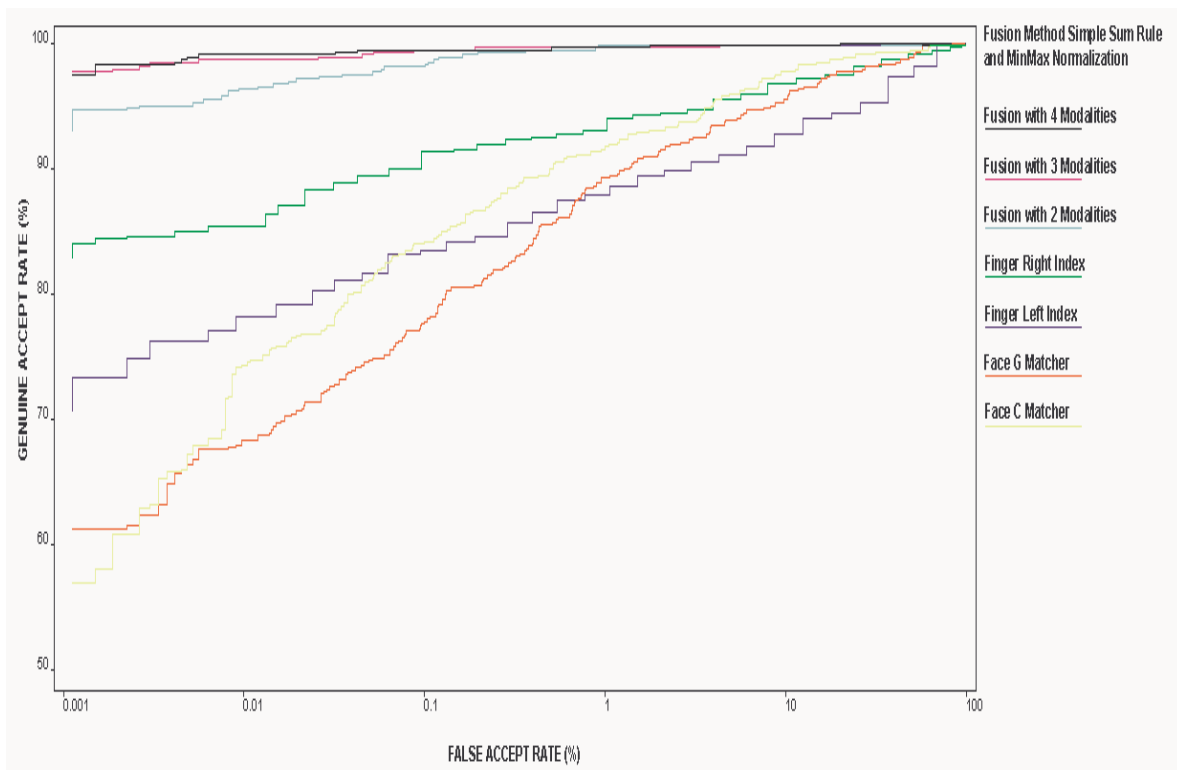


Figure 30: Fusion Simple Sum Rule with Min-Max normalization method on NIST dataset.

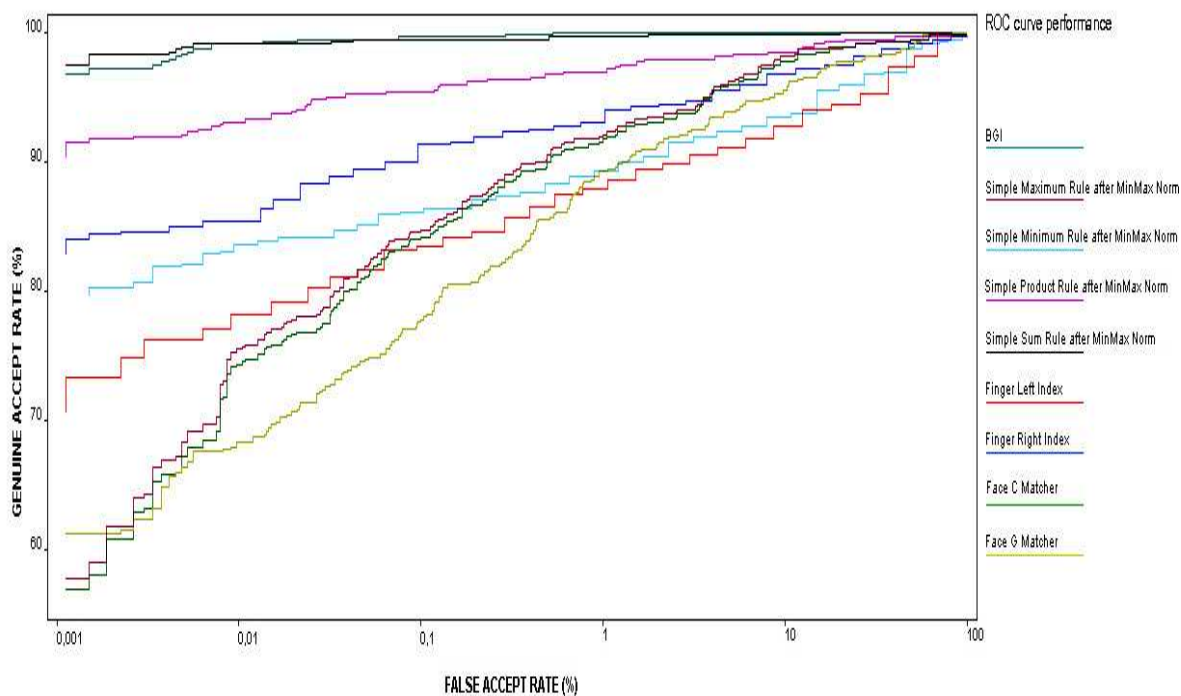


Figure 31: Different fusion methods after Min-Max normalization on NIST dataset.

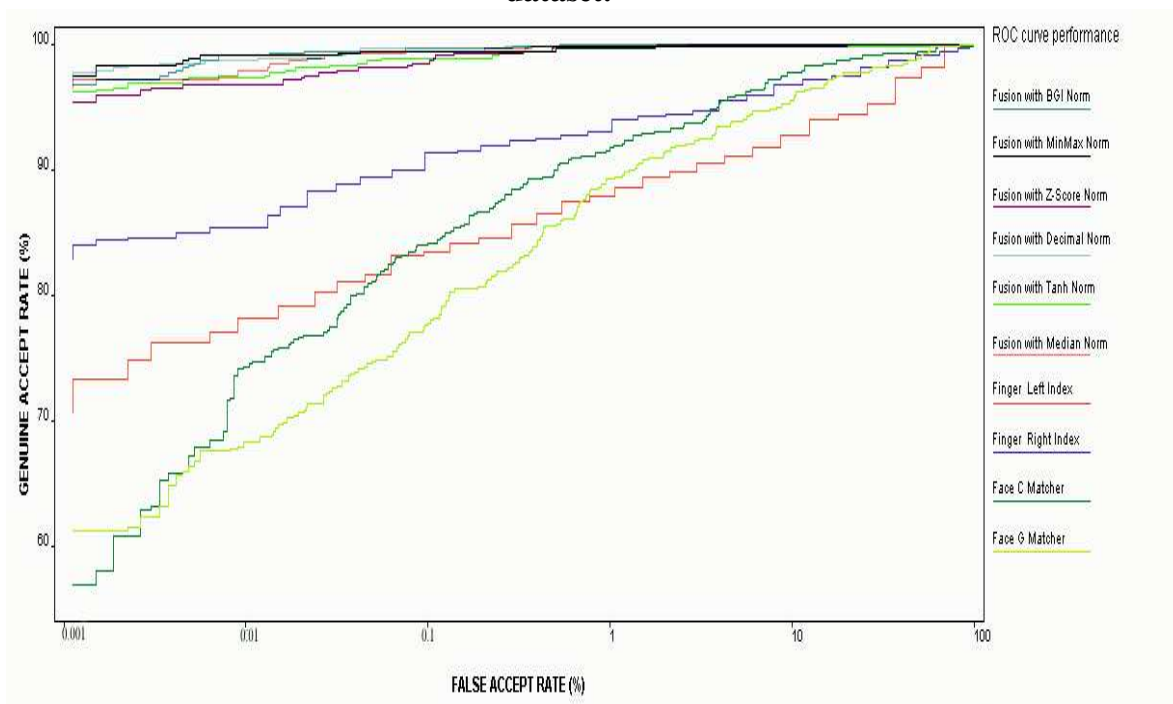


Figure 32: Different normalization methods followed by Simple Sum Rule on NIST dataset.

Obviously, the Simple Sum Rule after Min-Max normalization gives the best performance. The BGI fusion method gives almost as good a performance as the Simple Sum Rule after Min-Max normalization method. This method is convenient because it does not require prior normalization of the scores. However, the performance of this method is highly dependable on the type of the kernel function and the window size used for density estimation. In addition, it should be noted here that this method might require higher computational time because it requires the calculation of the density functions (pdfs) for each modality. This could be significant, especially, if the number of scores is large as is the case with this dataset (for each modality 517 genuine + 266772 imposter scores).

To observe the performance of the normalization methods - Min-Max, Z-Score, Tanh, Median normalization methods the Simple Sum Rule fusion was used and the results are presented in Figure 32. Min-Max normalization followed by the Simple Sum Rule gave the best ROC curve on the NIST dataset.

The overall conclusion is that Min-Max, Tanh and Z-Score normalization methods followed by the Simple Sum Rule give significant improvement in the performance of the multimodal biometric system.

Table 2 summarizes the Genuine Acceptance Rate (GAR) of the multimodal system for different normalization techniques followed by the Simple Sum Rule fusion method and BGI fusion at a False Acceptance Rate (FAR) of 0.1% on the NIST dataset.

Normalization techniques	Fusion techniques	
	Simple Sum Rule	BGI (average value)
Min-Max	99.398	99.525
Z-Score	98.343	
Tanh	98.795	
Median	99.398	
Decimal	99.548	

Table 2: Genuine Acceptance Rate (GAR) (%) of different normalization techniques followed by the Simple Sum Rule fusion method and the BGI method at 0.1% False Acceptance Rate (FAR).

5.3. Experiments with synthetically generated scores using the MSU and NIST datasets

Using goodness of fit statistic tests, we tried to guess the best distribution that describes the scores. Chi-Square and Kolomogorov-Smirnov tests were implemented in MUBI and the results are summarized in Table 3 for the MSU dataset and Table 4 for the NIST dataset. Four parametric distributions were considered Exponential, Normal, Gamma and Weibull. MUBI calculates the parameters for each distribution using MLE. The test statistic value and the p-value for each test has been imported in the fore mentioned tables.

MSU dataset		Exponential Distribution	Normal Distribution	Gamma Distribution	Weibull Distribution	Critical values at 0.05 and 0.01	Best Distribution
Finger Genuine Scores	Chi-Square test value / p value	157.888 / 0	93.472 / 0	136.24 / 0	114.856 / 0	32.671 29.615	normal
	K-s test value/ p value	0.143 / 0	0.066 / 0.025	0.129 / 0	0.093 / 0		
Finger Imposter Scores	Chi-Square test value / p value	162523.531 / 0	163223.598/0	161687.765 / 0	162724.764 / 0	568.902 556.534	exponential
	K-s test value/ p value	0.212 / 0	0.328 / 0	0.283 / 0	0.251 / 0		
Face Genuine Scores	Chi-Square test value / p value	104.384 / 0	230.224 / 0	50.44 / 0	60.12 / 0	32.671 29.615	gamma
	K-s test value/ p value	0.149 / 0	0.164 / 0	0.074 / 0.007	0.078 / 0.005		
Face Imposter Scores	Chi-Square test value / p value	11110.067 / 0	318.184 / 0	822.58 / 0	2557.149 / 0	568.902 556.534	normal
	K-s test value/ p value	0.285 / 0	0.021 / 0	0.051 / 0	0.176 / 0		
Hand Genuine Scores	Chi-Square test value / p value	123.48 / 0	228.552 / 0	72.704 / 0	66.554 / 0	32.671 29.615	weibull
	K-s test value/ p value	0.174 / 0	0.147 / 0	0.07 / 0.013	0.076 / 0.006		
Hand Imposter Scores	Chi-Square test value / p value	3938.947 / 0	5468 .279/0	837.252 / 0	2825.19 / 0	568.902 556.534	gamma
	K-s test value/ p value	0.186 / 0	0.116 / 0	0.035 / 0	0.145 / 0		

Table 3: Results from the Chi-Square and Kolomogorov-Smirnov test on MSU dataset.

The results according to the selected best distributions are marked with red in Table 3 and Table 4.

