

University of Southampton Research Repository

Copyright © and Moral Rights for this thesis and, where applicable, any accompanying data are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis and the accompanying data cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content of the thesis and accompanying research data (where applicable) must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holder/s.

When referring to this thesis and any accompanying data, full bibliographic details must be given, e.g.

Thesis: Author (Year of Submission) "Full thesis title", University of Southampton, name of the University Faculty or School or Department, PhD Thesis, pagination.

Data: Author (Year) Title. URI [dataset]

UNIVERSITY OF SOUTHAMPTON

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING

Electronics and Computer Science

Soft Biometric Fusion for Subject Recognition at a Distance

by

Bingchen Guo

Supervisors: Prof. Mark Nixon, Dr. John Carter

Internal examiner: Dr. Dasmahapatra

External examiner: Prof. Veldhuis

Thesis for the degree of Doctor of Philosophy

May 2018

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING

Electronics and Computer Science

Thesis for the degree of Doctor of Philosophy

Soft Biometric Fusion for Subject Recognition at a Distance

Bingchen Guo

Biometric recognition is an advanced technology that employs physical features (such as fingerprint, iris and face capture) and behavioural features (such as gait, signature and voice) to identify people. Biometric features are reliable and valid ways to describe the unique properties of individuals, but there are often rigorous requirements on the position and characteristics of devices used for data acquisition. Since biometric features can be difficult to capture at a distance, soft biometric features, such as height, weight, skin colour and gender, have received much attention. Although the uniqueness of soft biometric features is not as intuitively obvious as traditional biometric features, numerous experiments have demonstrated that the desired recognition accuracy can be achieved by using different soft biometric features. This thesis will propose state-of-the-art multimodal biometric fusion techniques to improve recognition performance of soft biometrics.

The first contribution of this thesis is to estimate fusion performance based on three types of soft biometrics - face, body and clothing. Feature level and score level fusion strategies will be employed to measure and analyse the influence of fusion on soft biometric recognition.

The second key contribution of this research is that the analysis of the influence of distance on soft biometric traits and an exploration of the potency

of recognition using fusion at varying distances have been performed. A new soft biometric database, containing images of the human face, body and clothing taken at three different distances, was created and used to obtain face, body and clothing attributes. First, this new database was constructed to explore the suitability of each modality at a distance: intuitively, the face is suitable for near field identification, and the body becomes optimal when the subject is further away. The new dataset is used to explore the potential of face, body and clothing for human recognition using fusion. In this section, some novel fusion techniques on different levels (feature, score and rank level) are proposed to improve soft biometric recognition performance.

A Supervised Generalised Canonical Correlation (SG-CCA) methodology is proposed to fuse the soft biometric features. The proposed SG-CCA is numerically validated to be the best fusion method compared with other multi-modal fusion methods. An SVM-weighted Likelihood Ratio Test (SVM-LRT) method is proposed for score level fusion. The experimental results demonstrate that SVM-LRT-based fusion significantly outperforms the single-mode recognition. A novel joint density distribution-based rank-score fusion is also proposed to combine rank and score information. Analysis using the new soft biometric database demonstrates that recognition performance is significantly improved by using the new methods over single modalities at different distances.

Table of Contents

Table of Contents.....	i
Table of Tables.....	v
Table of Figures.....	vii
Academic Thesis: Declaration of Authorship	xi
Acknowledgements.....	xiii
Chapter 1 Context and Contributions	1
1.1 Context	1
1.2 Contributions	3
1.3 Thesis outline	4
1.4 Publications	5
Chapter 2 Soft Biometrics	7
2.1 Existing soft biometrics datasets.....	7
2.1.1 Categorical dataset.....	7
2.1.2 Comparative dataset	13
2.2 Ranking inference	18
2.2.1 Elo rating system.....	19
2.2.2 Bradley-Terry ranking model	20
2.2.3 Results of comparisons	20
2.3 Analysis of single-mode recognition.....	22
2.3.1 Identification using the categorical datasets	22
2.3.2 Identification using the comparative datasets	23
2.4 Conclusions	25
Chapter 3 Soft Biometric Fusion	27
3.1 Fusion at feature level.....	27
3.1.1 Feature level fusion method	29
3.1.2 Feature level fusion experiments result.....	35

3.2	Fusion at score level.....	39
3.2.1	Similarity score calculation and normalization	39
3.2.2	Score-level fusion method.....	40
3.2.3	Score-level fusion experiment results	46
3.3	Conclusions	50
Chapter 4 Soft Biometric Dataset at Different Distances.....		51
4.1	Soft biometric dataset	51
4.1.1	Synthesising images	51
4.1.2	Soft biometric attributes and labels.....	55
4.1.3	Data acquisition via Crowdsourcing	58
4.2	Attributes Analysis	60
4.2.1	Ranking inference.....	60
4.2.2	Correlation analysis	60
4.2.3	Mutual information	62
4.3	Single-modal recognition.....	64
4.4	Conclusions	69
Chapter 5 Feature Level Fusion at Different Distances		71
5.1	Canonical Correlation Analysis	72
5.2	Generalized Canonical Correlation Analysis.....	73
5.3	Supervised Generalized Canonical Correlation Analysis	75
5.4	Experiments and analysis	77
5.4.1	Feature level fusion using supervised generalized canonical correlation analysis	77
5.4.2	Comparison with other linear dimensionality reduction method for fusion	78
5.5	Conclusions	79
Chapter 6 Score Level and Rank Level Fusion at Different Distances		81

6.1	Score level fusion.....	82
6.1.1	Estimation of similarity score densities	82
6.1.2	Score fusion using Bayesian theory	84
6.1.3	Score fusion using Likelihood Ratio Test.....	85
6.1.4	Score fusion using SVM-weighted Likelihood Ratio Test...	86
6.1.5	Experiments and discussion	88
6.2	Rank Level fusion.....	90
6.2.1	Borda count method.....	90
6.2.2	Logistic regression method	90
6.2.3	Nonlinear weight ranks method	90
6.2.4	PAV based rank fusion.....	91
6.2.5	Experiments and discussion	91
6.3	Conclusions	93
Chapter 7 Rank-score Fusion.....		95
7.1	Rank-score distribution	95
7.2	Normalization	97
7.3	Rank-score fusion	98
7.4	Experiment and discussion	99
7.4.1	Evaluation of rank-score fusion at three distances.....	99
7.4.2	Compared with single-modal recognition.....	100
7.4.3	Comparison with other fusion methods	101
7.5	Conclusions	102
Chapter 8 Conclusions and Future Work.....		105
8.1	Conclusions	105
8.2	Future work	107
8.2.1	Diversity of data collection	107
8.2.2	Automatic retrieval of biometric signatures.....	108
8.2.3	Descriptions from memory	108
References.....		109

Table of Tables

Table 2.1 Categorical body attributes and corresponding categorical labels..	8
Table 2.2 Categorical face attributes and corresponding categorical labels.	10
Table 2.3 Categorical clothing attributes and corresponding categorical labels	12
Table 2.4 Comparative body attributes and corresponding labels.	14
Table 2.5 Facial features used to compare subjects.....	16
Table 2.6 Clothing features used to compare subjects.	18
Table 2.7 Two rank list comparison (an example of Kendall tau distance calculation).....	21
Table 2.8 Difference between two rank lists (an example of Kendall Tau distance calculation).....	21
Table 3.1 Feature selected results by ANOVA.....	30
Table 3.2 Feature selected results by Pearson's r,	31
Table 3.3 Feature selected results by MI.	32
Table 3.4 Feature selected results by mRMR.	34
Table 3.5 Feature selected results by IFS.....	35
Table 3.6 Accuracies and sizes of five fusion methods (categorical dataset).....	36
Table 3.7 Accuracies and sizes of five feature selected methods (comparative dataset).....	37
Table 3.8 Accuracy comparison of simple average fusion.....	46
Table 4.1 Body traits and labels used to compare subjects.....	56

Table 4.2 Face attributes and corresponding categorical labels	57
Table 4.3 Clothing attributes and corresponding categorical labels	57
Table 4.4: Identification performance for single-modal methods	69
Table 5.1: Comparison of recognition results through SG-CCA with different numbers of features (d : number of features).....	77
Table 5.2 Identification performance using different methods (feature number=7).	79
Table 6.1: Identification performance using different methods	88
Table 6.2 <i>Gain</i> of different score fusion methods	89
Table 6.3 Identification performance using different rank-fusion methods ..	92
Table 6.4 <i>Gain</i> of different score fusion methods	92
Table 7.1 Accuracy rate using different normalisation methods.....	98
Table 7.2 Identification performance of signal-modal and rank-score fusion recognition	101
Table 7.3: Identification performance using different fusion methods.	102

Table of Figures

Figure 1.1 Single-mode biometric system.....	2
Figure 1.2 Multi-modal biometric system.	2
Figure 2.1 User gait annotation system interface [19].....	10
Figure 2.2 User interface of clothing label collection system	11
Figure 2.3 Interface bodily comparative label collection system.	15
Figure 2.4 Interface of facial comparative label collection system.....	16
Figure 2.5 Interface of clothing comparative label collection system.	18
Figure 2.6 Classification accuracy (up to rank 80) of soft categorical data when three modalities are used alone. Classification uses k NN (with $k=1$) and LoO for categorical body (Cat-body) categorical face (Cat-face) and categorical clothes (Cat-clothes).....	22
Figure 2.7 ROC performance of soft categorical traits. EER is calculated by investigating an equal number of False Positives and False Negatives, Categorical body (Cat-body) categorical face (Cat- face) and categorical clothes (Cat-clothes).	23
Figure 2.8 Recognition accuracy of each comparative dataset using Elo obtained from different numbers of comparisons. Accuracy was calculated using k NN and LoO for comparative body (Com- body) comparative face (Com-face) and comparative clothes (Com-clothes).....	24
Figure 2.9 ROC performance of soft comparative traits. EER was calculated by investigating an equal number of False Positives and False Negatives, comparative body (Com-body) comparative face (Com-face) and comparative clothes (Com-clothes).....	25
Figure 3.1 Feature level fusion.....	28
Figure 3.2 Feature level fusion flowchart.....	28

Figure 3.3 Comparison of recognition accuracy of five feature fusion methods (categorical dataset). Accuracy calculated using k NN (with $k=1$) and LoO classification tests.	36
Figure 3.4 Accuracy comparison of five feature selection methods (comparative dataset). Accuracy calculated using k NN (with $k=1$).	38
Figure 3.5 Score level fusion.	39
Figure 3.6 An example of simple average score fusion. Note that the match scores generated by the face and fingerprint matcher are similarity measurements. The range of match score is assumed to be $[0,1]$	41
Figure 3.7 An example of max-score fusion. Note that the match scores generated by the face and fingerprint matcher are similarity measurements. The range of match score is assumed to be $[0,1]$	42
Figure 3.8 Probability densities of comparative face and categorical body (estimated by a Parzen window), "Inter-class" and "intra-class": obtained according to the histogram; "PW-inter" and "PW-intra": estimated by a Parzen window.	44
Figure 3.9 Simple average fusion results. EER is calculated by investigating an equal number of False Positives and False Negatives, categorical face (Cat-face), comparative face (Com-face), simple average score fusion on categorical (Cat-Average Fusion) and comparative (Cat-Average Fusion).	47
Figure 3.10 Max score fusion results. EER calculated by finding an equal number of False Positives and False Negatives, categorical face (Cat-face), comparative face (Com-face), max-score fusion on categorical (Cat-max Fusion) and comparative (Com-max Fusion).	48
Figure 3.11 Bayesian fusion result, EER calculated by finding an equal number of False Positives and False Negatives, Bayesian score	

fusion on categorical (Cat-Bayes) and comparative (Com-Bayes)	49
Figure 4.1 Diagram of data acquisition environment.	52
Figure 4.2 Laboratory images after pre-processing.....	53
Figure 4.3 Image of the outdoor environment.	54
Figure 4.4 Synthesising images with outdoor environment.....	54
Figure 4.5 Examples of synthesising images at different distances.....	55
Figure 4.6 Interface of label collection system.....	59
Figure 4.7 Relationship between estimated and measured height	60
Figure 4.8 Pearson’s correlation coefficient for each trait in three groups (close and medium, close and far, medium and far).....	61
Figure 4.9 Mutual information of three datasets at three distances.	63
Figure 4.10 Body recognition accuracy obtained from different numbers of comparisons.	65
Figure 4.11 Face recognition accuracy obtained from different numbers of comparisons.	66
Figure 4.12 Cumulative match characteristic curves for the individual modalities	67
Figure 4.13 Accuracy of single-modal recognition (rank=1).	68
Figure 5.1 Recognition accuracy of the proposed SG-CCA algorithm using different number of features.	78
Figure 6.1 Probability density of face match score at three distance, estimated using a Parzen window.....	84
Figure 7.1 Overview of the Rank-Score distribution calculation framework. The notation sim_{X1}, S_i is used to denote the similarity score obtained by comparing an unknown subject S_i to the biometric sample 1 of a gallery subject in the face dataset.....	96

Figure 7.2 Fusion at three distances99

Figure 7.3 Accuracy for individual matcher (rank=1).....100

Academic Thesis: Declaration of Authorship

I, Bingchen Guo

declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

Fusion Analysis of Soft Biometrics for Recognition at a Distance

.....

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;

Signed:

Date:

Acknowledgements

Firstly, I would like to present my sincere gratitude to my supervisory team, Professor Mark Nixon, and Dr. John Carter. Under their considerate supervision, I managed to finish my Ph.D. work smoothly. In the past three years, I have benefitted greatly in terms of not only knowledge and skills but also research ability and academic temperament. Here, I want to give my sincere gratitude to these two knowledgeable, responsible and dependable supervisors. They are warm-hearted and care for my life in Southampton.

Secondly, I would give my gratitude to the staff in the Vision, Learning and Control Group as well as my colleagues and friends. They are Prof. Mahesan Niranjana, Dr. Bing Chu, Dr. Daniel Martinho-Corbishley, Mr. Thomas Ladyman. Moreover, I appreciate the fact that two internal examiners, Prof. Adam Prugel-Bennett and Dr. Srinandan Dasmahapatra, offered lots of constructive comments for my first year and upgrade exam.

Last but not least, I express my endless gratitude to my parents, who encouraged and financially supported me to go abroad and pursue my dreams.

Chapter 1 Context and Contributions

1.1 Context

Soft biometric recognition is a technique that uses semantic descriptions as features to identify subjects [1]. Soft biometric attributes include height, weight, gender and skin colour, which can be used to identify a person in practical applications [2]. In contrast, traditional biometrics overwhelmingly rely on sophisticated data collection devices. For example, facial recognition generally requires high image quality; however, image quality dramatically decreases as distance increases. Soft facial features, such as skin colour and face size, are relatively straightforward to perceive, even at a long distance. In comparison with traditional biometrics, soft biometric attributes are more easily understood.

Recognition performance using individual soft biometric datasets has been studied in previous research. A model using the human semantic description of soft biometrics to identify subjects was proposed in [3], where soft biometric features were used to enrich the recognition method. 19 body features were investigated in [4], and the results demonstrated that shoulder shape and arm length can aid recognition. 24 soft facial attributes are discussed in [5], and their performance is measured through analysis of variance, entropy and mutual information. Skin colour, eyebrow length and face length were demonstrated to be more reliable for use in recognition. In addition, 21 clothing features were reported in [6], demonstrating that clothing features can also be used for recognition (though clothing is innately short term as clothes can easily be changed). Furthermore, it was also demonstrated that head coverage, lower body clothing category and belt presence can greatly improve recognition.

Most traditional biometric features distinguish people by using their unique features, such as DNA and fingerprints, whilst soft biometric features are not so discriminative by their nature. Nonetheless, accurate recognition can be achieved by using multi-modal soft biometrics. Despite the research into multi-modal soft biometric recognition being at an early stage, some articles have reported results using the most advanced methods. A method proposed in [7] used soft biometric attributes to improve recognition performance of traditional biometrics. Prior work on soft biometrics at a distance is reported in [8], which demonstrated that the fusion of soft biometrics and traditional facial features could improve the performance of recognition based on a sparse representation. A fusion method proposed in [9] used soft biometrics (body, clothing and face) for identification at a distance. The results demonstrated that Bayesian fusion can greatly improve recognition performance.

In order to achieve more accurate recognition, recognition systems frequently employ multi-modal fusion. Fusion approaches are conventionally divided into five different levels: sensor, feature, score, rank and decision level [10]. Figure 1.1 and Figure 1.2 show block diagrams of recognition processes using single-mode and multi-modal biometrics.



Figure 1.1 Single-mode biometric system.

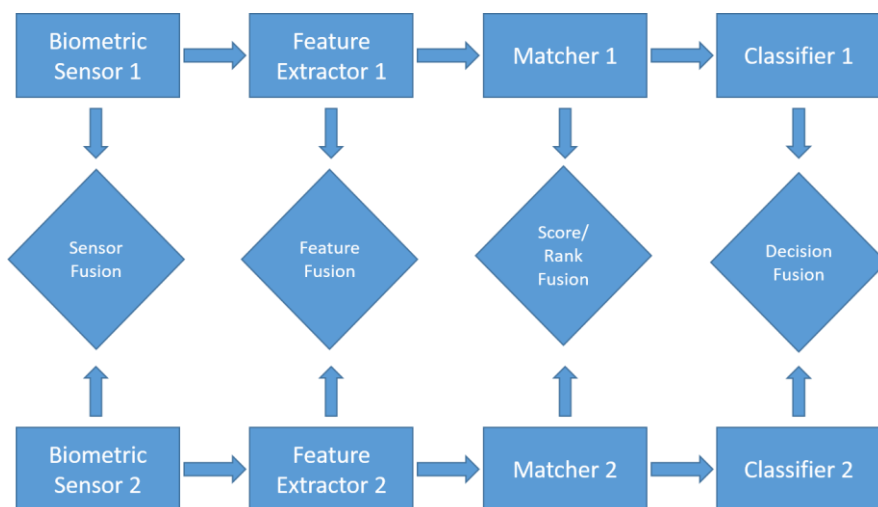


Figure 1.2 Multi-modal biometric system.

Feature level fusion based on feature extraction from multiple data sources is intended to create a new feature set to represent a subject. Therefore, the key requirement of the technology is to effectively describe feature information, in order to achieve the most accurate recognition. The general idea is to minimise the distance of feature information between intra-class samples, and maximise the distance between inter-class samples. Another important research field in the area of feature fusion focused on how to extract effective information by removing redundancy. Among a number of potential techniques used for feature extraction, linear feature extraction methods are widely used to reduce the dimensionality of the feature set. For example, a feature fusion method based on canonical correlation analysis (CCA) is introduced in [11]. Another feature level fusion technique, discriminant correlation analysis (DCA) [12], develops CCA by incorporating the class information into the correlation analysis of the feature sets. A multi-modal method, based on sparse representation, is proposed by Sumit, which significantly improved robustness and accuracy [13].

Score level fusion uses a combination of match scores from different biometric matchers, and then derives a new score from this information. Some simple methods, such as product rule, sum rule, max, medium and min rules, were introduced [14]. These methods can be readily implemented, since they do not require statistical information. Some score level fusion methods are based on the match score density distribution. A combinational method, using the Bayesian approach, was proposed in [14]. It estimated the genuine and impostor matching scores for each component modality. A support vector machine (SVM) based score level fusion was introduced and validated in [15]. The weighted score level fusion achieved a higher accuracy with the lower equal error rate (EER), compared with individual modalities [16].

1.2 Contributions

The main contributions of this research are:

1. To investigate human identification and verification performance by fusing different soft biometrics, provided by existing Southampton databases.

2. Creation of a new dataset of images, with controlled lighting and background at differing distances. A consistent set of labels for body, face and clothing at three distances is collected, and then used to investigate feature performance. The significance, stability and discriminative power of soft biometrics at different distances are analysed. Additionally, the contributions of attributes toward identification and verification are assessed.
3. An extended supervised generalised canonical correlation analysis (SG-CCA) method for soft biometric fusion is validated by three soft biometric datasets: body, face and clothing. The new dataset is first used for biometric fusion. The proposed method is not confined to biometrics, but generalised to use in pattern recognition.
4. An extended SVM-weighted likelihood ratio method for score level fusion. Results demonstrate the superiority of the fusion method in human recognition, when compared with single-modality.
5. A new rank and score level fusion method based on joint density distribution. Since rank is a linear description (i.e. 1, 2, 3, ...), it can be used to indicate the order of enrolled samples, but fails to describe the variations between adjacent samples. Thus, a novel technique is proposed to combine the effective information in rank and similarity scores, namely joint density-based rank-score fusion to consolidate the recognition result. The fusion effects over soft biometric characteristics and over distances are also analysed.

1.3 Thesis outline

This thesis is divided into two parts, corresponding to its two major contributions.

Part I focuses on the soft biometric fusion and validates the performance of soft biometric fusion using a Southampton dataset. The soft biometric database, including three datasets based on face, body and clothing, is introduced in Chapter 2. Each dataset is represented by categorical and comparative descriptions. It introduces the matching and recognition algorithms, and analyses recognition performance of different datasets. In Chapter 3, two fusion level methods (feature and score) are tested to

demonstrate the improvement of recognition performance using only soft biometrics.

Part 2 focuses on the new soft biometric dataset at different distances, and analyses the fusion preference at different distances. Chapter 4 introduces a new database of soft biometrics based on imagery collected using the University of Southampton Gait Tunnel. The images in the new database were synthesised in order to appear to be captured in an outdoor environment, and then labelled via human operators using the CrowdFlower system. The dataset's description, feature analysis, and the influence of the distance on soft biometric traits are all discussed in this chapter. Chapter 5 reviews the state-of-the-art techniques in feature-level fusion, and proposes a supervised generalised canonical correlation method to fuse soft biometric features. Chapter 6 introduces score level and rank level techniques, and the experiments are performed using a new soft biometric database. Chapter 7 presents a novel joint density distribution-based rank-score fusion strategy that combines rank and score information.

Finally, Chapter 8 summarises the results obtained so far and outlines the possible directions for future research.

1.4 Publications

1. Mark S. Nixon, Bingchen H. Guo, Sarah V. Stevenage, Emad S. Jaha, Nawaf Almodhahka, and Daniel Martinho-Corbishley. Towards automated eyewitness descriptions: describing the face, body and clothing for recognition. *Visual Cognition*, pages 1-15, 2016
2. Bingchen H. Guo, Mark S. Nixon, John N. Carter. Supervised generalized canonical correlation analysis of soft biometric fusion for recognition at a distance. *Proc. 8th International Conference on Imaging for Crime Prevention and Detection*, 2017
3. Bingchen H. Guo, Mark S. Nixon, John N. Carter. Fusion analysis of soft biometrics for recognition at a distance. *Proc. Identity, Security and Behaviour Analysis (ISBA)*, 2018 IEEE International Conference on, IEEE, 2018
4. Bingchen H. Guo, Mark S. Nixon, John N. Carter. A Joint Density Based Rank-Score Fusion for Soft Biometric Recognition at a Distance,

Accepted for 2018 *International Conference on Pattern Recognition (ICPR)*. IAPR/IEEE

5. Bingchen H. Guo, Mark S. Nixon, John N. Carter. Soft Biometric Fusion for Recognition at a Distance, *IEEE Trans on Biometrics, Behaviour and Identity Science* (to be submitted)

Chapter 2 **Soft Biometrics**

Soft biometrics use physical traits and behaviour characteristics that can be described using normal vocabulary. Soft biometric features include height, weight, hair length, arm length, skin colour, gender, race, and more. Compared with traditional biometrics, such as DNA, fingerprints and the iris, the features of soft biometrics are not unique. Nonetheless, when multiple features of soft biometrics are used for recognition at the same time, an accurate result can be obtained. In practical applications, the camera resolution significantly influences the performance of traditional biometric recognition, as CCTV may not provide detailed facial information. However, it remains straightforward to detect skin colour, gender and other soft biometric features [17]. Soft biometric traits have a clear advantage - that they are features that humans are more likely to use to describe another person. There is a gap between machines and people in identifying a person, and the features obtained from a computer algorithm are difficult to understand. However, soft biometrics bridge this gap [18].

This chapter introduces the database and the soft biometric features used in the experiments. The dataset includes categorical and comparative features for body, face and clothing datasets.

2.1 Existing soft biometrics datasets

2.1.1 Categorical dataset

Body categorical dataset

The body categorical dataset was originally collected by Samangooei [17]. The principles of trait selection are to choose features used in normal life. These

descriptions are relatively immutable, such as skin colour (except in rare cases, for example, if someone has had plastic surgery). Although it may be possible to change skin colour through a suntan, we are using skin colour to judge race. The chosen traits are easy to be collected and judged, especially from a remote distance, and the final traits used are listed in Table 2.1

Table 2.1 Categorical body attributes and corresponding categorical labels.

Body		Global	
Trait	Term	Trait	Term
0. Arm Length	(0.1) Very Short	12. Weight	(12.1) Very Thin
	(0.2) Short		(12.2) Thin
	(0.3) Average		(12.3) Average
	(0.4) Long		(12.4) Big
	(0.5) Very Long		(12.5) Very Big
1. Arm Thickness	(1.1) Very Thin	13. Age	(13.1) Infant
	(1.2) Thin		(13.2) Pre Adolescence
	(1.3) Average		(13.3) Adolescence
	(1.4) Thick		(13.4) Young Adult
	(1.5) Very Thick		(13.5) Adult
2. Chest	(2.1) Very Slim		(13.6) Middle Aged
	(2.2) Slim		(13.7) Senior
	(2.3) Average	14. Ethnicity	(14.1) European
	(2.4) Large		(14.2) Middle Eastern
	(2.5) Very Large		(14.3) Indian/Pakistan
3. Figure	(3.1) Very Small		(14.4) Far Eastern
	(3.2) Small		(14.5) Black
	(3.3) Average	(14.6) Mixed	
	(3.4) Large	(14.7) Other	
	(3.5) Very Large	15. Gender	(15.1) Female
4. Height	(4.1) Very Short		(15.2) Male
	(4.2) Short	Head	
	(4.3) Average	Trait	Term
	(4.4) Tall	16. Skin Colour	(16.1) White
	(4.5) Very Tall		(16.2) Tanned
5. Hips	(5.1) Very Narrow		(16.3) Oriental
	(5.2) Narrow		(16.4) Black
	(5.3) Average	17. Facial Hair Colour	(17.1) None
	(5.4) Broad		(17.2) Black

	(5.5) Very Broad		(17.3) Brown
6. Leg Length	(6.1) Very Short		(17.4) Red
	(6.2) Short		(17.5) Blond
	(6.3) Average		(17.6) Grey
	(6.4) Long	18. Facial Hair Length	(18.1) None
	(6.5) Very Long		(18.2) Stubble
7. Leg Direction	(7.1) Very Bowed		(18.3) Moustache
	(7.2) Bowed		(18.4) Goatee
	(7.3) Straight		(18.5) Full Beard
	(7.4) Knock Kneed	19. Hair Colour	(19.1) Black
	(7.5) Very Knock Kneed		(19.2) Brown
8 Leg Thickness	(8.1) Very Thin		(19.3) Red
	(8.2) Thin		(19.4) Blond
	(8.3) Average		(19.5) Grey
	(8.4) Thick		(19.6) Dyed
	(8.5) Very Thick	20. Hair Length	(20.1) None
9. Muscle Build	(9.1) Very Lean		(20.2) Shaven
	(9.2) Lean		(20.3) Short
	(9.3) Average		(20.4) Medium
	(9.4) Muscly		(20.5) Long
	(9.5) Very Muscly	21. Neck Length	(21.1) Very Short
10. Proportions	(10.1) Average		(21.1) Short
	(10.2) Unusual		(21.3) Average
11. Shoulder Shape	(11.1) Very Rounded		(21.4) Long
	(11.2) Rounded		(21.5) Very Long
	(11.3) Average	22. Neck Thickness	(22.1) Very Thin
	(11.4) Square		(22.2) Thin
	(11.5) Very Square		(22.3) Average
	(22.4) Thick		
		(22.5) Very Thick	

The body categorical dataset (Cat-body) describes 115 separate subjects. A Gait Annotation System (GAnn) [19] was used to make semantic annotations to each subject. Figure 2.1 shows an interface of GAnn.

