Covariate-invariant Gait Recognition using Random Subspace Method and its Extensions

by

Yu Guan

Thesis

Submitted to the University of Warwick

for the degree of Doctor of Philosophy

Doctor of Philosophy

The Department of Computer Science

....

THE UNIVERSITY OF
WARWICK

Contents

List of Tables

List of Figures

Acknowledgments

First and foremost, I would like to take this opportunity to express my deepest gratitude and respect to my supervisor Prof. Chang-Tsun Li, who constantly gave me support during my PhD time at the University of Warwick. I benefitted greatly from his encouragement, constructive advices, and financial support. I feel lucky to work with him, and look forward to maintaining the collaboration in the future.

My parents, Dr. Huaimin Guan, Mrs Cuihua Wang also deserve my cordial gratitude. Their love, support and encouragement have always been the source of my strength and the reason I have progressed this far. I also appreciate the support from my younger brother, Mr. Xin Guan, who cooks delicious food for me everyday during the last year of my PhD!

For my PhD research, I would like express my sincere thanks to Prof. Fabio Roli (PRA Lab, University of Cagliari, Italy), Prof. Massimo Tistarelli (CV Lab, University of Sassari, Italy), Prof. Liang Wang (NLPR, CASIA, China), and Prof. Tieniu Tan (NLPR, CASIA, China) for inviting me for academic visiting. I benefitted a lot from the valuable research experience from their labs. I would also like to thank the colleagues at Warwick, particularly, faculty members Dr. Abhir Bhalerao, Dr. Victor Sanchez and PhD students Miss Xingjie Wei, Mr. Yi Yao, Mr. Antonis Mouhtaropoulos, Mr. Ruizhe Li, Mr. Ning Jia, Mr. Xufeng Lin, Mr. Alaa Khadidos, Mr. Xin Guan, Mr. Muhammad Hilmi Kamarudin, Mr. Roberto Leyva, and Mr. Qiang Zhang. I get motivated a lot from them.

Yu Guan
Aug. 2014

Declarations

This thesis is submitted to the University of Warwick in partial fulfillment of the requirements for admission to the degree of Doctor of Philosophy. The work presented here is my own, except where specifically stated otherwise, and was performed in the Department of Computer Science at the University of Warwick under the supervision of Professor Chang-Tsun Li during the period Oct. 2010 to Aug. 2014. The research materials have not been submitted, either in the same or different form, to this or any other university for a degree. All sources of information are specifically acknowledged.

Abstract

Compared with other biometric traits like fingerprint or iris, the most significant advantage of gait is that it can be used for remote human identification without cooperation from the subjects. The technology of gait recognition may play an important role in crime prevention, law enforcement, etc. Yet the performance of automatic gait recognition may be affected by covariate factors such as speed, carrying condition, elapsed time, shoe, walking surface, clothing, camera viewpoint, video quality, etc. In this thesis, we propose a random subspace method (RSM) based classifier ensemble framework and its extensions for robust gait recognition.

Covariates change the human gait appearance in different ways. For example, speed may change the appearance of human arms or legs; camera viewpoint alters the human visual appearance in a global manner; carrying condition and clothing may change the appearance of any parts of the human body (depending on what is being carried/wore). Due to the unpredictable nature of covariates, it is difficult to collect all the representative training data. We claim overfitting may be the main problem that hampers the performance of gait recognition algorithms (that rely on learning). First, for speed-invariant gait recognition, we employ a basic RSM model, which can reduce the generalisation errors by combining a large number of weak classifiers in the decision level (i.e., by using majority voting).

We find that the performance of RSM decreases when the intra-class variations are large. In RSM, although weak classifiers with lower dimensionality tend to have better generalisation ability, they may have to contend with the underfitting problem if the dimensionality is too low. We thus enhance the RSM-based weak classifiers by extending RSM to multimodal-RSM. In tackling the elapsed time covariate, we use face information to enhance the RSM-based gait classifiers before the decision-level fusion. We find significant performance gain can be achieved when lower weight is assigned to the face information. We also employ a weak form of multimodal-RSM for gait recognition from low quality videos (with low resolution and low frame-rate) when other modalities are unavailable. In this case, model-based information is used to enhance the RSM-based weak classifiers. Then we point out the relationship of base classifier accuracy, classifier ensemble accuracy, and diversity among the base classifiers. By incorporating the model-based information (with lower weight) into the RSM-based weak classifiers, the diversity of the classifiers, which is positively correlated to the ensemble accuracy, can be enhanced.

In contrast to multimodal systems, large intra-class variations may have a significant impact on unimodal systems. We model the effect of various unknown covariates as a partial feature corruption problem with unknown locations in the spatial domain. By making some assumptions in ideal cases analysis, we provide the theoretical basis of RSM-based classifier ensemble in the application of covariate-invariant gait recognition. However, in real cases, these assumptions may not hold precisely, and the performance may be affected when the intra-class variations are large. We propose a criterion to address this issue. That is, in the decision-level fusion stage, for a query gait with unknown covariates, we need to dynamically suppress the ratio of the false votes and the true votes before the majority voting. Two strategies are employed, i.e., local enhancing (LE) which can increase true votes, and the proposed hybrid decision-level fusion (HDF) which can decrease false votes. Based on this criterion, the proposed RSM-based HDF (RSM-HDF) framework achieves very competitive performance in tackling the covariates such as walking surface, clothing, and elapsed time, which were deemed as the open questions.

The factor of camera viewpoint is different from other covariates. It alters the human appearance in a global manner. By employing unitary projection (UP), we form a new space, where the same subjects are closer from different views. However, it may also give rise to a large amount of feature distortions. We deem these distortions as the corrupted features with unknown locations in the new space (after UP), and use the RSM-HDF framework to address this issue. Robust view-invariant gait recognition can be achieved by using the UP-RSM-HDF framework.

In this thesis, we propose a RSM-based classifier ensemble framework and its extensions to realise the covariate-invariant gait recognition. It is less sensitive to most of the covariate factors such as speed, shoe, carrying condition, walking surface, video quality, clothing, elapsed time, camera viewpoint, etc., and it outperforms other state-of-the-art algorithms significantly on all the major public gait databases. Specifically, our method can achieve very competitive performance against (large changes in) view, clothing, walking surface, elapsed time, etc., which were deemed as the most difficult covariate factors.

Chapter 1

Introduction

1.1 Biometrics

Biometrics is the study of identifying subjects based on their physiological or behavioural traits. Physiological traits include face, iris, fingerprint, DNA, palm print, hand geometry, etc. while behavioral traits include gait, typing rhythm, signature, etc. A biometric trait needs to satisfy the following properties [Jain et al., 2004b]:

- *Universality*: each individual should have the trait.
- *Distinctiveness*: individuals can be well separated by the trait.
- *Permanence*: the trait should be sufficiently invariant over a period of time.
- *Collectability*: the trait can be measured quantitatively.

A biometric system based on a certain trait may have either a verification or an identification mode, depending on the application context [Jain et al., 2004b]. In the verification mode, the system validates the claimed identity (of a subject) by comparing the query biometric trait with his/her own reference trait stored in the system's database. The system conducts a one-to-one comparison to determine whether the claim is true or not. Biometric verification has widely been used in commercial applications such as access control.

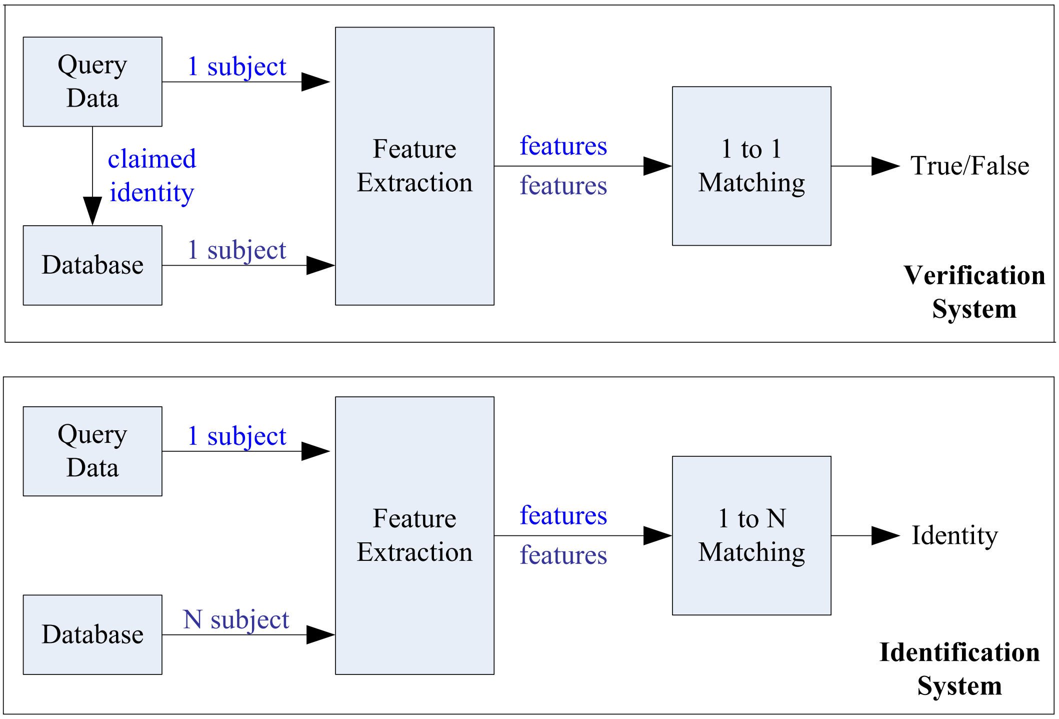
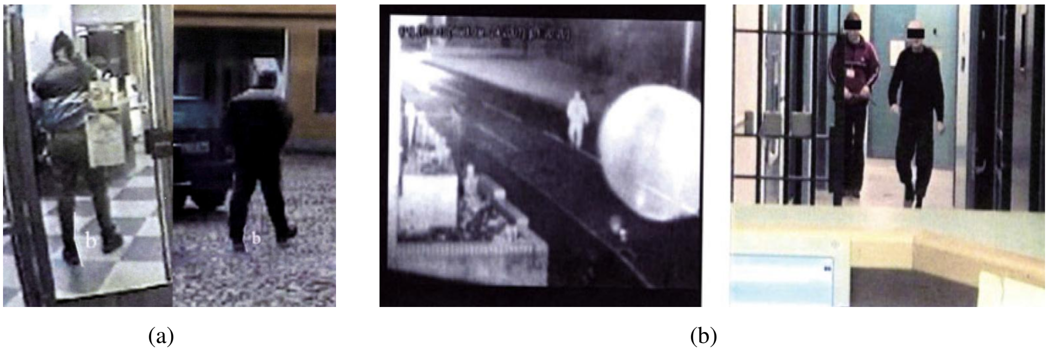


Figure 1.1: Biometric systems

In the identification mode, the system recognises a subject by searching all the biometric templates of all subjects in the database for a match. The system conducts a one-to-many comparison for a subject's identity. Biometric identification is used more frequently in the applications of law enforcement, e.g., latent fingerprint identification, human gait identification. The flowcharts of both modes are shown in Fig. 1.1, and this thesis falls into the category of biometric identification.

1.2 Human Identification at a Distance using Gait

Human identity recognition is fundamental to human life, and the technology of human identification and tracking from a distance may play an important role in crime prevention, law enforcement, search for missing people (e.g., missing children or people with dementia), etc. Nowadays, CCTV cameras are widely installed in public places such as airports,



(a) (b)

Figure 1.2: (a) CCTV images for the robbery case in Denmark [Larsen et al., 2008], left: the perpetrator, right: the suspect; (b) CCTV images for the burglary case in UK [Bouchrika et al., 2011], left: the perpetrator, right: the suspect.

government buildings, streets and shopping malls for the afore-mentioned purposes. In 2013, the British security industry authority (BSIA) estimated there are up to 5.9 million CCTV cameras nationwide, and that is around 1 every 11 people [Barrett, 2013]. Because of the need for sufficient manpower to supervise such a large number of CCTVs, the need for automatic human identification systems is acute.

Recently, a number of reports (e.g.,[Larsen et al., 2008] [Bouchrika et al., 2011]) suggested that behavioral biometrics, gait recognition, can be used for human identification from CCTV footage. In [Larsen et al., 2008], based on a checklist for forensic gait analysis, Larsen et al. managed to identify a bank robber in Denmark by matching surveillance footage, as illustrated in Fig. 1.2(a). Fig. 1.2(b) shows a gait recognition scenario in UK where a burglar was identified through gait analysis from a podiatrist [Bouchrika et al., 2011]. These pieces of gait-based evidences proved their usefulness by providing incriminating evidence, leading to convictions in a court of law. However, for automatic gait recognition, covariate factors like camera viewpoint, carrying condition, clothing, etc. may limit the performance. In this thesis, we aim to propose automatic gait recognition algorithms that are robust to these factors.

1.3 Contributions and Thesis Outline

Our objective in this thesis is to propose gait recognition algorithms that are less sensitive to covariate factors such as shoe, carrying condition, clothing, walking surface, view angle, elapsed time, speed, video quality, etc. In this thesis, we view the effect of most of the afore-mentioned covariates as a partial feature corruption problem with unknown locations, and propose a framework based on the concept of random subspace method (RSM) [Ho, 1998]. To address the problems caused by some covariates (such as clothing, walking surface, elapsed time, large view angel), we further propose several extensions of the RSM framework.

1.3.1 Contributions

The contributions of this thesis are listed as follows.

Gait Recognition Challenges Modelling and an Unified Solution

We treat the effect of most of covariates as a partial feature corruption problem with unknown locations. That is, various covariates (e.g., shoe, carrying condition, clothing, speed, elapsed time, etc.) only affect parts of the silhouette in the spatial domain. Due to unpredictable nature of the corrupted feature locations, it is difficult to use a single fixed classifier for robust gait recognition. We provide a classifier ensemble solution for addressing this issue, and point out its potential extensions. More theoretical details can be found in Chapter 6.

Gait-based Random Subspace Method (RSM): a Basic Model

We propose a basic RSM model for gait recognition. Initially, two dimensional principle component analysis (2DPCA) is used for feature decorrelation. A large number of random subspaces are constructed, with each subspace spanned by a subset of 2DPCA basis vectors chosen at random. The random features extracted from each subspace can be further enhanced with supervised methods (e.g., two-dimensional linear discriminant analysis (2DLDA) for the basic model), and we refer this process as to local enhancing (LE). The final classification decision is achieved by majority voting among the output labels from the base/weak classifiers corresponding to the random subspaces. We use this basic RSM model for speed-invariant gait recognition in Chapter 3. This basic model outperforms other state-of-the-art algorithms significantly, which demonstrates its effectiveness.

RSM Extension 1: Multimodal-RSM

We extend the basic RSM model to multimodal-RSM. That is, we enhance the RSM-based weak classifiers by using the classification score of other information to update the gait-based classification score for each random subspace. In Chapter 4, we fuse face information into the RSM-based gait recognition system to tackle the elapsed time covariate. In Chapter 5, for gait recognition in low-quality videos where face information is unreliable, we instead use model-based information to enhance the RSM-based weak classifiers. We also study the relationship of base classifier accuracy, classifier ensemble accuracy, and diversity among the base classifiers.

RSM Extension 2: RSM-based Hybrid Decision-level Fusion (RSM-HDF)

We propose RSM-based hybrid decision-level fusion (RSM-HDF) strategy for improving the performance of biometric (identification) systems. We construct a new criterion to $dynamically$ suppress the ratio of false votes and true votes before the majority voting. Two strategies are employed, i.e., the afore-mentioned LE which can increase true votes, and the proposed HDF which can decrease false votes.

For each subspace, the random features are further enhanced by LE using two different supervised methods (LE1, and LE2). In a random subspace with irrelevant features, LE would lead to label assignment (from the corresponding classifier) in a relatively random manner. Based on the same irrelevant features, a classifier pair corresponding to LE1 and LE2 would output two random labels, which are unlikely to be the same. Based on

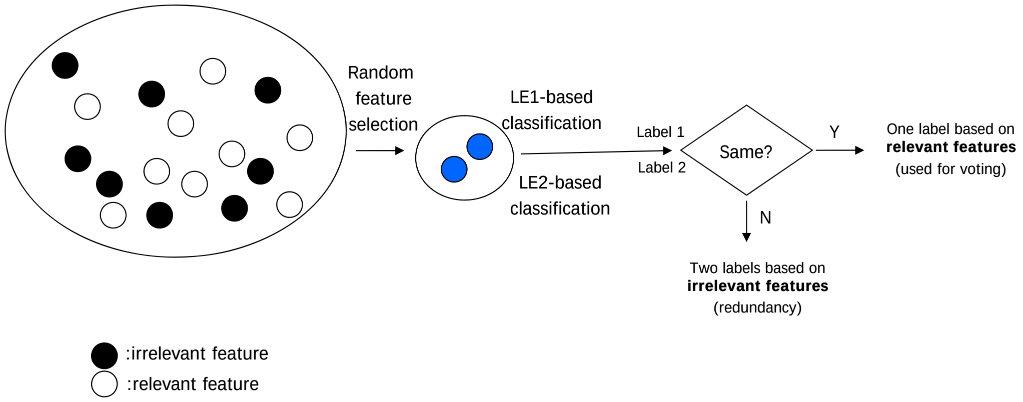


Figure 1.3: The process of suppressing false votes (i.e., the effect of irrelevant features)

the ”AND” rule, classifier pairs with different output labels are deemed as invalid votes and simply discarded, which can suppress the false votes effectively. The afore-mentioned process is illustrated in Fig. 1.3.1, and HDF is the majority voting after discarding the false votes. As presented in Chapter 6, we use RSM-HDF to tackle the covariates like walking surface, clothing, elapsed time, and observe that its performance is much higher than other state-of-the-art algorithms.

RSM Extension 3: Unitary Projected RSM-HDF (UP-RSM-HDF)

Camera viewpoint is different from other covariates, and cannot be simply deemed as a partial feature corruption problem in the spatial domain. Based on 2DLDA-based unitary projection (UP), we form a new space, where the same subjects are closer from different views. The trained UP matrix can transform gait from different views onto a common space before the matching is performed.

Although UP may make the multi-view samples of a subject closer in the new space, it may also give rise to significant feature distortions. We deem these distortions as the corrupted features with unknown locations in the new space (after UP) and use RSM-HDF to address this issue. In Chapter 7, we use UP-RSM-HDF for cross-view gait recognition, which significantly outperforms other state-of-the-art algorithms.

In Tackling the Open Issues

Covariates such as (large changes in) view, clothing, walking surface, elapsed time, etc. were deemed as the most challenging factors, which would significantly change the gait appearance. We model the effect of these factors as a partial feature corruption problem (in certain domains) with a large amount of irrelevant features. For query gaits in different walking conditions, these irrelevant features vary in terms of size and locations. We propose HDF to filter out the false votes (corresponding to the irrelevant features), before the majority voting. We employ the proposed RSM-HDF and UP-RSM-HDF to address the hard problems and our methods can achieve very competitive performance against the afore-mentioned challenging factors, much higher than other state-of-the-art algorithms.

1.3.2 Thesis Outline

The rest of this thesis is organised as follows. In Chapter 2, first we review the relevant literature of gait recognition on feature representation categories, challenges, databases, various algorithms against different covariates. We also introduce some fundamental knowledge on feature templates, subspace learning, and the concept of RSM. In Chapter 3, we introduce the basic model of gait-based RSM and use it for speed-invariant gait recognition. In Chapter 4, we extend RSM to multimodal-RSM. We incorporate face information into the RSM-based gait classifiers to tackle the elapsed time challenges. In Chapter 5, we use a similar form of multimodal-RSM. We incorporate model-based information into the RSM-based weak classifiers for robust gait recognition in low quality videos (low frame-rate and low resolution). A study on the relationship of base classifier accuracy, classifier ensemble accuracy, and diversity among the base classifiers is also provided. In Chapter 6, we model the effect of covariates as a partial feature corruption problem with unknown locations and provides the theoretical basis of RSM-based classifier ensemble method. A new criterion, which is to suppress the ratio of false votes and true votes, for improved decision-level fusion is proposed. This framework, namely RSM-HDF, is used to tackle difficult tasks with covariates such as walking surface, elapsed time, clothing, etc. In Chapter 7, we formulate cross-view gait recognition as a partial feature corruption problem with unknown locations in a feature space constructed through unitary projection. Chapter 8 concludes this thesis and suggests directions for future research.

1.4 List of Publications

We provide the full publication list for my PhD research on gait recognition as follows:

1. **Y. Guan,** C.-T. Li, and F. Roli, ”On Reducing the Effect of Covariate Factors in Gait Recognition: a Classifier Ensemble Method”. *IEEE Trans. Pattern Analysis and Machine Intelligence, (T-PAMI)*, vol.37, no. 99, 2015

2. **Y. Guan,** Y. Sun, C.-T. Li, and M. Tistarelli, ”Human Gait Identification from Extremely Low Quality Videos: an Enhanced Classifier Ensemble Method”. *IET Biometrics*. vol. 3, issue 2, pp 84-93, June, 2014

3. **Y. Guan,** X. Wei, and C.-T. Li, ”On the Generalization Power of Face and Gait in Gender Recognition”, *International Journal of Digital Crime and Forensics (IJDCF).* vol. 6, no. 1, pp 1-8, Jan. 2014

4. **Y. Guan,**, X. Wei, C.-T. Li, and Y. Keller, ”People Identification and Tracking through Fusion of Facial and Gait Features,” *International Workshop on Biometrics (BIOMET)*, Sofia, Bulgaria, 23-24 June, 2014 (Invited Paper)

5. **Y. Guan,** X. Wei, C.-T. Li, G. L. Marcialis, F. Roli and M. Tistarelli, ”Combining Gait and Face for Tackling the Elapsed Time Challenges” *in Proc. the 6th IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS),* Washington DC, USA, Sept. 2013

6. **Y. Guan,** and C.-T. Li “A Robust Speed-Invariant Gait Recognition System for Walker and Runner Identification” *in Proc. the 6th IAPR International Conference on Biometrics (ICB),* Madrid, Spain, June, 2013.

Thesis Chapters	Publications	Content
Chapter 3	Paper 6	RSM: the basic model
Chapter 4	Paper 5	Multimodal-RSM
Chapter 5	Paper 2, 7	A weak form of Multimodal-RSM
Chapter 6	Paper 1, 9, 10	The theoretical basis of RSM and RSM-HDF
Chapter 7	-	UP-RSM-HDF

Table 1.1: Thesis chapters and the corresponding publications

7. **Y. Guan,** C.-T. Li and S. D. Choudhury,“Robust Gait Recognition from Extremely Low Frame-Rate Videos,” *in Proc. International Workshop on Biometrics and Forensics (IWBF)*, Lisbon, Portugal, Apr. 2013.

8. **Y. Guan,** C.-T. Li and Y. Hu, “An Adaptive System for Gait Recognition in Multi-View Environments,” *in Proc. the 14th ACM Multimedia and Security Workshop (MMSec)*, Coventry, UK, Sept. 2012.

9. **Y. Guan,** C.-T. Li and Y. Hu, “Robust Clothing-Invariant Gait Recognition,” *in Proc. the 8th International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, Piraeus-Athens, Greece, July, 2012.

10. **Y. Guan,** C.-T. Li and Y. Hu, “Random Subspace Method For Gait Recognition,” *in Proc. IEEE International Conference on Multimedia and Expo Workshop (ICMEW)* , Melbourne, Australia, July 2012.

11. **Y. Guan,** C.-T. Li and Y. Hu, “Gait Recognition under Carrying Condition: a Static Dynamic Fusion Method,” *in Proc. SPIE Optics, Photonics and Digital Technologies for Multimedia Applications*, Brussels, Belgium, Apr. 2012.

Most chapters of this thesis (i.e., Chapters 3-6) are highly related to some of the afore-mentioned papers, as listed in Table 1.1.

Chapter 2

Literature Review

2.1 Categories of Gait Recognition Algorithms

Existing gait recognition algorithms can be roughly divided into two categories: model-based and appearance-based approaches. Model-based methods aim to model the human body structure for recognition, while appearance-based approaches can perform classification regardless of the underlying body structure. Although model-based methods may perform well in some challenging cases (e.g., when the view change is large [Goffredo et al., 2010]), they generally have lower performance than appearance-based methods. One major reason is that when affected by self-occlusion, low resolution or other factors, it is often difficult to estimate the body parameters precisely, and in this case they only provide limited information for recognition. Compared with model-based methods, appearance-based approaches can more general and most of them use the whole gait image as input features. Owing to the large number of features, a common approach adopted by appearance-based algorithms is to perform dimensionality reduction. In this section, we introduce the feature representations for both model-based and appearance-based methods.

2.1.1 Model-based Representation

Niyogi and Adelson proposed one of the earliest model-based representation, i.e., a five-stick model representing the body structure [Niyogi and Adelson, 1994]. In [Bobick and Johnson, 2001], a recognition model was proposed based on the static body and stride parameters. Lee and Grimson fitted ellipses to seven regions of body and used the corresponding parameters as gait signature [Lee and Grimson, 2002]. Cunado et al. modelled the gait as moving pendulum and extracted the features from the hip angular motion [Cunado et al., 2003]. Wagg and Nixon studied the model-based features through a statistical analysis and found that most recognition power lies in static body shape parameters and cadence [Wagg and Nixon, 2004]. Yam et al. proposed a model for walking gait and running gait recognition, which uses the signature derived from the angular motion of thighs and knees [Yam et al., 2004]. Wang et al. used distances between boundary pixels to silhouette centroid as model-based gait features [Wang et al., 2003a]. In [Kale et al., 2004b], the widths of outer contour were used to represent the body structure. Most recently, after estimating the poses of lower limbs, Goffredo et al. extracted model-based features based on the rectified angular measurements and trunk spatial displacements for view-invariant gait recognition [Goffredo et al., 2010].

2.1.2 Appearance-based Representation

One of the simplest appearance-based representation is the aligned binary gait silhouette. Sarkar et al. proposed the gait recognition baseline method, and they used spatial-temporal correlation on the aligned binary gait silhouettes [Sarkar et al., 2005]. However, such three dimensional video classification problems often incur high computational complexity and tend to be less robust to segmentation errors. To deal with these dilemmas, period-based gait appearance templates were proposed in recent works that encode the information of the frames from a gait cycle into a single image and formulate gait recognition as a two dimensional image classification problem. On the OU-ISIR-LP dataset, consisting of more than 3000 subjects, Iwama et al. [Iwama et al., 2012] conducted a study on six popular

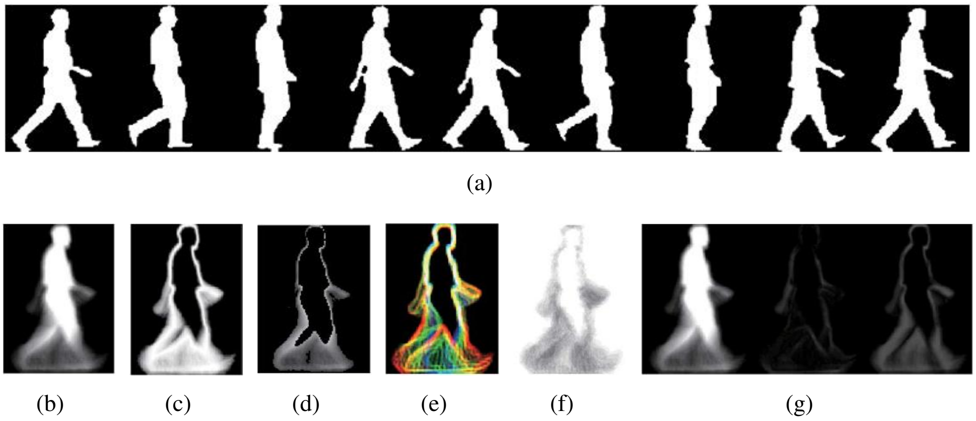


Figure 2.1: Gait representations for a subject on the OU-ISIR-LP dataset [Iwama et al., 2012], (a) the original gait silhouettes (b)-(g) the 6 period-based feature templates from left to right: GEI [Han and Bhanu, 2006], GEnI [Bashir et al., 2009], MGEI [Bashir et al., 2010a], CGI [Wang et al., 2012], GFI [Lam et al., 2011], and FDF (with 0, 1, and 2 times frequency elements)[Makihara et al., 2006]

period-based gait appearance templates including gait energy image (GEI)[Han and Bhanu, 2006], gait entropy image (GEnI) [Bashir et al., 2009], masked GEI based on GEnI (MGEI) [Bashir et al., 2010a], chrono-gait image (CGI) [Wang et al., 2012], gait flow image (G-FI)[Lam et al., 2011], and frequency-domain feature (FDF) [Makihara et al., 2006]. The gait silhouettes and the corresponding six appearance templates from a subject are listed in Fig. 2.1. Out of these six templates, GEI can be computed by averaging silhouettes over a gait cycle. GEnI can be obtained by calculating the entropy for every pixel over a gait cycle. MGEI can be created by masking the GEI with a pair-wise mask generated by each pair of probe and gallery GEnIs. After encoding the temporal information among gait frames by a colour mapping function, CGI can be obtained by compositing the colour encoded gait contour images in a gait cycle. Based on an optical flow field from silhouettes representing motion information, GFI can be computed by averaging the binarised flow images over a gait cycle. FDF can be generated by applying a discrete Fourier transform of the temporal axis to the silhouette images in a gait cycle. More details of these feature templates can be found in [Iwama et al., 2012],[Han and Bhanu, 2006],[Bashir et al., 2009],[Bashir et al., 2010a],[Wang et al., 2012],[Lam et al., 2011],[Makihara et al., 2006]. The results in [Iwama et al., 2012] showed that when there are no covariates, best performance can be achieved by using GEI.

Recently, several local feature based templates were also proposed. The local features of afore-mentioned basic templates (e.g., GEI) were extracted to form new gait templates for recognition. In [Liu et al., 2012], HOG-GEI was proposed by extracting histogram of oriented gradients (HOG) on GEI. In [Tao et al., 2007], [Huang et al., 2010], [Xu et al., 2012], [Guan et al., 2015], Gabor filters were used to extract the local features of GEI along five scales and eight orientations. Compared with the basic templates (e.g., GEI), the corresponding local feature based templates can yield certain level of performance improvement.

Since our contribution in this thesis is not to propose new gait templates, we only employ two most popular feature templates, i.e., GEI and Gabor-filtered GEI, and more details can be found in Chapter 2.4.1.

2.2 Gait Recognition Challenges and Databases

In this section, we introduce gait recognition challenges and the commonly used gait databases consisting of images with these challenging factors for algorithm evaluation.

2.2.1 Challenges

In real-world scenarios, the walking conditions between the probe and gallery images are not always the same. Walking conditions may change due to different covariate factors, which can be divided into four categories:

1. speed,
2. camera viewpoint,
3. covariates with unpredictable effect, e.g., shoe type, carrying condition, clothing, walking surface, elapsed time, etc.
4. video quality, e.g., low frame-rate and low resolution.

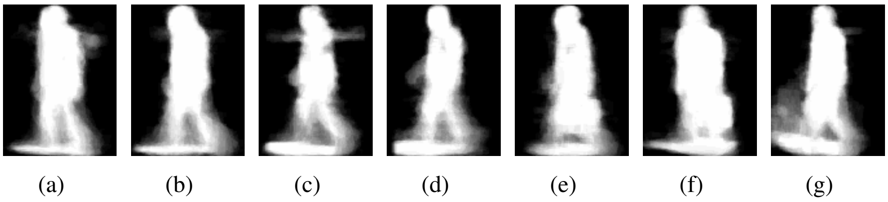


Figure 2.2: GEIs of one subject walking in different walking conditions from the USF gait dataset [Sarkar et al., 2005]. (a) is the GEI in normal condition. (b)-(g) are the GEIs under the influences of (b) viewpoint, (c) walking surface, (d) viewpoint and walking surface, (e) carrying condition, (f) carrying condition and viewpoint, (g) elapsed time, shoe type, clothing, and walking surface.

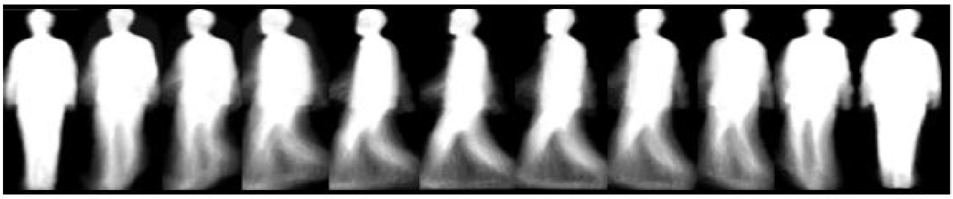


Figure 2.3: GEI examples from CASIA-B dataset [Yu et al., 2006] of a subject from view 0° to 180°, with an interval of 18°.

Since small changes in speed and camera viewpoint only have limited impact on gait recognition ([Tan et al., 2006], [Yu et al., 2006]), it is more reasonable to evaluate several representative speeds/views, instead of all the possible speeds/views. In OU-ISIR-A dataset [Makihara et al., 2012], the walking speeds for evaluation are from 2km/h to 7km/h, with an interval of 1km/h, while in CASIA-B dataset [Yu et al., 2006], the camera viewpoints for evaluation are from 0° to 180°, with an interval of 18°. For detailed studies on the effect of speed/viewpoint, we separate these two (relatively) controllable covariates from others.

Despite the effectiveness of GEIs, when the walking condition changes, matching GEIs directly makes the classification prone to errors. Fig. 2.2 shows some GEIs of the same subject walking in different walking conditions from the USF gait dataset [Sarkar et al., 2005], while Fig. 2.3 demonstrates several GEIs of the same subject in different views from the CASIA-B dataset [Yu et al., 2006]. We can see that covariates may substantially alter the human appearance, thus giving rise to recognition difficulties.

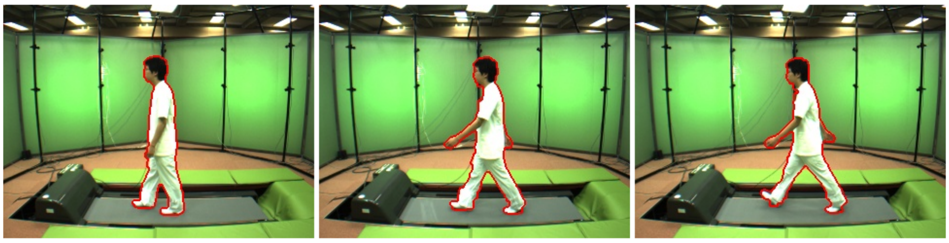


Figure 2.4: Gait images with different speeds from the OU-ISIR-A dataset[Makihara et al., 2012]

2.2.2 Databases

OU-ISIR-A

The OU-ISIR-A dataset [Makihara et al., 2012] was constructed for evaluating speed-invariant gait recognition algorithms. It was collected on a treadmill with a large range of speeds (from 2km/h to10km/h) in terms of walking or running for 34 subjects. The subjects were instructed to walk at six different speeds, ranging from 2km/h to 7km/h with an interval of 1km/h, and to run at three different speeds, ranging from 8km/h to 10km/h with an interval of 1km/h . Several gait images from this dataset are shown in Fig. 2.4.

OU-ISIR-B

The OU-ISIR-B dataset was constructed by Hossain et al. [Hossain et al., 2010] for studying the effect of clothing on gait recognition. It includes 68 subjects walking on a treadmill with up to 32 types of clothes combinations. Several gait images from this dataset are shown in Fig. 2.5.

OU-ISIR-D

The OU-ISIR-D database [Makihara et al., 2012] consists of two datasets, namely, DB-high (i.e., with small gait fluctuations) and DB-low (i.e., with large gait fluctuations). For DB-high/DB-low, there are 100 subjects (1 subject per sequence) for both the gallery and probe. The original resolution and frame-rate in OU-ISIR-D database are 128 × 88 pixels and 60

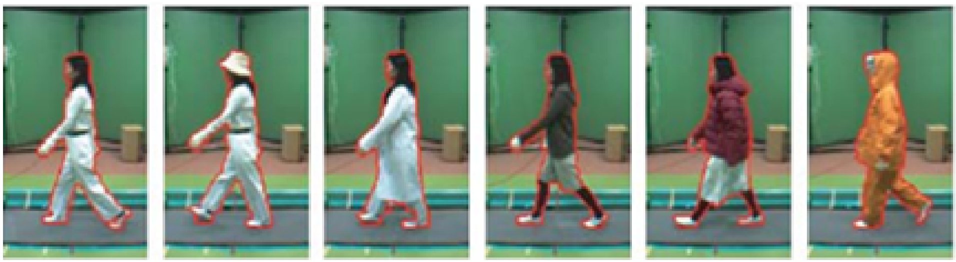


Figure 2.5: Gait images with several different clothes types from the OU-ISIR-B dataset [Hossain et al., 2010]

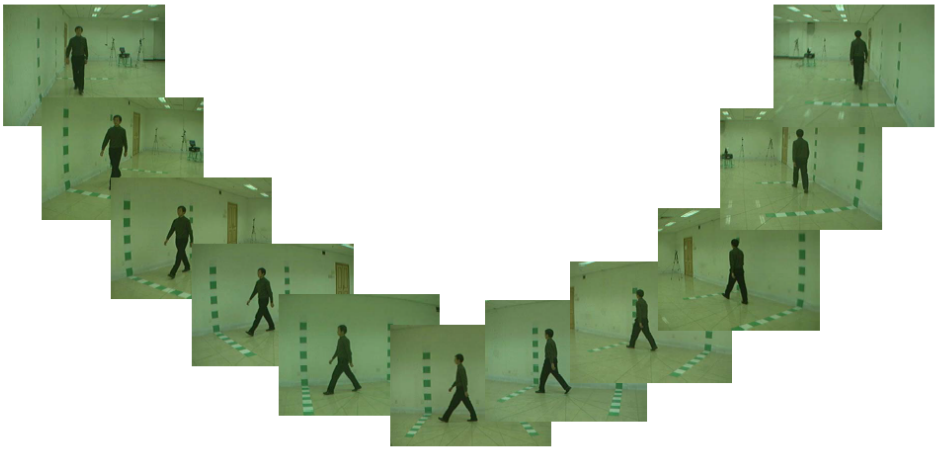


Figure 2.6: Gait images from CASIA-B dataset [Yu et al., 2006] of a subject from view 0° to 180°, with an interval of 18°;

fps. This dataset is normally down-sampled in a spatial and temporal manner, and used to test gait recognition algorithms on extremely low quality videos.

CASIA-B

The CASIA-B gait dataset [Yu et al., 2006] is a large multi-view gait dataset, which consists of 124 subjects walking in the indoor environment with the cameras fixed at 11 viewpoints (from 0° to 180° with an interval of 18°), as shown in Fig. 2.6.