NIST dataset		Exponential Distribution	Normal Distribution	Gamma Distribution	Weibull Distribution	Critical values at 0.05 and 0.01	Best Distribution
Finger Left Index Genuine Scores	Chi-Square test value / p value	97.34 / 0	89 / 0	77.596 / 0	42.021 / 0.004	32.671 29.615	Weibull
	K-s test value/ p value	0.112 / 0	0.08 / 0.003	0.088 / 0.001	0.077 / 0.004		
Finger Left Index Imposter Scores	Chi-Square test value / p value	15686774.70 / 8 / 0	15712887.708 / 0	15687478.39 / 6 / 0	15699752.03 / 0	568.902 556.534	Normal
	K-s test value/ p value	0.388 / 0	0.162 / 0	0.159 / 0	0.196 / 0		
Finger Right Index Genuine Scores	Chi-Square test value / p value	175.638 / 0	68.064 / 0	69.936 / 0	21.085 / 0.454	32.671 29.615	Weibull
	K-s test value/ p value	0.193 / 0	0.068 / 0.017	0.087 / 0.001	0.026 / 0.874		
Finger Right Index Imposter Scores	Chi-Square test value / p value	15708968.87 / 0	15755990.445 / 0	15708992.33 / 3 / 0	15721143.87 / 5 / 0	568.902 556.534	Gamma
	K-s test value/ p value	0.38 / 0	0.161 / 0	0.271 / 0	0.193 / 0		
Face C Matcher Genuine Scores	Chi-Square test value / p value	5215.553 / 0	101.34 / 0	2393.085 / 0	10815.043 / 0	32.671 29.615	Normal
	K-s test value/ p value	0.586 / 0	0.093 / 0	0.443 / 0	1 / 0		
Face C Matcher Imposter Scores	Chi-Square test value / p value	9757343.81 / 0	205280.135 / 0	3061979.339 / 0	136856615.0 / 1 / 0	568.902 556.534	Normal
	K-s test value/ p value	0.609 / 0	0.216 / 0	0.488 / 0	1 / 0		
Face G Matcher Genuine Scores	Chi-Square test value / p value	1381.936 / 0	23.979 / 0.294	49.766 / 0	10857 / 0	32.671 29.615	Normal
	K-s test value/ p value	0.426 / 0	0.058 / 0.06	0.088 / 0.001	1 / 0		
Face G Matcher Imposter Scores	Chi-Square test value / p value	1293650.557 / 0	242532.293 / 0	263207.276 / 0	137 387 / 579.993 / 0	568.902 556.534	gamma
	K-s test value/ p value	0.456 / 0	0.218 / 0	0.185 / 0	1 / 0		

Table 4 Results from the Chi-Square and Kolomogorov-Smirnov test on NIST dataset.

For every modality from the MSU and NIST dataset, genuine and imposter distributions with parameters are specified in Table 5 and Table 6. For the MSU dataset, 500 and 12250 scores were generated for each genuine distribution and each imposter distribution, respectively, using the parametric distribution. For the NIST dataset contains 517 genuine scores 10000 imposter scores were generated.

For both datasets MSU and NIST, multiple synthetic datasets were generated. In this thesis the results from four such experiments are presented as MSU - case study 1 and 2 and NIST - case study 1 and 2. The range for the scores is taken from the original scores and the shift parameter is calculated according to the domain of the distribution tested.

MSU Scores	Distribution	Parameters	Shift	Range
finger genuine scores	Normal	normal(307.226, 203,681)	0	0 – 966
finger imposter scores	Exponential	exp(6.166)	0	0 – 126
face genuine scores	Gamma	gamma(1.426,14.636)	0	0.6838 - 105.537
Face imposter scores	Normal	normal(114.066, 39.195)	0	17.835033 - 267.889435
hand genuine scores	Weibull	weib(0.003, 1.438)	0	0 – 266
hand imposter scores	Gamma	gamma(2.305, 71.205)	0	25 – 852

Table 5: Summary for the selected distributions for MSU dataset.

The density functions of the generated scores compared with density functions of the original scores are shown in figures 33, 34 and 35 for case study 1 and figures 39, 40 and 41 for case study 2. We can observe that the generated distributions are very similar to the original distributions. The overlapping area between the genuine and imposter distributions of generated scores is almost the same as the overlapping area between the genuine and imposter distributions of original scores. For the MSU dataset, the finger modality has the smallest overlapping area compared to the area for

face and hand geometry in both case studies. In Figure 36, for case study 1 and in Figure 42, for case study 2, it can be observed that the performance of the finger modality is the best of all the three modalities. It should be noted that the performance of generated scores for face and hand in both case studies give worst performance than the modalities from the original data, due to the increased overlapping area between genuine and imposter distributions.

NIST Scores	Distribution	Parameters	Shift	Range
Finger Right Index genuine scores	Weibull	weibull (0.000346, 1.741)	0	0 – 257
Finger Right Index imposter scores	Normal	normal(7.003, 3.096)	0	0 – 43
Finger Left Index genuine scores	Weibull	weibull (0.003, 1.347)	-4	4 - 246
Finger Left Index imposter scores	Gamma	Gamma(3.813, 1.899)	0	0 – 45
Face C Matcher genuine scores	Normal	normal(0.715, 0.105)	0	-1 – 0.89818
Face C Matcher imposter scores	Normal	normal(0.529, 0.074)	0	-1 – 0.73154
Face G Matcher genuine scores	Normal	normal(76.783, 2.883)	0	64.80565 – 83.49406
Face G Matcher imposter scores	Gamma	gamma(15.323, 0.786)	-54.835	54.83538 – 76.48196

Table 6: Summary for the selected distributions for NIST dataset.

To evaluate the impact of the normalization method on the generated dataset all normalization methods implemented in MUBI were tested and the Simple Sum Rule fusion method was applied to combine the scores. Figure 36 the summarizes the fusion ROC curves along with the ROC curves from the individual modalities for the case study 1 and in Figure 42 for case study 2. All fusion combinations give better performance than each individual modality. The best result presented for the case study 1 is the BGI fusion method with Gaussian kernel function and window size 10. Tanh and Z-Score normalization followed by the Simple Sum Rule fusion method

also gave very good performance. The best result for the case study 2 is the Simple Sum Rule fusion method with Tanh and Median normalization method.

In order to see if the fusion combinations on generated data give better performance than the same fusion combination on the original data, one of the best fusion combinations is selected and compared with the same combination on the original data. Figure 37 shows one of the best methods on generated data – BGI compared with the BGI on original data from MSU dataset. We can see that BGI on generated data outperforms the BGI on original data.

In addition, to compare the fusion combination with normalization included, Min-Max normalization method followed by the Simple Sum Rule fusion method is compared with the same combination on the original MSU dataset in Figure 38. We can see that the fusion combination on the original data gives better performance.

The performance evaluation of the case study 2 dataset is presented in Figure 43 and 44. Figure 43 shows BGI fusion method performed on the generated data compared with the BGI on the original data from MSU dataset. We can see that BGI on generated data outperforms the BGI on original data on some parts of the ROC curve. In Figure 44, the Simple Sum Rule fusion method after Z-score normalization is compared with the same combination on the original data from the MSU dataset. The fusion combination on the original data gives better performance. However, the fusion combination from the generated data is very close to the original one and provides better performance than each individual modality.

Conclusion

Experiments above show that the best performance fusion combination in the both cases for generated data did not necessary gave the same combination for fusion on the original dataset. Only BGI fusion method outperformed the BGI from original dataset in the both case studies.

Table 7 summarizes the Genuine Acceptance Rate (GAR) of the multimodal system for different normalization techniques followed by the Simple Sum Rule fusion

method and BGI fusion at a False Acceptance Rate (FAR) of 0.1% for both case studies.

All normalization methods followed by the Simple Sum Rule gave better performance on the original MSU dataset. In addition, interestingly, though the datasets in both case studies were generated by the same distributions and same parameters, the combination for the overall winner was observed to be different in both the cases. In general BGI performs very good for all datasets. Tanh and Z-Score normalization followed by the Simple Sum Rule perform well in the first case study where Median normalization followed by the Simple Sum Rule performed well in the second.

Normalization Techniques	MSU casa study 1		MSU casa study 2	
	Simple Sum Rule	BGI (average value)	Simple Sum Rule	BGI (average value)
Z-Score	99.398	99.3974	97.59	98.291
Min-Max	97.544		96.837	
Tanh	99.398		97.892	
Median	98.343		98.193	
Decimal	97.741		96.687	

Table 7: Genuine Acceptance Rate (GAR) (%) of different normalization techniques followed by the Simple Sum Rule fusion method and the BGI method at 0.1% False Acceptance Rate (FAR).

5.3.1 Case study 1 – MSU dataset

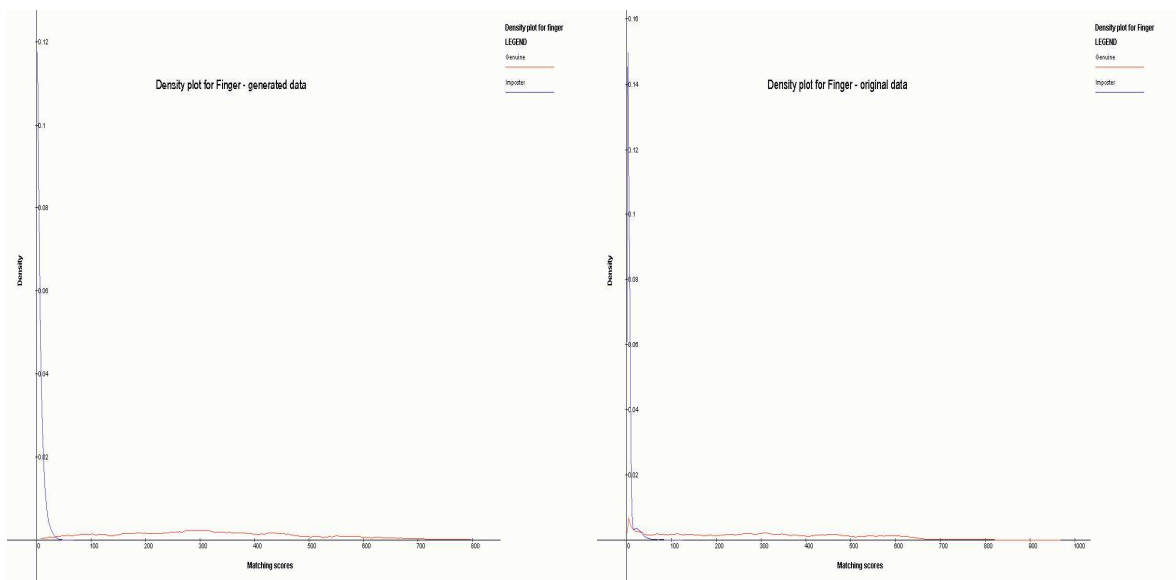


Figure 33: Density plots of the Finger scores for generated data and original data.

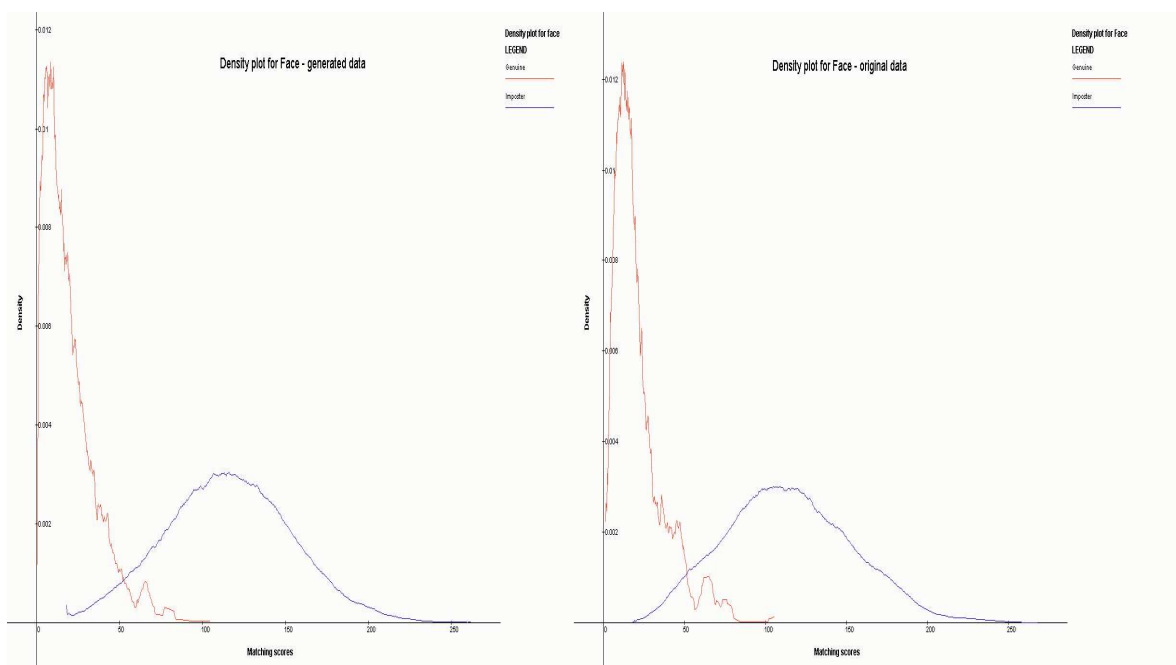


Figure 34: Density plots of the Face scores for generated data and original data.

