Human gait analysis and recognition

Liu, Nini

2012

Liu, N. (2012). Human gait analysis and recognition. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/50745>

<https://doi.org/10.32657/10356/50745>

Human Gait Analysis and Recognition

Liu Nini

School of Electrical & Electronic Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy

2012

Acknowledgments

First of all, I would like to thank my supervisor, Associate Professor Tan Yap-Peng, for his valuable instructions on my research and continuous encouragement through the time of my study towards the Doctor of Philosophy degree. His knowledge, wisdom, kindness and spirit have benefited me greatly.

My sincere thanks also go to my fellow colleague Lu Jiwen for his advice and encouragement.

Further, I would also like to thank Prof. Yao Jianchao, Doctor Chen Zhenzhong, my colleagues Yang Gao and Li Maodong for their help in my research work. I am also thankful to the technicians both in the Media Technology Lab and the Information System Research Lab for their kind assistance, support and understanding.

Finally, I am very grateful to my family for their constant love, encouragement and support. I would like to thank my boyfriend, Tian Yushuang, whose love and support are always the greatest inspiration for me. Appreciation also goes to all my friends who have given me the strength to go on.

Contents

Acknowledgments	i
List of Abbreviations	ix
List of Figures	xi
List of Tables	xvi
1 Introduction	1
1.1 Motivation	1
1.2 The Challenges	2
1.3 Contributions	4
1.4 Organization	6
2 Background and Literature Survey	10
2.1 Generic Gait Recognition System	11
2.1.1 Background Subtraction	12
2.1.2 Databases	15
2.2 Gait Feature Extraction	21
2.2.1 Appearance-based Descriptors	22
2.2.2 Model-based Descriptors	33
2.3 Gait Recognition Across Varying Views	36
2.3.1 Multi-view Gait Recognition	36
2.3.2 Cross-view Gait Recognition	38

2.4	Gait Recognition Across Varying Walking Conditions	41
2.4.1	Walking Speed Variation	41
2.4.2	Walking with Carrying or Wearing Condition Variation	43
2.5	Conclusion	44
3	Gait Recognition Based on Dynamic Feature Analysis	45
3.1	Introduction	45
3.2	Spatial Dynamic Feature Analysis Based on GEI	48
3.2.1	Discriminant Feature Extraction	48
3.2.2	Dynamic Feature Selection	50
3.2.3	Spatial Dynamic Feature Generation	52
3.3	Spatial-Temporal Dynamic Feature Analysis	53
3.3.1	Temporal Information by Color	54
3.3.2	Spatial-Temporal Dynamic Feature Generation	55
3.4	Feature Space Dimension Reduction	57
3.5	Classification	58
3.6	Experiments	58
3.6.1	Database	58
3.6.2	Parameter Optimization	59
3.6.3	Results on CASIA-B Database	59
3.6.4	Results on USF Database	62
3.6.5	Further Discussion	65
3.6.6	Summary	66
3.7	Conclusion	67
4	Gait Recognition Across Varying Views	68
4.1	Introduction	68

4.2	Gait Feature Descriptor	71
4.3	Joint Subspace Learning (JSL)	73
4.3.1	Motivation	73
4.3.2	Training	74
4.3.3	Recognition	77
4.3.4	Patch-based JSL	78
4.4	Experimental Results	79
4.4.1	Database	79
4.4.2	Experimental Settings & Parameter Tuning	79
4.4.3	Gait Feature Comparison	80
4.4.4	View-Invariant Gait Recognition Results	81
4.5	Conclusion	84
5	View Recognition of Human Gait Sequences in Videos	88
5.1	Introduction	88
5.2	Motivation	90
5.3	Proposed Approach	92
5.3.1	Gait Sequence Representation	94
5.3.2	Adaptive Discriminant Analysis with Enhanced Multiple Kernel Learning (ADA-EMKL)	95
5.3.3	View Recognition with ADA-EMKL	99
5.4	Experiments	100
5.4.1	Database	100
5.4.2	Experimental Settings	101
5.4.3	Toy Example Illustration	101
5.4.4	Results on Normal Gait Sequences	102
5.4.5	Results on Gait Sequences Carrying a Bag	105

5.4.6	Results on Gait Sequences Wearing a Coat	106
5.4.7	Application of View Recognition from Human Gait Sequences .	107
5.4.8	Summary	109
5.5	Conclusion	109
6	Gait Recognition Across Multiple Varying Walking Conditions	111
6.1	Introduction	111
6.2	Proposed Approach	115
6.2.1	Feature Representation	116
6.2.2	Multiview Subspace Representation (MSR)	116
6.2.3	Marginal Canonical Correlation Analysis (MCCA)	119
6.2.4	Classification	121
6.3	Experimental Results	122
6.3.1	Databases	122
6.3.2	Experiments and Results	122
6.3.3	Discussions	130
6.4	Conclusions	131
7	Conclusions and Future Work	133
7.1	Conclusions	133
7.1.1	Dynamic Feature Analysis in Gait Recognition	133
7.1.2	Gait Recognition Across Varying Views	134
7.1.3	Gait Recognition Across Varying Walking Conditions	135
7.2	Future Work	136
7.2.1	Fully Automatic View-Invariant Gait Recognition System	136
7.2.2	Soft Biometrics From Gait	137
7.2.3	Generic Database in Gait Analysis	139
7.2.4	Pre-Processing of Gait Video	140

Summary

This thesis explores the research of *human gait analysis and recognition*. It is still a hot research area due to the great demand for automatic human identification at a distance in many security-sensitive environments. Till now, most of the existing methods concentrate on gait recognition under controlled environments. Although a few researchers have paid attention to the problems of gait recognition under varying walking conditions, including viewing angle variation, clothing or carrying condition variation, and so on; however, the performance still leaves much to be desired. Therefore in this thesis, we aim to improve the gait recognition results under these varying walking conditions and thus facilitate the implementation of gait recognition research in real applications.

Firstly, we address the problem of gait recognition under clothing or carrying condition variations and propose two simple and effective gait feature descriptors to enhance the performance. It is reported that either selecting relevant dynamic features from the popular gait appearance feature, Gait Energy Image (GEI), or incorporating temporal information into GEI is successful in alleviating the effects caused by clothing and carrying variations. Therefore, we firstly propose a new feature, called Dynamic GEI (DGEI), to select the most relevant spatial dynamic features from GEI for recognition. Then, we also propose a spatial-temporal dynamic gait image (STDGI) to further incorporate the temporal information into GEI through color interpolation.

Secondly, we focus on the problem of gait recognition across varying views and propose a new joint subspace learning (JSL) solution. Our proposed method is inspired by the fact that if a three-dimensional (3-D) object can be well represented by a linear combination of a small number of prototypes from the same view, then the representation coefficients with the same prototypes remain fairly similar across different views. We propose to use these representation coefficients as view-invariant features for recognition. To obtain the coefficients, we first conduct JSL to learn the prototypes under different views. Then we represent the samples as a linear combination of the corresponding prototypes and extract the coefficients for recognition.

Thirdly, we investigate the view recognition problem of human gait sequences from videos. To our knowledge, little previous work has focused on this problem, yet it is important because most existing view-invariant gait recognition methods assume that the views of the testing gait sequences are known before recognition which may not hold in many practical applications. To address this problem, we propose a new subspace learning method, adaptive discriminant analysis with enhanced multiple kernel learning (ADA-EMKL), which could combine multiple gait descriptors and extract low-dimensional features for view recognition. In our proposed method, we generate high order kernels to model the non-linear relationship between base kernels and apply pairwise data correlation to alleviate the confusion among samples that could be easily misclassified.

Last but not least, we develop a new Multiview Subspace Representation (MSR) method which can handle multiple intra-subject variations during walking. By assuming that the gait data from different views of the same subject lie in a low-dimensional linear subspace, we propose to use the subspace basis to represent the data set.

Extensive experiments conducted on popular gait databases demonstrate the effectiveness and robustness of all the above proposed methods.