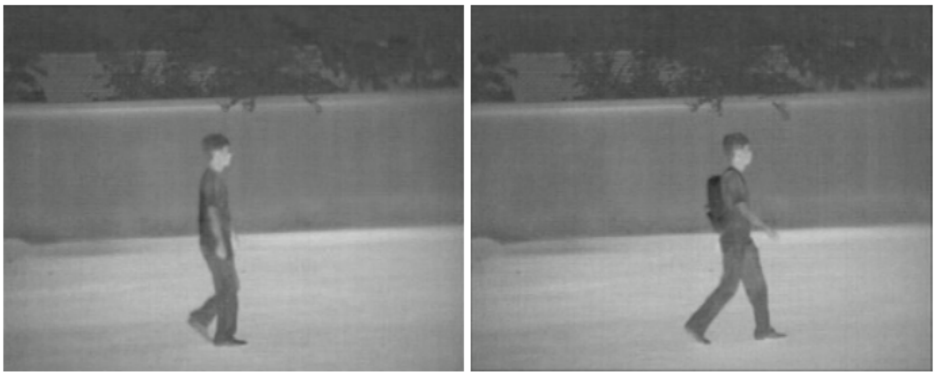


Figure 2.7: Gait images from CASIA-C dataset [Tan et al., 2006] of a subject collected at night environment using infrared cameras two different walking conditions: normal walking and carrying condition, from left to right

CASIA-C

The CASIA-C dataset [Tan et al., 2006] was collected at night time using infrared cameras, with 153 subjects in three different speeds (i.e., slow/normal/fast walking) and a carrying condition. Two gait images from this dataset are shown in Fig. 2.7.

TUM-GAID

This TUM-GAID dataset [Hofmann et al., 2014] simultaneously contains RGB images, depth images, and audio of 305 subjects in total. In [Hofmann et al., 2014], Hofmann et al. designed an experimental protocol (based on 155 subjects) to evaluate the robustness of algorithms against covariate factors like shoe, carrying condition (5kg backpack), elapsed time (January/April) which also potentially includes changes in clothing, lighting condition, etc. Several gait images from this dataset are shown in Fig. 2.8.

USF

The USF dataset [Sarkar et al., 2005] is a large outdoor gait database consisting of 122 subjects. A number of covariate factors are considered: camera viewpoints, shoes, surface types, carrying conditions, elapsed time, and clothing. Several gait images from this dataset are shown in Fig. 2.9.

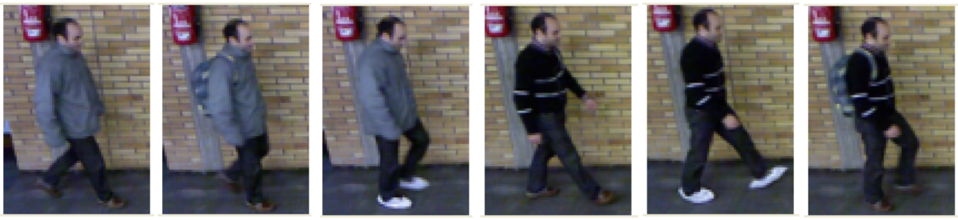


Figure 2.8: Gait images from TUM-GAID dataset [Hofmann et al., 2014] of a subject with six different walking conditions: normal walking (in Jan.), with a backpack (in Jan.), with a different pair of shoes (in Jan.), normal walking (in April), with a backpack (in April), with a different pair of shoes (in April), from left to right

Figure 2.9: Gait images in the outdoor environment from the USF dataset [Sarkar et al., 2005]

2.3 Related Work

In this section, we mainly introduce gait recognition algorithms that were designed to reduce the effect of the afore-mentioned covariates, i.e., speed, view, covariates caused by low video quality, and other covariates with unpredictable nature such as clothing, carrying condition, etc. We also introduce several methods on gait-based multimodal fusion, which drew increasingly attentions recently.

2.3.1 Gait Recognition with Speed Covariate

Current speed-invariant gait recognition algorithms fall into two categories [Kusakunniran et al., 2012a]: 1) methods relying on speed normalisation and, 2) methods relying on speed-invariant features. The first category is to transform the features from different speeds into a common one before matching. In [Tanawongsuwan and Bobick, 2004], Tanawongsuwa and

Bobick developed a stride normalisation procedure to map gait sequences across speeds. In [Tsuji et al., 2010], Tsuji et al. applied the view transformation model concept [Makihara et al., 2006] for walking speed-invariant gait recognition and claimed that the effect of speed changes is similar to camera viewpoint changes to some extent.

The second category is to employ (relatively) walking speed-invariant features. Liu and Sarkar developed a time normalised gait feature based on the hidden Markov model, which suggests certain insensitiveness to walking speed changes [Liu and Sarkar, 2006]. In order to handle variations in walking speed, the feature template, head and torso image (HTI) was proposed, which removes the unstable leg parts from silhouettes [Tan et al., 2006]. In [Tan et al., 2007c],[Tan et al., 2007a], Tan et al. defined gait signatures through projecting the silhouette onto different directions, and claimed that these signatures lead to reasonable results in cross-speed walking gait recognition experiments. In [Kusakunniran et al., 2009a], an adaptive weighting technique named weighted binary pattern (WBP) was applied to the rescaled GEI. Competitive performance in the cross-speed walking gait recognition tasks was observed. More recently, methods based on procrustes shape analysis (PSA) [Kusakunniran et al., 2011][Kusakunniran et al., 2012a] have shown great potentials in handling the large changes of walking speed. Kusakunniran et al. proposed higher-order derivative shape configuration (HSC) to extract speed-invariant gait features from the P-SA descriptors [Kusakunniran et al., 2011]. Based on the HSC framework, a differential composition model (DCM) was proposed, which can adaptively assign weights to different body parts [Kusakunniran et al., 2012a]. Compared with HSC, the introduction of DCM delivers significant performance improvement (against large walking speed changes), yet this adaptive weighting scheme has two limitations: 1) it requires an additional training set that covers all the representative speeds; 2) it is highly correlated to the degree of speed changes and required external information (e.g., video frame-rate) when the absolute walking speed is not available [Kusakunniran et al., 2012a]. Nevertheless, methods belonging to the second category can achieve reasonable performance in some gait recognition scenarios under the influences of walking speed changes.

Compared with walking gait recognition, there are only a few works on running gait recognition [Yam et al., 2001],[Yam et al., 2002b],[Yam et al., 2002a],[Yam et al., 2004],[Iosifidis et al., 2012], which was claimed to be more potent (than walking gait recognition) [Yam et al., 2002a],[Yam et al., 2004],[Iosifidis et al., 2012]. For the cross-mode gait recognition (e.g., to identify an unknown runner solely using the walking gallery), Yam et al. [Yam et al., 2002a] reported that there is a unique mapping between the walking and the running gait patterns for each subject. However, they also pointed out that the generic mapping across the population does not exist, since walking/running is highly individual-related. Based on their approach [Yam et al., 2002a], it is unlikely for the unknown runner to be identified given only the walking gait gallery.

2.3.2 Gait Recognition with Camera Viewpoint Covariate

The covariate camera viewpoint may affect the gait features in a global manner. Although performance may be relatively stable when the view changes are small (e.g., less than 18°), it may drop significantly when the view differences between the reference gait (i.e., gallery) and query gait (i.e., probe) are large (e.g., more than 36°) [Yu et al., 2006]. Given the fact that camera viewpoint is one of the most common covariates in real-world scenarios, it is desirable to have gait recognition algorithms that are less sensitive to large view changes.

Cross-view gait recognition methods can be roughly divided into three categories. Methods belong to the first category [Ariyanto and Nixon, 2011] [Zhao et al., 2006] are based on 3D reconstruction through images from multiple calibrated cameras. Although promising performance can be achieved, one major limitation is that a fully controlled and cooperative multiple camera environment is required, which may be less practical in real-world surveillance scenarios.

Methods in the second category perform view normalisation on gait features before matching. In [Goffredo et al., 2010], after estimating the poses of lower limbs, Goffredo et al. extracted the rectified angular measurements and trunk spatial displacements as gait features. Although competitive performance can be achieved when the view change is large, this method has two disadvantages: 1) It has much lower accuracies than many other methods (e.g., [Bashir et al., 2010b],[Kusakunniran et al., 2009b],[Kusakunniran et al., 2010]) when the view change is small, since such model-based features only provide limited information. 2) It is not applicable when the poses of the lower limbs are hard to estimate (e.g., frontal/back view). Recently, Kusakunniran et al. [Kusakunniran et al., 2013] proposed a view normalisation framework based on domain transformation obtained through invariant low-rank textures (TILT). They claimed that this method can normalise gaits from arbitrary views to a common canonical view (close to the side view). Although reasonable results can be achieved, it is not applicable when views are too different from the canonical view, e.g., frontal/back view.

The third category is to learn the mapping/projection relationships of gaits across views. The learning process relies on the training data that covers the views appearing in the gallery and probe. Through the learned metric(s), gaits from two different views can be projected onto the common subspace for matching. In [Makihara et al., 2006], Makihara et al. introduced the SVD-based view transformation model (VTM) to project gait features from one view into another. To avoid oversizing and overfitting of VTM, Kusakunniran et al. used truncated SVD (TSVD) [Kusakunniran et al., 2009b]. After pointing out the limitations of SVD-based VTM, Kusakunniran et al. reformulated VTM construction as a regression problem [Kusakunniran et al., 2010]. Instead of using the global features (e.g.,[Makihara et al., 2006],[Kusakunniran et al., 2009b]), local region of interest (ROI) was selected based on local motion relationship to build VTMs through support vector regression (SVR). In [Kusakunniran et al., 2012b], the performance was further improved by replacing SVR with sparse regression (SR). Instead of projecting gait features onto a common space, Bashir et al. [Bashir et al., 2010b] used canonical correlation analysis (CCA) to project gaits from two different views onto two subspaces with maximal correlation. The correlation strength was employed as the similarity measure for identification. In [Kusakunniran et al., 2014], after claiming there may exist some weakly or uncorrelated information in the global gaits across views, Kusakunniran et al. carried out motion co-clustering to partition the global gaits into multiple groups of gait segments. For feature extraction, they performed CCA on these multiple groups [Kusakunniran et al., 2014], instead of the global gait features as in [Bashir et al., 2010b]. Different from most works with multiple trained projection matrices for different view pairs (e.g., [Makihara et al., 2006][Kusakunniran et al., 2009b],[Kusakunniran et al., 2010],[Kusakunniran et al., 2012b],[Kusakunniran et al., 2014],[Bashir et al., 2010b]), recently, Hu et al. proposed a novel unitary linear projection named view-invariant discriminative projection (ViDP) [Hu et al., 2013]. The unitary nature of ViDP makes cross-view gait recognition possible without knowing the query gait views. For cross-view gait recognition, Hu proposed a novel gait representation named enhanced Gabor gait (EGG) [Hu, 2013], which encodes both statistical property and structure characteristics through a non-linear mapping. They used the regularised local tensor discriminant analysis (RLTDA) for dimensionality reduction, which can capture the nonlinear manifolds that are robust against view changes. Since RLTDA is sensitive to the initialisation, a number of RLTDA learners were further fused at score-level for higher performance. In [Hu, 2014], by considering the spatial structure information within each gait sample and local geometry information among multiple gait samples, a classification method named uncorrelated multilinear sparse local discriminative canonical correlation analysis (UMSLDCCA) was proposed for cross-view gait recognition. Compared with the other two categories of cross-view gait recognition, the third category can be performed 1) in a less controlled and non-cooperative environment and 2) when the views are significantly different from the side view (e.g., frontal/back view).

2.3.3 Gait Recognition with other Unknown Covariates

Based on concatenated GEIs, Han and Bhanu used PCA and LDA for feature extraction [Han and Bhanu, 2006]. In [Li et al., 2008], Li et al. proposed a discriminant locally linear embedding (DLLE) framework for feature extraction, which can preserve the local manifold structure. Both methods yield higher performance than the GEI matching method [Liu and Sarkar, 2004]. By using two subspace learning methods named coupled subspaces analysis (CSA) and discriminant analysis with tensor representation (DATER), Xu et al. extracted features directly from GEIs [Xu et al., 2006]. They demonstrated that the matrix representation can yield much higher performance than the vector representation reported in [Han and Bhanu, 2006]. Similarly, matrix-based marginal fisher analysis (MMFA) was proposed to extract more discriminant features from GEIs to boost the performance [Xu et al., 2007]. In [Tao et al., 2007], after convolving a number of Gabor functions with the GEI representation, Tao et al. used the Gabor-filtered GEI as a new gait feature template. They also proposed the general tensor discriminant analysis (GTDA) for extracting on the high-dimensional Gabor features [Tao et al., 2007]. To preserve the local manifold structure of the high-dimensional Gabor features, Chen et al. proposed a tensor-based Riemannian manifold distance-approximating projection (TRIMAP) framework [Chen et al., 2010]. Since spatial misalignment may degrade the performance, based on Gabor representation, image-to-class distance was utilised in [Huang et al., 2010] to allow feature matching to be carried out within a spatial neighborhood. By using the techniques of universal background model (UBM) learning and maximum a posteriori (MAP) adaptation, Xu et al. proposed the Gabor-based patch distribution feature (Gabor-PDF) in [Xu et al., 2012], and the classification was performed based on locality-constrained group sparse representation (LGSR). Compared with GEI features (e.g., [Han and Bhanu, 2006],[Xu et al., 2006]), Gabor features (e.g., [Tao et al., 2007],[Huang et al., 2010],[Xu et al., 2012]) tend to be more discriminant and can yield higher accuracies. However, due to the high dimensionality of Gabor features, these methods normally require high computational costs. Recently, several gait recognition methods based on sparse representation classification (SRC) [Wright et al., 2009] were proposed. Lu et al. proposed the sparse reconstruction based metric learning (SRML) method for gait recognition, which can minimise the intra-class sparse reconstruction errors and maximise the inter-class sparse reconstruction errors simultaneously [Lu et al., 2014]. Sparse bilinear discriminant analysis (SBDA), which extends the matrix representation based discriminant analysis methods (e.g.,[Xu et al., 2006],[Xu et al., 2007]) to sparse cases, was proposed in [Lai et al., 2014]. However, higher performance can only be gained for SRC-based methods when the training data is extensive enough to span the conditions that might occur in the test set [Wright et al., 2009]. Based on gait training data obtained in the normal condition, the accuracies in [Lu et al., 2014],[Lai et al., 2014] may drop rapidly when the query gaits in significantly different walking conditions.

By using the ”cutting and fitting” scheme, Han and Bhanu [Han and Bhanu, 2006] generated synthetic GEI templates to simulate the effect of the walking surface covariate. In [Wang et al., 2012], Wang et al. proposed a feature template, named chrono-gait image (C-GI), with the temporal information of the gait sequence encoded. Although results reported in [Iwama et al., 2012],[Wang et al., 2012] suggest that CGI and GEI may have similar performance in most cases, in tackling the carrying condition covariate, CGI outperforms GEI significantly [Wang et al., 2012]. Liu and Sarker [Liu and Sarkar, 2006] employed the population hidden Markov model (pHMM) to generate dynamics-normalised gait cycles for gait recognition, and encouraging performance against the walking surface covariate was observed.

Most existing algorithms perform unsatisfactorily when elapsed time is taken into consideration, as elapsed time potentially also includes the changes of clothing, walking surface, etc. Irrespective of other covariates, in [Matovski et al., 2012], Matovski et al. investigated the effect of elapsed time on a small gait dataset and found that short term elapsed time does not affect the recognition significantly. They claimed that clothing may be the most challenging covariate factor for appearance-based methods [Matovski et al., 2012]. Based on a newly constructed gait dataset consisting 32 different clothes combinations, Hossain et al. proposed an adaptive scheme for weighting different body parts in order to reduce the effect of clothing [Hossain et al., 2010]. However, this method requires an additional training set that covers all the possible clothes types, which is less practical in real-world applications.

2.3.4 Gait-based Fusion Methods

The performance of a gait recognition system is limited when facing large intra-class gait variations, which may be attributed to walking surface, clothing, elapsed time, etc. Multimodal fusion is a solution to reduce the error rate and has been widely applied to the biometrics field [Jain et al., 2004b], e.g., face+fingerprint [Rattani et al., 2007], face+iris [Wang et al., 2003b]. For human identification at a distance, it is natural to fuse gait and face, which can be acquired using the same camera. Intuitively, gait is less sensitive to low resolution, illumination, hair style, while face is robust to the changes or walking surface and clothing. As such, gait and face may be complementary for recognition.

Compared with remote human identification based on gait or face recognition, the technology for fusing gait and face is still at the early stage. In the work [Shakhnarovich and Darrell, 2002], after applying canonical view rendering technique (CVRT), face and gait information from multiple camera views were fused at the score level. By doing so, significant performance improvement is observed. In [Kale et al., 2004a], Kale et al. showed that even in a single camera environment, directly combining the scores of face and gait can boost the overall performance. Based on population hidden Markov model (pHMM), Liu and Sarkar selected gait stances for recognition in the outdoor environment [Liu and Sarkar, 2007]. Based on different fusion strategies, extensive experiments were conducted on images with variations in the walking surface and time. They found that the performance is higher when fusing gait and face than intra-model fusion (i.e., face+face or gait+gait) [Liu and Sarkar, 2007]. Since the reliability of face and gait varies with different subject-camera distances, Geng et al. proposed an adaptive score-level fusion scheme [Geng et al., 2007]. The weights of the face and gait scores are distance-driven. It has been empirically shown to outperform score-level fusion with fixed weights in the multi-view environment. In [Zhou and Bhanu, 2007], Zhou and Bhanu performed a score-level fusion of gait and the enhanced side face image (ESFI). Compared with original side face image (OSFI), they found that improving face image quality can further enhance the recognition rate. They further applied feature-level fusion by concatenating the ESFI and gait [Zhou and Bhanu, 2008]. Recently, Alpha matte was used by Hofmann et al. to segment gait and face images with better qualities, before a score-level fusion [Hofmann et al., 2012].

Gait can also be combined with gait from different feature spaces [Han and Bhanu, 2006] or cameras/sources [Matovski et al., 2012],[Hofmann et al., 2014]. In [Han and Bhanu, 2006], GEI and synthetic GEI templates were generated from the same silhouettes to different feature spaces, and they were fused in the score level. In [Hofmann et al., 2014], GEI, depth, and audio information were fused to tackle the elapsed time challenges and encouraging performance was achieved. In [Matovski et al., 2012], by concatenating gait information from three cameras from different views into a new template, promising results were achieved on a small temporal dataset.

2.3.5 Gait Recognition from Low Frame-rate Videos

In [Mori et al., 2010], phase synchronisation was used and gait probe videos with low frame-rate were matched with high frame-rate gallery videos. However, this approach fails whenever both the probe and gallery samples are from low frame-rate videos. Several temporal reconstruction-based methods were proposed to deal with such dual low frame-rate problem. In [Al-Huseiny et al., 2010], Al-Huseiny et al. proposed a level-set morphing approach for temporal interpolation. In [Makihara et al., 2011], a temporal super resolution (SR) method was employed to build a high frame-rate gait sequence period based on the multiple periods of low frame-rate gait sequences. Based on [Makihara et al., 2011], Akae et al. [Akae et al., 2011] applied an exemplar of high frame-rate image sequences to improve the temporal SR quality. A unified framework of example-based and reconstruction-based periodic temporal SR was proposed in [Akae et al., 2012]. It works well for gait recognition from (dual) extremely low frame-rate videos (with small gait fluctuations). Most of these works attempt to recover the high frame-rate gait sequences in the first place. However, in the applications of human identification given extremely low frame-rate videos (e.g., 1 fps), they either have low performance due to the generated high levels of reconstruction artifacts [Al-Huseiny et al., 2010],[Akae et al., 2011] or have to assume that the amount of motion among the gait periods is the same, which is not feasible when there are large gait fluctuations [Akae et al., 2012].

2.4 Fundamental Knowledge for This Thesis

2.4.1 Gait Feature Templates

Given the binary aligned silhouettes, several feature templates can be defined for classification, as reviewed in Chapter 2.1.2. Since our contribution in this thesis is not to propose new gait templates, we only employ two most popular feature templates, i.e., GEI and Gabor-filtered GEI, which are introduced below.

GEI

GEI is a popular feature template widely used in recent gait recognition algorithms due to simplicity and effectiveness. GEI is the average silhouette over one gait cycle, which encodes a number of binary silhouettes into a grayscale image, and it has two main advantages: 1) the segmentation noises can be smoothed; 2) the computational cost can be significantly reduced [Han and Bhanu, 2006]. Given the aligned binary gait silhouette images $\mathbf{B}_t$ for the t^{th} frame in a sequence, GEI is defined as follows:

$$\mathbf{GEI} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{B}_t,$$

where T is the number of frames in a gait cycle [Han and Bhanu, 2006].

Another form of GEI is to average the silhouettes over the whole gait sequence, instead of cycles. It is often referred to as average gait image (AGI) [Veres et al., 2004], which is defined as:

$$\mathbf{AGI} = \frac{1}{\hat{T}}\sum_{t=1}^{\hat{T}}\mathbf{B}_t,$$

where $\hat{T}$ is the total number of frames in a gait sequence. It is suitable for the cases when there are only a few frames or the gait cycle is hard to estimate. We will use AGI in Chapter 5 for low frame-rate gait recognition.

In this thesis, unless otherwise specified, we use the GEI feature template (e.g., in datasets OU-ISIR-A, OU-ISIR-B, CASIA-B, CASIA-C, TUM-GAID, and USF) with default resolution 128×88 pixels.

Gabor filtered GEI (Gabor)

For the Gabor template, a family of Gabor functions at a given pixel z are defined as:

$$\Psi_{\tau,\nu}(z) = \frac{\|\gamma_{\tau,\nu}\|^2}{\delta^2} e^{-\frac{\|\gamma_{\tau,\nu}\|^2 \|z\|^2}{2\delta^2}} [e^{i\gamma_{\tau,\nu} z} - e^{-\frac{\delta^2}{2}}],$$

where $\gamma_{\tau,\nu} = \theta_\tau e^{i\phi_\nu}$ is the frequency vector that determines the scale and the orientation of the Gabor functions. $\theta_\tau = 2^{-(\tau+2)/2}\pi$ determines the scale and $\phi_\nu = \pi\nu/8$ determines the direction. As in [Tao et al., 2007], we let $\delta = 2\pi$, $\tau = \{0, 1, 2, 3, 4\}$, and $\nu = \{0, 1, 2, 3, 4, 5, 6, 7\}$. With 5 scales and 8 orientations, we can get a total of 40 Gabor functions. For each GEI with the size of 128×88 pixels, we acquire 40 Gabor-filtered images after convolving the GEI with the 40 Gabor functions. Since the entries of each filtered image are complex numbers, the corresponding magnitude values are adopted. These 40 Gabor-filtered images are used to form a Gabor of 640×704 pixels. Moreover, as suggested in [Xu et al., 2012], for computational efficiency we further downsample each Gabor to a size of 320×352 pixels, since the Gabor features extracted from the neighboring pixels are generally redundant to some extent [Xu et al., 2012].

In this thesis, unless otherwise specified, we use the Gabor feature templates (e.g., in datasets OU-ISIR-A, OU-ISIR-B, CASIA-B, CASIA-C, TUM-GAID, and USF) with default resolution 320×352 pixels.

2.4.2 Subspace Learning

We also introduce several subspace learning methods including unsupervised methods principle component analysis (PCA), two-dimensional PCA (2DPCA) and a supervised method two-dimensional linear discriminant analysis (2DLDA).

PCA

PCA aims to find a transformation matrix $\mathbf{W}_{pca}$ for dimensionality reduction and feature decorrelation. Assume there are n samples in the training set, such that $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n] \in \mathbb{R}^{m \times n}$, where m is the original dimensionality. The covariance matrix $\mathbf{S}$ can be calculated:

$$\mathbf{S} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T, \tag{2.1}$$

where $\bar{\mathbf{x}} = \dfrac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i$. Let $\{\boldsymbol{\phi}_k\}_{k=1}^{d}$ be the eigenvectors associated with eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_d (d \leq m)$ of the following generalised eigenvalue problem: $\mathbf{S}\boldsymbol{\phi} = \lambda\boldsymbol{\phi}$. The solution $\mathbf{W}_{pca}$ to this maximisation problem is given by $\mathbf{W}_{pca} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, ..., \boldsymbol{\phi}_d]$.

2DPCA

Different from PCA, 2DPCA [Yang et al., 2004] takes the two-dimensional data (e.g., image) as the input template. It also has a different form in calculating the covariance matrix. Given n images $\mathbf{I}_i \in \mathbb{R}^{N_1 \times N_2}, i = 1, 2, ..., n$ in the training set, the covariance matrix $\mathbf{S}$ can be estimated as:

$$\mathbf{S} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{I}_i - \bar{\mathbf{I}})^T(\mathbf{I}_i - \bar{\mathbf{I}}), \tag{2.2}$$

where $\bar{\mathbf{I}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}_i$. Let $\{\boldsymbol{\phi}_k\}_{k=1}^{d}$ be the eigenvectors associated with eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_d (d \leq N_2)$ of the following generalised eigenvalue problem: $\mathbf{S}\boldsymbol{\phi} = \lambda\boldsymbol{\phi}$. The solution $\mathbf{W}_{2dpca}$ to this maximisation problem is given by $\mathbf{W}_{2dpca} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, ..., \boldsymbol{\phi}_d]$.

2DLDA

Here we introduce the training process of 2DLDA [Li and Yuan, 2005]. Given n images $\mathbf{I}_i \in \mathbb{R}^{N_1 \times N_2}, i = 1, 2, ..., n$ in the training set, it aims to find a transformation matrix $\mathbf{W}_{2dlda} \in \mathbb{R}^{N_1 \times M}, (M \leq N_1)$ that maximises the ratio of the between-class scatter matrix $\mathbf{S}_b$ to the within-class scatter matrix $\mathbf{S}_w$, i.e.,

$$\mathbf{W}_{2dlda} = \underset{\mathbf{W}^T\mathbf{W}=\mathbf{I}}{\text{argmax}} \ \text{trace}((\mathbf{W}^T\mathbf{S}_w\mathbf{W})^{-1}(\mathbf{W}^T\mathbf{S}_b\mathbf{W})), \tag{2.3}$$

where $\mathbf{I}$ is the identity matrix. Let $\boldsymbol{\mu} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{I}_i$ be the global centroid, $\mathcal{D}_j$ be the j^{th} class (out of c classes) with n_j samples, such that $n = \sum_{j=1}^{c} n_j$. The between-class scatter matrix $\mathbf{S}_b$ and the within-class scatter matrix $\mathbf{S}_w$ in Eq.(2.3) are defined as follows:

$$\begin{aligned} \mathbf{S}_b &= \sum_{j=1}^{c} n_j(\mathbf{m}_j - \boldsymbol{\mu})(\mathbf{m}_j - \boldsymbol{\mu})^T, \\ \mathbf{S}_w &= \sum_{j=1}^{c} \sum_{\mathbf{I}_i \in \mathcal{D}_j} (\mathbf{I}_i - \mathbf{m}_j)(\mathbf{I}_i - \mathbf{m}_j)^T, \end{aligned} \tag{2.4}$$

where $\mathbf{m}_j$ is the within-class centroid for $\mathcal{D}_j$. Let $\{\boldsymbol{\phi}_k\}_{k=1}^{M}$ be the eigenvectors associated with eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_M (M \leq N_1)$ of the following generalised eigenvalue problem: $\mathbf{S}_b\boldsymbol{\phi} = \lambda\mathbf{S}_w\boldsymbol{\phi}$. The solution $\mathbf{W}_{2dlda}$ to this maximisation problem is given by $\mathbf{W}_{2dlda} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, ..., \boldsymbol{\phi}_M]$, which can perform feature extraction in the row direction of the images.

2.5 Random Subspace Method (RSM): Properties and Applications

Overfitting is a common problem for learning-based methods when the training data is less representative, or has large dimensionality-to-sample ratio (also known as ”curse of dimensionality”). The concept of random subspace method (RSM) offers a way to address this issue [Ho, 1998]. Based on the stochastic discrimination theory [Kleinbery, 1996], RSM is constructed by combining many classifiers that have weak discriminant ability but can generalise well.

By splitting the original feature space randomly into a number of subspaces with lower dimensions, random sampling strategy is adopted for constructing decision forests [Ho, 1998]. For each base classifier, only the selected features have nonzero contribution to the similarity measurement, and the final classification decision is made by aggregating the results from the weak classifiers. By ignoring some dimensions in the feature space, the classification is less sensitive to query samples that are different from the training samples only in the unselected dimensions [Ho, 1998]. RSM can effectively deal with the ”curse of dimensionality”, since the dimensionality is greatly reduced by using different feature subsets to generate multiple classifiers [Ho, 1998]. In [Skurichina and Duin, 2002], LDA was further employed in the random feature subsets to improve the performance of the weak classifiers. Since each subspace has much lower feature dimensions after random sampling, the undersampled problem of LDA can be avoided [Skurichina and Duin, 2002]. In the context of face recognition, eigenfaces were employed as candidates to construct the random subspaces for feature projection [Wang and Tang, 2004], [Wang and Tang, 2006]. In [Chawla and Bowyer, 2005], more experimental results were reported to support the effectiveness of RSM in the application of face recognition, which has great generalisation ability to unseen data. In [Kuncheva et al., 2010], RSM was employed to classify functional magnetic resonance imaging (fMRI) data, which faces an overfitting problem due to the extremely large dimensionality-to-sample ratio. Most recently, RSM was used to address partial occlusions for human detection in still images [Marin et al., 2014]. Based on a representative validation set, a classifier filtering scheme is employed to select the most discriminant ones to form the optimal ensemble [Marin et al., 2014].

RSM is a general concept and there are various ways for implementation. The random subspaces can be constructed by sampling the original features [Ho, 1998], holistic features (e.g., eigenface [Chawla and Bowyer, 2005], [Wang and Tang, 2006]), or image blocks [Marin et al., 2014]. In each random subspace, the features can be further extracted using supervised methods such as LDA [Skurichina and Duin, 2002], [Wang and Tang, 2006]. The base classifiers can be decision trees [Ho, 1998], nearest neighbour (NN) classifiers [Chawla and Bowyer, 2005], [Wang and Tang, 2006], support vector machine (SVM) based classifiers [Marin et al., 2014], etc. There are also several popular classifier combination rules, such as (weighted) sum rule [Chawla and Bowyer, 2005], [Wang and Tang, 2006], [Marin et al., 2014] or majority voting [Wang and Tang, 2006].

Chapter 3

Robust Speed-Invariant Gait Recognition using Random Subspace Method (RSM)

In real-world scenarios, walking/running speed is one of the most common covariate factors that can affect the performance of gait recognition systems. By assuming the effect caused by the speed changes (from the query walkers/runners) are intra-class variations that the training data (i.e., gallery) fails to capture, overfitting the less representative training data may be the main problem that degrades the performance. In this chapter, we introduce the gait-based random subspace method (RSM) to solve this problem. For query gaits with unknown walking/running speeds, we try to reduce the generalisation errors by combining a large number of weak classifiers. We evaluate our method on two benchmark databases, i.e., CASIA-C dataset and OU-ISIR-A dataset. For the cross-speed walking/running gait recognition experiments, nearly perfect results are achieved, significantly higher than other state-of-the-art algorithms. We also study the unknown-speed runner identification solely using the walking gait gallery, and encouraging experimental results suggest the effectiveness of our method in such cross-mode gait recognition tasks.

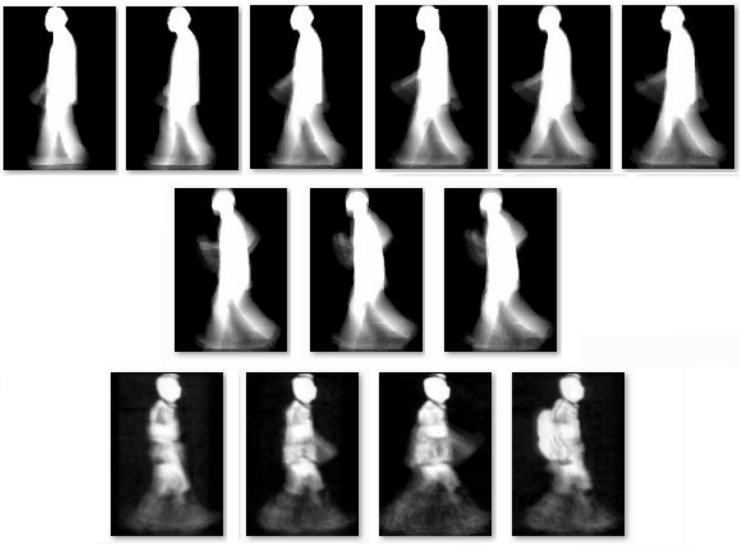


Figure 3.1: GEI samples from the OU-ISIR-A and CASIA-C datasets. Top row: GEIs from OU-ISIR-A dataset with walking speed ranging from 2km/h to 7km/h (from left to right with an 1km/h interval); middle row: GEI samples from OU-ISIR-A dataset with running speed ranging from 8km/h to 10 km/h (from left to right with 1km/h interval); bottom row: from CASIA-C dataset, the GEIs with slow/normal/fast walking speed, and a GEI sample with a bag (from left to right).

3.1 Problem Statement and Motivation of using RSM

In real-world scenarios, speed is one of the most common covariate factors. In recent years several methods have been proposed to solve the cross-speed (walking) gait recognition, yet most of them are not applicable when the speed changes are large. Although it was claimed that differential composition model (DCM) [Kusakunniran et al., 2012a] can tackle the large speed changes, this adaptive weighting scheme has two main limitations. That is, it requires 1) an additional multi-speed training set and 2) external information when the absolute walking speed is not available [Kusakunniran et al., 2012a], as reviewed in Chapter 2.3.1. These two requirements make DCM sensitive to other covariates (e.g., bag), which limits its application in real-world scenarios. It is desirable to build a robust system with higher performance and less limitations.

To begin with, we summarise the characteristics of speed changes by using the GEI examples from OU-ISIR-A dataset [Tsuji et al., 2010] and CASIA-C dataset [Tan et al., 2006]. Fig. 3.1 illustrates the effect of speed variations on GEIs, from which we can observe:

1. For the fixed-mode gait recognition (walking only or running only), the static parts of the body are relatively independent of walking/running speed changes.
2. For the cross-mode gait recognition, there may be some similar patterns between fast walking and running, e.g., head, neck, and hip.

These observations are consistent with the claims in [Tsuji et al., 2010], i.e., although the dynamic gait features can be significantly affected by speed changes, the static features can be relatively stable. Intuitively, we may perform classification based on static features only. However, they tend to be sensitive to other covariates, e.g., bag [Tan et al., 2006]. On the other hand, when the speed change is small, it may be beneficial to take advantage of the useful dynamic information (e.g., the thigh area).

We notice that for most algorithms, performance is high when the training data (i.e., gallery) and test data (i.e., probe) are of similar speeds, and vice versa. This phenomenon is a typical overfitting problem from the perspective of learning. For test data of various walking speeds (e.g., 2km/h or 7km/h), the training data of a certain speed (e.g.,4km/h) does not represent the whole population. For gait recognition in real-world scenarios, it is difficult to collect representative training data that covers all possible covariates, especially for the subject-related covariates such as speed, carrying condition, clothing, etc. In this case, overfitting the less representative training data may be the major problem that hamper the performance of the methods that relying on learning. To address this issue, we employ the RSM concept to enhance the generalisation accuracy [Ho, 1998]. As introduced in Chapter 2.5, RSM was initially proposed by Ho [Ho, 1998] to build random decision forests and later this concept was applied to different classification tasks such as face recognition, [Wang and Tang, 2006], fMRI classification, [Kuncheva et al., 2010], human detection in still images [Marin et al., 2014], etc. For each base classifier, only the randomly selected features have nonzero contribution to the similarity measurement, and the classification is less sensitive to query samples that are different from the training samples only in the unselected dimensions. We take advantage of this property of RSM to design a speed-invariant gait recognition system in this chapter.

3.2 Speed-invariant Gait Recognition System

We use Gabor-filtered GEIs (referred to as Gabor) as the input feature template, which was regarded as an effective feature template for human gait recognition [Tao et al., 2007] [Xu et al., 2012]. Given a GEI sample, Gabor functions from five scales and eight orientations are employed to generate the Gabor feature template. For computational efficiency, similar to [Xu et al., 2012], we use the downsampled Gabor templates in this thesis. More details about Gabor templates can be found in Chapter 2.4.1.

3.2.1 Random Subspace Construction

Due to the high dimensionality of Gabor templates, we apply 2DPCA to decorrelate the features (in column direction), before using the corresponding basis vectors for random subspace construction. The reasons of using 2DPCA, instead of conventional PCA are two-fold: 1) 2DPCA has a much lower time complexity [Yang et al., 2004]; 2) the input features of 2DPCA are of matrix form, which may preserve the data structure to some extent [Yang et al., 2004], and empirical results shows that 2DPCA normally outperforms the conventional PCA in face recognition [Yang et al., 2004] and gait recognition [Zhang et al., 2010].

Given n gait templates (e.g., Gabor templates) $\mathbf{I}_i \in \mathbb{R}^{N_1 \times N_2}, i = 1, 2, ..., n$ in the training set, the covariance matrix $\mathbf{S}^*$ can be estimated as:

$$\mathbf{S}^* = \frac{1}{n} \sum_{i=1}^{n} (\mathbf{I}_i - \bar{\mathbf{I}})^T (\mathbf{I}_i - \bar{\mathbf{I}}), \tag{3.1}$$

where $\bar{\mathbf{I}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{I}_i$. The eigenvectors of $\mathbf{S}^*$ can be computed to decorrelate the features. d eigenvectors associated with the non-zero eigenvalues are retained as candidates $\mathbf{T} =$

$[\mathbf{t}_1, \mathbf{t}_2, ..., \mathbf{t}_d] \in \mathbb{R}^{N_2 \times d}$ to construct the random subspaces. By repeating L times the process of randomly selecting subsets of $\mathbf{T}$ (with size $N \ll d$), the random subspaces $\mathbf{R}^l \in \mathbb{R}^{N_2 \times N}, l = 1, 2, ..., L$ are generated and can be used as random feature extractors. The process of random subspace construction is summarised in Algorithm 3.1.

Algorithm 3.1 Random Subspace Construction

```
Input: Training data (i.e., gallery) $\{\mathbf{I}_i \in \mathbb{R}^{N_1 \times N_2}\}_{i=1}^n$, subspace dimensionality $N$, subspace number $L$;
Output: Random transformation matrices: $\mathbf{R}^l \in \mathbb{R}^{N_2 \times N}, l = 1, 2, ..., L$;
  Step 1: Calculating $\mathbf{S}^* = \frac{1}{n}\sum_{i=1}^n (\mathbf{I}_i - \bar{\mathbf{I}})^T (\mathbf{I}_i - \bar{\mathbf{I}})$, where $\bar{\mathbf{I}} = \frac{1}{n}\sum_{i=1}^n \mathbf{I}_i$;
  Step 2: For $\mathbf{S}^*$, calculating its $d$ largest non-zero eigenvectors $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, ..., \mathbf{t}_d] \in \mathbb{R}^{N_2 \times d}$;
  for $l = 1$ to $L$ do
    Step 3: Forming $\mathbf{R}^l \in \mathbb{R}^{N_2 \times N}$ by randomly selecting subsets of $\mathbf{T}$ (with size $N \ll d$);
  end for
```

Given the random feature extractors $\mathbf{R}^l, l = 1, 2, ..., L$, a gait template $\mathbf{I}$ can be represented as L sets of projection coefficients $\mathbf{X}^l \in \mathbb{R}^{N_1 \times N}, l = 1, 2, ..., L$, i.e.,

$$\mathbf{X}^l = \mathbf{I}\mathbf{R}^l, \qquad l = 1, 2, ..., L. \tag{3.2}$$

Because the random feature extraction process is only performed in the column direction, the dimensionality is reduced from $N_1 \times N_2$ to $N_1 \times N$.