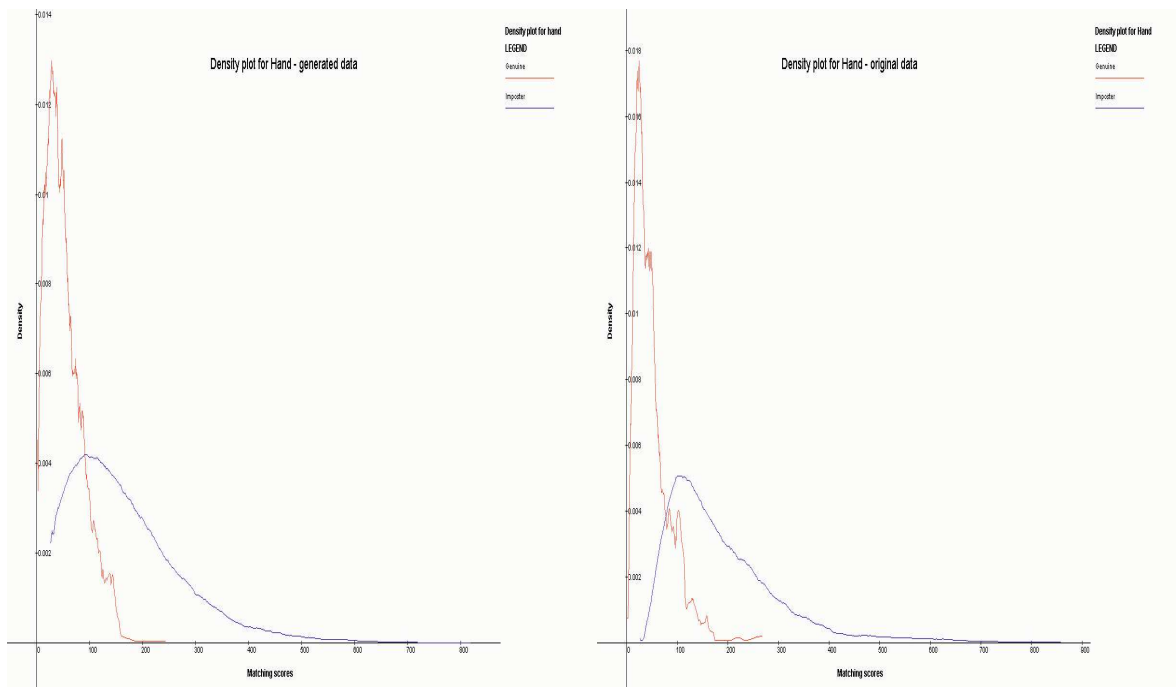


Figure 35: Density plot of the Hand scores for generated data and original data.

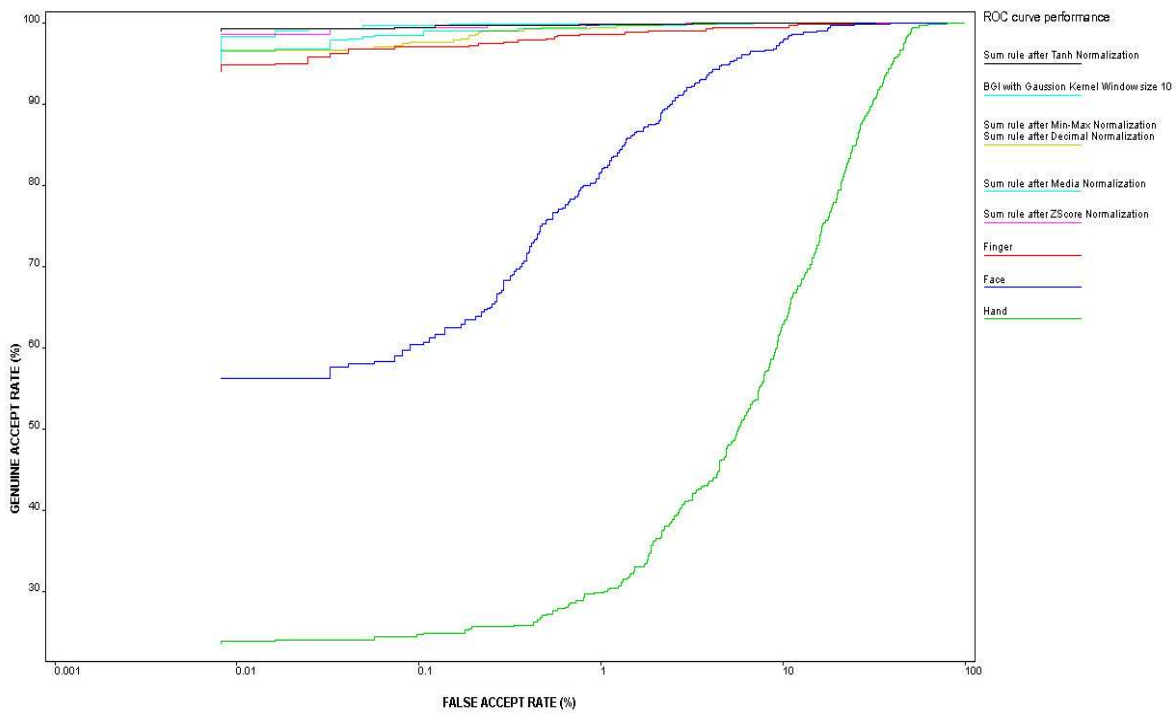


Figure 36: Fusion methods with different normalizations on the generated data.

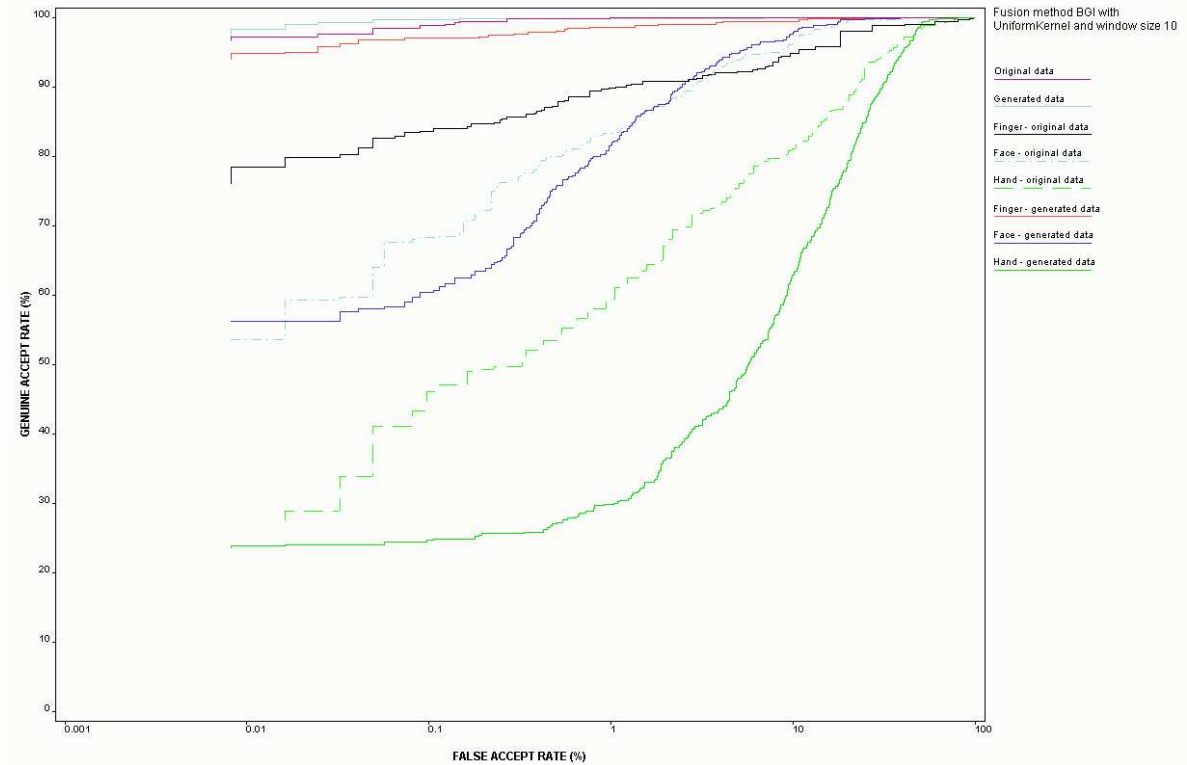


Figure 37: BGI Fusion method on generated data compared with BGI method on original data.

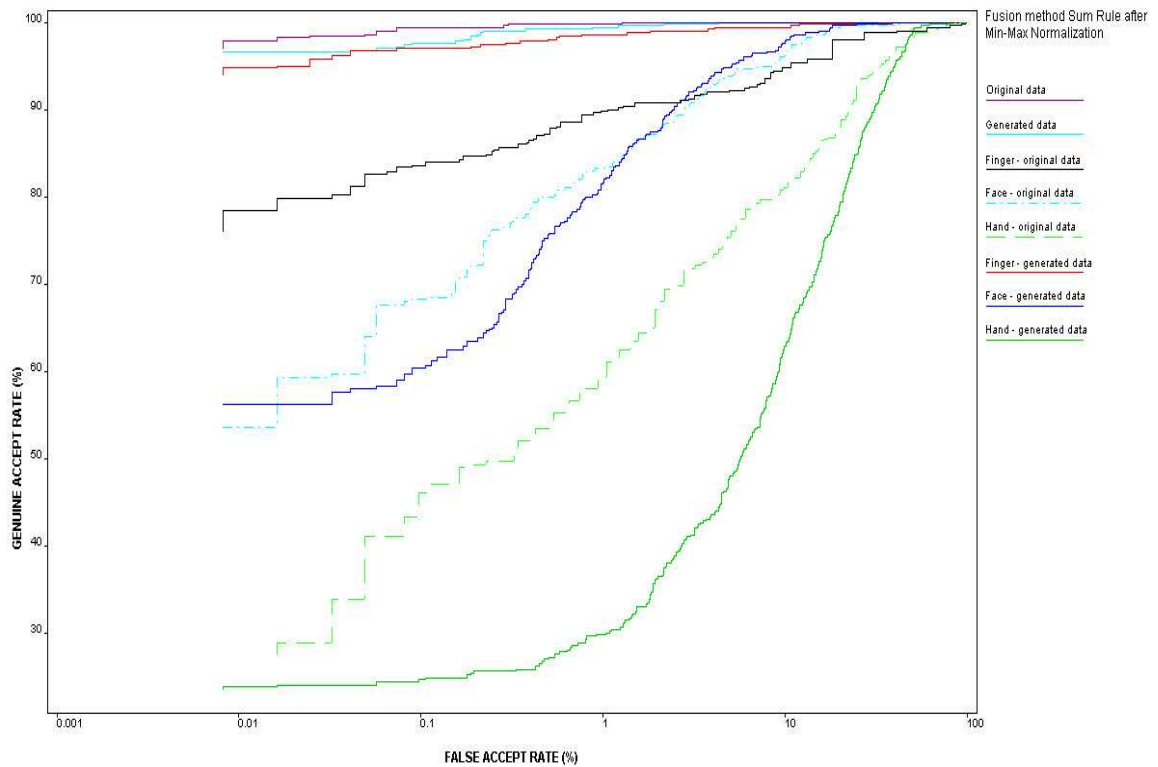


Figure 38: Simple Sum Rule Fusion method after Min-Max normalization on generated data compared with same method on original data.

5.3.2 Case study 2 – MSU dataset

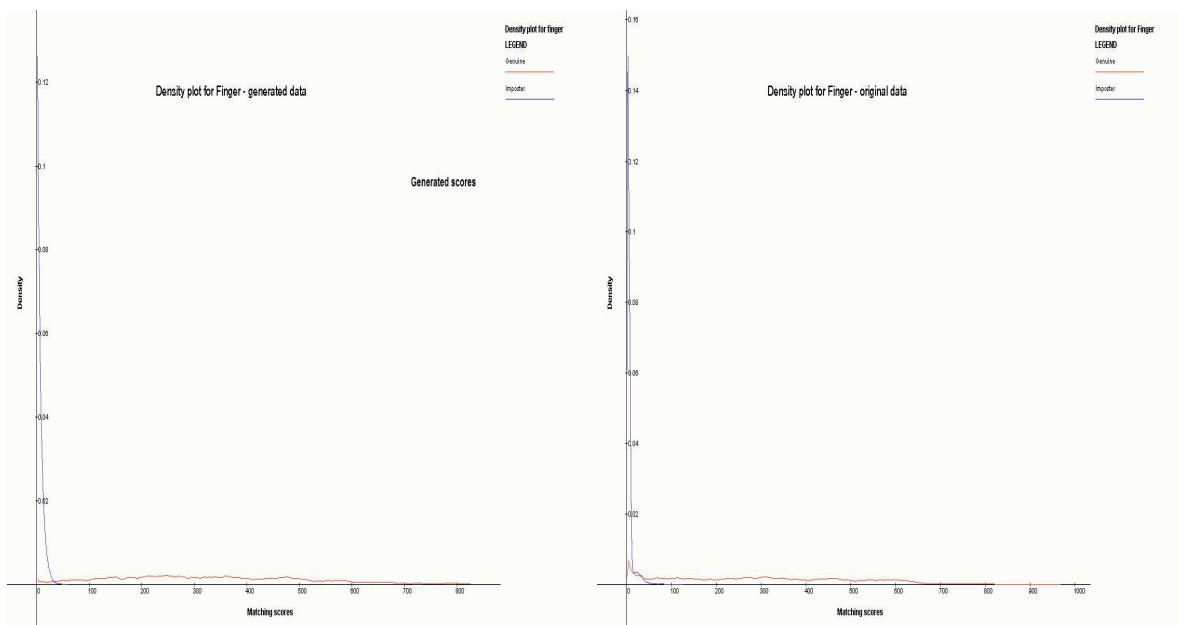


Figure 39: Density plots of the Finger scores for generated data and original data.