List of Abbreviations

ABBREVIATION	FULL EXPRESSION
3-D	Three Dimensional
ADA	Adaptive Discriminant Analysis
CCA	Canonical Correlation Analysis
CGI	Chrono-Gait Image
CHLAC	Cubic Higher-order Local Auto-Correlation
DCV	Discriminant Common Vector
DFT	Discrete Fourier Transform
DTW	Dynamic Time Warping
EDI	Energy Deviation Image
FDA	Fisher Discriminant Analysis
GEI	Gait Energy Image
GHI	Gait History Image
HumanID	Human Identification at a Distance
HMM	Hidden Markov Model
HOK	High Order Kernel

IGEI	Imagery part of the Fourier Transform of GEI
JSL	Joint Subspace Learning
KFD	Key Fourier Descriptor
LDA	Linear Discriminant Analysis
LOC	Linear Object Class
MEI	Motion Energy Image
MHI	Motion History Image
MKL	Multiple Kernel Learning
MSR	Multi-view Subspace Representation
PCA	Principle Component Analysis
PDS	Positive Definite Symmetric
pHMM	population Hidden Markov Model
RBF	Radial Basis Function
REI	Radon-transform-based Energy Image
SSR	Subspace-based Sequence Representation
STDGI	Spatial-Temporal Dynamic Gait Image
SVD	Singular Value Decomposition
SVM	Support Vector Machine
SVR	Support Vector Regression
VTM	View Transform Model
WEI	Width Energy Image
WSD	Weighted Subspace Distance

List of Figures

2.1	Block diagram of a generic gait recognition system.	11
2.2	Illustration of background subtraction in one frame of the CASIA-B gait database [17].	12
2.3	Illustration of two sample silhouette images with different segmentation errors.	14
2.4	Sample frames of the Gatech gait database [15].	16
2.5	Sample frames of CMU database [13] from six different views and four different walking conditions.	17
2.6	Sample frames of UMD gait database [18].	18
2.7	Sample frames of the SOTON gait database [14].	19
2.8	Sample frames from concrete surface and grass surface under two different views in the USF database [12].	20
2.9	Sample frames of CASIA-B database [17]. (a) 11 different capturing views. (b) 3 different walking conditions.	21
2.10	Summarization of appearance-based gait feature descriptors.	23
2.11	Samples of aligned silhouette images of two different subjects. The rightmost image in each row is the corresponding GEI.	27

2.12	Illustration of four projective gait features. (a) original silhouette, (b) horizontal projection, (c) vertical projection, (d) positive diagonal projection (top left to bottom right) and (e) negative diagonal projection (top right to bottom left).	27
2.13	Energy image examples.	28
2.14	(a) The ρ -s coordinate system, (b) Radon transform and (c) $\Re$ transform.	29
2.15	Example of the CGI temporal template (b) with its corresponding GEI (a)	32
2.16	Structure of hidden Markov Model.	34
2.17	Examples of appearance feature (GEI) from two different subjects under 4 different views.	37
2.18	Example of walking condition variation on gait appearance feature (GEI). (a) Fast (left) vs. Slow (right), (b) Normal (left), Carrying a bag (middle) and Wearing a coat (right).	41
3.1	Examples of relevant feature selection from GEI. (a) and (b) are from [91], (c) is from [108] and (d) is from [109].	46
3.2	Flow chart of spatial dynamic feature extraction.	48
3.3	Feature weight comparison. For better visualization, histograms of each weight are calculated which are shown on the right side of each weight.	50
3.4	Illustration of feature selection mask and the new proposed feature selection weight.	51
3.5	Different features extracted from GEI	52
3.6	Flow chart of spatial-temporal dynamic feature extraction.	53
3.7	(a) width vector of a gait sequence (b) the corresponding autocorrelation.	54
3.8	Visualization of the interpolation functions in Eq. 3.13 ~ Eq. 3.15.	56
3.9	Illustration of the spatial-temporal dynamic feature generation.	57

3.10	Illustration of example GEIs obtained from the two chosen databases. .	59
3.11	Illustration of the effect from the design parameters on the final recognition results. (a) θ in Eq. 3.7 and (b) λ in Eq. 3.8.	60
3.12	Illustration of the selected feature descriptors on example sequence of the two databases. Samples in the first row are from CASIA-B database and those in the second row are from USF database.	61
4.1	Illustration of the line defined by (ρ, θ) (left), the ρ -s coordinate system (middle) and the result of the transform (right).	71
4.2	The flowchart of the proposed approach.	72
4.3	Illustration of gait appearances of two different subjects under different views.	74
4.4	Illustration of the idea indicated by LOC. For each equation, the left hand side is the GEI under a certain view and the right hand side are prototypes under the same view (GEI is used for illustration here because it is better than REI for visualization of view variation).	75
4.5	Illustration of projections in the learned subspace from different feature representations. (a) GEI, (b) REI. For each class, samples come from two different views but share the same color.	81
4.6	First rank recognition rates. The gallery view is 90° . The proposed approach contains 5 important parts which are denoted as “Radon + patch + JSL + ‘ $n_c = 24$ ’ + ‘ $l < k$ ’”. In each of the figures shown above, we replace the method of each part with the alternative method and show the results of the modified approach.	86
4.7	Experimental results of other 10 views except 90° as the gallery view for view-invariant gait recognition.	87

5.1	Flow chart of the proposed approach to view recognition of human gait sequences in videos.	92
5.2	Illustration of the two features GEI and IGEI under 11 different views and 3 different walking conditions.	93
5.3	Illustration of the mean absolute value of the IGEI feature under 11 different views and 3 different walking conditions.	94
5.4	Illustration of the proposed idea for multiple kernel-based dimension reduction.	95
5.5	Illustration of samples from three different walking conditions and 11 different views.	100
5.6	Illustration of a toy example to show the role of pairwise data correlation in subspace learning.	102
5.7	Illustration of confusion matrices (%) of the results in Table 5.3. Only the values greater than 1 are shown in the figure.	103
5.8	The learned kernel weights for each method.	104
5.9	Illustration of confusion matrices (%) of the results in Table 5.4. Only the values greater than 1 are shown in the figure.	105
5.10	Illustration of confusion matrices (%) of the results in Table 5.5. Only the values greater than 1 are shown in the figure.	107
5.11	Performance of a fully automatic view-invariant gait recognition system with the gallery view from 90°.	108
6.1	Flowchart of the proposed approach MSR. The leftmost figures are the gait features ordered by view.	115

6.2	Comparison of original representation and subspace representation. (a) Original samples from 11 different views and 3 different walking conditions (normal (upper row), carrying a bag (middle row) and wearing a coat (bottom row)) represented by GEI. (b) The corresponding first six principle components computed from Eq. 6.2.	118
6.3	Illustration of data capturing and sample frames captured by each camera.	123
6.4	Experimental results to show the influence of the free parameters on the final recognition results. (a) Feature dimension of linear subspaces (k_1). (b) Number of canonical correlations for similarity measure (k_2). .	124
6.5	Illustration of the experimental setting for view variations in a multi-camera system.	126
6.6	Experimental results to show MSR's tolerance on different walking condition variations. Both training views and testing views are from camera 2,4,6,8,10. (a) Carrying a bag. (b) Wearing a coat.	127
6.7	Experimental results to show MSR's tolerance on both view variation and walking condition variation. Training views are from camera 2,4,6,8,10; testing views are from camera 1,3,5,7,9. (a) Carrying a bag. (b) Wearing a coat.	128
7.1	The diagram of a fully automatic view-invariant gait recognition system.	136
7.2	A sample confusion matrix from view recognition of human gait sequences.	138

List of Tables

2.1	Popular databases in gait recognition.	15
2.2	The settings of probe sets. The settings for gallery set is (G, A, R, NB, M/N)	21
2.3	Categories of methods in gait recognition across varying views.	36
3.1	Experimental results (%) using original feature descriptors on CASIA-B database.	60
3.2	Experimental results (%) using original feature descriptors and LDA on CASIA-B database.	61
3.3	Experimental results (%) using original feature descriptors and LDA on USF database.	63
3.4	Experimental results (%) using the upper half frames in CASIA-B database.	65
4.1	First rank recognition rates (%) under different views. The gallery view is 90°.	80
4.2	First rank recognition rates (%) achieved by different methods. The gallery view is 90°	82
4.3	First rank recognition rates (%) achieved by different combinations of the gallery views which are indicated as in the first column.	84
5.1	Training and testing stage of the proposed ADA-EMKL.	99

5.2	Experimental settings for the three test sets.	101
5.3	Experimental results (%) of Exp. 1 with test samples from normal walking sequences.	103
5.4	Experimental results (%) of Exp. 2 with test samples from walking sequences carrying a bag.	105
5.5	Experimental results (%) of Exp. 3 with test samples from walking sequences wearing a coat.	106
6.1	First rank recognition rates of different methods under conventional multiple view gait recognition. The available views are as indicated in the first column of each row, for which the number represents the camera in Figure 6.3.	124
6.2	Cumulative recognition rates of different methods under view variation. Training views are from camera 2,4,6,8,10 and testing views are from 1,3,5,7,9 with different orders.	127
6.3	Experimental results on CMU MoBo database.	128
6.4	First rank recognition rate of the proposed framework with missing data. Training views are from cameras 2, 4, 6, 8, 10 and testing views are from cameras as indicated in the table.	129
6.5	Performance of methods using different lengths of a gait cycle.	130

Chapter 1

Introduction

1.1 Motivation

Identifying people at a distance has potential applications in security-sensitive places such as banks, malls, car parks, stations and so on. However, cameras in these places nowadays only play the role of straightforward recording and processing (such as compression) rather than analyzing and alerting, which makes the system far from intelligent. Boulgouris *et. al* described in [1] an attractive application of a real intelligent surveillance system: with a camera installed at the entrance of a building, the surveillance system would open the door automatically only for the approaching people who are permitted to enter the building. With the goal of realizing what have been mentioned above, researchers are paying more and more attention to the challenges of identifying a person at a distance and without contact.