3.2.2 Local Enhancer 1 (LE1)

The random features can be used for classification directly. However, these features may be redundant and less discriminant since 1) the random feature extraction process based on Eq.(3.2) is performed only in the column direction, 2) the random feature extractor is trained only in an unsupervised manner, without using the label information. As a result, these features may lead to low performance in terms of both computational costs and recognition accuracies. To address this issue, we take advantage of the label information. Supervised learning can be used for each subspace to extract more discriminant features, and we name this process as local enhancing (LE). In this chapter, two-dimensional LDA (2DLDA) [Li and Yuan, 2005] is employed for LE, and the corresponding feature extractor is referred to as local enhancer 1 (LE1). [1] The process of generating the 2DLDA-based LE1 is summarised in Algorithm 3.2, and more details about 2DLDA can be found at [Li and Yuan, 2005].

Algorithm 3.2 LE1

```
Input: Gallery $\{\mathbf{I}_i \in \mathbb{R}^{N_1 \times N_2}\}_{i=1}^{n}$, random transformation matrices: $\mathbf{R}^l \in \mathbb{R}^{N_2 \times N}, l = 1, 2, ..., L$, and the number of LE1 projection directions $M$;
Output: LE1-based transformation matrices: $\hat{\mathbf{W}}^l \in \mathbb{R}^{N_1 \times M}, l = 1, 2, ..., L$;
  Step 1: Random feature extraction on gallery $\mathbf{X}_i^l = \mathbf{I}_i \mathbf{R}^l, \quad i = 1, 2, ...n, \quad l \in [1, L]$;
  for $l = 1$ to $L$ do
    Step 2: Letting $\bar{\mathbf{X}}^l = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i^l$ be the global centroid, $\mathcal{D}_j^l$ be the $j^{th}$ class (out of $c$ classes) with $n_j$ samples, $\bar{\mathbf{M}}_j^l$ be the within-class centroid for $\mathcal{D}_j^l$;
    Step 3: Calculating $\mathbf{S}_b^l = \sum_{j=1}^{c} n_j(\bar{\mathbf{M}}_j^l - \bar{\mathbf{X}}^l)(\bar{\mathbf{M}}_j^l - \bar{\mathbf{X}}^l)^T$;
    Step 4: Calculating $\mathbf{S}_w^l = \sum_{j=1}^{c} \sum_{\mathbf{X}_i^l \in \mathcal{D}_j^l} (\mathbf{X}_i^l - \bar{\mathbf{M}}_j^l)(\mathbf{X}_i^l - \bar{\mathbf{M}}_j^l)^T$;
    Step 5: Setting $\hat{\mathbf{W}}^l = [\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_M]$, which are the $M$ largest eigenvectors of $(\mathbf{S}_w^l)^{-1}\mathbf{S}_b^l$;
  end for
```

In the l^{th} subspace, given the trained $\hat{\mathbf{W}}^l \in \mathbb{R}^{N_1 \times M}$ (the output of Algorithm 3.2) and a gait sample with random features $\mathbf{X}^l \in \mathbb{R}^{N_1 \times N}$, the new feature representation $\mathbf{Y}^l \in \mathbb{R}^{M \times N}$ can be extracted according to

$$\mathbf{Y}^l = (\hat{\mathbf{W}}^l)^T \mathbf{X}^l, \quad l \in [1, L]. \tag{3.3}$$

In contrast to the random feature extraction which is performed in the column direction, from Eq.(3.3) we can see that LE1 extracts the features in the row direction, and the dimensionality is further reduced to $M \times N$.

3.2.3 Classification by Majority Voting

Based on the afore-mentioned random feature extraction (by using Eq.(3.2)) and local enhancing (by using Eq.(3.3)), feature matrices can be extracted from the original gait images (i.e., Gabor templates). We further concatenate each feature matrix into its corresponding

[1] In contrast to LE2 based on IDR/QR (see Chapter 6), and LE3 based on ULDA (see Chapter 7).

vector, before classification is performed.

Nearest Mean Classifier

For gait images (e.g., GEI or Gabor templates) collected in the gallery, one assumption is often made that each subject is walking without changing the pose significantly. Based on this assumption, nearest mean (NM) classifier can be used, which takes the centroid of the each subject's samples (in the gallery) as reference [Han and Bhanu, 2006]. It is worth noting that other type of classifiers can also be used such as SVM based classifiers [Marin et al., 2014], max-margin based classiers [Maji and Berg, 2009], etc. For simplicity, here we employ the NM classifier for each subspace.

For the l^{th} subspace, let $[\mathbf{u}_1^l, \mathbf{u}_2^l, ..., \mathbf{u}_c^l]$ be the centroids of feature vectors corresponding to the c subjects in the gallery. For a query gait sequence $\mathbf{P}^l$ with n_p samples (e.g., GEI or Gabor templates) with the corresponding feature vectors $[\mathbf{p}_1^l, \mathbf{p}_2^l, ..., \mathbf{p}_{n_p}^l]$, the distance between $\mathbf{P}^l$ and the j^{th} class centroid $\mathbf{u}_j^l$ is defined as:

$$\delta(\mathbf{P}^l, \mathbf{u}_j^l) = \frac{1}{n_p} \sum_{i=1}^{n_p} \|\mathbf{p}_i^l - \mathbf{u}_j^l\|, \quad j \in [1, c]. \tag{3.4}$$

Then we assign the label to the class with the shortest Euclidean distance. Let $\{\omega_j\}_{j=1}^c$ be class labels of c subjects in the gallery, then the output label $\Omega^l(\mathbf{P}^l) \in \{\omega_j\}_{j=1}^c$ of the l^{th} base classifier can be expressed as:

$$\Omega^l(\mathbf{P}^l) = \underset{\omega_j}{\text{argmin}}\, \delta(\mathbf{P}^l, \mathbf{u}_j^l), \quad j \in [1, c]. \tag{3.5}$$

It is worth noting that other similarity measures (e.g., Cosine, Mahalanobis distance, etc.) can also be used. However, previous works suggest Euclidean distance is most effective and efficient one in gait recognition [Liu and Sarkar, 2004],[Sarkar et al., 2005],[Han and Bhanu, 2006],[Bashir et al., 2009],[Zhang et al., 2010],[Guan et al., 2012b],[Wang et al., 2012],[Guan et al., 2012a].

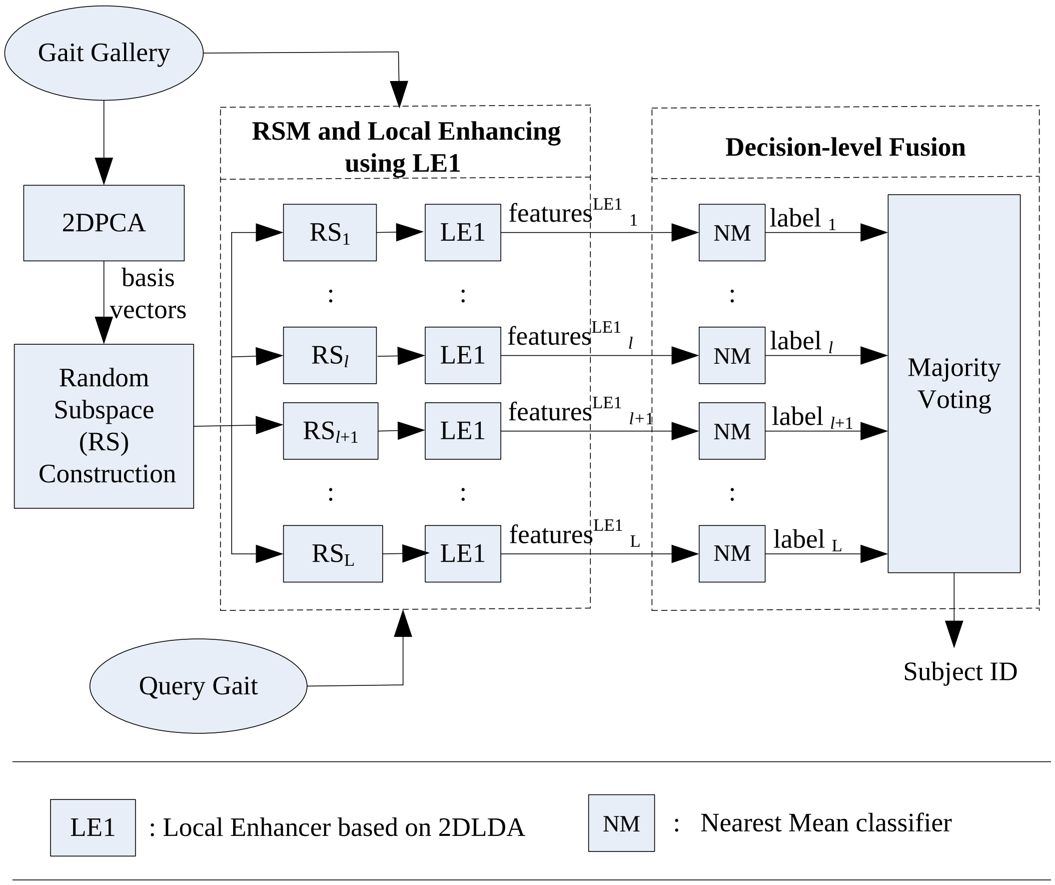


Figure 3.2: Gait-based RSM

Majority Voting

The final classification decision can be made through majority voting [Kittler et al., 1998] among the predicted labels of the L classifiers. Given a query gait $\mathbf{P}$, the optimal class label $\hat{\Omega}(\mathbf{P})$ is:

$$\hat{\Omega}(\mathbf{P}) = \underset{\omega_j}{\text{argmax}} \sum_{l=1}^{L} \triangle_{\omega_j}^{l}, \quad j \in [1, c],$$

where

$$\triangle_{\omega_j}^{l} = \begin{cases} 1, & \text{if } \Omega^l(\mathbf{P}^l) = \omega_j, \\ 0, & \text{otherwise,} \end{cases} \qquad j \in [1, c]. \tag{3.6}$$

A flowchart of the gait-based RSM is shown in Fig. 3.2.

3.3 Experiments

To evaluate the robustness of our method, two benchmark datasets are used in our experiments, i.e., CASIA-C [Tan et al., 2006] and OU-ISIR-A [Tsuji et al., 2010]. The CASIA-C dataset was collected at night time using infrared cameras, with a large number of subjects (153) at three different speeds (*i.e.*, slow/normal/fast walking) and a carrying condition. The OU-ISIR-A dataset was collected on a treadmill with a large range of speeds (from 2km/h to10km/h) in terms of walking or running for 34 subjects. In this section, three experiments are designed as follows:

1. Cross-speed walker identification on the CASIA-C dataset
2. On the OU-ISIR-A dataset, fixed-mode cross-speed gait recognition (i.e., cross-speed walker/runner identification)
3. On the OU-ISIR-A dataset, cross-mode cross-speed gait recognition (i.e., cross-speed runner identification solely using walking gallery)

There are three parameters in our method, i.e., the random subspace/base classifier number (L), the random subspace dimensionality (N), and the number of LE1 projection directions (M). In [Ho, 1998], it was claimed that the accuracy does not decrease with respect to the increasing number of classifiers, and we empirically set $L = 1000$. For the random subspace dimensionality N, since the generalisation ability is inversely proportional to N [Ho, 1998], we empirically set $N = 5$. We also empirically set $M = 40$. A full discussion of the sensitivity of these three parameters are provided in Chapter 6.3.2.

We use the rank-1 correct classification rate (CCR) to evaluate the performance. Considering the random nature of our method, we run each experiments 10 times, with the mean rank-1 CCR (with the standard deviation) reported as the result. To demonstrate the effectiveness of RSM, we also implement 2DPCA+2DLDA (based on Gabor templates) for comparison in the three experiments. Specifically, we preserve 99% of the 2DPCA variance and set the number of 2DLDA's projection directions as $M = 40$.

3.3.1 Cross-speed Walker Identification on the CASIA-C Dataset

The CASIA-C dataset contains three different walking speeds and one carrying condition, i.e., slow walking (*fs*), normal walking (*fn*), fast walking (*fq*), and carrying a bag (*fb*). The gait images of all the 153 subjects are used in our experiments. For each subject, there are two sequences of *fs*, four sequences of *fn*, two sequences of *fq*, and two sequences of *fb*. For each subject, we include three *fn* sequences in the gallery set, and use the rest sequences as probes. This dataset can be used to evaluate the algorithms against the (small) walking speed changes and carrying condition. We compare our method with other classical methods, *i.e.*, gait curves (GCV) [DeCann and Ross, 2010], normalised dual-diagonal projections (NDDP) [Tan et al., 2007d], orthogonal diagonal projections (ODP) [Tan et al., 2007c], wavelet packet silhouette representation (WPSR) [Dadashi et al., 2009], HTI [Tan et al., 2006], horizontal direction projection (HDP) [Tan et al., 2007a], active energy image (AEI) [Zhang et al., 2010], Pseudoshape [Tan et al., 2007b], WBP [Kusakunniran et al., 2009a], HSC [Kusakunniran et al., 2011], DCM [Kusakunniran et al., 2012a], and the method 2DPCA+2DLDA. The corresponding experimental results in terms of rank-1 CCR are reported in Table 3.1. We can see that the rank-1 CCRs of our method are nearly 100% in all the three tasks with probe sets in different speeds (i.e., *fn*, *fs*, and *fq*), which are significantly higher than those of other state-of-the-art algorithms.

However, in this dataset, since the variations of walking speed are relatively small, the appearance of the silhouettes is less affected (see Fig. 3.1). In this case, most of the algorithms can achieve competitive performance. For robustness evaluation, we also conduct another experiment on probe *fb*, with the a different type covariate, i.e., bag. In this case, the performance of most algorithms decreases significantly. Our method consistently yields high performance, with a mean rank-1 CCR of 96.2%, significantly higher than other methods. These experimental results suggest the robustness and effectiveness of the RSM framework.

-	# subject	*fn*	*fs*	*fq*	*fb*
GCV [DeCann and Ross, 2010]	153	91	65	70	26
NDDP [Tan et al., 2007d]	153	98	84	84	16
ODP [Tan et al., 2007c]	153	98	80	80	16
WPSR [Dadashi et al., 2009]	153	93	83	85	20
HTI [Tan et al., 2006]	46	94	85	88	51
HDP [Tan et al., 2007a]	153	98	84	88	36
AEI [Zhang et al., 2010]	153	89	89	90	80
Pseudoshape [Tan et al., 2007b]	153	98	91	94	25
WBP [Kusakunniran et al., 2009a]	153	99	86	90	81
HSC [Kusakunniran et al., 2011]	50	98	92	92	-
DCM [Kusakunniran et al., 2012a]	120	97	92	93	-
2DPCA+2DLDA	153	100	97	97	71
RSM (our method)	153	**100**±0.00	**99.7**±0.24	**99.6**±0.14	**96.2**±0.86

Table 3.1: Algorithms comparison in terms of rank-1 CCR(%) on the CASIA-C dataset. *fn*, *fs*, and *fq* denote the speed types of the three probe sets (i.e., normal, slow, and fast); *fb* is the bag-carrying probe set; the speed type of the gallery set is *fn* (i.e., normal).

P G	2km/h	3km/h	4km/h	5km/h	6km/h	7km/h
2km/h	**100**	**100**	88	80	80	84
3km/h	**100**	**100**	**100**	88	84	80
4km/h	88	96	**100**	92	92	84
5km/h	**96**	96	96	96	**100**	**96**
6km/h	84	84	96	96	**100**	**100**
7km/h	84	88	84	96	**100**	**100**

Table 3.2: The rank-1 CCR(%) distribution of DCM [Kusakunniran et al., 2012a] in the cross-speed walking gait recognition. G/P denotes the speeds of the gallery/probe.

P G	2km/h	3km/h	4km/h	5km/h	6km/h	7km/h
2km/h	**100**	96	96	96	**100**	84
3km/h	**100**	**100**	96	**100**	**100**	84
4km/h	**100**	**100**	**100**	**100**	**100**	84
5km/h	92	96	**100**	**100**	**100**	**96**
6km/h	**92**	92	96	**100**	**100**	**100**
7km/h	88	84	72	92	**100**	**100**

Table 3.3: The rank-1 CCR(%) distribution of the method 2DPCA+2DLDA in the cross-speed walking gait recognition. G/P denotes the speeds of the gallery/probe.

P G	2km/h	3km/h	4km/h	5km/h	6km/h	7km/h
2km/h	**100**±0.00	**100**±0.00	**100**±0.00	**97.6**±2.07	97.6±2.80	**94**±2.83
3km/h	**100**±0.00	**100**±0.00	**100**±0.00	**100**±0.00	**100**±0.00	**98.4**±2.07
4km/h	**100**±0.00	**100**±0.00	**100**±0.00	**100**±0.00	**100**±0.00	**90.4**±2.80
5km/h	92.8±1.69	**96.4**±1.26	**100**±0.00	**100**±0.00	**100**±0.00	**96**±0.00
6km/h	**92**±0.00	**94.4**±2.07	**100**±0.00	**100**±0.00	**100**±0.00	**100**±0.00
7km/h	**92**±0.00	**94**±2.11	**94.8**±1.93	**100**±0.00	**100**±0.00	**100**±0.00

Table 3.4: The rank-1 CCR(%) distribution of RSM (our method) in the cross-speed walking gait recognition. G/P denotes the speeds of the gallery/probe.

3.3.2 Cross-speed Walker/Runner Identification in the Fixed-mode on the OU-ISIR-A Dataset

Compared with the CASIA-C dataset, the OU-ISIR-A dataset contains less subjects (34) but broader range of walking/running speeds. There are six different walking speeds from 2km/h to 7km/h with an interval of 1km/h, and three different running speeds from 8km/h to 10km/h with an interval of 1km/h.

On the OU-ISIR-A dataset, we compare RSM with DCM [Kusakunniran et al., 2012a] and the method 2DPCA+2DLDA in all the cross-speed walking gait recognition tasks. Note DCM [Kusakunniran et al., 2012a] used 9 subjects (covering all the possible 6 walking speeds in this dataset) for training, while RSM and the method 2DPCA+2DLDA, only need gallery with a specific speed for training. Based on the rest 25 subjects, the results of the three methods are reported in Table 3.2-3.4. Out of the total 36 tasks including different levels of walking speed changes, the performance of our method is generally better than the other two, especially when speed changes are large (e.g., between 2km/h and 7km/h).

Based on the 25 subjects, we also apply our method and 2DPCA+2DLDA to the cross-speed running gait recognition tasks. To the best of our knowledge, the effect of the running speed has not been studied in other existing works. The recognition accuracies for these two methods are illustrated in Table 3.5-3.6. Both methods can achieve high performance, since in this dataset the running speed range is 8km/h-10km/h and the corresponding

P G	8km/h	9km/h	10km/h
8km/h	**100**	**96**	96
9km/h	96	**100**	**100**
10km/h	96	**100**	**100**

Table 3.5: The rank-1 CCR(%) distribution of the method 2DPCA+2DLDA in the cross-speed running gait recognition. G/P denotes the speeds of the gallery/probe.

P G	8km/h	9km/h	10km/h
8km/h	**100**±0.00	**96**±0.00	**100**±0.00
9km/h	**97.2**±1.93	**100**±0.00	**100**±0.00
10km/h	**99.2**±1.69	**100**±0.00	**100**±0.00

Table 3.6: The rank-1 CCR(%) distribution of RSM (our method) in the cross-speed running gait recognition. G/P denotes the speeds of the gallery/probe.

intra-class variations are small (see Fig. 3.1). From the experimental results on CASIA-C and fixed-mode OU-ISIR-A datasets, we can see that when the intra-class variations are relatively small, although method 2DPCA+2DLDA is generally less effective than RSM, it can still yield competitive performance.

Table 3.7 lists the average rank-1 CCRs of our method in the fixed-mode gait recognition tasks, i.e., 98.07% average rank-1 CCR (corresponding to Table 3.4) for walker identification, and 99.16% average CCR (corresponding to Table 3.6) for runner identification. The nearly perfect performance suggests the effectiveness of the RSM framework in speed-invariant gait recognition. It is also worth noting that the average identification rate of runner is slightly higher than walker, and this observation is consistent with the claims in [Yam et al., 2002a],[Yam et al., 2004],[Iosifidis et al., 2012]. In the context of fixed-mode gait recognition, the explanations of this phenomenon are two-fold: 1) the running subjects may have smaller inter-class similarities than the walking subjects [Yam et al., 2002a][Yam et al., 2004]; 2) the running subjects may also have smaller intra-class variations which contribute positively to recognition, since normally in real-world scenarios (at least in this dataset), running tends to have smaller speed range (e.g., 8km/h-10km/h) than walking (e.g., 2km/h-7km/h). Nevertheless, for the fixed-mode gait recognition, our method can achieve

-	walking	running
average	98.07	99.16

Table 3.7: The general average rank-1 CCR(%) of RSM (our method) in the cross-speed walking (Table 3.4) and running (Table 3.6) gait recognition.

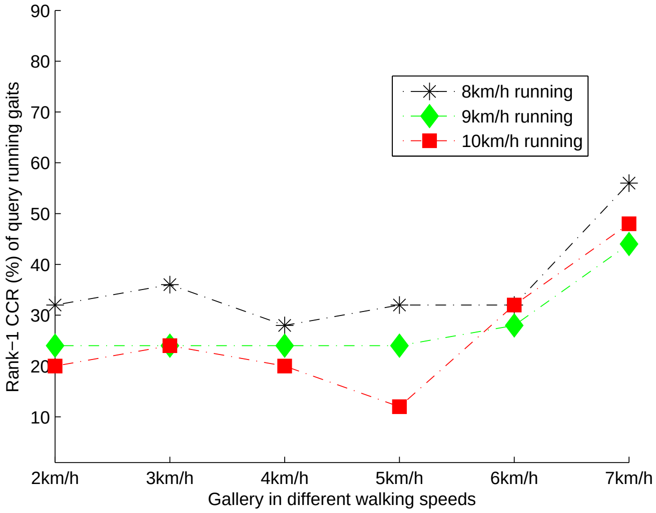


Figure 3.3: The rank-1 CCR (%) distribution of the method 2DPCA+2DLDA in cross-mode gait recognition, i.e., to identify unknown runners given the gallery of walkers.

nearly perfect performance.

3.3.3 Cross-mode Runner Identification on the OU-ISIR-A Dataset

In real-world scenarios, cross-mode gait recognition is often required, e.g., identifying a runner when only the walking gallery is available. Although Yam et al. [Yam et al., 2002a] claimed that cross-mode gait recognition is not feasible due to the lack of generic mapping between walking and running subjects/gaits, encouraging identification rates are still achieved by our method on the 25 subjects in the OU-ISIR-A dataset, as shown in Fig. 3.4. To evaluate the effectiveness of RSM, we also conduct the cross-mode gait recognition experiments using method 2DPCA+2DLDA and the corresponding performance are reported in Fig. 3.3. From it we can see that the general performance of method 2DPCA+2DLDA

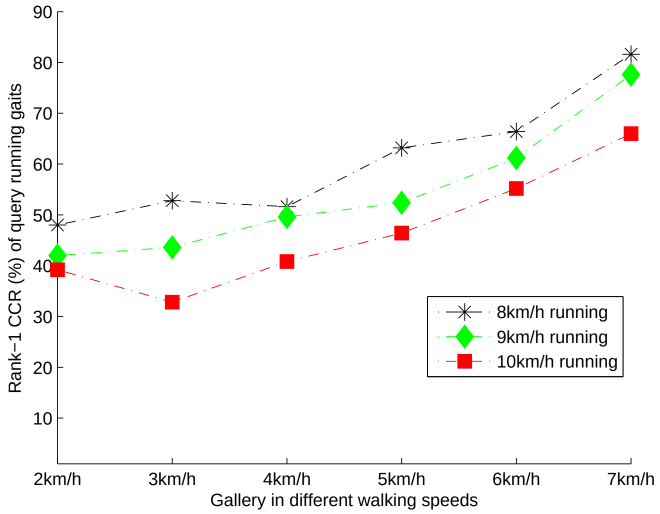


Figure 3.4: The rank-1 CCR (%) distribution of RSM (our method) in cross-mode gait recognition, i.e., to identify unknown runners given the gallery of walkers.

is significantly worse than RSM. The experimental results in Chapter 3.3.1 suggest when the intra-class variations are small, method 2DPCA+2DLDA can yield reasonable performance, yet its performance drops significantly when the intra-class variations are large. In this case, the walking gallery (used for training) becomes less representative when the query gait is from a different mode (i.e., running). To prevent overfitting the less representative training data from happening, the RSM framework combines a large number of weak classifiers and it achieves encouraging performance in these challenging cross-mode gait recognition tasks.

Based on the RSM framework, matching query running gaits of different speeds to the fast walking gallery sets (e.g., in 7km/h) tends to deliver higher performance. Generally, for cross-mode gait recognition, the identification rate is higher when the speed difference is smaller, since there may be more effective features in the common space between (faster) walking and (slower) running. For example, for runner identification at 8km/h, the rank-1 CCR based on a 7km/h walking gallery is 81.6%, and it is significantly higher than using a 3km/h walking gallery with a rank-1 CCR of 52.8%, as shown in Fig. 3.4. For real-world applications, the results in Fig.3.4 suggest that when the running gallery is unavailable, a faster walking gallery is more suitable for runner identification.

3.3.4 Discussion

For the fixed-mode gait recognition, our RSM framework copes very well with speed variation. However, for the more challenging cross-mode scenario, when the speed differences (between the running probe and walking gallery) are large, the performance is unsatisfactory. In this case, the intra-class variations become extremely large because we deem running and walking as the same modality. It is an open question to achieve satisfactory performance with the existence of extremely large intra-class variations. Nevertheless, in real-world applications, it is possible to reduce such speed differences by using the faster walking gait gallery, e.g., with a walking speed of 7km/h, which has nearly perfect performance in walker identification and competitive performance in runner identification.

3.4 Summary

In this chapter, we present a classifier ensemble method based on RSM concept to solve the cross-speed gait recognition problems in both fixed-mode and cross-mode. For fixed-mode gait recognition, compared with the previous cross-speed walker identification algorithms, our method delivers a significant improvement. Our method also achieves nearly perfect accuracies in the cross-speed runner identification. Different from the fixed-mode, the cross-mode gait recognition is challenging due to the significant differences between walking and running. We study the cross-speed runner identification using solely the walking gallery. The experimental results suggest that a faster walking gallery (e.g., 7km/h) is suitable for cross-speed human identification in unknown mode (i.e., running mode or walking mode). More details of this chapter can be found in one of our publications in [Guan and Li, 2013].

Chapter 4

Tackling the Elapsed Time Covariate using Multimodal-RSM

In Chapter 3, we demonstrated that RSM is an effective framework for gait recognition in tackling several simple covariates like speed, and carrying condition. Through combining a large number of weak classifiers, the generalisation errors can be greatly reduced. However, an unimodal biometric system's capability is limited when facing extremely large intra-class variations. One of the major challenges is that the human walking style may change over time. To tackle this challenge, in this chapter we propose a multimodal-RSM framework with faces taken into account in order to reinforce the weak classifiers without compromising the generalisation ability of the system as a whole. The evaluations of our method on the TUM-GAID dataset suggest the effectiveness of multimodal-RSM for tackling the most challenging elapsed time covariate, which also includes the changes in shoe, carrying status, clothing, lighting condition, etc.

4.1 Problem Statement and Motivation

Gait recognition with the elapsed time covariate is one of the most challenging tasks because subjects' walking styles change over time. In Fig. 4.1, we illustrate several GEI samples

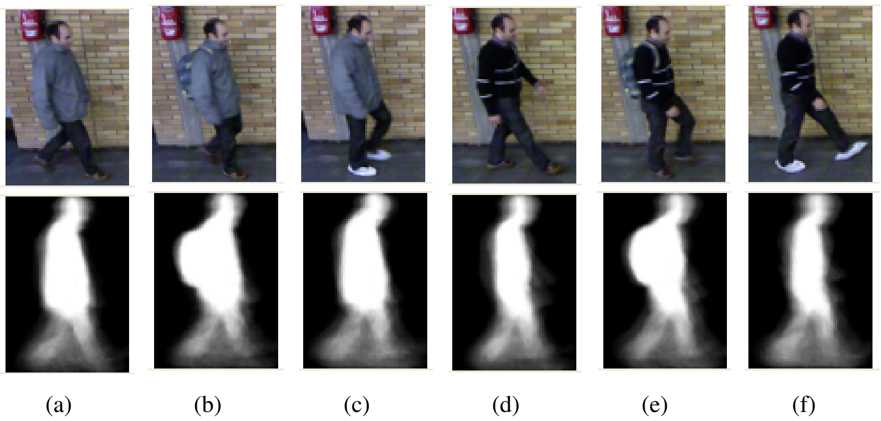


Figure 4.1: Gait images from the TUM-GAID dataset for a subject in 6 different conditions. (a): normal; (b): backpack; (c): coating shoes; (d): elapsed time + normal; (e): elapsed time + backpack; (f): elapsed time + coating shoes. Top row includes the gait RGB images while the bottom row includes the corresponding GEIs.

from the TUM-GAID database [Hofmann et al., 2014]. Compared with normal condition (e.g.,in Fig. 4.1(a)), the human appearance may change significantly as time elapses (e.g., in Fig. 4.1(d), Fig. 4.1(e), Fig. 4.1(f)). Moreover, as time elapses, the same subject's clothing, carrying condition, weight, fatigue, etc. are likely to change. It is an open question for unimodal biometric systems to handle extremely large intra-class variations, and in this case, building multimodal systems could be an effective way to enhance the performance [Jain et al., 2004b]. In this chapter, we aim to extend the previously proposed RSM to a multimodal-RSM framework so as to fuse gait and face information. We use multi-class kernel Fisher analysis (KFA)[Liu, 2006] for face feature extraction. The corresponding face score is then used to strengthen the gait-based weak classifiers before the majority voting.

Compared with other covariate factors, only a few works systematically studied the effect of elapsed time. In [Matovski et al., 2012], Matovski et al. investigated the effect of elapsed time (up to nine months) based on 25 subjects by fusing the gait information from three cameras in different views. The results suggest that: 1) irrespective of other covariates, short term elapsed time does not affect the recognition significantly; 2) the accuracies may drop rapidly when other covariates (e.g., clothing) are included. Since, in real-world scenarios, it is unrealistic to have other covariates perfectly controlled, our objective in this chapter is to tackle the elapsed time challenge in a less constrained environment. Besides, the modalities we are fusing (i.e., gait and face) can be easily collected using one camera, in contrast to three in [Matovski et al., 2012]. In RSM systems, weak classifiers with lower dimensions tend to have better generalisation ability [Ho, 1998]. However, they may encounter an underfitting problem if the dimensionality is too low. It is desirable to use information from other sources (e.g., other biometric modalities) to strengthen the weak classifiers. Although face at a distance may be less reliable, it may provide some complementary information for gait. In [Jain et al., 2004a], Jain et al. demonstrated that the error rate of a fingerprint recognition system can be further reduced by integrating soft biometric information. Similarly, in this chapter we treat the less reliable face as a ”soft” biometric trait by assigning lower weight to its score. After summing up the weighted face score and each gait score (corresponding to each weak classifier), the final result is obtained by majority voting among the output labels of the updated classifiers. It is also worth mentioning that by assigning lower weight to the face score, it is less likely to smooth the diversity (among the updated weak classifiers), which is important for multiple classifier systems [Kuncheva and Whitaker, 2003]. A flowchart of the multimodal-RSM framework is shown in Fig. 4.2.

4.2 Gait Recognition

4.2.1 Gait Feature Templates

Gabor-filtered GEI has been demonstrated to be an effective feature template for gait recognition [Tao et al., 2007],[Xu et al., 2012]. Given a GEI sample, Gabor functions of five scales and eight orientations are employed to generate the Gabor template, as introduced in Chapter 2.4.1. For the gait recognition system, we use the basic RSM model introduced in Chapter 3, which can extract the random features in the *column* direction and perform local enhancing (LE) in the *row* direction of the silhouettes. We also attempt to employ the RSM basic model in a different way, i.e., extracting the random features in the *row* direction and performing LE in the *column* direction of the silhouettes. One simple way to implement this

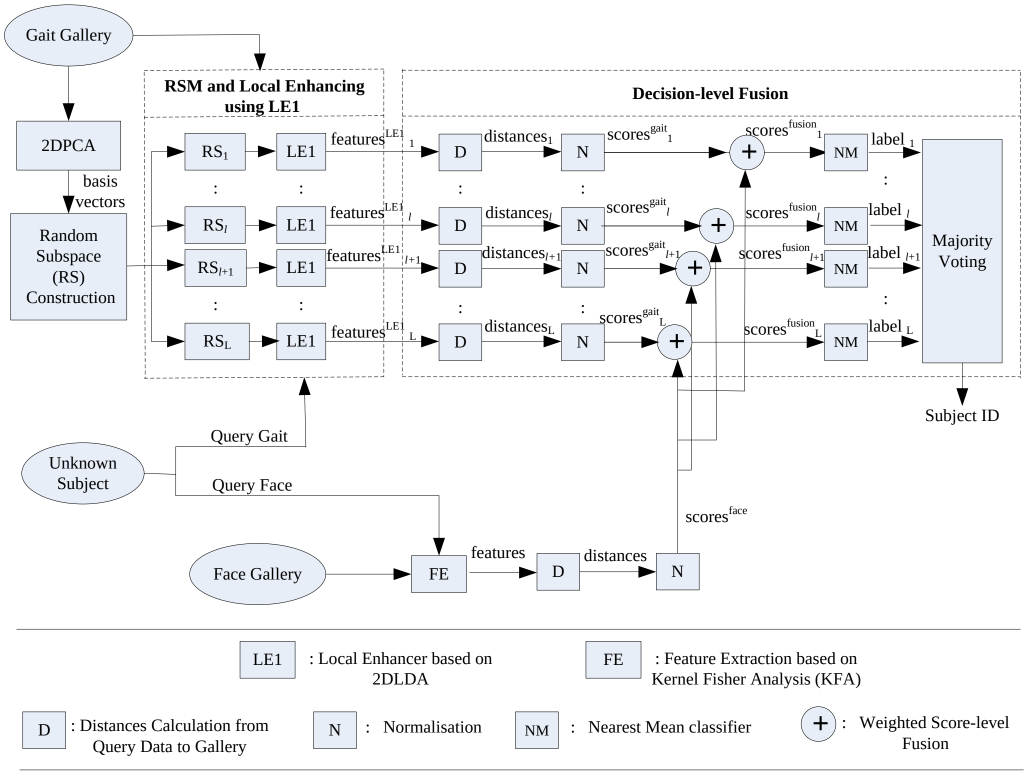


Figure 4.2: The multimodal-RSM framework

is to transpose the gait silhouettes in the first place, and thus we introduce another Gabor-based feature template in this chapter, i.e., transposed gait template. After transposing the optimised-GEI defined in [Kusakunniran et al., 2009b], we use the corresponding Gabor-filtered features as the second gait template. For computational efficiency, similar to [Xu et al., 2012], we use the downsampled version of both templates. In this chapter, these two Gabor feature templates are referred to as *Gait* and *T-Gait*, and the process of generating both templates is shown in Fig. 4.3.

4.2.2 RSM for Gait Recognition

RSM is used for gait feature extraction, and the corresponding training process is summarised in Algorithm 3.1 and Algorithm 3.2, which can be found in Chapter 3.2.1 and Chapter 3.2.2. Assuming for the l^{th} subspace, we get $\mathbf{R}^l$ and $\hat{\mathbf{W}}^l$ from Algorithm 3.1 and

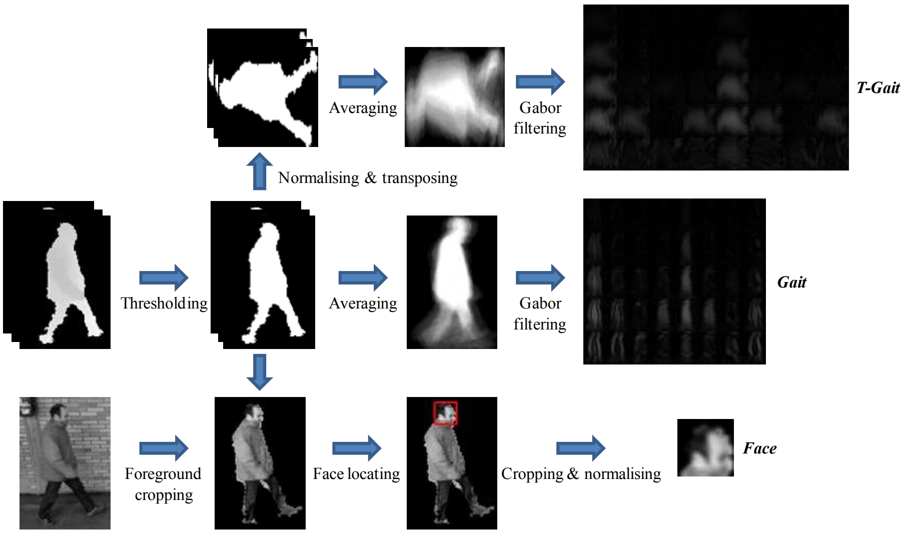


Figure 4.3: The process of generating the 3 feature templates, i.e., *T-Gait, Gait, Face*

Algorithm 3.2, then feature extraction can be performed on a gait image $\mathbf{I}$ by using

$$\mathbf{X}^l = (\hat{\mathbf{W}}^l)^T(\mathbf{I}\mathbf{R}^l), \quad l \in [1, L]. \tag{4.1}$$

After concatenating the feature matrices (the output of Eq.(4.1)) into the corresponding vectors, the dissimilarity between different samples can be calculated based on Euclidean distance, as introduced in Chapter 3.2.3. For the l^{th} subspace, let $[\mathbf{u}_1^l, \mathbf{u}_2^l, ..., \mathbf{u}_c^l]$ be the centroids of the c subjects in the gallery. For a query gait sequence $\mathbf{P}^l$ with n_p samples $[\mathbf{p}_1^l, \mathbf{p}_2^l, ..., \mathbf{p}_{n_p}^l]$, the distance between $\mathbf{P}^l$ and the j^{th} class centroid $\mathbf{u}_j^l$ is defined as:

$$\delta(\mathbf{P}^l, \mathbf{u}_j^l) = \frac{1}{n_p}\sum_{i=1}^{n_p} \|\mathbf{p}_i^l - \mathbf{u}_j^l\|, \quad j \in [1, c]. \tag{4.2}$$

This dissimilarity for the l^{th} subspace can be normalised as the gait score, before being combined with the face information.