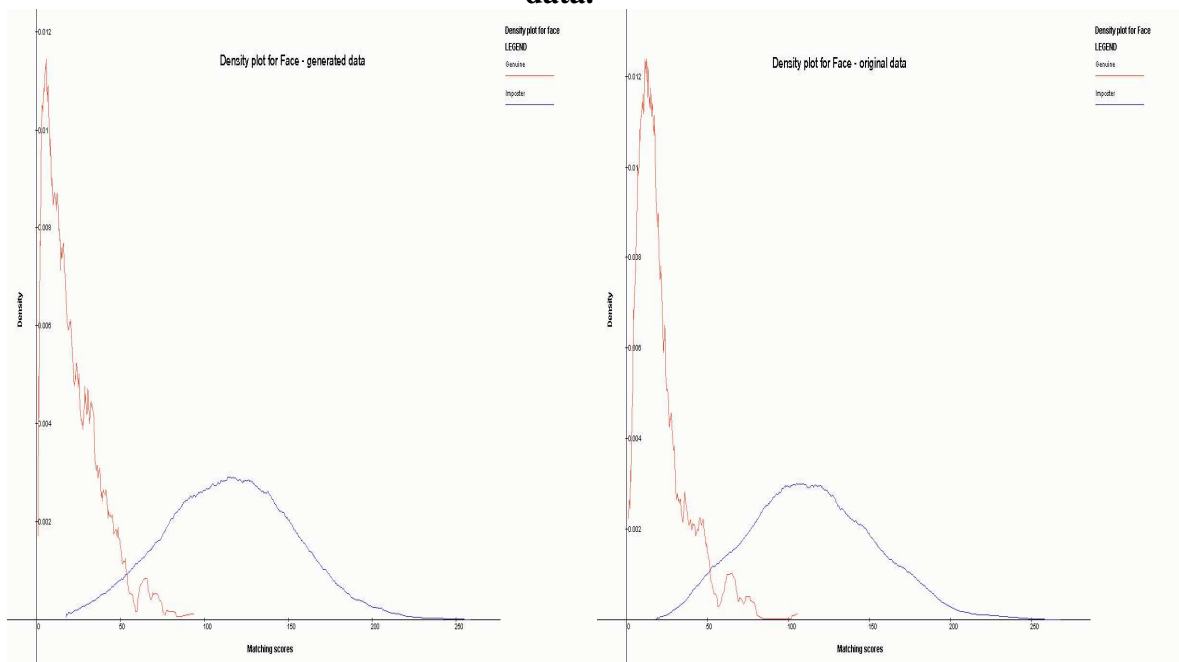


Figure 40: Density plot of the Face scores for generated data and original data.

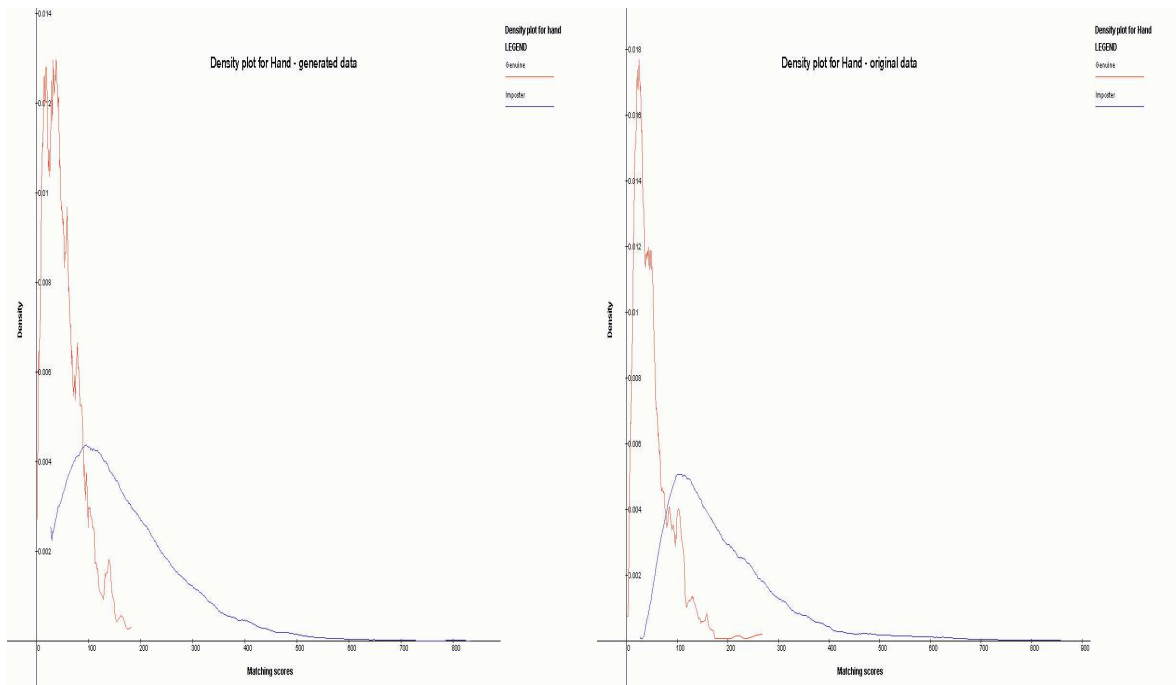


Figure 41: Density plot of the Hand scores for generated data and original data.

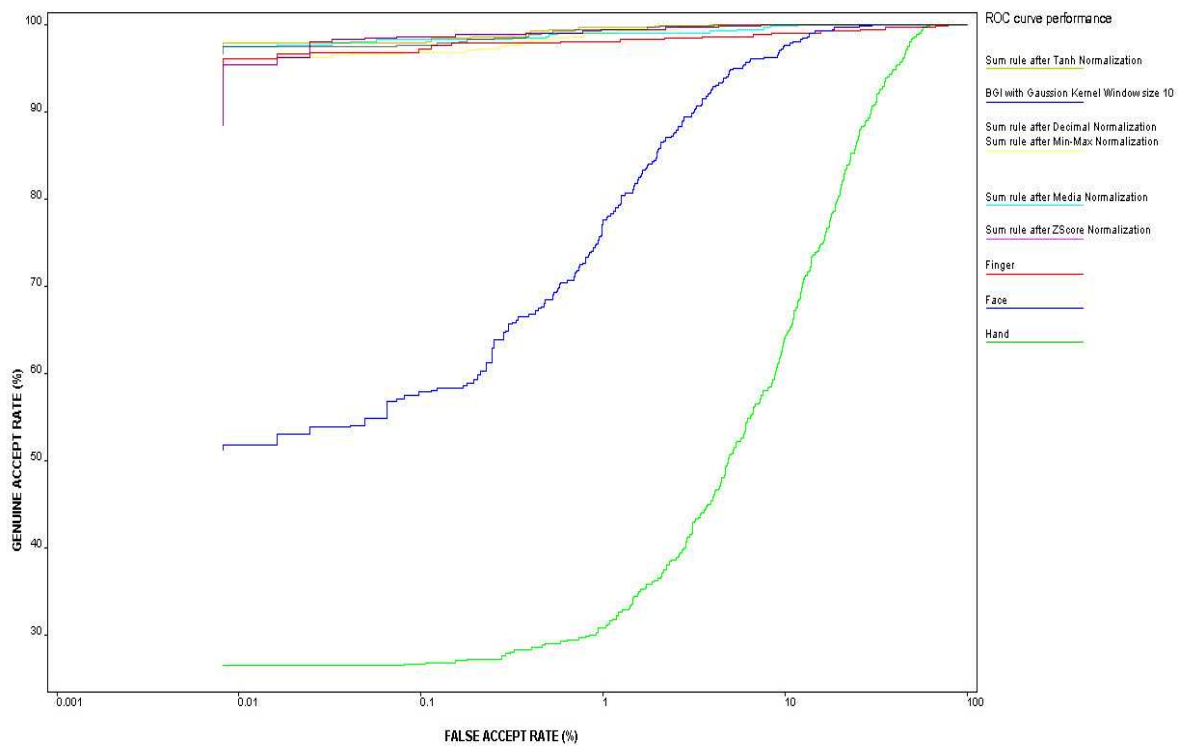


Figure 42: Fusion methods with different normalizations on the generated data.

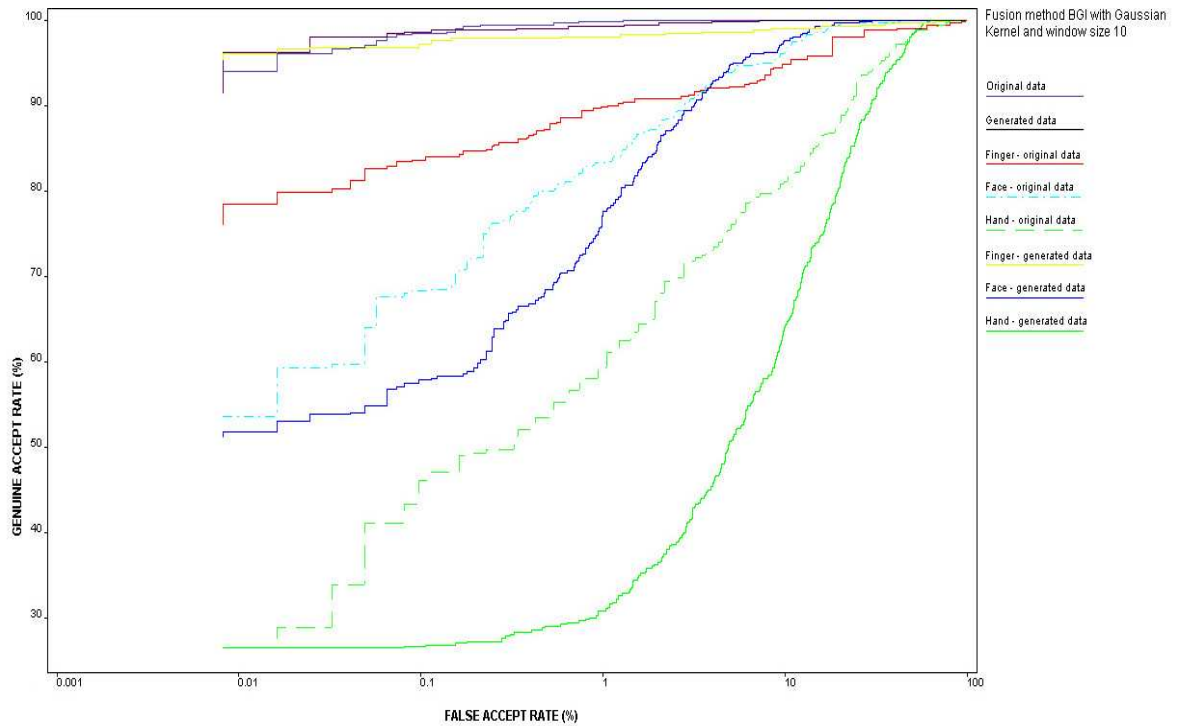


Figure 43: BGI Fusion method on generated data compared with BGI method on original data.

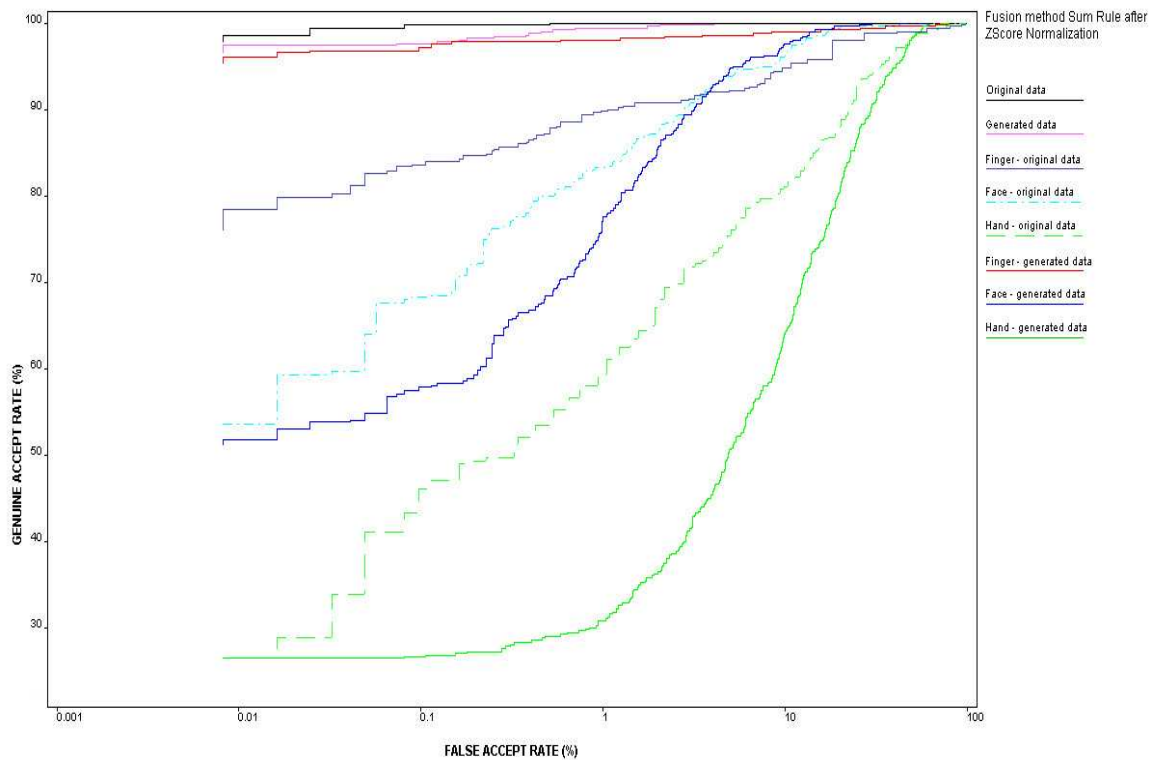


Figure 44: Performance of the Simple Sum Rule Fusion method after Z-score normalization on generated data compared with the same method on the original data.