It is however not easy to identify people at a distance because most popular biometrics (iris, fingerprint, palm print and voice) are hard to obtain under this situation. Although faces can be obtained from a distance, the distance still can not be too far due to the fact that the recognition of face depends greatly on image resolution. As a result, the implementation of an intelligent video surveillance system could rely on human gait analysis as gait has been reported to be an effective biometric that can be obtained at a distance.

Generally, gait is defined as the way people walk. A gait video not only contains the dynamic information in the time domain, but also includes the static information of height, limb length, body shape, shoulder width in the spatial domain. Gait is believed to be as unique to a person as fingerprint is and thus can be used for the purpose of identification. The studies in [2, 3, 4, 5] across a few decades have demonstrated this from the medical and behavioral viewpoint. Later, the effectiveness of gait as a biometric was also demonstrated by the analysis of gait signals [6, 7, 8, 9, 10].

As a biometric, gait has the following advantages for identification in video surveillance systems,

- (1) Gait can be detected at a distance and even from low resolution videos.
- (2) The data collection of gait does not require a cooperating subject and doesn't require close contact. Thus, it won't make people feel disturbed.
- (3) Gait is relatively difficult to hide, steal or fake compared with other biometrics.

Besides identification, gait is also an effective cue for other soft-biometrics recognition, such as age [103], gender [98, 99, 101, 102], and race [100]. In fact, the knowledge from these soft-biometrics can also contribute to identification to some extent.

In other words, gait is a potential biometric that can facilitate the development of intelligent video surveillance system. And the study of gait, as well as its applications for identification and other soft-biometrics, is currently an active research area.

1.2 The Challenges

While many research groups have put efforts on gait analysis ever since it was studied, the Human Identification at a Distance (HumanID) project sponsored by the U.S. Defense Advanced Research Project Agency (DARPA) plays an important part. With

the help of this project, the research groups in many institutions started to investigate the appropriateness of gait as a biometric for human identification and achieved promising results. Moreover, it has also stimulated and attracted increasing efforts from researchers in other countries. So far, the recognition rates on small and simple database could generally achieve high accuracy. For larger and more complicated databases which are collected under controlled environments, the results can also be promising.

Despite of the significant achievements gained in gait recognition ever since its inception, there still exist a lot of challenges that prevent gait recognition from real applications. The reason is that the gait features adopted by most existing methods are sensitive to the following factors:

- (1) *Clothing variation.* It is easy to imagine that the type of clothes worn could affect the body shape which is part of a gait signal. In addition, dresses and overcoats could cover the dynamics from the lower limbs. Meanwhile, heavy clothes could also affect the dynamics of gait because of the weight difference.
- (2) *Carrying objects.* When the walking subject is carrying objects, the body shape could change greatly especially from some specific views. And such effects are usually very difficult to overcome.
- (3) *View variation.* In gait data collection from real-world, the subjects are usually unaware of being recorded and thus it is impossible to ask cooperation from them to walk in the direction desired. Therefore, when the walking direction changes, the captured gait appearance can also change. This is a very tough problem as we generally can not store the gait information of all possible directions in the database.

- (4) *Time*. A person's gait may also change over time due to growth or aging.
- (5) *Body condition variation*. In the condition of injury, pregnancy, or extreme mood, the gait dynamics could be very different from the normal state.
- (6) *Environments variation*. In a gait recognition system, extracting the walking subject from the background is an important step as it determines the quality of the gait feature which can notably influence the final recognition results. Therefore, when the background is complicated (crowded scene) or the illumination varies, the obtained gait feature would also be different.

Moreover, a general problem for biometric recognition is that the database is usually not large enough for practical applications. Thus, even if the proposed feature or algorithm improves the recognition results in empirical studies, we may not know for sure whether they are still effective in a large scale database of a real-world application.

1.3 Contributions

To the best of our knowledge, there have been efforts concentrating on the challenges of view, clothing and carrying variation in the literature. Although several solutions have been proposed, the recognition results are not promising and the performance still leaves much to be desired. Therefore, in this thesis, we aim to improve the performance under these challenging walking conditions and thus facilitate the application of intelligent surveillance systems in the future. Specifically, our contributions can be summarized as solving the following four challenging problems in gait recognition:

- (1) *How to effectively extract useful information from a gait sequence so that it is also robust under clothing or carrying condition variation?* Although the popular Gait Energy Image (GEI) descriptor is simple and effective under normal walking

conditions, it is sensitive to walking condition variations which greatly limits its application. However, it is reported that either conducting dynamic analysis on it or incorporating temporal information into it can alleviate the effects caused by clothing or carrying variation. In this thesis, we first propose a new spatial dynamic feature extraction scheme to select the most relevant features from GEI for recognition. Then, we further incorporate temporal information through color interpolation. As a result, both the spatial dynamic feature and temporal dynamic information are retained and demonstrated to be effective as well as robust under carrying and clothing condition variations.

- (2) *How can we recognize a person under view variations?* View variation is reported to be a very difficult problem in gait recognition because the gait appearances under different views could be totally different; however, this usually happens in real scenarios. To address this challenge, we propose to learn common representations for samples from different views of the same subject. This is achieved by applying the proposed joint subspace learning (JSL) method to learn the representative basis for different views. It can deal with both single and multiple-view recognition applications.
- (3) *How to recognize the walking direction of the captured gait sequences?* Most of the existing view-invariant gait recognition algorithms assume that the walking direction of the test sample is known. However, this assumption may not hold in many real applications. In this thesis, we address the view recognition problem and propose a new subspace learning method, called adaptive discriminant analysis with enhanced multiple kernel learning (ADA-EMKL), which can combine multiple descriptors and extract low-dimensional features effectively. In our proposed method, both high order kernels and pairwise data correlation are introduced in the subspace learning process. Extensive experiments are also conducted

to demonstrate the effectiveness as well as robustness of the proposed algorithm in the task of view recognition of human gait sequences.

- (4) *How to recognize a person under multiple walking condition variations?* Multiple walking condition variations frequently occur in real applications concurrently; thus the gait appearances are distorted much more significantly. To our knowledge, little previous work has addressed this problem, and existing methods as well as our previously proposed methods are not suitable for this problem because they usually attend to one variation only. Therefore, our last contribution in this thesis is to develop a multi-view subspace-based representation (MSR) method for representation, and a learning based algorithm, marginal canonical correlation analysis (MCCA), to enhance the classification accuracy. By representing samples from different views of the same subject using the corresponding subspace basis, MSR is robust to multiple intra-subject variations. We also propose to extend the MSR to represent a gait sequence to solve the problem of gait recognition with different types of missing data.

1.4 Organization

The remaining of this thesis is organized as follows:

In Chapter 2, the background of gait recognition as well as related work is introduced. We first give a brief introduction to a generic gait recognition system. Then we provide detailed descriptions of the background subtraction step as well as some popular gait databases. As for related work, we focus on the existing work that closely relates to the problems of interest in this thesis. Firstly, most of the existing gait feature descriptors are reviewed. Secondly, the research work that attempts to solve the view variation problem is summarized. Lastly, we show some existing work that attempts to address the problem of walking condition variations.

In Chapter 3, we exploit both spatial and temporal dynamic information to further enhance the gait recognition performance under clothing and carrying condition variations. Having noted that either selecting relevant features from the popular GEI or incorporating temporal information into it can enhance the recognition performance, we first propose a new spatial dynamic feature extraction scheme to reduce the information redundancy in GEI. It selects and weights the spatial dynamic feature by applying the class-dependent feature selection mask and the sample-independent discriminative feature weight. Then, we propose a new gait descriptor, called Spatial-Temporal Dynamic Gait Image (STDGI), to extract both spatial and temporal dynamic information by using the learned feature weight and color interpolation, separately. Experimental results on two popular gait databases demonstrate the effectiveness and robustness of the proposed dynamic features especially under wearing and carrying condition variations.