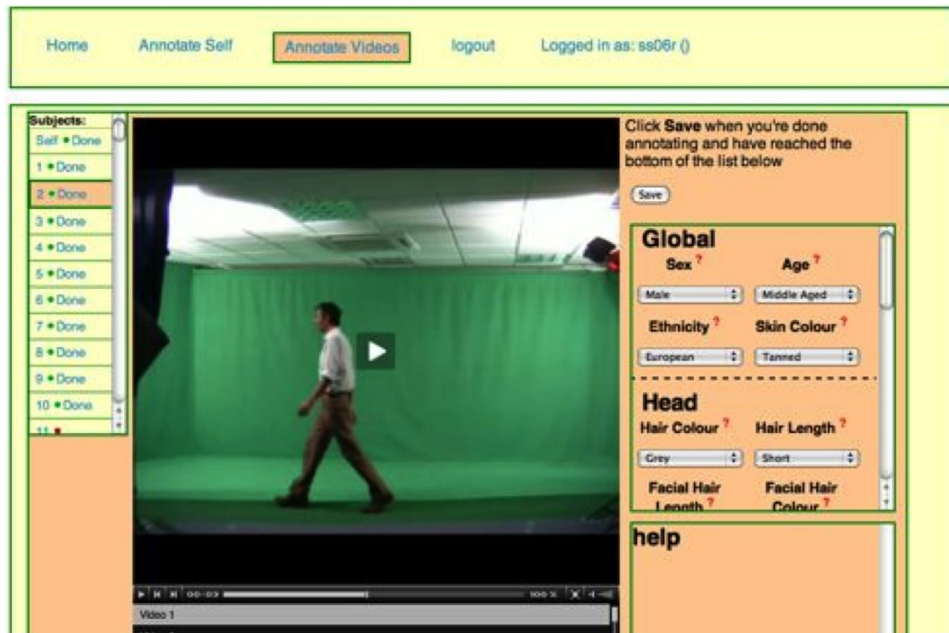


Figure 2.1 User gait annotation system interface [19].

Face categorical dataset

Table 2.2 Categorical face attributes and corresponding categorical labels.

Trait	Low Label	high label	Trait	Low Label	high label
Face	Short	Long	Eyebrows	Close Together	Far Apart
Face	Narrow	Wide	Eyebrows	Straight	Arched
Face	Bony	Fleshy	Eyes	Small	Large
Skin	Light	Dark	Eyes	Slanted	Round
Skin	Smooth	Wrinkled	Ears	Small	Large
Skin	Clear	Pimpled	Ears	Close to head	Sticking out
Hair	Short	Long	Ears	Hidden	Evident
Hair	Straight	Curly	Nose	Flat	Protruding
Hair	Thin	Thick	Nose	Short	Long
Forehead	Small	Large	Nose	Narrow	Wide
Forehead	Straight	Hairline	Nose	Upturned	Hooked
Eyebrows	Thin	Bushy	Lips	Lips Thin Thick	Thick
Eyebrows	Low	High	Chin and Jaw	Angular	Round
			Chin and Jaw	Receding	Protruding

The same approach was used to choose facial and clothing features. However, more details are used for the facial features, including the shape of the eyebrows, the length of the face, and more. Table 2.2 provides a full list of the facial features used.

Clothing categorical dataset

All subjects used for the face and clothing dataset (Cat-face and Cat-clothes) are the same as those for the body dataset. Each of the 115 individuals was described and labelled by multiple users [20]. The clothing features are listed in Table 2.3. The user interface developed to view attributes and obtain clothing labels is shown in Figure 2.2.

IN THIS TASK: You have labeled: 0 of 10 subjects

Please select an appropriate label for each (clothing/person) attribute to best describe the given subject.

NOTE: in all the given attributes, please describe what you see not what you infer. For example a rolled-up long sleeve is described based on its current situation of arm exposure to maybe (medium, or short).

FOR HINTS: MOVE YOUR MOUSE CURSOR OVER THIS SYMBOL ⓘ


Subject 014	Body part	Attribute	Annotation
	Head	Head clothing category ⓘ	<input type="radio"/> Cap <input type="radio"/> Mask <input type="radio"/> Scarf <input type="radio"/> Hat <input type="radio"/> None
		Head coverage ⓘ	<input type="radio"/> All <input type="radio"/> Most <input type="radio"/> Fair <input type="radio"/> Slight <input type="radio"/> None
		Face covered ⓘ	<input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Don't know
		Hat ⓘ	<input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Don't know
	Upper body	Upper body clothing category ⓘ	<input type="radio"/> Jacket <input type="radio"/> Jumper <input type="radio"/> T-shirt <input type="radio"/> Shirt <input type="radio"/> Blouse <input type="radio"/> Sweater <input type="radio"/> Coat <input type="radio"/> Other
		Neckline shape ⓘ	<input type="radio"/> Strapless <input type="radio"/> V-shape <input type="radio"/> Round <input type="radio"/> Shirt collar <input type="radio"/> Don't know
		Neckline size ⓘ	<input type="radio"/> Very small <input type="radio"/> Small <input type="radio"/> Medium <input type="radio"/> Large <input type="radio"/> Very Large
		Sleeve length ⓘ	<input type="radio"/> Very short <input type="radio"/> Short <input type="radio"/> Medium <input type="radio"/> Long <input type="radio"/> Very Long
	Lower body	Lower body clothing category ⓘ	<input type="radio"/> Trouser <input type="radio"/> Skirt <input type="radio"/> Dress
		Shape ⓘ	<input type="radio"/> Straight <input type="radio"/> Skinny <input type="radio"/> Wide <input type="radio"/> Tight <input type="radio"/> Loose
Leg length ⓘ		<input type="radio"/> Very short <input type="radio"/> Short <input type="radio"/> Medium <input type="radio"/> Long <input type="radio"/> Very Long	

Figure 2.2 User interface of clothing label collection system

Table 2.3 Categorical clothing attributes and corresponding categorical labels

Body zone	Semantic Attribute	Categorical Labels
Head	1. Head clothing category	[None, Hat, Scarf, Mask, Cap]
	2. Head coverage	[None, Slight, Fair, Most, All]
	3. Face covered	[Yes, No, Don't know]
	4. Hat	[Yes, No, Don't know]
Upper body	5. Upper body clothing category	[Jacket, Jumper, T-shirt, Shirt, Blouse, Sweater, Coat, Other]
	6. Neckline shape	[Strapless, V-shape, Round, Shirt collar, Don't know]
	7. Neckline size	[Very Small, Small, Medium, Large, Very Large]
	8. Sleeve length	[Very Short, Short, Medium, Long, Very Long]
Lower body	9. Lower body clothing category	[Trousers, Skirt, Dress]
	10. Shape	[Straight, Skinny, Wide, Tight, Loose]
	11. Leg length (of lower clothing)	[Very Short, Short, Medium, Long, Very Long]
	12. Belt presence	[Yes, No, Don't know]
Foot	13. Shoes category	[Heels, Flip ops, Boot, Trainer, Shoe]
	14. Heel level	[Flat/low, Medium, High, Very high]
Attached to body	15. Attached object category	[None, Bag, Gun, Object in hand, gloves]
	16. Bag (size)	[None, Side-bag, Cross-bag, Handbag, Backpack, Satchel]
	17. Gun	[Yes, No, Don't know]
	18. Object in hand	[Yes, No, Don't know]
	19. Gloves	[Yes, No, Don't know]
General style	20. Style category	[Well-dressed, Business, Sporty, Fashionable, Casual, Nerd, Hippy, Religious, Tramp, Other]
Permanent	21. Tattoos	[Yes, No, Don't know]

2.1.2 Comparative dataset

Another type of the features is labelled by comparing two objects using human descriptions. For example, it is easy to tell that one person is taller than another by observation, as it is easy to make a rough guess of height, but hard to estimate someone's exact height. The comparative dataset is used to study a more reliable description.

Comparative descriptions use features that are easily understood and annotated. The information for identification needs multiple comparisons between objects. Comparative descriptions can deliver more accurate descriptions.

The human categorical description was acquired based on subjective measurements, and varies with different people due to various standards. These standards are dictated by personal opinion, and are usually based on the understanding of the commentator of the population average value and the population difference, with this variation leading to diverse annotations. Comparative descriptions are generalised based on visual comparisons between two subjects, which makes the labels more consistent.

Each comparison describes the difference of each feature between two subjects, such as height, weight and the length of the arm. The comparison for each feature is labelled using three classes: shorter, the same or taller, based on the observation that a scale of 3 could lead to positive discrimination [5]. Each level is denoted by a signed integer, for example, when comparing height between two subjects, -1 means shorter, 0 represents the same and +1 means taller.

Body comparisons

Body and global features have the highest occurrence in the description of witnesses, including height, weight, race and gender. These features are obvious and easy to remember. In this study, the features were mainly based

on the work of Macleod [21], which provided the best physical features for the human description.

Some traits, such as the shape of the legs, have been removed because they are difficult to observe from a side view image. Some head features that can be only extracted from a minority of subjects were removed [4]. 16 out of 19 characteristics are compared, with every feature being labelled with three tags. Three characteristics are used as the absolute markers of gender, race and skin colour, since these three special traits are not suitable for comparisons, or lack a suitable standard to compare. However, these features provide useful information, so will not be removed. All the features used to describe the body are shown in Table 2.4.

Table 2.4 Comparative body attributes and corresponding labels.

Trait	Type	Labels
Arm Length	Comparative	Shorter, Same, Longer
Arm Thickness	Comparative	Thinner, Same, Thicker
Chest	Comparative	Smaller, Same, Bigger
Figure	Comparative	Smaller, Same, Larger
Height	Comparative	Shorter, Same, Taller
Hips	Comparative	Shorter, Same, Taller
Leg Length	Comparative	Shorter, Same, Longer
Leg Thickness	Comparative	Thinner, Same, Thicker
Muscle Build	Comparative	Leaner, Same, More Muscular
Shoulder Shape	Comparative	More Square, Same, More Rounded
Weight	Comparative	Thinner, Same, Fatter
Age	Comparative	Younger, Same, Older
Ethnicity	Absolute	European, Middle Eastern, Indian/Pakistan, Black, Mixed, Other
Gender	Absolute	Female, Male
Skin Colour	Absolute	White, Tanned, Oriental, Black
Hair Colour	Comparative	Lighter, Same, Darker

Hair Length	Comparative	Shorter, Same, Longer
Neck Length	Comparative	Shorter, Same, Longer
Neck Thickness	Comparative	Thinner, Same, Thicker



Figure 2.3 Interface bodily comparative label collection system.

Figure 2.3 shows the interface of the comparative label collection system for body traits. There are 572 sets of comparisons in the database. The comparisons are made using 80 sub-subjects and 20 targets. This meant 2 subjects' responses could be compared using the same target. In these circumstances, a new comparison can be inferred from those two comparisons.

Facial comparisons

Research in the field of psychology demonstrates that detailed descriptions of facial features by people are often wrong, and they are rarely used in the description of suspects, because people's vocabulary for facial features is deficient [22], and facial features are difficult to remember [23].

The comparing vision allows comparative labels to be used to describe facial features in a natural way. It needs meaningful words to describe the features and to avoid subjective labels [4]. This can improve the accuracy of the description. Although facial features described by witnesses are not used as often as body features, they play an important role in many criminal investigations.

Figure 2.4 provides an overview of the interface of facial comparative label collection. There are 292 sets of comparisons in the database. The following comparisons use 40 subjects and 10 targets.



Figure 2.4 Interface of facial comparative label collection system

The facial comparison was performed using 27 features, listed in Table 2.5. Each feature was labelled by one of three grades. The descriptions of each feature are listed in Table 2.5.

Table 2.5 Facial features used to compare subjects

Feature	Labels
Face	Shorter, Same, Longer
Face	Narrower, Same, Wider
Face	More Bony, Same, More Fleshy

Skin	Lighter, Same, Darker
Skin	Smoother, Same, More Wrinkles
Skin	Clearer, Same, More Pimples
Hair	Shorter, Same, Longer
Hair	Straighter, Same, Curlier
Hair	Thinner, Same, Thicker
Forehead	Smaller, Same, Larger
Forehead	Straighter Hairline, Same, More Receded Hairline
Eyebrows	Thinner, Same, Bushier
Eyebrows	Lower, Same, Higher
Eyebrows	Closer Together, Same, Further Apart
Eyebrows	Straighter, Same, More Arched
Eyes	Smaller, Same, Larger
Eyes	More Slanted, Same, Rounder
Ears	Smaller, Same, Larger
Ears	Close to Head, Same, Further from Head
Ears	More Hidden, Same, More Evident
Nose	Flatter, Same, More Protruding
Nose	Shorter, Same, Longer
Nose	Narrower, Same, Wider
Nose	More Uprturned, Same, More Hooked
Lips	Thinner, Same, Thicker
Chin and Jaw	More Angular, Same, More Round
Chin and Jaw	More Receding, Same, More Protruding

Clothing comparisons

Clothing comparison was made using 7 traits, as listed in Table 2.6. Each feature is divided into three grades. Figure 2.5 shows the interface of the clothing comparative label collection system. There were 317 sets of comparisons in the database.


Table 2.6 Clothing features used to compare subjects.

Body zone	Semantic Attribute	Labels
Head	Head coverage	Less, Same, More
	Face covered	Less, Same, More
Upper body	Neckline size	Smaller, Same, Larger
	Sleeve length	Shorter, Same, Longer
Lower body	Leg length	Shorter, Same, Longer
Foot	Heel level	Lower, Same, Higher
Attached to body	Bag (size)	Smaller, Same, Larger


You have done: 0 of 10 comparisons

IN THIS TASK:
Please compare subject A to subject B with selecting an appropriate comparative label for each (clothing/person) attribute.
NOTE: if a compare attribute is not available in both subjects set the degree of comparison to (SAME).
FOR HINTS: MOVE YOUR MOUSE CURSOR OVER THIS SYMBOL ⓘ

Subject A - (043)



Subject B - (017)



Body part	Attribute	Annotation
Head	Head coverage ⓘ	<input type="radio"/> Much less <input type="radio"/> Less <input type="radio"/> Same <input type="radio"/> More <input type="radio"/> Much more
	Face covered ⓘ	<input type="radio"/> Much less <input type="radio"/> Less <input type="radio"/> Same <input type="radio"/> More <input type="radio"/> Much more
Upper body	Neckline size ⓘ	<input type="radio"/> Much smaller <input type="radio"/> Smaller <input type="radio"/> Same <input type="radio"/> Larger <input type="radio"/> Much larger
	Sleeve length ⓘ	<input type="radio"/> Much shorter <input type="radio"/> Shorter <input type="radio"/> Same <input type="radio"/> Longer <input type="radio"/> Much longer
Lower body	Leg length (of lower clothing) ⓘ	<input type="radio"/> Much shorter <input type="radio"/> Shorter <input type="radio"/> Same <input type="radio"/> Longer <input type="radio"/> Much Longer
Shoes	Heel level ⓘ	<input type="radio"/> Much lower <input type="radio"/> Lower <input type="radio"/> Same <input type="radio"/> Higher <input type="radio"/> Much higher

Figure 2.5 Interface of clothing comparative label collection system.

2.2 Ranking inference

The comparison between two subjects was introduced as a more robust method for description. It was then considered to apply to the identification applications. Each subject was described using another subject as a benchmark. Comparative annotations need to be transformed to convey meaningful subject invariant information. The resulting value is defined as a

relative measurement. It can be used as a biometric feature for recognition. In essence, the rating method provides a relative measurement by comparison.

2.2.1 Elo rating system

The Elo rating system provides a ranking method based on Thurstone's case for comparative descriptions [24]. The Elo system was initially invented to quantify chess players' skill. Each chess player's capability cannot be measured directly, but is usually judged during chess games against other players. Playing chess is much like comparative labels. Relative measurements are made by comparing features, which is the same as comparing the skill of two chess players.

In the Elo rating system, the game was originally defined as a comparison between two players, A and B for chess games. For biometric identification, it is a visual comparison instead. The result of the comparison is the sample that indicates the difference between two players. The result is used to adjust the player's level.

$$Q_A = 10^{R_A/U} \quad (2.1)$$

$$Q_B = 10^{R_B/U} \quad (2.2)$$

$$E_A = \frac{Q_A}{Q_A + Q_B} \quad (2.3)$$

$$E_B = \frac{Q_B}{Q_A + Q_B} \quad (2.4)$$

$$R'_A = R_A + K(S_A - E_A) \quad (2.5)$$

$$R'_B = R_B + K(S_B - E_B) \quad (2.6)$$

This system uses the result S of the game to adjust the player's level. S_A is a compared result between A and B , S_B is the inverse of S_A . Different game results will update a player's level. When S equals 1, it means winning the game, 0.5 for a draw and 0 for losing. E is the mathematical expectation of the game's result, which can be calculated based on the player's level using Eq.(2.3) and (2.4). The adjusted difference is controlled by K , and K defines the maximum adjusted level value. A constant U reflects how the player's level

impacts the expectation, of which the value is chosen by experience; and set to 400.

2.2.2 Bradley-Terry ranking model

Bradley-Terry [25] [26] is a widely used ranking method. Supposing there is a competition between two players i and j , ($i, j \in \{1, \dots, K\}$), the probability that i defeats j is

$$P(i \text{ beats } j) = \frac{p_i}{p_i + p_j} \quad (2.7)$$

where p_i and p_j can be thought of as the ‘ability’ of each player. p_i and p_j are positive-valued parameters. There is a ‘contest’ between i and j , and ‘contests’ are comparisons. All ‘contests’ are independent. The parameters can be estimated by maximum likelihood.

Denoting τ_{ij} as the number of times that i beats j , then the negative log-likelihood takes the form

$$l(p) = -\sum_{i < j} \left(\tau_{ij} \log \frac{p_i}{p_i + p_j} + \tau_{ji} \log \frac{p_j}{p_i + p_j} \right) \quad (2.8)$$

$l(p)$ is scale-invariant and $l(p) = l(\alpha p)$ for any $\alpha > 0$. It is convenient to assume $\sum_{i=1}^K p_i = 1$, and then p_i is estimated by

$$\begin{aligned} p &= \operatorname{argmin} l(p) \\ \text{s. t. } & 0 \leq p_j, j = 1, \dots, K, \quad \sum_{i=1}^K p_j = 1 \end{aligned} \quad (2.9)$$

2.2.3 Results of comparisons

The principle of Kendall Tau rank distance [27] is to count the number of pairwise conformities between two ranking lists. The larger the distance is between them, the less similar the two ranks are.

Assuming that there are two rank lists, τ_1 and τ_2 , i is the element in τ_1 and τ_2 .

$$D(\tau_1, \tau_2) = \frac{|\{(i, j): i < j, (\tau_1(i) < \tau_1(j) \wedge \tau_2(i) > \tau_2(j)) \vee (\tau_1(i) > \tau_1(j) \wedge \tau_2(i) < \tau_2(j))\}|}{n(n-1)/2} \quad (2.10)$$

Eq.(2.10) is normalised by $\frac{n(n-1)}{2}$. $D(\tau_1, \tau_2)$ is equal to 0 if two ranks are identical, or equal to 1 if they are inverse to each other.

Supposing a group of five people are ranked by their marks of physics and mathematics examinations, respectively. Person A has highest mark in physics and the third highest mark in mathematics.

Table 2.7 Two rank list comparison (an example of Kendall tau distance calculation).

Person	A	B	C	D	E
Rank by physics	1	2	3	4	5
Rank by mathematics	3	1	4	2	5

In order to calculate the Kendall Tau distance, each person is compared with each other person, counting the number of pairwise in rank 1 that is opposite to rank 2.

Table 2.8 Difference between two rank lists (an example of Kendall Tau distance calculation).

Pair	(A,B)	(A,C)	(A,D)	(A,E)	(B,C)	(B,D)	(B,E)	(C,D)	(C,E)	(D,E)
Physics	1<2	1<3	1<4	1<5	2<3	2<4	2<5	3<4	3<5	4<5
Math	3>1	3<4	3>2	3<5	1<4	1<2	1<5	4>2	4<5	2<5
Count	×		×					×		

The normalised Kendall Tau distance is:

$$D = \frac{3}{5(5-1)/2} = 0.3 \quad (2.11)$$

The Elo rating system and Bradley-Terry ranking model were used to rate the comparative datasets, and Kendall tau distance is used to measure the distance of feature ranking results between ranks in the Elo rating system and Bradley-Terry ranking model. The maximum distance is 0.23, so the rank results of those two methods are similar. As such, the Elo system will be used in the remainder of this thesis.

2.3 Analysis of single-mode recognition

2.3.1 Identification using the categorical datasets

A Leave-one-Out validation (LoO) approach was used to validate the classification ability of categorical data. Each test included a group of LoO classifications, and each of the datasets (Cat-body, Cat-face and Cat-clothes) was independently used for human identification. In the test, k -nearest neighbours (k NN) was used for matching the target with $k = 1$. The EER for each test was calculated through a ROC curve.

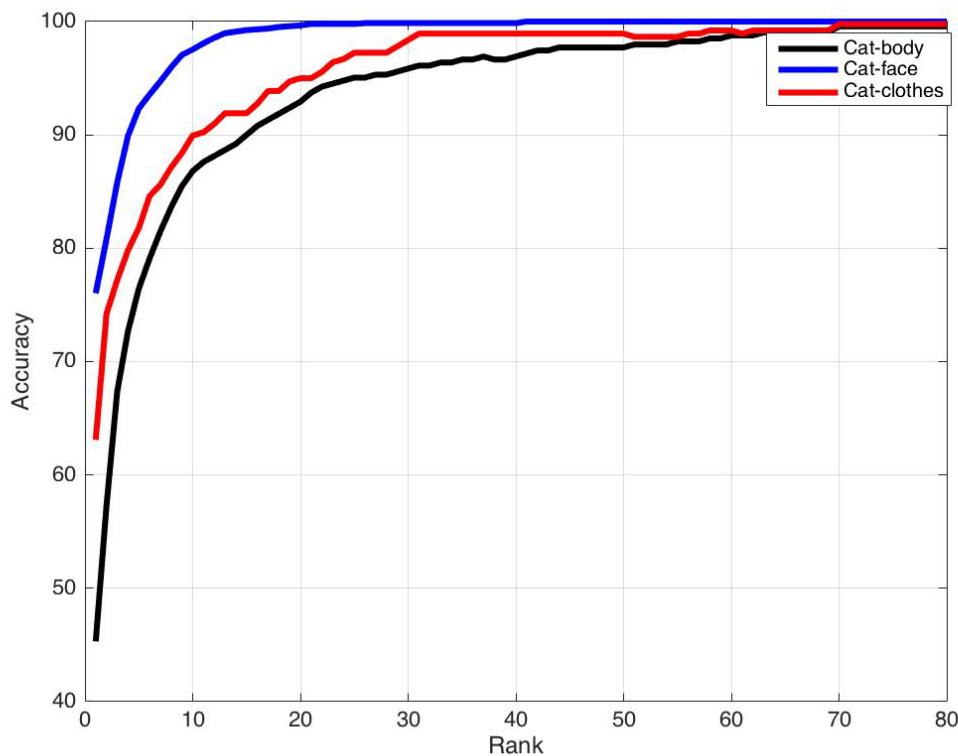


Figure 2.6 Classification accuracy (up to rank 80) of soft categorical data when three modalities are used alone. Classification uses k NN (with $k=1$) and LoO for categorical body (Cat-body) categorical face (Cat-face) and categorical clothes (Cat-clothes).

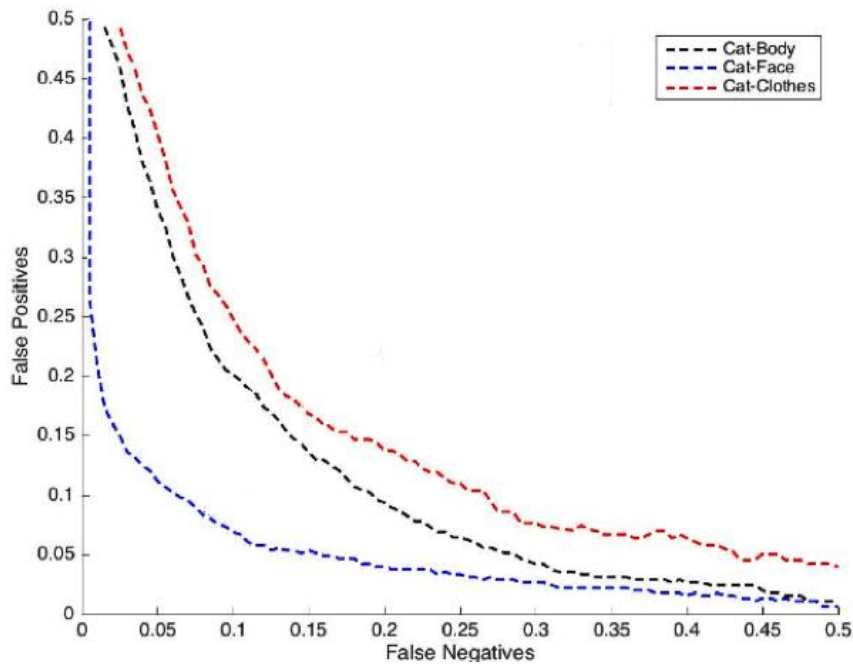


Figure 2.7 ROC performance of soft categorical traits. EER is calculated by investigating an equal number of False Positives and False Negatives, Categorical body (Cat-body) categorical face (Cat-face) and categorical clothes (Cat-clothes).

Figure 2.6 and Figure 2.7 show the classification accuracy and ROC curve results for the three categorical datasets. The different traits vary in each performance curve. It shows that the accuracy of recognition increases gradually with the growth of the rank, and that the face appears the most potent for identification.

2.3.2 Identification using the comparative datasets

The comparative data cannot be used directly. The Elo rating system was therefore used to calculate a relative score for each subject. The feature vector used to make classifications was based on the Elo score.

The recognition experiment is intended to retrieve the subject from the 40 existing subjects in the database. All comparative features were considered here. kNN was used to match subjects, where $k=1$ and the LoO mode was used to conduct tests. The results, in terms of recognition accuracy, vary with the number of comparative features, which are shown in Figure 2.8. In addition, the comparative facial dataset achieves a smooth curve first, which means

fewer comparative factors are needed to achieve higher classification accuracy. The comparative clothing only reaches 35% when in Figure 2.6 it can reach 100%. This is similar to the performance observed by Jaha [28], where the accuracy using categorical traits is much better than comparative traits.

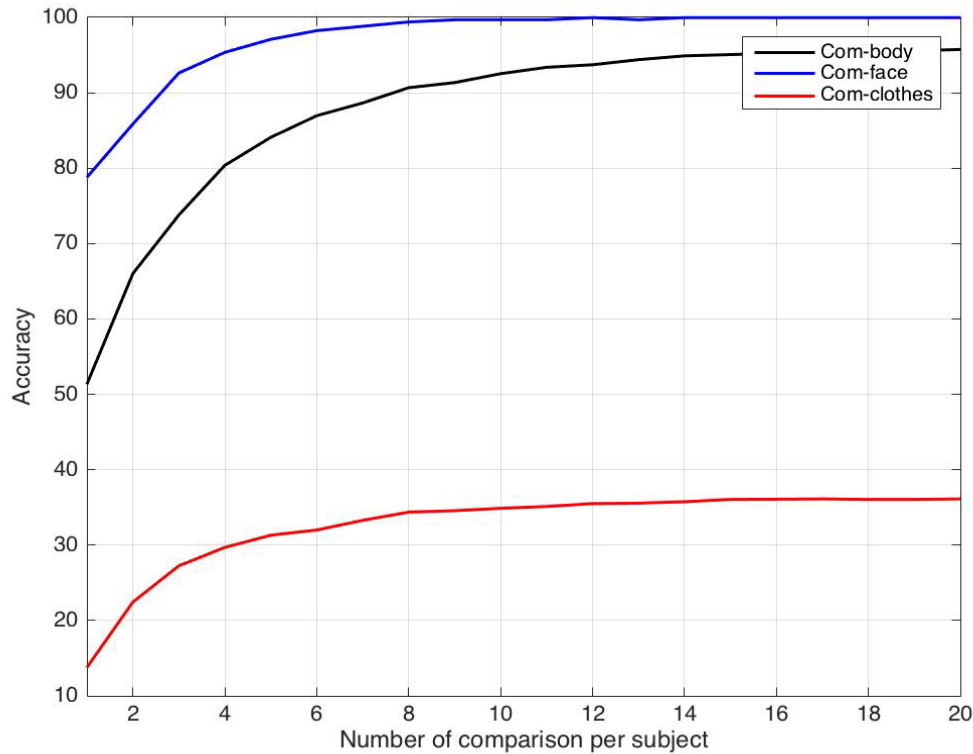


Figure 2.8 Recognition accuracy of each comparative dataset using Elo obtained from different numbers of comparisons. Accuracy was calculated using *k*NN and LoO for comparative body (Com-body) comparative face (Com-face) and comparative clothes (Com-clothes)

When analysing the comparative data, 5 comparisons are randomly selected for each subject. ROC curve was then calculated for each dataset. Figure 2.9 shows the ROC curve results from the three datasets. In the three single-mode biometric features (body, face and clothes), the facial verification performance is the highest on both categorical and comparative labels.