NIST dataset

Similar experiments were conducted for the NIST dataset. According to Table 5 two datasets were generated. The density functions of the generated scores compared with density functions of original scores are shown in Figure 45 for finger left index, Figure 46 for finger right index, Figure 47 for face C matcher and Figure 48 for face G matcher for case study 1 and in Figure 51 for finger left index, Figure 52 for finger right index, Figure 53 for face C matcher and Figure 54 for face G matcher for case study 2. As before, the distributions are very similar to the original ones, and the overlapping area is almost the same in both cases. Right Finger Index is the best modality in both case studies and the ROC curves for the performance of each modality can be seen in Figure 49 and Figure 55. Generated datasets for the finger modality give better performance than the original ones, which is not the case with the face modality. The density distributions for the face C matcher and face D matcher deviate from the original distributions. The overlapping area between the genuine and imposter scores for these two modalities is larger compared with the overlapping area for the original scores, and the ROC curves for both modalities in the case studies gives inferior performance.

In Figure 49 for case study 1 and in Figure 55 for case study 2, the ROC curves generated by testing all normalization methods followed by the Simple Sum Rule fusion method is presented. It can be observed that the ROC curve of the fusion scenarios gives better performance than each individual modality. The best result in both cases was obtained with the Simple Sum Rule fusion method and the Median normalization method.

The Simple Sum Rule fusion method after Median normalization is compared with the same method of the original data from the NIST dataset. The ROC curve of the generated data looks better than the ROC curve of the same combination on original data in Figure 50 for case study 1 and in Figure 56 for case study 2.

Conclusion

From the experiments on the NIST dataset, it can be observed that again as for the MSU dataset, the synthetic data does not result in the same performance as the original dataset for the same normalization/fusion combination. For both case studies, involving synthetic scores the best combination method is Median and MAD normalization combined with the Simple Sum Rule fusion. In both the cases, this combination outperformed the same combination from the original NIST dataset. The experiments also illustrate that the ROC curves of all normalization techniques in combination with the Simple Sum Rule fusion method are very close to each other and yield very good performance in both case studies.

Table 8 summarizes the Genuine Acceptance Rate (GAR) of the multimodal system for different normalization techniques followed by the Simple Sum Rule fusion method and BGI fusion method at a False Acceptance Rate (FAR) of 0.1% for both the case studies.

Normalization Techniques	NIST casa study 1		NIST casa study 2	
	Simple Sum Rule	BGI (average value)	Simple Sum Rule	BGI (average value)
Z-Score	99.699	99.533	99.548	99.652
Min-Max	98.398		98.946	
Tanh	97.59		98.042	
Median	99.849		99.699	
Decimal	99.699		99.247	

Table 8 Genuine Acceptance Rate (GAR) (%) of different normalization techniques followed by the Simple Sum Rule fusion method and BGI at 0.1% False Acceptance Rate (FAR).

5.3.3 Case study 1 - NIST dataset

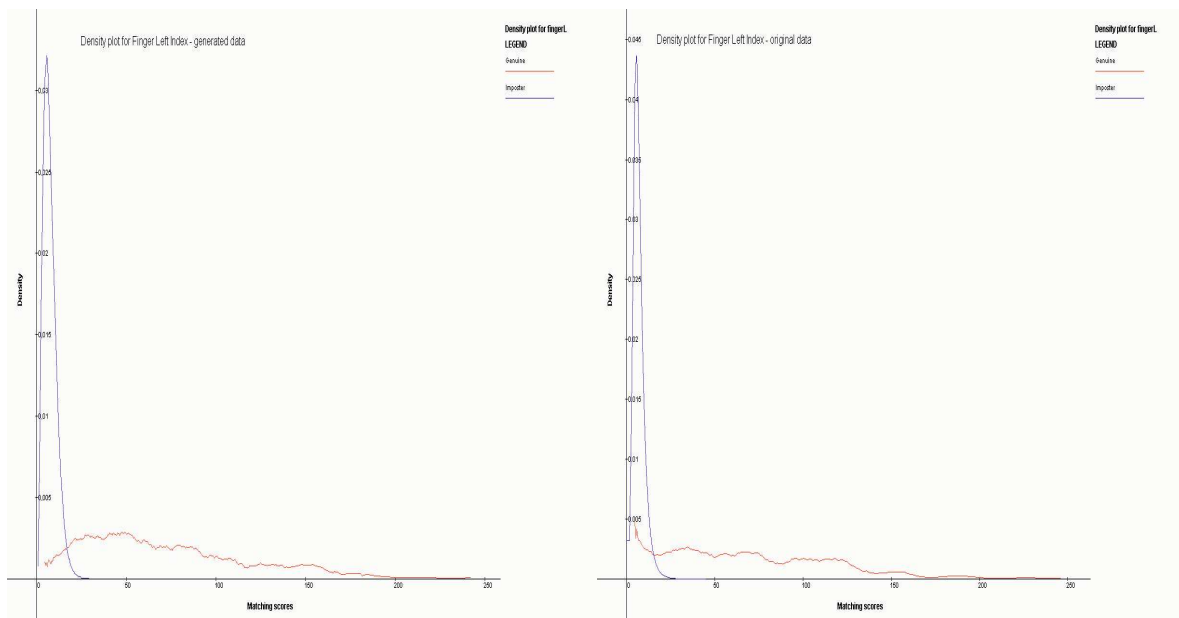


Figure 45: Density plots of the Finger Left Index scores for generated data and original data.

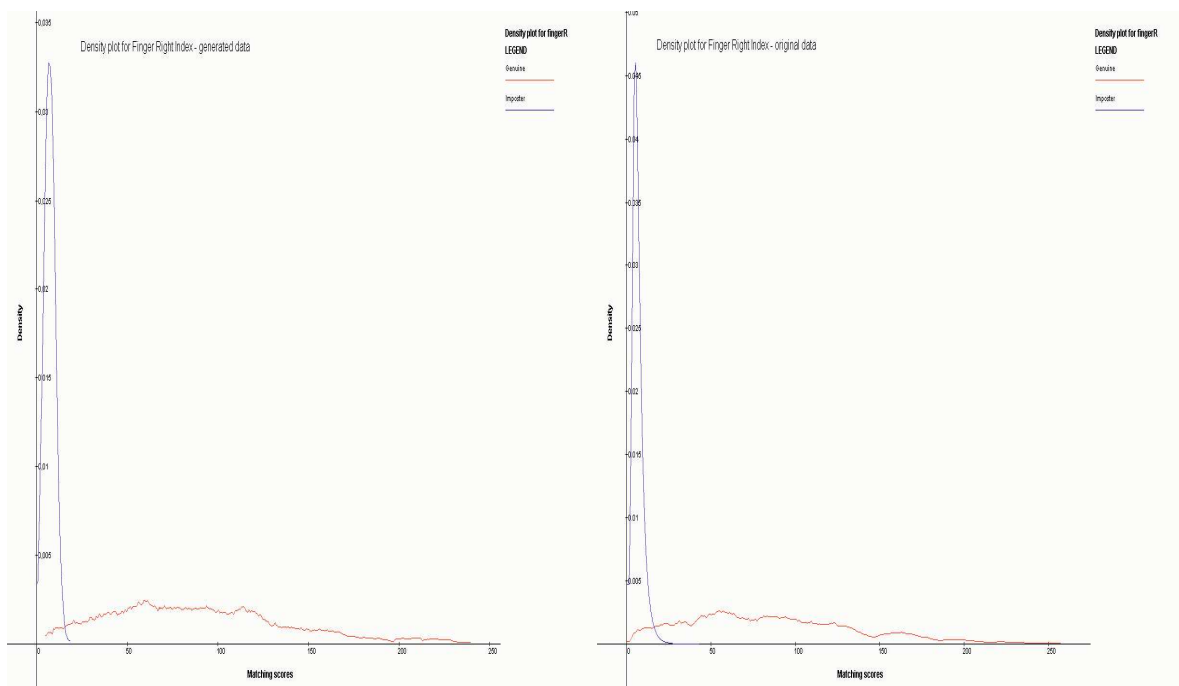


Figure 46: Density plots of the Finger Right Index scores for generated data and original data.

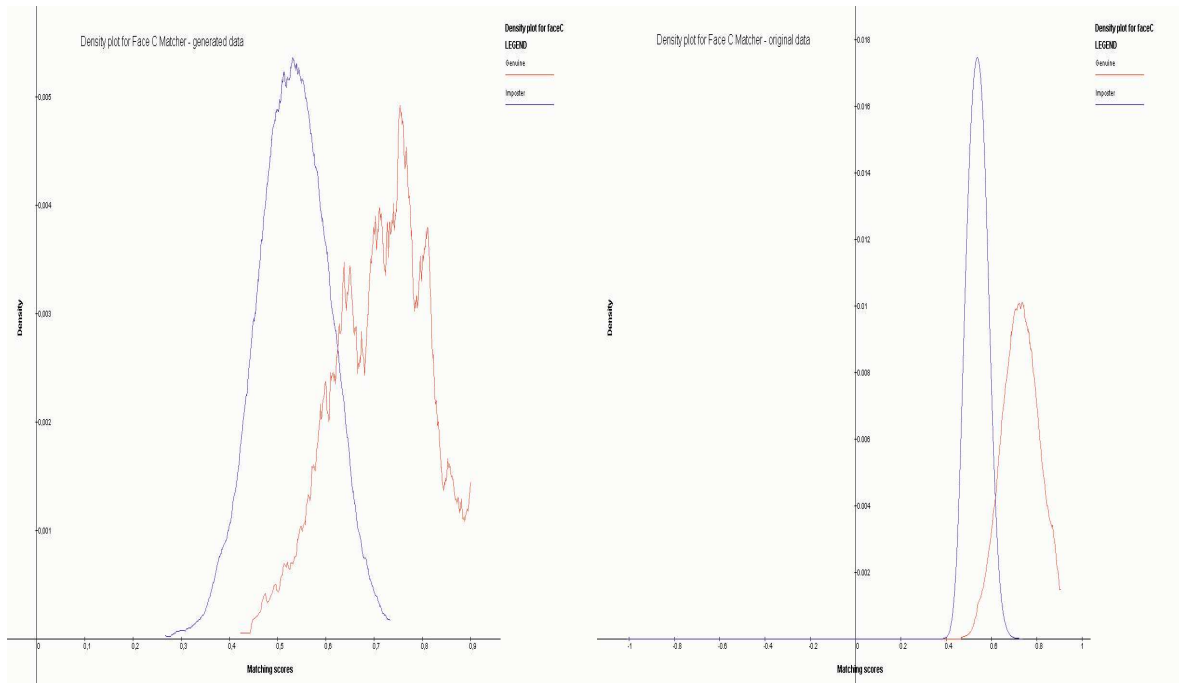


Figure 47: Density plots of the Face C Matcher scores for generated data and original data.

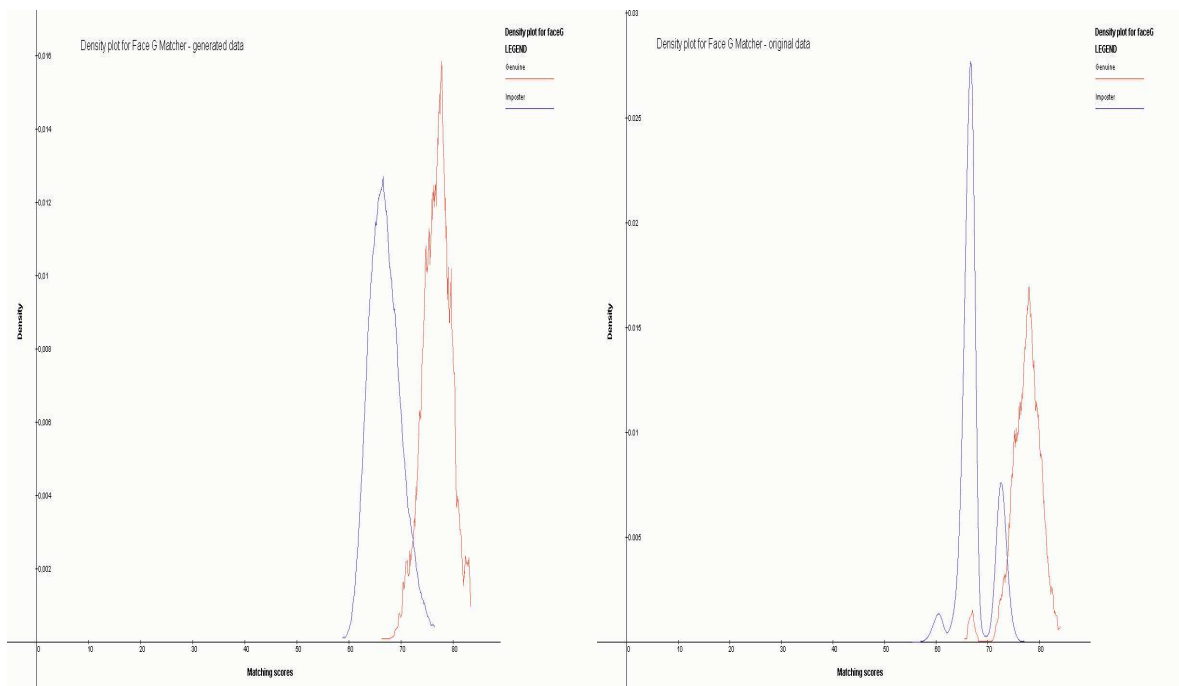


Figure 48: Density plots of the Face G Matcher scores for generated data and original data.