In Chapter 4, we formula the problem of gait recognition across varying views and propose a new Joint Subspace Learning (JSL) method to solve it. Our proposed method is inspired by the fact that if a three-dimensional (3-D) object can be well represented by a linear combination of a small number of prototypes from the same view, then the representation coefficients with the same prototypes should remain fairly similar across different views. Therefore, we propose to use these representation coefficients as view-invariant features for recognition. To obtain the coefficients, we first conduct JSL to learn the prototypes of different views. Then we represent the samples in both the gallery set and probe set acquired from different views as a linear combination of these prototypes in the corresponding views and then extract the coefficients for recognition. In addition, we also develop a new descriptor, named Radon-transform-based Energy Image(REI), and divide it into patches to further enhance the performance. Experimental results on the widely used CASIA-B gait database are presented to demonstrate the performance of the proposed method.

In Chapter 5, we study the problem of view recognition from human gait sequences in videos. This problem has rarely been formally studied in the literature, but it has great importance because most of the existing motion-based methods assume that the viewing angles of the testing gait sequences are known before identification, but this assumption may not hold in many real applications. Motivated by the fact that in the problem of view recognition, the intra-class variation could be different due to identity variation, we propose to use multiple descriptors to represent the gait sequences. The reason is that multiple feature representations which characterize the data from various aspects are usually more precise than a single representation. To better combine these multiple gait descriptors, we also propose a new subspace learning method, called adaptive discriminant analysis with enhanced multiple kernel learning (ADA-EMKL), to extract low-dimensional features for view recognition. In our proposed method, high order kernels are generated to model the non-linear relationship among the base kernels. Moreover, pairwise data correlation is applied to impose penalty to sample pairs that can be easily mis-classified. Experimental results on a popular multiview gait database are presented to show both the efficacy and the robustness of the proposed approach.

In Chapter 6, we investigate the problem of gait recognition across multiple intra-subject variations and propose a new Multiview Subspace Representation (MSR) method as the solution. By assuming that the gait data from different viewing angles of the same subject lie in a low-dimensional linear subspace, we treat them as a feature set and apply Singular Value Decomposition (SVD) to calculate the underlying subspaces. Then the subspace basis is used for representing the feature set. Thus, the distances between feature sets become the distances between subspaces, which are calculated by our new learning based algorithm MCCA. Experimental results on two popular gait databases taken under varying walking conditions have demonstrated

that the proposed framework is both effective and robust to multiple intra-subject variations.

Finally, Chapter 7 summarizes all our works in this thesis. Several promising future research directions for human gait analysis are also discussed.

Chapter 2

Background and Literature Survey

It is easy to imagine that the gait signal is recorded in the form of video. This makes it very different from other biometrics (face, fingerprint and iris) which are based on 2D images. Therefore, the processing speed as well as the captured data volume becomes main considerations for gait data preprocessing and storage. Generally speaking, there are two directions for preprocessing after walking subject detection. One fits a predefined adaptive model, usually 3D, to the body figure in each frame of the video, and the other uses a binary image to differentiate the subject from the background. Obviously, model fitting is usually time consuming because a gait video usually consists of quite a number of frames. Hence, most of the research work in human gait analysis have adopted the latter strategy for preprocessing the captured gait videos.

In this chapter, we will review the basics of a generic gait recognition system adopted by most of the existing methods as well as our proposed methods in the subsequent chapters. We will also give a literature survey on topics that are related to gait signal representation and gait recognition under different walking condition variations.

Figure 2.1: Block diagram of a generic gait recognition system.

2.1 Generic Gait Recognition System

Gait recognition is a multistage process and we illustrate in Figure 2.1 a general gait recognition system which is adopted by most of the existing motion-based gait recognition research. It mainly contains five parts: *gait signal capturing*, *background subtraction*, *feature extraction*, *recognition* as well as the enrolled *database*. Since most of the challenging problems we mentioned in the previous chapter haven't been completely solved yet, it is important for the current systems that the *gait signal capturing* is performed under a uniform background and only one subject is in the scene at a time. Moreover, the capturing viewpoint as well as the subject's walking condition should be accordant with those in the database. Once the walking video is obtained, the walking subject is detected and separated from the background in the *background subtraction* step. After this, each frame becomes a binary image in which the white part denotes the subject and the black part denotes the background. *Feature extraction* is a critical step in gait recognition as it can compact as much discriminative information as possible for recognition. Finally, the *recognition* step is conducted to identify or classify the walking subject by comparing the extracted gait features with those that are stored in the database.

Figure 2.2: Illustration of background subtraction in one frame of the CASIA-B gait database [17].

In the rest of this section, we will provide a brief introduction to the background subtraction step as well as some popular databases used in gait recognition research. In the following sections of this chapter, we will survey literature on the feature extraction and recognition steps which are also the main focus of this thesis.

2.1.1 Background Subtraction

In gait recognition, most of the proposed algorithms rely on the silhouette template. In this thesis, we also adopt this popular strategy. It needs several steps to obtain the silhouette template for each frame in a gait sequence [11, 12].

- (1) *Background reconstruction.* In cases that the background image is not available, we need to reconstruct it from the captured walking videos. A simple and effective solution regards the median value of the gait sequence as the estimated background. Let G denotes a gait sequence containing N frames, then

the background image B at the particular position (x, y) can be obtained as $B(x, y) = \text{median}\{G_1(x, y), G_2(x, y), \dots, G_N(x, y)\}$. We show in Figure 2.2(b) the estimated background image from a given gait sequence. It can be seen that the simple algorithm can reconstruct a fixed background effectively if the subject keeps moving from one side to another. However, if the walking direction is towards or away from the camera, this algorithm will fail and a Gaussian background model should be adopted for reconstruction [12].

- (2) *Foreground extraction.* Given the background image, we subtract it from the original image to extract the walking subject. The result is set to be a binary image to eliminate the depth information as well as the effect from the color of clothes. Usually, a threshold T is learned through conventional histogram method using the detected changing pixels in each frame. Then, the binary silhouette of the i th frame is obtained as:

$$D_i(x, y) = \begin{cases} 1, & \text{if } |G_i(x, y) - B(x, y)| \geq T \\ 0, & \text{otherwise} \end{cases} \quad (\text{Eq. 2.1})$$

For real RGB image, the three color channels are regarded to be independent and a pixel in the binary silhouette is determined as the foreground pixel as long as the corresponding pixel in at least one channel accords with Eq. (2.1). Figure 2.2(c) shows the silhouette image after foreground extraction.

- (3) *Shadow and noise elimination.* Obviously, the silhouette image after foreground extraction contains shadow and noise which will degrade the recognition performance if they are not suppressed. To alleviate such effects, we apply a simple shadow elimination method based on the RGB space model. For each pixel in the foreground area extracted in the above step, if the angle of the two RGB vectors (from the background image and the original image separately) is larger than a

Figure 2.3: Illustration of two sample silhouette images with different segmentation errors.

predefined threshold, the corresponding pixel is reset to belong to the background. After the above shadow elimination step, several filters, such as erosion, dilation, connected component analysis, are applied to eliminate some small regions or noises. Figure 2.2(d) shows the silhouette after shadow and noise elimination.

- (4) *Alignment.* To facilitates the computation in similarity measure, the above extracted silhouette is finally aligned to a fixed size (suppose 128×88) to form the template. The alignment process is conducted as follows: firstly, the silhouette is scaled to a height of 128 pixels; secondly, the silhouette is shifted along the horizontal direction so that the center of the silhouette is at column 44. Noted that, the height to width ratio is maintained during alignment. Figure 2.2(e) shows the final aligned silhouette.

In most cases, good quality silhouettes will be obtained after the above mentioned steps, but sometimes the silhouettes still occur with errors. We show in Figure 2.3 two sample silhouettes with different segmentation errors. In fact, such errors are inevitable in real applications as the real background is usually more complicated than that in the experimental study. Therefore, existing algorithms mainly focus on later steps of the gait recognition system.

Table 2.1: Popular databases in gait recognition.

DATA	SUBJECTS	ENVIRONMENT	VARYING PARAMETERS
USF v.1.7 [12]	71	Outdoor	View, surface, shoe
USF v.2.1 [12]	122	Outdoor	View, surface, shoe, briefcase, clothing, time
CMU [13]	25	Indoor	View, speed, carrying objects, surface
SOTON [14]	118	Indoor & outdoor	View, treadmill
Gatech [15]	20	Indoor & outdoor	View
CASIA-A [16]	20	Outdoor	View
CASIA-B [17]	124	Indoor	View, clothing, bag
UMD-1 [18]	25	Outdoor	View
UMD-2 [18]	55	Outdoor	View, time, clothing
MIT [19]	25	Indoor	Time

2.1.2 Databases

Existing gait recognition system can only recognize known persons or classify unknown subjects. Hence, the enrolled database plays an important part. When gait was first studied for the recognition task [3], the database only contained 6 walking subjects, 3 men and 3 women, with lighting sources mounted on the joints prominent during walking. Later, gait data were captured for the analysis of biomechanics [20], in which 3D marker positions, 3D ground reaction force as well as joint-angle trajectories of 5 participants were available for study. Recently, most studies of human gait were based on the databases that captured the body shape. Here, we summarize some popular ones in Table 2.1 and will give detail descriptions to them.