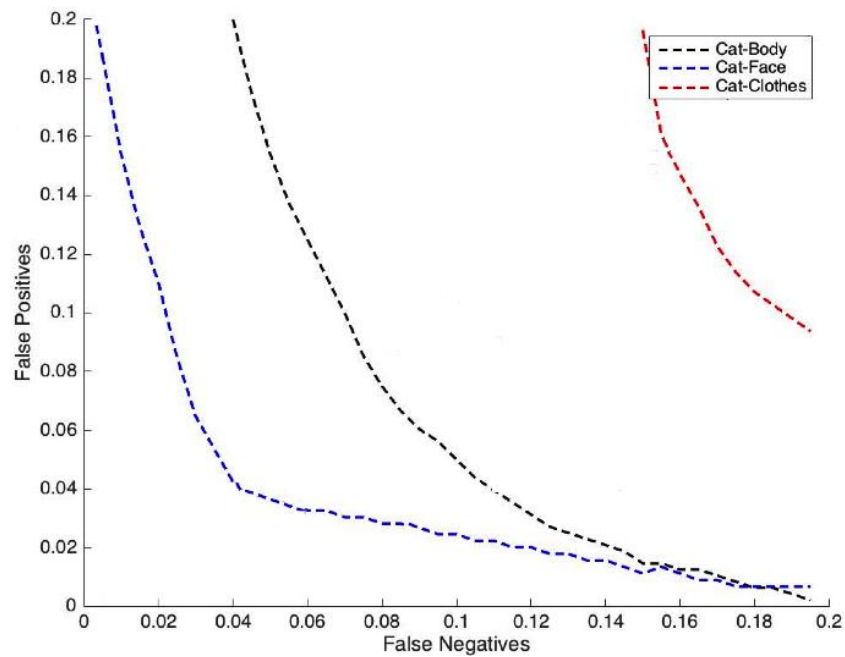


Figure 2.9 ROC performance of soft comparative traits. EER was calculated by investigating an equal number of False Positives and False Negatives, comparative body (Com-body) comparative face (Com-face) and comparative clothes (Com-clothes)

2.4 Conclusions

This chapter provided an overview of soft biometrics and highlighted their important role in subject recognition. These modalities were first analysed together. Previous analyses of the soft biometric datasets were implemented by different researchers, and using different ranking approaches and classifiers. Here, the same ranking method and classification was applied across all datasets.

First, it introduced the definition of soft biometrics, and outlined several sets of soft biometric databases. Soft biometric features are physical traits and human behaviour characteristics, which are labelled using normal vocabulary. The definition and selection of semantic features was then specified. The categorical datasets based on the body, face and clothes were introduced and then analysed. The recognition results demonstrated that categorical features have acceptable recognition capabilities. In order to improve the accuracy of recognition, another type of soft biometric feature, the comparative feature,

was introduced. The accuracy was improved when comparative features were used in recognition. The recognition accuracy based on the face was highest on both categorical and comparative labels.

The analysis showed differing performance for the approaches when used in isolation and in a consistent way. It seems prudent to investigate the fusion of the data, so that differing properties of the features might be taken advantage of in different scenarios.

Chapter 3 **Soft Biometric Fusion**

Fusion approaches are conventionally divided into five different levels: sensor, feature, score, rank and decision level [10].

Sensor layer fusion has a large volume of raw data from sensors, which are sensitive to the data transmission environment. The sensor-level data is always acquired with poor stability, and gives rise to unsatisfactory performance [29]. The soft biometrics dataset described in the previous chapter was labelled by humans (sensor), but the sensor level fusion cannot be performed here. The feature layer needs to be fused in the stage of feature extraction. Extracting useful information from different biometric models is a basic idea of feature level fusion. Feature layer fusion is designed to achieve better recognition results by maximising the discriminative performance of various features [29] [30] [31]. The fundamental goal of both match score level and decision level fusion is to fuse the results (match score or identify results) of different biometric modes. Calculating weights for different modes is a popular method. This chapter will test recognition performance of soft biometric fusion at the feature and the score level.

3.1 Fusion at feature level

Features were extracted from the raw data acquired by sensors, and were then subject to comprehensive analysis for pre-processing. Feature layer fusion was proposed to fuse features after their extraction. This fusion belongs to the middle level, implementing objective compression of information, which is advantageous to the real-time processing. The extracted features are directly related to the labels of objects. The feature fusion results can, therefore, give the most feature information needed by classification analysis, and, in theory, this will achieve the best level of recognition [30] [32].

Figure 3.1 shows the block diagram of a feature level fusion. At present, the main methods of feature layer fusion are serial strategy [33], parallel strategy [31] and weighted stack [34].

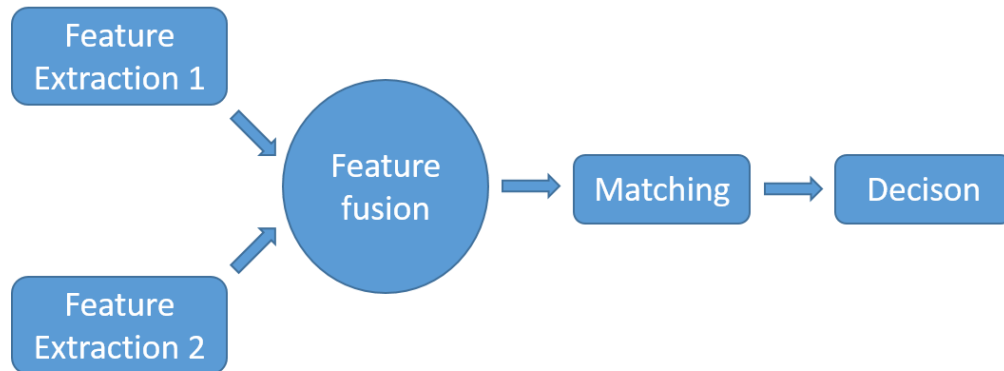


Figure 3.1 Feature level fusion.

The method used in the experiments was named as serial strategy, which means to install the features of the body, face and clothes in a series of separate feature matrices. Some features have only a small influence on the classification result; therefore, a robust feature selection process should be applied to select the most relevant features for classification, so as to achieve the approximate or a better result of the classification task before feature selection [29].

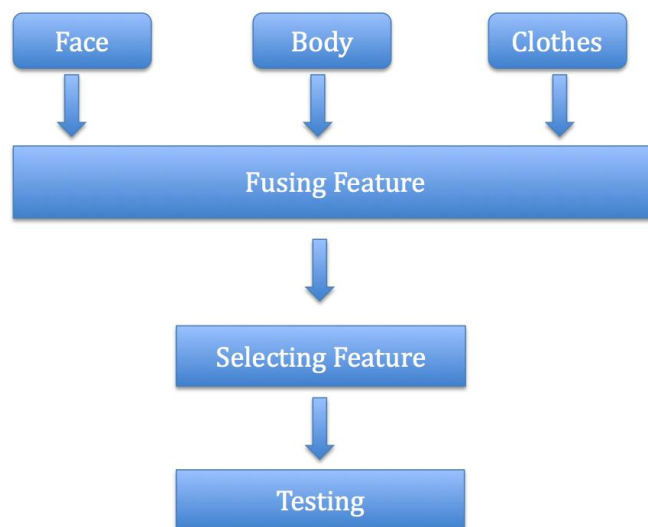


Figure 3.2 Feature level fusion flowchart.

Figure 3.2 shows the flowchart of feature level fusion in the experiment. Feature selection techniques can be divided into three classes: filter methods, which are only based on properties of features but ignore learning, like the Mutual Information; wrappers, which score a set of features using classifiers, for example, decision-tree-based wrapper genetic [35]; embedded methods, which inject the selection process into the learning of the classifier, as Minimum Redundancy Maximum Relevance (mRMR) describes in [36]. A feature will be chosen if it can maximise the mutual Information between the feature and the subject's label while the mutual Information between the selected feature and the subset of the selected features is minimal. Here there are five methods of feature selection: analysis of variance (ANOVA), Pearson's r , mutual Information, mRMR and infinite feature selection [37]. Those five methods are involved in filter, wrappers and embedded classes.

3.1.1 Feature level fusion method

3.1.1.1 ANOVA

In statistics, it is necessary to define the importance of different factors. One way to achieve this is the ANOVA, which is a type of method that can measure the importance of the single variable [38]. It can also be used to calculate the F ratio as:

$$F = \frac{B}{W} = \frac{\sum_i n_i (\bar{x}_i - \bar{x})^2 / (K-1)}{\sum_{ij} (x_{ij} - \bar{x}_i)^2 / (N-K)} \quad (3.1)$$

where B and W are total between-group variance and total within-group variance, respectively. x_{ij} represents the j^{th} sample in the i^{th} group. Similarly, \bar{x}_i represents the mean value of samples in the i^{th} group. \bar{x} represents the mean value of all samples. K refers to the number of groups, and N refers to the total number of samples. Therefore, the F ratio is the ratio of intra-class variance to inter-class variance. Degrees of freedom $K - 1$ and $N - K$ are used to weight the calculations. In order to separate different factors, the intra-class variance is as small as possible, whilst the inter-class variance is as large as possible. F-ratio increases when the inter-class variance is bigger than the intra-class variance. It was identified by calculating the F-ratio of each feature.

The features are more important (as measured using ANOVA) when the F ratio increases.

When we use ANOVA to select features, the F value of each feature was calculated and the results were sorted in descending order. A feature is selected first if it has the maximum F . Table 3.1 shows first five features selected by ANOVA. It is demonstrated that the face features have superior performance in both the categorical and the comparative datasets.

Table 3.1 Feature selected results by ANOVA.

a) Categorical database

	Feature (example labels)	Dataset
1	Hair (short/long)	Face
2	Face (bony/fleshy)	Face
3	Face (narrow/wide)	Face
4	Chin and Jaw (angular/round)	Face
5	Nose (flat/protruding)	Face

b) Comparative database

	Feature (example labels)	Dataset
1	Hair (shorter/longer)	Face
2	Ear (smaller/larger)	Face
3	Forehead (smaller/larger)	Face
4	Ear (more hidden/more evident)	Face
5	Eyebrows (closer together/further apart)	Face

3.1.1.2 Pearson's r

Eq. (3.2) shows the Pearson's r correlation.

$$r = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3.2)$$

X and Y represent two sets of features. x_i and y_i are i^{th} annotations in these two feature sets. Each group of features has n annotations. σ_{XY} is the

covariance of two variables X and Y , σ_X and σ_Y are, respectively, the standard deviations of those two variables.

In computation of correlation, the annotators of each object are randomly assigned to two groups, whose descriptions are averaged. Each object will produce two sets of descriptions. They produce 100 random allocations. This can be used to interpret the most correlated features by calculating the correlation coefficient of each semantic feature given by these random groups, to determine the most relevant one. After calculating the Pearson value of each feature, they are sorted in descending order, and the feature will be selected first that has the highest correlation.

Table 3.2 shows the first five features selected by Pearson's r . Compared with the first five results of ANOVA, there are two clothing features in the categorical labels, and one body feature in the comparative labels for Pearson's r correlation. However, compared with the whole order of effective features, their results are very similar.

Table 3.2 Feature selected results by Pearson's r ,

a) Categorical database

	Feature (example labels)	Dataset
1	Hair (short/long)	Face
2	Face (narrow/wide)	Face
3	Head cover (slight/all)	Clothing
4	Face (bony/fleshy)	Face
5	Sleeve length (Short/long)	Clothing

b) Comparative database

	Feature (example labels)	Dataset
1	Hair (shorter/longer)	Face
2	Height (shorter/taller)	Body
3	Ear (smaller/larger)	Face
4	Ears (more hidden/more evident)	Face
5	Forehead(smaller/larger)	Face

3.1.1.3 Mutual Information

Feature selection based on mutual information is described in [39]. Assuming that there are two discrete random variables X and Y with marginal probability distribution functions $p(x)$, $p(y)$ and the joint probability distribution function of $p(x, y)$, the mutual Information of X and Y is calculated by Eq.(3.3):

$$I(x, y) = \sum_{i,j} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)} \quad (3.3)$$

Mutual information can describe the relationship of the selected features and the output labels. The features with the most contribution to classification can be selected by calculating the mutual information with subjects' labels. The feature that has the maximum mutual information for a subject's labels will be selected first and will be added to the selected feature subset.

Table 3.3 Feature selected results by mutual Information.

a) Categorical database

	Feature (example labels)	Dataset
1	Hair (short/long)	Face
2	Face (short/long)	Face
3	Eyebrows (close together/far apart)	Face
4	Head cover (slight/all)	Clothing
5	Lower body clothing category (trouser, skirt, dress)	Clothing

b) Comparative database

	Feature (example labels)	Dataset
1	Hair (shorter/longer)	Face
2	Ear (smaller/larger)	Face
3	Forehead (smaller/larger)	Face
4	Ear(more hidden/more evident)	Face
5	Eyebrows (closer together/further apart)	Face

Table 3.3 shows the first five features selected by mutual Information. Those three methods provided the same result on hair (short/long), making it the feature that has the best recognition performance. The capability of the face is better than body and clothing, because there is more detail in facial features.

3.1.1.4 Minimal Redundancy Maximal Relevance

The above three feature selection algorithms only consider the degree of relevance between the features and categories, but do not consider the redundancy between the features. Minimal Redundancy Maximal Relevance (mRMR) combines the feature relevance with redundancy [40]. The feature subsets are the optimal feature subsets for filtering the redundant information.

In the analysis of feature relevance and redundancy, the selected measurement tool for relevance is very important for mRMR. In this experiment, mutual information is employed as the relevance measurement factor, and the mutual information can be calculated through Eq.(3.3):

Using the maximum relevance to measure the relationship between features and subjects' labels requires maximum correlations between them, specifically, the maximum mutual information of the features and the targets [36]. Minimum redundancy is a description of the dependent relationship of the features, which requires the minimum relevance of each feature attribute. The minimum redundancy is achieved by minimising mutual information among the features. The key concept of the mRMR algorithm is to combine the selection criteria of the maximum relevance between the features, and categorise these with the selection criteria of the minimum redundancy among the features.

In mRMR, the first important feature can be selected with maximum mutual information, because it can further reduce the uncertainty of other features in the feature sets. In other words, the feature providing the most information to the recognition system is selected first. The later features can be selected according to the formula Eq. (3.4).

$$f_{m+1} = \arg \max \left\{ I(f_i, c) - \frac{1}{m} \sum_{f_t \in S} I(f_i, f_t) \right\} \quad (3.4)$$

where C is the subjects' label information and S is the selected feature subset. f_t is the feature existed in S , f_i is the remaining feature.

Table 3.4 shows first five features selected by mRMR. Facial features still have the highest descriptive power. For comparative datasets, the features of body and face are better than clothing. As the results show in Chapter 2, the recognition result provided by comparative clothing datasets is not as good as that of the face and body.

Table 3.4 Feature selected results by mRMR.

a) Categorical database

	Feature (example labels)	Dataset
1	Hair (short/long)	Face
2	Head cover (slight/all)	Clothing
3	Chin and Jaw (angular/round)	Face
4	Nose (short/long)	Face
5	Face (bony/eshy)	Face

b) Comparative database

	Feature (example labels)	Dataset
1	Ears (smaller/larger)	Face
2	Weight (thinner/fatter)	Body
3	Hair Colour (lighter/darker)	Body
4	Hair (shorter/longer)	Face
5	Skin (clearer/more pimples)	Face

3.1.1.5 Infinite feature selection

In paper [37], a filter-based feature selection algorithm, called infinite feature selection (IFS) is proposed. It performs the feature sorting in an unsupervised method, and selects the best m features using a cross-validation strategy. This algorithm assumes that each feature is a node in the graph, and weights are given by the mixture of Spearman's rank correlation coefficients and standard deviations between feature distributions. A path over the graph is seen as a possible feature selection. An integral path process is then applied. For the result, it evaluates a single feature energy score for each feature while

considering all the possible subsets of features as paths on a graph. The higher the final score is, the more important the feature is.

Table 3.5 shows the first five features selected by IFS. For the categorical database, hair length is still the most stable feature, and a body feature first occurs in the categorical database.

Table 3.5 Feature selected results by IFS.

a) Categorical database

	Feature (example labels)	Dataset
1	Hair (short/long)	Face
2	Skin (clear/pimpled)	Face
3	Ethnicity (European/far eastern)	Face
4	Belt presence (yes/no/don` t know)	Clothing
5	Arm length (short/long)	Body

b) Comparative database

	Feature (example labels)	Dataset
1	Ears (smaller/larger)	Face
2	Height (shorter/taller)	Body
3	Chest (smaller/larger)	Body
4	Forehead (straighter hairline/more receded hairline)	Face
5	Face (narrower/wider)	Face

3.1.2 Feature level fusion experiments result

The method used in the experiments is a serial strategy, which intends to concatenate the features of the body, face and clothes in a single feature matrix. Tests were conducted using five methods (ANOVA, Pearson's r , mutual Information, mRMR and IFS) on the categorical and comparative datasets.

3.1.2.1 Fusion on categorical dataset result

For the categorical dataset, there are 22 features in the body dataset, 27 features in the face dataset and 21 features in the clothing dataset. After concatenation, the most effective features are chosen from the 80 features

using five methods. A test is conducted with classification to compare the performance of the five feature selection methods.

Figure 3.3 shows that the classification accuracy increases gradually with the growth of the feature subset size, and levels off at the end.

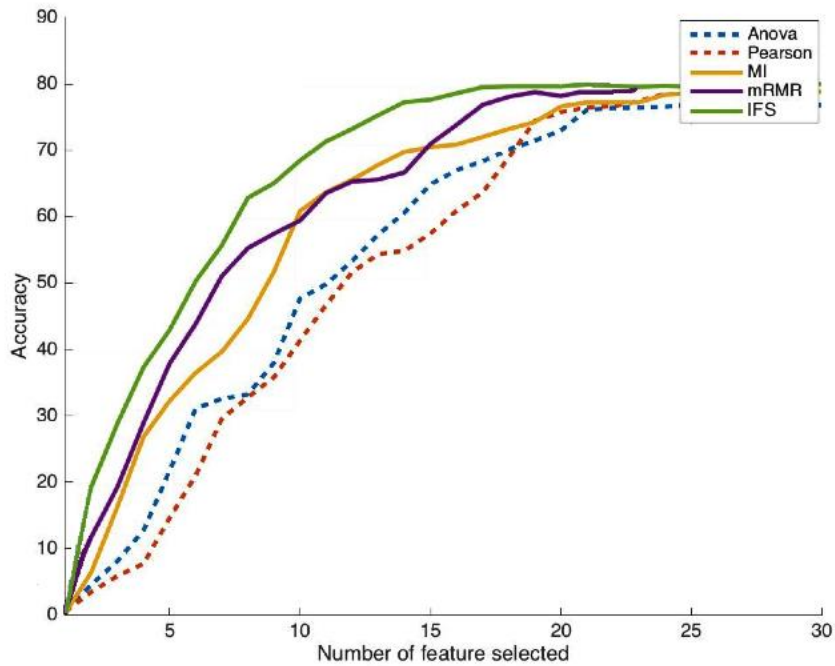


Figure 3.3 Comparison of recognition accuracy of five feature fusion methods (categorical dataset). Accuracy calculated using k NN (with $k=1$) and LoO classification tests.

Table 3.6 Accuracies and sizes of five fusion methods (categorical dataset).

Method	Accuracy exceeding 80% (number of features)	Feature Number=			Min EER
		5	15	30	
ANOVA	30	12.57%	63.32%	77.83%	0.10
Pearson's r	29	11.90%	58.27%	80.02%	0.09
MI	25	39.04%	69.82%	80.62%	0.06
mRMR	23	40.97%	70.34%	81.12%	0.06
IFS	17	45.13%	77.15%	81.57%	0.03

Figure 3.3 shows the classification accuracy gained by the five feature selection algorithms varying with the feature subset dimensions. Different selection algorithms lead to different performances. Figure 3.3 also demonstrates that, with the expansion of feature subsets, the classification accuracy becomes more stable. When 15 features were selected from the feature subsets, the classification accuracies using mRMR, ANOVA and IFS algorithms were 70.34%, 63.32% and 77.15% respectively.

In addition, Figure 3.3 shows that the IFS algorithm achieved stability first. In other words, it achieves higher classification accuracy with smaller feature subsets. By analysing the data in Table 3.6, the number of features (when accuracy > 80%) indicates the minimum number of features for each feature selection algorithm needed to obtain stable and desirable classification performance. The ratio shows the number of the selected feature subsets and the sum of the sizes of the original feature sets. It is clear that IFS consistently obtains the best performance, the EER is minimised and the accuracy is over 80%, when using 17 features. Compared with ANOVA and Pearson's *r*, the advantages of IFS are more pronounced when only using 5 features.

3.1.2.2 Fusion on comparative dataset

In the feature fusion tests, there were 16, 27 and 7 features for the body, face and clothing respectively. After feature fusion, there were 50 features in total.

Table 3.7 Accuracies and sizes of five feature selected methods (comparative dataset).

Method	Accuracy exceeding 80% (number of features)	Feature Number=			Min EER
		5	15	30	
ANOVA	32	40.99%	83.46%	98.14%	0.04
Pearson's <i>r</i>	32	39.44%	78.14%	98.36%	0.04
MI	27	56.58%	91.85%	100%	0.02
mRMR	26	64.53%	95.89%	100%	0.016
IFS	16	79.81%	98.69%	100%	0.008

In Table 3.6 and Table 3.7 the feature subsets selected by the IFS algorithm appear to offer the best performance. Five feature selection algorithms all

perform well, and can achieve a high classification accuracy. The IFS algorithm can obtain the best classification accuracy through the minimum number of feature subsets. The classification accuracies of mRMR and mutual Information algorithm are higher than ANOVA and Pearson's r .

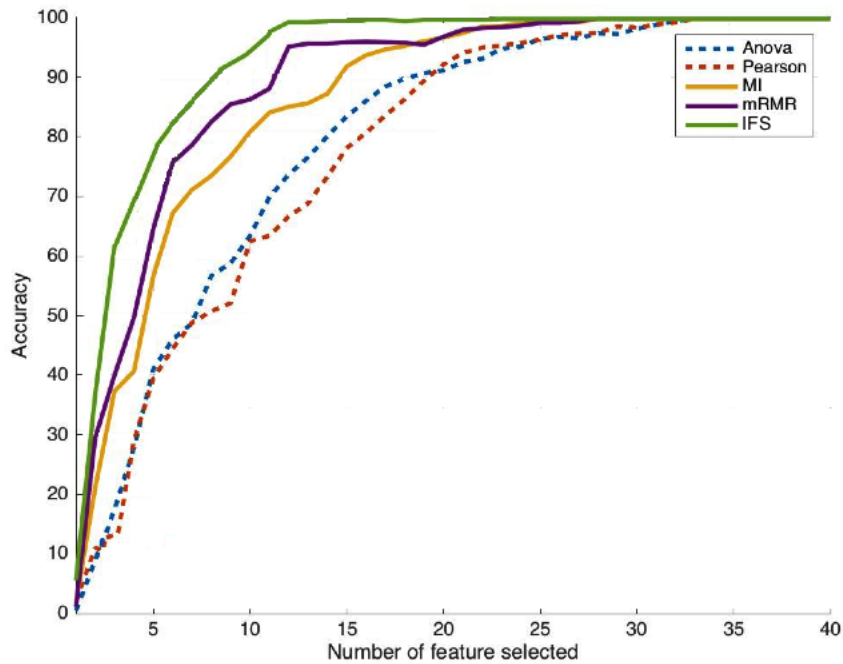


Figure 3.4 Accuracy comparison of five feature selection methods (comparative dataset). Accuracy calculated using k NN (with $k=1$).

ANOVA and Pearson's r feature selection approaches both measure the relationship between each feature and label independently. Pearson's r is sensitive to linearity, which means that if it is a non-linear relationship, even if the two variables are in one-to-one corresponding relations, Pearson's r correlation result might be close to 0. If it is only judged according to the values of ANOVA or Pearson's r , the results could be misleading. It would achieve a good effect if the features and labels are all in a linear relationship and the features are independent. However, with respect to the data, it is clear that the features are in close correlation. There are multiple correlated features in the data, and some features are redundant. Meanwhile, the features are not completely in linear correlation with labels. mutual Information is used to measure the nonlinear relationship of two variables. mRMR not only considers

the correlation between features and labels, but also takes the relationship between features into consideration, which is more suitable for analysis of the data. IFS can obtain the best results.

3.2 Fusion at score level

Score level fusion is a combination of match scores from different biometric matchers, which then derives a new score. Each biometric sample calculates the match score independently, and score level fusion combines all match scores into a single score through an algorithm. Figure 3.5 shows a block diagram of a feature level fusion.

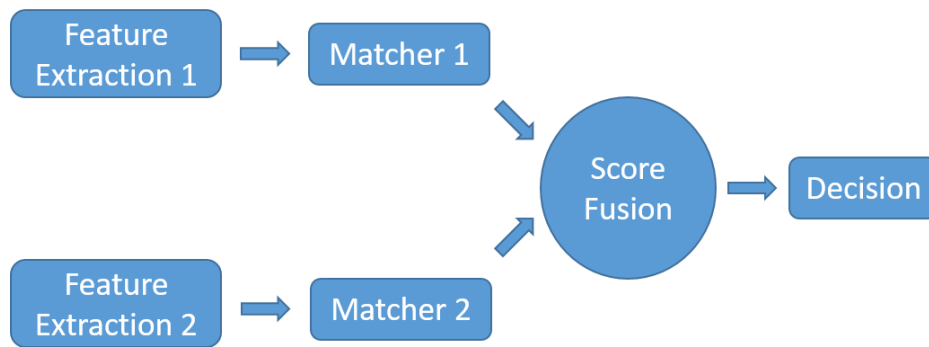


Figure 3.5 Score level fusion.

3.2.1 Similarity score calculation and normalization

For a biometric matcher, the outputs of the classifier are always comprised of a match score list and a rank list. Match score describes a distance or a similarity between the testing subject with registered subjects. The distance is normally calculated by Euclidean distance, Mahalanobis Distance or other algorithms. The similarity score can be then computed by Eq.(3.8).

$$sim(x, y) = \frac{1}{1+D(x,y)} \quad (3.5)$$

where $D(x,y)$ is the distance between two samples. A distance of 0 corresponds to a similarity of 1 (the largest possible value); a distance of infinity corresponds to a similarity of 0 (the smallest possible value). The rank list is obtained by the decreasing order of similarity scores.

The range of match scores for individual matchers may not be homogeneous, and the scores may follow different statistical distributions. Score normalisation is a necessary step before the combination scheme, as it is designed to scale the range of individual match scores and transfer them into a common range.

A widely used technique is Z-score normalisation. This method requires the mathematic mean and standard deviation of a given score list. The normalisation is conducted:

$$s'_k = \frac{s_k - \mu}{\sigma} \quad (3.6)$$

where s'_k is an updated score, μ is the arithmetic mean and σ is the standard deviation.

In order to improve the estimation ability of the marginal distribution of similarity scores [41], the following order preserving transformation was recommended:

$$T(s_k) = \log \frac{s_k - a}{b - s_k} \quad (3.7)$$

Where s_k is similarity score and it is bound between a and b .

3.2.2 Score-level fusion method

3.2.2.1 Simple average score fusion

Match score fusion is implemented by calculating the average value of each matching parameter. Figure 3.6 shows an example of simple average score fusion.

$$y_i = \frac{1}{M} \sum_{j=1}^M s_i^j, \quad \forall i \quad (3.8)$$

In Eq. (3.8), M is the number of the multimode biometric matchers for fusion. It denotes M matchers in a given multimode biometric system. Each matcher is a single mode biometric system and labelled by numerical indicator $j \in 1, 2 \dots M$ and s_i^j is the i^{th} match score of the j^{th} matcher. Figure 3.6 shows an example of simple average score fusion, for $M = 2$ and $i = 1, 2, 3, 4$.

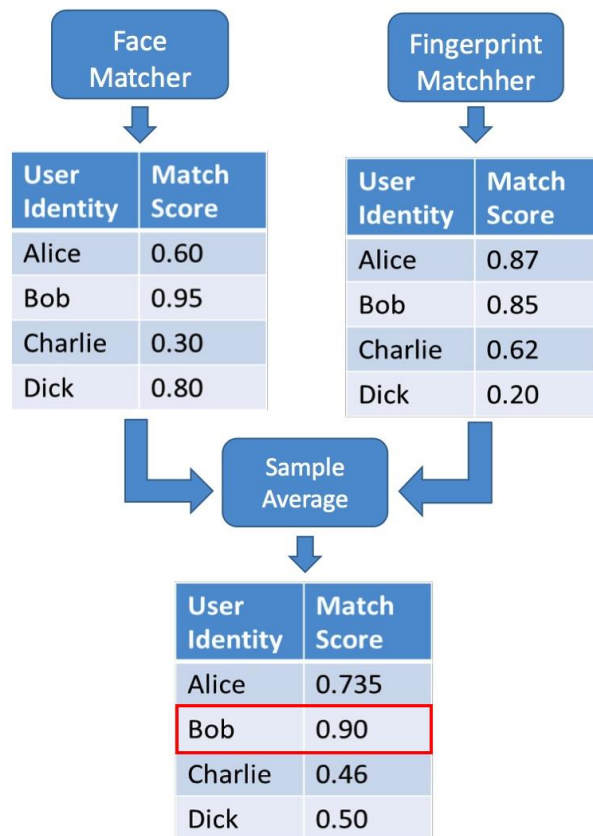


Figure 3.6 An example of simple average score fusion. Note that the match scores generated by the face and fingerprint matcher are similarity measurements. The range of match score is assumed to be $[0,1]$.