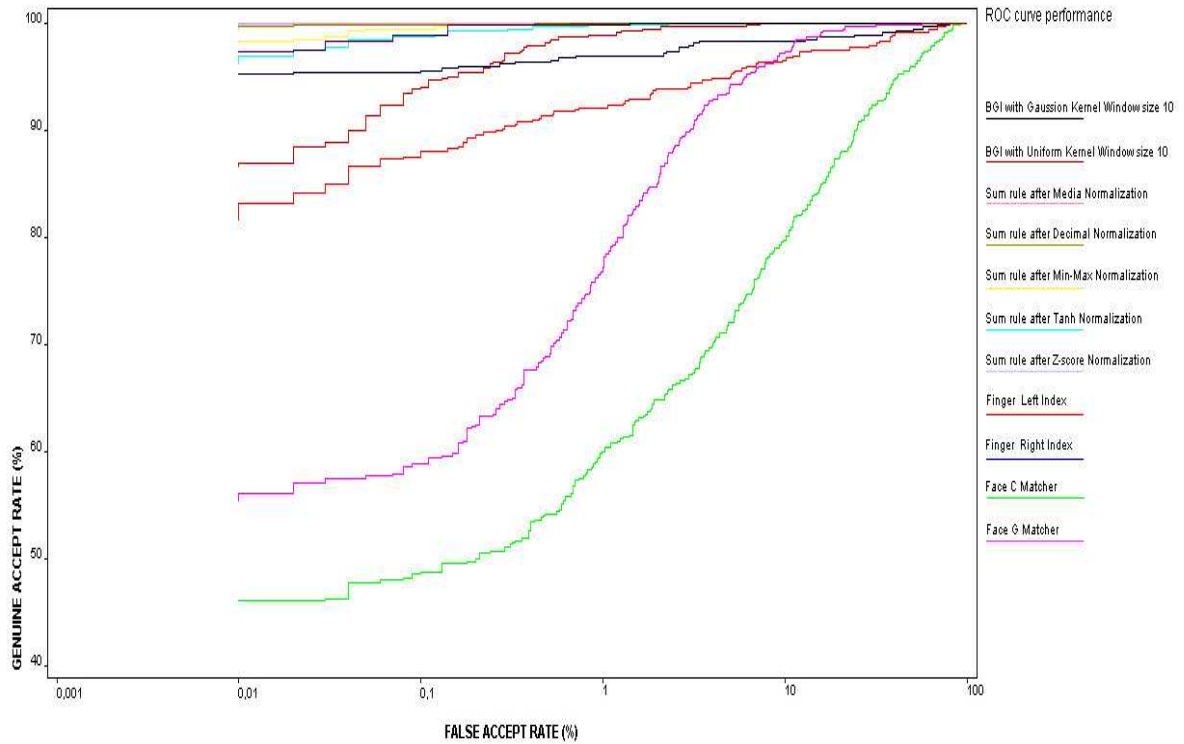


Figure 49: Fusion methods with different normalizations on the generated data.

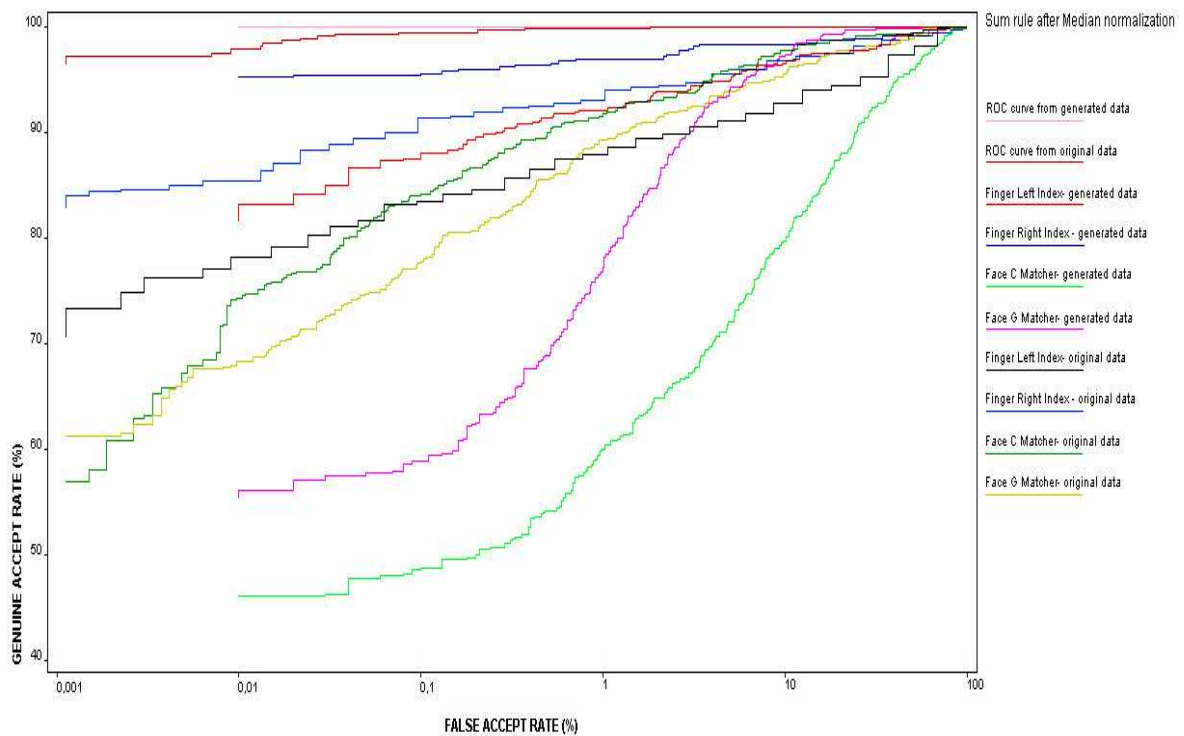


Figure 50: Simple Sum Rule fusion method after Median normalization on generated data compared with same method on original data.

5.3.4 Case study 2 – NIST dataset

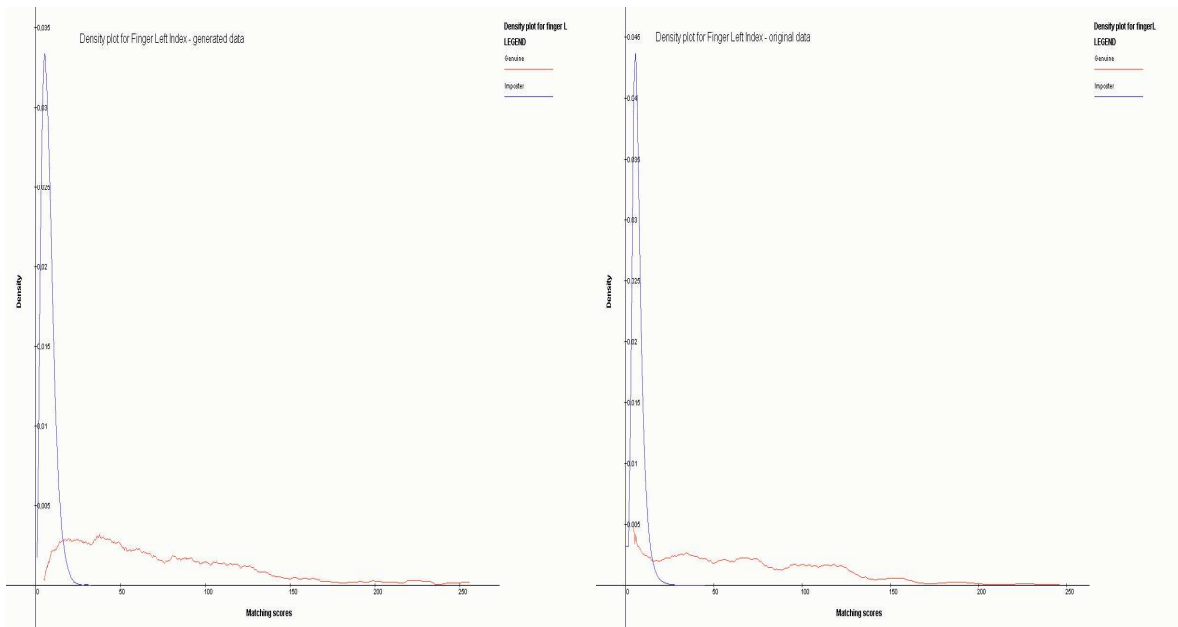


Figure 51: Density plots of the Finger Left Index scores for generated data and original data.

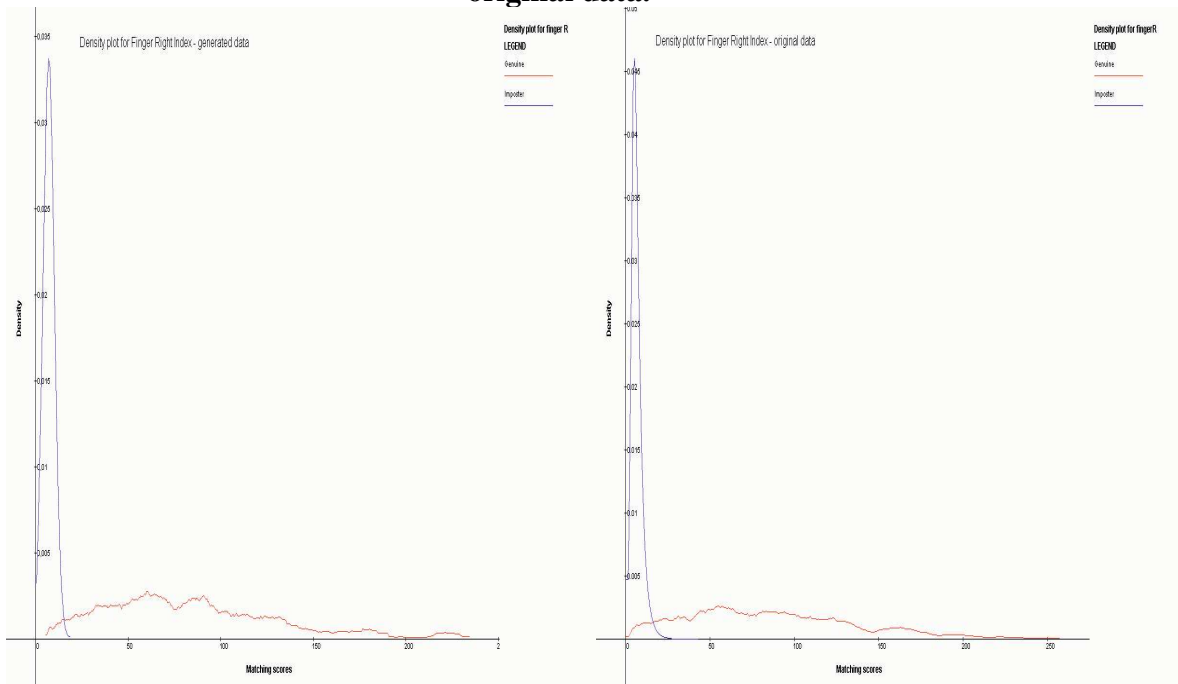


Figure 52: Density plots of the Finger Right Index scores for generated data and original data.