4.3 Face Recognition[2]

4.3.1 Face Cropping

The process of face cropping is demonstrated in Fig. 4.3: first we use the binarised depth mask on the corresponding grayscale image to get the whole human body from the background. Then, a line-by-line scanning is performed to locate two landmarks (i.e., the top-most pixel and the right-most pixel of the upper body area). A pre-defined face area is then cropped based on the two landmarks. Finally the cropped faces are aligned by the landmarks and normalised to 18×18 pixels. They are referred to as *Face* templates in this chapter.

4.3.2 Face Identification

For face feature extraction, multi-class kernel Fisher analysis (KFA) [Liu, 2006] is used. KFA first performs nonlinear mapping from the input space to a high dimensional feature space. Then LDA can be employed in the new feature space.

Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n] \in \mathbb{R}^{m \times n}$ be the matrix of n training samples (i.e., gallery) in the input space, and $\mathbf{x}_i$ denotes the concatenated vector from the i^{th} *Face* template. Assume there are c classes and $n_1, n_2, ..., n_c$ are the number of training samples for each class and $\sum_{i=1}^{c} n_i = n$. Let $f : \mathbb{R}^m \to F$ be a nonlinear mapping from the input space to the feature space. Then the matrix in the feature space can be represented as: $\mathbf{Y} = [f(\mathbf{x}_1), f(\mathbf{x}_2), ..., f(\mathbf{x}_n)]$. Generally, we expect different classes to be well separated while samples within the same class to be tightly related. This leads to optimising $J_1 = \mathrm{trace}(\mathbf{S}_m^{-1}\mathbf{S}_b)$ where $\mathbf{S}_b$ is the between-class scatter matrix while $\mathbf{S}_m$ is the mixture scatter matrix in the feature space.

However, it is difficult to evaluate $\mathbf{S}_m$ and $\mathbf{S}_b$ in the high dimensional feature space. In KFA, a kernel matrix $\mathbf{K}$ is defined as: $\mathbf{K} = \mathbf{Y}^T\mathbf{Y}$ where $\mathbf{K}_{ij} = (f(\mathbf{x}_i) \cdot f(\mathbf{x}_j)), i, j = 1, 2, ..., n$. So optimising J_1 involves solving the following eigenvalue problem by replacing

[2] This work presented in this section was carried out by my colleague Xingjie Wei (x.wei@warwick.ac.uk), whose contribution is highly appreciated.

-	Gallery	Probe					
Walking Condition	N	N	B	S	TN	TB	TS
#Seq.	155×4	155×2	155×2	155×2	16×2	16×2	16×2
#*Gait/T-Gait* per Seq.	1	1	1	1	1	1	1
#*Face* per Seq.	4	1	1	1	1	1	1

Table 4.1: Experimental settings on the TUM-GAID dataset [Hofmann et al., 2014]. Abbreviation note : N - Normal, B - Backpack, S - Shoe, TN - Time+Normal, TB - Time+Backpack, TS - Time+Shoe.

$\mathbf{S}_m$ and $\mathbf{S}_b$ with the kernel matrix $\mathbf{K}$:

$$\mathbf{KZK}\alpha = \lambda\mathbf{KK}\alpha, \tag{4.3}$$

where α (resp. λ) denotes the eigenvector (resp. eigenvalue). Here $\mathbf{Z} \in \mathbb{R}^{n\times n}$ is a block diagonal matrix: $\mathbf{Z} = diag\{\mathbf{Z}_1, \mathbf{Z}_2, ..., \mathbf{Z}_c\}$ where $\mathbf{Z}_j$ is a $n_j \times n_j$ matrix with elements all equal to $\frac{1}{n_j}, j = 1, 2, ..., c$. Let $\mathbf{A} = [\alpha_1, \alpha_2, ..., \alpha_r] \in \mathbb{R}^{n\times r}$ be the eigenvectors corresponding to the $r(r \leqslant c - 1)$ largest eigenvalues. For a face vector $\mathbf{x}$, the KFA features can be extracted by using:

$$\mathbf{F} = \mathbf{A}^T\mathbf{B} \tag{4.4}$$

where $\mathbf{B} = [f(\mathbf{x}_1) \cdot f(\mathbf{x}) \quad f(\mathbf{x}_2) \cdot f(\mathbf{x}) \quad ... \quad f(\mathbf{x}_n) \cdot f(\mathbf{x})]^T$. Actually, the kernel matrix $\mathbf{K}$ can be computed with a kernel function, instead of explicitly performing the nonlinear mapping. Here we use the fractional power polynomial kernel function[Liu, 2006] as:

$$k(\mathbf{a}, \mathbf{b}) = (f(\mathbf{a}) \cdot f(\mathbf{b})) = sign(\mathbf{a} \cdot \mathbf{b})(abs(\mathbf{a} \cdot \mathbf{b}))^\gamma \tag{4.5}$$

where $sign(\cdot)$ is the sign function and $abs(\cdot)$ is the absolute value operator. We empirically set $\gamma = 0.8$. By using KFA, the linear model is able to capture the nonlinear patterns in the original data. After feature extraction, we can get the face score with a nearest mean (NM) classifier.

4.4 Fusion Strategy

In the context of the RSM framework, we update the voters/base classifiers by fusing the face score and each gait score out of the L subspaces, before the majority voting. Score normalisation is required before the fusion, and there are several popular techniques such as min-max, z-score, tanh normalisation, etc. [Jain et al., 2005] and here we only employ the simple min-max rule. In the l^{th} subspace, given a query gait $\mathbf{P}^l$ and the gallery with c classes $\{\mathbf{u}_j^l\}_{j=1}^c$, through Eq.(4.2) we can get the corresponding distance vector $\delta(\mathbf{P}^l, \mathbf{u}_j^l), j = 1, 2, ..., c$. For simplicity, we write $\mathbf{d}^l = \delta(\mathbf{P}^l, \mathbf{u}_j^l), j = 1, 2, ..., c$. Given that, the normalised gait score $\mathbf{S}_{gait}^l$ can be defined as:

$$\mathbf{S}_{gait}^l = \frac{\mathbf{d}^l - \min(\mathbf{d}^l)}{\max(\mathbf{d}^l) - \min(\mathbf{d}^l)}, \quad l \in [1, L]. \tag{4.6}$$

The face score $\mathbf{S}_{face}$ can also be obtained through min-max normalisation. Then the classifiers are updated using weighted sum rule. Specifically, the updated score $\mathbf{S}_{face+gait}^l$ corresponding to the l^{th} classifier is defined as:

$$\mathbf{S}_{face+gait}^l = \omega \mathbf{S}_{face} + (1 - \omega)\mathbf{S}_{gait}^l. \tag{4.7}$$

where $\omega \in [0, 1]$ is the weight for the face information. In [Jain et al., 2004a], when the less reliable soft biometric traits (gender, ethnicity, and height) with low weights assigned were fused in a fingerprint recognition system, significant performance gain was achieved. In the context of human identification at a distance, face is also less reliable due to low resolution or the presence of other covariates. Similar to [Jain et al., 2004a], we use face as the ancillary information with a low weight for its score. After the score is updated for each voter, majority voting is applied for the final classification decision.

4.5 Experiments

We conduct experiments on the newly released TUM-GAID dataset [Hofmann et al., 2014]. This dataset simultaneously contains RGB video, depth and audio with 305 subjects in total. In [Hofmann et al., 2014], Hofmann *et al.* designed an experimental protocol (based on 155 subjects) to evaluate the robustness of algorithms against covariate factors.

In the TUM-GAID dataset, the depth images in the tracked bounding box are provided and we can get the corresponding binary silhouettes/masks by thresholding and aligning. Then we can acquire the *Face* templates using the method introduced in Chapter 4.3.1. The process of getting the 3 feature templates used in this chapter is illustrated in Fig. 4.3. There is one gait template (i.e.,*Gait/T-Gait*) corresponding to a gait sequence. For the face, since there are small view changes in a sequence, in the gallery we select 4 *Face* templates from each sequence to capture more intra-class variations. For the probe, only one *Face* template is used from each sequence. The experimental settings of the gallery and probe sets are shown in Table 4.1.

There are three parameters in the RSM framework, namely, number of the random subspaces/base classifiers(L), the dimensionality of the random subspace (N), and number of LE1 projection directions (M). It was verified that in the RSM framework the performance does not decrease with the increasing number of classifiers[Ho, 1998], and we empirically set $L = 1000$. In the RSM framework when the subspace dimension N is too large it encounters the overfitting problem and its performance converges to the one of traditional 2DPCA+2DLDA. Classifiers with small value of N can usually generalise well [Ho, 1998], but underfitting may occur when N is too small. Our aim is to strengthen the weak classifiers by fusing the additional face information (to avoid underfitting). In this case, it does not sacrifice the generalisation ability of the whole system by using a small value of N. We set $N = 2$ in this chapter. We also empirically set $M = 40$. A full discussion of the sensitivity of these three parameters are provided in Chapter 6.3.2.

To evaluate the performance of the algorithms, we use the rank-1/rank-5 correct

Experiment	N	B	S	TN	TB	TS
#Seq.	310	310	310	32	32	32
Rank-1 CCRs						
Face + KFA	89	72	71	44	**38**	44
Gait + RSM	**100**	**79**	**97**	57	**38**	**57**
T-Gait + RSM	99	62	92	**60**	28	56

Table 4.2: The rank-1 CCRs (%) by using single modality on the 6 probes from the TUM-GAID dataset.

classification rate (CCR). Rank-1 (resp. rank-5) CCR shows rate that the correct subject is ranked as the top 1 candidate (resp. top 5 candidates). Due to the random nature, the results of different runs may vary to some extent. We repeat all the experiments (over the 6 probes, i.e., N, B, S, TN, TB, and TS) 10 times and the overall rank-1 performance statistics (mean, standard deviation, maxima and minima) of the proposed multimodal-RSM system are reported in Table 4.3. For the rest of the chapter, we only report the mean values.

4.5.1 Identification using Single Modality

Experiments based on 3 single modalities (i.e., *Face* only, *Gait* only, and *T-Gait* only) are conducted, respectively. The rank-1 CCRs corresponding to the 6 probe sets are illustrated in Table 4.2. Generally, based on a certain modality, reasonable results can be achieved on probe sets N, B, and S. However, when elapsed time is taken into account (i.e., on probe sets TN, TB, and TS), the rank-1 CCRs decrease significantly. It was experimentally verified in Chapter 3.3.1 that carrying condition has little impact on RSM-based gait recognition algorithms. However, when the object carried is heavy (e.g., 5kg backpack in TUM-GAID dataset[Hofmann et al., 2014]), it may change the whole walking style to some extent and thus affects the performance. Similarly, subjects' walking styles may change over time in an unpredictable manner due to potential changes in carrying status, shoe, clothing, emotion, fatigue, etc. The coupled effect may have significant impact on the recognition accuracies when only the gait trait is used. Although facial features may be affected by lighting conditions and low resolution, intuitively it is less sensitive to the heavy object carried, shoe, clothing, etc. Therefore faces may provide some additional information to enhance the performance of the gait recognition system. We will show how we can capitalise on this and leverage the benefit of fusing multiple modalities in the next section.

4.5.2 Tackling the Elapsed Time Covariate using Multimodal-RSM

In this section, by using face as ancillary information, we apply the proposed multimodal-RSM to tackle the challenging elapsed time covariate. Over probes TN, TB, and TS, the rank-1 CCR distributions with respect to face score weight are reported in Fig.4.5(a)-4.5(c). From these figures, we can observe:

1. Higher performance can be achieved when the weight of the face score is relatively low (*e.g.*, $0 < \omega \leq 0.2$). For example, by fusing *T-Gait* and *Face* with $\omega = 0.2$, the performance gains over gait-based RSM (i.e., with $\omega = 0$) are upto $15\%, 32\%$, and 15% for probes TN, TB, and TS, respectively.
2. Although generally *T-Gait* has lower performance than *Gait* (i.e.,with $\omega = 0$), its performance gain is more significant when fusing face information. For example, with $\omega = 0.2$, in terms of rank-1 accuracies, *T-Gait* + *Face* is 6% and 10% higher than *Gait* + *Face* for probes TN, and TS, respectively.

Generally, the performance is very competitive when the weight is within a certain range of small values (e.g., $0 < \omega \leq 0.2$). It is not suitable to set ω too high or too low. There are two extreme cases according to Eq (4.7): when $\omega = 1$, the performance is equivalent to that of face recognition systems (i.e., *Face* + KFA), and when $\omega = 0$, it becomes a gait-based RSM system. For a general multimodal-RSM system, given the fact that it is difficult to collect representative validation data (which covers all the possible covariates) for parameter tuning, it remains an open question to find the optimal ω. Nevertheless, experimental results suggest that very significant performance gains can be achieved by assigning face score a relatively low weight. In this case, the weak classifiers are strengthened without sacrificing the diversity of the whole multiple classifier system. We also test this multimodal-RSM scheme on probes N, B, and S, and the performance distributions shown in Fig. 4.6 also suggest that a lower weight of face information is more suitable for perfor-

-	Mean	Std	Max	Min
Multimodal-RSM (*Gait* + *Face*)	94.73	0.41	95.32	94.15
Multimodal-RSM (*T-Gait* + *Face*)	94.76	0.56	95.61	93.76

Table 4.3: The rank-1 CCR statistics (%) over 10 runs of our multimodal-RSM system

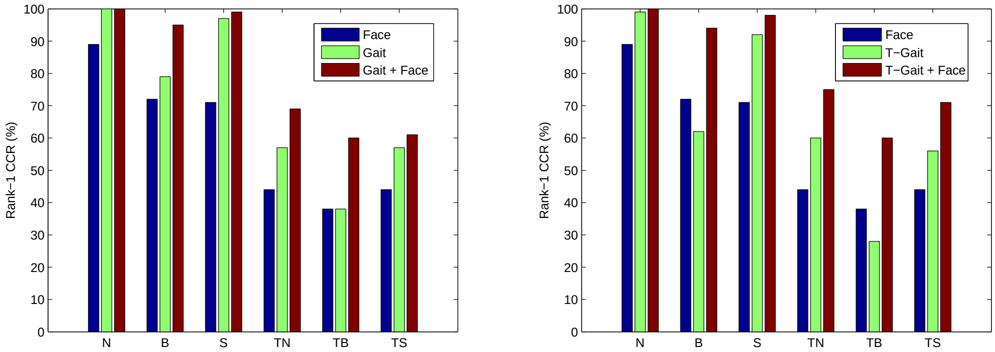


Figure 4.4: Unimodal vs. Multimodal.

mance enhancement. For the rest of this chapter, we only report the results corresponding to $\omega = 0.2$.

4.5.3 Algorithms Comparison

Over the 6 probe sets, we compare our multimodal-RSM (i.e., *Gait* + *Face*, and *T-Gait* + *Face*) with the 3 unimodal-based methods (from Chapter 4.5.1), as shown in Fig.4.4. Experimental results have clearly indicated the effectiveness of our fusion method in tackling the most challenging elapsed time covariate. It also has superb performance when the subject carries 5kg backpack (probe B). Moreover, under the multimodal-RSM scheme, the performance gain is more significant for *T-Gait* + *Face*.

We also compare our method with two recently proposed multimodal methods [Hofmann et al., 2012], [Hofmann et al., 2014]. We implement the GEI + Eigenface [Hofmann et al., 2012] and quote the results of Audio + Depth + GEI from [Hofmann et al., 2014]. Table 4.4 illustrates the performance of the three methods in terms of rank-1/rank-5 CCRs, and the performance of our method is generally much higher. Specifically, for tackling

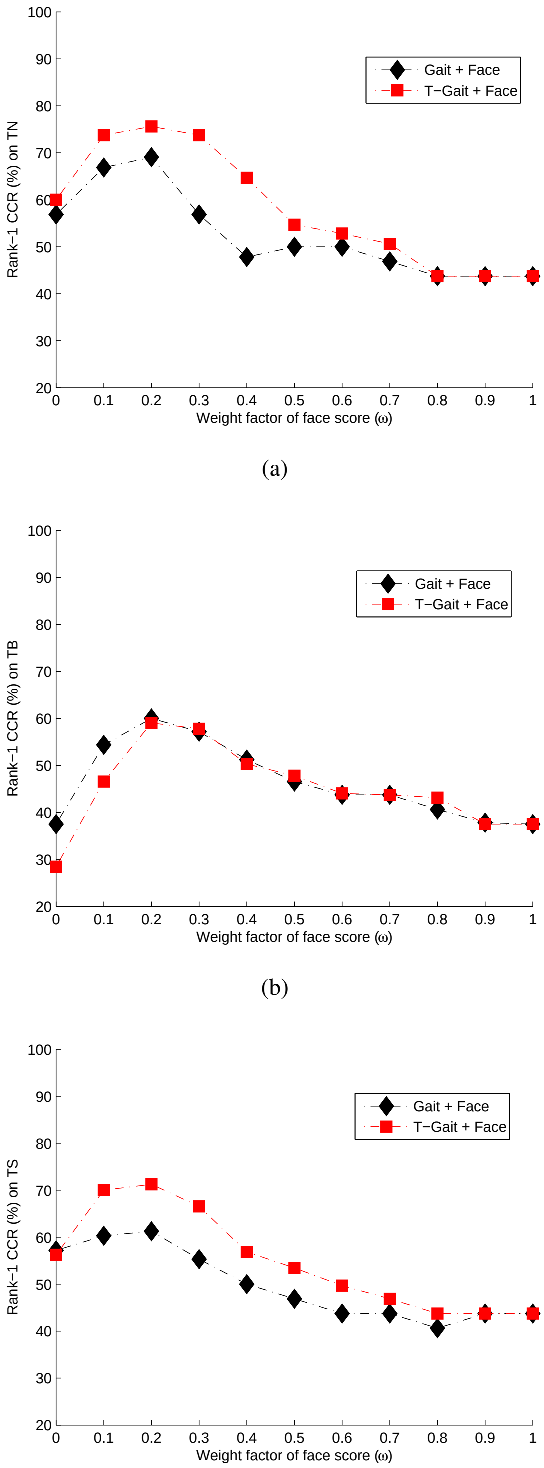
Experiment	N	B	S	TN	TB	TS	Avg.
#Seq.	310	310	310	32	32	32	-
Rank-1 CCRs							
GEI + Eigenface [Hofmann et al., 2012]	97	63	65	47	50	28	71.9
Audio + Depth + GEI[Hofmann et al., 2014]	99	59	95	66	3	50	80.2
Multimodal-RSM (*Gait* + *Face*)	**100**	**95**	**99**	69	**60**	61	**94.8**
Multimodal-RSM (*T-Gait* + *Face*)	**100**	94	98	**75**	**60**	**71**	94.7
Rank-5 CCRs							
GEI + Eigenface [Hofmann et al., 2012]	**100**	84	79	63	63	50	85.0
Audio + Depth + GEI [Hofmann et al., 2014]	**100**	85	99	**81**	28	72	91.5
Multimodal-RSM (*Gait* + *Face*)	**100**	**98**	99	76	71	76	96.7
Multimodal-RSM (*T-Gait* + *Face*)	**100**	**98**	**100**	**81**	**74**	**82**	**97.4**

Table 4.4: Algorithms comparison in terms of rank-1/rank-5 CCRs (%). Avg. denotes the weighted average.

the elapsed time, the method in [Hofmann et al., 2012] has lower performance in probe TS while the method in [Hofmann et al., 2014] (when face information is not fused) only has 3% rank-1 CCR in probe TB. Compared with them, our method consistently has much higher performance in these cases. However, the coupled effect of elapsed time and heavy backpack (probe TB) has larger impact on our system. For example, for *T-Gait* + *Face*, the rank-1 CCR is only 60%, much lower than the ones on probe TN (75%) and probe TS (71%). Nevertheless, compared with the gait-based RSM systems, fusing face as additional information can dramatically reduce the error rates.

4.6 Summary

In this chapter, we extend RSM to multimodal-RSM by allowing other modalities (e.g., face) to be fused. Face provides additional information to strengthen the gait-based weak classifiers in terms of discrimination capability, without compromising the generalisation ability of the whole system. The proposed multimodal-RSM system has much higher performance than the unimodal systems and other multimodal methods. Although additional experiments and theoretical findings are necessary to draw the final conclusions on the benefit of fusing face information, this work empirically demonstrates an effective way on combining multi-modalities information to tackle the most challenging elapsed time covariate, which may also include the changes of clothing, shoe, carrying status, etc. In the future, we will explore how to effectively fuse other information (e.g., age, gender, height, etc.) into this RSM-based system to further improve the performance on challenging problems. More details of this chapter can be found in one of our publications in [Guan et al., 2013b].

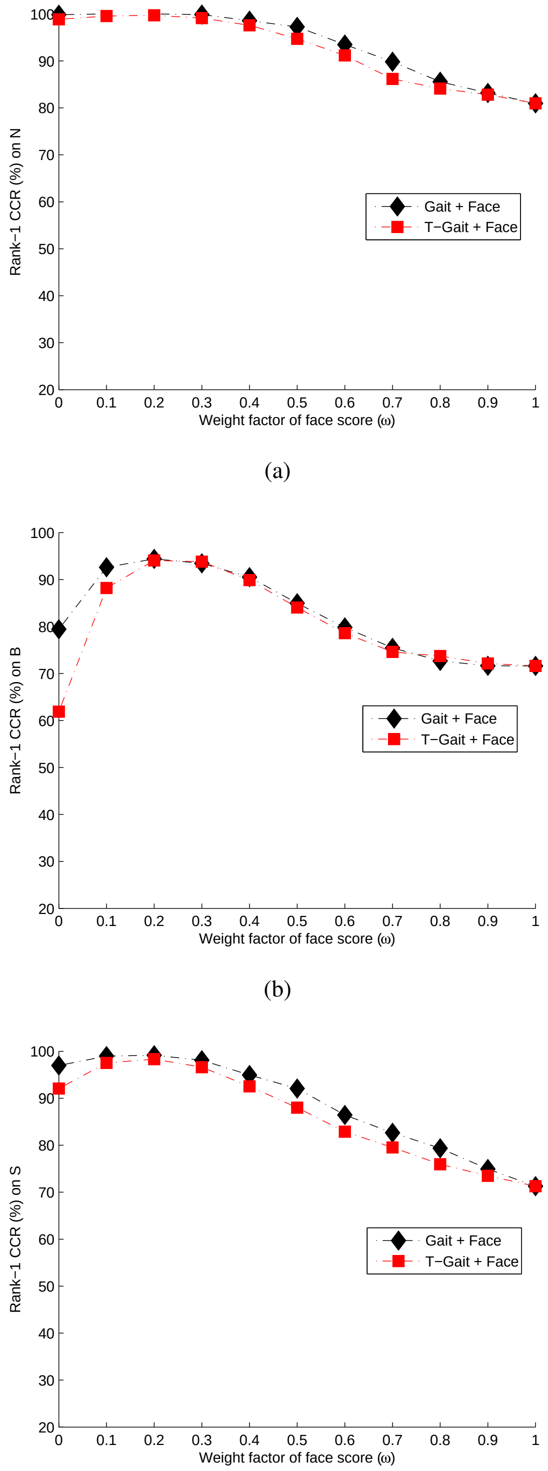


(a)

(b)

(c)

Figure 4.5: Using multimodal-RSM to tackle the elapsed time challenges on probes TN, TB, and TS.



(a)

(b)

(c)

Figure 4.6: Using multimodal-RSM on probes N, B, and S.

Chapter 5

Gait Recognition from Extremely Low Quality Videos through Incorporating Model-based Information into RSM

Nowadays, surveillance cameras are widely installed in public places for security and law enforcement, but the video quality may be low due to limited transmission bandwidth or storage capacity. In this chapter, we apply the concept of multimodal-RSM to gait recognition from low quality videos, which have a frame-rate at 1 fps and resolution of 32×22 pixels. Different from popular temporal reconstruction-based methods, we use the average gait image (AGI) over the whole sequence as the appearance-based feature description. After using RSM-based feature extraction on the AGIs, we incorporate the model-based information to enhance the weak classifiers. We find that the performance improvement is directly proportional to the average disagreement level of weak classifiers (i.e., diversity), which can be increased by using the model-based information. We evaluate our method on both indoor and outdoor databases (i.e., the low quality versions of OU-ISIR-D and USF databases), and the results suggest that our method is more general and effective than other state-of-the-art algorithms.

5.1 Problem Statement and Motivation

Surveillance cameras are widely installed in public places such as airports, government buildings, streets and shopping malls to prevent criminal activities nowadays. However, the video quality may be low due to limited transmission bandwidth or storage capacity. The corresponding gait sequences extracted from such low quality videos may have much lower resolution if the subjects are far away from the cameras, and much fewer ”clean” gait frames available if the subjects are significantly suffered from the occlusions in the crowded areas. The lack of frames may have a similar effect as extremely low frame-rate (e.g., 1 fps). As such, gait recognition algorithms that are robust to low quality gait videos are desirable.

As introduced in Chapter 2.1, gait recognition methods can be roughly divided into two categories: model-based and appearance-based approaches. Model-based methods use human body parameters as features for recognition, while appearance-based methods are more general, and most of them perform classification based on pixel intensities. Appearance-based approaches can work well with relatively low resolution gait videos, in contrast to model-based methods with the less reliable estimated body parameters. In this chapter, we use average gait image (AGI) [Veres et al., 2004] as the appearance-based feature template, which has similar properties to GEI. It is the average image over the whole gait sequence, instead of over a single gait cycle. Compared with GEI, there is no requirement for gait cycle detection, which is difficult to perform given the low frame-rate [Guan et al., 2013a]. Several silhouettes and their AGI samples (derived from extremely low quality videos) from the OU-ISIR-D [Makihara et al., 2012] and USF [Sarkar et al., 2005] databases are shown in Fig. 5.1.

To combat overfitting, RSM-based feature extraction is used on AGIs. Motivated by the multimodal-RSM introduced in Chapter 4, we aim to incorporate other information

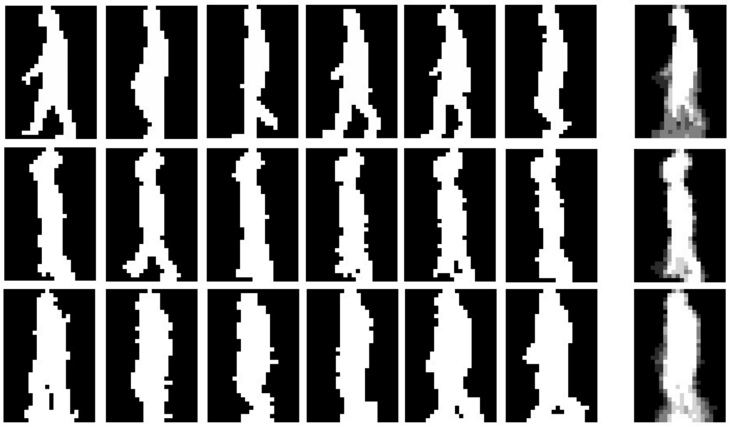


Figure 5.1: Extremely low quality gait sequences (derived from videos with frame-rate at 1 fps and resolution of 32×22 pixels). Top/middle row: from the indoor OU-ISIR-D [Makihara et al., 2012] database with low/large gait fluctuations; bottom row: from the outdoor USF database [Sarkar et al., 2005]. In each row, the rightmost image is the AGI corresponding to the whole gait sequence.

to strengthen the weak classifiers for higher performance. Although face information from surveillance cameras is useful to some extent for multimodal-RSM, when subjects are too far away from the camera with extremely low resolution, face information may become extremely unreliable. As afore-mentioned, for low-resolution videos model-based gait information may be less reliable, yet it may provide some body structure information from a different perspective. In this chapter, we use the model-based information for the enhancement of the appearance-based weak classifiers. Since both types of information can be derived from the same gait video, this is especially useful when the footage quality is extremely low with other modalities/information unavailable.

For classifier ensembles, the diversity of the predictions (i.e., predicted labels by the base classifiers) is important and can be measured by means of different metrics [Kuncheva and Whitaker, 2003]. One of the most popular metrics is the *pair-wise disagreement measure*, which is proportional to the number of *different* outputs (correct or wrong) between any pair of base classifiers in the whole multiple classifier system [Kuncheva and Whitaker, 2003]. This measure was initially proposed by Skalak [Skalak, 1996]. Ho [Ho, 1998] applied it to random decision forests. In this chapter, we use this metric to explore the relationship among diversity, individual classifier accuracy, and ensemble accuracy. From the perspective of individual classifier accuracy and diversity, we explain the performance gain (i.e., enhanced ensemble accuracy) through incorporating model-based information into RSM.

5.2 Incorporating Model-based Information into RSM

Different from the previous two chapters (i.e., Chapters 3-4) which used 2DPCA for random subspace construction, we employ PCA in this chapter. The reason is straight forward: With low resolution images, the trained 2DPCA eigenspace would be spanned by only a small number of basis vectors, whose upper bound is the column number (of one low resolution image). This would limit the number of constructed random subspaces to some extent. On the other hand, PCA is unlikely to face this problem, since for PCA eigenspace, the corresponding basis vector number's upper bound is much larger, i.e., the pixel number of one gait image. Therefore, we can generate a large number of random subspaces/base classifiers for higher performance. As such, we use PCA for such low resolution gait images.

In this section, we first introduce how to use PCA-based RSM for the feature extraction, then we employ two simple model-based methods for extremely low quality videos. Model-based information can be used to enhance the weak classifiers without sacrificing the diversity of the whole multiple classifier system. Finally, the diversity measurement used to evaluate our system is described.

5.2.1 AGI for Low Frame-rate Videos

In this chapter, we use average gait image (AGI) [Veres et al., 2004] as the appearance-based feature description, as introduced in Chapter 2.4.1. In extremely low frame-rate environments (e.g., 1 fps gait videos), the major benefit of using AGI is that the gait period, which is difficult to estimate, is not required. However, when the video recording time is also short, the major problem is the lack of frames (e.g., 5 frames for a sequence), which makes the averaging operation less effective. For example, when the walking starting stances of the probe and reference sequence are different, through averaging operation, although the static parts (e.g., head) can be relatively stable [Guan et al., 2013a], the dynamic parts (e.g., legs or arms) can be rather diverse. Assuming such effect caused by extremely low frame-rate are intra-class variations that the gallery data fails to capture and in this case, the RSM concept can be used to reduce such generalisation errors.

5.2.2 RSM for Low Resolution Videos

PCA is used to decorrelate the feature space, before the random subspace construction. Given c classes/gait sequences in the gallery, there are c AGIs. Let m be the pixel number of an AGI, after concatenating each two-dimensional AGI, the gallery can be represented as $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_c] \in \mathbb{R}^{m \times c}$. Then the covariance matrix $\mathbf{S}$ can be estimated:

$$\mathbf{S} = \frac{1}{c}\sum_{i=1}^{c}(\mathbf{a}_i - \bar{\mathbf{a}})(\mathbf{a}_i - \bar{\mathbf{a}})^T, \tag{5.1}$$

where $\bar{\mathbf{a}} = \frac{1}{c}\sum_{i=1}^{c} \mathbf{a}_i$. The eigenvectors of $\mathbf{S}$ can be computed and the leading v eigenvectors are retained as candidates to span the random subspaces.

L random subspaces can be generated and each projection matrix (e.g., feature extractor) is formed by randomly selecting s eigenvectors (from the v candidates). Let the projection matrix be $\mathbf{R}^l \in \mathbb{R}^{s \times m}, l \in [1, L]$, and each gallery sample $\mathbf{a}_i \in \mathbb{R}^m, i \in [1, c]$ can be projected as L sets of coefficients $\mathbf{u}_i^l \in \mathbb{R}^s, l = 1, 2, ..., L$ as the new gait representations:

$$\mathbf{u}_i^l = \mathbf{R}^l \mathbf{a}_i, \quad l = 1, 2, ..., L, \quad i \in [1, c]. \tag{5.2}$$

During a comparison trial, for the l^{th} subspace first $\mathbf{R}^l$ is used for feature extraction, and then Euclidean distance is adopted to measure the dissimilarity. For example, the distance

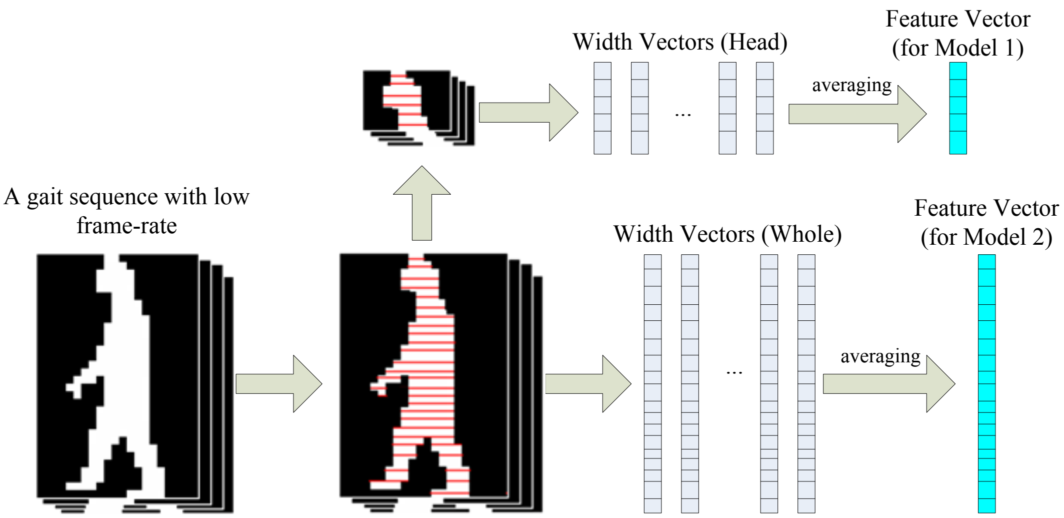

Width Vectors (Head)
Feature Vector (for Model 1)
averaging
...
A gait sequence with low frame-rate
Width Vectors (Whole)
Feature Vector (for Model 2)
averaging
...

Figure 5.2: The process of generating feature vectors for model-based methods.

between a query AGI $\mathbf{x} \in \mathbb{R}^m$ and a certain class $\mathbf{u}_i^l, i \in [1, c]$ is:

$$d(\mathbf{x}, \mathbf{u}_i^l) = \|\mathbf{R}^l\mathbf{x} - \mathbf{u}_i^l\|, \quad i \in [1, c], \quad l \in [1, L], \tag{5.3}$$

which can be updated using the model-based information for further processing.

5.2.3 Model-based Method for Low Quality Videos

From low resolution images (e.g., 32 × 22 pixels, see Fig. 5.1), it is difficult to estimate the body parameters that are used as classical model-based features, e.g., angles of hip and thigh [Cunado et al., 1997], or stride and leg length [Bobick and Johnson, 2001], etc. For gait images with poor segmentation quality, in [Kale et al., 2004b], Kale et al. used the silhouette widths (of each row) as model-based features, which are easy to estimate.

Motivated by [Kale et al., 2004b], we use widths from the binary silhouette as model-based features, under the assumption that low resolution and poor segmentation quality may have similar effect on gait images. We also employ the widths (of each row) from the static silhouette area (i.e., the head area) as model-based features, which are less affected by the low frame-rate [Guan et al., 2013a]. Since each gait sequence may have several frames, the corresponding average width vectors for *the whole silhouette* and *the head area* are used as input feature vectors for model-based classification. In this chapter, the classification models based on head width vector and silhouette width vector are referred to as Model 1 and Model 2, respectively. The process of generating the model-based feature vectors for Model 1 and Model 2 is illustrated in Fig. 5.2.

During a comparison trial, the distance between two sequences can be measured based on the Euclidean distance. Let the dimensionality of the feature vector (for Model 1 or Model 2) be r, given the gallery consisting of c classes/sequences $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_c] \in \mathbb{R}^{r \times c}$, and distance between a query gait $\mathbf{y} \in \mathbb{R}^r$ and a certain class $\mathbf{b}_i, i \in [1, c]$ is:

$$d(\mathbf{y}, \mathbf{b}_i) = \|\mathbf{y} - \mathbf{b}_i\|, \quad i \in [1, c], \tag{5.4}$$

which can be used as model-based information to update RSM-based classifiers.

Since our contribution in this chapter is to study the effect of fusing model-based information into RSM system for gait recognition in low quality videos, for simplicity other covariates (such as carrying condition, view, etc.) are not considered. The simplest direct matching (as shown in Eq.(5.4)) is used to generate the corresponding model-based scores, which is less robust to covariates. In the futher, we will use more effective feature extraction methods to further improved the robustness of the model-based information.

5.2.4 Fusion with Model-based Information

Multimodal fusion is a popular way to enhance the performance by combining two or more biometric modalities, and it has demonstrated its effectiveness in a large number of biometric applications [Jain et al., 2004b]. However, it assumes that multiple modalities are available, which may not hold in less controlled environments. For single modality, one popular form of fusion is to combine the classification scores derived from different feature spaces, and this is especially useful for biometric data with limited information. In [Han and Bhanu, 2006], real GEI and synthetic GEI were generated from the same gait silhouettes, before their classification scores were fused at a score level. Although these two sources may be highly correlated, the enhanced performance suggests that there may exist some complementary power between the two different feature representations [Han and Bhanu, 2006]. Due to the fact that correlated information may be combined to boost the performance, we fuse model-based and appearance-based methods in this chapter. We aim to enhance the performance of RSM, an appearance-based method, by incorporating the model-based information, which is used to update the classifiers of the L subspaces.

Score normalisation is required before the fusion, and there are several popular techniques such as min-max, z-score, tanh normalisation, etc. [Jain et al., 2005] and here we only use the simple min-max rule. In the l^{th} subspace, Given a query gait $\mathbf{x}$ and the gallery with c classes of feature vectors $\{\mathbf{u}_i^l\}_{i=1}^c$, through Eq.(5.3) we can get the corresponding distance vector $d(\mathbf{x}, \mathbf{u}_i^l), i = 1, 2, ..., c$. For simplicity, we write $\mathbf{d}^l = d(\mathbf{x}, \mathbf{u}_i^l), i = 1, 2, ..., c$. Given that, the normalised gait score $\mathbf{d}_{rsm}^l$ can be defined as:

$$\mathbf{d}_{rsm}^l = \frac{\mathbf{d}^l - \min(\mathbf{d}^l)}{\max(\mathbf{d}^l) - \min(\mathbf{d}^l)}, \quad l \in [1, L]. \tag{5.5}$$

Similarly, through (5.4) we can get the distance vector for the model-based methods. Let the distance vector be $\mathbf{d}_{model}$ after min-max normalisation, then the fused distance vector (for the l^{th} subspace) $\mathbf{d}_{fusion}^l$ can be updated using the weighted sum rule:

$$\mathbf{d}_{fusion}^l = \omega \mathbf{d}_{model} + (1 - \omega)\mathbf{d}_{rsm}^l, \quad l \in [1, L], \tag{5.6}$$

where $\omega \in [0, 1]$ is the weight for the model-based information. In the l^{th} subspace, based on the fused distance vector through Eq.(5.6), we assign the class label to the one with the shortest distance. The final classification decision is achieved through majority voting [Kittler et al., 1998] among all the L classifiers. Given a query gait and c classes in the gallery with labels $[W_1, W_2, ..., W_c]$, the optimal class label $\hat{W}_i$ is:

$$\hat{W}_i = \underset{W_i}{\text{argmax}} \sum_{l=1}^{L} \triangle_{W_i}^{l}, \quad i \in [1, c], \tag{5.7}$$

where

$$\triangle_{W_i}^{l} = \begin{cases} 1, & \text{if } \mathbf{d}_{fusion}^{l}(W_i) = \min(\mathbf{d}_{fusion}^{l}), \\ 0, & \text{otherwise}, \end{cases} \quad i \in [1, c]. \tag{5.8}$$

The weight ω is important for the score-level fusion for each subspace. It is not appropriate to set ω too high or too low. According to Eq.(5.6), we can see that our method becomes the conventional RSM system when $\omega = 0$. On the other hand, when $\omega = 1$, it becomes the model-based method. Intuitively, ω should be a small value, given the fact that model-based features in such low quality videos are less reliable. A detailed evaluation of ω is provided in Chapter 5.3.3.