3.2.2.2 Max-Score fusion

The fusion function of Max-Score (MAS) is presented in Eq. (3.9):

$$y_i = \max(s_i^1, s_i^2, \dots, s_i^j, \dots, s_i^M), \quad \forall i \quad (3.9)$$

j is the serial number of matchers for fusion. M is the number of multimode biometric matchers. s_i^j is the score of matching with i_{th} subject using j_{th} matcher. The output of the maximum score method fusion is the maximum of the match scores of M matchers. Figure 3.7 shows an example of Max-Score fusion.

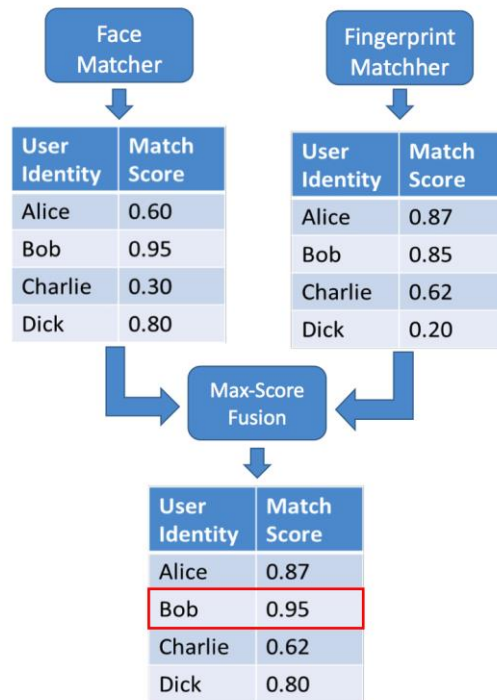


Figure 3.7 An example of max-score fusion. Note that the match scores generated by the face and fingerprint matcher are similarity measurements. The range of match score is assumed to be $[0,1]$.

3.2.2.3 Density-based score fusion using Bayesian theory

3.2.2.3.1 Estimation of similarity score densities

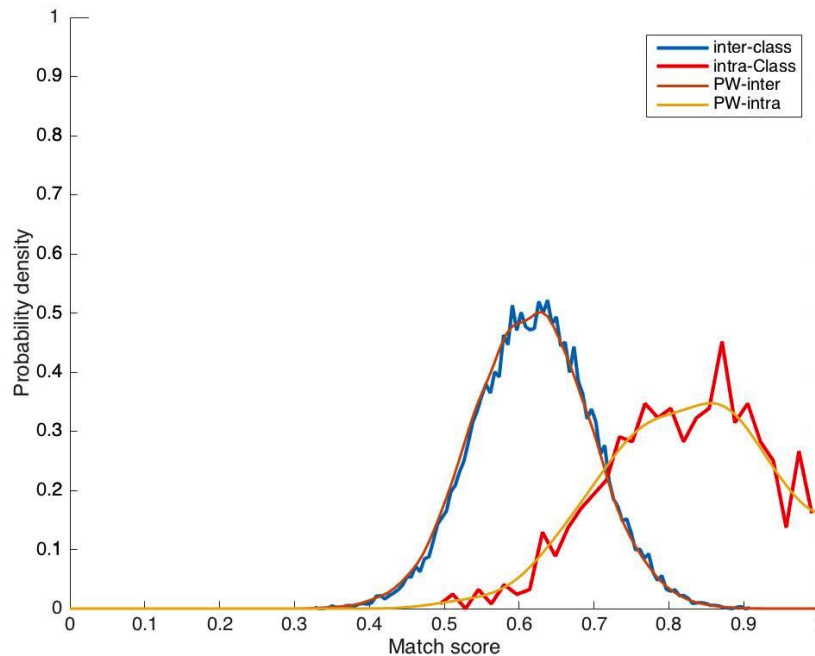
The method of Parzen Window is used in this analysis to estimate the probability density of the match score [42] [43] [44]. It is a nonparametric method to estimate Probability Density Functions (PDF). In order to smooth the estimation result, the Gaussian Kernel function will be applied. For Parzen Window estimation, the bandwidth (window size) is fixed, while the size of bandwidth has great influence on estimations. If the bandwidth is excessively large, the estimation result would be over-smoothed, which loses data characteristics. However, if bandwidth is too small, the estimation would be excessively limited to observation data, and many incorrect peak values are likely to appear. Therefore, the bandwidth needs to be selected according to the size of the dataset. One method of selecting bandwidth is to use the Asymptotic Mean Integrated Squared Error (AMISE), which selects the bandwidth that can best minimise AMISE [45].

There are three databases for human body, face and clothing, and the Parzen Window is used to estimate the PDF of each of them. In Eq.(3.14), a Gaussian Kernel function is used to smooth the estimation result.

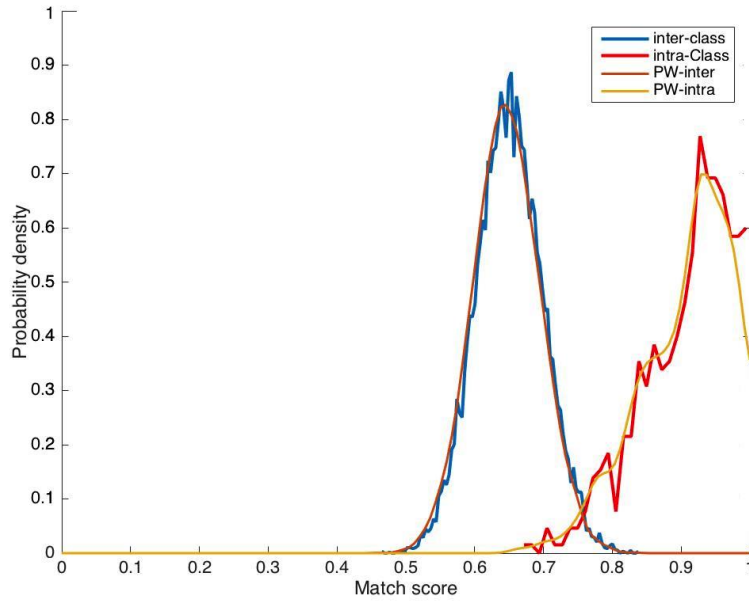
$$p(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \left(-\frac{(x_i-x)^2}{2\sigma^2} \right) \quad (3.10)$$

where x_1, \dots, x_n are n data samples and x is the centre point, σ needs to be selected. This is the average of these Gaussian functions, with each data point as a centre point.

By taking the match scores of intra-class and inter-class as the inputs of a Parzen window, one can estimate two pmfs; the pmf for intra-class, and the pmf for inter-class. Figure 3.8 is an example of prediction results of probability densities of the comparative face and categorical body datasets, which are the best and the worst single identification mode, by using a Parzen window. In Figure 3.8, "Inter-class" and "intra-class" are the results obtained according to the histogram; "PW-inter" and "PW-intra" are the results using Parzen window.



(a) Categorical body dataset



(b) Comparative face dataset

Figure 3.8 Probability densities of comparative face and categorical body (estimated by a Parzen window), "Inter-class" and "intra-class": obtained according to the histogram; "PW-inter" and "PW-intra": estimated by a Parzen window.

3.2.2.3.2 Score fusion using Bayesian theory

Three independent databases include the human body, face and clothing. Each has 2 sets of features: comparative and categorical. First, the match score of each database is calculated to estimate its probability mass function (pmf) and fuse the match scores according to Bayes' theorem.

In probability theory, Bayes' theorem is a way of understanding the probability that an event is affected by another piece of evidence [46]. Bayes' theorem is given in Eq. (3.11).

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (3.11)$$

where $P(A)$ and $P(B)$ are the probabilities of observing A and B without regard to each other. $P(A|B)$, a conditional probability, is the probability of observing

event A given that B is true. While $P(B|A)$ is the probability of observing event B given that A is true.

The verification problem can be viewed as a two-class classification issue. If the users are correctly identified, they are genuine; if the users are mistakenly identified, they are impostors.

There are 40 subjects in three databases. Taking the body database as an example, the 27 samples of each person are different observers' descriptions of this person. The samples of each subject are assigned into two groups randomly, among which half of the samples in one group are for training, while the remaining are for testing. When testing, matching one sample to any one of the 40 templates results in a match score; the probability of this match score in intra-class and inter-class can be obtained based on the two pmfs obtained from the training sample. Taking this probability as the prior probability, the probability of this matching fraction in intra-class and inter-class can be estimated according to the Bayesian Theory. $P(\text{genuine}|s)$ is the probability that the label of the test sample is the same as the label of the template. The formula represented in Eq.(3.12), $P(\text{impostor}|s)$ is considered as the probability of inter-class, which means the probability that the label of the test sample is different from the label of the template. The formula is presented Eq.(3.13).

$$P(\text{genuine}|s) = \frac{P(s|\text{genuine})P(\text{genuine})}{p(s)} \quad (3.12)$$

$$P(\text{impostor}|s) = \frac{P(s|\text{impostor})P(\text{impostor})}{p(s)} \quad (3.13)$$

$$p(s) = P(s|\text{genuine})P(\text{genuine}) + P(s|\text{impostor})P(\text{impostor}) \quad (3.14)$$

The same calculation is performed on all three databases.

Because the three databases are completely independent, in Eq.(3.15) and (3.16), M is the number of sample matchers (face matcher, body matcher, clothing matcher); in this case, $M = 3$.

$$P(\text{genuine}|S_1, \dots, S_M) = \prod_{j=1}^M P(\text{genuine}|S_j) \quad (3.15)$$

$$P(\text{impostor}|S_1, \dots, S_M) = \prod_{j=1}^M P(\text{impostor}|S_j) \quad (3.16)$$

In the final judgment, D in Eq.(3.17) is used to obtain the decision based on the Bayesian decision rule. If the output is 1, it indicates that the match score belongs to a genuine user. In other words, the label of the test sample is the same as the label of the template. If the output is 0, the score belongs to an impostor. The label of the test sample is different from the label of the template.

$$D = \begin{cases} 1 & P(\text{genuine}|S_1, \dots, S_N) \geq P(\text{impostor}|S_1, \dots, S_N) \\ 0 & \text{otherwise} \end{cases} \quad (3.17)$$

3.2.3 Score-level fusion experiment results

3.2.3.1 Simple average score fusion

In the three single-mode biometric features, body, face and clothing, the recognition performance based on the face is highest, while the fusion performance outperforms the facial recognition. Furthermore, the feature-fusion recognition experiments of body, face and clothing are based on the simple average fusion method. ROC curves are shown in Figure 3.9. It can be observed that there is no significant reduction of ERR for both datasets. The purpose of the multimodal biometric recognition is to improve the reliability of the identification SA (Simple Average) fusion algorithm, and can significantly improve the accuracy of the recognition system.

Table 3.8 Accuracy comparison of simple average fusion

Mode	Accuracy (rank=1)	
	Categorical dataset	comparative dataset
Body	45.32%	97.09%
Face	76.03%	84.09%
Clothing	63.09%	48.42%
Simple Average Fusion	81.02%	98.23%

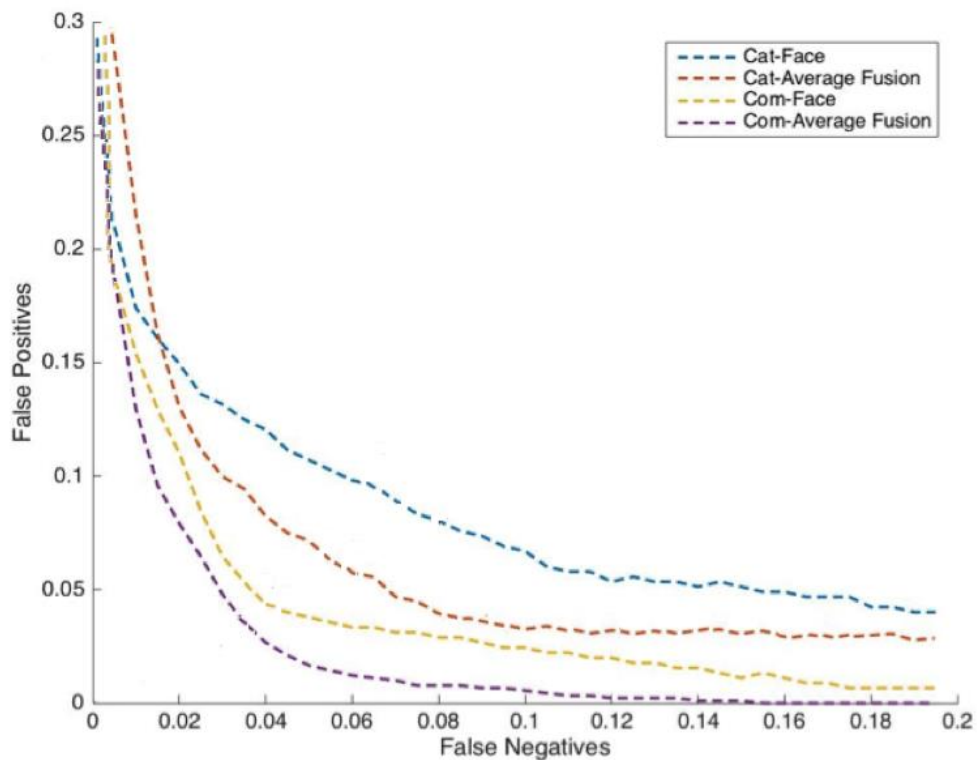


Figure 3.9 Simple average fusion results. EER is calculated by investigating an equal number of False Positives and False Negatives, categorical face (Cat-face), comparative face (Com-face), simple average score fusion on categorical (Cat-Average Fusion) and comparative (Cat-Average Fusion).

3.2.3.2 Max-Score fusion

The ROC recognition curves before and after MAS fusion are shown in Figure 3.10. It shows that there is no significant reduction of the MAS fusion recognition. However, False Negative Rate (FNR) reduces whilst False Positive Rate (FPR) rises for MAS fusion recognition.

The output result of max-score fusion is the maximum value of a group of matching scores. The mathematical expectation of match score after fusion shall be greater than each individual matcher. Thus, under the condition of obtaining the same threshold, the FPR will increase while FNR will decrease after fusion [47].

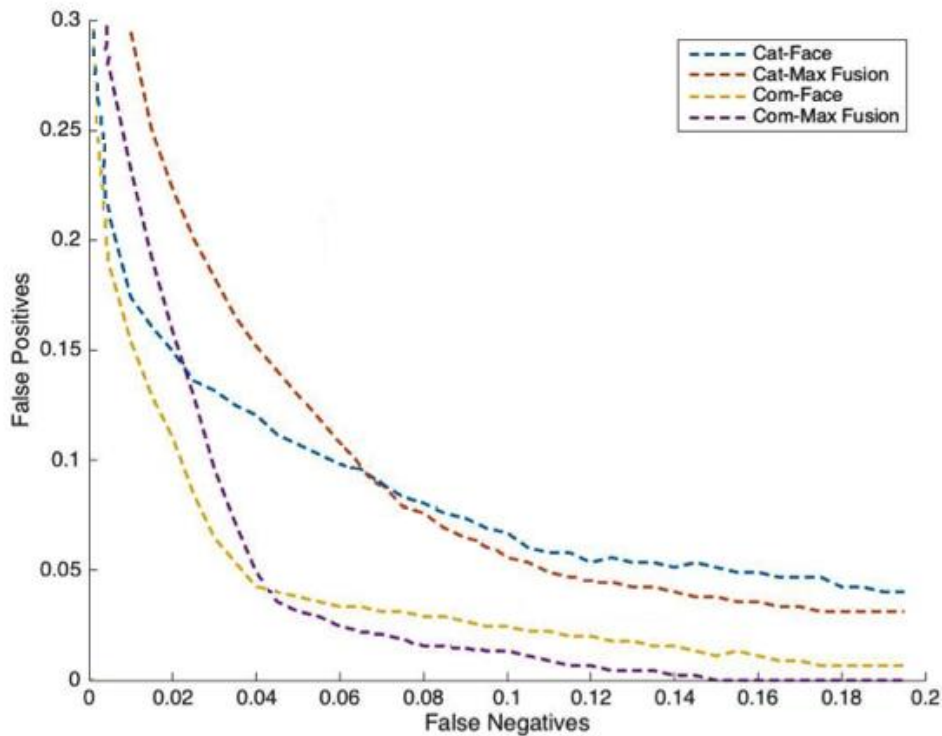


Figure 3.10 Max score fusion results. EER calculated by finding an equal number of False Positives and False Negatives, categorical face (Cat-face), comparative face (Com-face), max-score fusion on categorical (Cat-max Fusion) and comparative (Com-max Fusion).

3.2.3.3 Score fusion using Bayesian theory

The verification tests of the samples of 40 users for comparative data were conducted with the Bayesian fusion algorithm to calculate the ROC curve, as shown in Figure 3.11. The verification performance is improved with fusion. In order to analyse the comparative data, 5 comparisons were chosen randomly for each subject, and then the ROC curve was calculated for each case. When 5 comparisons were used, the classification accuracy of facial data was 96.2% (EER was 0.040). The accuracy of body data was 84.9% (EER was 0.79), and that of the clothing dataset was 31.9% (EER was 0.159). The results were improved by 6.1% from Samangoei [17], 1.9 % from Reid [4] and 31.4% from Jaha [20]. Compared with these studies, the number of subjects is different, but the features are the same. The slight increase is likely to be due to the potential

differences in a smaller population derived from the same dataset. After Bayesian fusion, verification accuracy is 99.42% and the EER is 0.14%.

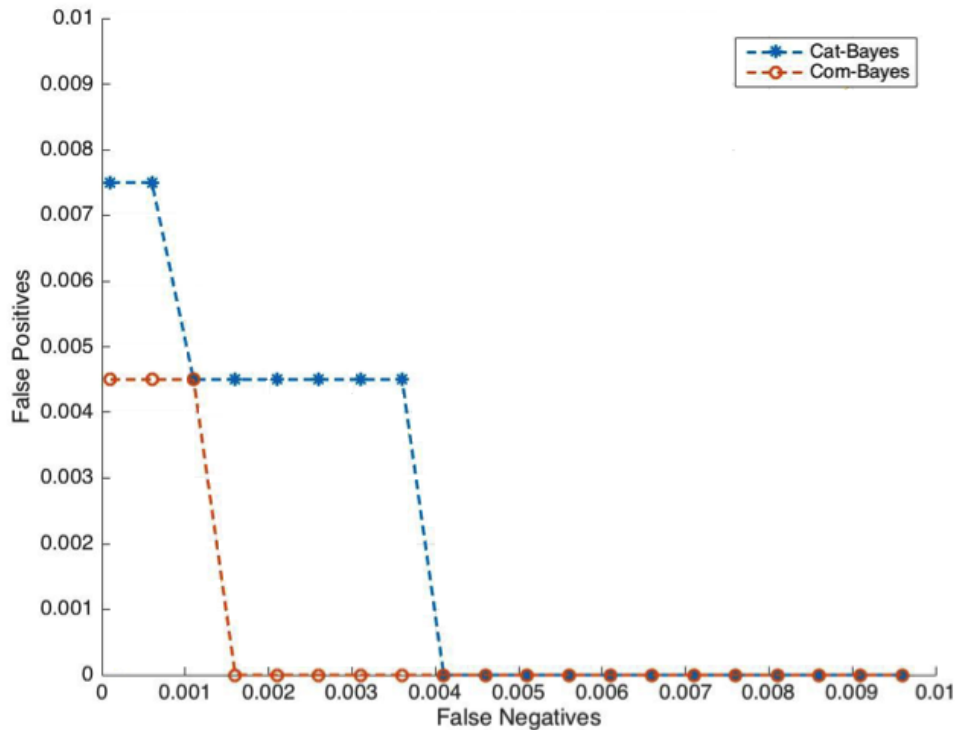


Figure 3.11 Bayesian fusion result, EER calculated by finding an equal number of False Positives and False Negatives, Bayesian score fusion on categorical (Cat-Bayes) and comparative (Com-Bayes)

The laboratory data was derived using moderate lighting, and the labels had high descriptive power. Since the data was derived in laboratory conditions (rather than surveillance video), good verification performance was achieved, and the results suggest that fusion would further improve performance.

Another important reason that Bayesian fusion can achieve such a good result is that the determination is based not only on the information of the test samples, but also on prior probabilities after the Bayesian estimation. The prior probability distribution represents a powerful mechanism to combine information with previous match scores. The capacity to consider uncertainty is a major strength of Bayesian fusion.

3.3 Conclusions

This chapter introduced the basic theory and function of the feature fusion algorithm, and the basic algorithms of feature selection. ANOVA and Pearson's r are two simple feature selection approaches with smaller computational load, though they have limitations. For example, Pearson's r can calculate correct linearity, but it is sensitive to noise. In addition, this chapter introduced three feature selection approaches based on information theory. First, it introduced the basic concepts of the information measurement standard in information theory, such as information entropy and mutual information. Then it described three algorithms: MI, mRMR and IFS. The experiments demonstrated that feature selection based on information theory obtains better results, because the features are relevant. It also proved that IFS obtains the feature subset with the best classification performance.

The last part of this chapter explored identification by the fusion of face, body, and clothing features, based on several fusion theories at match score level. It analysed the fusion algorithm of two classic matching levels: sample average score fusion and max-score fusion, and it then implemented experimental analysis. In order to improve the recognition performance of multi-biometric methods, a fusion algorithm based on Bayesian theory was proposed. According to the experiment, the EER of fusion based on Bayesian theory was 0.0014 for comparative data and 0.0036 for categorical data. This result is much better than the recognition performance based on the face, which obtained high recognition performance using single-mode.

The results suggest that fusion is a suitable approach, since it can exploit properties of different features and in different scenarios. However, the data were not designed to be consistent across the three modalities of face, body and clothing. This suggests that the next stage of the investigation into soft biometric fusion is to collect a consistent set of data, so that the properties of fusion can be better explored.

Chapter 4 **Soft Biometric Dataset at Different Distances**

For soft biometric recognition, there is no standardised dataset to evaluate recognition performance, and this is especially problematic in the research area of soft biometrics at different distances. One significant advantage of soft biometrics is that it does not have rigorous requirements on the resolution of images collected by cameras. The label of soft biometrics can be therefore collected at a distance. The research into the influence of distance on feature annotation can give more useful information about features to be used at far or close distances. A new soft biometric database, based on different distances, must be built. In order to approach the identification in real life, the images in this database were simulated in an outdoor environment. Compared with the background controlled in a laboratory, there are more objects, such as buildings and cars, in the outdoor background, which can be used as points of reference when the soft traits are labelled. The accuracy of feature annotation can be improved with the help of reference objects.

4.1 Soft biometric dataset

4.1.1 Synthesising images

The new database, comprising of 131 male and 69 female subjects, was built and labelled with face, body and clothing traits. The original images were collected from the University of Southampton Gait Tunnel [48], and then

synthesised with an outdoor environment. There were 12 cameras deployed in different positions in the gait laboratory. The resolution of cameras is 640×480 and the capture rate is 30 frames per second. The viewpoint, at which the images collected provide the maximum face and body information of subjects was selected, as shown in Figure 4.1. The length of the gait laboratory is 7 meters. The minimum distance for the whole-body observation is 2 meters away from the camera. The image acquisition is therefore conducted between 2 and 7 meters. Three points (2, 4.5 and 7 meters away from the camera) are marked as close, medium and far respectively. Three images in which the subjects are stood in those three positions are selected, and then simulated in an outdoor environment.

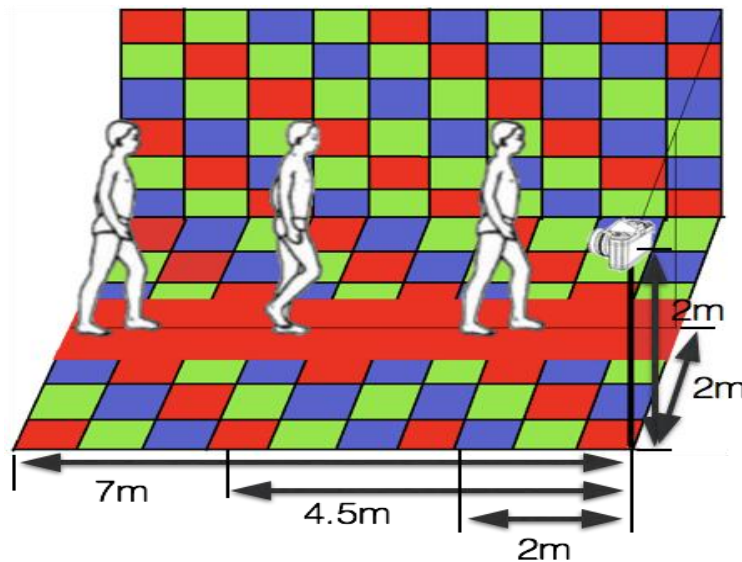


Figure 4.1 Diagram of data acquisition environment.

In order to bridge the gap between laboratory and outdoor images, the illumination and shadows need to be controlled. A cloudy day was chosen to eliminate shadows in the outdoor environment. Moreover, the brightness and contrast of laboratory images were improved to resemble outdoor conditions. The laboratory images after pre-processing are shown in Figure 4.2, in which subjects are located at far, medium and close distances respectively.



(a) Far



(b) Medium



(c) Close

Figure 4.2 Laboratory images after pre-processing.

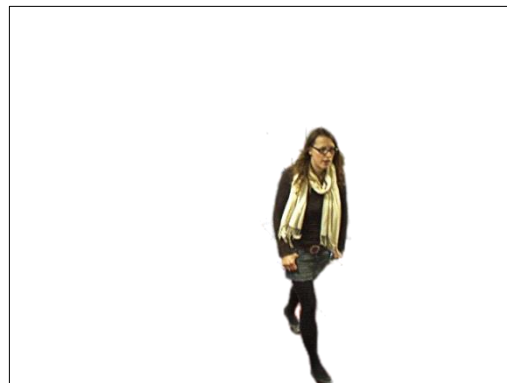
Taking a set of pictures of the outdoor environment is the first step towards synthesising images for the new dataset. The position of the camera in the outdoor environment should be the same as that used in the laboratory. Three points at the close, medium and far distance were marked in advance. The outdoor background image is shown in Figure 4.3.



Figure 4.3 Image of the outdoor environment.



(a) Laboratory - close



(b) Subject extraction



(c) Outdoor background



(d) Synthetic - close

Figure 4.4 Synthesising images with outdoor environment

The next step is to segment subjects from laboratory images and to place them at appropriate locations in the outdoor background image. First, the subject's image is extracted from the laboratory image (Figure 4.4 (a), (b)) and then added to the background (Figure 4.4 (c)) where the imaging geometry was the same as in the view in the laboratory. In this way, the controlled laboratory background is replaced with a consistent outdoor background. A synthesised image is shown in Figure 4.4 (d). This process is applied to laboratory images acquired at the close, medium or far distances from the camera. A few of examples of synthesising images are shown in Figure 4.5.

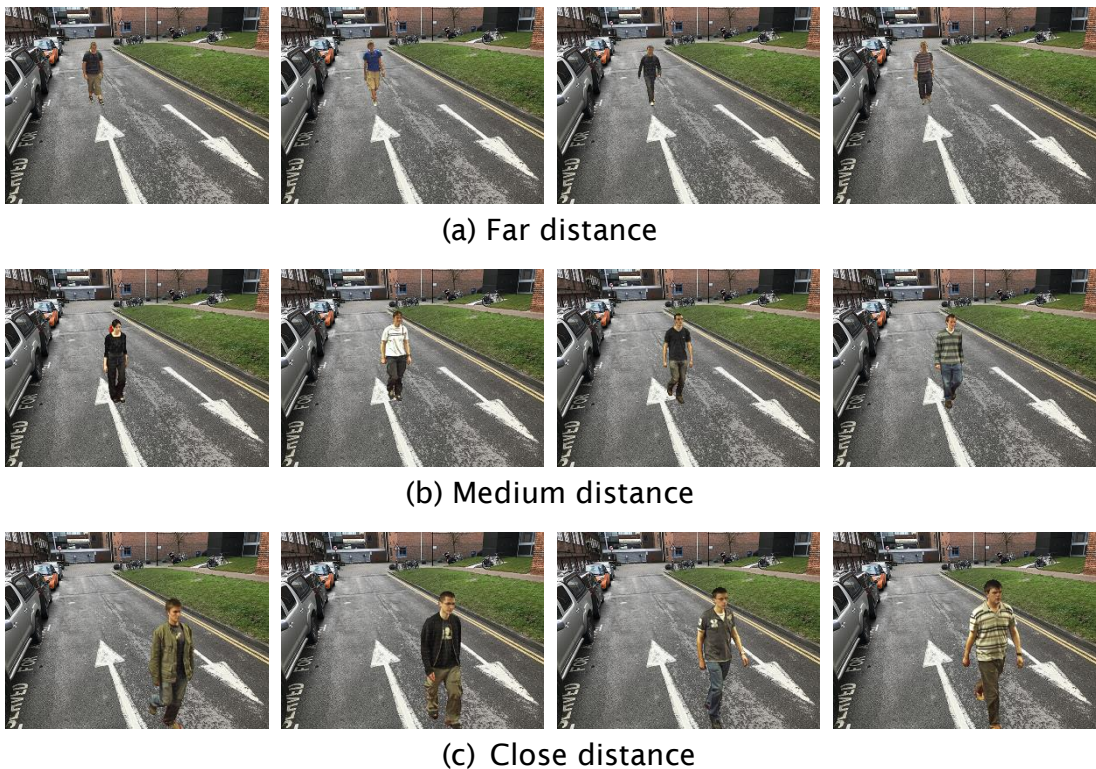


Figure 4.5 Examples of synthesising images at different distances.