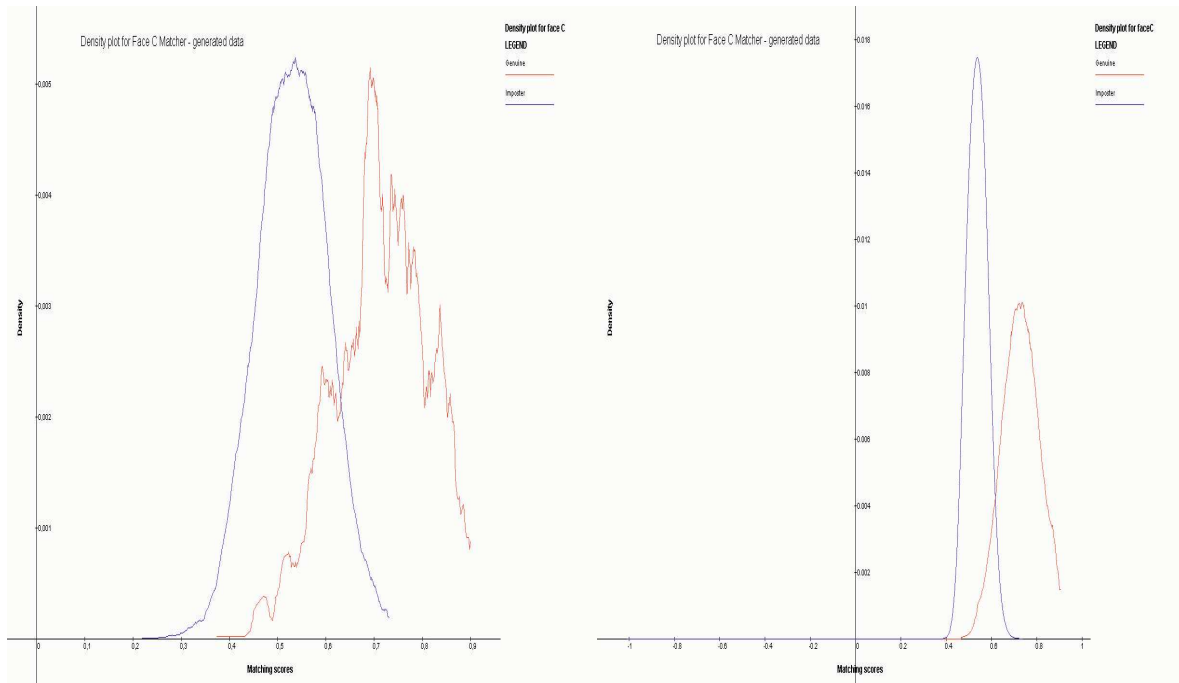


Figure 53: Density plots of the Face C Matcher scores for generated data and original data.

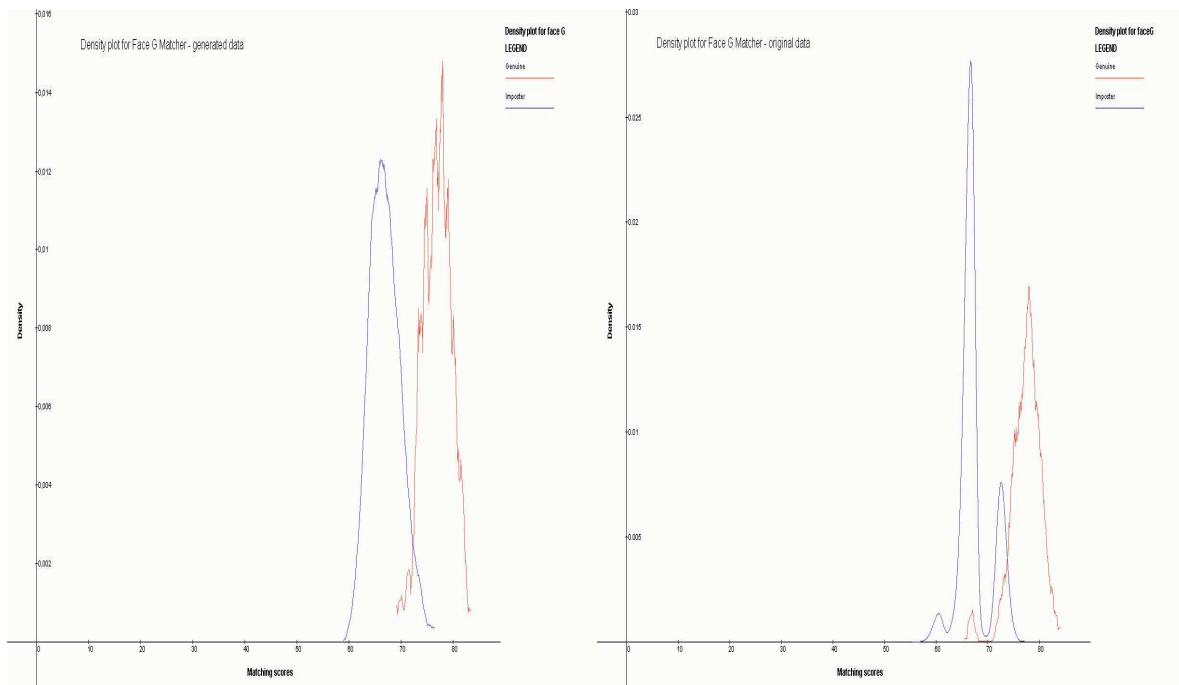


Figure 54: Density plots of the Face C Matcher scores for generated data and original data.

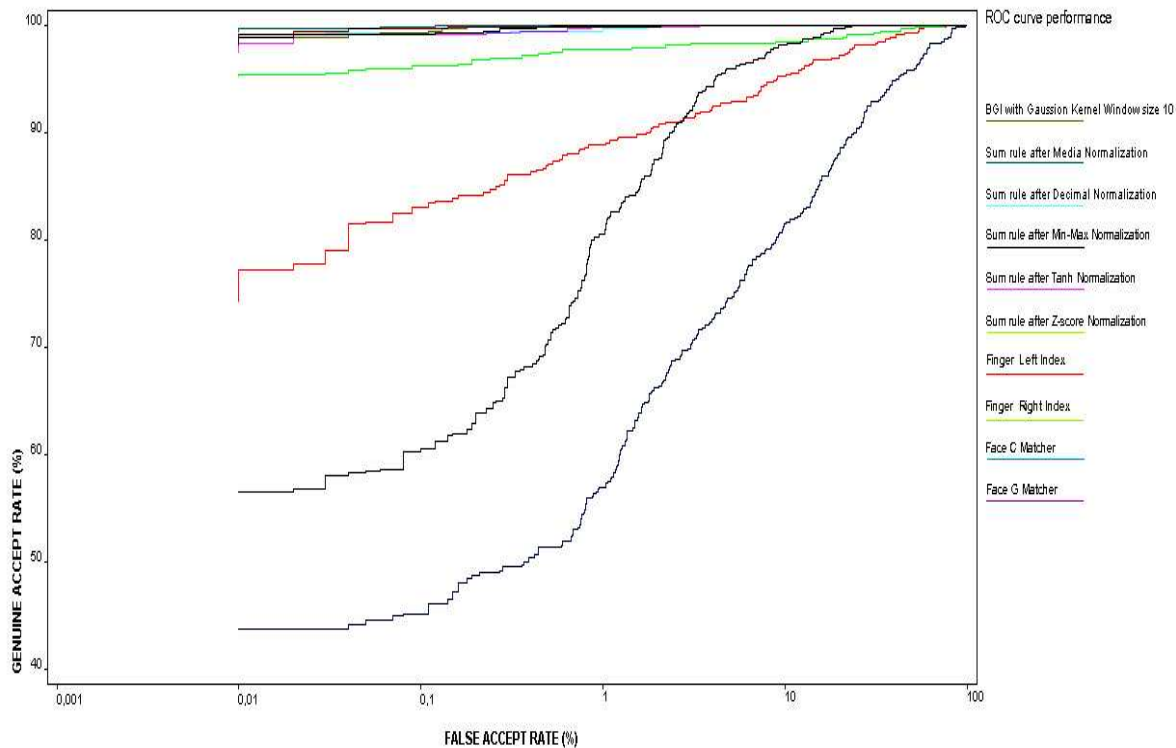


Figure 55: Fusion methods with different normalizations on the generated data.

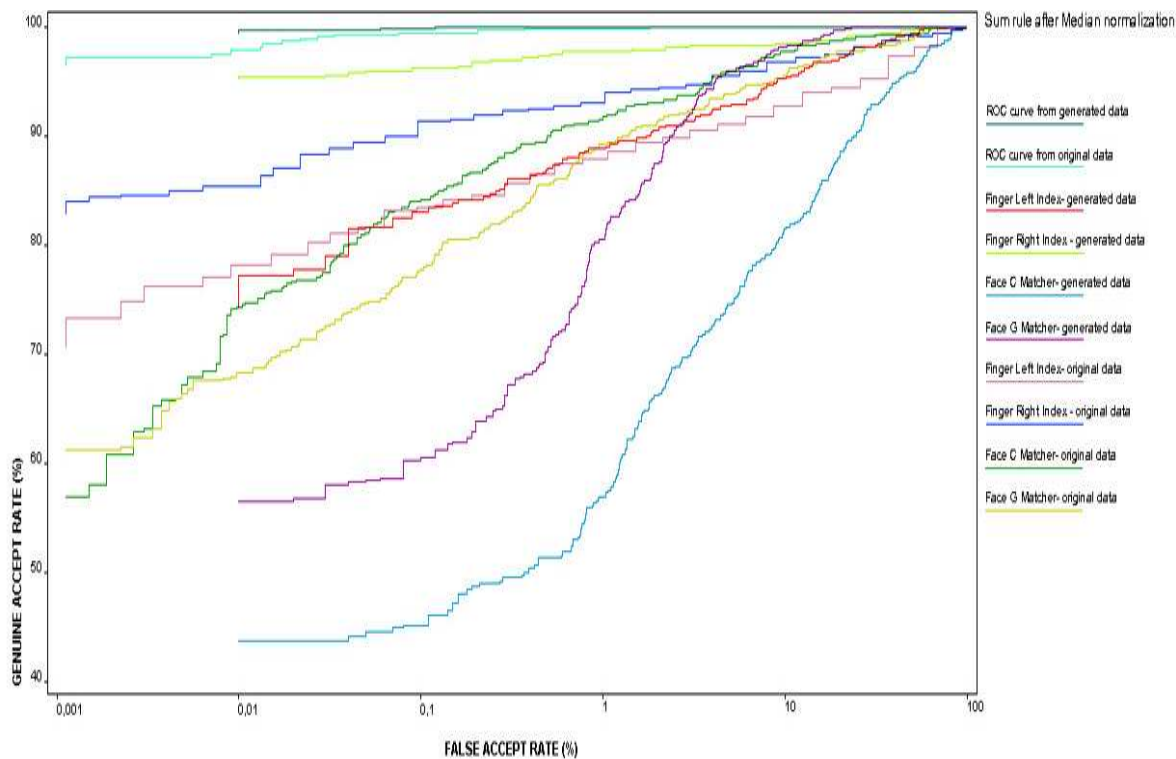


Figure 56: Simple Sum Rule fusion method after Median normalization on generated data compared with same method on original data.

Chapter 6 Discussion and Conclusion

Chapter 6

Discussion and Conclusion

From the experiments and the results it is demonstrated that combining multiple sources of evidence at the score level improves the performance of the biometric system. Two important aspects of fusion at this level are the normalization of the data to obtain values or scores in a common range and the fusion rule used to combine the scores. It is demonstrated here and in [5] that the normalization of scores prior to combining them improves the recognition performance of a multimodal biometric system that uses the face, fingerprint and hand-geometry traits for user authentication.

A methodology for testing multimodal biometric systems introduced in [5] is implemented into the MUBI tool. This methodology provides a general framework for conducting normalization and fusion evaluations. The basis of this methodology is applying fusion schemes after the individual biometric match-scores are determined. Fusion at this stage does not affect the existing proprietary biometric systems allowing a common middle layer to handle the multimodal application but with a small amount of common information. Another advantage of using match scores is that the data from prior evaluations of single-modal biometric systems can be reused. This avoids live testing or re-running individual biometric algorithms. The advantage of fusion at the match-score level is that existing and proprietary single-model

biometric systems can easily be integrated into a multimodal biometric environment if some basic information is provided by these existing systems. The required information does not expose the internal operations of these systems.

Scores output by individual matchers are non-homogeneous and they can be distance or similarity scores. Also, ranges of the scores may be different, for example [0,100] or [0,1000] along with their distributions.

For predictive purposes, it is often desirable to understand the shape of the underlying distribution of the matching scores. The tool plots the genuine and imposter distribution for each modality. In a perfect system, the genuine distribution would be non-overlapping with the imposter distribution. If the users are more habituated with the application, it will be easier for the users to give same samples and this will have influence on the genuine distribution moving it away from the imposter distribution and reducing the overlapping area between them. This movement will have direct impact on the imposter distribution because they are in tight relationship. At this moment, there is no way how we can predict the translation of the distributions under various operational scenarios.

In addition, the tool can perform distribution fitting. To determine the underlying distribution, the tool fits the observed distribution to a theoretical distribution by comparing the frequencies observed in the data to the expected frequencies of the theoretical distribution.

The tool searches for the set of parameters that make the closest match between the distribution function and each genuine and imposter dataset. It will identify a distribution that is most likely to produce a particular match score dataset. The results should be evaluated quantitatively and qualitatively, before using it to generate synthetic scores. Two methods are used for calculating the best distributions for the datasets. For sample data, distribution parameters are estimated using Maximum Likelihood Estimators (MLEs). Datasets are compared against four distributions: Gamma, Normal, Exponential, and Weibull.

After determining which distribution fits best, the scores are generated with the suggested distribution and compared with the fusion methods of the original datasets. In this work, it is shown that even though the scores follow the same distribution the

performance of the system for the same normalization/fusion combination is different. This suggests that the performance of the multimodal biometric system completely depends of the dataset. Even small changes in the scores result in different normalization/fusion combinations leading to better performances. It implies that general rules that would give best performance on different datasets cannot be developed. From the experiments, it can be concluded that in general the Simple Sum Rule and BGI perform well over all datasets. However, no inferences can be made to predict the performance of the system if it is not tested on the representative data. This suggests the importance of a tool that would allow a practitioner to experiment with different combinations of normalization and fusion schemes, analyze the data and directly compare different fitting techniques. One can choose a normalization and fusion scheme after analyzing the genuine and imposter score distributions of the individual matchers. It is necessary to develop a strategy to choose the design parameters of normalization. With this work, we have addressed some limitations of deploying of multimodal systems like the lack of a common testing framework and the absence of tools to evaluate and build such systems.