2.1.2.1 MIT gait database

The gait data in the MIT database contains 24 participants, 10 women and 14 men, walking in indoor environments with the camera placed perpendicular to the walking path. In the data collection, the subjects were asked to walk twice, back and forth, with their normal strides and speeds. It was captured on four separate days in two months. Except time variations, the wearing condition also varied from sweaters and coats to t-shirts and shorts. In all, 194 gait sequences were captured, with each of which contained at least 3 complete walking cycles.

Figure 2.4: Sample frames of the Gatech gait database [15].

2.1.2.2 Gatech gait database

The Georgia Tech gait database contains two kinds of gait data from 20 participants: indoor and outdoor gait sequences as well as indoor 3D trajectories. All kinds of data were captured in different view angles. Sample images are shown in Figure 2.4.

2.1.2.3 CMU Mobo gait database

The CMU Mobo gait database was also collected in indoor environment. It contained sequences from 25 participants walking on a treadmill. During the record, each participant was asked to perform four different kinds of walking: slow walk (2.06 miles/hr), fast walk (2.82 miles/hr), walk on an inclined plane and slow walk holding a ball. Six cameras were set up around the walker. Both the six views and the four walking conditions are shown in Figure 2.5.

Figure 2.5: Sample frames of CMU database [13] from six different views and four different walking conditions.

2.1.2.4 UMD gait database

The UMD database was designed for the human identification at a distance project. It contained two parts which were collected on different months of the same year. The first part contained 25 subjects who were asked to walk in 4 different views, while the second part contained 55 walking subjects captured by 2 orthogonal cameras. All the collections were done in outdoor environment. Sample images from the two datasets are shown in Figure 2.6.

Later, a small dataset containing only 12 subjects was also gathered for view-invariant gait recognition research. The walking directions captured were 0, 15, 30, 45 and 60 degrees to the camera.

(a) Four different poses of the first dataset.

(b) The two orthogonal views of the second dataset.

Figure 2.6: Sample frames of UMD gait database [18].

2.1.2.5 SOTON gait database

The Southampton Human ID at a distance gait database consists of two major segments for different purposes of research. To address the problems of whether gait signature is distinguishable across a large number of people in normal walking conditions, and to what extent researchers could devote to accurate extraction of subjects, a large population ($\simeq 100$ subjects), but basic, database was collected from six different views under three scenarios (see Figure 2.7(a)). Meanwhile, to investigate the robustness of proposed biometric techniques to various common walking condition variations, such as carrying objects, wearing different clothes or footwear, a small population (12 subjects), but more detailed, database was collected with variation in footwear, clothes, carrying conditions as well as speeds (see Figure 2.7(b)).

2.1.2.6 USF gait database

The USF HumanID outdoor gait database (version 2.1) was designed to put forward the gait recognition research under challenging walking conditions with the combina-

(a) Sample images from the large dataset.

(b) Sample images from the small dataset.

Figure 2.7: Sample frames of the SOTON gait database [14].

tion of the following five factors: (1) type of walking surface by G for grass and C for concrete, (2) camera position by R for right and L for left, (3) shoe type by A or B, (4) carriages by NB for not carrying a briefcase and BF for carrying a briefcase, and (5) the acquisition time by M for May and N for November. At last, gait data of 122 subjects under 32 different walking conditions were collected. Each subject was asked to walk multiple times counterclockwise along two elliptical courses laid on concrete ground and grass lawn separately and viewed by two cameras, whose common viewing angle was approximately 30 degrees. Samples are shown in Figure 2.8. At last, 1870 sequences were captured and divided into one gallery set for training and 12 probe sets labeled from "A" to "L" for testing. The differences of the 12 probe sets from the gallery set are shown in Table 2.2.

2.1.2.7 CASIA-B gait database

This CASIA-B database was taken in an indoor environment and with 11 cameras around the left hand side of the subject when he/she was walking. The angle between

Figure 2.8: Sample frames from concrete surface and grass surface under two different views in the USF database [12].

two nearest view directions was 18° . Figure 2.9(a) shows the captured image from the 11 cameras and the viewing angle is named as 0° , 18° , 36° , 54° , 72° , 90° , 108° , 126° , 144° , 162° , 180° from left to right. Gait data of 124 subjects were captured eventually, among whom 93 were men and 31 were women. Every subject was asked to walked 10 times in the scene(6 normal + 2 with a bag + 2 with a coat, example frames are shown in Figure 2.9(b)). Thus there were a total of $10 \times 11 \times 124 = 13640$ video sequences in the database.

Figure 2.9: Sample frames of CASIA-B database [17]. (a) 11 different capturing views. (b) 3 different walking conditions.

Table 2.2: The settings of probe sets. The settings for gallery set is (G, A, R, NB, M/N)

Exp.	Probe	Subjects	Differences
A	(G,A,L,NB,M or N)	122	View
B	(G,B,R,NB,M or N)	54	Shoe
C	(G,B,L,NB,M or N)	54	View, Shoe
D	(C,A,R,NB,M or N)	121	Surface
E	(C,A,R,NB,M or N)	60	Surface, Shoe
F	(C,A,L,NB,M or N)	121	Surface, View
G	(C,b,L,NB,M or N)	60	Surface, View, Shoe
H	(G,A,R,BF,M or N)	120	Briefcase
I	(G,B,R,BF,M or N)	60	Shoe, Briefcase
J	(G,A,R,BF,M or N)	120	View, Briefcase
K	(G,A/B,R,NB,N)	33	Time (Shoe, Clothing)
L	(G,A/B,R,NB,N)	33	Surface, Time (Shoe, Clothing)

2.2 Gait Feature Extraction

Gait feature extraction plays an important part in a gait recognition system. An appropriate gait feature should not only capture the gait characteristics that are discriminant across individuals but also be robust to operating conditions. Apparently, the binary silhouette appears to be a good feature as it not only excludes the influences of the background and color information from the skin, clothes and hair, but also captures the motion of most of the body parts. However, it is not effective because a gait

sequence usually contains quite a few frames. Therefore, how to effectively represent a gait sequence attracts lots of attentions.

Generally, all the proposed gait descriptors can be divided into two kinds: appearance-based and model-based. In comparison, the appearance-based representations extract useful information from the 2D binary silhouettes and they are relatively simple and effective, however, most of them are view-dependent. While for the model-based descriptors, they use a 3D model to represent the walking subject and thus are robust to view variation. But they usually require samples from multiple views and need to track the movements of certain body parts, both of which make it very difficult and complicated. In the following sections, we will provide detailed descriptions of these methods.

2.2.1 Appearance-based Descriptors

Currently, a lot of appearance-based descriptors have been proposed and some popular ones are summarized in Figure 2.10. We will briefly illustrate these descriptors in the following sections.

2.2.1.1 Gait characteristic descriptors

In earlier gait recognition research, light sources were mounted on the joints of the walking person and the moving trajectories played as a cue for recognition [2, 3]. Later, binary silhouettes were popular among researchers and dynamic time warping (DTW) [21] was applied to calculate the similarity of gait sequences. However, the recognition results were not satisfying because the discriminant gait feature was not fully explored. As a result, researchers began to seek effective gait characteristic descriptors to represent the gait sequence.

Polana et al. [22] proposed to use the intensity values of certain pixel over time as a 1D gait signal for recognition; Cutler and Davis [23, 95] calculated the self-

Figure 2.10: Summarization of appearance-based gait feature descriptors.

similarity plot as a gait sequence representation while Haritaoglu *et al.* [24] introduced the autocorrelation function for representation. BenAbdelkader *et al.* [25] regarded the stride and height as discriminant features which were estimated based on the estimated 3D trajectories of the movement. Classical moments were also applied to gait recognition [96, 97].

The gait characteristics extraction reduces the feature dimension to some extent, however, they are still not effective and normalization along time is needed for fair similarity measure because different sequences usually have different periods. More-

over, some of these characteristics are not easy to extract especially under extreme views. Consequently, simple and more effective descriptors are still in great demand.