4.1.2 Soft biometric attributes and labels

The next step was to select soft biometric features for the body, face and the clothes that could be observed and described precisely and conventionally at different distances (i.e. skin colour and height). The facial features (i.e. the shape of the eyebrows and the length of the face) have strong discriminating power, while body and clothing traits such as gender and the majority colour of clothes, are more straightforward to observe. The earlier research studied

recognition capability, and this part will utilise that work. It would be possible to prune the feature set according to their recognition capabilities.

19 body attributes were analysed in [4]. The recognition performance of 12 body attributes was investigated in [49]. Considering the results in those two articles, 10 effective attributes were selected as the body attributes for the new dataset. The attributes and labels are listed in Table 4.1.

Table 4.1 Body traits and labels used to compare subjects

Body traits	Labels
Gender	More feminine, Same, More masculine
Age	Older, Same, Younger
Height	Taller, Same, Shorter
Weight	Fatter, Same, Thinner
Shoulder shape	More square, Same, Rounder
Hair colour	Lighter, Same, Darker
Hair length	Shorter, Same, Longer
Neck length	Shorter, Same, Longer
Humpback	More straight, Same, More curved
Arm length	Longer, Same, Shorter

24 facial attributes were compared in [5]. ANOVA, Entropy and mutual information were employed to analyse the recognition performance of the attributes. The results demonstrated that Skin Colour, Eyebrow Length, Lip Thickness, and Face Length have better recognition performance and are the most consistent. The facial traits were selected based on [5]. In this paper, the face images were collected at a close distance with high-quality images. Some features which cannot be observed from far away, such as Eye-to-Eyebrow Distance and Inter Eyebrow Distance, are modified in the new dataset. For example, Eyebrow Thickness and Eyebrow Length are replaced with Eyebrow Shape. The attributes and labels are listed in Table 4.2.

Table 4.2 Face attributes and corresponding categorical labels

Face traits	Labels
Eye-brow shape	More straight, Same, More curved
Nose shape	More flatter, Same, More protruding
Forehead	Straighter hairline, Same, More receded hairline
Eyes	Smaller, Same, Larger
Ears	More hidden, Same, More evident
Skin colour	Lighter, Same, Darker
Face size	Shorter, Same, Longer
Face	More bony, Same, Fleshier
Lips	Thinner, Same, Thicker
Chin and jaw	More angular, Same, Rounder

In [6], 21 categorical traits and 7 comparative traits of clothing are analysed. The results demonstrated that clothing attributes can achieve good results when they are used for recognition. Furthermore, the accuracy using categorical traits is better than comparative traits. ANOVA was used to analyse the performance of different traits, and the results show that head coverage, lower body clothing category and belt presence are better identifiers than other traits. Thus, categorical traits are used for the new clothing dataset. 7 features are investigated in [6], which have good recognition performance and are straightforward to observe at different distances, plus 3 new attributes: the majority colour of upper body and of lower body, and the presence of glasses, constitute the new clothing feature set. The attributes and labels for clothing are listed in Table 4.3.

Table 4.3 Clothing attributes and corresponding categorical labels

Clothing traits	Labels
Upper body clothing category	Jumper, T-shirt, Shirt, Blouse, Sweater, Coat, Hoodie, Other
Lower body clothing category	Trousers, Skirt, Dress
Any attached object category	None, Bag, Gloves, Hat, Scarf, Necktie, Other

Clothing style	Well-dressed, Business, Sporty, Fashionable, Casual, Other
The majority colour of upper body	Grey, Black, White, Jeans blue, Others
The majority colour of lower body	Grey, Black, White, Jeans blue, Others
Face coverage	Yes, No
head coverage	Yes, No
Presence of belt	Yes, No, Unsure
Wear glasses	Yes, No

4.1.3 Data acquisition via Crowdsourcing

The soft features were labelled by human operators. In order to increase efficiency and to obtain high-quality results, the features were collected through crowdsourcing. A crowdsourcing task needed to be built for the large collection of high-quality comparative annotations. The CrowdFlower platform was used to build and run the crowdsourced annotation task. CrowdFlower provides comprehensive data analysis and quality control tools, allowing acceptance of a range of responses, whilst rejecting non-genuine answers.

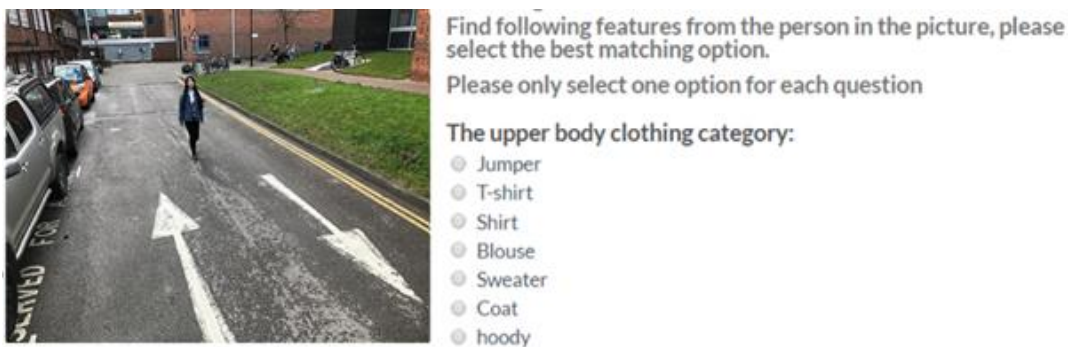
Each of the new images was labelled with all 30 soft biometric traits. In data collection, categorical descriptions of clothing and comparative descriptions of the face and body traits were used. It was demonstrated that the clothing traits could be used for recognition, and it is properly described using categorical labels [28]. In terms of face and body traits, comparative descriptions can convey more accurate descriptions, since observers can easily perceive differences between two subjects, for example, one person being taller than the other [4] [17], This eliminates known psychological effects, such as owner variables and confirmation bias.

Each comparison describes the difference of each feature between two subjects, such as height, weight and the length of an arm. The comparison for each feature is labelled using three classes: shorter, the same or higher, following the observation that a scale of 3 could lead to better discriminative capability [5]. Each level is denoted by a signed integer, for example, when comparing the height of two subjects, -1 means shorter, 0 represents the same

and +1 means taller. The labels for the new database based on the traits were collected using CrowdFlower. The interface for the collection system is shown Figure 4.6. Each of the 200 individuals was labelled by 20 people for the categorical clothing labels. The face and body were labelled by the comparison between each of the 200 subjects and 20 randomly chosen subjects. The total number of comparisons is 4000, each one labelled by 20 people.



(a) Body labels



(b) Clothing labels

Figure 4.6 Interface of label collection system

4.2 Attributes Analysis

4.2.1 Ranking inference

Elo was used to rank the comparative data. It is necessary to ensure the Elo rating results are trustworthy. Pixel height is used to estimate the actual height of a subject in an image. Meanwhile, an Elo ranking result, in terms of human height, represents a systematic judgement of height according to the comparative description. The positive correlation of pixel height and Elo rating results can validate the accuracy of Elo rating results, as shown in Figure 4.7.

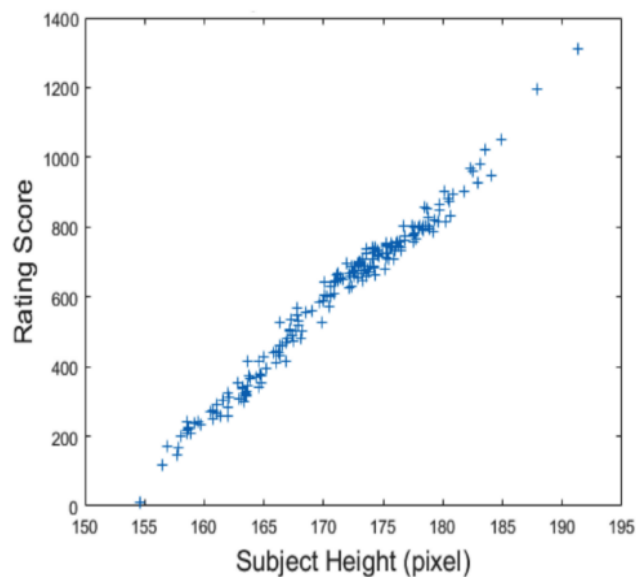


Figure 4.7 Relationship between estimated and measured height

The Pearson's correlation coefficient is 0.93 using this method, and outperforms the result 0.87 obtained previously [50]. This is likely due to the fixed geometry used from an outdoor environment, which allows labellers to pay more attention to the comparison of height.

4.2.2 Correlation analysis

The correlation coefficient (Pearson's) of each semantic feature in three groups (close and medium, close and far, medium and far) was employed to measure the stability of the ranking system. Theoretically, a larger coefficient indicates that a particular trait is less sensitive to distance.

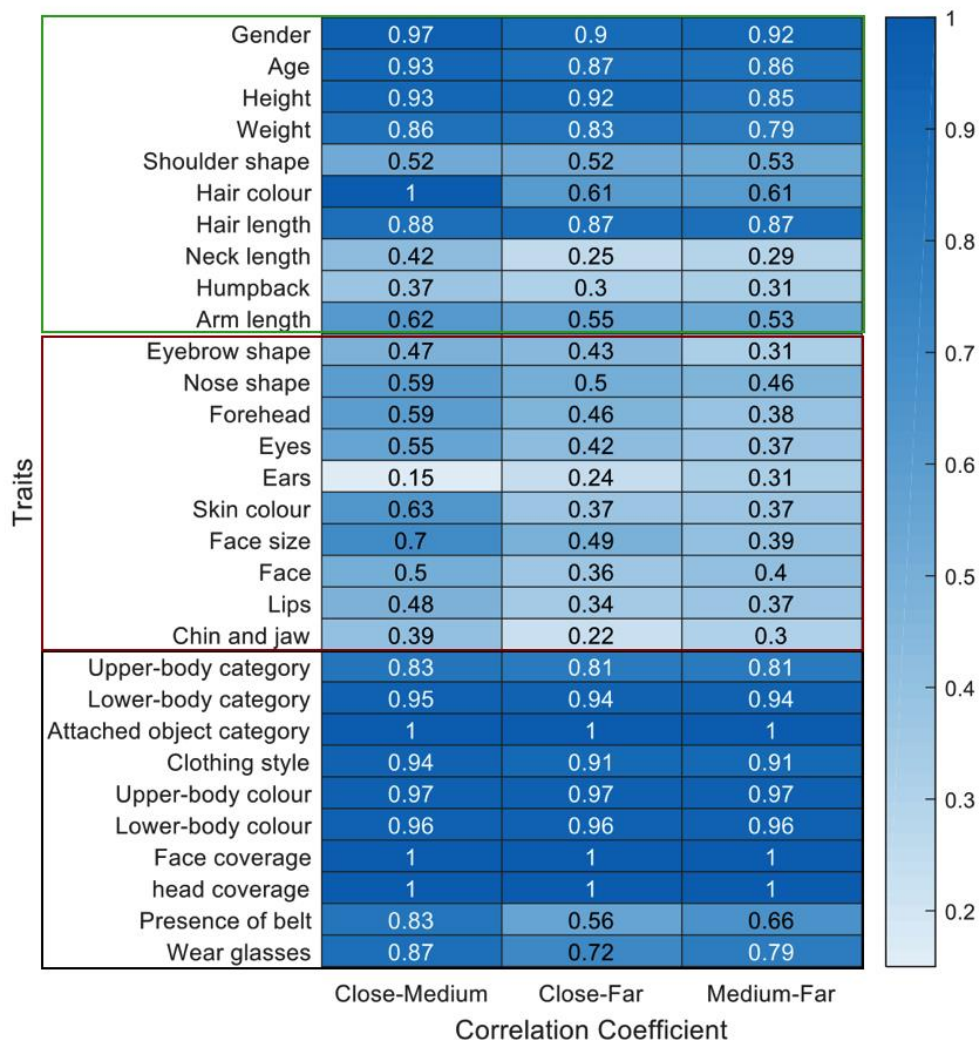


Figure 4.8 Pearson's correlation coefficient for each trait in three groups (close and medium, close and far, medium and far)

The correlation coefficient of each trait in three groups is depicted in Figure 4.8. It shows that the stability of most of the clothing traits, and some of the body traits, is relatively high, whilst the stability of face traits varies substantially at different distances. For example, in the body feature set, humpback and neck length are the most sensitive traits. Hair colour is good between close and medium distance, but worse at close-far and medium-far. The result demonstrated that the hair colour is less stable at a far distance. In the facial feature set, the ear is weak at all three groups, which means it is a sensitive feature, whilst face size has the highest stability in the face trait set. In clothing traits, face coverage and head coverage are equal to one at all three

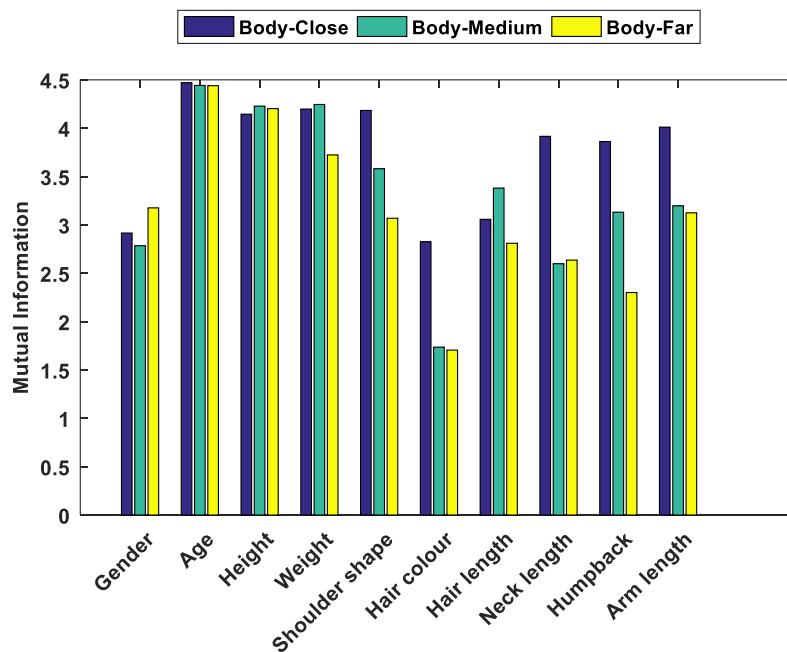
groups. In other words, these two features are the most straightforward to be observed.

4.2.3 Mutual information

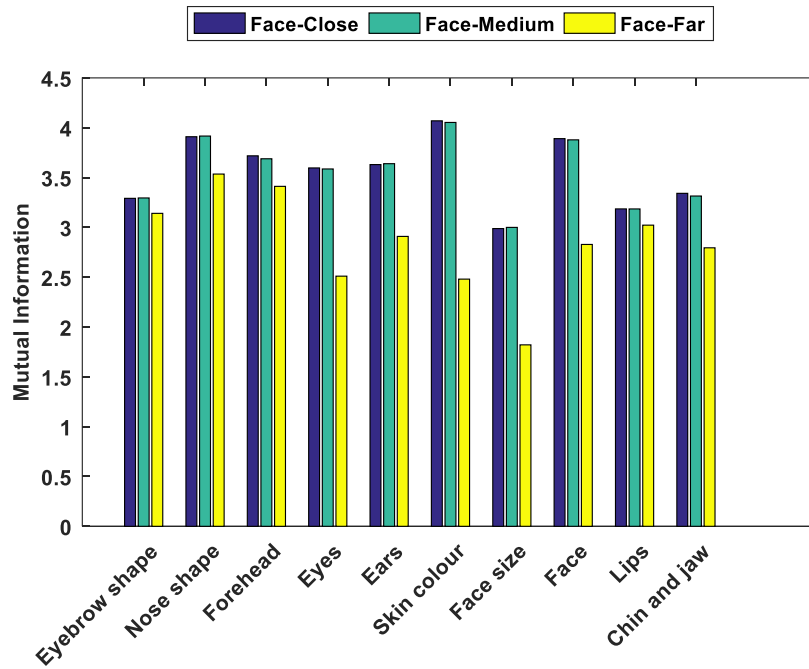
Mutual information was introduced to measure the intensity of the correlation between two variables. Given two random variables, X and Y , whose marginal probability distribution functions are $p(x)$ and $p(y)$ respectively, the mutual information $I(X; Y)$ was given in Eq.(3.3).

Here mutual information is used to measure the relevance between each trait and subject ID. Since subject ID presents the explicit differences of each subject, larger mutual information demonstrates the stronger discriminating capacity of the trait. Moreover, small differences of mutual information for each trait at three distances reflect superior reliability.

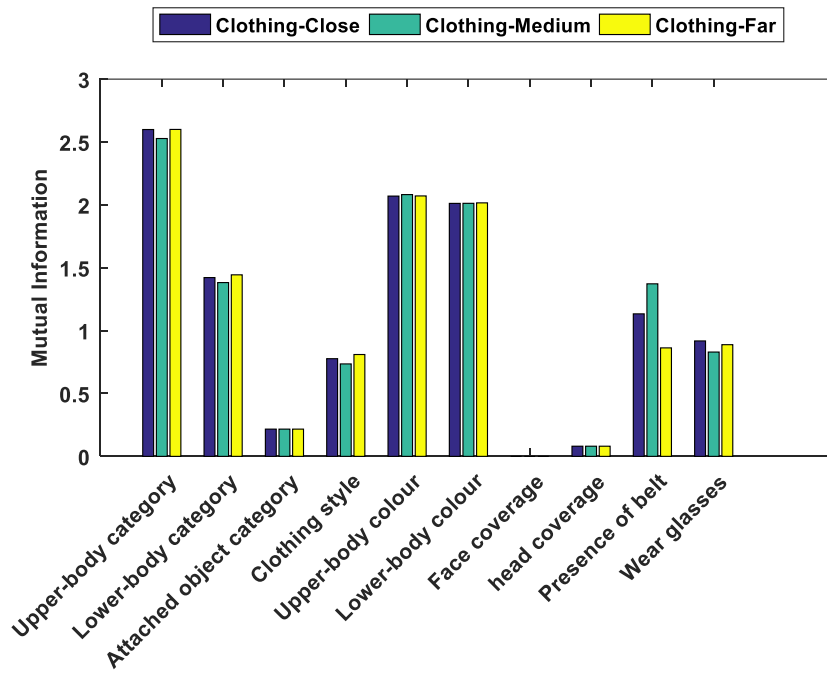
Figure 4.9 shows that age and upper body clothing category have the highest discriminatory differentiating power on body and clothing trait sets, respectively.



(a) Body traits



(b) Face traits



(c) Clothing traits

Figure 4.9 Mutual information of three datasets at three distances.

Despite skin colour showing the highest discriminatory power at close and medium distance, it is relatively weak at a far distance. It can be concluded

that body traits have the highest discriminating power at different distances, whilst the clothing traits are more stable across all distances. It appears that clothing traits are more easily observed in comparison with the other two types of traits. However, they have less uniqueness in recognition, which reduces discriminatory power. Further, clothing is an innately short-term biometric, since clothing can easily be changed. Moreover, since body traits have more detail than clothing traits, they have stronger differentiating power. It is known that the resolution of the face varies greatly with distance. Thus, the discriminatory power of facial traits decreases sharply at far distances. Furthermore, the trait stability evaluated by mutual information is in accordance with the results given by correlation analysis. The clothing traits appear to be the most stable traits for recognition.

4.3 Single-modal recognition

The purpose of the identification experiment is to assess the effectiveness of the proposed attributes (listed in Table 4.1, Table 4.2 and Table 4.3) for identification using the new dataset, and show the applicability of single-modal soft biometrics. The experiments simulate a realistic scenario that aims to retrieve the identity of an unknown subject (or probe) from a soft biometric database using verbal descriptions for the probe (i.e. eyewitness statement).

The experiments used LoO cross-validation, and were implemented using 200 subjects, in which 100 subjects were randomly chosen as training samples, and the remaining samples were used for testing.

The identification of unknown subjects was performed by calculating the Euclidean distance, d , between the biometric signature of the probe and the biometric signature of each subject in the gallery as follows:

$$d = \sqrt{\sum_{i=1}^T (X(i) - Y(i))^2} \quad (4.1)$$

where X is a vector that represents the biometric signature of one subject. For example, the unknown subject, Y , is a vector that represents the biometric signature of another subject (the subject in the gallery that is compared with the unknown subject), and $T = 10$ is the number of soft biometric attributes

composing the biometric signatures. The nearest neighbour was used here for classification: subjects were sorted in ascending order according to their distances from the probe, and the rank of the correct match was used to report the identification performance.

For face and body, the comparative labels are used. The recognition accuracy (rank=1) over varying numbers of probe comparisons (n) is shown in Figure 4.10 and Figure 4.11. . It is easy to obtain that with increase in the numbers of comparisons; the recognition accuracy is improved accordingly. The similarity of Figure 4.10 and Figure 4.11 demonstrated that the number of comparisons directly influenced the recognition performance.

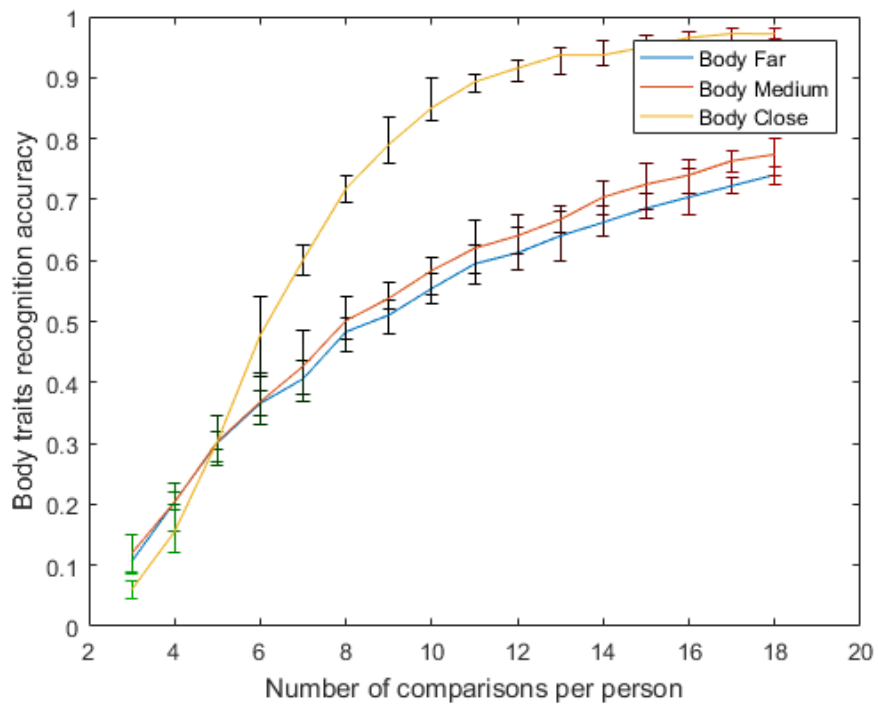


Figure 4.10 Body recognition accuracy obtained from different numbers of comparisons.

Figure 4.10 shows body recognition accuracy obtained from different numbers of comparisons. The Police and Criminal Evidence Act [51] explains that a good identity system should consist of 8 to 12 compared people. In this case, 9 comparisons are required for recognition. The recognition accuracy only used one comparison to construct the accuracy, of roughly 10% at all three distances. The recognition performance continues to increase over the range. At close

distance, the accuracy improvement is noticeable when the number of people compared is lower than 12 and achieves ~98% with 18 comparisons. For medium and far distances, it is achieving a ~85% correct recognition rate with 18 comparisons.

In comparison with [49], the database is comprised of 100 subjects and 12 body traits represented by comparative distributions. Their identification rate with 10 comparisons is ~97%, which is higher than that of the new method (~85%). Nevertheless, the dataset used in this chapter is twice as large as the dataset in [49] and employs less features. This study concentrates on the fusion approaches and on whether fusion itself can be used for recognition as well as its properties, whilst assuming that all single mode approaches can be improved independently.

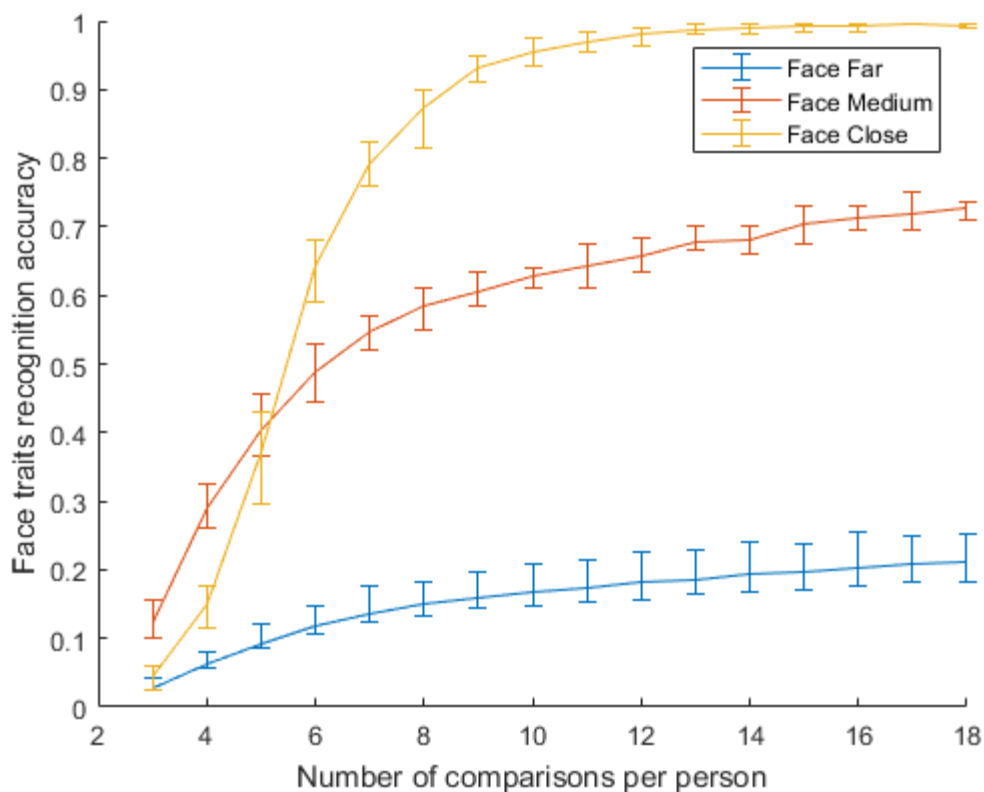


Figure 4.11 Face recognition accuracy obtained from different numbers of comparisons.

The facial recognition accuracy over different numbers of comparisons is shown in Figure 4.11. It shows that, at close distances, the accuracy of facial

descriptions greatly outperforms that of body descriptions. The facial description achieves an ~80% identification accuracy with 7 comparisons, whilst body only achieves ~60%. The recognition performance at a close distance is much better than that at medium and far distances. A ~97% recognition accuracy is obtained with ten comparisons, obtaining a maximum of a 100% accuracy at 18 comparisons. The recognition performance is limited at a far distance, with the accuracy rate achieving only ~19% when using 18 comparisons. This suggests that distance is more important than the number of comparisons when using the face for recognition.

By comparing the identification performance at close distance obtained from [5], which achieved an accuracy of 100% with 10 comparisons by using 24 attributes with 4038 subjects, the accuracy of the proposed method using 10 comparisons is ~97%, which demonstrates that 10 traits used here include enough information for classification.