5.2.5 Diversity Measurement

Diversity among the classifiers is deemed to be a key issue in classifier ensemble [Kuncheva and Whitaker, 2003]. Yet the relationship between ensemble accuracy and diversity is still unclear, which may depend on the specific applications and the metrics used for diversity measurement [Kuncheva and Whitaker, 2003]. In this chapter, we use the *pair-wise disagreement measure* [Skalak, 1996], [Ho, 1998] to explore the relationship of diversity and the ensemble accuracy of our proposed gait recognition system.

Let $\mathcal{Z} = \{\ddagger_1, \ddagger_2, ..., \ddagger_N\}$ be the test set, and each sample $\ddagger_j$ includes both AGI vector $\mathbf{x}_j \in \mathbb{R}^m$ and model-based feature vector $\mathbf{y}_j \in \mathbb{R}^r$, i.e., $\ddagger_j = \{\mathbf{x}_j, \mathbf{y}_j\} \in \mathbb{R}^{m+r}, j = 1, 2, ..., N$. After fusing model-based information into the RSM system through Eq.(5.6), for simplicity we can represent the l^{th} classifier as $D_l : \mathbb{R}^{m+r} \longrightarrow \{0, 1\}, l \in [1, L]$ such that for $\ddagger_j \in \mathcal{Z}$, $D_l(\ddagger_j) = 1$ if the classification is correct, and $D_l(\ddagger_j) = 0$, otherwise. Based on the classification results from two different classifiers D_l and $D_k, l, k \in [1, L], l \neq$

k, we can count the numbers with respect to four different output scenarios as follows:

$$N^{11} = \forall_{j\in[1,N]} \quad \{\#(D_l(\ddagger_j) = 1 \wedge D_k(\ddagger_j) = 1)\}, \tag{5.9}$$

$$N^{10} = \forall_{j\in[1,N]} \quad \{\#(D_l(\ddagger_j) = 1 \wedge D_k(\ddagger_j) = 0)\}, \tag{5.10}$$

$$N^{01} = \forall_{j\in[1,N]} \quad \{\#(D_l(\ddagger_j) = 0 \wedge D_k(\ddagger_j) = 1)\}, \tag{5.11}$$

$$N^{00} = \forall_{j\in[1,N]} \quad \{\#(D_l(\ddagger_j) = 0 \wedge D_k(\ddagger_j) = 0)\}. \tag{5.12}$$

The disagreement of two classifiers D_k and D_l is equal to the ratio between the number of cases on which D_k and D_l make different predictions (i.e., case N^{10} and case N^{01}) to the total number of test samples N [Kuncheva and Whitaker, 2003], i.e.,

$$Dis(D_k, D_l) = \frac{N^{10} + N^{01}}{N}, \qquad k, l \in [1, L], \tag{5.13}$$

where $N = N^{11} + N^{10} + N^{01} + N^{00}$. We can also write Eq.(5.13) as:

$$Dis(D_k, D_l) = 1 - \frac{(N^{00} + N^{11})}{N}, \qquad k, l \in [1, L]. \tag{5.14}$$

For a multiple classifier system D consisting of L base classifiers, the diversity $Div(D)$ is defined as the average disagreement level of all the $L(L-1)/2$ classifier pairs, i.e.,

$$Div(D) = \frac{2}{L(L-1)} \sum_{l=1,l<k}^{L} Dis(D_k, D_l). \tag{5.15}$$

$Div(D)$ tends to be low when the average base classifiers are either *too weak* or *too strong*, since on test set $\mathcal{Z}$ the outputs of most classifier pairs will be more likely to be either *Both Wrong* (i.e., case N^{00}) or *Both Correct* (i.e., case N^{11}), which are inversely proportional to *Disagreement*, according to Eq.(5.14).

Our aim is to enhance the ensemble accuracy by strengthening the weak classifiers (through fusing the model-based information) without sacrificing the diversity. The experimental evaluation of the relationship among diversity, individual classifier accuracy and ensemble accuracy is provided in Chapter 5.3.3.

5.3 Experimental Evaluation

In this section, we first introduce the datasets and experimental settings used. Then we discuss the performance sensitivity with respect to the random feature number s used for each classifier. By using the model-based information, we explain the enhanced ensemble accuracy in terms of the individual classifier accuracy and diversity. Finally, we compare our system with other state-of-the-art methods for gait recognition from videos with extremely low quality.

5.3.1 Dataset and Configuration

The proposed method is evaluated on the *extremely low quality versions* of the indoor OU-ISIR-D database [Makihara et al., 2012] and the outdoor USF database [Sarkar et al., 2005]. Both databases provide the binary aligned silhouettes. The original resolution and frame-rate in OU-ISIR-D database are 128×88 pixels and 60 fps [Makihara et al., 2012], while they are 128×88 pixels and 30 fps in the USF database [Sarkar et al., 2005]. The intention of this work is to propose a system that is capable of dealing with extremely low video quality. Therefore, we downsample the afore-mentioned databases to create *extremely low quality versions* with lower resolution (i.e., 32×22 pixels) and frame-rate (i.e., 1 fps) in a manner similar to [Mori et al., 2010],[Akae et al., 2012].

The OU-ISIR-D database consists of two datasets, namely, DB-high (i.e., with small gait fluctuations) and DB-low (i.e., with large gait fluctuations). For DB-high/DB-low, there are 100 subjects (1 subject per sequence) for both the gallery and probe. For the outdoor USF database, 12 experiments were initially designed by Sarkar et al. for algorithm evaluations against covariate factors such as camera viewpoint, shoe, carrying condition, walking surface, and elapsed time, etc. Since in this chapter we focus on human gait recog-

Table 5.1: Datasets configuration. DB-high/DB-low has low/high gait fluctuations; DB-outdoor has high levels of segmentation errors, and camera viewpoint covariate.

Dataset	DB-high	DB-low	DB-outdoor
#Subject	100	100	122
#Seq. per Subject	1	1	1
Resolution (pixels)	32×22	32×22	32×22
Frame-rate (fps)	1	1	1
#Frames per Seq.	6	6	$4 \sim 7$

nition from extremely low quality videos, instead of evaluating our method on the 12 experiments (with default quality), only the extremely low quality version of USF dataset A is used, and we refer to it as DB-outdoor in this chapter. DB-outdoor includes 122 subjects (1 subject per sequence) for both the gallery and probe, which are captured in different camera viewpoints (about 30° difference). A summary of the datasets configuration is shown in Table 5.1.

The proposed model-based methods (Model 1 or Model 2) use the average width vector of a certain area (head or the whole silhouette) as feature template, as shown in Fig. 5.2. For Model 1, we simply define the head area as roughly the topmost $1/3$ of the whole silhouette. Specifically, for videos with resolution 32×22 pixels, the dimensionality for feature vector corresponding to Model 1 (resp. Model 2) is $r = 10$ (resp. $r = 32$).

The rank-1 correct classification rate (CCR) is used to measure the performance. We repeat all the experiments 10 times and report the mean values. In Table 5.3, we also report results in terms of the mean and standard deviation.

5.3.2 The Effect of Random Feature Number

For the initial eigenspace construction, we choose eigenvectors corresponding to the largest 200 eigenvalues (i.e., $v = 200$), which preserves nearly 100% of the variance. For each random subspace, the corresponding base classifier has some generalisation ability for the unselected features (i.e., unselected subspaces) [Ho, 1998]. However, the underfitting problem may arise if number of random features s is too small. On the other hand, the base

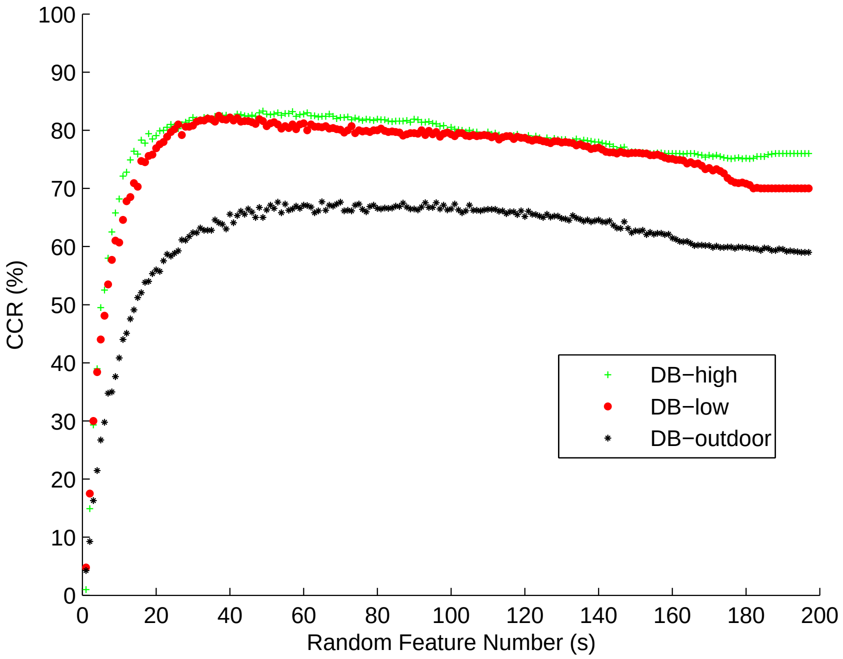


Figure 5.3: The rank-1 CCR distribution (%) with respect to the number of random features (s), given both the probe and gallery videos with frame-rate at 1 fps and resolution of 32×22 pixels.

classifiers may be overfitted if s is too large.

By setting the classifier number $L = 1000$, we check the system's sensitivity to s within the range $[2, v - 2]$ on the three datasets (i.e., DB-high, DB-low, and DB-outdoor). The rank-1 CCR distribution with respect to s shown in Fig. 5.3 clearly indicates the effect of fusion based on underfitted (e.g., with $s \leq 20$) or overfitted (e.g., with $s \geq 160$) classifiers. They tend to have lower accuracies, and the reasons may be: 1) for underfitted/weak classifiers, there is not enough information for them to make the correct classification; 2) for overfitted/strong classifiers, due to the highly overlapped feature set, they tend to make the same prediction (i.e., lack of diversity), which makes the fusion less effective.

Table 5.2: Rank-1 CCR (%) of Model 1/Model 2 on the three datasets, given both the probe and gallery videos with frame-rate at 1 fps and resolution of 32×22 pixels.

-	DB-high	DB-low	DB-outdoor
Model 1	57	51	6.56
Model 2	59	62	35.25

5.3.3 Performance Gain Analysis

To enhance the overall fusion performance, we aim to use ancillary information (from model-based methods) to enhance the underfitted classifiers. Although Model 1 and Model 2 may be less reliable (see Table 5.2), they may provide information from a different perspective.

To measure the effect through fusing model-based information, three metrics are used, i.e., rank-1 CCR of the ensemble ($CCR_{ensemble}$), rank-1 CCR of individual classifier (CCR_{indiv}), and diversity (Div). $CCR_{ensemble}$ is the performance through majority voting (see Eq.(5.7)). CCR_{indiv} is the average performance of the L classifiers. Given a test set $\mathcal{Z} = [\ddagger_1, \ddagger_2, ..., \ddagger_N]$ and classifier $D_l : \mathbb{R}^{m+r} \longrightarrow \{0, 1\}$, we have

$$CCR_{indiv} = \frac{1}{L} \sum_{l=1}^{L} \sum_{j=1}^{N} \frac{D_l(\ddagger_j)}{N}. \tag{5.16}$$

Div is the average of the disagreement levels between any two different classifiers in a multiple classifier system (see Eq.(5.15)).

Based on different number of random features $s = \{20, 40, 80, 160\}$, we conduct experiments on various values of the weight ω within the range $[0, 1]$ with an interval of 0.1. It is worth noting that when $\omega = 0$, the proposed system is the same as the conventional RSM system without fusing model-based information. On the three datasets (both the probe and gallery videos with frame-rate at 1 fps and resolution of 32×22 pixels), the performance distributions with respect to ω are shown in Fig. 5.4-5.6. Note we do not report the results corresponding to Model 1 on DB-outdoor, which provides extremely unreliable information in the outdoor environment (with 6.56% accuracy, see Table 5.2). From Fig. 5.4-5.6, we can observe:

1. Generally, $CCR_{ensemble} \propto Div$.
2. Without fusing model-based information (i.e., $\omega = 0$), Div is relatively low (e.g., $Div \approx 10\%$) when the base classifiers are either too strong (e.g., see Fig. 5.4(d), 5.5(d) when $CCR_{indiv} \approx 70\%$) or too weak (e.g.,see Fig. 5.6(a) when $CCR_{indiv} \approx 6\%$).
3. $Div \propto CCR_{indiv}$ holds only for weak classifier ensemble (e.g.,with $CCR_{indiv} \leq 30\%$). When $CCR_{indiv} \geq 30\%$, however, Div starts to decrease with respect to the strengthening base classifiers.
4. Model-based information may enhance CCR_{indiv} when low weight is assigned (e.g., $\omega \leq 0.5$). The enhancements tend to be more significant for underfitted classifiers than overfitted classifiers.

Diversity is important for classifier ensembles, and our experimental results suggest that $CCR_{ensemble} \propto Div$ in our gait recognition scenarios. Diversity can be increased by enhancing the weak classifiers (with lower CCR_{indiv}), which are constructed based on a small number of random features, e.g., $s = 20$. Model-based information can then be used to increase the diversity, and thus enhance $CCR_{ensemble}$. Given the fact it is not beneficial to fuse stronger base classifiers, it is preferable to assign the model-based information a lower weight in order to preserve the diversity.

For model-based information, although Model 1 has lower rank-1 CCRs in the indoor datasets (see Table 5.2), it can provide some ancillary information to the RSM system (with higher $CCR_{ensemble}$). One explanation is that compared with Model 2 (based on the whole silhouette), Model 1 (based on the head area) is less correlated with the RSM-based weak classifiers, which are derived from the whole body. DB-outdoor is more challenging due to higher levels of segmentation errors and camera viewpoint covariate, and in the case the model-based information can be extremely unreliable (see Table 5.2). Nevertheless, compared with the best results from conventional RSM without fusing model-based information (see Fig. 5.3), incorporating Model 2 (with 35.25% rank-1 CCR) into the underfitted classifiers can still increase the performance by up to 5%.

Gait identification from extremely low quality videos is a challenging task due to the lack of information. In the low resolution condition, it is also difficult to capture other modalities (e.g., face) to enhance the performance of the gait recognition system. Experimental results suggest the effectiveness of our method in this limited condition, since both model-based information and RSM-based classifiers can be derived from the low quality silhouettes. Although model-based information may be less reliable, they may reveal the data structure from a different perspective. Using such information may enhance the RSM-based weak classifiers without sacrificing the diversity.

5.3.4 Algorithms Comparison

On DB-high and DB-low, we compare our method with other algorithms, i.e., morphing-based reconstruction (Morph) [Al-Huseiny et al., 2010], periodic temporal SR (PTSR) [Akae et al., 2011], and example-based and reconstruction-based temporal SR (ERTSR) [Akae et al., 2012]. We directly quote the results of Morph, PTSR, and ERTSR from [Akae et al., 2012], which are based on the same experimental settings as ours.

As stated in Chapter 5.3.3, higher performance can be achieved by combining low-weighted model-based information (Model 1) with a weaker classifier ensemble. In Table 5.3, we report our results (mean and standard deviation of 10 runs) based on $s = \{20, 40\}$, $\omega = 0.2$. It is also worth noting that our system is less sensitive to s and ω within a certain range, as shown Fig. 5.4-5.5.

From Table 5.3, we can see that reconstruction-based methods ([Al-Huseiny et al., 2010],[Akae et al., 2011]) tend to have low rank-1 CCRs. This is because significant amount of artifacts can be generated due to the extremely low frame-rate and low resolution, and reconstruction-based methods [Al-Huseiny et al., 2010],[Akae et al., 2011] are not able to cope with those artifacts effectively. EPTSR [Akae et al., 2012] can greatly improve the

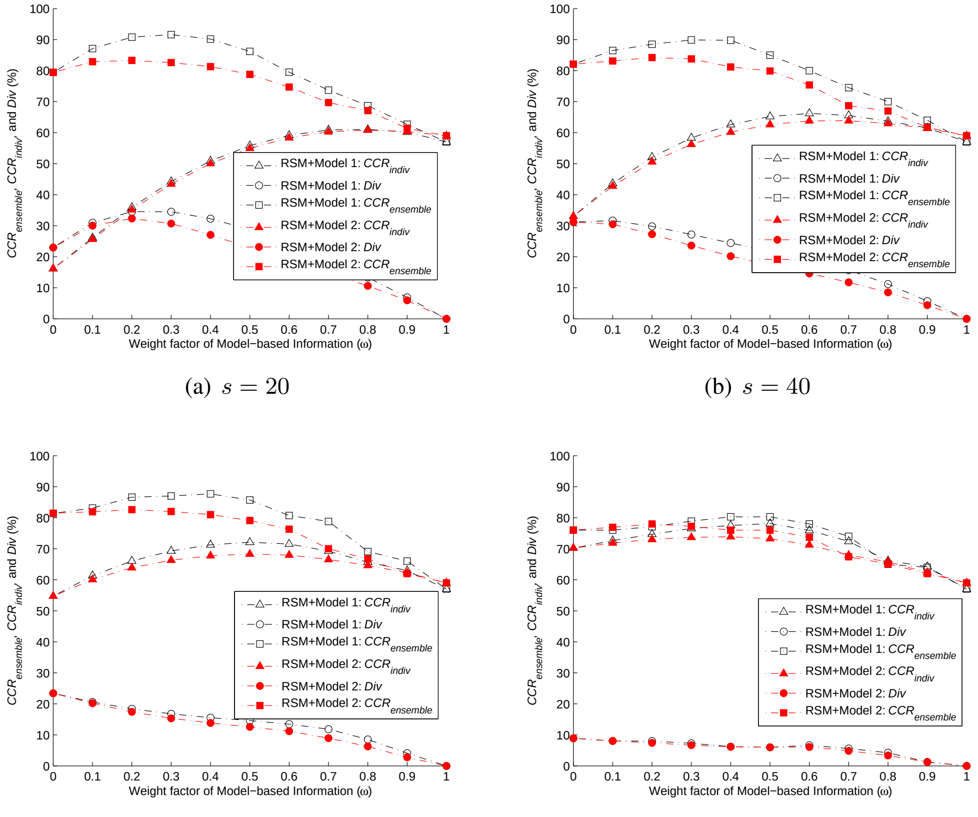
Table 5.3: Algorithms comparisons in terms of rank-1 CCR (%) on DB-high/DB-low, given both the probe and gallery videos with frame-rate at 1 fps and resolution of 32×22 pixels.

-	DB-high	DB-low
Morph [Al-Huseiny et al., 2010]	52	N/A
PTSR [Akae et al., 2011]	44	N/A
ERTSR [Akae et al., 2012]	87	N/A
RSM ($s = 20$)	79.50±1.90	75.20±2.49
RSM ($s = 40$)	82.10±0.57	81.80±1.75
RSM+Model 1($s = 20$)	**90.80±1.48**	**88.40±1.43**
RSM+Model 1($s = 40$)	88.50±1.35	87.50±1.18

accuracy by assuming that the degree of motion is the same among gait cycles. However, this assumption does not hold when there are large gait fluctuations (e.g., on DB-low) [Akae et al., 2012]. Compared with the three methods, the RSM-based method is more adaptive and can be applied to both DB-high an DB-low with reasonable accuracies. In this chapter, fusing model-based information into the RSM system can further reduce the error rates. This effect is more significant for weaker classifier ensembles (e.g., with $s = 20$).

5.4 Summary

In this chapter, we propose an enhanced classifier ensemble method for gait identification from extremely low quality videos. By incorporating the model-based information into the RSM-based weak classifiers, the diversity of the classifiers can be enhanced, which is positively correlated to the ensemble accuracy. We also find that it is less beneficial to combine stronger base classifiers with the model-based information, since in this case they tend to have the same prediction, which contributes negatively to the diversity of the whole multiple classifier system. Compared with other state-of-the-art algorithms, our method delivers significant improvements in terms of identification accuracy and generalisation capability. More details of this chapter can be found in our publications in [Guan et al., 2013a], and [Guan et al., 2014].



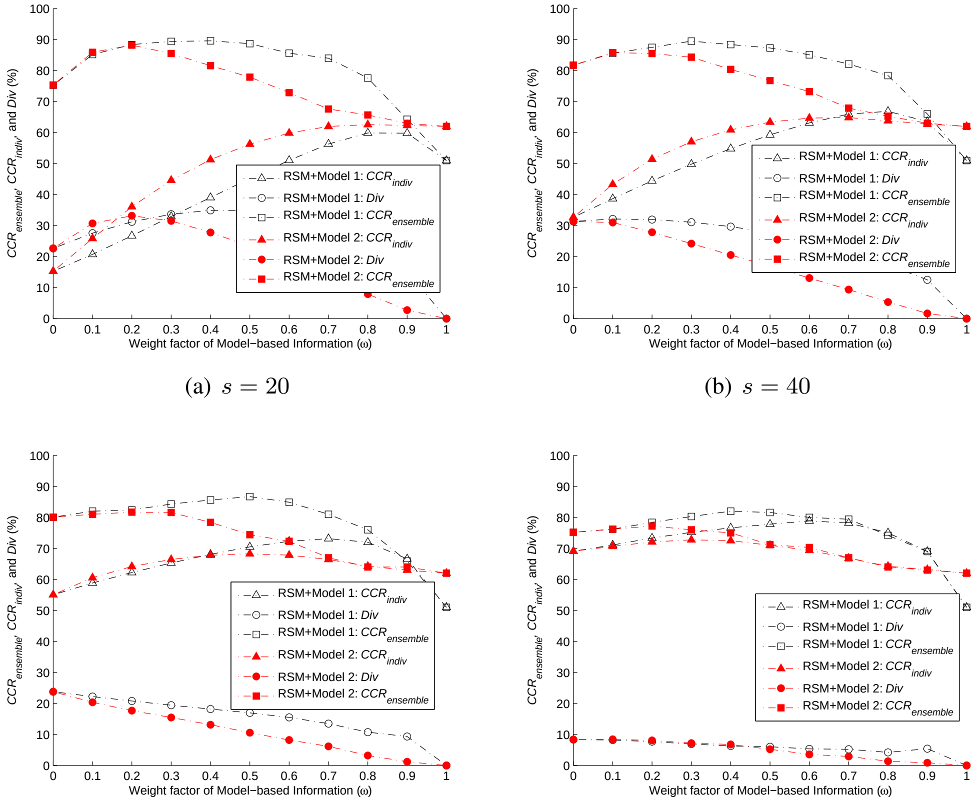
(a) $s = 20$

(b) $s = 40$

(c) $s = 80$

(d) $s = 160$

Figure 5.4: On DB-high, performance distributions with respect to the weight of model-based information (ω). (a)-(d) are based on classifier ensemble with random feature number $s = \{20, 40, 80, 160\}$, respectively.



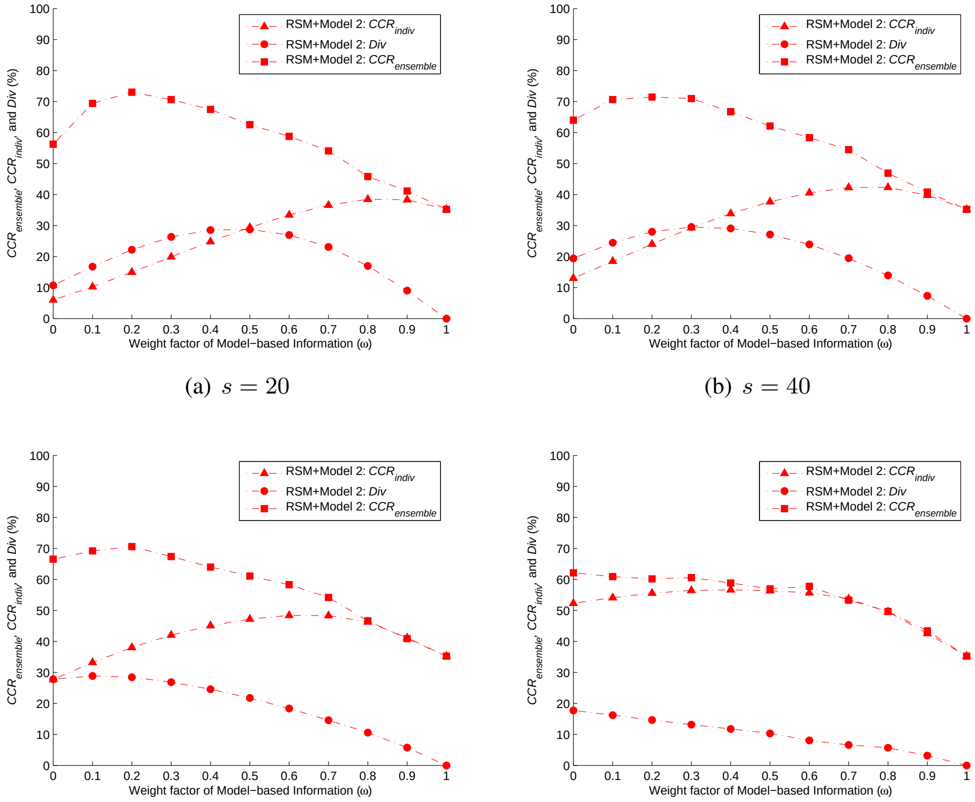
(a) $s = 20$

(b) $s = 40$

(c) $s = 80$

(d) $s = 160$

Figure 5.5: On DB-low, performance distributions with respect to the weight of model-based information (ω). (a)-(d) are based on classifier ensemble with random feature number $s = \{20, 40, 80, 160\}$, respectively.



(a) $s = 20$

(b) $s = 40$

(c) $s = 80$

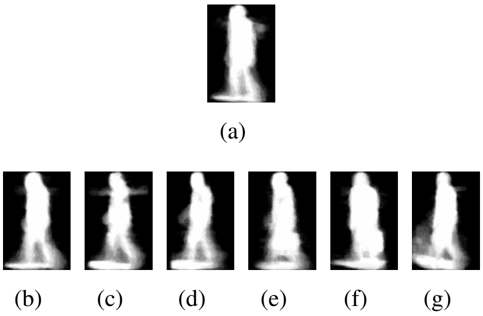
(d) $s = 160$

Figure 5.6: On DB-outdoor, performance distributions with respect to the weight of model-based information (ω). (a)-(d) are based on classifier ensemble with random feature number $s = \{20, 40, 80, 160\}$, respectively.

Chapter 6

Reducing the Effect of Unknown Covariates using RSM-based Hybrid Decision-level Fusion (RSM-HDF)

Compared with covariates like speed or camera viewpoint, it is difficult to predict the effect of covariates like clothing, walking surface, elapsed time, etc. In this chapter, we aim to reduce the effect of such unknown covariates. We model the effect as an unknown partial feature corruption problem. Since the locations of corruptions may differ for different query gaits, relevant features may dynamically turn into irrelevant features when walking condition changes. In this case, it is difficult to train one fixed classifier that is robust to a large number of different covariates. First, we give a detailed analysis of RSM-based classifier ensemble for tackling unknown covariates with different feature corruption locations. According to the analysis, then we propose a hybrid decision-level fusion (HDF) on the RSM framework on solving the hard problems, e.g., gait recognition against the most challenging covariates like clothing, walking surface, and elapsed time. Finally, we evaluate our method on the USF dataset and OU-ISIR-B dataset, and it has much higher performance than other state-of-the-art algorithms.



(a)

(b) (c) (d) (e) (f) (g)

Figure 6.1: GEI samples of the same subject from the USF gait dataset [Sarkar et al., 2005]. First row: (a) gallery sample in normal condition. Second row: probe samples under the influences of (b) viewpoint, (c) walking surface, (d) viewpoint and walking surface, (e) carrying condition, (f) carrying condition and viewpoint, (g) elapsed time, shoe type, clothing, and walking surface.

6.1 Challenges, Problem Formulation, and Solution

For unimodal biometric systems, one of the major limitations is that the performance can be significantly affected by large intra-class variations [Jain et al., 2004b]. In the context of gait recognition, large intra-class variations are caused by unknown covariates, such as carrying condition, clothing, walking surface, camera viewpoint, elapsed time, etc. Fig. 6.1 demonstrates several GEI samples from the USF dataset [Sarkar et al., 2005] in different walking conditions. We can see that covariates may significantly change the appearance of the GEIs, and it is desirable to extract covariates-invariant features for classification against different covariates in unknown walking conditions.

For feature extraction, previous works (e.g., [Han and Bhanu, 2006] [Li et al., 2008] [Xu et al., 2006][Xu et al., 2007][Tao et al., 2007] [Chen et al., 2010] [Lai et al., 2014] [Lu et al., 2014]) learn metrics from training data (i.e., gallery in normal condition), and the effect of covariates can be reduced to some extent. However, an effective metric can only be learned based on representative training data. Unfortunately, getting hold of a representative training set is often difficult, in real-world gait recognition scenarios. Since most of the walking conditions of query gaits are unknown, the training data collected in normal condition cannot represent the whole population. In this case, overfitting the less representative training data is the major problem of the traditional methods relying on

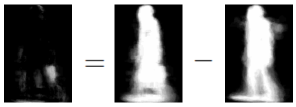


Figure 6.2: An example on modelling the covariate effect by image difference between the gallery GEI (i.e., Fig.6.1(a)) and a probe GEI (i.e., Fig.6.1(e))

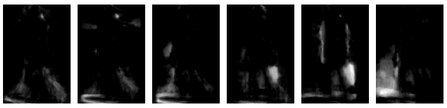


Figure 6.3: Difference images between the gallery GEI Fig.6.1(a) and the probe images Fig.6.1(b)-Fig.6.1(g)

learning.

In this chapter, by using GEIs, we model the effect caused by various covariates as an unknown partial feature corruption problem, and deem the solution of robust gait recognition as a *dynamic* feature selection problem. Through ideal cases analysis, we provide the general classifier ensemble solution. To tackle the hard problems in real cases, we employ two strategies, namely, local enhancing (LE) (as introduced in Chapter 3.2.2) and hybrid decision-level fusion (HDF).

6.1.1 Gait Recognition Challenges

Walking conditions with covariates may affect the human gait patterns. Based on the image-based representations (e.g., GEI), such effect of covariates can be expressed as the image difference between the gallery and probe. Given a subject with the gallery GEI sample $\mathbf{I}_{gallery}$ in normal condition, the probe GEI samples $\{\mathbf{I}_i^{probe}\}_{i=1}^{F}$ in F different walking conditions, we define the corresponding difference images as:

$$\hat{\mathbf{I}}_i = \mathbf{I}_i^{probe} - \mathbf{I}^{gallery}, \quad i \in [1, F], \tag{6.1}$$

as illustrated in Fig. 6.2. Several examples of $\hat{\mathbf{I}}_i$ corresponding to different walking conditions are shown in Fig. 6.3. For an unknown walking condition i, $\hat{\mathbf{I}}_i$ indicates the corrupted gait features with unknown locations. Before matching, we need to find a feature extractor

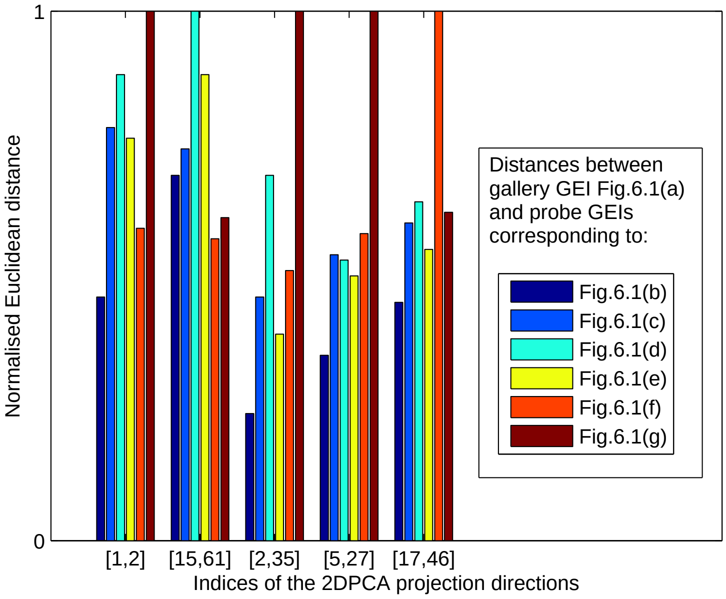


Figure 6.4: Normalised Euclidean distances from gallery sample Fig. 6.1(a) to probe samples Fig.6.1(b)-Fig.6.1(g) based on different projection directions

$\mathbf{T}^*$ that can suppress $\hat{\mathbf{I}}_i$, i.e.,

$$\mathbf{T}^* = \text{argmin}\, \|\hat{\mathbf{I}}_i \mathbf{T}\|^2, \quad i \in [1, F], \tag{6.2}$$

where $\mathbf{T}$ can be defined as some transformation matrices that can extract features in the column direction of $\hat{\mathbf{I}}_i$. However, the locations of corruptions may differ for different walking conditions, as shown in Fig. 6.3. Given such *non-deterministic* nature of $\hat{\mathbf{I}}_i$, from Eq.(6.2) we can see that it is difficult to find a *fixed* $\mathbf{T}^*$ that can extract effective features that can generalise to a large number of different covariates. In light of this, an effective $\mathbf{T}^*$ should only extract the relevant features *dynamically* with respect to different walking conditions.

6.1.2 Problem Formulation: A Dynamic Feature Selection Problem

In this chapter we use 2DPCA to project the gait data. Assume $\mathbf{T}$ is the 2DPCA transformation matrix, which consists of the largest d non-zero principle components such that

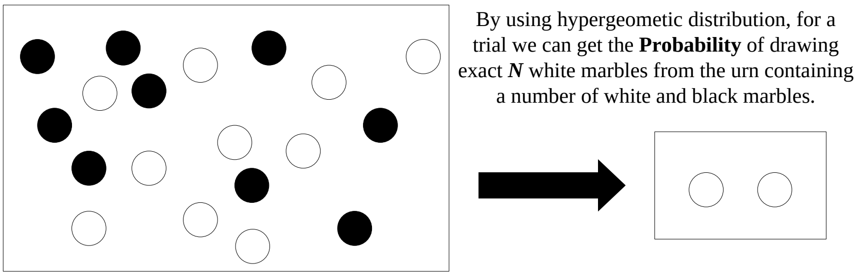
$\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, ..., \mathbf{t}_d]$. Since $[\mathbf{t}_1, \mathbf{t}_2, ..., \mathbf{t}_d]$ are pairwise orthogonal vectors, Eq.(6.2) can be written as:

$$\begin{aligned} \mathbf{T}^* &= \text{argmin} \, ||\hat{\mathbf{I}}_i[\mathbf{t}_1, \mathbf{t}_2, ..., \mathbf{t}_d]||^2 \\ &= \text{argmin} \, ||\hat{\mathbf{I}}_i\mathbf{t}_1||^2 + ||\hat{\mathbf{I}}_i\mathbf{t}_2||^2 \\ &\qquad + ... + ||\hat{\mathbf{I}}_i\mathbf{t}_d||^2, \quad i \in [1, F]. \end{aligned} \tag{6.3}$$

It is difficult for traditional 2DPCA with a fixed $\mathbf{T}^* = [\mathbf{t}_1, \mathbf{t}_2, ..., \mathbf{t}_d]$ to reduce the effect of F different walking conditions. In Eq.(6.3), since the terms $||\hat{\mathbf{I}}_i\mathbf{t}_j||^2, j \in [1, d], i \in [1, F]$ are non-negative terms (corresponding to the holistic features), it is possible to reduce the effect of covariates by selecting a subset of $[\mathbf{t}_1, \mathbf{t}_2, ..., \mathbf{t}_d]$ to form a new feature extractor.

To see the reduced effect of covariates, we form different feature extractors by randomly selecting 2 projection directions from $[\mathbf{t}_1, \mathbf{t}_2, ..., \mathbf{t}_d]$. In Fig. 6.4 we report the corresponding normalised Euclidean distances from gallery sample Fig.6.1(a) to probe samples Fig.6.1(b)-Fig.6.1(g), respectively. We can see that the effect of covariates can be reduced based on certain projection directions. For example, it is better to use projection directions $[\mathbf{t}_2, \mathbf{t}_{35}]$ to reduce the effect of walking conditions in Fig.6.1(b) and Fig.6.1(e), while it is preferable to use projection directions $[\mathbf{t}_{15}, \mathbf{t}_{61}]$ to reduce the effect of walking conditions in Fig. 6.1(f) and Fig. 6.1(g), as shown in Fig.6.4. Although based on holistic features, these observations empirically validate that it is possible to select a subset of $[\mathbf{t}_1, \mathbf{t}_2, ..., \mathbf{t}_d]$ as feature extractors to reduce the effect of certain covariates. However, the optimal subset may vary given different walking conditions due to the *non-deterministic* nature of $\hat{\mathbf{I}}_i$.

Therefore, we formulate the problem of reducing the effect of unknown covariates as a *dynamic* feature selection problem. For an unknown walking condition, we assume the irrelevant features (mainly corresponding to the corrupted pixels, as indicated by $\hat{\mathbf{I}}_i$) exist in $m(0 < m < d)$ unknown non-negative terms $||\hat{\mathbf{I}}_i t_j||^2, j \in [1, d]$ in Eq.(6.3). We aim to *dynamically* select $N \in [1, d]$ relevant features (mainly corresponding to the uncorrupted pixels) for classification. It is obvious that the probability of selecting N relevant features is 0 when $N > d - m$. When $N \in [1, d - m]$, by randomly selecting N features out of



By using hypergeometic distribution, for a trial we can get the **Probability** of drawing exact ***N*** white marbles from the urn containing a number of white and black marbles.

Figure 6.5: An example of hypergeometric distribution application in classical urn problem

d features, the probability of *not* choosing the m irrelevant features follows the hypergeometric distribution [Walck, 2007], which is defined as a discrete probability distribution that describes the probability of k successes in N draws, without replacement, from a finite population of size d containing exactly m failures and $d-m$ successes. The corresponding probability mass function [Walck, 2007] is given by the ratio of binomial coefficients:

$$P(k) = \frac{\binom{d-m}{k}\binom{d-(d-m)}{N-k}}{\binom{d}{N}}. \tag{6.4}$$

In our case the probability of not selecting the m irrelevant features (i.e., selecting $k = N$ relevant features for N draws) can be expressed as:

$$\begin{aligned} P(k=N) &= \frac{\binom{d-m}{N}\binom{d-(d-m)}{N-N}}{\binom{d}{N}} \\ &= \frac{(d-m)!(d-N)!}{d!(d-m-N)!}, \qquad N \in [1, d-m]. \end{aligned} \tag{6.5}$$

We also provide a similar example of hypergeometric distribution application in classical urn problem in Fig. 6.5. For simplicity, we use $P(N)$ instead of $P(k = N)$ for the rest of this chapter. $P(N)$ is a generalisation measure that is related, someway, to the performance of the matching algorithm. Classifiers corresponding to greater $P(N)$ tend to generalise well to unseen covariates, and vice versa.