2.2.1.2 Contour-based descriptors

Later, researchers proposed to extract the contour of binary silhouettes for feature representation [16, 26, 27, 28] which greatly reduced the feature dimension. Yu *et al.* [26] extracted the contour of each frame and placed them in the complex plane, whose origin was set as the centroid of the contour. Then each point on the contour can be expressed as a complex number

$$s_i = x_i + j \cdot y_i, \text{ for } i = 0, 1, \dots, N - 1 \quad (\text{Eq. 2.2})$$

where N was the pre-defined number of points on the contour. Then the contour was unwrapped counterclockwise from the top point and then represented by a complex vector $[s_0, s_1, \dots, s_{N-1}]$.

By applying certain transformation on the contour vector, some contour-based descriptors were proposed later, such as the distance transform descriptor [27], key Fourier descriptor [16] and procrustes mean shape descriptor [28].

In distance transform, the relative distances rather than the coordinates of the contour points were calculated for representation. After the silhouette contour was obtained, the centroid (x_c, y_c) of the contour was first calculated. And then the outer contour was turned into a distance signal $S = (d_0, d_2, \dots, d_{N-1})$ where $d_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2}$. Thus the original 2D silhouette contour was represented by a 1D signal. Finally, normalization was also needed to eliminate the influence of spatial scale.

For Fourier transform descriptor, given the contour vector, its Discrete Fourier Transform (DFT) was first calculated, $f_n = \sum_{m=0}^{N-1} s_m e^{-2j\pi \frac{nm}{N}}$, $n = 0, 1, \dots, N - 1$. For

each gait cycle, there should be $N * T$ descriptors, where T is the period. Among these descriptors, the ones whose amplitudes reached the peaks were defined as Key Fourier Descriptors (KFD). And it had been shown that KFD had better discriminatory capability than all Fourier descriptors and the computational complexity was also greatly reduced.

To further reduce the computational complexity, Wang *et al.* [28] proposed to use a mean shape to represent all the shapes in a gait cycle. Given a set of shapes $\mathbf{s} = [s_1, s_2, \dots, s_k]^T$, they were first centralized as $\mathbf{u} = [u_1, u_2, \dots, u_k]^T$, where $u_i = s_i - \bar{s}$ and $\bar{s} = \sum_{i=1}^k s_i / k$, then their mean shape $\hat{\mathbf{u}}$ can be obtained by computing the dominant eigenvectors of the following matrix:

$$\mathbf{S}_{\mathbf{u}} = \sum_{i=1}^k \frac{(\mathbf{u}_i \mathbf{u}_i^*)}{(\mathbf{u}_i^* \mathbf{u}_i)} \quad (\text{Eq. 2.3})$$

where the superscript $*$ represents the complex conjugation transpose. The full Procrustes distance between two mean shapes $\hat{\mathbf{u}}_1$ and $\hat{\mathbf{u}}_2$ was defined as

$$d_F(\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2) = 1 - \frac{|\hat{\mathbf{u}}_1^* \hat{\mathbf{u}}_2|^2}{\|\hat{\mathbf{u}}_1\|^2 \|\hat{\mathbf{u}}_2\|^2} \quad (\text{Eq. 2.4})$$

For contour-based features descriptors, the quality of the binary silhouettes plays an important part because it determines the exact extraction of the outer contour. However, for low quality or low resolution data which are frequently obtained in realistic applications, the extracted outer contour from the silhouette could be greatly distorted. Although attention has been paid to the silhouette quality problem [88, 93, 94], difficulties that usually exist in real applications (as indicated in Figure 2.3) remain unsolved.

2.2.1.3 Energy image-based descriptor

A common problem for previously mentioned descriptors is that they need time normalization to alleviate the period difference which may generate artificial information.

To address this difficulty, Lee et al. [46] proposed an appearance histogram descriptor as well as a harmonic decomposition in frequency domain to represent a gait sequence. Later, energy image descriptors are proposed which represent a gait cycle using the mean frame. According to the representation of gait frames, several energy image descriptors have been proposed: Gait Energy Image (GEI) [29], Width Energy Image (WEI) [36, 37, 38, 111], X-T plane and Y-T plane energy images [39] and Radon transform template [32]. In the following paragraphs, we will introduce these descriptors along with some extensions, which are related to our studies in this thesis.

GEI: GEI is the most popular descriptor in gait recognition due to its simplicity and effectiveness. It represents a gait cycle by its mean frame. Given the aligned binary silhouette images $B(x, y)$, GEI is defined as follows:

$$G(x, y) = \frac{1}{N} \sum_{t=1}^N B_t(x, y), \quad (\text{Eq. 2.5})$$

where N is the number of frames in the complete cycle, t is the frame number, and x and y are pixel position in the 2-D image coordinate. Figure 2.11 shows selected sample frames from the gait cycle of two people in different rows separately, and the corresponding GEIs are shown in the rightmost image of each row. As we can see from the figure, GEI captures both the major shapes of silhouettes and the main body movement over time. It is also demonstrated in several experiments that GEI is very competitive in recognition task.

WEI: WEI is obtained based on the projective descriptor which uses the total number of bright pixels in certain direction to represent the silhouette. There are mainly four directions [36, 37, 38, 111] that we can project the binary silhouette, namely, horizontal, vertical, positive diagonal and negative diagonal. All the four projections are shown in Figure 2.12.

After the projective vector $P_t(j)(j = 1, \dots, n)$ is obtained (where n is the dimension), WEI is computed by the equation as follows:

Figure 2.11: Samples of aligned silhouette images of two different subjects. The rightmost image in each row is the corresponding GEI.

Figure 2.12: Illustration of four projective gait features. (a) original silhouette, (b) horizontal projection, (c) vertical projection, (d) positive diagonal projection (top left to bottom right) and (e) negative diagonal projection (top right to bottom left).

$$W(j) = \frac{1}{N} \sum_{t=1}^N P_t(j), \quad (\text{Eq. 2.6})$$

where N is the number of frames in the gait sequence and t is the frame number.

X-T plane and Y-T plane energy images: A gait sequence is actually a spatial-temporal (XYT) volume. For GEI ($\frac{1}{T} \sum_{t=1}^T B_t(x, y)$, T is the period), it is the energy image along the time axis (X-Y plane). Similarly, there should be other two energy images along the x -axis (Y-T EI) ($\frac{1}{X} \sum_{x=1}^X B_x(y, t)$, X is the number of columns in each frame) and the y -axis (X-T EI) ($\frac{1}{Y} \sum_{y=1}^Y B_y(x, t)$, Y is the number of rows in each frame) separately. Examples of the three kinds of energy images are shown in Figure 2.13(b)-(d). In [39], the X-T plane energy image is used as the gait sequence representation; besides, it also segments the thigh and leg regions from the main body region to preserve the dynamic features. Experimental results in [39] showed that

Figure 2.13: Energy image examples.

the recognition rates using X-T EI are a little worse than those using GEI on the same database. Meanwhile, as different sequences have different number of frames, period estimation as well as time normalization are needed to ensure the same feature dimension among different gait sequences.

Radon transform template: Representing the binary silhouettes with certain transformations, we can obtain other kinds of energy image descriptors besides GEI. Radon transform template [32] is one of the examples. By defining the center of the silhouette as the reference point, its radon transform can be calculated as follows:

$$R_f(\rho, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} B(x, y) \delta(\rho - x \cos \theta - y \sin \theta) dx dy \quad (\text{Eq. 2.7})$$

where $R_f(\rho, \theta)$ is the line integral of a 2-D image $B(x, y)$ along a line from $-\infty$ to ∞ . The position of the line is determined by two parameters ρ and θ . All the parameters are shown in Figure 2.14(a). The corresponding radon transform is shown in Figure 2.14(b).

Given the radon transforms of all binary silhouettes, radon transform template first estimates the continuous gait signal within a gait cycle and then calculates the mean to represent the gait cycle.

Extensions: The energy image descriptors are important in gait recognition because many researchers exploit different discriminative information based on them

Figure 2.14: (a) The ρ - s coordinate system, (b) Radon transform and (c) $\mathfrak{R}$ transform.

[30, 31, 32, 33, 34, 35, 105, 106, 107]. For example, authors in [106] and [107] proposed to use the local binary pattern (LBP) to record the neighborhood relation of a GEI pixel. Another example is the popular Gabor feature descriptor [31] which describes GEI in different scales and orientations. The gait representation methods which apply certain transformation on GEI are powerful, but their computational cost is very high compared with the original GEI.