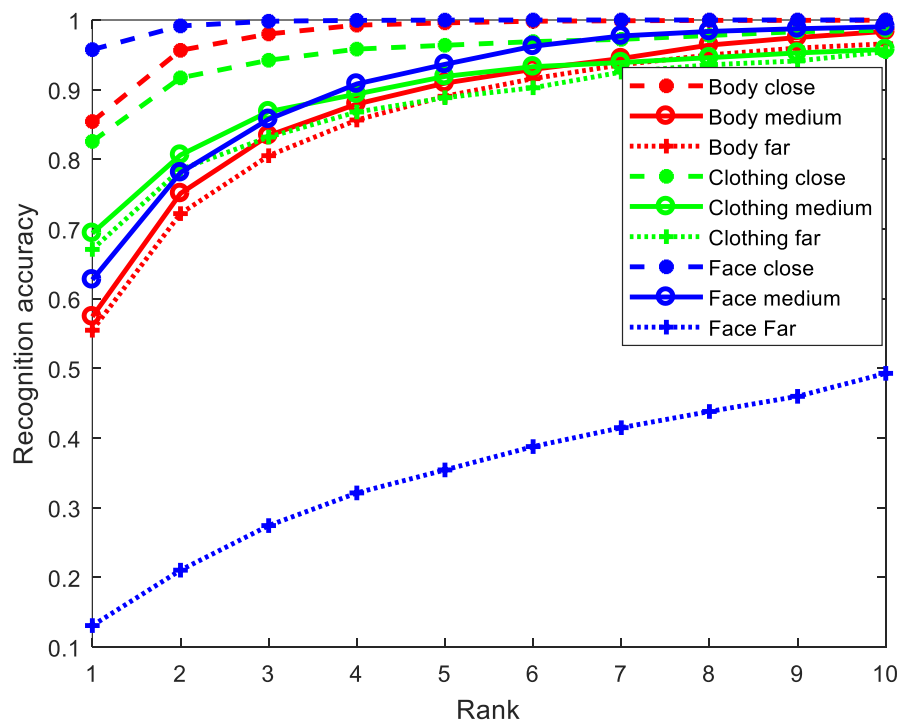


Figure 4.12 Cumulative match characteristic curves for the individual modalities

The cumulative match characteristic curves (CMC) for single-modal soft biometric feature sets are depicted in Figure 4.12. At a close distance, the recognition accuracy of facial traits is ~96% (for rank 1 identification), but it

falls sharply to ~63% at a medium distance and ~13% for a far distance. The recognition by clothing is the most consistent, with accuracies of ~83% ~69% and ~67% at the three distances respectively.

For facial traits, the recognition accuracy at close distance is 99.2% (rank=2) and 100% (rank=6). At medium and far distances, the recognition accuracy is 99.03% (rank=10) and 49.3% (rank=10) respectively. In terms of recognition using clothing traits, the accuracy at medium (95.8%) and far (95.9%) distances is very similar. It performs higher at a close distance (98.5%). For the body traits, the recognition accuracy is 99.3% (rank=4) at a close distance and decreases slightly to 98.4% and 96.6% (rank=10) at medium and far distances. In summary, at a close distance, the facial traits, as well as body and clothing traits, show high performance. At medium distance, three feature sets obtain similar accuracy, but at a far distance, the accuracy of clothing and body traits significantly outperforms that of the face.

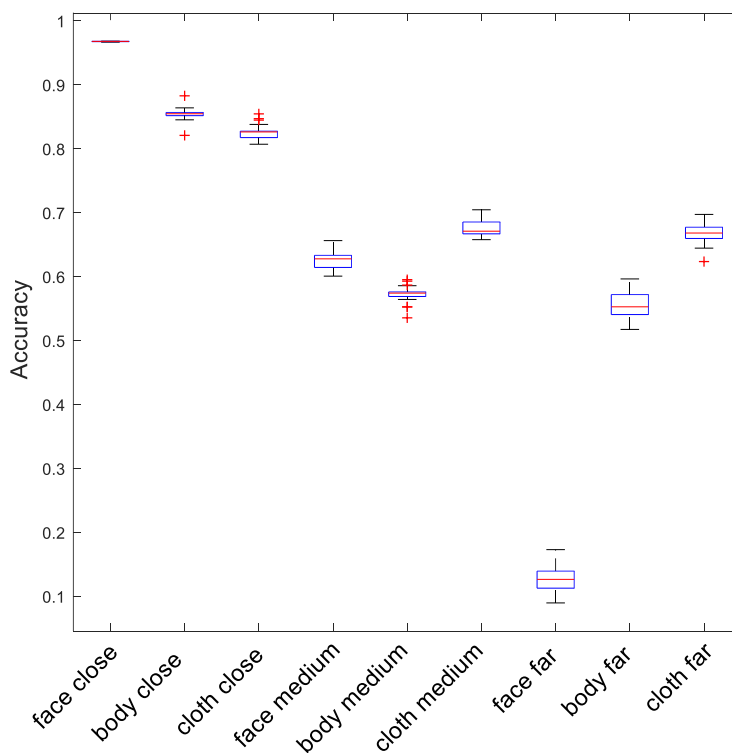


Figure 4.13 Accuracy of single-modal recognition (rank=1).

The process is repeated 20 times, and the boxplot of recognition results is shown Figure 4.13. It shows the extent of the accuracy over 9 single-modal methods. It shows that the recognition performance of clothing and body traits is relatively stable, whilst the performance of facial traits varies substantially at different distances. As we know, the resolution of face images is greatly influenced by distance. Thus, the recognition accuracy of facial traits decreases sharply at far distances.

The average recognition rates at three distances for each feature set are illustrated in Table 4.4. Recognition accuracy and equal error rate (EER) are employed to evaluate the performance of different biometric modals.

Table 4.4: Identification performance for single-modal methods

	Close		Medium		Far	
	Accuracy (%)	EER (%)	Accuracy (%)	EER (%)	Accuracy (%)	EER (%)
Face	95.7	0.45	62.7%	3.89	13.1%	22.76
Body	85.4	1.77	57.4%	3.96	55.5%	4.33
Clothing	82.6	2.76	69.4%	4.00	67.1%	3.59

4.4 Conclusions

This chapter introduces a new database for soft biometrics based on imagery collected from the University of Southampton Gait Tunnel, synthesised so as to appear to be from an outdoor environment, and then labelled using CrowdFlower. The influence of distance on soft biometric traits was firstly analysed. In terms of single-modal recognition, facial traits achieve the best result at close distance but are not stable, they fall sharply with an increase in distance. Compared with facial traits, as clothing and body traits have less uniqueness, the accuracies of the two trait sets are lower than that of facial traits at close distances, whereas the stability of body and clothing is much better than stability for the face. They can achieve good recognition results, even at a far distance.

Chapter 4

The following chapters further analyse soft biometric features through application of fusion. This research will evaluate the capability of the soft biometric traits in multi-modal recognition at different distances.

Chapter 5 Feature Level Fusion at Different Distances

In order to improve biometric system performance, information fusion is a key technique in multi-modal biometric systems. Multi-modal biometric fusion is conventionally divided into five levels: sensor, feature, score, rank and decision levels. In this chapter, we will conduct an analysis of the recognition performance with different statistical feature fusion methods.

Feature level fusion is based on the feature sets extracted from multiple data sources to create a new feature set that represents a subject. Therefore, the core goal is to describe the most effective feature information, so as to achieve superior recognition performance. The general idea is to minimise the distance of feature information for intra-class samples and maximise the distance of inter-class. Another important research area of feature fusion is how to extract effective information by minimising or removing redundancy.

Among a number of techniques implemented for feature level fusion, linear feature extraction methods are widely used to reduce the dimensionality of the feature set. Principal Component Analysis (PCA) is one of the most popular methods, used mainly for dimensionality reduction in compression and recognition problems [52] [53]. One method for feature level fusion, based on PCA [10], is to combine hand and face features. Another powerful dimensionality reduction technique is the Linear Discriminant Analysis (LDA) [54] [55]. A feature fusion method based on CCA is introduced in [11]. Another feature level fusion technique, DCA [12], improved CCA by incorporating class information into the correlation analysis of the feature sets. A multi-modal method based on sparse representation was proposed by Sumit, which significantly improved the robustness and accuracy [13].

This chapter will propose a supervised generalised canonical correlation (SG-CCA) method to fuse soft biometric features. The experiments were performed using the new soft biometric database, which contains the human face, body and clothing traits at three different distances. Furthermore, will explore the potential of face, body and clothing for human recognition using SG-CCA fusion compared with other linear dimensionality reduction fusion methods. The results will demonstrate the superiority of soft biometric fusion using SG-CCA method for human recognition.

5.1 Canonical Correlation Analysis

Canonical Correlation Analysis (CCA), proposed by Hotelling in 1936 [56], is a general method for studying the correlating linear relationship between two sets of random variables. The pairwise variables are employed together to investigate projections from two feature spaces that maximise the correlation between the projected representations [57]. By using CCA for feature fusion, the two datasets are regarded as two sets of variables.

Given two datasets X_1 and X_2 , each has n samples as $X_1 = [x_1^1, x_1^2, \dots, x_1^n]$ and $X_2 = [x_2^1, x_2^2, \dots, x_2^n]$, where x_j^i notes the i^{th} sample in the j^{th} set, $j = 1$ or 2 . We intend to map two datasets along two directions ω_1 and ω_2 that can maximise the correlation, project x_1 onto directions ω_1 can be described as:

$$x_1 \rightarrow \langle \omega_1, x_1 \rangle \quad (5.1)$$

$$X_1, \omega_1 = (\langle \omega_1, x_1^1 \rangle, \dots, \langle \omega_1, x_1^n \rangle) \quad (5.2)$$

and the corresponding value of x_2 is

$$X_2, \omega_2 = (\langle \omega_2, x_2^1 \rangle, \dots, \langle \omega_2, x_2^n \rangle) \quad (5.3)$$

The function that maximises the correlation between the two vectors is described as:

$$\begin{aligned} J(\omega_1, \omega_2) &= \max_{\omega_1, \omega_2} \text{corr}(X_1 \omega_1, X_2 \omega_2) \\ &= \max_{\omega_1, \omega_2} \frac{\omega_1^T C_{12} \omega_2}{\sqrt{\omega_1^T C_{11} \omega_1 \times \omega_2^T C_{22} \omega_2}} \end{aligned} \quad (5.4)$$

where C is the covariance matrix, specified as $C_{11} = X_1^T X_1$, $C_{22} = X_2^T X_2$ and $C_{12} = X_1^T X_2 = C_{21}^T$.

Since the optimisation of the objective function in Eq.(5.4) is invariant with respect to the scaling of ω_1 and ω_2 ($J(\omega_1, \omega_2) = J(\alpha\omega_1, \beta\omega_2)$) [57], assuming $\omega_1^T C_{11} \omega_1 = \omega_2^T C_{22} \omega_2 = 1$, the maximum of Eq.(5.4) can be computed by the following equations:

$$\begin{aligned} \omega_1 \omega_2 &= \operatorname{argmax} \omega_1^T C_{12} \omega_2 \\ \text{s. t. } \omega_1^T C_{11} \omega_1 &= \omega_2^T C_{22} \omega_2 = 1 \end{aligned} \quad (5.5)$$

By using the Lagrange multiplier method, Eq.(5.5) can be written as.

$$L(\lambda, \omega_1, \omega_2) = \omega_1^T C_{12} \omega_2 - \frac{\lambda_1}{2} (\omega_1^T C_{11} \omega_1 - 1) - \frac{\lambda_2}{2} (\omega_2^T C_{22} \omega_2 - 1) \quad (5.6)$$

Taking derivatives in respect to ω_1 and ω_2 ,

$$\frac{\partial L}{\partial \omega_1} = C_{12} \omega_2 - \lambda_1 C_{11} \omega_1 = 0 \quad (5.7)$$

$$\frac{\partial L}{\partial \omega_2} = C_{21} \omega_1 - \lambda_2 C_{22} \omega_2 = 0 \quad (5.8)$$

Let $\lambda = \lambda_1 = \lambda_2$

$$\begin{bmatrix} 0 & C_{12} \\ C_{21} & 0 \end{bmatrix} \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} = \lambda \begin{bmatrix} C_{11} & 0 \\ 0 & C_{22} \end{bmatrix} \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} \quad (5.9)$$

where λ is a Lagrange multiplier, which is equal to the correlation coefficient of $\omega_1^T x_1$ and $\omega_2^T x_2$. This is a standard eigenvalue problem and ω_1 and ω_2 can be determined by the eigenvector corresponding to the largest eigenvalue in Eq.(5.9).

5.2 Generalized Canonical Correlation Analysis

CCA finds a linear relationship between two views. It cannot be directly applied to multi-view data. In practical applications, the number of variables is always more than two. Thus, in this section, we will review a possible framework for generalising Canonical Correlation Analysis (gCCA), and the number of datasets is extended to m [58]. gCCA is used to investigate a set of directions

that maximise the total correlation. Each correlation is calculated by any two sets of variables.

Given m sets of variables in total, each set has n samples. The samples in one set are $X_j = [x_j^1, x_j^2, \dots, x_j^k, \dots, x_j^n]$, where j is one of the m ($j = 1, \dots, m$). x_j^i is i^{th} sample in j^{th} set of variables. The canonical correlation between j^{th} and k^{th} set of variables can be calculated using Eq.(5.4). gCCA is supposed aimed to derive $W = [\omega_1, \dots, \omega_m]^T$ that maximises the sum of correlations. Therefore Eq.(5.4) is extended to the following equation.

$$\omega = \operatorname{argmax} \left(\sum_{j=1}^m \sum_{k=i+1}^m \frac{\omega_j^T C_{jk} \omega_k}{\sqrt{\omega_j^T C_{jj} \omega_j \times \omega_k^T C_{kk} \omega_k}} \right) \quad (5.10)$$

Similar to CCA, the optimal value of W can be derived by solving the following constrained problem:

$$\begin{aligned} \omega &= \operatorname{argmax} (\sum_{j=1}^m \sum_{k=i+1}^m \omega_j^T C_{jk} \omega_k) \\ \text{s. t. } &\omega_j^T C_{jj} \omega_j = 1, \quad \forall i = 1, \dots, m \end{aligned} \quad (5.11)$$

Given $\omega_j^T C_{jk} \omega_k = \omega_k^T C_{kj} \omega_j$, we can transform Eq. (5.11) as follows:

$$\begin{aligned} W &= \operatorname{argmax} W^T \tilde{C} W \\ \text{s. t. } &W^T C_d W = I \\ &\omega_1^T C_{11} \omega_1 = \dots = \omega_m^T C_{mm} \omega_m \end{aligned} \quad (5.12)$$

where I is an $n \times n$ identity matrix and

$$\begin{aligned} \tilde{C} &= \begin{bmatrix} C_{11} & \dots & C_{1m} \\ \vdots & \ddots & \vdots \\ C_{m1} & \dots & C_{mm} \end{bmatrix} - C_d \\ C_d &= \begin{bmatrix} C_{11} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & C_{mm} \end{bmatrix} \end{aligned} \quad (5.13)$$

The optimal $W = [\omega_1, \dots, \omega_m]^T$ can be calculated through Eq.(5.12). By using Lagrange multipliers, Eq.(5.12) can be written as:

$$\tilde{C} \begin{bmatrix} \omega_1 \\ \vdots \\ \omega_m \end{bmatrix} = \lambda C_d \begin{bmatrix} \omega_1 \\ \vdots \\ \omega_m \end{bmatrix} \quad (5.14)$$

5.3 Supervised Generalized Canonical Correlation Analysis

As CCA only focuses on the correlations between the pairwise variables from different datasets, the class information of the samples is ignored, which adversely affects the recognition performance [59]. In order to overcome this issue, this chapter proposes a supervised generalised canonical correlation analysis (SG-CCA) by incorporating label information of samples into the gCCA model. The SG-CCA is a supervised method for multi-view data. It is easier to be used in practical applications and the result is theoretically improved compared with gCCA.

SG-CCA is supposed to provide a set of projection directions, $W = [\omega_1, \dots, \omega_n]^T$, to maximise intra-class correlations between two different datasets (as justified in [60], the inter-class correlation is automatically minimised when intra-class correlation is maximised).

Before we derive SG-CCA, supervised CCA is described. Given two sets of variables X_1 and X_2 , each set has n pairs of samples and is labelled as c classes. We define α as the number of classes ($\alpha = 1, \dots, c$), and each class has n_α samples ($\sum_{\alpha=1}^c n_\alpha = n$). The samples are listed as $(x_{1,i}^{(\alpha)}, x_{2,i}^{(\alpha)})$, where $i = [1, 2, \dots, n_\alpha]$ and $x_{m,i}^{(\alpha)}$ denotes i^{th} sample in α^{th} class at m^{th} dataset. Expressing $X_1^{(\alpha)}$ and $X_2^{(\alpha)}$ as the sample sets of α^{th} class, and $X_1^{(\alpha)} = [x_{1,1}^{(\alpha)}, \dots, x_{1,n_\alpha}^{(\alpha)}]$ and $X_2^{(\alpha)} = [x_{2,1}^{(\alpha)}, \dots, x_{2,n_\alpha}^{(\alpha)}]$. $\widehat{X}_1^{(\alpha)}$ and $\widehat{X}_2^{(\alpha)}$ are used to represent the variables, $X_1^{(\alpha)}$ and $X_2^{(\alpha)}$, after projection on to directions ω_X and ω_Y . The intra-class correlation between $\widehat{X}^{(\alpha)}$ and $\widehat{Y}^{(\alpha)}$ can be then defined as:

$$\begin{aligned} \rho^{(\alpha)} &= \frac{\widehat{X}_1^{(\alpha)} \widehat{X}_2^{(\alpha)T}}{\sqrt{\widehat{X}_1^{(\alpha)} \widehat{X}_1^{(\alpha)T} \times \widehat{X}_2^{(\alpha)} \widehat{X}_2^{(\alpha)T}}} \\ &= \frac{\omega_{X_1}^T X_1^{(\alpha)} X_2^{(\alpha)} \omega_{X_2}}{\sqrt{\omega_{X_1}^T X_1^{(\alpha)} X_1^{(\alpha)T} \omega_{X_1} \times \omega_{X_2}^T X_2^{(\alpha)} X_2^{(\alpha)T} \omega_{X_2}}} \end{aligned} \quad (5.15)$$

ω_X and ω_Y can be calculated by maximising the sum of all the intra-class correlations which is denoted as:

$$\omega_X, \omega_Y = \operatorname{argmax} \left(\sum_{\alpha=1}^c \frac{\omega_{X_1}^T X_1^{(\alpha)} X_2^{(\alpha)} \omega_{X_2}}{\sqrt{\omega_{X_1}^T X_1^{(\alpha)} X_1^{(\alpha)T} \omega_{X_1} \times \omega_{X_2}^T X_2^{(\alpha)} X_2^{(\alpha)T} \omega_{X_2}}} \right) \quad (5.16)$$

Eq.(5.16) is specified as following,

$$\begin{aligned} \omega_X, \omega_Y &= \operatorname{argmax} \left(\omega_{X_1}^T \sum_{\alpha=1}^c X_1^{(\alpha)} X_2^{(\alpha)} \omega_{X_2} \right) \\ \text{s. t. } \omega_{X_1}^T X_1^{(\alpha)} X_1^{(\alpha)T} \omega_{X_1} &= 1 \\ \omega_{X_2}^T X_2^{(\alpha)} X_2^{(\alpha)T} \omega_{X_2} &= 1 \\ i &= 1, \dots, c \end{aligned} \quad (5.17)$$

Then supervised CCA is extended to the generalised case, namely SG-CCA. Given m different datasets, each one has n samples labelled into c classes, Given α as the number of classes ($\alpha = 1, \dots, c$) and n_α denotes the total number of training samples in each class ($\sum_{\alpha=1}^c n_\alpha = n$). In order to clarify the samples, i is used to denote the number of samples in one class ($i = 1, \dots, n_\alpha$). The number of dataset is represented by j , and j is in the range of 1 to m . Those samples in α^{th} class in different dataset are given as $X^{(\alpha)} = \{(X_1^{(\alpha)}, \dots, X_j^{(\alpha)}, \dots, X_m^{(\alpha)})\}$, $X_i^{(\alpha)} = \{(x_{j,1}^{(\alpha)}, \dots, x_{j,i}^{(\alpha)}, \dots, x_{j,m_\alpha}^{(\alpha)})\}$.

We combine all the intra-class correlations together and maximise the sum. The formulation of SG-CCA is as follows:

$$\begin{aligned} W &= \operatorname{argmax} \left(\sum_{\alpha=1}^c W^T \tilde{C}^{(\alpha)} W \right) \\ \text{s. t. } W^T C_d^{(\alpha)} W &= 1 \end{aligned} \quad (5.18)$$

where

$$\tilde{C}^{(\alpha)} = \begin{bmatrix} C_{11}^{(\alpha)} & \dots & C_{1m}^{(\alpha)} \\ \vdots & \ddots & \vdots \\ C_{m1}^{(\alpha)} & \dots & C_{mm}^{(\alpha)} \end{bmatrix} - \begin{bmatrix} C_{1m}^{(\alpha)} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & C_{mm}^{(\alpha)} \end{bmatrix} \quad (5.19)$$

$$C_d^{(\alpha)} = \begin{bmatrix} C_{11}^{(\alpha)} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & C_{mm}^{(\alpha)} \end{bmatrix} \quad (5.20)$$

5.4 Experiments and analysis

5.4.1 Feature level fusion using supervised generalized canonical correlation analysis

This section will validate the performance of the proposed SG-CCA algorithm with different dimensions. The feature sets used are the same as those described in 0. The experiment was implemented using 200 subjects, and each subject had 20 samples, in which 10 samples were randomly chosen as testing data to train SG-CCA and obtain the transformation matrices. The remaining 10 samples were regarded as testing data and were employed for evaluation.

Table 5.1: Comparison of recognition results through SG-CCA with different numbers of features (d : number of features).

	close	medium	far
SG-CCA($d=5$)	95.5%	78.0%	66.6%
SG-CCA($d=7$)	98.4%	86.9%	69.3%
SG-CCA($d=10$)	99.6%	92.3%	75.2%

The recognition accuracies with different numbers of features are shown in Table 5.1 and Figure 5.1. The results show that accuracy is proportional to the number of features used. When the number of features is 5, the average recognition accuracy achieves 95.5% at close distance, which is similar to the accuracy of single face modality at close distance, but higher than that of body and clothing traits. With increase in dimensions for training, the accuracy at a close distance increases to 98.4%, which is higher than all the results of single-modal methods. Clothing traits achieved the highest recognition rate at medium distance, which is increased by 17.5% using SG-CCA. It can be also concluded that high dimensional fusion contributes to the distinctive recognition accuracies. For example, the accuracy is boosted to 99.6% at a close distance when the number of features equals 7. This phenomenon is clearly shown in the Figure 5.1.

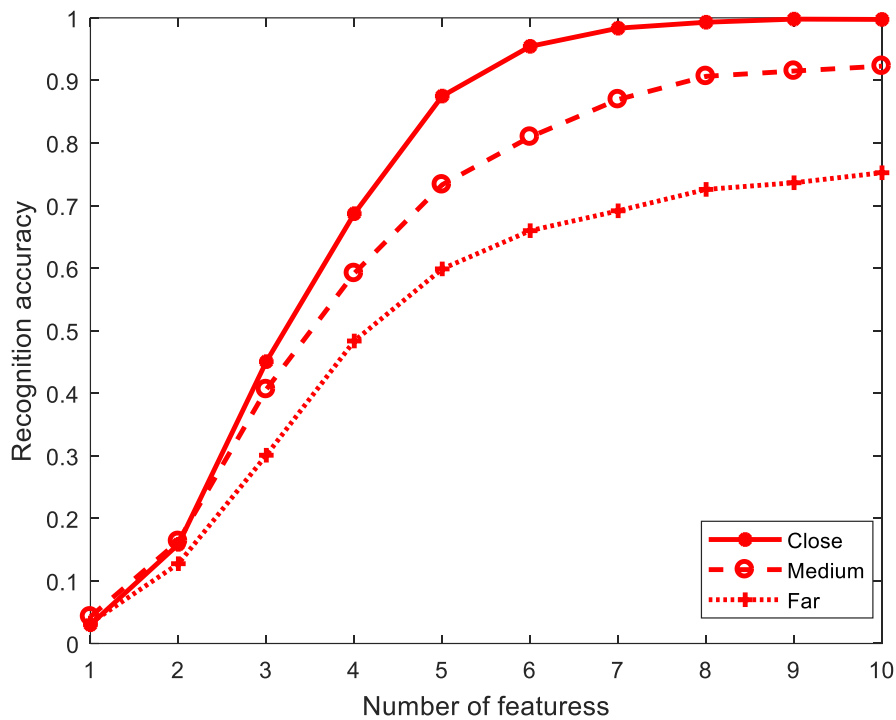


Figure 5.1 Recognition accuracy of the proposed SG-CCA algorithm using different number of features.

5.4.2 Comparison with other linear dimensionality reduction method for fusion

As multi-modal methods aim to fuse several single modalities for recognition, the weak (face traits at a far distance) and strong (face traits at close distance) modalities are normally considered together. However, they are not expected to obtain an inferior performance due to weak modalities. The experiments in this section will evaluate the robustness of various linear dimensionality reduction methods (PCA, LDA, gCCA and SG-CCA).

The recognition accuracy at three distances using different feature fusion methods is listed in Table 5.2. The proposed method clearly outperforms other feature level fusion techniques. At a close distance, the recognition accuracy of the fusion method is always superior to that of single-modal methods, except for the result of PCA, which is the same as that of face traits. At a medium distance, all the fusion methods achieve excellent recognition rate, compared with the best result given by single-modal method - 69.4%. LDA

improves the accuracy by 4.4%, while SG-CCA contributes to the highest rate - 86.9%. At a far distance, the accuracy of PCA and LDA is higher than that of a single face or body traits, but lower than that of single clothing features, because the accuracy of the face at a far distance is only 13.1%, which lowers the fusion results. Alongside this, the fusion results of gCCA are slightly improved, and the accuracy of the proposed SG-CCA increases to 69.3%.

Table 5.2 Identification performance using different methods (feature number=7).

	Close		Medium		Far	
	Accuracy (%)	EER (%)	Accuracy (%)	EER (%)	Accuracy (%)	EER (%)
Face	95.7	0.45	62.7	3.89	13.1	22.76
Body	85.4	1.77	57.4	3.96	55.5	4.33
Clothing	82.6	2.76	69.4	4.00	67.1	3.59
PCA	95.8	0.40	74.3	2.77	61.0	4.05
LDA	96.4	0.51	73.8	2.91	61.6	3.00
gCCA	97.1	0.39	82.3	2.06	67.9	2.98
SG-CCA	98.4	0.37	86.9	1.53	69.3	2.79

The development of performance fusion at close and far distances is not apparent as medium distance, as single modalities can achieve desirable results at a close distance, while facial attributes might reduce the fusion performance at a far distance. However, the numerical evidence shows that the proposed SG-CCA performs the best at all three distances, especially at close and medium distances, and with a slight improvement (over clothing) at the far distance.

5.5 Conclusions

In this chapter, a novel fusion method named SG-CCA was studied. When using SG-CCA for fusion, increasing the number of features helped to achieve higher recognition accuracy. The experiments were also implemented with other multi-modal fusion methods (PCA, LDA, and gCCA) for comparison. Recognition accuracy and EER were employed to evaluate performance. The comparisons between multi-modal and the single-modal biometric methods

demonstrated that recognition performance can be improved by multi-biometrics. The proposed SG-CCA was validated numerically and shown to be the best fusion method available, especially at close and medium distances.

In summary, this chapter demonstrated the improved performance at different distances using soft biometrics feature level fusion. A method for score level and rank level fusion will be investigated in the next chapter.

Chapter 6 Score Level and Rank Level Fusion at Different Distances

In order to achieve high recognition accuracy, multi-modal fusion is frequently employed in recognition systems. In this section, we will focus on score and rank level.

Normally, the identification system provides two types of result: the match scores of enrolled subjects and the subjects' ranks. Match scores are used to measure the similarity between testing and enrolled subjects. Score level fusion is a combination of match scores from different biometric matchers, which then derives a new score. Some simple methods, such as product rule, sum rule, max, medium and minimum rules, were introduced [14]. These methods could be implemented readily, since no statistical information is required. Part of score level fusion is based on the match score density distribution. A combination method using the likelihood ratio test is proposed in [61] which estimates the genuine and impostor matching scores, and then calculates likelihood ratios for each component modality. Similarly, the Bayesian approach [14] is also widely used for score level fusion. A SVM based score level fusion is introduced and validated in [15], which demonstrates that the weighted score level fusion can achieve a higher accuracy with lower EER compared with individual modalities [16].