Lemma 1. *Let d, m, N be the numbers of the total features, irrelevant features (with $0 <$*

$m < d$), and randomly selected features, respectively, then $P(N)$ in Eq.(6.5) is inversely proportional to N in the range $N \in [1, d-m]$.

Proof. We use mathematical induction for this proof, and we need to prove $P(N+1)/P(N) < 1$ for all the $N \in [1, d-m]$. According to Eq.(6.5), we have

$$\begin{aligned}\frac{P(N+1)}{P(N)} =& \frac{(d-N-1)!(d-m-N)!}{(d-m-N-1)!(d-N)!} \\ =& 1 - \frac{m}{d-N} < 1,\end{aligned} \tag{6.6}$$

where the last inequality follows since we have $0 < \dfrac{m}{d-N} < 1$ given $N \in [1, d-m]$. This completes the proof of the lemma. □

According to Lemma 1, we can see that classifiers corresponding to smaller values of N tend to generalise well (i.e., with greater $P(N)$). Therefore, it is preferable to set N to smaller values. In this case, the general form of $P(N)$ in Eq.(6.5) can also be simplified. Since when $N \ll d$, the hypergeometric distribution can be deemed as a binomial distribution [Walck, 2007], we have

$$\begin{aligned}P(N) &= \binom{N}{N} p^N (1-p)^{(N-N)} \\ &= p^N, \quad N \ll d,\end{aligned} \tag{6.7}$$

where p is the probability of randomly selecting one relevant feature. Since $p = 1 - m/d$, Eq.(6.7) can be written as:

$$P(N) = (1 - \frac{m}{d})^N, \quad N \ll d, \tag{6.8}$$

which clearly reflects the simple relationship between $P(N)$ and the other parameters when $N \ll d$.

6.1.3 Classifier Ensemble Solution and its Extensions

We use a classifier ensemble strategy for covariate-invariant gait recognition. After repeating the random feature selection process L times (L is a *large number*), L base classifiers with generalisation measure $P(N)$ (see Eq.(6.8)) are generated. Classification can then be performed by majority voting [Kittler et al., 1998] among the labeling results of the base classifiers.

Given a query gait and the gallery consisting of c classes, we define the true votes V_{true} as the number of classifiers with correct prediction, while the false votes $\{V_{false}^i\}_{i=1}^{c-1}$ as the incorrectly predicted classifier number distribution over the other $c-1$ classes. Given that, through majority voting, the final correct classification can be achieved only when $V_{true} > \max\{V_{false}^i\}_{i=1}^{c-1}$.

Ideal Cases Analysis:

We make two assumptions in ideal cases:

1. When irrelevant features are not selected by a classifier (i.e., the unaffected classifier), correct classification should be achieved.
2. When irrelevant features are selected by a classifier (i.e., the affected classifier), the output label should be a random guess from the c classes, with a probability of $1/c$.

Given L classifiers, in the ideal cases the true votes $\bar{V}_{true}$ can be approximated as the sum of the number of unaffected classifiers $P(N)L$ (based on *the law of large numbers* [Grinstead and Snell, 1997]) and the number of affected classifiers:

$$\bar{V}_{true} \approx \text{round}\left(P(N)L + \frac{(1-P(N))L}{c} \right). \tag{6.9}$$

Similarly, in the ideal cases the false votes should correspond to the number of affected classifiers for the other $c-1$ classes. Based on assumption 2, its maximum value, max$\{\bar{V}^i_{false}\}^{c-1}_{i=1}$, can be roughly expressed as the corresponding mean value:

$$\max\{\bar{V}^i_{false}\}^{c-1}_{i=1} \approx \mathrm{mean}\{\bar{V}^i_{false}\}^{c-1}_{i=1} \approx \mathrm{round}\left(\frac{(1-P(N))L}{c}\right). \tag{6.10}$$

Correct classification can be achieved when $\bar{V}_{true} >$ max$\{\bar{V}^i_{false}\}^{c-1}_{i=1}$. From Eq.(6.9) and Eq.(6.10), it is obvious that this condition can be met when $P(N) > \varepsilon$ (ε is a small positive number).

From ideal cases analysis, it can be seen that we ***only care*** *about the number of unaffected classifiers, instead of which ones are. According to Eq.(6.8), by setting N to a small number to make $P(N) > \varepsilon$, the corresponding classifier ensemble method* ***can be insensitive to the locations of corrupted gait features*** *(since $\bar{V}_{true} >$ max$\{\bar{V}^i_{false}\}^{c-1}_{i=1}$, as supported by Eq.(6.9) and Eq.(6.10)). Therefore, it is robust to a large number of covariates as long as the number of irrelevant features is not extremely large (i.e., $m < d$ in Eq.(6.8)).*

Real Cases Analysis:

Here the two assumptions made in the ideal cases analysis do not hold precisely:

1. Unaffected classifiers do not always make the correct classification. The violation of this assumption may decrease the true votes i.e.,($V_{true} \leq \bar{V}_{true}$) and increase the total number of false votes, from which we can also have mean$\{V^i_{false}\}^{c-1}_{i=1} \geq$ mean$\{\bar{V}^i_{false}\}^{c-1}_{i=1}$.

2. Although the affected classifiers would result in label assigning in a relatively random manner, it is difficult to estimate max$\{V^i_{false}\}^{c-1}_{i=1}$ precisely. We can only get max$\{V^i_{false}\}^{c-1}_{i=1} \geq$ mean$\{V^i_{false}\}^{c-1}_{i=1}$.

From above it is easy to see that max$\{V^i_{false}\}^{c-1}_{i=1} \geq$ mean$\{\bar{V}^i_{false}\}^{c-1}_{i=1}$. Since in ideal cases we also have max$\{\bar{V}^i_{false}\}^{c-1}_{i=1} \approx$ mean$\{\bar{V}^i_{false}\}^{c-1}_{i=1}$, in real cases we can deem $\bar{V}_{true}$ and max$\{\bar{V}^i_{false}\}^{c-1}_{i=1}$ from ideal cases analysis as the upper bound and lower bound, respectively, i.e.,

$$\begin{aligned} V_{true} &\leq \bar{V}_{true}, \\ \max\{V_{false}^i\}_{i=1}^{c-1} &\geq \max\{\bar{V}_{false}^i\}_{i=1}^{c-1}. \end{aligned} \tag{6.11}$$

Correct classification can be achieved only when $V_{true} > \max\{V_{false}^i\}_{i=1}^{c-1}$, our objective is to increase V_{true} and decrease $\max\{V_{false}^i\}_{i=1}^{c-1}$. Since it is difficult to estimate $\max\{V_{false}^i\}_{i=1}^{c-1}$ precisely, we relax the objective from decreasing $\max\{V_{false}^i\}_{i=1}^{c-1}$ to decreasing $\sum_{i=1}^{c-1} V_{false}^i$. We define a variable Γ as the ratio of false votes to true votes, and our objective become a problem of suppressing Γ, which is defined as:

$$\Gamma = \frac{\sum_{i=1}^{c-1} V_{false}^i}{V_{true} + \epsilon}, \tag{6.12}$$

where ϵ is a small positive number to avoid trivial results. To suppress Γ in Eq.(6.12), we extend the classifier ensemble method by proposing two strategies:

1. **Local Enhancing**

   If we use the output labels of all the L classifiers as valid votes such that $L = V_{true} + \sum_{i=1}^{c-1} V_{false}^i$, then according to (6.12) we can get $\Gamma \propto L/(V_{true} + \epsilon)$. Since L is a constant, the first strategy is to increase V_{true}. To realise it, based on supervised learning, for each subspace we extract more discriminant features in order to enhance the classification performance. We name this second-stage supervised feature extraction (after the first-stage random feature selection) as local enhancing (LE). In Chapters 3-4, we have used 2DLDA-based LE1 for local enhancing [3].

2. **Hybrid Decision-level Fusion (HDF)**

   This strategy is to decrease $\sum_{i=1}^{c-1} V_{false}^i$ by *dynamically* filtering out classifiers corresponding to the irrelevant features, before majority voting. Irrelevant features would

[3] It is worth noting, a different form of LE can also be to fuse other information (e.g., face information, model-based information) with the base classifiers, as shown in Chapters 4-5.

lead to label assignment in a relatively random manner. Based on the same irrelevant features, a classifier pair corresponding to two different LE methods would output two random labels, which are unlikely to be the same. Based on the "AND" rule, classifier pairs with different output labels are deemed as invalid votes and simply discarded. Although this scheme may also decrease V_{true} to some extent, it significantly reduces the value of $\sum_{i=1}^{c-1} V_{false}^{i}$, which can suppress Γ in Eq.(6.12).

We model the effect of covariates as an unknown partial feature corruption problem. Since the locations of corruptions may differ from one query gait to another, we deem it as a *dynamic* feature selection problem. Our scheme combines a large number of weak classifiers with generalisation measure $P(N)$. In ideal case analysis, we have demonstrated this method is insensitive to the locations of corrupted gait features. This property is the foundation of generalisation to unseen covariates. However, the performance may suffer when the number of corrupted features is large, since the generalisation measure, $P(N)$, of the base classifiers is inversely proportional to the number of irrelevant features m (see Eq.(6.8)), which would result in a high Γ in Eq.(6.12). To solve this problem, we further extend the RSM by using LE and HDF to suppress Γ. The flowchart of RSM-based HDF is shown in Fig. 6.6. More details about HDF can be found in Chapter 6.2.2.

6.2 Extensions of RSM

In this section, we extend RSM to RSM-HDF. Details of random subspace construction and LE1 can be found in Chapter 3.2.1. HDF requires two types of local enhancers (as shown in Fig. 6.6), and in this chapter we use 2DLDA-based LE1 and IDR/QR-based [Ye et al., 2005] local enhancer 2 (LE2). Compared with LE1, LE2 has a much lower time complexity (when the class number is relatively small).

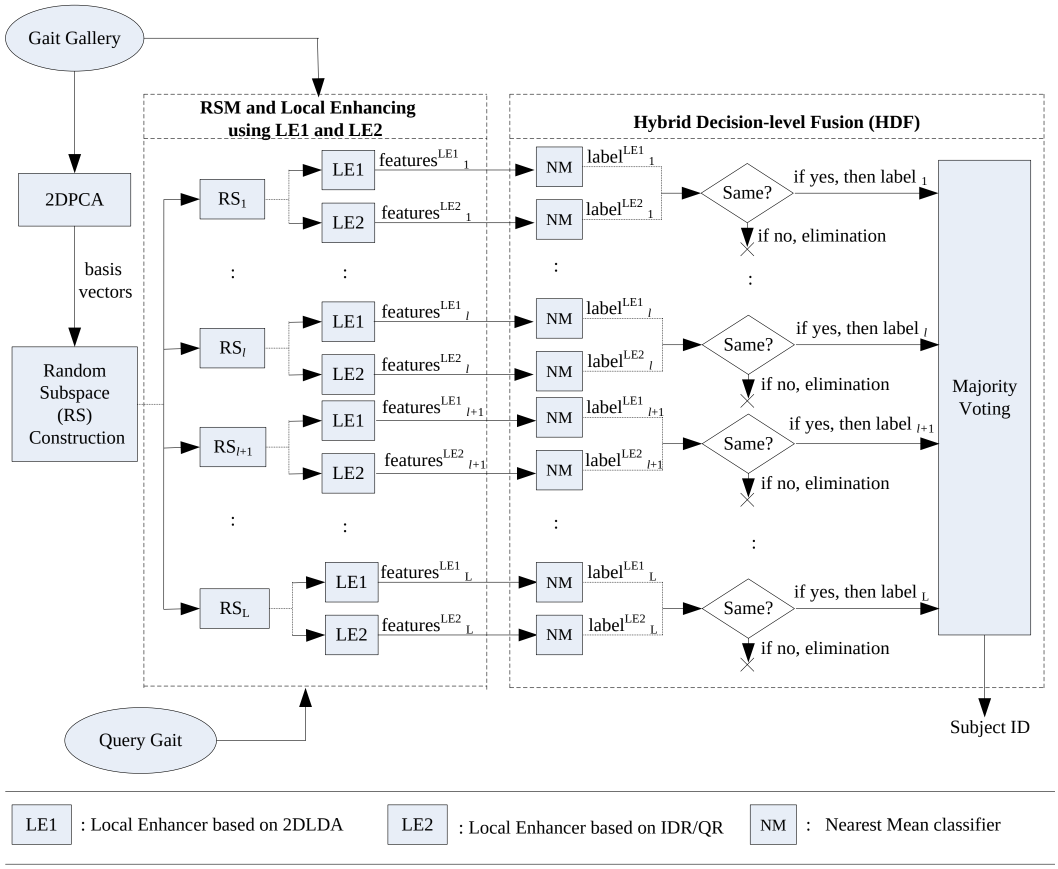


Figure 6.6: RSM-based HDF

6.2.1 Local Enhancer 2 (LE2)

We train LE2 based on IDR/QR, and interested readers may refer to [Ye et al., 2005] for more details about IDR/QR. Compared with LE1, LE2 has a much lower time complexity (see Chapter 6.3.3 for details). The process of generating LE2 is summarised in Algorithm 6.1.

In the l^{th} subspace, given the trained $\hat{\mathbf{V}}^l \in \mathbb{R}^{S\times M}$ (the output of Algorithm 6.1) and a gait sample with the concatenated random feature vector $\hat{\mathbf{X}}^l \in \mathbb{R}^{S\times 1}$, the new feature representation $\hat{\mathbf{x}}^l \in \mathbb{R}^{M\times 1}$ can be extracted by

$$\hat{\mathbf{x}}^l = (\hat{\mathbf{V}}^l)^T \hat{\mathbf{X}}^l, \quad l \in [1, L]. \tag{6.13}$$

Algorithm 6.1 LE2

```
Input: Gallery $\{\mathbf{I}_i \in \mathbb{R}^{N_1 \times N_2}\}_{i=1}^{n}$, random transformation matrices: $\mathbf{R}^l \in \mathbb{R}^{N_2 \times N}, l = 1, 2, ..., L$, and the number of LE2 projection directions $M$;
Output: LE2-based transformation matrices: $\hat{\mathbf{V}}^l \in \mathbb{R}^{S \times M}, l = 1, 2, ..., L$, where $S = N_1 N$;
  Step 1: Random feature extraction on gallery $\mathbf{X}_i^l = \mathbf{I}_i \mathbf{R}^l, \quad i = 1, 2, ...n, l \in [1, L]$;
  Step 2: Concatenating $\mathbf{X}_i^l \in \mathbb{R}^{N_1 \times N}$ to $\hat{\mathbf{X}}_i^l \in \mathbb{R}^{S \times 1}, i = 1, 2, ..., n, l \in [1, L]$;
  for $l = 1$ to $L$ do
    Step 3: For $\hat{\mathbf{X}}_i^l, i = 1, 2, ..., n$ corresponding to $c$ classes,
    1) calculating the global centroid $\bar{\mathbf{m}}^l$;
    2) letting $\hat{\mathbf{D}}_j^l$, $\hat{\mathbf{m}}_j^l$, and $n_j$ be the set of within-class samples, within-class centroid, and sample number for the $j^{th}$ class, respectively;
    Step 4: Constructing the set of within-class centroids: $\mathbf{C} = [\hat{\mathbf{m}}_1^l, \hat{\mathbf{m}}_2^l, ..., \hat{\mathbf{m}}_c^l]$, and performing QR decomposition of $\mathbf{C}$ as $\mathbf{C} = \mathbf{QR}$, where $\mathbf{Q} \in \mathbb{R}^{S \times c}$;
    Step 5: After setting $\mathbf{e}_j = (1, 1, ..., 1)^T \in \mathbb{R}^{n_j}$, computing
    $\mathbf{H}_b^l = [\sqrt{n_1}(\hat{\mathbf{m}}_1^l - \bar{\mathbf{m}}^l), \sqrt{n_2}(\hat{\mathbf{m}}_2^l - \bar{\mathbf{m}}^l), ..., \sqrt{n_c}(\hat{\mathbf{m}}_c^l - \bar{\mathbf{m}}^l)]$,
    $\mathbf{H}_w^l = [\hat{\mathbf{D}}_1^l - \hat{\mathbf{m}}_1^l \mathbf{e}_1{}^T, \hat{\mathbf{D}}_2^l - \hat{\mathbf{m}}_2^l \mathbf{e}_2{}^T, ..., \hat{\mathbf{D}}_c^l - \hat{\mathbf{m}}_c^l \mathbf{e}_c{}^T]$;
    Step 7: Calculating
    $\hat{\mathbf{S}_B^l} = \mathbf{Y}^T \mathbf{Y}$, where $\mathbf{Y} = (\mathbf{H}_b^l)^T \mathbf{Q}$;
    Step 8: Calculating
    $\hat{\mathbf{S}_W^l} = \mathbf{Z}^T \mathbf{Z}$, where $\mathbf{Z} = (\mathbf{H}_w^l)^T \mathbf{Q}$;
    Step 9: For $(\hat{\mathbf{S}_W^l})^{-1} \hat{\mathbf{S}_B^l}$, calculating its $M$ largest eigenvectors, $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_M] \in \mathbb{R}^{c \times M}$.
    Step 10: Setting $\hat{\mathbf{V}}^l = \mathbf{QU}$;
  end for
```

6.2.2 Classification by Hybrid Decision-level Fusion (HDF)

Given L random feature extractors $\mathbf{R}^l, l = 1, 2, ..., L$ and the corresponding LE1 and LE2 ($\hat{\mathbf{W}}^l$ and $\hat{\mathbf{V}}^l, l = 1, 2, ..., L$), the sets of features used for classification can be extracted by using Eq.(3.2), Eq.(3.3), and Eq.(6.13). Then human gait recognition can be performed based on HDF to suppress Γ in Eq.(6.12).

LE1 or LE2 can minimise Γ in Eq.(6.12) mainly by increasing V_{true} and has its own performance upper bound. Based on the output labels from both the LE1-based and LE2-based nearest mean (NM) classifiers (as introduced in Chapter 3.2.3), we aim to *dynamically* eliminate the voters corresponding to the irrelevant features.

Given the query gait $\mathbf{P} = \{\mathbf{P}^l\}_{l=1}^L$ and c classes in the gallery with labels $\{\omega_j\}_{j=1}^c$, for simplicity reasons, let $\Omega_{LE1}^l(\mathbf{P}^l) \in \{\omega_j\}_{j=1}^c$ and $\Omega_{LE2}^l(\mathbf{P}^l) \in \{\omega_j\}_{j=1}^c, l = 1, 2, ..., L$ be the output labels of the NM classifier pairs based on LE1 and LE2, respectively. We define HDF as the classifier ensemble method that outputs the optimal class label $\bar{\Omega}(P)$ such that:

$$\bar{\Omega}(P) = \underset{\omega_j}{\operatorname{argmax}} \sum_{l=1}^{L} \Theta_{\omega_j}^l, \quad j \in [1, c],$$

where

$$\Theta_{\omega_j}^l = \begin{cases} 1, & \text{if } \Omega_{LE1}^l(\mathbf{P}^l) = \Omega_{LE2}^l(\mathbf{P}^l) = \omega_j, \\ 0, & \text{otherwise,} \end{cases} \qquad j \in [1, c]. \tag{6.14}$$

It can be seen that $\sum_{j=1}^c \sum_{l=1}^L \Theta_{\omega_j}^l \leq L$. When there are a large number of irrelevant features for a query gait, most of the false votes can be eliminated such that $\sum_{j=1}^c \sum_{l=1}^L \Theta_{\omega_j}^l \ll L$. In this case, $\sum_{i=1}^{c-1} V_{false}^i$ of Eq.(6.12) can be significantly reduced while V_{true} of Eq.(6.12) are less affected. As such, Γ of Eq.(6.12) can be effectively suppressed. In Chapter 6.3.4, we will also provide the performance gain analysis to study the relationship between Γ and accuracy. The experimental results suggest that, through suppressing Γ, the proposed HDF scheme may lead to a much improved performance against the most challenging covariates like clothing, walking surface, and elapsed time, as demonstrated in Chapter 6.3.4.

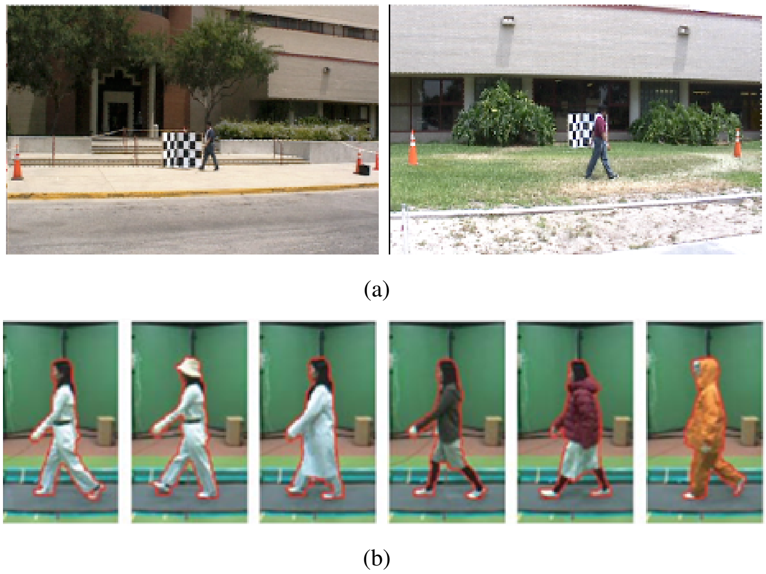
Table 6.1: 12 pre-designed experiments on the USF dataset

Experiment	A	B	C	D	E	F	G	H	I	J	K	L
Probe Size	122	54	54	121	60	121	60	120	60	120	33	33
Covariates	V	H	VH	S	SH	SV	SHV	B	BH	BV	THC	STHC

Abbreviation note: V-View, H-Shoe, S-Surface, B-Briefcase, T-Time, C-Clothing

Table 6.2: List of clothes used in the OU-ISIR-B dataset (Abbreviation: Name)

RP: Regular pants	BP: Baggy pants	SP: Short pants	CP: Casual pants	Sk: Skirt
HS: Half shirt	FS: Full shirt	LC: Long coat	Pk: Parker	DJ: Down jacket
CW: Casual wear	RC: Rain coat	Ht: Hat	Cs: Casquette cap	Mf: Muffler



(a)

(b)

Figure 6.7: Data samples from (a) the USF dataset [Sarkar et al., 2005], and (b) the OU-ISIR-B dataset[Hossain et al., 2010]

Table 6.3: Different clothing combinations in the OU-ISIR-B dataset

Type	s_1	s_2	s_3	Type	s_1	s_2	Type	s_1	s_2	Type	s_1	s_2
3	RP	HS	Ht	0	CP	CW	F	CP	FS	N	SP	HS
4	RP	HS	Cs	2	RP	HS	G	CP	Pk	P	SP	Pk
6	RP	LC	Mf	5	RP	LC	H	CP	DJ	R	RC	-
7	RP	LC	Ht	9	RP	FS	I	BP	HS	S	Sk	HS
8	RP	LC	Cs	A	RP	Pk	J	BP	LC	T	Sk	FS
C	RP	DJ	Mf	B	RP	Dj	K	BP	FS	U	Sk	PK
X	RP	FS	Ht	D	CP	HS	L	BP	Pk	V	Sk	DJ
Y	RP	FS	Cs	E	CP	LC	M	BP	DJ	Z	SP	FS

s_i: i^{th} clothes slot

6.3 Experiments

In this section, first we describe the two benchmark databases, namely USF [Sarkar et al., 2005] and OU-ISIR-B [Hossain et al., 2010] datasets. Then we evaluate our method on different parameter settings. After providing the time complexity of our method, we discuss the effectiveness of the proposed LE and HDF strategies. Finally, we compare the performance of our method with other state-of-the-art algorithms on both datasets.

6.3.1 Datasets

The USF dataset [Sarkar et al., 2005] is a large publicly available human gait database, consisting of 122 subjects walking in outdoor environments. A number of covariates are presented in this dataset: camera viewpoints (left/right), shoes (type A/type B), surface types (grass/concrete), carrying conditions (with/without a briefcase), elapsed time (May/Nov.), and clothing. There are twelve pre-designed experiments for the purpose of testing a single covariate or a combination of covariates, with the size of the probe set ranging from 33 to 122, as shown in Table 6.1. The gallery has all the 122 subjects in normal condition. The OU-ISIR-B dataset was constructed by Hossain et al. [Hossain et al., 2010] for studying the effect of clothing on gait recognition. The evaluation set includes 48 subjects walking on a treadmill with up to 32 types of clothes combinations. Table 6.2 lists the clothes types involved in the dataset, while Table 6.3 gives the types of different clothes combinations. The gallery consists of the 48 subjects in standard clothes (i.e., type 9: regular pants + full shirt), whereas the probe set includes 856 gait sequences of the same 48 subjects in the other 31 clothes types. Several data samples of the USF dataset and OU-ISIR-B dataset are shown in Fig. 6.7. For both datasets, the gallery is used as training set.

In this chapter, we use GEI and the downsampled Gabor-filtered GEI (referred to as Gabor) as the input feature templates. In the USF dataset, the GEIs are available while in the OU-ISIR-B dataset, the aligned silhouettes are provided. After estimating the gait periods using the baseline method [Sarkar et al., 2005], we can construct the GEI samples for the OU-ISIR-B dataset. The size of the GEI (resp. Gabor) templates is 128×88 (resp. 320×352) pixels for both datasets.

To evaluate the performance of the algorithms, we employ rank-1/rank-5 correct classification rate (CCR). Considering the random nature of our method, the results of different runs may vary to some extent. We repeat all the experiments 10 times and the statistics (mean, std, maxima and minima) are reported in Table 6.4 and Table 6.8 for both datasets. The small values of std (listed in Table 6.4 and Table 6.8) of the 10 runs indicate the stability of our method. Therefore, throughout the rest of the chapter, we only report the mean values.

6.3.2 Parameter Settings

There are 3 parameters in our method, namely, the dimensionality of random subspace N, the base classifier number L, and the number of projection directions M for LE1/LE2. Through Lemma 1, we can see that $P(N)$ of each base classifier is inversely proportional to N. As discussed in Chapter 6.1.2, it is preferable to set N to a small number. On the other hand, the classifier ensemble solution works better when the classifier number L is large, since in this case we can roughly estimate the unaffected classifier number as $P(N)L$ based on *the law of large numbers* [Grinstead and Snell, 1997] (see Chapter 6.1.3). Like most subspace learning methods, the performance should not be sensitive to the number of projection directions M for LE1/LE2 (if it is not extremely small). Given these, L should be large while N should be small. On the USF dataset (Experiment A-L) we check the average performance sensitivity to N, M, and L, based on two types of input feature templates (i.e., Gabor and GEI).

1. By empirically setting $L = 1000$ and $M = 20$, we check the sensitivity of the performance against N within the range $[2, 6]$. Fig. 6.8(a) illustrates the average rank-1 CCRs with respect to N. We can see that the recognition accuracies are not sensitive to N.

Table 6.4: Performance statistics in terms of rank-1 CCRs (%) for 10 runs on the USF dataset

-	maxima	minima	std	mean
GEI + RSM	52.44	50.73	0.62	51.45
GEI + RSM(LE1)	64.08	62.36	0.49	63.01
GEI + RSM(LE2)	62.79	59.81	0.91	61.72
GEI + RSM-HDF(LE1,LE2)	70.90	69.20	0.50	70.01
Gabor + RSM	67.58	66.30	0.46	67.06
Gabor + RSM(LE1)	75.29	73.89	0.49	74.56
Gabor + RSM(LE2)	74.95	73.45	0.36	74.27
Gabor + RSM-HDF(LE1,LE2)	82.12	80.30	0.53	81.17

Table 6.5: Running Time (Seconds) on the USF dataset

-	Training Time	Query Time Per Sequence
GEI + RSM(LE1)	91.42	0.58
GEI + RSM(LE2)	28.66	0.26
Gabor + RSM(LE1)	320.09	0.60
Gabor + RSM(LE2)	32.79	0.31

2. Based on $L = 1000$, and $N = 2$, we evaluate the performance distribution by setting $M = [20, 40, 60, 80, 100]$. The results in Fig. 6.8(b) suggest that the performance is not sensitive to M.

3. By setting $N = 2$, and $M = 20$, we can also see from Fig. 6.8(c) that the performance is not decreasing with the increasing number of classifiers.

Given these observations, for the rest of this chapter, we only report the results based on $N = 2$, $M = 20$, and $L = 1000$.

Table 6.6: Rank-1 CCRs (%) of our methods on the USF dataset

Experiment	A	B	C	D	E	F	G	H	I	J	K	L	Avg.
GEI + RSM	89	93	82	24	27	16	16	83	69	54	18	9	51.36
GEI + RSM(LE1)	95	94	86	42	50	23	35	88	88	72	26	19	62.88
GEI + RSM(LE2)	96	94	82	40	46	27	29	87	83	72	19	18	61.70
GEI + RSM-HDF(LE1,LE2)	98	95	88	54	60	37	44	90	93	83	33	21	70.16
Gabor + RSM	96	94	87	47	45	24	38	96	97	85	25	27	67.13
Gabor + RSM(LE1)	100	95	93	62	63	42	50	97	96	89	23	29	74.65
Gabor + RSM(LE2)	98	94	93	60	58	39	47	97	97	92	34	31	74.09
Gabor + RSM-HDF(LE1,LE2)	100	95	94	73	73	55	64	97	99	94	42	42	81.15

6.3.3 Time Complexity Analysis

We analyse the time complexity for generating L classifiers based on LE1/LE2. For a LE1-based classifier in Algorithm 3.2, it takes $O(nNN_1^2)$ (resp. $O(cNN_1^2)$) for $\mathbf{S}_w^l$ (resp. $\mathbf{S}_b^l$), and $O(N_1^3)$ for the eigenvalue decomposition problem. N_1 is the row number of the input feature template (i.e., $N_1 = 128$ for GEI and $N_1 = 320$ for Gabor), and n and c are the training sample number and class number, respectively. Since in our case $n > c$ and $n > N_1$, the time complexity for generating L LE1-based classifiers can be written as $O(LnNN_1^2)$.

For a LE2-based classifier in Algorithm 6.1, it takes $O(c^2S)$ for the QR decomposition, where $S = NN_1$ is the feature number. Calculating $\hat{\mathbf{S}_B^l}$ and $\hat{\mathbf{S}_W^l}$ requires $O(c^3)$ and $O(c^2n)$, while calculating $\mathbf{Z}$ and $\mathbf{Y}$ requires $O(nSc)$ and $O(Sc^2)$. Solving the eigenvalue decomposition problem of $(\hat{\mathbf{S}_W^l})^{-1}\hat{\mathbf{S}_B^l}$ takes $O(c^3)$ and the final solution $\hat{\mathbf{V}}^l$ is obtained by matrix multiplication, which takes $O(cSM)$. Since in our case $n > c$ and $S > c$, the time complexity for generating L LE2-based classifiers is $O(LnSc)$, which can also be written as $O(LnNN_1c)$.

We run the matlab code of our method on a PC with an Intel Core i5 3.10 GHz processor and 16GB RAM. The training time (i.e., generating $L = 1000$ classifiers) and query time per sequence are reported in Table 6.5. It is clear that LE2 is very efficient when the dimensionality is large, and we can see that Gabor + RSM(LE2) only takes about $1/10$ of Gabor + RSM(LE1) in terms of training time.

6.3.4 Performance Gain Analysis

In Chapter 6.1.3, we claim that the performance gain can be achieved by using LE1/LE2 and HDF, which can effectively suppress the ratio Γ of false votes $\sum_{i=1}^{c-1} V_{false}^i$ to true votes

V_{true}. For evaluation purpose, on a probe set with K gait sequences, we define $\hat{\Gamma}$ as

$$\hat{\Gamma} = \text{median}\left\{ \frac{\sum_{i=1}^{c-1} V_{false}^{i}}{V_{true} + \epsilon} \right\}_{k=1}^{K}, \tag{6.15}$$

which is used to reflect the general ratio of false votes to true votes over the whole probe set. We set $\epsilon = 1$ to avoid the trivial results. Based on GEI and Gabor templates, we conduct experiments using the proposed LE1/LE2 and HDF. Over the 12 probe sets (A-L) on the USF dataset, the distribution of $\hat{\Gamma}$ and the performance are reported in Fig. 6.9 and Table 6.6, from which we can observe the following:

1. Generally, by employing LE1/LE2 and HDF, $\hat{\Gamma}$ can be greatly reduced, and significant performance gain can be achieved.

2. Compared with GEI template, Gabor template tend to be more discriminant and to yield much lower $\hat{\Gamma}$, which contributes positively to the improved performance.

3. HDF can greatly reduce $\hat{\Gamma}$ and has very competitive performance in tackling the challenging covariates like walking surface (D-G) and elapsed time (K-L).

The performance gain by employing LE1/LE2 is large, and methods based on LE1/LE2 are robust to covariates such as viewpoint, shoe, and briefcase (A-C, H-J). However, the performance is less satisfactory for the challenging covariates like walking surface and elapsed time, which can significantly corrupt the gait features (i.e., the number of corrupted features m is large in Eq.(6.8)). In this case, HDF can effectively eliminate the false votes, hence yielding competitive accuracies in tackling the hard problems caused by walking surface or elapsed time.

6.3.5 Algorithms Comparison

On the USF dataset, we compare our method Gabor + RSM-HDF(LE1,LE2) with the recently published works, with the experimental results reported in Table 6.7. These algorithms in the previous works include Baseline [Sarkar et al., 2005], hidden Markov models(HMM) [Kale et al., 2004b], GEI+LDA [Han and Bhanu, 2006], synthetic GEI (Syn) +LDA [Han and Bhanu, 2006], GEI+Fusion [Han and Bhanu, 2006], CSA+DATER [Xu et al., 2006], dynamic-normalised gait recognition (DNGR) [Liu and Sarkar, 2006], MMFA [Xu et al., 2007], GTDA [Tao et al., 2007], linearisation of DLLE (DLLE/L) [Li et al., 2008], TRIMAP [Chen et al., 2010], Image-to-Class [Huang et al., 2010], CGI+Fusion [Wang et al., 2012], Gabor-PDF+LGSR [Xu et al., 2012], SRML [Lu et al., 2014], and S-BDA [Lai et al., 2014]. For the algorithms with multiple results due to different parameter settings, only the best ones are reported.

In terms of rank-1 CCRs, our method outperforms other algorithms on all the 12 probe sets. It has an average accuracy more than 10% higher than the second best method, Gabor-PDF+LGSR [Xu et al., 2012]. In terms of average rank-5 CCRs, our method also outperforms other algorithms. Our method is significantly superior than others on the challenging tasks D-G, and K-L, which are under the influences of walking surface, elapsed time and the combination of other covariates. Although these walking conditions may significantly corrupt the gait features, our proposed HDF scheme (based on LE1 and LE2) can still suppress Γ of Eq.(6.12) to some extent, leading to much higher performance than other algorithms. We notice that our method only has 42% rank-1 CCRs for probe sets K-L on the USF dataset. In these cases, elapsed time is coupled with other covariates like walking surface, clothing, and shoe, as listed in Table 6.1. These walking conditions may significantly increase the number of irrelevant features m, which would result in a lower $P(N)$ in Eq.(6.8). According to Chapter 6.1.3, a lower $P(N)$ would lead to a higher Γ in Eq.(6.12), which contribute negatively to the performance. Since extremely large intra-class variations may significantly limit the performance of unimodal biometric systems [Jain et al., 2004b], fusing soft biometrics [Reid et al., 2014] into our gait recognition method will be investigated in the future.

6.3.6 In Tackling the Clothing Challenges

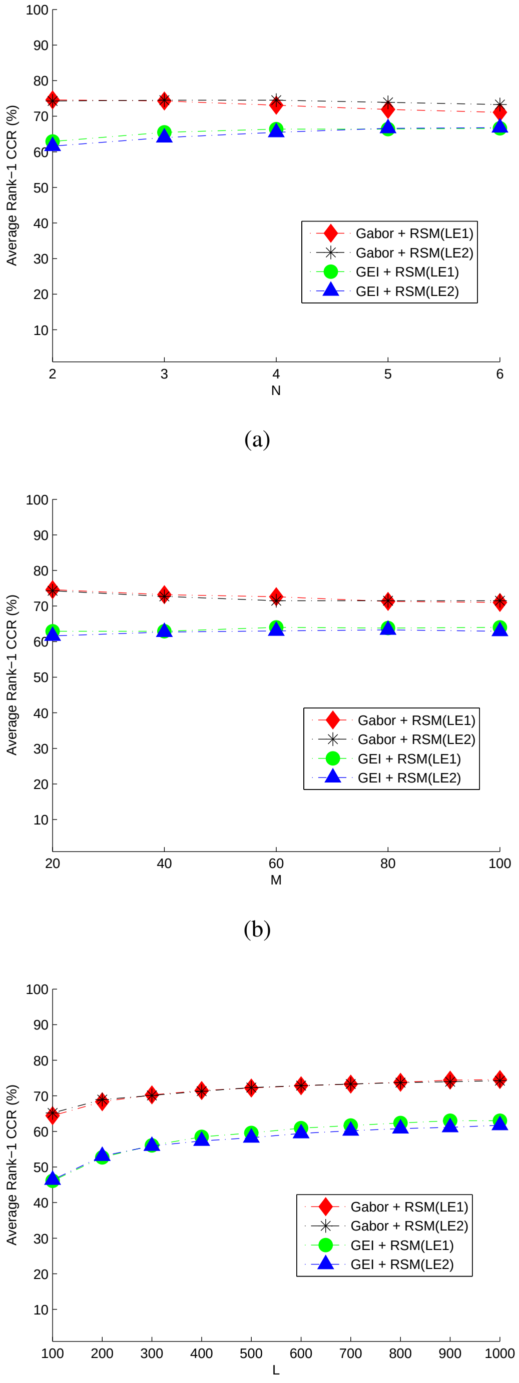
Clothing is deemed as one of the most challenging covariates [Matovski et al., 2012], since this subject-related covariate may corrupt the gait features in unpredictable locations. Compared with other covariates, there are only a few works, that have studied the effect of clothing based on a large number of clothes types. Recently, Hossain et al. [Hossain et al., 2010] built the OU-ISIR-B dataset, which consists of 32 combinations of clothes types (see Table 6.3). Based on an additional training data that covers all the possible clothes types, they proposed an adaptive part-based method [Hossain et al., 2010] to reduce the effect of different clothes types in the probes.