Meanwhile, several researchers proposed to extract dynamic information from GEI to make it robust under walking condition variations. The stimulation to study the gait dynamic should be the work done by Foster *et al.* [80]. They used different area masks, such as horizontal line, vertical line, bottom half and full body, to perform recognition experiments and found out that there was sufficient information in the dynamical time signature of gait for recognizing people with a high accuracy. Later, Bashir *et al.* [109] proposed to select the dynamic part as those whose intensity values were lower than a pre-defined threshold. Obviously, this feature selection mask was person dependent and thus it recorded the within-class dynamic information. Moreover, they also proposed two learning based methods, supervised and unsupervised, to generate the mask [91]. These masks attempted to select a predefined number of features that possess the most discriminative power in the learning process. While authors in

[108] proposed to extract dynamic information by multiplying the sample variances (weight) with the original intensity values pixel by pixel, thus a new gait feature, defined as EGFI, was formed. Compared with the mask proposed in [109], the two feature selection masks [91] and the feature weight [108] were all obtained through learning and thus they were sample independent. Moreover, they also selected the between-class dynamic information.

Last but not least, $\mathfrak{R}$ transform [33] is also another example which is based on the Radon transform template:

$$\mathfrak{R}_f(\theta) = \int_{-\infty}^{\infty} T_{Rf}^2(\rho, \theta) d\rho \quad (\text{Eq. 2.8})$$

where T_{Rf} is the Radon transform of f . Figure 2.14(c) shows the corresponding $\mathfrak{R}$ -transform of the middle figure. The advantage of using $\mathfrak{R}$ transform is that it is invariant to translation and scaling when normalized by a scaling factor (the area of the silhouette shape). Meanwhile, it has also been shown in [33] that a rotation of the shape results in a translation of the $\mathfrak{R}$ transform along the horizontal axis. Therefore, only rotations modify the function.

Energy image based descriptors are not only robust to noise but also efficient and effective. However, they ignore the temporal information when representing a gait cycle by a compact representation. Nevertheless, they are still widely used because of their promising recognition results in experimental studies.

2.2.1.4 Temporal template descriptor

Temporal template makes it possible to preserve temporal information within a compact representation. Let $B_t(x, y)$ be an image sequence with t as the frame number and let $D_t(x, y)$ be a binary image in which the regions of motion at time t compared with that at time $t - 1$ are set to 1; for many applications image differentiating is adequate to generate D .

(1) *MEI*

The binary cumulative motion images are referred to as motion-energy image (MEI) [40]. It is defined as:

$$E_{\tau}^{MEI}(x, y) = \bigcup_{t=1}^{\tau} D_t(x, y) \quad (\text{Eq. 2.9})$$

Here τ is the time duration of the sequence. However, MEI only records where the motion shows in the image, rather than how the motion changes along time. Although it is termed as energy image, we regard it as temporal template descriptor because its representative ability depends on the time duration.

(2) *MHI*

To address the problem of MEI, motion-history image (MHI) was presented in [41]. The pixel intensity of MHI is defined to record the temporal history of motion at that point. Moreover, a simple replacement and decay operator is used to generate MHI:

$$E_{\tau}^{MHI}(x, y) = \begin{cases} \tau, & \text{if } D_t(x, y) = 1 \\ \max(0, E_{\tau}^{MHI}(x, y, \tau - 1) - 1), & \text{otherwise} \end{cases} \quad (\text{Eq. 2.10})$$

Accordingly in MHI, the more recently moving pixels are brighter. In [41], it is shown that the combination of MEI and MHI possesses better discriminative ability than either alone.

(3) *GHI*

The Gait History Image (GHI) [42] is another temporal template which is defined as follows:

$$E_{\tau}^{GHI}(x, y) = \begin{cases} \tau, & \text{if } S(x, y) = 1 \\ \sum_{t=1}^{\tau} D_t(x, y)(t - 1), & \text{otherwise} \end{cases} \quad (\text{Eq. 2.11})$$

where $S(x, y)$ represents the common regions among all the frame of a gait sequence; that is, pixels which are always foreground are bright in S . To obtain S , if the silhouettes have good quality, intersection is used:

$$S(x, y) = \bigcup_{t=1}^{\tau} B_t(x, y) \quad (\text{Eq. 2.12})$$

Figure 2.15: Example of the CGI temporal template (b) with its corresponding GEI (a)

While if the silhouettes have bad quality, S will be defined based on the GEI with a threshold th :

$$S(x, y) = \begin{cases} 1, & \text{if } G(x, y) \geq th \\ 0, & \text{otherwise} \end{cases} \quad (\text{Eq. 2.13})$$

For comparison of the above temporal templates, MEI and MHI can't record static information of gait. MHI can record more detailed motion information by highlighting recent moving pixels. GHI integrates the advantages of both MHI and GEI, by recording both the static information and the detailed motion information.

Recently, a novel temporal template, Chrono Gait Image (CGI) was proposed for gait recognition which employed color interpolation to incorporate the temporal information [43]. Figure 2.15 shows an example of GEI and the corresponding CGI temporal template. Obviously, CGI is much more informative than GEI.

By incorporating temporal information in the compact representation, temporal templates are proved to be robust in wearing and carrying condition variation compared with GEI. However, there are much to be desired because spatial dynamic information is not fully exploited in these temporal templates.

2.2.2 Model-based Descriptors

Generally speaking, model-based methods can be divided into two directions: one models the body of the person and the other models the walk of the person. For body models, parameters that define the body are calculated for each frame of the walking sequence. While for walking models, usually a model of how the subject moves is generated and it is subject-dependent.

Several body models have been developed so far [44, 45, 46, 47, 48, 49, 50, 58] and they can be further divided into three categories. The first category is the joint-based model. In [58], 24 degree of freedom are defined to measure the position, orientation as well as movement of the body model. In [45], the body model is presented by the static body parameters which were defined as the distances between different joints. The second category models different body parts as different simple 2D shapes such as trapezoid [44], ellipse [46, 48], cylinder [51], truncated cone [52], and super-quadrics [53]. The shape parameters which measure the fitness of the proposed shape with the corresponding body parts are calculated for every frame in the gait sequence. The third category uses different 3D shapes to model the human body [47] which is robust to view-point. Interested readers can refer to relevant reference papers for the illustration of these body models.

However, in real application such as video surveillance, the body models mentioned above are not attractive mainly due to their complexity. The reason is that the modeling process is usually based on tracking which is computational complexity and time consuming.

For walking models, usually Hidden Markov Model (HMM) is applied for the moving parameter estimation [54, 55, 56, 57, 59, 60]. The structure of HMM is shown in Figure 2.16. The HMM framework is suitable because the gait postures that the individual subject shows in a gait sequence can be regarded as the states (S_t) of the

Figure 2.16: Structure of hidden Markov Model.

HMM and are identical to that subject. In conventional HMM-based gait recognition research, the model is learned for each subject in the database. And given the learned HMM parameters (λ_i) for the i th person ($i=1,2,\dots,N$, where N is the total number of subjects), the identity of an enquiry sequence is usually obtained by the maximum likelihood criterion.

$$\hat{i} = \arg \max_i \mathbf{P}(Y|\lambda_i) \quad (\text{Eq. 2.14})$$

where Y is the representation of the input inquiry sequence.

Kale et al. [54] introduced HMM to gait recognition and proposed both indirect and direct approaches to train HMM. Firstly, a feature vector called frame to exemplar distance (FED) was formed to capture both the shape and the motion information, then HMM is trained to capture the information in the FED vector sequences. To further enhance the recognition performance, Chen et al. [57] proposed a two-level time series model based on dynamic Bayesian network. In the first level, several temporally adjacent clusters modeled by a Dynamic Texture or logistic Dynamic Texture were formed for each feature cycle. In the second level, HMM was trained to capture the relationship among the above textures. Noticing that improvements are limited if efforts were only put on the gait feature representation, the authors in [56] focused on the adaption of the HMM to the gait dynamics and designed the topology of the HMM to be a circular ring based on the fact of periodical in human gait. While in [60],

the HMM was given by $\lambda_i = (A_i, B_i, \pi_i)$ where A_i is the transition matrix, B_i is the probability distributions of a feature vector conditional on the state index and π_i is the initial distribution. All the parameters are refined iteratively in the training process. Later, extensions of HMM with more complex structure were proposed to achieve better gait recognition performance. In [59], the authors explored two extended HMM frameworks, the factorial HMM (FHMM) and the parallel HMM (PHMM), for gait recognition using multiple feature representations. FHMM composes of several independent HMM but the observation vector depends upon the state of all the layers. While for PHMM, the observation vector of each layer is independent. Generally, FHMM can be regarded as a feature-level fusion of HMM representation and PHMM can be regarded as a decision-level fusion of HMM structure.