A rank list of enrolled subjects is another output of a biometric matcher. The goal of rank level fusion is to derive a mutually agreed rank for each identity through the consolidation of the rank's output by individual identification models. Regarding rank level fusion, a number of techniques have been researched recently. A review of different rank level fusion approaches was reported in [62] [63]. Highest Rank and Borda Count are frequently employed

to fuse ranks, since they do not require training or prior statistical information. The Borda Count allocates weight for different matchers based on their recognition performances. Another fusion method proposed in [64] generalised the Mallows model on permutations. A mixed group ranks method is introduced in [65], which is a combination of Borda Count, Logistic Regression, and Highest Rank methods. The experiments demonstrate that this method can provide superior recognition performance. An effective rank fusion scheme for multi-modal biometric is proposed in [66] using Markov chain, and it achieved a high performance by fusing the face, ear, and iris. A nonlinear method is studied in [67] for ranking combination. An unsupervised rank fusion method, Inverse Square Rank fusion, is proposed in [68]. The algorithm is based on quadratic decay and logarithmic document frequency normalisation. However, it should be noted that rank fusion requires access to the whole set of results while score fusion does not, which can be regarded as a disadvantage of the approach.

6.1 Score level fusion

6.1.1 Estimation of similarity score densities

There are three databases for human body, face and clothing, and the Parzen Window is used to estimate the PDF of each. In Eq. (6.1), a Gaussian kernel function is used to smooth the estimation result.

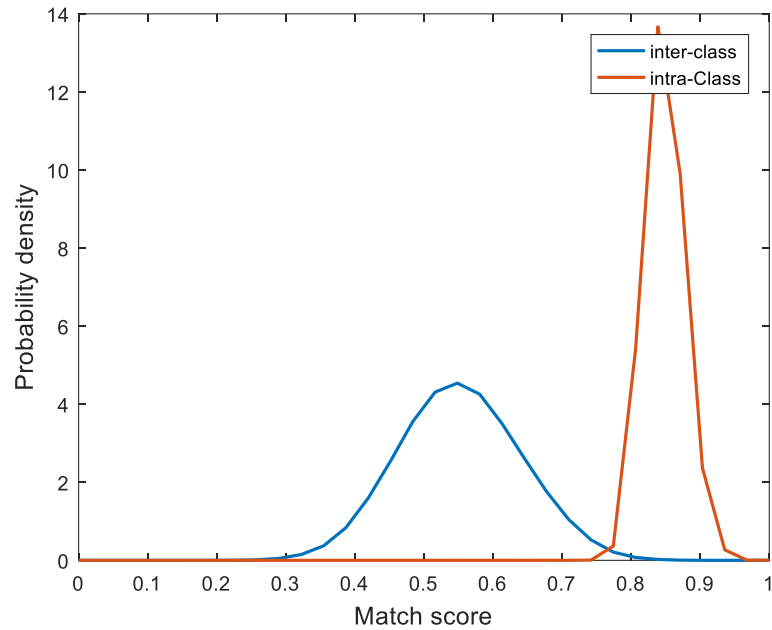
$$p(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \left(-\frac{(x_i-x)^2}{2\sigma^2} \right) \quad (6.1)$$

where x_1, \dots, x_n are n data samples and x is the centre point, σ is variance.

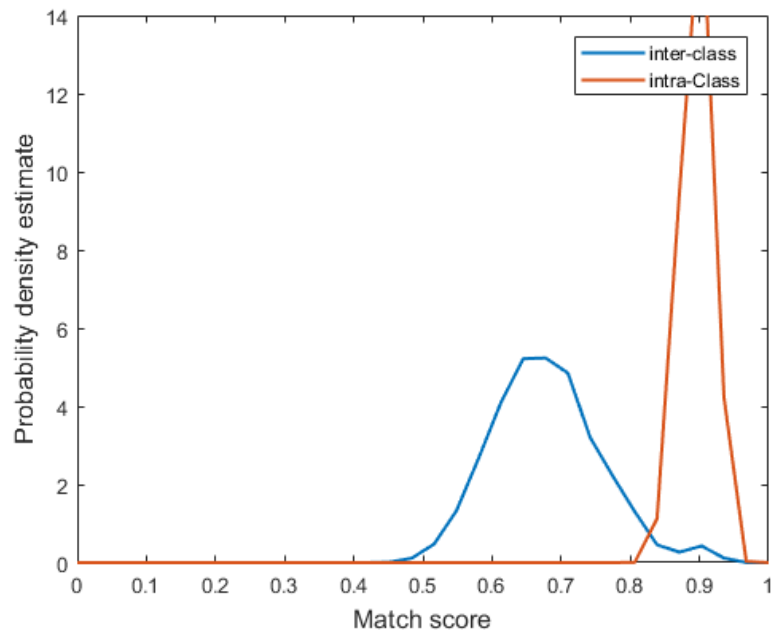
In our new database, each of the 200 subjects has 20 samples. For each subject, the intra-class similarity score can be obtained by comparing 20 samples of the same subject, whilst the inter-class similarity score is calculated by comparing any two samples from different subjects. By considering the similarity scores of intra-class and inter-class as the inputs of a Parzen Window, we can estimate two PDFs for intra-class and inter-class.

Figure 6.1 depicts an example probability density distribution of the face datasets at three different distances when using a Parzen Window. The close

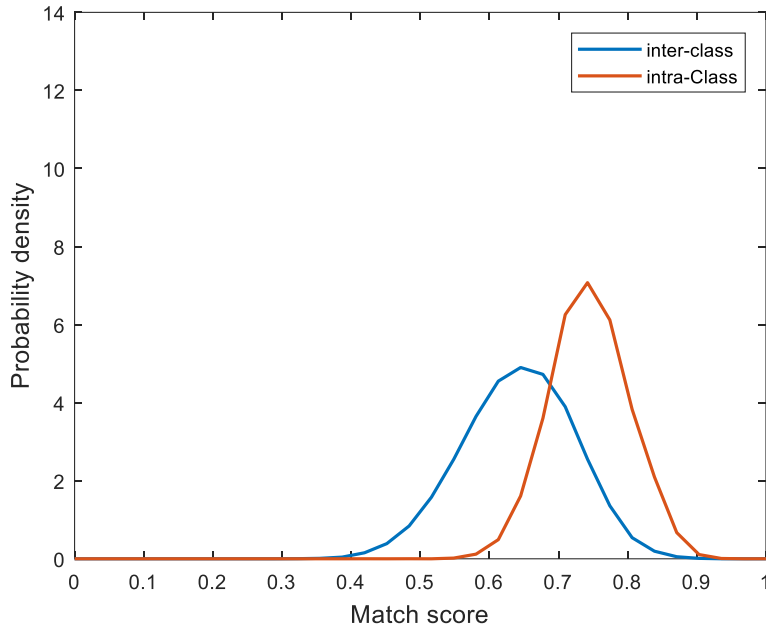
face density distribution appears to offer best recognition in our dataset - the overlap between inter-class and intra-class is small, which is why recognition by face at close distances achieves the best result. The face at a far distance is the worst since the overlap is much larger than that of the other two.



(a) Close



(b) Medium



(c) Far

Figure 6.1 Probability density of face match score at three distance, estimated using a Parzen window

6.1.2 Score fusion using Bayesian theory

As one of the most popular biometric recognition methods, score-level fusion, has satisfactory performance in traditional biometric recognition, such as logistic regression. By combining several match scores given by single-modal methods, score-level fusion can significantly improve recognition accuracy.

It is well known that Bayesian analysis (Eq.(6.2)) is able to enhance the posterior probability by means of the prior probability.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (6.2)$$

where $P(A)$ and $P(B)$ are the probabilities of A and B occurring independently and $P(A|B)$ represents the probability of event A occurring on condition of the fact that B is true. Similarly, $P(B|A)$ denotes the probability of event B occurring on condition of the fact that A is true.

The identification of an input sample can be seen as a two-class (genuine or impostor user) issue. For any given match score, the probability of this score belonging to a genuine or an impostor can be calculated by Eq. (6.3) and (6.4).

$$P(g|S) = \frac{p(S|g)P(g)}{p(S)} \quad (6.3)$$

$$P(i|S) = \frac{p(S|i)P(i)}{p(S)} \quad (6.4)$$

$$p(S) = p(S|g)P(g) + p(S|i)P(i) \quad (6.5)$$

where g represents genuine and i represents the impostor event, $P(g|S)$ describes the probability that the event g is true on condition of the event S . $p(S|g)$ denotes a distribution of event S on condition of event g .

For multi-biometrics, there are several independent matchers. The probability of a score belonging to a genuine or an impostor is denoted in Eq.(6.6), where N is the number of sample matchers

$$P(k|S_1, \dots, S_M) = \prod_{j=1}^N P(k|S_j) \quad k \in g, i \quad (6.6)$$

The final judgement can be made through Eq.(6.7). An output equalling 1 indicates that the match score belongs to a genuine user, while the score belongs to an impostor user if the output is 0.

$$D = \begin{cases} 1 & P(g|S_1, \dots, S_M) \geq P(i|S_1, \dots, S_M) \\ 0 & \text{otherwise} \end{cases} \quad (6.7)$$

6.1.3 Score fusion using Likelihood Ratio Test

According to Neyman-Pearson theory [69], given a FPR, when the FNR reaches the minimum, the Likelihood Ratio Test (LRT) is represented by:

$$\lambda(s) = \frac{f_{gen}(s)}{f_{imp}(s)} \quad (6.8)$$

where s is the input score, $f_{gen}(s)$ is the densities of genuine training data and $f_{imp}(s)$ is the impostor. In Eq.(6.9), D is the verification result and η is the classification threshold, which is determined based on a given FAR.

$$D = \begin{cases} \text{genuine class} & \text{if } \lambda(s) \geq \eta \\ \text{impostor class} & \text{if } \lambda(s) < \eta \end{cases} \quad (6.9)$$

The vector $[s_1, \dots, s_N]$ denotes a set of match scores calculated through N different biometric matchers,

$$\lambda(s) = \frac{\prod_{i=1}^N f_{gen}(s_i)}{\prod_{i=1}^N f_{imp}(s_i)} = \prod_{i=1}^N \lambda(s_i) \quad (6.10)$$

6.1.4 Score fusion using SVM-weighted Likelihood Ratio Test

The logarithm of the likelihood ratio is denoted as:

$$\rho(s) = \log(\lambda) = \sum_{i=1}^N \rho(s_i) \quad (6.11)$$

and the weighted likelihood ratio is defined as:

$$\lambda_w(s) = \prod_{i=1}^N \lambda_w(s_i) = \prod_{i=1}^N (\lambda(s_i))^{w_i} \quad (6.12)$$

where w_i is the weight of the i^{th} matcher and each $\lambda_w(s_i)$ is computed by $(\lambda(s_i))^{w_i}$. The logarithm of the weighted likelihood ratio is represented as:

$$\begin{aligned} \rho_w(s) &= \sum_{i=1}^N w_i \rho(s_i) = [w_1 \quad \dots \quad w_N] \begin{bmatrix} \rho(s_1) \\ \vdots \\ \rho(s_N) \end{bmatrix} \\ &= W^T P \end{aligned} \quad (6.13)$$

Given a threshold θ , the weighted likelihood ratio test can be written as:

$$D = W^T P - \theta \quad (6.14)$$

If $D \geq 0$, the result is accepted, otherwise it is rejected. We use a kernel SVM to optimise W and θ and then obtain the optimal verification, called the SVM-weighted likelihood ratio test (SVM-LRT). An SVM was proposed to investigate an optimal hyperplane that can accurately classify samples into two classes [70]. Given a training set $S = \{x_i, y_i\}_{i=1}^m$, with input the vector $x_i \in R^n$, and $y_i \in \{-1, +1\}$, the hyperplane for a linear case is defined as:

$$g(x) = \omega^T x + b = 0 \quad (6.15)$$

$$y_i(x_i \omega + b) - 1 \geq 0, \forall i \in [1, \dots, m] \quad (6.16)$$

A hyperplane with maximum margin (the maximum distance between closest samples and hyperplane) can be calculated by solving the optimisation problem:

$$\begin{aligned} \omega, b &= \operatorname{argmax} \frac{2}{\|\omega\|} \\ \text{s. t. } & y_i(x_i\omega + b) - 1 \geq 0, \forall i \in [1, \dots, m] \end{aligned} \quad (6.17)$$

$\max_{\omega, b} \frac{2}{\|\omega\|}$ can be represented as $\min_{\omega, b} \frac{1}{2} \|\omega\|^2$. Using the Lagrange multiplier for it, the Lagrange function can be described as:

$$L(\omega, b, \alpha) = \frac{1}{2} \|\omega\|^2 - \sum_{i=1}^m \alpha_i [y_i(x_i\omega + b) - 1] \quad (6.18)$$

where α_i is Lagrange variables and $\alpha_i \geq 0 \forall i \in [1, \dots, m]$.

This quadratic optimisation problem can be solved by removing the gradient of the Lagrange function with respect to the variables ω and b .

$$\nabla_{\omega} L = \omega - \sum_{i=1}^m \alpha_i y_i x_i = 0 \Rightarrow \omega = \sum_{i=1}^m \alpha_i y_i x_i \quad (6.19)$$

$$\nabla_b L = -\sum_{i=1}^m \alpha_i y_i = 0 \Rightarrow \sum_{i=1}^m \alpha_i y_i = 0 \quad (6.20)$$

$$\begin{aligned} \forall i, \alpha_i [y_i(x_i\omega + b) - 1] &= 0 \Rightarrow \\ \alpha_i &= 0 \vee y_i(x_i\omega + b) - 1 = 0 \end{aligned} \quad (6.21)$$

Then, ω in Eq.(6.18) can be replaced by Eq.(6.19), Eq.(6.18) is therefore rewritten as,

$$\begin{aligned} \alpha &= \operatorname{argmax} \left(\sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \right) \\ \text{s. t. } & \sum_{i=1}^m \alpha_i y_i = 0 \\ & \alpha_i \geq 0, \forall i \in [1, \dots, m] \end{aligned} \quad (6.22)$$

According to Eq.(6.19) and (6.21), the optimal values for ω and b can be calculated by α , which is the solution of Eq.(6.22).

Nevertheless, the real data cannot be classified using linear classifier. Thus, a non-linear kernel is used to map the data into a high-dimensional space. The optimal hyperplane can be therefore obtained by solving Eq. (6.23), which is regarded as an extension of Eq.(6.22).

$$\begin{aligned}
\alpha &= \operatorname{argmax} \left(\sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j K(x_i \cdot x_j) \right) \\
s. t. \quad &\sum_{i=1}^m \alpha_i y_i = 0 \\
&\alpha_i \geq 0, \forall i \in [1, \dots, m]
\end{aligned} \tag{6.23}$$

where $K(x_i \cdot x_j)$ is the kernel function, selected based on the dataset.

6.1.5 Experiments and discussion

In the testing dataset, each subject has 20 samples, and half of them are randomly chosen as testing data, and the remaining samples are regarded as training data. During testing, one sample in the testing dataset is used to match any one randomly chosen sample in the training dataset. The probability distribution of this similarity score for genuine and impostor can be obtained based on the two probability densities obtained from the training sample.

The results of multi-modal recognition using Bayesian theory, likelihood ratio test and SVM-weighted likelihood ratio test are listed in Table 6.1. Recognition accuracy, and EER are employed to evaluate the performance of different biometric models.

Table 6.1: Identification performance using different methods

	Close		Medium		Far	
	Accuracy (%)	EER (%)	Accuracy (%)	EER (%)	Accuracy (%)	EER (%)
Face	95.7	0.45	62.7	3.89	13.1	22.76
Body	85.4	1.77	57.4	3.96	55.4	4.33
Clothing	82.5	2.76	69.4	4.00	67.0	3.59
Bayes	96.3	0.38	84.6	1.07	78.1	2.57
LRT	96.1	1.07	87.7	2.46	76.5	3.02
SVM-LRT	97.0	0.32	92.5	0.89	80.8	2.08

It is clear that recognition accuracy is improved greatly using three fusion algorithms, in comparison with three single-model methods. The advantages of the proposed methods are more prominent in the experiments at medium and far distances.

According to EER, *Gain* is used to quantify the improvement of EER using fusion methods, in comparison with single-model methods. The improvement factor *Gain* is defined as Eq.(6.24).

$$Gain = \frac{\min(EER_1, \dots, EER_N)}{EER_f} \quad (6.24)$$

where EER_1, \dots, EER_N are the EER of N single-mode biometric identification methods, and EER_f is the EER of fusion system. The relevant results are listed in Table 6.2.

Table 6.2 *Gain* of different score fusion methods

	Close	Medium	Far
Bayesian	1.18	3.64	1.40
LRT	0.42	1.58	1.19
SVM-LRT	1.41	4.37	1.73

In comparison with a fusion based on LRT, the algorithms based on Bayesian theory and SVM-LRT both achieve superior results. Bayesian theory not only considers the distribution of match score, but also incorporates prior probability, both of which contribute to an excellent recognition result. In terms of fusion based on SVM-LRT, the optimal threshold and weights for various distances can be derived through training sets to improve the performance of classification. For example, at a close distance, the weight of face is enhanced, while those of body and clothing are reduced.

In comparison with [9], the database was comprised of 40 subjects and 27 faces, 19 body and 7 clothing traits, which were represented by comparative distributions. The data did not use the concepts of distance from camera exposed in this chapter, rather opting to delete the facial traits when they cannot be seen. In terms of the result of the fusion recognition, although the number of facial and body traits is lower, and the total number of subjects has increased to 200, this chapter's approach performs better than that of [9].

6.2 Rank Level fusion

6.2.1 Borda count method

Borda count [71] [72] is an unsupervised rank level fusion method with many possible applications. Given m different matchers, each has n enrolled samples on training dataset. A vote is given to each rank. For example, the rank-one of each matcher is given N votes, and rank-two is given $N - 1$. Also, the rank- n will be assigned 1 vote. Then the sum of votes of individual matchers for one user is used for identification. For example, the final vote for user k is:

$$s(k) = \sum_{j=1}^m v_{j,k} \quad (6.25)$$

where $v_{j,k}$ is the vote for user k given by j^{th} matcher. The winner - the true user identity - is determined based on the highest votes.

The Borda count methods assume that the votes assigned to the user by different matchers are independent, and the performance of matchers is the same. The training phase is not required.

6.2.2 Logistic regression method

The Logistic Regression Method [73] is a generalisation of Borda count where the expected performance of different matchers is not the same. It can be formulated as the weighted Borda count, i.e. the statistic for user k is:

$$s(k) = \sum_{j=1}^m \omega_j v_{j,k} \quad (6.26)$$

The weight, ω_j , for j^{th} matcher can be calculated using logistic regression, empirical values or more sophisticated machine learning approaches. Compared with Borda count, a training phase is required for weight calculation.

6.2.3 Nonlinear weight ranks method

A nonlinear method to calculate the weights of different matchers is investigated in [67]. Three nonlinear combinations are introduced to generate the consolidated ranks; the equations are listed as follows, where ω_j is weight

for j^{th} matcher and $r_{j,k}$ is the k^{th} rank given by j^{th} matcher. The weights reflect the significance of different matchers.

$$s(k) = \sum_{j=1}^m \tanh(\omega_j r_{j,k}) \quad (6.27)$$

$$s(k) = \sum_{j=1}^m \exp(\omega_j r_{j,k}) \quad (6.28)$$

$$s(k) = \sum_{j=1}^m \omega_j \exp(r_{j,k}) \quad (6.29)$$

6.2.4 PAV based rank fusion

In a new rank fusion method, based on Pool Adjacent Violators (PAV) algorithm [74], rank is converted for each matcher to an approximate score using Eq.(6.30),

$$s_{j,k} = \frac{(n_{train} + 1 - r_{j,k})}{n_{train}} \quad (6.30)$$

where n_{train} is the total number of identities in the training dataset and $r_{j,k}$ is the rank for user k given by j^{th} matcher. The score after conversion is within the range from $\frac{1}{n_{train}}$ to 1. The log likelihood ratio was then calculated for each score value using the PAV algorithm.

The sum of a score of individual matches for one user is used for identification. For example, the final score for the k^{th} rank is:

$$s(k) = \sum_{j=1}^m s_{j,k}^{LR} \quad (6.31)$$

where $s_{j,k}^{LR}$ is the score after computing the log likelihood ratio. The winner's identity is determined based on the highest score.

6.2.5 Experiments and discussion

The recognition results at three distances using different score-level fusion methods are listed in Table 6.3. The progress is the same as in the rank-level fusion test. For each subject, half of 20 samples are randomly chosen to be used as testing data, and the remaining samples used as training data.

Table 6.3 Identification performance using different rank-fusion methods

	<i>Close</i>		<i>Medium</i>		<i>Far</i>	
	<i>Accuracy</i> (%)	<i>EER</i> (%)	<i>Accuracy</i> (%)	<i>EER</i> (%)	<i>Accuracy</i> (%)	<i>EER</i> (%)
Face	95.7	0.45	62.7	3.89	13.1	22.76
Body	85.4	1.77	57.4	3.96	55.5	4.33
Clothing	82.6	2.76	69.4	4.00	67.1	3.59
<i>Borda count method</i> [72]	95.8	0.45	76.4	3.64	73.3	4.17
<i>Logistic regression</i> [73]	96.4	0.39	82.3	3.73	75.5	3.83
<i>Nonlinear weight ranks</i> [67]	96.9	0.39	86.2	3.48	79.3	3.44
<i>PAV based</i> [74]	97.0	0.38	86.0	3.01	79.1	3.33

It is clear that the nonlinear weight ranks and PAV based rank fusion outperform the other two rank fusion techniques at medium and far distances. At a close distance, all the fusion methods achieve excellent recognition rates. Compared with the best results given by single-mode methods (95.8%), PAV based fusion improves the accuracy by 1.2%. The accuracy improvement is also prominent when using Borda count method. At a medium distance, the recognition performance of nonlinear weight rank fusion improves significantly. The accuracy increases to 86.2%. And the EER achieves 14.1%. At a far distance, the accuracy improvement is also satisfactory, especially for nonlinear weight ranks and PAV based rank fusion.

Table 6.4 *Gain* of different score fusion methods

	Close	Medium	Far
Borda count method	1	1.07	0.86
Logistic regression [73]	1.15	1.04	0.94
Nonlinear weight ranks [67]	1.15	1.12	1.04
PAV based [74]	1.18	1.29	1.09

Gain of different score fusion methods are listed in Table 6.4, compare the *Gain* of score and rank level fusion, the results of score level fusion methods outperform rank fusion techniques.

6.3 Conclusions

Three different fusion methods in rank and score level were introduced, and their performance was validated by three sets of soft biometric datasets. Results of the comparison between multi-model biometric methods with single-model biometric methods showed that recognition performance was significantly improved by using multi-model biometrics. The proposed SVM-LRT was numerically demonstrated to be the best fusion method available, with a recognition accuracy of over 98% at all distances. Clearly, the performance can be improved by fusion but the current approach requires access to all the data. As such it appeared prudent to investigate an approach which could deliver the same performance, but without this restriction.

Chapter 7 Rank-score Fusion

The results presented in the Chapter 6 demonstrated the effectiveness of fusion in rank and score level using soft biometrics for subject identification. The aforementioned methods demonstrated that recognition performance can be improved by rank or score fusion. In this chapter, an improved fusion method based on rank and score level fusion will be proposed. The difference between testing samples and enrolled samples can be observed intuitively using similarity scores, based on which ranks are sorted. However, the rank and similarity score information is different. Since rank is a linear description (i.e. 1, 2, 3, ...), it can be used to indicate the order of enrolled samples, but does not describe the variation between adjacent samples. A novel technique will, therefore, be proposed to combine the effective information in rank and similarity scores, namely rank-score fusion, and then to consolidate the recognition results.

7.1 Rank-score distribution

The outputs of a classifier are always comprised of a match score list or a rank list. The match score describes a distance or a similarity between the testing subject and the registered subject. The distance between subjects is normally calculated by Euclidean distance, Mahalanobis Distance or other metrics. The rank list is then obtained by sorting the similarity score in descending order.

Given k samples as registered samples for each matcher, a rank-score distribution can be constructed after training. In order to reduce the influence of outliers on the experiment and achieve more reliable results, each subject is used to match all remaining subjects. During the training process, similarity score matrix $s_i = \{s_{i,1}, \dots, s_{i,k-1}\}$, and rank matrix $r_i = \{r_{i,1}, \dots, r_{i,k-1}\}$ are obtained,

where i is the subject ID in the range of 1 to k , and $s_{i,n}$ is the similarity score between i^{th} subject with n^{th} remaining subject. The variables in the rank and score matrices are in one-to-one correspondence ($r_{i,n}$ is the rank order of $s_{i,n}$). For multi-modal biometrics, there are m different matchers, each of which produces one rank matrix and one score matrix. $\mathcal{S} = \{[s_{i,1}^j, \dots, s_{i,k-1}^j]\}$ and $\mathcal{R} = \{[r_{i,1}^j, \dots, r_{i,k-1}^j]\}$ are used to denote the sets of scores and rank matrices separately, where j is the number of biometric matcher ($j = 1, \dots, m$) and $i = 1, \dots, k$.

A two-dimensional density distribution with two corresponding variables in \mathcal{S} and \mathcal{R} is estimated. The calculation is constructed using a Gaussian Kernel function to smooth the result. The density function with rank and score is estimated by:

$$p(s, r) = \frac{1}{2\pi\sigma_r\sigma_s} \exp \left[- \left(\frac{(r-\mu_r)^2}{\sigma_r} + \frac{(s-\mu_s)^2}{\sigma_s} \right) \right] \quad (7.1)$$

Modality	Gallery	Matching score	Rank
Face	X_1	$sim(X_1, S_i)=0.3$	3
	X_2	$sim(X_2, S_i)=0.8$	1
	X_3	$sim(X_3, S_i)=0.4$	2
Body	Y_1	$sim(Y_1, S_i)=0.5$	3
	Y_2	$sim(Y_2, S_i)=0.6$	2
	Y_3	$sim(Y_3, S_i)=0.9$	1
Clothing	Z_1	$sim(Z_1, S_i)=0.4$	2
	Z_2	$sim(Z_2, S_i)=0.2$	3
	Z_3	$sim(Z_3, S_i)=0.7$	1

1. Compute the similarity score and rank for each gallery subject
2. Create one rank matrix and one score matrix

$$\mathcal{S} = \begin{bmatrix} 0.3 & 0.8 & 0.4 \\ 0.5 & 0.6 & 0.9 \\ 0.4 & 0.2 & 0.7 \end{bmatrix} \text{ and } \mathcal{R} = \begin{bmatrix} 3 & 1 & 2 \\ 3 & 2 & 1 \\ 2 & 3 & 1 \end{bmatrix}$$
3. Estimate the density distribution

Figure 7.1 Overview of the Rank-Score distribution calculation framework. The notation $sim(X_1, S_i)$ is used to denote the similarity score obtained by comparing an unknown subject S_i to the biometric sample 1 of a gallery subject in the face dataset.

7.2 Normalization

After $p(s, r)$ is obtained, the joint density of a pair of match score and rank, $p(s_m, r_m)$, is calculated and employed as measurement of weights. In order to further improve the method, the density is normalised before it is used as a weight. Z-score normalisation has been described previously in Section 3.2.1

Min-max normalisation: Min-max normalisation is suitable for the case that the boundaries (minimum and maximum values) of scores are known. After normalisation, all the scores are transferred into a common range $[0,1]$. Given a set of match scores $S = \{s_1, \dots, s_k\}$, the normalisation scores are given by:

$$s'_k = \frac{s_k - \min}{\max - \min} \quad (7.2)$$

Medium and median absolute deviation normalisation (MAD): Compared with Z-score normalisation, MAD is a robust method, since it is insensitive to outliers. The scheme of median and MAD is given by:

$$s'_k = \frac{s_k - \text{med}}{MAD} \quad (7.3)$$

where $\text{med} = \text{median}\{s_1, \dots, s_k\}$ and $MAD = \text{median}\{|s_k - \text{med}|\}$. This technique does not retain input distribution and does not transfer scores in a common range.

Tanh-estimators: The tanh-estimator proposed in [75] is a robust and efficient method. The normalised scores are given by:

$$s'_k = \frac{1}{2} \left\{ \tanh \left(0.01 \left(\frac{s_k - \mu}{\sigma} \right) \right) + 1 \right\} \quad (7.4)$$

where μ is the arithmetic mean and σ is standard deviation.

The recognition performance of the rank-score fusion method using different normalised algorithms is shown in Table 7.1. Tanh-estimators is demonstrated

to provide the best result across different normalisation techniques. At medium and far distances, the results given by tanh-estimators has a distinct advantage, and is as good as MAD at a close distance.

Table 7.1 Accuracy rate using different normalisation methods.