We evaluate our methods Gabor + RSM(LE1), Gabor + RSM(LE2), and Gabor + RSM-HDF(LE1,LE2) on this dataset. The statistics of our methods over 10 runs are reported in Table 6.8. We compare our method Gabor + RSM-HDF(LE1,LE2) with the part-based method [Hossain et al., 2010] in Table 6.9. Our method outperforms the part-based method by more than 25% in terms of rank-1 CCR. It is worth noting that different from [Hossain et al., 2010], our method does not require any additional training data which covers all the possible clothes types and can generalise well to unseen clothes types.

We also study the effect of different clothes types, and the rank-1 CCRs with respect to different clothes types are reported in Fig. 6.10. For most of the clothes types, our method can achieve more than 90% rank-1 accuracies. However, when the clothes types of the probes cover large parts of the human body, the recognition tasks become harder. Although HDF can effectively enhance the performance in these challenging tasks, the results are less satisfactory when the following 3 clothes types are encountered: 1) clothes type R, (i.e., raincoat) with a CCR of 63.3%; 2) clothes type H, (i.e., casual pants + down jacket) with a CCR of 52.1%; 3) clothes type V, (i.e., skirt + down jacket) with a CCR of 52.2%. Nevertheless, in general the results suggest that our method is robust to clothing.

6.4 Summary

In this chapter, we model the gait recognition challenges caused by various covariate factors as an unknown partial feature corruption problem and propose a classifier ensemble method to address this issue. Our method can be used as a gait recognition framework since it is insensitive to the corruption locations. To tackle the hard problems, we further extend our method by incorporating two proposed strategies: local enhancing and hybrid decision-level fusion. The extended method can effectively suppress the ratio of false votes to true votes before the majority voting. It is less sensitive to the most challenging covariates like clothing, walking surface, and elapsed time. Our method has only 3 parameters, to which the performance is not sensitive. It can be trained within minutes and perform real-time recognition in less than 1 second, which suggests that it is practical in the real-world scenarios. More details of this chapter can be found in our publications in [Guan et al., 2012a],[Guan et al., 2012b], and [Guan et al., 2015].



(a)

(b)

(c)

Figure 6.8: On the performance sensitivity to the parameters on the USF dataset: (a) N is the dimensionality of the random subspace; (b) M is the number of projection directions of LE1/LE2; (c) L is the classifier number

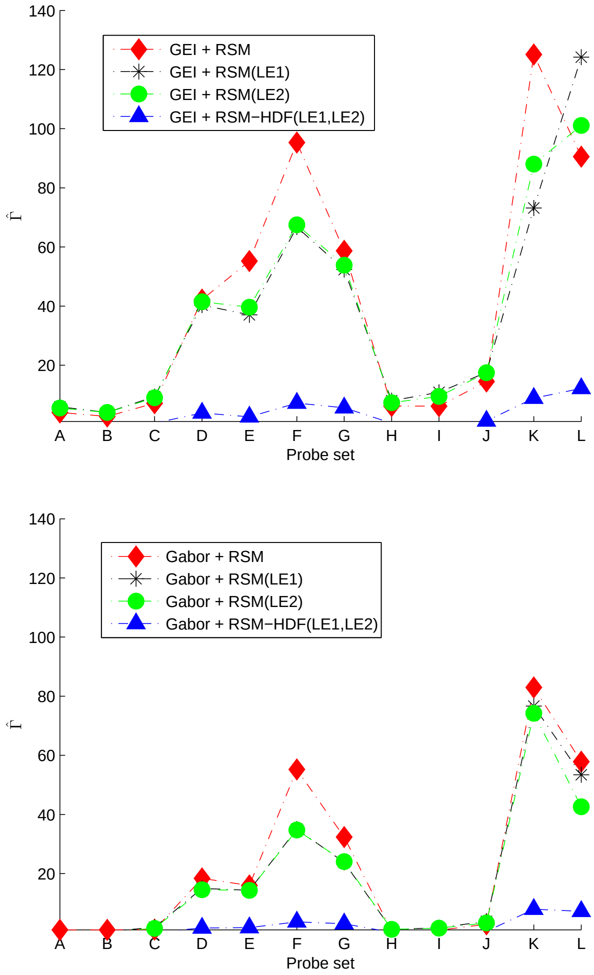


Figure 6.9: The general ratio $\hat{\Gamma}$ of between false votes to true votes distribution over the 12 probe sets on the USF dataset

Table 6.7: Algorithms comparison in terms of rank-1/rank-5 CCRs (%) on the USF dataset

Experiment	A	B	C	D	E	F	G	H	I	J	K	L	Avg.
Rank-1 CCRs													
Baseline[Sarkar et al., 2005]	73	78	48	32	22	17	17	61	57	36	3	3	40.96
HMM [Kale et al., 2004b]	89	88	68	35	28	15	21	85	80	58	17	15	53.54
GEI + LDA [Han and Bhanu, 2006]	89	87	78	36	38	20	28	62	59	59	3	6	51.00
Syn + LDA [Han and Bhanu, 2006]	84	93	67	53	45	30	34	48	57	39	21	24	51.16
GEI + Fusion [Han and Bhanu, 2006]	90	91	81	56	64	25	36	64	60	60	6	15	57.66
CSA + DATER [Xu et al., 2006]	89	93	80	44	45	25	33	80	79	60	18	21	58.51
DNGR [Liu and Sarkar, 2006]	85	89	72	57	66	46	41	83	79	52	15	24	62.81
MMFA [Xu et al., 2007]	89	94	80	44	47	25	33	85	83	60	27	21	59.90
GTDA [Tao et al., 2007]	91	93	86	32	47	21	32	95	90	68	16	19	60.58
DLLE/L[Li et al., 2008]	90	89	81	40	50	27	26	65	67	57	12	18	51.83
TRIMAP [Chen et al., 2010]	92	94	86	44	52	27	33	78	74	65	21	15	59.66
Image-to-Class [Huang et al., 2010]	93	89	81	54	52	32	34	81	78	62	12	9	61.19
Gabor-PDF + LGSR [Xu et al., 2012]	95	93	89	62	62	39	38	94	91	78	21	21	70.07
CGI + Fusion [Wang et al., 2012]	91	93	78	51	53	35	38	84	78	64	3	9	61.69
SRML [Lu et al., 2014]	93	94	85	52	52	37	40	86	85	68	18	15	66.50
SBDA [Lai et al., 2014]	93	94	85	51	50	29	36	85	83	68	18	24	61.35
Gabor + RSM-HDF(LE1,LE2)	**100**	**95**	**94**	**73**	**73**	**55**	**64**	**97**	**99**	**94**	**42**	**42**	**81.15**
Rank-5 CCRs													
Baseline [Sarkar et al., 2005]	88	93	78	66	55	42	38	85	78	62	12	15	64.54
HMM [Kale et al., 2004b]	-	-	-	-	-	-	-	-	-	-	-	-	-
GEI + LDA [Han and Bhanu, 2006]	93	93	89	65	60	42	45	88	79	80	6	9	68.70
Syn + LDA [Han and Bhanu, 2006]	93	96	93	75	71	54	53	78	82	64	33	42	72.06
GEI + Fusion [Han and Bhanu, 2006]	94	94	93	78	81	56	53	90	83	82	27	21	76.23
CSA + DATER [Xu et al., 2006]	96	96	94	74	79	53	57	93	91	83	40	36	77.86
DNGR [Liu and Sarkar, 2006]	96	94	89	85	81	68	69	96	95	79	46	39	82.05
MMFA [Xu et al., 2007]	98	98	94	76	76	57	60	95	93	84	48	39	79.90
GTDA [Tao et al., 2007]	98	**99**	97	68	68	50	56	95	**99**	84	40	40	77.58
DLLE/L[Li et al., 2008]	95	96	93	74	78	50	53	90	90	83	33	27	71.83
TRIMAP [Chen et al., 2010]	96	**99**	95	75	72	54	58	93	88	85	43	36	77.75
Image-to-Class [Huang et al., 2010]	97	98	93	81	74	59	55	94	95	83	30	33	79.17
Gabor-PDF + LGSR [Xu et al., 2012]	99	94	96	**89**	**91**	64	64	**99**	98	92	39	45	85.31
CGI + Fusion [Wang et al., 2012]	97	96	94	77	77	56	58	98	97	86	27	24	79.12
SRML [Lu et al., 2014]	-	-	-	-	-	-	-	-	-	-	-	-	-
SBDA [Lai et al., 2014]	98	98	94	74	79	57	60	95	95	84	40	40	79.93
Gabor + RSM-HDF(LE1,LE2)	**100**	98	**98**	85	84	**73**	**79**	98	**99**	**98**	**55**	**58**	**88.59**

Table 6.8: Performance statistics in terms of rank-1 CCRs (%) for 10 runs on the OU-ISIR-B dataset

-	maxima	minima	std	mean
Gabor + RSM(LE1)	89.49	86.92	0.76	87.92
Gabor + RSM(LE2)	87.97	86.80	0.36	87.52
Gabor + RSM-HDF(LE1,LE2)	91.00	90.07	0.32	90.72

Table 6.9: Algorithms comparison in terms of rank-1 CCRs (%) on the OU-ISIR-B dataset

Part-based method [Hossain et al., 2010]	Gabor + RSM-HDF(LE1,LE2)
63.9	90.7

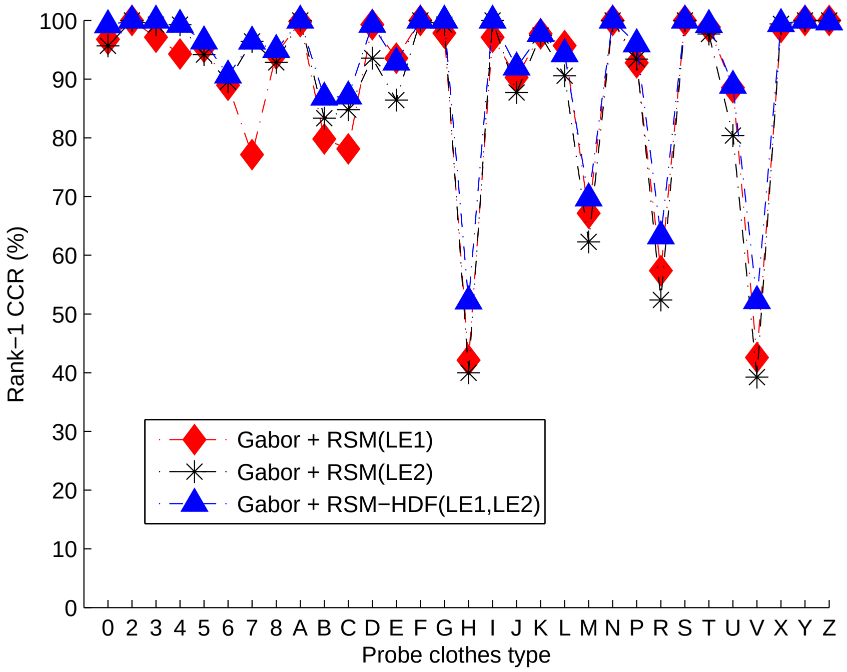


Figure 6.10: Performance distribution with respect to 31 probe clothes types on the OU-ISIR-B dataset

Chapter 7

Robust View-Invariant Gait Recognition through Unitary Projected RSM-HDF (UP-RSM-HDF)

7.1 Problem Statement and Proposed UP-RSM-HDF

The method proposed in this chapter is an extension of the RSM-HDF, which can be applied to address the cross-view gait recognition. Most of the covariates can be deemed as a partial feature corruption problem with unknown locations in the spatial domain. However, this is not applicable when the change of camera view is large, which may alter the gait appearance in a global manner, as shown in Fig. 7.2.

To address this issue, the concept of transfer learning (i.e., inductive transfer learning) [Pan and Yang, 2010] was used to learn a low-dimensional representation that is shared across different views [Makihara et al., 2006],[Kusakunniran et al., 2009b],[Kusakunniran et al., 2012b],[Zheng et al., 2011],[Bashir et al., 2010b]. They are referred to as view trans-

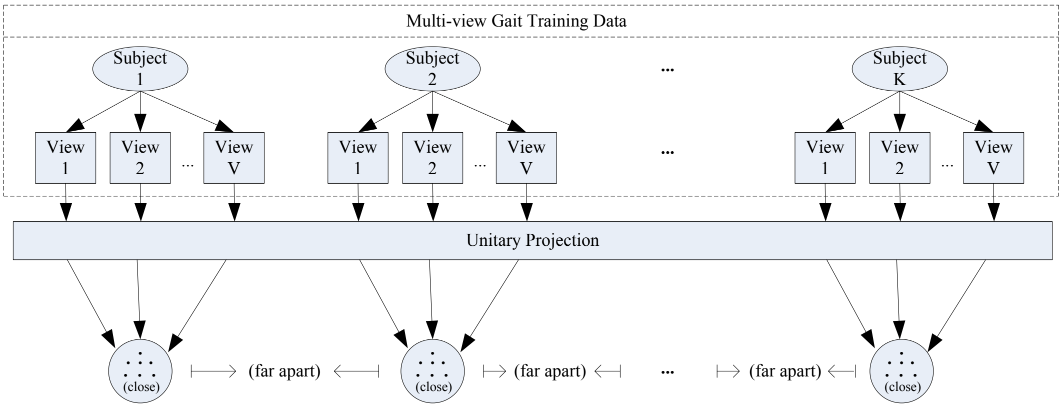


Figure 7.1: The concept of unitary projection

formation model (VTM) based methods. The learning process relies on the training data that covers the view pairs appearing in the gallery and probe. Through the learned VTMs, gaits from two different views can be projected onto the common subspace for matching. However, for multiple cross-view gait recognition tasks, there may be a large number of VTMs to be learned for different view pairs (from the gallery and probe).

We aim to form a domain, in which the effect of view changes can be deemed as a partial feature corruption problem. In this new space, samples of the same subject are closer regardless of views. In [Hu et al., 2013], the concept of unitary projection was used. Based on the general Fisher's criterion that within-class distance (in different views) should be minimised while the between-class distance should be maximised, the unitary projected (UP) matrix was trained using a multi-view training set [1], as shown in Fig. 7.1. The UP matrix can transform gait from different views into a common space, before the matching can be performed for cross-view gait recognition [Hu et al., 2013].

One of the main advantages of UP is that only one VTM is required for different cross-view gait recognition scenarios, in contrast to multiple VTMs for different view pairs [Makihara et al., 2006],[Kusakunniran et al., 2009b],[Kusakunniran et al., 2012b],[Zheng et al., 2011],[Bashir et al., 2010b]. Although UP may make the samples (from different views) of a subject closer in the new space, it may also give rise to a large amount of feature

[1] Multi-view is referred to as all the representative views appearing in the cross-view gait recognition tasks

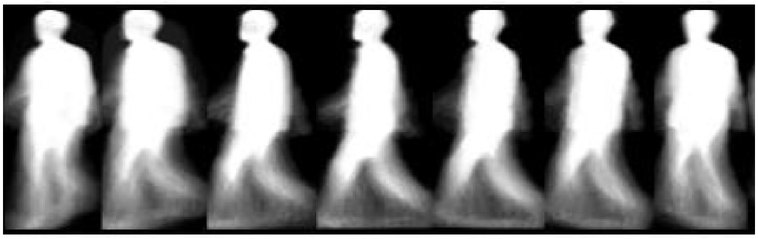


Figure 7.2: GEIs in the CASIA-B dataset [Yu et al., 2006] from 36° to 144° with an interval of 18°

distortions. These distortions may vary according to different view pairs, and can affect the performance significantly [Hu et al., 2013]. In this case, we deem these distortions as the corrupted features with unknown locations in the new space (after unitary projection) and use RSM-HDF to address this issue. We name this process UP-RSM-HDF, and its flowchart is shown in Fig. 7.3.

7.1.1 UP-RSM

In the previous chapters (i.e., Chapters 3-6), we use gallery as training set for random subspace construction, yet in this chapter we use a different setting. We split the database into 2 parts, i.e., multi-view training set (for UP matrix training), and evaluation set (including gallery for local enhancers training, and probe). Note the subjects in the multi-view training set and evaluation set are not overlapping. Based on the multi-view training set, we can train an UP matrix to generate the random subspaces. We name this process as UP-RSM. Then we use the evaluation set (including gallery and probe from various views) for cross-view gait recognition. After UP-RSM, gallery and probe from two unknown views can be projected onto multiple common (sub)spaces, before the matching can be performed.

Assuming there are K subjects in the multi-view training set $\mathcal{D}$, i.e., $\mathcal{D} = \{\mathcal{D}_j\}_{j=1}^{K}$. Let the data of the j^{th} subject $\mathcal{D}_j$ be $\mathcal{D}_j = \{\{\mathbf{I}_j^{i,v} \in \mathbb{R}^{N_1 \times N_2}\}_{i=1}^{k_j(v)}\}_{v=1}^{V}$, where v is a certain view out of V views, and for the j^{th} subject in view v, $\mathbf{I}_j^{i,v}$ is the i^{th} image (e.g., GEI or Gabor, with a size of $N_1 \times N_2$ pixels) out of a total number of $k_j(v)$ samples. In this thesis, the UP matrix is trained using 2DLDA. We summarise the corresponding UP-based random

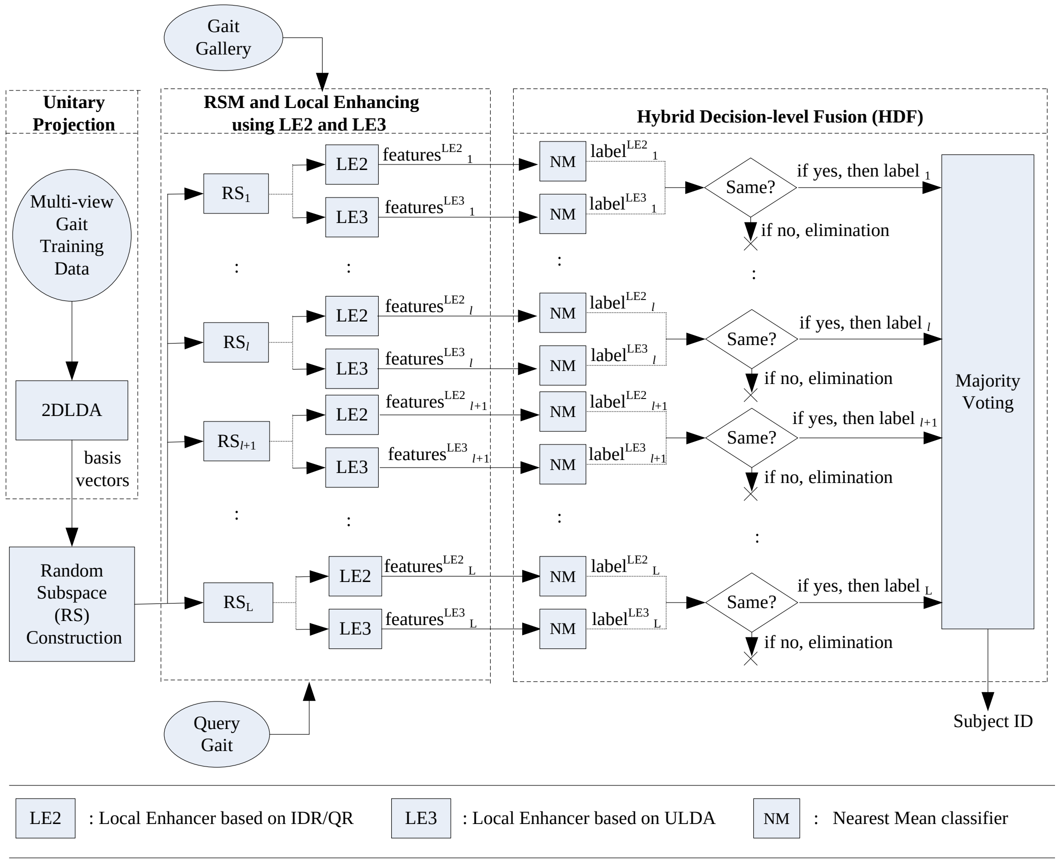


Figure 7.3: UP-RSM-HDF

subspace construction in Algorithm 7.1.

7.1.2 UP-RSM based LE2 and LE3

As discussed in the Chapter 6, we need two local enhancers for HDF. In this chapter, we use LE2 based on IDR/QR [Ye et al., 2005]. We also use uncorrelated LDA (ULDA)[Ye, 2005] for local enhancing, which is referred to as LE3. Note different from the UP matrix, LE2 and LE3 are trained using the gallery. The processes of generating UP-RSM-based LE2 and LE3 are summarised in Algorithm 7.2 and Algorithm 7.3, respectively.

For a gait sample taken from an unknown view $\mathbf{I}^v \in \mathbb{R}^{N_1 \times N_2}, v \in [1, V]$, given the random transformation matrices (output from Algorithm 7.1) $\hat{\mathbf{R}}^l \in \mathbb{R}^{N_1 \times N}, l = 1, 2, ..., L$, UP-based feature extraction can be performed by using $\mathbf{X}^l = \{\hat{\mathbf{R}}^l\}^T \mathbf{I}^v, l \in [1, L]$. After

Algorithm 7.1 UP-based Random Subspace Construction

```
Input: Multi-view training set $\mathcal{D} = \{\mathcal{D}_j = \{\{\mathbf{I}_j^{i,v} \in \mathbb{R}^{N_1 \times N_2}\}_{i=1}^{k_j(v)}\}_{v=1}^{V}\}_{j=1}^{K}$, subspace dimensionality $N$, and subspace number $L$;
Output: Random transformation matrices: $\hat{\mathbf{R}}^l \in \mathbb{R}^{N_1 \times N}, l = 1, 2, ..., L$;
  Step 1: Calculating the global centroid
  $\mathbf{m} = \sum_{j=1}^{K} \sum_{v=1}^{V} \sum_{i=1}^{k_j(v)} \mathbf{I}_j^{i,v} / (\sum_{j=1}^{K} \sum_{v=1}^{V} k_j(v))$;
  Step 2: Calculating the within class centroids $\mathbf{m}_j$ for $\mathcal{D}_j, j = 1, 2, ..., K$ such that
  $\mathbf{m}_j = \sum_{v=1}^{V} \sum_{i=1}^{k_j(v)} \mathbf{I}_j^{i,v} / (\sum_{v=1}^{V} k_j(v))$;
  Step 3: Calculating $\mathbf{S}_b = \sum_{j=1}^{K} \sum_{v=1}^{V} k_j(v)(\mathbf{m}_j - \mathbf{m})(\mathbf{m}_j - \mathbf{m})^T$;
  Step 4: Calculating $\mathbf{S}_w = \sum_{j=1}^{K} \sum_{v=1}^{V} \sum_{i=1}^{k_j(v)} (\mathbf{I}_j^{i,v} - \mathbf{m}_j)(\mathbf{I}_j^{i,v} - \mathbf{m}_j)^T$;
  Step 5: For $(\mathbf{S}_w)^{-1}\mathbf{S}_b$, calculating its $d$ leading non-zero eigenvectors $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, ..., \mathbf{t}_d] \in \mathbb{R}^{N_1 \times d}$;
  for $l = 1$ to $L$ do
    Step 6: Forming $\hat{\mathbf{R}}^l \in \mathbb{R}^{N_1 \times N}$ by randomly selecting subsets of $\mathbf{T}$ (with size $N \ll d$);
  end for
```

Algorithm 7.2 UP-RSM-based LE2

```
Input: Gallery gait data in view $v_g$, $\{\mathbf{I}_i^{v_g} \in \mathbb{R}^{N_1 \times N_2}\}_{i=1}^{n}, v_g \in [1, V]$,UP-based random transformation matrices: $\hat{\mathbf{R}}^l \in \mathbb{R}^{N_1 \times N}, l = 1, 2, ..., L$, and the number of LE2 projection directions $M$;
Output: UP-RSM-based LE2 transformation matrices: $\hat{\mathbf{V}}^l \in \mathbb{R}^{S \times M}, l = 1, 2, ..., L$, where $S = NN_2$;
  Step 1: UP-based random feature extraction on gallery $\mathbf{X}_i^l = \{\hat{\mathbf{R}}^l\}^T \mathbf{I}_i^{v_g}, i = 1, 2, ...n, l \in [1, L]$;
  Step 2: Concatenating $\mathbf{X}_i^l \in \mathbb{R}^{N \times N_2}$ to $\hat{\mathbf{X}}_i^l \in \mathbb{R}^{S \times 1}, i = 1, 2, ..., n, l \in [1, L]$;
  for $l = 1$ to $L$ do
    Steps 3-10 in Algorithm 6.1.
  end for
```

Algorithm 7.3 UP-RSM-based LE3

```
Input: Gallery gait data in view $v_g$, $\{\mathbf{I}_i^{v_g} \in \mathbb{R}^{N_1 \times N_2}\}_{i=1}^{n}, v_g \in [1, V]$,UP-based random transformation matrices: $\hat{\mathbf{R}}^l \in \mathbb{R}^{N_1 \times N}, l = 1, 2, ..., L$, and the number of LE3 projection directions $M$;
Output: UP-RSM-based LE3 transformation matrices: $\hat{\mathbf{G}}^l \in \mathbb{R}^{S \times M}, l = 1, 2, ..., L$, where $S = NN_2$;
  Step 1: UP-based random feature extraction on gallery $\mathbf{X}_i^l = \{\hat{\mathbf{R}}^l\}^T \mathbf{I}_i^{v_g}, i = 1, 2, ...n, l \in [1, L]$;
  Step 2: Concatenating $\mathbf{X}_i^l \in \mathbb{R}^{N \times N_2}$ to $\hat{\mathbf{X}}_i^l \in \mathbb{R}^{S \times 1}, i = 1, 2, ..., n, l \in [1, L]$;
  for $l = 1$ to $L$ do
    Step 3: For $\hat{\mathbf{X}}_i^l, i = 1, 2, ..., n$ corresponding to $c$ classes,
    1) calculating the global centroid $\hat{\mathbf{m}}^l$;
    2) letting $\hat{\mathbf{D}}_j^l$, $\hat{\mathbf{m}}_j^l$, and $n_j$ be the set of within-class samples, within-class centroid, and sample number for the $j^{th}$ class, respectively;
    Step 4: After setting $\mathbf{e}_j = (1, 1, ..., 1)^T \in \mathbb{R}^{n_j}$,$\mathbf{e} = (1, 1, ..., 1)^T \in \mathbb{R}^n$, computing:
    $\mathbf{H}_b^l = [\sqrt{n_1}(\hat{\mathbf{m}}_1^l - \hat{\mathbf{m}}^l), \sqrt{n_2}(\hat{\mathbf{m}}_2^l - \hat{\mathbf{m}}^l), ..., \sqrt{n_c}(\hat{\mathbf{m}}_c^l - \hat{\mathbf{m}}^l)]$,
    $\mathbf{H}_w^l = [\hat{\mathbf{D}}_1^l - \hat{\mathbf{m}}_1^l \mathbf{e}_1{}^T, \hat{\mathbf{D}}_2^l - \hat{\mathbf{m}}_2^l \mathbf{e}_2{}^T, ..., \hat{\mathbf{D}}_c^l - \hat{\mathbf{m}}_c^l \mathbf{e}_c{}^T]$;
    $\mathbf{H}_t^l = [\hat{\mathbf{X}}_i^l, \hat{\mathbf{X}}_2^l, ..., \hat{\mathbf{X}}_n^l] - \hat{\mathbf{m}}^l \mathbf{e}^T$;
    Step 5: Performing reduced SVD of $\mathbf{H}_t^l$, as $\mathbf{H}_t^l = \mathbf{U}_1 \mathbf{\Sigma}_t \mathbf{V}_1^T$, $\mathbf{U}_1 \in \mathbb{R}^{S \times t}, \mathbf{\Sigma}_t \in \mathbb{R}^{t \times t}$, where $t = rank(\mathbf{H}_t^l (\mathbf{H}_t^l)^T)$;
    Step 6: Calculating $\mathbf{B} = \mathbf{\Sigma}_t^{-1} \mathbf{U}_1^T \mathbf{H}_b^l, \mathbf{B} \in \mathbb{R}^{t \times c}$;
    Step 7: Performing SVD of $\mathbf{B}$ as $\mathbf{B} = \mathbf{P \Sigma Q}^T, \mathbf{P} \in \mathbb{R}^{t \times t}$;
    Step 8: Calculating $\mathbf{G} = \mathbf{U}_1 \mathbf{\Sigma}_t^{-1} \mathbf{P}, \mathbf{G} \in \mathbb{R}^{S \times t}$;
    Step 9: Setting $\hat{\mathbf{G}}^l = \mathbf{G}(:, 1 : M)$, (i.e., letting $\hat{\mathbf{G}}^l$ be the first $M$ columns of $\mathbf{G}$), where $M \leq t$;
  end for
```

concatenating $\mathbf{X}^l \in \mathbb{R}^{N \times N_2}$ to $\hat{\mathbf{X}}^l \in \mathbb{R}^{S \times 1}, S = NN_2, l \in [1, L]$, local enhancing can be performed.

Given UP-RSM-based LE2 transformation matrices $\hat{\mathbf{V}}^l \in \mathbb{R}^{S \times M}, l = 1, 2, ..., L$ (output from Algorithm 7.2), the new feature representation $\hat{\mathbf{x}}^l_{LE2} \in \mathbb{R}^{M \times 1}$ can be extracted by

$$\hat{\mathbf{x}}^l_{LE2} = (\hat{\mathbf{V}}^l)^T \hat{\mathbf{X}}^l, \quad l \in [1, L]. \tag{7.1}$$

Similarly, given UP-RSM-based LE3 transformation matrices $\hat{\mathbf{G}}^l \in \mathbb{R}^{S \times M}, l = 1, 2, ..., L$ (from Algorithm 7.3), the new feature representation $\hat{\mathbf{x}}^l_{LE3} \in \mathbb{R}^{M \times 1}$ can be extracted by

$$\hat{\mathbf{x}}^l_{LE3} = (\hat{\mathbf{G}}^l)^T \hat{\mathbf{X}}^l, \quad l \in [1, L]. \tag{7.2}$$

In the l^{th} subspace, if the random features (from query gait) is relevant, they should still remain discriminant after local enhancing by LE2 or LE3. Based on such discriminant features, the corresponding (LE2/LE3-based) classifier pair would output the same label (i.e., valid vote). On the other hand, in the l^{th} subspace if the random features are irrelevant (i.e., distorted features caused by UP), the classifier pair (based on LE2 and LE3) would output two random labels, which are more likely to be different. Based on the ”AND” rule, we can eliminate such false vote corresponding to this subspace. Majority voting is only performed among the valid votes, which correspond to different subsets of the relevant features for the query gait. More details about RSM-HDF can be found in Chapter 6.2.2. The flowchart of UP-RSM-HDF is demonstrated in Fig. 7.3.

7.2 Experiments

7.2.1 Dataset and Experimental Settings

The CASIA-B dataset [Yu et al., 2006] is a large multi-view database consisting of 124 subjects in the indoor environment. In this chapter, samples taken from seven views (36°, 54°, 72°, 90°, 108°, 126°, and 144°) are used to evaluate the cross-view gait recognition algorithms. For each subject in each view, there are six gait sequences walking in normal condition. Recently, Zheng et al. made the CASIA-B GEIs available [Zheng et al., 2011]. For each gait sequence, they generated one GEI with a size of 240×240 pixels. For computational efficiency, in this chapter we cut and resize these GEIs into 128×88 pixels, before convolving with the Gabor filters. Following the previous chapters, we use the downsampled Gabor features as input feature templates (with a size of 320×352 pixels).

We divide these 124 subjects into two sets, namely, training set, and evaluation set (i.e., gallery and probe). Following the works [Kusakunniran et al., 2009b], [Kusakunniran et al., 2010], [Zheng et al., 2011], [Kusakunniran et al., 2012b], [Hu et al., 2013], etc., we use 24 subjects for training while 100 subjects for evaluation. The UP matrix is trained based on multi-view training set (e.g., all the afore-mentioned seven views from $36°$ to $144°$), while the cross-view recognition is performed based on the evaluation set. Specifically, there are six samples (i.e., Gabor templates in this chapter) per subject per view for UP matrix training. During evaluation, for a subject in a view, four samples are used as gallery and the rest two are used as probe. These settings are summarised in Table 7.1.

We use rank-1 correct classification rate (CCR) to measure the performance. In Chapter 6.1, for RSM we claim that N (subspace dimensionality) should be a small number, and L (number of subspaces/classifiers) should be a large number. In the experiments in Chapter 6.3.2, we empirically set $N = 2$, and $L = 1000$. We have also found that the recognition performance is not sensitive to the number of local enhancers' projection directions M, and setting $M = 20$ makes the system accurate and computationally efficient, as shown in Chapter 6.3.2. However, in the cross-view gait recognition experiments, we found $M = 40$ can yield higher performance in terms of accuracy than $M = 20$ with a slightly higher computational cost. For cross-view gait recognition, maybe more information should be encoded for local enhancers by employing more projection directions, which contributes positively to the recognition.

With $N = 2$, $L = 1000$, and $M = 40$, we run our method (UP-RSM-HDF) 10 times, and the means and standard deviations of the CCRs for the cross-view gait recogni-

Multi-view training set	Evaluation set for cross-view gait recognition	
	100 subjects, 7 views	
24 subjects, 7 views	Gallery	Probe
6 samples per subject per view	4 samples per subject per view	2 samples per subject per view

Table 7.1: Settings of training/evaluation set

P G	36°	54°	72°	90°	108°	126°	144°
36°	99.1±0.15	97.8±0.25	81.5±1.92	58.8±1.68	57.7±1.64	81.3±1.33	84.4±1.09
54°	97.4±0.23	99.0±0.00	95.0±0.47	85.9±1.10	87.1±1.37	89.8±1.17	81.0±0.65
72°	86.5±1.24	97.8±0.81	99.9±0.20	99.4±0.39	97.6±0.42	90.1±1.15	60.9±1.36
90°	55.9±1.64	84.6±1.04	100±0.00	100±0.00	99.7±0.33	89.1±1.31	53.2±1.49
108°	65.1±2.02	88.3±0.81	98.0±0.42	99.6±0.15	99.9±0.23	97.2±0.23	84.9±1.53
126°	82.9±1.25	92.4±0.75	91.2±1.02	90.7±0.96	99.1±0.32	100±0.00	98.7±0.25
144°	84.7±1.38	85.5±0.82	63.3±2.53	64.1±2.24	87.5±0.91	97.8±0.25	99.8±0.25

Table 7.2: The rank-1 CCR (%) of cross-view gait recognition using UP-RSM-HDF(LE2,LE3); G/P denotes the views of gallery/probe.

tion are reported in Table 7.2. The lower standard deviations indicate the stability of our method, and we only report the means for the rest of this chapter.

7.2.2 Cross-view Gait Recognition

We perform cross-view gait recognition using our UP-RSM-HDF algorithm, and the results are shown in Table 7.2. We can see that for probes with view of 54° or 126°, very high performance (at least more than 80% rank-1 CCR) can be achieved against gallery in any views (out of the seven views). For other query views, when the view difference is between 54°-72°, the accuracy become lower, e.g., 53.2% rank-1 CCR for probe in 144° while gallery in 90°, 57.7% rank-1 CCR for probe in 108° while gallery in 36°, etc. On the other hand, the performance is high when the view changes are lower (e.g., less than 54°) or higher (e.g., more than 72°). It is interesting to see that the performance is high when the view changes are very large, e.g., 84.4% rank-1 CCR for probe in 144° while gallery in 36°, 85.5% rank-1 CCR for probe in 54° while gallery in 144°, etc. These view angel pairs look roughly symmetric (e.g., see Fig. 7.2 for 36° and 144°), and maybe in the feature space (after UP) they have more overlapping areas with less unknown distorted features. In this

Probe view	36°					
Gallery view	54°	72°	90°	108°	126°	144°
ViDP [Hu et al., 2013]	97	78	41	50	60	54
CCA [Bashir et al., 2010b]	**98**	72	45	35	33	37
View-rectification [Goffredo et al., 2010]	65	60	**57**	57	58	-
UP-RSM(LE2)	97	70	40	47	72	76
UP-RSM(LE3)	97	78	47	51	68	74
UP-RSM-HDF(LE2,LE3)	97	**87**	56	**65**	**83**	**85**

Table 7.3: Algorithm comparison with rank-1 CCR (%) for probe view 36°

case, RSM-HDF can effectively address this problem.

We also compare our methods (UP-RSM(LE2), UP-RSM(LE3), and UP-RSM-HDF(LE2,LE3)) with other state-of-the-art algorithms in different cross-view gait recognition scenarios. These methods are FT-SVD [Makihara et al., 2006], TSVD [Kusakunniran et al., 2009b], RBF-SVR [Kusakunniran et al., 2010], Robust VTM [Zheng et al., 2011], ROI-SR [Kusakunniran et al., 2012b], ViDP [Hu et al., 2013], Co-clustering [Kusakunniran et al., 2014], UMSLDCCA [Hu, 2014], EGG-RLTDA-A [Hu, 2013], TILT [Kusakunniran et al., 2013], CCA [Bashir et al., 2010b], View-rectification [Goffredo et al., 2010], SRML [Lu et al., 2014]. The results of most algorithms are based on the settings similar to Table 7.1, such as FT-SVD, TSVD, RBF-SVR, Robust VTM, ROI-SR, ViDP, Co-clustering, and UMSLDCCA, while the results of the rest methods are based on smaller evaluation sets, e.g., CCA used 74 subjects for training and 50 subjects for evaluation, View-rectification and SRML used 65 subjects for evaluation.

We report the results in Table 7.3-7.9, from which we can see that our methods generally outperform other methods, and the performance improvement is more substantial when the view difference is large. Significant performance gain can be achieved by the hybrid decision-level fusion of LE2 and LE3, compared with the results of LE2/LE3-based UP-RSM followed by a majority voting, which confirms the effectiveness of HDF. Motivated by this, to tackle some difficult cross-view scenarios, in the future we will investigate into new ways for fusing more (e.g., ≥ 3) local enhancers, which may eliminate the false votes in a more strict manner.