	MSU - overall winners	NIST - overall winners
Original dataset	Tanh + Simple Sum Rule BGI	Min-Max + Simple Sum Rule BGI
Case study 1	Tanh + Simple Sum Rule Z-Score + Simple Sum Rule BGI	Median + Simple Sum Rule Z-Score + Simple Sum Rule BGI
Case study 2	Median + Simple Sum Rule Tanh + Simple Sum Rule BGI	Median + Simple Sum Rule Z-Score + Simple Sum Rule BGI

Table 9: Summary of the experiments on generated data.

Finally, MUBI can be used one to speculate the number of required subjects to get statistically significant results for a specific application and to decide the size of the

test set. The size of the test size affects the performance accuracy. The larger the test the more accurate the results are expected to be.

Consequently, MUBI is a performance prediction tool under General Public License that is available to everyone who wants to evaluate multimodal biometric systems. In addition, the source code of MUBI is available and its modular structure will allow users to upgrade the tool and to implement more methods if needed.

Future work

Currently, there are thousands of users who successfully use biometric devices in a wide variety of applications. The proposed tool provides system integrations with the ability to test and explore different fusion schemes on different datasets. Because the datasets are fixed, results of technology tests are repeatable. The goal is to provide a tool for testing results, which will aid current and prospective users in selecting and employing biometric technologies in a secure, user-friendly, and cost-effective manner. MUBI allows a system designer to model hypothetical multimodal biometric systems that can vary the biometric modalities, matching algorithm, normalization and fusion techniques, and sample databases. Given this tool, systems can be built to optimally match specific application requirements.

In this thesis, we tried to model the distributions of the genuine and imposter scores in order to generate synthetic datasets that would represent the original ones. From the experiments was demonstrated that even small variations in the tail of the modeled distribution have a great impact on the fusion performance. Synthetic datasets can provide an initial estimation of the matching performance. However, this performance may not represent the actual performance if the tails of the distribution are not accurately modeled. Thus it is important to model the tails of the distribution of the genuine and imposter scores appropriately in order to derive benefits from synthetic datasets.

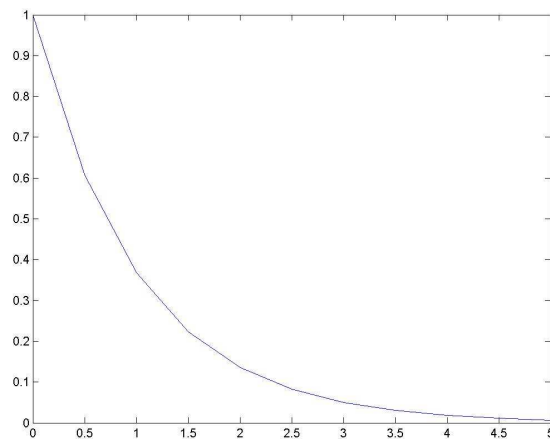
The idea presented in this thesis can be extended as the science of biometric testing becomes more mature. Currently there is no scientific-based general biometric testing

protocol that ensures valid results, which may result in partial or sometimes misleading results. This tool will aid in developing standard testing and reporting protocols for multimodal biometric systems and it will help testers to achieve the best possible estimation of the performance while applying minimum effort to perform the evaluation, itself. In future, one can include the testing of representative data from large populations and expanding MUBI by incorporating novel fusion methods that can be evaluated against current fusion schemes.

Appendix A

Exponential Distribution

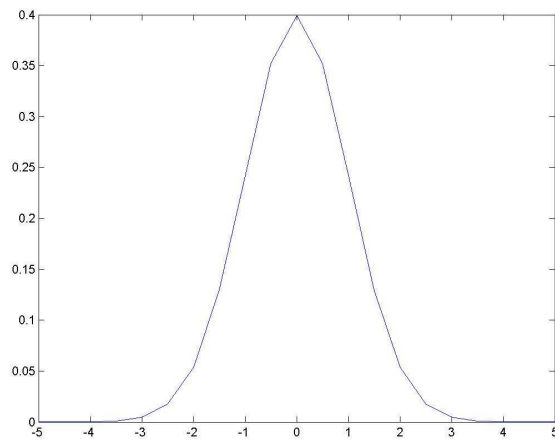
Parameters: $\beta > 0$
Domain: $0 \leq x < +\infty$
Density and Cumulative functions: $f(x) = \frac{e^{-x/\beta}}{\beta}$, $F(x) = 1 - e^{-x/\beta}$
Mean: β
Variance: β^2
Skewness: 2
Kurtosis: 9



PDF – exp(1).

Normal Distribution

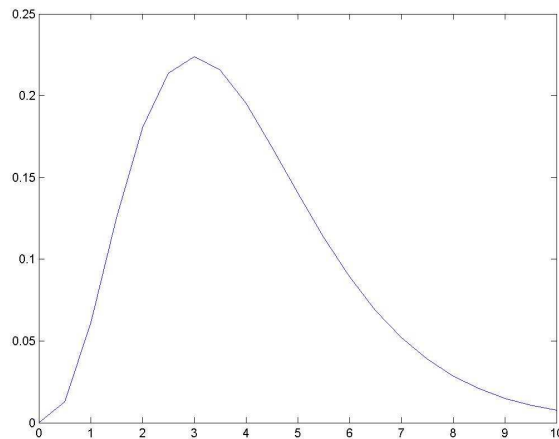
Parameters: $\mu, \sigma > 0$
Domain: $-\infty \leq x \leq +\infty$
Density and Cumulative functions: $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$, $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$
Mean: μ
Variance: σ^2
Skewness: 0
Kurtosis: 3



PDF – normal (0,1).

Gamma Distribution

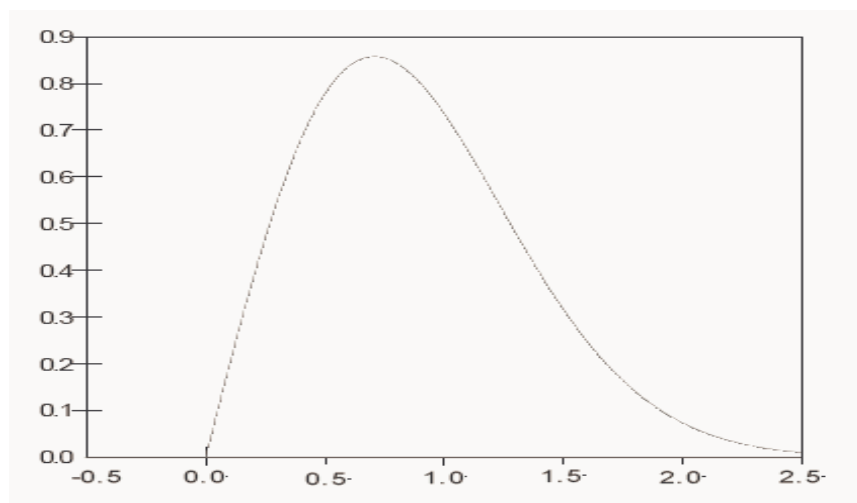
Parameters: $\alpha > 0, \beta > 0$
Domain: $0 < x < +\infty$
Density and Cumulative functions: $f(x) = \frac{1}{\beta\Gamma(\alpha)} \left(\frac{x}{\beta}\right)^{\alpha-1} e^{-x/\beta}$, $F(x) = \frac{\Gamma_{x/\beta}(\alpha)}{\Gamma(\alpha)}$
Mean: $\beta\alpha$
Variance: $\beta^2\alpha$
Skewness: $\frac{2}{\sqrt{\alpha}}$
Kurtosis: $3 + \frac{6}{\alpha}$



PDF – gamma(4,1).

Weibull Distribution

Parameters: $\alpha > 0, \beta > 0$
Domain: $0 \leq x < +\infty$
Density and Cumulative functions: $f(x) = \frac{\alpha x^{\alpha-1}}{\beta^\alpha} e^{-(x/\beta)^\alpha}$, $F(x) = 1 - e^{-(x/\beta)^\alpha}$
Mean: $\beta \Gamma(1 + \frac{1}{\alpha})$ where Γ is the Gamma Function
Variance: $\beta^2 \left[\Gamma(1 + \frac{2}{\alpha}) - \Gamma^2(1 + \frac{1}{\alpha}) \right]$ where Γ is the Gamma Function
Skewness: $\frac{\Gamma(1 + \frac{3}{\alpha}) + 3\Gamma(1 + \frac{2}{\alpha})\Gamma(1 + \frac{1}{\alpha}) + 2\Gamma^3(1 + \frac{1}{\alpha})}{\left[\Gamma(1 + \frac{2}{\alpha}) - \Gamma^2(1 + \frac{1}{\alpha}) \right]^{3/2}}$ where Γ is the Gamma Function
Function



PDF – weibull(2,1).

References

1. Arun A. Ross, Karthik Nandakumar, Anil K. Jain “Handbook of Multibiometrics”, Springer; 1 edition (May 24, 2006)
2. A. K. Jain and A. Ross, “Learning User-Specific Parameters in Multibiometric System”, Proc. of International Conference on Image Processing (ICIP), September 2002
3. M. Indovina, U. Uludag, R. Snelick, A. Mink and A. Jain, "Multimodal Biometric Authentication Methods: A COTS Approach", Proc. MMUA 2003, Workshop on Multimodal User Authentication, pp. 99-106, Santa Barbara, CA, December 11-12, 2003.
4. A. Ross and A. K. Jain, “Information Fusion in Biometrics”, Pattern Recognition Letters, Vol 24, Issue 13, September 2003, pps 2115–2125.
5. A. Jain, K. Nandakumar, A. Ross, “Score normalization in multimodal biometric systems” Proc. of Audio- and Video-based Biometric Person Authentication (AVBPA) 2005, pp. 1049-1058, Rye Brook, NY, July 2005.
6. A. K. Jain, A. Ross, and S. Prabhakar, “An Introduction to Biometric Recognition”, IEEE Transactions On Circuits And Systems For Video Technology, vol. 14, no. 1, pp. 4–21, January 2004,
7. S. C. Dass, K. Nandakumar, and A. K. Jain “A Principled Approach to Score Level Fusion in Multimodal Biometric Systems”, Proc. of Audio- and Video-based Biometric Person Authentication (AVBPA) 2005, pp. 1049-1058, Rye Brook, NY, July 2005.

8. National Institute of Standards and Technology: NIST Biometric Scores Set. Available at <http://www.itl.nist.gov/iad/894.03/biometricscores> (2004)
9. M. Thieme, "Multimodal Biometric Systems: Applications and Usage Scenarios", Biometric Consortium Conference 2003, Arlington, VA
10. R. M. Bolle, J. H. Connell, S. Pankanti, N. K. Ratha, A. W. Senior, "Guide to Biometrics", Springer, 2003
11. C. Beumier, "Facial surface identification", Available at <http://www.sic.rma.ac.be/beumier/3D/3d.html>.
12. A. K. Jain and A. Ross, "Multibiometric Systems", Communications of the ACM, Special Issue on Multimodal Interfaces, Vol. 47, No. 1, pp. 34-40, January 2004.
13. S. C. Dass, K. Nandakumar, A. K. Jain, "A Principled Approach to Score Level Fusion in Multimodal Biometric Systems", AVBPA 2005
14. A. K. Jain, R. Bolle and S. Pankanti, "Personal Identification in Networked Society", Kluwer Academic Publishers, 2002
15. L. Howells, "Fusion Comes in from the Cold", Technical paper, October 2005, Available at http://www.europeanbiometrics.info/images/resources/77_922_file.pdf
16. S. K. Dahel, Q. Xiao, "Accuracy Performance Analysis of Multimodal Biometrics", Proceedings of 2003 IEEE Workshop on Information Assurance, United States Military Academy, West Point, NY June 2003

17. M. C. Fairhurst, F. Deravi, J. George, "Towards Optimized Implementations of Multimodal Biometric Configurations", CMSPS2004 - IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety Venice, Italy, 21-22 July 2004
18. R. Snelick, Umut Uludag, Alan Mink, Michael Indovina, and Anil 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
19. R. Lazarick, "Multibiometric Techniques and Standards Activities", 39th Annual International Carnahan Conference on Security Technology, CCST '05, October 2005
20. N. Sedgwick, "The Need for Standardization of Multi-Modal Biometric Combination", Business/technical presentation, Cambridge Algorithmica Limited, Available at http://www.camalg.co.uk/s03017_pr0/pr0_040216a.pdf
21. R. Snelick, M. Indovina, J. Yen, A. Mink, "Multimodal Biometrics: Issues in Design and Testing", National Institute of Standards and Technology, *ICMI'03*, November 5-7, 2003, Vancouver, British Columbia, Canada.
22. N. Sedgwick, "Multi-Modal Combination using the Biometric Gain Concept", Business/technical presentation to the Biometrics 2004 Conference, Available at http://www.camalg.co.uk/s03017_pr1/s03017_pr1.pdf
23. L. Hong, A. Jain and S. Pankanti, "Can Multibiometrics Improve performance?", Proceedings AutoID'99, Summit, NJ, Oct 1999, PP. 59-64.

24. R. O. Duda, P. E. Hart, and D. G. StorkDuda, "Pattern Classification", John Wiley & Sons, 2001