	Close	Medium	Far
Non-normalised	96.5%	84.9%	74.0%
Min-max	96.9%	85.7%	77.0%
Z-score	96.7%	90.8%	78.8%
MAD	98.6%	86.9%	74.8%
Tanh-estimators	98.5%	92.5%	82.6%

7.3 Rank-score fusion

The parameters of the rank-score distribution specified in Eq.(7.1) are calculated from training data before testing. During testing, an unknown user is matched with all registered users. For each matcher, one similarity score list and one rank list is obtained. Each registered user has a corresponding similarity score and a rank order, which are used as the inputs in Eq.(7.1). The joint density of similarity score and rank order is calculated. After normalisation, it is used as a weight to update the similarity score. The final similarity score of an unknown user and the k^{th} enrolled user can be calculated using Eq.(7.5), which is a weighted sum of different matchers.

$$ws_s(k) = \sum_{j=1}^m w_{p_{j,k}} s_{j,k} \quad (7.5)$$

where $s_{j,k}$ is the similarity score of enrolled user k and $w_{p_{j,k}}$ is normalised joint density. The unknown user is labelled as the class that has the maximum values of ws_s .

7.4 Experiment and discussion

7.4.1 Evaluation of rank-score fusion at three distances

In this section, experiments are performed to validate the performance of the proposed rank-score fusion algorithm at three distances.

The feature sets used are the same as those described in Chapter 4. The experiments are implemented using 200 subjects. Each subject has 20 samples, in which 5 samples are randomly chosen to train the joint density function and to obtain normalisation weights. The remaining samples are used for testing. The experiments were repeated 20 times, and the box-plot of recognition results are shown in Figure 7.2.

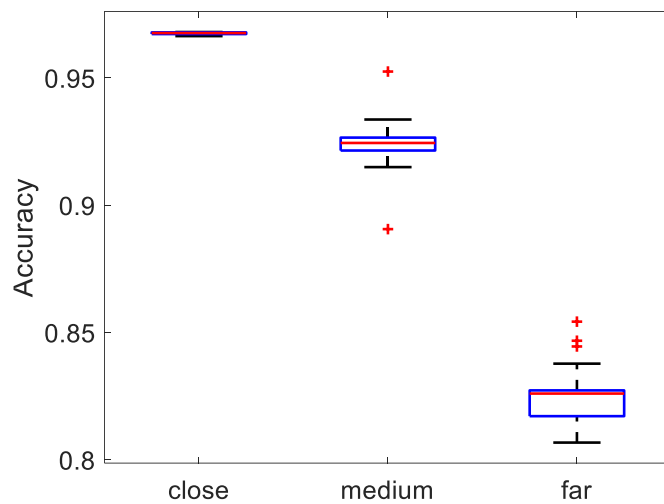


Figure 7.2 Fusion at three distances

It is clear that at the close distance is the most consistent, with an average recognition accuracy of 98.5% at close distance. The recognition accuracy at a medium distance slightly decreases to 92.5% on average, with the maximum 95.3% and minimum 89.1%. The accuracy and the uncertainty at a far distance are worse than that at close and medium distances.

7.4.2 Compared with single-modal recognition

Figure 7.3 is a vertical boxplot that shows the extent of the accuracy over 9 single-modal methods. The recognition results at three distances using single-modal methods and rank-score fusion method are listed in Table 7.2. At a close distance, the recognition accuracy of facial traits is 95.7%, which is the best over three signal-modal methods. The average recognition accuracy is 98.5% after rank-score fusion, which is 2.8% higher than using single facial traits. The fusion result at a close distance is the most consistent, and the variance is smallest compared with the other two distances. The results demonstrate that clothing traits achieve the highest recognition rate at a medium distance (69.4%), which is increased by 23.1% using rank-score fusion. The stability of fusion results at a far distance is not as good as the other two distances, because the accuracy of facial traits at a far distance is only 13.1%, which lowers the fusion result. Meanwhile, the fusion results of rank-score fusion are slightly improved at a far distance, and the accuracy of the proposed rank-score fusion increases to 82.6%.

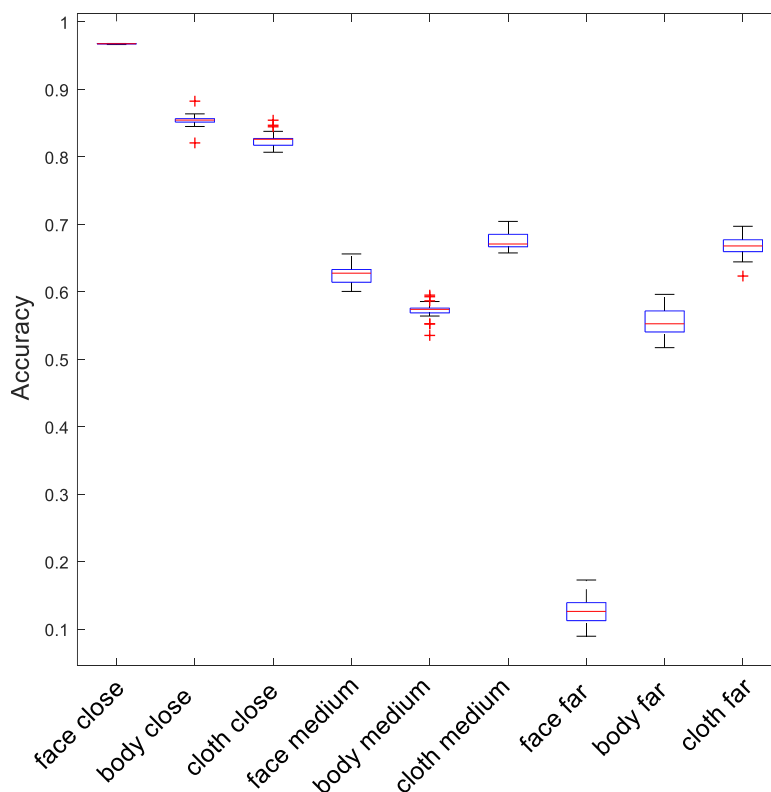


Figure 7.3 Accuracy for individual matcher (rank=1).

Table 7.2 Identification performance of signal-modal and rank-score fusion recognition

	Close		Medium		Far	
	Accuracy (%)	EER (%)	Accuracy (%)	EER (%)	Accuracy (%)	EER (%)
Face	95.7	0.45	62.7	3.89	13.1	22.76
Body	85.4	1.77	57.4	3.96	55.5	4.33
Clothing	82.6	2.76	69.4	4.00	67.1	3.59
rank-score fusion	98.5	0.33	92.5	0.69	82.6%	2.42

The comparisons between multi-modal and single-modal biometric methods demonstrate that recognition performance can be improved by multi-biometrics. The *gain* of rank-score fusion at three distances is 1.36, 5.64 and 1.48, respectively. The fusion performance at close and far distances is not improved as apparent as that at medium distance. As we mentioned before, the single-mode methods can achieve desirable results at a close distance, while facial attributes might reduce the fusion performance at a far distance.

7.4.3 Comparison with other fusion methods

The recognition results at three distances using different fusion methods are listed in Table 7.3. It is clear that the proposed method outperforms other rank and score level fusion techniques. At all three distances, the recognition accuracy of the proposed fusion method is always superior to that of other methods. At a close distance, all the fusion methods achieve excellent recognition rates. Compared with the best results given by other methods (97.0%), rank-score fusion improves the accuracy by 1.5%. At a medium distance, the recognition performance of rank-score fusion improves significantly. The accuracy increases to 92.5%, and the EER achieves 0.69%. The improvement of fusion performance at close and far distances is not as apparent as at a medium distance, which improves recognition accuracy by 1.8% (SVM-LRT).

Table 7.3: Identification performance using different fusion methods.

	Close		Medium		Far	
	Accuracy (%)	EER (%)	Accuracy (%)	EER (%)	Accuracy (%)	EER (%)
Bayesian theory [9]	96.3	0.38	84.6	1.07	78.1	2.57
Log likelihood ratio [69]	96.1	1.07	87.7	2.46	76.5	3.02
SVM-LRT	97.0	0.32	92.5	0.89	80.8	2.08
Borda count method [72]	95.8	0.45	76.4	3.64	73.3	4.17
Logistic regression [73]	96.4	0.39	82.3	3.73	75.5	3.83
Nonlinear weight ranks [67]	96.9	0.39	86.2	3.48	79.3	3.44
PAV based [74]	97.0	0.38	86.0	3.01	79.1	3.33
Rank-score fusion	98.5	0.33	92.5	0.69	82.6	2.42

The recognition results at three distances using different fusion methods are listed in Table 7.3. It is clear that the proposed method outperforms other rank and score level fusion techniques. At all three distances, the recognition accuracy of the proposed fusion method is always superior to that of other methods. At a close distance, all the fusion methods achieve excellent recognition rates. Compared with the best results given by other methods (97.0%), rank-score fusion improves the accuracy by 1.5%. At a medium distance, the recognition performance of rank-score fusion improves significantly. The accuracy increases to 92.5%, and the EER achieves 0.69%. The improvement in fusion performance at close and far distances is not as apparent as at a medium distance, which improves recognition accuracy by 1.8% (SVM-LRT).

7.5 Conclusions

A novel joint density-based rank-score fusion technique to fuse three soft biometric methods was proposed in this chapter. The experiments were conducted with other multi-modal fusion methods to make comparisons. Accuracy and EER were employed to evaluate their performance. The result of rank-score fusion demonstrates that at a close distance the soft biometric recognition performance is the most consistent. Compared with other fusion

methods, the proposed rank-score fusion is numerically demonstrated to be able to obtain the best results at all three distances, particularly at a medium distance. This leads to more general conclusions on this work.

Chapter 8 Conclusions and Future Work

8.1 Conclusions

The aim of this research is to improve the recognition performance of soft biometrics through different fusion techniques.

The first part of the work validated and justified the recognition performance of soft biometric fusion. Forthcoming research on soft biometric fusion focuses on the combination of soft biometrics and traditional biometrics. Thus, this research intended to test the applicability of fusion methods using soft biometric attributes.

In the research, three datasets were used which were collected from the University of Southampton Tunnel Laboratory database. The three datasets comprised of face, body and clothing traits, each of which included two types of attributes: categorical and comparative. The results of the single-modal recognition tests demonstrated that comparative attributes provide better results for the face and body. Compared with the label of categorical attributes, which are subjective descriptions, the objective descriptions between two subjects proved more reliable. For the clothing dataset, the comparative attributes were inferior to categorical attributes.

Two-level (feature and score) fusion methods were used to test the fusion capability of soft biometrics. In feature level fusion, several methods were used to find the feature which has the maximum effective information. A feature selected method, IFS, provided the best results with both categorical and comparative datasets. For score level fusion, the excellent performance was achieved by using a method based on Bayes theorem. In this section, the

improved recognition performance of soft biometrics fusion method was confirmed.

The second part of this thesis analysed the influence of different distances on soft biometric attributes, and on fused soft biometrics at different distances. The motivation was that the quality of face images decreases sharply when the distance between the subject and the camera increases. Face images have lots of detail, which cannot be obtained at a far distance. In order to test the effectiveness of attributes at different distances on consistent data, a new dataset of soft biometrics at different distances was created. The images in the new dataset were also collected from University of Southampton Tunnel Laboratory, by synthesising with an appropriate outdoor environment. Reference objects, such as vehicles, in the actual environment, were employed to aid the (calibration of the human) labelling of soft biometric traits. Three modalities of features were used, each of them with ten attributes which had three descriptions. Based on previous research, comparative attributes are best for face and body, and categorical is best for clothing. The labels were collected by crowdsourcing task using CrowdFlower.

Pearson's correlation coefficient and mutual information were used to analyse the stability and discriminating capacity of features at different distances. The results highlighted the individual advantages and disadvantages of face, body and clothing traits. The uniqueness of facial features leads to a high discriminatory power, but the details of facial features cannot be recognised when the subject is far away from the camera. It is concluded that the stability of face features is insufficient for recognition at distance. Compared with facial features, body and clothing features have less uniqueness in recognition, which reduces their differentiating power, although their consistency across distance is superior.

Three new fusion methods on the feature, score and rank-score levels were proposed, and the experiments were performed using the new soft biometric database. For feature level fusion, an SG-CCA method was used to fuse soft biometric features, and the results demonstrated the superiority of soft biometric fusion using the SG-CCA method for human recognition. For score level fusion, the proposed SVM-LRT was numerically demonstrated to be the best fusion method, when compared with fusion using Bayesian and LRT. In

addition, a novel joint density distribution-based rank-score fusion was studied, which overcame the shortcomings of the single-level fusion method, and the recognition performance was significantly improved by this method.

In conclusion, compared with single-mode biometric methods, the experimental results demonstrate clearly that the recognition performance of multi-modal soft biometrics is significantly improved by using biometric fusion. Naturally as recognition at a close distance is generally good, it can be seen that fusion largely improves the recognition at medium and far distances. There is still much room for improvement at the far distance. It would appear that the fusion process selects information that is best for recognition at any distance and so the effects are most dramatic when appropriate features are weighted more favourably for recognition purposes,

8.2 Future work

This section provides a non-exhaustive list of future work related to soft biometric recognition. Some of these issues were encountered during the research for this thesis, but have yet to be fully addressed.

8.2.1 Diversity of data collection

The images in the new dataset were collected from the University of Southampton tunnel laboratory. Because of the limitation on laboratory cameras, all the images used in the new dataset were taken from a single viewpoint. The whole body of the subject can be seen from this viewpoint, which helps Crowdfunder users to compare subjects. However, in a real-life situation, the viewpoint is rarely fixed. The viewpoint has an influence on soft biometric feature collection and fusion and this effect should be investigated in future.

In addition, all the subjects used in this thesis were captured when they were at pre-marked distances and were compared with subjects at the same distance. In practical applications, it is often required to compare two people at the different distances from the camera. The locations of subjects will also influence soft biometric labels. For example, it is hard to compare the height

of two subjects if they are at different distances. It is necessary to perform more experiments for the proposed fusion methods.

8.2.2 Automatic retrieval of biometric signatures

This thesis demonstrated the effectiveness of soft biometrics for subject identification. The labels of the soft biometric database were achieved by human comparisons and descriptions. In real life, identification of an unknown suspect is always based on verbal descriptions from eyewitnesses. It is important to bridge the semantic gap between humans and machines. In order to have more practical applications, automatic retrieval of biometric signatures is required. Some relevant work has been done in [28] [76].

8.2.3 Descriptions from memory

The comparisons between two subjects were collected by annotators while subjects' images were presented. However, in particular applications, the features used for identification are always described by eyewitness based on their memory. Some details of soft biometric features could be confused with the passage of time. Future research into the influence of memory on comparative labels should be conducted.

References

- [1] D. Reid, S. Samangoei, C. Chen, M. Nixon and A. Ross, "Soft biometrics for surveillance: an overview," in *Handbook of statistics*, vol. 31, Elsevier, 2013, pp. 327--352.
- [2] M. S. Nixon, P. L. Correia, K. Nasrollahi, T. B. Moeslund, A. Hadid and M. Tistarelli, "On soft biometrics," *Pattern Recognition Letters*, vol. 68, pp. 218--230, 2015.
- [3] S. Samangoei, B. Guo and M. S. Nixon, "The use of semantic human description as a soft biometric," in *Biometrics: Theory, Applications and Systems, 2008. BTAS 2008. 2nd IEEE International Conference on*, IEEE, 2008, pp. 1--7.
- [4] D. A. Reid, M. S. Nixon and S. V. Stevenage, "Soft biometrics; human identification using comparative descriptions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 6, pp. 1216--1228, 2014.
- [5] N. Almodhahka, M. S. Nixon and J. Hare, "Human face identification via comparative soft biometrics," in *Identity, Security and Behavior Analysis (ISBA), 2016 IEEE International Conference on*, IEEE, 2016, pp. 1--6.
- [6] E. S. Jaha and M. S. Nixon, "Clothing Analysis for Subject Identification and Retrieval," in *Recent Advances in Intelligent Image Search and Video Retrieval*, Springer, 2017, pp. 167--211.
- [7] A. K. Jain, S. C. Dass and K. Nandakumar, "Soft biometric traits for personal recognition systems," in *Biometric Authentication*, Springer, 2004, pp. 731--738.
- [8] P. Tome, J. Fierrez, R. Vera-Rodriguez and M. S. Nixon, "Soft biometrics and their application in person recognition at a distance," *IEEE Trans. IFS*, vol. 9, no. 3, pp. 464--475, 2014.

- [9] M. S. Nixon, B. H. Guo, S. V. Stevenage, E. S. Jaha, N. Almodhahka and D. Martinho-Corbishley, "Towards automated eyewitness descriptions: describing the face, body and clothing for recognition," *Visual Cognition*, vol. 25, no. 4-6, pp. 1--15, 2016.
- [10] A. Ross and R. Govindarajan, "feature level fusion of hand and face biometrics," in *Biometric Technology for Human Identification II*, vol. 5779, International Society for Optics and Photonics, 2005, pp. 196--204.
- [11] Q.-S. Sun, S.-G. Zeng, Y. Liu, P.-A. Heng and D.-S. Xia, "A new method of feature fusion and its application in image recognition," *Pattern Recognition*, vol. 38, no. 12, pp. 2437--2448, 2005.
- [12] M. Haghghat, M. Abdel-Mottaleb and W. Alhalabi, "Discriminant correlation analysis: Real-time feature level fusion for multimodal biometric recognition," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 9, pp. 1984--1996, 2016.
- [13] S. Shekhar, V. M. Patel, N. M. Nasrabadi and R. Chellappa, "Joint sparse representation for robust multimodal biometrics recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 1, pp. 113--126, 2014.
- [14] A. A. Ross, A. K. Jain and K. Nandakumar, "Score level fusion," in *Handbook of Multibiometrics*, Springer, 2006, pp. 91--142.
- [15] F. Wang and J. Han, "Multimodal biometric authentication based on score level fusion using support vector machine," *Opto-electronics review*, vol. 17, no. 1, pp. 59--64, 2009.
- [16] H. M. Sim, H. Asmuni, R. Hassan and R. M. Othman, "Multimodal biometrics: Weighted score level fusion based on non-ideal iris and face images," *Expert Systems with Applications*, vol. 41, no. 11, pp. 5390--5404, 2014.

- [17] S. Samangooei, B. Guo and M. S. Nixon, "The use of semantic human description as a soft biometric," in *Biometrics: Theory, Applications and Systems*, IEEE, 2008, pp. 1--7.
- [18] A. K. Jain, S. C. Dass and K. Nandakumar, "Can soft biometric traits assist user recognition?," in *Biometric Technology for Human Identification*, vol. 5404, International Society for Optics and Photonics, 2004, p. 39.
- [19] S. Samangooei, "Semantic Biometrics," University of Southampton, 2010.
- [20] E. S. Jaha and M. S. Nixon, "Analysing soft clothing biometrics for retrieval," in *International Workshop on Biometric Authentication*, Springer, 2014, pp. 234--245.
- [21] M. D. MacLeod, J. N. Frowley and J. W. Shepherd, *Whole body information: Its relevance to eyewitnesses*, Cambridge University Press, 1994.
- [22] C. A. Meissner, S. L. Sporer and J. W. Schooler, "Person descriptions as eyewitness evidence," *Handbook of eyewitness psychology: Memory for people*, pp. 1--34, 2007.
- [23] L. L. Kuehn, "Looking down a gun barrel: Person perception and violent crime," *Perceptual and Motor Skills*, vol. 39, no. 3, pp. 1159--1164, 1974.
- [24] K. Tsukida and M. R. Gupta, "How to analyze paired comparison data," in *University of Washington*, 2011.
- [25] R. R. Davidson, "On extending the Bradley-Terry model to accommodate ties in paired comparison experiments," *Journal of the American Statistical Association*, vol. 65, no. 329, pp. 317--328, 1970.
- [26] D. R. Hunter, "MM algorithms for generalized Bradley-Terry models," *The annals of statistics*, vol. 32, no. 1, pp. 384--406, 2004.

- [27] D. Sculley, "Rank aggregation for similar items," in *Proceedings of the 2007 SIAM international conference on data mining*, SIAM, 2007, pp. 587--592.
- [28] E. S. Jaha and N. S. Mark, "From Clothing to Identity: Manual and Automatic Soft Biometrics," *IEEE Trans. on IFS*, vol. 11, no. 10, pp. 2377--2390, 2016.
- [29] A. K. Jain and A. Ross, "Multibiometric systems," *Communications of the ACM*, vol. 47, no. 1, pp. 34-40, 2004.
- [30] A. Ross and A. Jain, "Information fusion in biometrics," *Pattern recognition letters*, vol. 24, no. 13, pp. 2115--2125, 2003.
- [31] J. Yang, J.-y. Yang, D. Zhang and J.-f. Lu, "Feature fusion: parallel strategy vs. serial strategy," *Pattern recognition*, vol. 36, no. 6, pp. 1369--1381, 2003.
- [32] V. Subbarayudu and M. V. Prasad, "Multimodal biometric system," in *First International Conference on merging Trends in Engineering and Technology*, IEEE, 2008, pp. 635--640.
- [33] A. K. Rattani, R. Dakshina, M. Bicego and M. Tistarelli, "Robust feature-level multibiometric classification," in *Biometric Consortium Conference, 2006 Biometrics Symposium: Special Session on Research at the*, IEEE, 2006, pp. 1--6.
- [34] Y.-F. Yao, X.-Y. Jing and H.-S. Wong, "Face and palmprint feature level fusion for single sample biometrics recognition," *Neurocomputing*, vol. 70, no. 7-9, pp. 1582--1586, 2007.
- [35] S. C. Shah and A. Kusiak, "Data mining and genetic algorithm based gene/SNP selection," *Artificial intelligence in medicine*, vol. 31, no. 3, pp. 183--196, 2004.

- [36] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," *IEEE Transactions on neural networks*, vol. 5, no. 4, pp. 537--550, 1994.
- [37] G. Roffo, S. Melzi and M. Cristani, "Infinite feature selection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 4202--4210.
- [38] R. M. Heiberger and E. Neuwirth, "One-way anova," in *R through excel*, Springer, 2009, pp. 165--191.
- [39] B. Guo and M. S. Nixon, "Gait feature subset selection by mutual information," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 39, no. 1, pp. 36--46, 2009.
- [40] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *Journal of bioinformatics and computational biology*, vol. 3, no. 2, pp. 185--205, 2005.
- [41] S. C. Dass, Y. Zhu and A. K. Jain, "Validating a biometric authentication system: Sample size requirements," *IEEE Transactions on PAMI*, vol. 28, no. 12, pp. 1902--1319, 2006.
- [42] P. Vincent and Y. Bengio, "Manifold parzen windows," in *Advances in neural information processing systems*, 2003, pp. 849--856.
- [43] D. Greene, P. Cunningham and R. Mayer, "Unsupervised learning and clustering," in *Machine learning techniques for multimedia*, Springer, 2008, pp. 51--90.
- [44] N. Kwak and C.-H. Choi, "Input feature selection by mutual information based on Parzen window," *IEEE transactions on pattern analysis and machine intelligence*, vol. 24, no. 12, pp. 1667--1671, 2002.
- [45] A. C. Guidoum, "Kernel estimator and bandwidth selection for density and its derivatives the kedd package version 1.0.3," Oct 2015. [Online].

Available: <http://cran.us.r-project.org/web/packages/kedd/vignettes/kedd.pdf>..

- [46] C. M. Jones, J. S. Marron and S. j. Sheather, "A brief survey of bandwidth selection for density estimation," *Journal of the American Statistical Association*, vol. 91, no. 433, pp. 401--407, 1996.
- [47] X. Li and T. Liu, *Research on Personal Identity Recognition Method Based on Multi-Biometric*, Tianjin: Tianjin University, 2010.
- [48] R. D. Seely, S. Samangoeei, M. Lee, J. N. Carter and M. S. Nixon, "The university of southampton multi-biometric tunnel and introducing a novel 3d gait dataset}," in *Biometrics: Theory, Applications and Systems, 2008. BTAS 2008. 2nd IEEE International Conference on*, IEEE, 2008, pp. 1--6.
- [49] D. Martinho-Corbishley, M. S. Nixon and J. N. Carter, "Analysing comparative soft biometrics from crowdsourced annotations," *IET Biometrics*, vol. 5, no. 4, pp. 276--283, 2016.
- [50] D. A. Reid and M. S. Nixon, "Using comparative human descriptions for soft biometrics," in *Biometrics (IJCB), 2011 International Joint Conference on*, IEEE, 2011, pp. 1--6.
- [51] M. D. Freeman, *The Police and Criminal Evidence Act 1984*, Sweet & Maxwell, 1985.
- [52] S. Wold, K. Esbensen and P. Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37--52, 1987.
- [53] C. He, Q. Liu, H. Li and H. Wang, "Multimodal medical image fusion based on IHS and PCA," *Procedia Engineering*, vol. 7, pp. 280--285, 2010.
- [54] S. Mika, G. Ratsch, J. Weston, B. Scholkopf and K.-R. Mullers, "Fisher discriminant analysis with kernels," in *Neural networks for signal*

processing IX, 1999. Proceedings of the 1999 IEEE signal processing society workshop., IEEE, 1999, pp. 41--48.

- [55] O. Toygar and A. Adnan, "Face recognition using PCA, LDA and ICA approaches on colored images," *Istanbul University-Journal of Electrical & Electronics Engineering*, vol. 3, no. 1, pp. 735--743, 2012.
- [56] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3/4, pp. 321--377, 1936.
- [57] L. Sun, S. Ji and J. Ye, "Canonical correlation analysis for multilabel classification: A least-squares formulation, extensions, and analysis," *IEEE Transactions on PAMI*, vol. 33, no. 1, pp. 194--200, 2011.
- [58] J. Rupnik and J. Shawe-Taylor, "Multi-view canonical correlation analysis," in *Conference on Data Mining and Data Warehouses*, 2010.
- [59] X.-Y. Jing, R. Hu, Y.-P. Zhu, S. Wu, C. Liang and J.-Y. Yang, "Intra-View and Inter-View Supervised Correlation Analysis for Multi-View Feature Learning," in *AAAI*, vol. 14, 2014, pp. 1882--1889.
- [60] T. Sun, S. Chen, J. Yang and P. Shi, "A novel method of combined feature extraction for recognition," in *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*, IEEE, 2008, pp. 1043--1048.
- [61] S. C. Dass, K. Nandakumar and A. K. Jain, "A principled approach to score level fusion in multimodal biometric systems," in *International conference on audio-and video-based biometric person authentication*, Springer, 2005, pp. 1049--1058.
- [62] A. Kumar and S. Shekhar, "Palmpoint recognition using rank level fusion," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*, 2010.
- [63] R. Sharma, S. Das and P. Joshi, "Rank level fusion in multibiometric systems," in *Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), 2015 Fifth National Conference on*, IEEE, 2015, pp. 1--4.

- [64] G. Lebanon and J. Lafferty, "Cranking: Combining rankings using conditional probability models on permutations," in *ICML*, Citeseer, 2002, pp. 363--370.
- [65] O. Melnik, Y. Vardi and C.-H. Zhang, "Mixed group ranks: Preference and confidence in classifier combination," *IEEE Transactions on PAMI*, vol. 26, no. 8, pp. 973--981, 2004.
- [66] M. M. Monwar and M. Gavrilova, "Markov chain model for multimodal biometric rank fusion," *Signal, Image and Video Processing*, vol. 7, no. 1, pp. 137--149, 2013.
- [67] A. Kumar and S. Shekhar, "Personal identification using multibiometrics rank-level fusion," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 41, no. 5, pp. 743--752, 2011.
- [68] A. Mourao, F. Martins and J. Magalhaes, "Inverse square rank fusion for multimodal search," in *International Workshop on Content-Based Multimedia Indexing*, IEEE, 2014, pp. 1--6.
- [69] K. Nandakumar, Y. Chen, S. C. Dass and A. Jain, "Likelihood ratio-based biometric score fusion," *IEEE transactions on PAMI*, vol. 30, no. 2, pp. 342--347, 2008.
- [70] S. Suthaharan, "Support vector machine," in *Machine Learning Models and Algorithms for Big Data Classification*, Springer, 2016, pp. 207--235.
- [71] D. Black, "Partial justification of the Borda count," *Public Choice*, vol. 28, no. 1, pp. 1--15, 1976.
- [72] K. Mc Donald and A. F. Smeaton, "A comparison of score, rank and probability-based fusion methods for video shot retrieval," in *International Conference on Image and Video Retrieval*, 2005.
- [73] A. A. Ross, K. Nandakumar and A. Jain, "Rank level fusion," in *Handbook of multibiometrics*, Springer Science & Business Media, 2006, pp. 70-72.

- [74] S. Nanang, "Pool Adjacent Violators Based Biometric Rank Level Fusion," in *International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2017.
- [75] A. Jain, K. Nandakumar and A. Ross, "Score normalization in multimodal biometric systems," *Pattern recognition*, vol. 38, no. 12, pp. 2270--2285, 2005.
- [76] N. Y. Almudhahka, M. S. Nixon and . J. S. Hare, "Semantic Face Signatures: Recognizing and Retrieving Faces by Verbal Descriptions," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 3, pp. 706--716, 2018.