Probe view	54°					
Gallery view	36°	72°	90°	108°	126°	144°
FT-SVD [Makihara et al., 2006]	72	43	28	19	24	18
TSVD [Kusakunniran et al., 2009b]	87	81	49	31	27	19
RBF-SVR [Kusakunniran et al., 2010]	**99**	97	70	55	45	30
ROI-SR [Kusakunniran et al., 2012b]	**99**	**98**	74	58	48	34
EGG-RLTDA-A [Hu, 2013]	98	**98**	64	65	64	30
TILT [Kusakunniran et al., 2013]	-	77	68	54	56	-
ViDP [Hu et al., 2013]	97	94	**87**	**88**	72	63
Co-clustering [Kusakunniran et al., 2014]	97	95	63	53	48	34
UMSLDCCA [Hu, 2014]	95	96	72	62	54	54
CCA [Bashir et al., 2010b]	95	93	60	43	45	40
View-rectification [Goffredo et al., 2010]	57	65	62	63	63	-
UP-RSM(LE2)	98	93	74	80	86	76
UP-RSM(LE3)	97	96	75	80	88	79
UP-RSM-HDF(LE2,LE3)	98	**98**	85	**88**	**92**	**86**

Table 7.4: Algorithm comparison with rank-1 CCR (%) for probe view 54°

Probe view	72°					
Gallery view	36°	54°	90°	108°	126°	144°
TILT [Kusakunniran et al., 2013]	-	79	96	81	54	-
ViDP [Hu et al., 2013]	71	91	99	97	79	43
View-rectification [Goffredo et al., 2010]	53	72	64	70	67	-
UP-RSM(LE2)	71	94	**100**	**98**	86	51
UP-RSM(LE3)	79	94	99	95	86	50
UP-RSM-HDF(LE2,LE3)	**82**	**95**	**100**	**98**	**91**	**63**

Table 7.5: Algorithm comparison with rank-1 CCR (%) for probe view 72°

Probe view	90°					
Gallery view	36°	54°	72°	108°	126°	144°
FT-SVD [Makihara et al., 2006]	8	27	36	58	28	21
TSVD [Kusakunniran et al., 2009b]	22	52	75	79	45	26
RBF-SVR [Kusakunniran et al., 2010]	35	63	95	95	65	38
Robust VTM [Zheng et al., 2011]	-	42	86	88	50	26
ROI-SR [Kusakunniran et al., 2012b]	38	66	97	96	70	43
EGG-RLTDA-A [Hu, 2013]	56	74	98	97	84	38
TILT [Kusakunniran et al., 2013]	-	70	97	93	55	-
ViDP [Hu et al., 2013]	47	80	**100**	**100**	82	51
Co-clustering [Kusakunniran et al., 2014]	41	66	96	95	68	41
UMSLDCCA [Hu, 2014]	47	64	95	94	90	52
CCA [Bashir et al., 2010b]	35	52	93	93	78	45
View-rectification [Goffredo et al., 2010]	53	74	73	69	67	-
SRML [Lu et al., 2014]	-	68	94	96	70	-
UP-RSM(LE2)	43	79	99	99	83	47
UP-RSM(LE3)	51	81	99	99	85	55
UP-RSM-HDF(LE2,LE3)	**59**	**86**	99	**100**	**91**	**64**

Table 7.6: Algorithm comparison with rank-1 CCR (%) for probe view 90°

Probe view	108°					
Gallery view	36°	54°	72°	90°	126°	144°
TILT [Kusakunniran et al., 2013]	-	47	80	95	78	-
ViDP [Hu et al., 2013]	47	80	97	99	**99**	80
CCA [Bashir et al., 2010b]	33	45	75	95	92	73
View-rectification [Goffredo et al., 2010]	47	70	59	64	73	-
UP-RSM(LE2)	46	84	97	**100**	**99**	81
UP-RSM(LE3)	50	79	96	99	98	85
UP-RSM-HDF(LE2,LE3)	**58**	**87**	**98**	**100**	**99**	**88**

Table 7.7: Algorithm comparison with rank-1 CCR (%) for probe view 108°

Probe view	126°					
Gallery view	36°	54°	72°	90°	108°	144°
FT-SVD [Makihara et al., 2006]	7	17	16	22	55	76
TSVD [Kusakunniran et al., 2009b]	21	31	42	53	80	95
RBF-SVR [Kusakunniran et al., 2010]	26	42	57	78	98	**98**
Robust VTM [Zheng et al., 2011]	-	31	60	**89**	-	89
ROI-SR [Kusakunniran et al., 2012b]	33	46	58	80	**99**	**98**
EGG-RLTDA-A [Hu, 2013]	49	58	66	75	97	**98**
TILT [Kusakunniran et al., 2013]	-	58	53	55	77	-
ViDP [Hu et al., 2013]	58	70	81	81	98	96
Co-clustering [Kusakunniran et al., 2014]	35	49	60	78	98	**98**
UMSLDCCA [Hu, 2014]	46	57	69	75	97	97
CCA [Bashir et al., 2010b]	25	45	75	78	93	**98**
View-rectification [Goffredo et al., 2010]	45	57	60	70	68	-
UP-RSM(LE2)	71	86	84	79	97	97
UP-RSM(LE3)	72	87	86	81	96	97
UP-RSM-HDF(LE2,LE3)	**81**	**90**	**90**	**89**	97	**98**

Table 7.8: Algorithm comparison with rank-1 CCR (%) for probe view 126°

Probe view	144°					
Gallery view	36°	54°	72°	90°	108°	126°
ViDP [Hu et al., 2013]	58	69	45	42	80	98
CCA [Bashir et al., 2010b]	35	40	52	42	75	95
UP-RSM(LE2)	74	72	49	38	71	**99**
UP-RSM(LE3)	75	73	48	39	81	98
UP-RSM-HDF(LE2,LE3)	**84**	**81**	**61**	**53**	**85**	**99**

Table 7.9: Algorithm comparison with rank-1 CCR (%) for probe view 144°

7.3 Summary

In this chapter, we simply extend RSM-HDF to UP-RSM-HDF for cross-view gait recognition. Through unitary projection, we can turn cross-view gait recognition problem into a partial feature corruption problem in the new space and address it by using RSM-HDF. The performance of UP-RSM-HDF is much higher than other existing state-of-the-art algorithms, which suggests that it is an effective method for the view covariate.

Chapter 8

Conclusion and Further Works

8.1 Thesis Summary

In this thesis, we have addressed the gait recognition challenges caused by various covariates, including speed, carrying condition, elapsed time, shoe, walking surface, clothing, camera viewpoint, video quality, etc. In tackling these covariates, our framework generally outperforms other state-of-the-art algorithms significantly in all the databases evaluated in this thesis.

Covariates change the human gait appearance in different ways. For example, speed may change the appearance of human arms or legs; camera viewpoint alters the human visual appearance in a global manner; carrying condition and clothing may change the appearance of any parts of the human body (depending on what is carried/wore). We view these effect caused by covariates as a partial feature corruption problem with unknown locations in a certain domain. We claim a dynamic feature selection process could address this issue, i.e., given a query gait with unknown covariates, only the relevant gait features are dynamically chosen and used for classification. However, it is difficult to measure these relevant features in the training stage, and relevant features could become irrelevant due to feature corruption caused by unknown covariates. We use a RSM-based classifier ensemble to address this issue, the classifiers with irrelevant features can be eliminated in

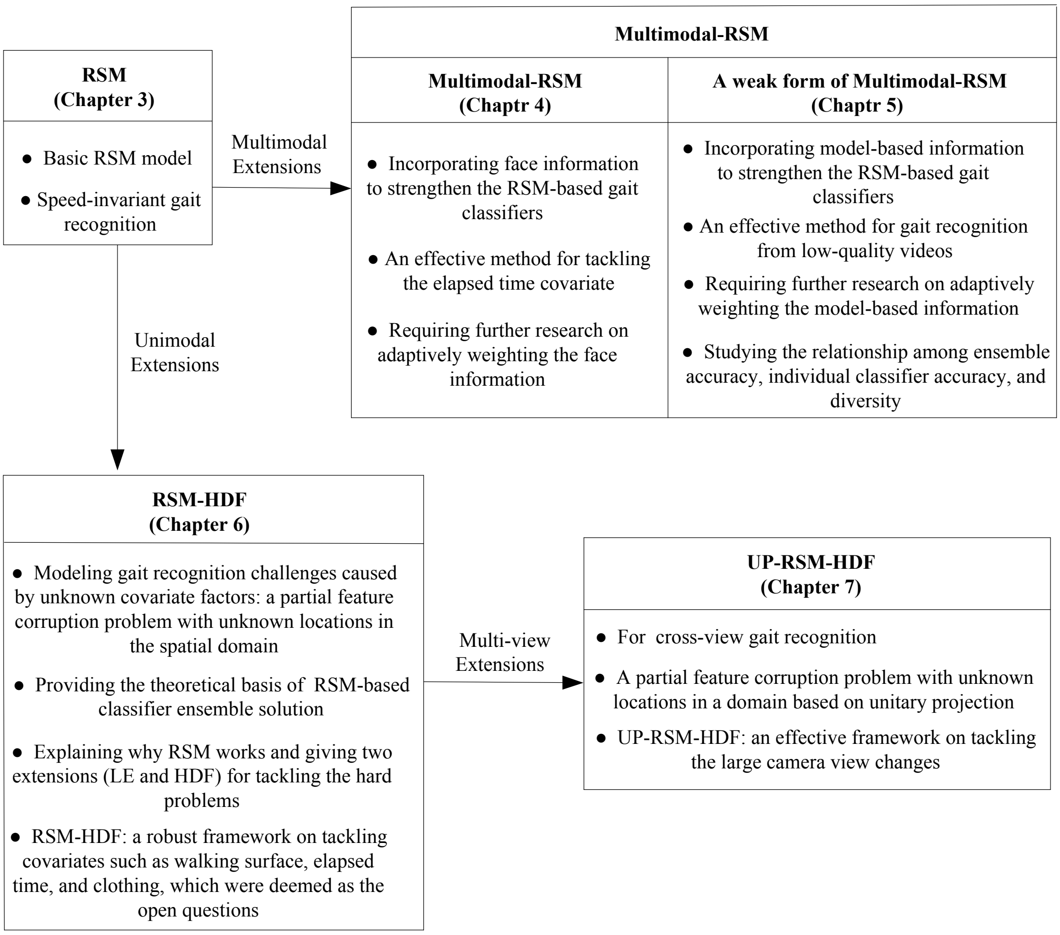

RSM
(Chapter 3)
● Basic RSM model
● Speed-invariant gait recognition
Multimodal Extensions
Multimodal-RSM
Multimodal-RSM
(Chaptr 4)
● Incorporating face information to strengthen the RSM-based gait classifiers
● An effective method for tackling the elapsed time covariate
● Requiring further research on adaptively weighting the face information
A weak form of Multimodal-RSM
(Chaptr 5)
● Incorporating model-based information to strengthen the RSM-based gait classifiers
● An effective method for gait recognition from low-quality videos
● Requiring further research on adaptively weighting the model-based information
● Studying the relationship among ensemble accuracy, individual classifier accuracy, and diversity
Unimodal Extensions
RSM-HDF
(Chapter 6)
● Modeling gait recognition challenges caused by unknown covariate factors: a partial feature corruption problem with unknown locations in the spatial domain
● Providing the theoretical basis of RSM-based classifier ensemble solution
● Explaining why RSM works and giving two extensions (LE and HDF) for tackling the hard problems
● RSM-HDF: a robust framework on tackling covariates such as walking surface, elapsed time, and clothing, which were deemed as the open questions
Multi-view Extensions
UP-RSM-HDF
(Chapter 7)
● For cross-view gait recognition
● A partial feature corruption problem with unknown locations in a domain based on unitary projection
● UP-RSM-HDF: an effective framework on tackling the large camera view changes

Figure 8.1: Summary of the thesis

the stage of decision-level fusion.We also extend RSM to multimodal-RSM for performance improvement. After providing the theoretical basis of the RSM-based classifier ensemble solution, we further propose several extensions of RSM, such as RSM-HDF, UP-RSM-HDF to address the hard problems such as gait recognition against walking surface, elapsed time, clothing, camera viewpoint, etc. We summarise the works of this thesis (i.e., Chapters 3-7) in Fig. 8.1. Detailed descriptions are given as follows.

In Chapter 3, we introduce the basic RSM model, and use it for speed-invariant gait recognition. Initially, 2DPCA is used for feature decorrelation. A large number of random subspaces are constructed, with each subspace spanned by a subset consisting of several randomly selected 2DPCA basis vectors. We further use 2DLDA-based LE1 to enhance the random features for each subspace before majority voting is applied among the corresponding classifiers for making the final classification decision. We use this basic RSM model for speed-invariant gait recognition, and it outperforms other state-of-the-art algorithms significantly. We also test the basic RSM model for runner identification solely using the walker gallery. Although it can yield some encouraging accuracies, the general experimental results suggest that several extensions should be done when facing large intra-class variations.

In Chapter 4, to address the large intra-calss variations caused by elapsed time, we extend the basic RSM model to multimodal-RSM. In RSM systems, weak classifiers with lower dimensionality tend to have better generalisation ability [Ho, 1998]. However, they may face the underfitting problem if the dimensionality is too low. We enhance the RSM-based weak classifiers by using the face information. Although face information is also less reliable due to non-uniform illumination, low resolution, it may provide some complementary information for gait. We find significant performance gain can be achieved when lower weight is assigned to the face information in the multimodal-RSM system. Multimodal-RSM produces very competitive results against elapsed time covariate, which also includes the changes of clothing, shoe, carrying condition, etc.

In Chpater 5, we instead use model-based information to enhance the RSM-based weak classifiers for gait recognition in low-quality videos. Although face information from surveillance videos is useful to some extent for multimodal-RSM, when subjects are too far away from the camera with extremely low resolution, face information may become extremely unreliable. Model-based gait information, on the other hand, may provide some body structure information from a different perspective. Since both information can be derived from the same gait video, this method is especially useful when the video footage quality is extremely low with other modalities unavailable. We also study the relationship of base classifier accuracy, classifier ensemble accuracy, and diversity among the base classifiers. We find that by incorporating the model-based information (with lower weight) into the RSM-based weak classifiers, the diversity of the classifiers, which is positively correlated to the ensemble accuracy, can be enhanced.

In Chapter 6, we model the effect of various unknown covariates as a partial feature corruption problem with unknown locations in the spatial domain. By making some assumptions in ideal cases analysis, we provide the theoretical basis of robust gait recognition using RSM-based classifier ensemble. However, in real cases, these assumptions may not hold precisely, and the performance may be affected when intra-class variations are large. We propose a criterion to address this issue. That is, in the decision-level fusion stage, for a query gait with unknown covariates, we need to $dynamically$ suppress the ratio of false votes to the true votes, before the majority voting. Two strategies are employed, i.e., LE which can increase true votes, and the proposed HDF which can decrease false votes. Extensive experiments are conducted, and performance gain is significant by employing this proposed criterion. We use this RSM-HDF framework in tackling the covariates like walking surface, clothing, and elapsed time, and its performance is much higher than other state-of-the-art algorithms.

In Chapter 7, we extend RSM-HDF to UP-RSM-HDF for cross-view gait recognition. The factor of camera viewpoint is different from the other covariates and cannot be simply deemed as a partial feature corruption problem in the spatial domain. Based on the proposed 2DLDA-based UP, we form a new space, where the same subjects are closer from different views. However, it may also give rise to a large amount of feature distortions. We then use RSM-HDF to address this issue, since in this case, we may deem these distortions as the corrupted features with unknown locations in the new space (after UP). We use this UP-RSM-HDF framework for cross-view gait recognition and it significantly outperforms other state-of-the-art algorithms. It is worth mentioning that it can perform very well even when the view changes are large, e.g., for probe view $144°$ and gallery view $36°$.

8.2 Further Works

Although we have proposed a number of approaches in this thesis for addressing some challenges in gait recognition, more work is yet to be done in order to bring the performance of gait recognition systems to a greater level. We list some possible new lines of investigations for future research.

1. For hybrid decision-level fusion (HDF), we will develop more local enhancers with lower time complexity. We will also investigate new techniques of HDF on combining more than 3 types of local enhancers.

2. For cross-view gait recognition, in the (multi-view) training process, we will explore new ways for taking advantage of the supervised local structure preserving technologies, e.g., local fisher discriminant analysis (LFDA) [Sugiyama, 2007]. For example, instead of 2DLDA-based unitary projection (UP), we will employ 2DLFDA-based UP for UP-RSM-HDF.

3. We will also study the feasibility of applying the proposed UP-RSM-HDF concept-s to other applications of heterogeneous recognition. These applications could be: cross-pose face recognition, cross-age face recognition, cross-spectral face recognition, sketch face recognition (using normal face images as gallery), low resolution image recognition (using high resolution image as gallery), cross-view action recognition, etc. For above cases, a (cross-domain) training set is required to (roughly) build the relationship of the different domains (e.g., multi-pose, multi-age, multi-spectral, sketch-face, multi-resolution, multi-view, etc.). After feature projection into the common space, feature distortions are likely to occur. Then it becomes a partial feature corruption problem with unknown locations, and the RSM-HDF concept can be used to address it. It is worth noting the cross-domain training process can be application-specific, which is worth further research.

Bibliography

N. Akae, Y. Makihara, and Y. Yagi. Gait recognition using periodic temporal super resolution for low frame-rate videos. In *Proc. Int'l Joint Conf. Biometrics (IJCB)*, pages 1–7, 2011.

N. Akae, A. Mansur, Y. Makihara, and Y. Yagi. Video from nearly still: An application to low frame-rate gait recognition. In *Proc. Conf. Computer Vision and Pattern Recognition (CVPR)*, pages 1537–1543, 2012.

M.S. Al-Huseiny, S. Mahmoodi, and M.S. Nixon. Gait learning-based regenerative model: A level set approach. In *Proc. Int'l Conf. Pattern Recognition (ICPR)*, pages 2644–2647, 2010.

G. Ariyanto and M.S. Nixon. Model-based 3d gait biometrics. In *Proc. of Int'l Joint Conf. on Biometrics (IJCB)*, pages 1–7, 2011.

D. Barrett. One surveillance camera for every 11 people in britain, says cctv survey. http://www.telegraph.co.uk/technology/10172298/One-surveillance-camera-for-every-11-people-in-Britain-says-CCTV-survey.html, 2013.

K. Bashir, T. Xiang, and S. Gong. Gait recognition using gait entropy image. In *Proc. of Int'l Conf. on Crime Detection and Prevention*, pages 1–6, 2009.

K. Bashir, T. Xiang, and S. Gong. Gait recognition without subject cooperation. *Pattern Recognition Letters (PRL)*, 31(13):2052–2060, 2010a.

K. Bashir, T. Xiang, and S. Gong. Cross-view gait recognition using correlation strength. In *Proc. of the British Machine Vision Conf. (BMVC)*, pages 1–11, 2010b.

A.F. Bobick and A.Y. Johnson. Gait recognition using static, activity-specific parameters. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, pages I–423–I–430, 2001.

I. Bouchrika, M. Goffredo, J. Carter, and M. S. Nixon. On using gait in forensic biometrics. *Journal of Forensic Sciences*, 56(4):882–889, 2011.

N.V. Chawla and K.W. Bowyer. Random subspaces and subsampling for 2-d face recognition. In *Proc. Conf. Computer Vision and Pattern Recognition (CVPR)*, pages 582–589, 2005.

C. Chen, J. Zhang, and R. Fleischer. Distance approximating dimension reduction of riemannian manifolds. *IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics (T-SMC-B)*, 40(1):208–217, 2010.

D. Cunado, M.S. Nixon, and J.N. Carter. Using gait as a biometric, via phase-weighted magnitude spectra. In *Proc. of Int'l Conf. on Audio- and Video-Based Biometric Person Authentication*, pages 95–102, March 1997.

D. Cunado, M. S. Nixon, and J. N. Carter. Automatic extraction and description of human gait models for recognition purposes. *Computer Vision and Image Understanding (CVIU)*, 90(1):1 – 41, 2003.

F. Dadashi, B.N. Araabi, and H. Soltanian-Zadeh. Gait recognition using wavelet packet silhouette representation and transductive support vector machines. In *Proc. of Int'l Congress on Image and Signal Processing*, pages 1–5, Oct. 2009.

B. DeCann and A. Ross. Gait curves for human recognition, backpack detection, and silhouette correction in a nighttime environment. In *Proc. of SPIE conf. on Biometric Technology for Human Identification*, pages 76670Q1–76670Q–13, 2010.

X. Geng, L. Wang, M. Li, Q. Wu, and K. Smith-Miles. Distance-driven fusion of gait and face for human identification in video. In *Proc. of Image and Vision Computing New Zealand*, pages 19–24, 2007.

M. Goffredo, I. Bouchrika, J.N. Carter, and M.S. Nixon. Self-calibrating view-invariant gait biometrics. *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics (T-SMC-B)*, 40(4):997–1008, Aug 2010.

C. M. Grinstead and J. L. Snell. *Introduction to Probability*. American Mathematical Society, 1997.

Y. Guan and C.-T. Li. A robust speed-invariant gait recognition system for walker and runner identification. In *Proc. of IAPR Int'l Conf. on Biometrics (ICB)*, pages 1–8, 2013.

Y. Guan, C.-T. Li, and Y. Hu. Random subspace method for gait recognition. In *Proc. of IEEE Int'l Conf. on Multimedia and Expo Workshops*, pages 284–289, July 2012a.

Y. Guan, C.-T. Li, and Y. Hu. Robust clothing-invariant gait recognition. In *Proc. of Int'l Conf. on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, pages 321–324, July 2012b.

Y. Guan, C.-T. Li, and S.D. Choudhury. Robust gait recognition from extremely low frame-rate videos. In *Proc. Int'l Workshop on Biometrics and Forensics*, pages 1–4, 2013a.

Y. Guan, X. Wei, F. Roli C.-T. Li, G. L. Marcialis, and M. Tistarelli. Combining gait and face for tackling the elapsed time challenges. In *Proc. Int'l Conf. Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–8, 2013b.

Y. Guan, Y. Sun, C.-T. Li, and M. Tistarelli. Human gait identification from extremely low-quality videos: an enhanced classifier ensemble method. *IET Biometrics*, 3(2):84–93, 2014.

Y. Guan, C.-T. Li, and F. Roli. On reducing the effect of covariate factors in gait recognition:

a classifier ensemble method. *IEEE Trans. Pattern Analysis and Machine Intelligence (T-PAMI)*, 37(99), 2015.

J. Han and B. Bhanu. Individual recognition using gait energy image. *IEEE Trans. on Pattern Analysis and Machine Intelligence (T-PAMI)*, 28(2):316–322, Feb. 2006.

T. K. Ho. The random subspace method for constructing decision forests. *IEEE Trans. on Pattern Analysis and Machine Intelligence (T-PAMI)*, 20(8):832–844, Aug. 1998.

M. Hofmann, S.M. Schmidt, A. N. Rajagopalan, and G. Rigoll. Combined face and gait recognition using alpha matte preprocessing. In *Proc. of IAPR Int'l Conf. on Biometrics (ICB)*, pages 390–395, 2012.

M. Hofmann, J. Geiger, S. Bachmann, B. Schuller, and G. Rigoll. The tum gait from audio, image and depth (gaid) database: Multimodal recognition of subjects and traits. *Journal of Visual Communication and Image Representation*, 25(1):195 – 206, 2014.

M.A. Hossain, Y. Makihara, J. Wang, and Y. Yagi. Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control. *Pattern Recognition (PR)*, 43(6):2281–2291, 2010.

H. Hu. Enhanced gabor feature based classification using a regularized locally tensor discriminant model for multiview gait recognition. *IEEE Trans. on Circuits and Systems for Video Technology (T-CSVT)*, 23(7):1274–1286, 2013.

H. Hu. Multiview gait recognition based on patch distribution features and uncorrelated multilinear sparse local discriminant canonical correlation analysis. *IEEE Trans. on Circuits and Systems for Video Technology (T-CSVT)*, 24(4):617–630, April 2014.

M. Hu, Y. Wang, Z. Zhang, J.J. Little, and D. Huang. View-invariant discriminative projection for multi-view gait-based human identification. *IEEE Trans. on Information Forensics and Security (T-IFS)*, 8(12):2034–2045, Dec 2013.

Y. Huang, D. Xu, and T. Cham. Face and human gait recognition using image-to-class distance. *IEEE Trans. Circuits and Systems for Video Technology (T-CSVT)*, 20(3):431–438, 2010.

A. Iosifidis, A. Tefas, and I. Pitas. Activity-based person identification using fuzzy representation and discriminant learning. *IEEE Trans. on Information Forensics and Security (T-IFS)*, 7(2):530–542, April 2012.

H. Iwama, M. Okumura, Y. Makihara, and Y. Yagi. The ou-isir gait database comprising the large population dataset and performance evaluation of gait recognition. *IEEE Trans. Information Forensics and Security (T-IFS)*, 7(5):1511–1521, 2012.

A.K. Jain, S.C. Dass, and K. Nandakumar. Soft biometric traits for personal recognition systems. In *Proc. of the Int'l Conf. on Biometric Authentication*, pages 731–738, 2004a.

A.K. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Trans. on Circuits and Systems for Video Technology (T-CSVT)*, 14(1):4–20, Jan. 2004b.

A.K. Jain, K. Nandakumar, and A. Ross. Score normalization in multimodal biometric systems. *Pattern Recognition (PR)*, 38(12):2270–2285, 2005.

A. Kale, A.K. Roychowdhury, and R. Chellappa. Fusion of gait and face for human identification. In *Proc. of IEEE Int'l Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 5, pages V–901–4, May 2004a.

A Kale, A Sundaresan, A N. Rajagopalan, N.P. Cuntoor, AK. Roy-Chowdhury, V. Kruger, and R. Chellappa. Identification of humans using gait. *IEEE Trans. on Image Processing (T-IP)*, 13:1163–1173, 2004b.

J. Kittler, M. Hatef, R. P.W. Duin, and J. Matas. On combining classifiers. *IEEE Trans. Pattern Analysis and Machine Intelligence (T-PAMI)*, 20(3):226–239, 1998.

E. Kleinbery. An overtraining-resistant stochastic modeling method for pattern recognition. *Annals of Statistics*, 4(6):2319–2349, 1996.

L. Kuncheva, J. Rodriguez, C. Plumpton, D. Linden, and S. Johnston. Random subspace ensembles for fmri classification. *IEEE Trans. Medical Imaging (T-MI)*, 29(2):531–542, 2010.

L. I. Kuncheva and C. J. Whitaker. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Mach. Learn.*, 51:181–207, 2003.

W. Kusakunniran, Q. Wu, H. Li, and J. Zhang. Automatic gait recognition using weighted binary pattern on video. In *Proc. of IEEE Int'l Conf. on Advanced Video and Signal Based Surveillance*, pages 49–54, Sept. 2009a.

W. Kusakunniran, Q. Wu, H. Li, and J. Zhang. Multiple views gait recognition using view transformation model based on optimized gait energy image. In *Proc. of IEEE Int'l Conf. on of Computer Vision Workshops*, pages 1058–1064, 2009b.

W. Kusakunniran, Q. Wu, J. Zhang, and H. Li. Support vector regression for multi-view gait recognition based on local motion feature selection. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, pages 974–981, 2010.

W. Kusakunniran, Q. Wu, J. Zhang, and H. Li. Speed-invariant gait recognition based on procrustes shape analysis using higher-order shape configuration. In *Proc. of IEEE Int'l Conf. on Image Processing (ICIP)*, pages 545–548, Sept. 2011.

W. Kusakunniran, Q. Wu, J. Zhang, and H. Li. Gait recognition across various walking speeds using higher order shape configuration based on a differential composition model. *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics (T-SMC-B)*, 42(6): 1654–1668, Dec. 2012a.

W. Kusakunniran, Q. Wu, J. Zhang, and H. Li. Gait recognition under various viewing angles based on correlated motion regression. *IEEE Trans. on Circuits and Systems for Video Technology (T-CSVT)*, 22(6):966–980, 2012b.

W. Kusakunniran, Q. Wu, J. Zhang, Y. Ma, and H. Li. A new view-invariant feature for cross-view gait recognition. *IEEE Trans. on Information Forensics and Security (T-IFS)*, 8(10):1642–1653, Oct 2013.

W. Kusakunniran, Q. Wu, J. Zhang, H. Li, and L. Wang. Recognizing gaits across views through correlated motion co-clustering. *IEEE Trans. on Image Processing (T-IP)*, 23 (2):696–709, Feb 2014.

Z. Lai, Y. Xu, Z. Jin, and D. Zhang. Human gait recognition via sparse discriminant projection learning. *IEEE Trans. Circuits and Systems for Video Technology (T-CSVT)*, 2014.

T. H. W. Lam, K. H. Cheung, and J. N. K. Liu. Gait flow image: A silhouette-based gait representation for human identification. *Pattern Recognition (PR)*, 44(4):973–987, 2011.

P.K. Larsen, E.B. Simonsen, and N. Lynnerup. Gait analysis in forensic medicine. *Journal of Forensic Sciences*, 53:1149–1153, 2008.

L. Lee and W. E L Grimson. Gait analysis for recognition and classification. In *Proc. of IEEE Int'l Conf. on Automatic Face and Gesture Recognition (FG)*, pages 148–155, 2002.

M. Li and B. Yuan. 2d-lda: A statistical linear discriminant analysis for image matrix. *Pattern Recognition Letters*, 26:527–532, 2005.

X. Li, S. Lin, S. Yan, and D. Xu. Discriminant locally linear embedding with high-order tensor data. *IEEE Trans Systems, Man, and Cybernetics, Part B: Cybernetics (T-SMC-B)*, 38(2):342–352, 2008.

C. Liu. Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance. *IEEE Trans. on Pattern Analysis and Machine Intelligence (T-PAMI)*, 28(5):725–737, 2006.

Y. Liu, J. Zhang, C. Wang, and L. Wang. Multiple hog templates for gait recognition. In *Proc. of Int'l Conf. Pattern Recognition (ICPR)*, pages 2930–2933, 2012.

Z. Liu and S. Sarkar. Simplest representation yet for gait recognition: averaged silhouette. In *Proc. of the Int'l Conf. on Pattern Recognition (ICPR)*, volume 4, pages 211–214, 2004.

Z. Liu and S. Sarkar. Improved gait recognition by gait dynamics normalization. *IEEE Trans. on Pattern Analysis and Machine Intelligence (T-PAMI)*, 28(6):863–876, June 2006.

Z. Liu and S. Sarkar. Outdoor recognition at a distance by fusing gait and face. *Image Vision Computing*, 25(6):817–832, 2007.

J. Lu, G. Wang, and P. Moulin. Human identity and gender recognition from gait sequences with arbitrary walking directions. *IEEE Trans. on Information Forensics and Security (T-IFS)*, 9(1):51–61, 2014.

S. Maji and A.C. Berg. Max-margin additive classifiers for detection. In *Proc. Int'l Conf. Computer Vision (ICCV)*, pages 40–47, 2009.

Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi. Gait recognition using a view transformation model in the frequency domain. In *Proc. of European Conf. on Computer Vision (ECCV)*, volume 3953, pages 151–163, 2006.

Y. Makihara, A. Mori, and Y. Yagi. Temporal super resolution from a single quasi-periodic image sequence based on phase registration. In *Proc. Asian Conf. Computer Vision (ACCV)*, pages 107–120, 2011.

Y. Makihara, H. Mannami, A. Tsuji, M.A. Hossain, K. Sugiura, A. Mori, and Y. Yagi. The ou-isir gait database comprising the treadmill dataset. *IPSJ Trans. on Computer Vision and Applications*, 4:53–62, 2012.

J. Marin, D. Vazquez, A. Lopez, J. Amores, and L. Kuncheva. Occlusion handling via random subspace classifiers for human detection. *, IEEE Trans. Cybernetics*, 44(3):342–354, 2014.

D.S. Matovski, M.S. Nixon, S. Mahmoodi, and J.N. Carter. The effect of time on gait recognition performance. *IEEE Trans. on Information Forensics and Security (T-IFS)*, 7 (2):543–552, 2012.

A. Mori, Y. Makihara, and Y. Yagi. Gait recognition using period-based phase synchronization for low frame-rate videos. In *Proc. Int'l Conf. Pattern Recognition (ICPR)*, pages 2194–2197, 2010.

S.A Niyogi and E.H. Adelson. Analyzing and recognizing walking figures in xyt. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, pages 469–474, 1994.

S. Pan and Q. Yang. A survey on transfer learning. *IEEE Trans. Knowledge and Data Engineering (T-KDE)*, 22(10):1345–1359, 2010.

A. Rattani, D.R. Kisku, M. Bicego, and M. Tistarelli. Feature level fusion of face and fingerprint biometrics. In *Proc. of Int'l Conf. Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–6, Sept 2007.

D. Reid, M. S. Nixon, and S. Stevenage. Soft biometrics; human identification using comparative descriptions. *IEEE Trans. Pattern Analysis and Machine Intelligence (T-PAMI)*, 36(6):1216–1228, June 2014.

S. Sarkar, P.J. Phillips, Z. Liu, I.R. Vega, P. Grother, and E. Ortiz. The humanid gait challenge problem: data sets, performance, and analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence (T-PAMI)*, 27(2):162–177, 2005.

G. Shakhnarovich and T. Darrell. On probabilistic combination of face and gait cues for identification. In *Proc. of Int'l Conf. on Automatic Face and Gesture Recognition (FG)*, pages 169–174, May 2002.

D. B. Skalak. The sources of increased accuracy for two proposed boosting algorithms. In *Proc. American Association for Arti Intelligence Workshop*, pages 120–125, 1996.

M. Skurichina and R.P.W. Duin. Bagging, boosting and the random subspace method for linear classifiers. *Pattern Analysis and Applications*, 5(2):121–135, 2002.

M. Sugiyama. Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. *J. Mach. Learn. Res.*, 8:1027–1061, 2007.

D. Tan, K. Huang, S. Yu, and T. Tan. Efficient night gait recognition based on template matching. In *Proc. of Int'l Conf. on Pattern Recognition (ICPR)*, volume 3, pages 1000 –1003, 2006.

D. Tan, K. Huang, S. Yu, and T. Tan. Uniprojective features for gait recognition. In *Proc. of Int'l conf. on Advances in Biometrics (ICB)*, pages 673–682, 2007a.

D. Tan, K. Huang, S. Yu, and T. Tan. Recognizing night walkers based on one pseudoshape representation of gait. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, June 2007b.

D. Tan, K. Huang, S. Yu, and T. Tan. Orthogonal diagonal projections for gait recognition. In *Proc. of IEEE Int'l Conf. on Image Processing (ICIP)*, volume 1, pages 337–340, Oct. 2007c.

D. Tan, S. Yu, K. Huang, and T. Tan. Walker recognition without gait cycle estimation. In *Proc. of Int'l conf. on Biometrics (ICB)*, pages 222–231, 2007d.

R. Tanawongsuwan and A. Bobick. Modelling the effects of walking speed on appearance-based gait recognition. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 783–790, 2004.

D. Tao, X. Li, X. Wu, and S.J. Maybank. General tensor discriminant analysis and gabor features for gait recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence (T-PAMI)*, 29(10):1700–1715, Oct. 2007.

A. Tsuji, Y. Makihara, and Y. Yagi. Silhouette transformation based on walking speed for gait identification. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 717–722, June 2010.

G.V. Veres, L. Gordon, J.N. Carter, and M.S. Nixon. What image information is important in silhouette-based gait recognition? In *Proc. Conf. Computer Vision and Pattern Recognition (CVPR)*, pages 776–782, 2004.

D.K. Wagg and M.S. Nixon. On automated model-based extraction and analysis of gait. In *Proc. of Int'l Conf. on Automatic Face and Gesture Recognition (FG)*, pages 11–16, 2004.

C. Walck. *Handbook on Statistical Distributions for Experimentalists*. University of Stockholm Press, 2007.

C. Wang, J. Zhang, L. Wang, J. Pu, and X. Yuan. Human identification using temporal information preserving gait template. *IEEE Trans. Pattern Analysis and Machine Intelligence (T-PAMI)*, 34(11):2164–2176, 2012.

L. Wang, T. Tan, H. Ning, and W. Hu. Silhouette analysis-based gait recognition for human identification. *IEEE Trans. on Pattern Analysis and Machine Intelligence (T-PAMI)*, 25 (12):1505–1518, Dec 2003a.

X. Wang and X. Tang. Random sampling lda for face recognition. In *Proc. Int'l Conf. Computer Vision and Pattern Recognition (CVPR)*, pages 259–265, 2004.

X. Wang and X. Tang. Random sampling for subspace face recognition. *Int'l Journal of Computer Vision (IJCV)*, 70(1):91–104, 2006.

Y. Wang, T. Tan, and A. K. Jain. Combining face and iris biometrics for identity verification. In *Proc. of Int'l Conf. on Audio- and video-based biometric person authentication*, pages 805–813, 2003b.

J. Wright, A.Y. Yang, A Ganesh, S.S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence (T-PAMI)*, 31 (2):210–227, 2009.

D. Xu, S. Yan, D. Tao, L. Zhang, X. Li, and H.J. Zhang. Human gait recognition with matrix representation. *IEEE Trans. Circuits and Systems for Video Technology (T-CSVT)*, 16(7): 896–903, 2006.

D. Xu, S. Yan, D. Tao, S. Lin, and H. Zhang. Marginal fisher analysis and its variants for human gait recognition and content-based image retrieval. *IEEE Trans. Image Processing (T-IP)*, 16(11):2811–2821, 2007.

D. Xu, Y. Huang, Z. Zeng, and X. Xu. Human gait recognition using patch distribution feature and locality-constrained group sparse representation. *IEEE Trans. on Image Processing (T-IP)*, 21(1):316–326, Jan. 2012.

C. Yam, M. S. Nixon, and J. N. Carter. Extended model-based automatic gait recognition of walking and running. In *Proc. of Int'l Conf. on Audio- and Video-Based Biometric Person Authentication*, pages 278–283, June 2001.

C. Yam, M. S. Nixon, and J. N. Carter. On the relationship of human walking and running: automatic person identification by gait. In *Proc. of Int'l Conf. on Pattern Recognition (ICPR)*, volume 1, pages 287–290, 2002a.

C. Yam, M. S. Nixon, and J. N. Carter. Gait recognition by walking and running: a model-based approach. In *Proc. of Asian Conf. on Computer Vision (ACCV)*, pages 1–6, Jan. 2002b.

C. Yam, M. S. Nixon, and J. N. Carter. Automated person recognition by walking and running via model-based approaches. *Pattern Recognition (PR)*, 37(5):1057–1072, 2004.

J. Yang, D. Zhang, A.F. Frangi, and J. Yang. Two-dimensional pca: a new approach to appearance-based face representation and recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence (T-PAMI)*, 26(1):131–137, Jan. 2004.

J. Ye. Characterization of a family of algorithms for generalized discriminant analysis on undersampled problems. *J. Mach. Learn. Res.*, 6:483–502, 2005.

J. Ye, Q. Li, H. Xiong, H. Park, R. Janardan, and V. Kumar. Idr/qr: An incremental dimension reduction algorithm via qr decomposition. *IEEE Trans. Knowledge and Data Engineering (T-KDE)*, 17(9):1208–1222, 2005.

S. Yu, D. Tan, and T. Tan. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In *Proc. of Int'l Conf. on Pattern Recognition (ICPR)*, volume 4, pages 441–444, 2006.

E. Zhang, Y. Zhao, and W. Xiong. Active energy image plus 2dlpp for gait recognition. *Signal Processing*, 90(7):2295 – 2302, 2010.

G. Zhao, G. Liu, H. Li, and M. Pietikainen. 3d gait recognition using multiple cameras. In *Proc. of Int'l Conf. on Automatic Face and Gesture Recognition (FG)*, pages 529–534, 2006.

S. Zheng, J. Zhang, K. Huang, R. He, and T. Tan. Robust view transformation model for gait recognition. In *Proc. of Int'l Conf. on Image Processing (ICIP)*, pages 2073–2076, Sept 2011.

X. Zhou and B. Bhanu. Integrating face and gait for human recognition at a distance in video. *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics (T-SMC-B)*, 37(5):1119–1137, 2007.

X. Zhou and B. Bhanu. Feature fusion of side face and gait for video-based human identification. *Pattern Recognition (PR)*, 41(3):778–795, March 2008.