Different from the previously mentioned work which models the gait dynamic of each subject by an individual HMM, Liu *et. al* [55] proposed a population Hidden Markov Model (pHMM) which captures the gait dynamics from a set of training subjects. Then, it was used as a generic model to normalize the gait dynamics in each testing sequence. Meanwhile, linear discriminant analysis (LDA) was also applied to maximize the inter-subject distances and minimize the intra-subject variations.

Besides the HMM model, Veeraraghavan et al. [61] proposed a new walking model to achieve rate-invariant recognition which can be applied to gait recognition. The model contains a nominal motion trajectory and a function space which captures the probability distribution of motion-specific time warping transformations. Then, recognition is achieved by a Bayesian algorithm.

Generally, the walking model can better capture the gait dynamics in time domain. However, it is still view-dependent and needs large number of samples for exact parameter estimation.

2.3 Gait Recognition Across Varying Views

It has been demonstrated that gait is an efficient biometric for human identification at a distance. However, the performance can be easily affected when the viewing angle of the test sequence is different from what we have stored in the database. We illustrate in Figure 2.17 examples of GEI from two different subjects under different views. Obviously, the intra-subject differences under large view variation (row) are much larger than inter-subject differences of the same view (column). That’s why gait recognition across large view variation is still challenging and remains to be solved.

Existing methods that attempt to solve the view problem can be generally divided into three categories, the first one explores different discriminative information from multiple views of gait sequences and applies effective fusion methods to perform recognition [11, 62, 66, 71]; the second category builds 3-D model or visual hull for the walking subject according to the samples from multiple views[65, 70, 72]; the last one extracts view-invariant features or learns view transform model to realize view-invariance, in which the training and testing sequences are usually from different views [63, 64, 67, 68, 69, 73, 74, 75]. We summarize these in Table 2.3 and will provide detail descriptions for each category in later sections.

Table 2.3: Categories of methods in gait recognition across varying views.

Category	Training View	Testing View	Idea
1	Multiple	Multiple	Fusion (feature or score level)
2	Multiple	Multiple or single	3-D model or visual hull
3	Single	Single	View-invariant feature or view transform model

2.3.1 Multi-view Gait Recognition

Gait recognition based on side view (90° in Figure 2.17) was once popular because it was believed that this particular view contains the richest gait information of the waking subject. However, there has been experimental study in [125] which showed

Figure 2.17: Examples of appearance feature (GEI) from two different subjects under 4 different views.

the gait recognition performance from the frontal-parallel view was sometimes better than that from the side view. Therefore, researchers believed that the combination of multiple views could further enhance the recognition performance. The authors of [66] extracted useful features from gait sequences of 5 different views and proposed a learned weighted sum fusion rule to combine these features. In [71], a new weight for score level feature fusion based on Dempster-Shafer theory was proposed and samples from two different views were used to perform recognition. These methods belong to the first category, but they are still view-dependent because the testing views should be accordant with the training views.

Meanwhile, several researchers proposed to generate a 3-D model or visual hull from the available multi-view data samples. In [70], multiple images from different views were used to generate a visual hull for each subject and such visual hull was further utilized for recognition. Tyagi *et. al* [65] presented a framework to generate the canonical view through multiple views. Firstly, the sagittal plane of the object

was identified through the cameras and scene geometry. Then, each of the obtained multiple views was warped to the canonical view through learnt planar homographies. Finally, evidence fusion was applied to recover the object’s shape. Bodor *et al.*[72] learned a three-dimensional (3-D) model of the walking person from samples acquired in four different views and used image-based rendering techniques to generate gait feature under arbitrary views. Like previously mentioned model-based descriptors, these methods in the second category can solve the view problem to a large extent. However, the modeling process is usually complicated and time consuming.

2.3.2 Cross-view Gait Recognition

The third category should be the most difficult because the appearance features in different views could be totally different (see Figure 2.17). To solve this challenging problem, some researchers proposed to extract view-invariant features for gait sequence representation. In [76], the author proposed to estimate the real height of a person through the height in the silhouette. However, this calculation depended on the parameters of the camera, the depth information of each pixel and the walking direction of the person. Goffredo *et al.* [73] built up a markerless joint motion system based on 2D silhouettes of human gait sequences, and rectified these features across different views to achieve view-invariance. While reasonably good recognition rates can be attained for large view variation, its performance degrades under small view variations. Moreover, it can’t handle extreme views. Lu and Tan [74] proposed to project gait features from two different views into a dimension reduced feature subspace, so as to preserve the intra-class geometrical structures and maximize the interclass distances simultaneously. Bashir *et al.* also proposed to apply CCA to maximize the correlation between samples from two different views. In fact, these methods only have tolerance on small view variations.

Meanwhile, several researchers proposed to perform view transformation to realize view-invariance [64, 77, 78, 79] and several view transformation methods were applied, such as image geometry, Singular Value Decomposition (SVD) and Support Vector Regression (SVR). For image geometry, one important job is to estimate the viewing angle which is defined as the angle (θ) between the walking direction and the image plane. In [77], the author proposed two methods to estimate the parameter θ . One was the perspective projection approach and the other was the optical flow based Shape from Motion approach. After θ was obtained, gait feature captured at the non-canonical view was then projected to the canonical view through the following rules:

$$x_0 = f \frac{x_\theta \cos(\theta) - f \sin(\theta)}{-x_\theta \sin(\theta) + f \cos(\theta)}, \quad y_0 = f \frac{y_\theta}{-x_\theta \sin(\theta) + f \cos(\theta)}, \quad (\text{Eq. 2.15})$$

Clearly, camera parameters are required for the calculation which makes the method not applicable. Although the authors showed an optimization criterion for the estimation of the focal length f through experimental study, estimation errors may be significant due to limitations of image processing techniques. Hence, the quality of the transformed feature will be greatly degraded.

For SVD based VTM [64], the main supporting idea is that each feature obtained is composed of a view-independent vector and a subject-independent vector. To compute these view-independent vectors and subject-independent vectors, the features from multiple views of different subjects are gathered to build a mixture gait matrix in which each row contains samples (column vector) from the same views of different subjects and each column contains samples from different views of the same subject. Then SVD is applied to get the solution. Mathematically, the above steps can be

expressed as:

$$\begin{bmatrix} g_{\theta_1}^1 & \dots & g_{\theta_1}^K \\ \vdots & \vdots & \vdots \\ g_{\theta_I}^1 & \dots & g_{\theta_I}^K \end{bmatrix} = USV^T = \begin{bmatrix} P_{\theta_1} \\ \vdots \\ P_{\theta_I} \end{bmatrix} [v^1 \dots v^K] \quad (\text{Eq. 2.16})$$

where $g_{\theta_i}^k$ is the column vector of a gait sample representation of subject k under view θ_i , $P = US$ is regarded as the view-dependent matrix and V is the subject-dependent matrix. Accordingly, each gait feature can be factorized as: $g_{\theta_i}^k = P_{\theta_i} v^k$. Therefore, it is easy to conclude that any gait feature transformation from viewing angle θ_i to θ_j can be obtained by: $g_{\theta_j}^k = P_{\theta_j} P_{\theta_i}^+ g_{\theta_i}^k$, where $P_{\theta_i}^+$ is the pseudo inverse matrix of P_{θ_i} . The main problem of SVD-based VTM is that its main supporting idea lacks mathematical proof. Moreover, global feature is used for representation which makes the algorithm sensitive to background noise and partial occlusion.

Later, authors in [79] proposed a SVR-based VTM which is more effective than SVD-based VTM. To transform a GEI obtained in view θ_j to θ_i , multiple SVR processes are needed with each of the regression defined as follows:

$$p_{\theta_i}^k \approx f(\theta_i, \theta_j, k, p) = \langle w, ROI_{p_{\theta_j}^k}^{\theta_i} \rangle + b \quad (\text{Eq. 2.17})$$

where $p_{\theta_i}^k$ denotes the p th pixel of a GEI $g_{\theta_i}^k$ from view θ_i of subject k , $ROI_{p_{\theta_j}^k}^{\theta_i}$ denotes the pixels in $g_{\theta_j}^k$ that are relevant to the pixel $p_{\theta_j}^k$. Obviously, the effectiveness of the above SVR model needs sufficient numbers of training pair $(g_{\theta_i}^k; g_{\theta_j}^k)$ from different subjects.

A main problem for VTM is that they need the exact view of both training and testing sequences and wrong estimation of the viewing angle would generate large transformation error.

Figure 2.18: Example of walking condition variation on gait appearance feature (GEI). (a) Fast (left) vs. Slow (right), (b) Normal (left), Carrying a bag (middle) and Wearing a coat (right).

2.4 Gait Recognition Across Varying Walking Conditions

As we have indicated in the previous chapter, several walking conditions can affect gait appearance, including shoes and clothing, views, walking surfaces, carriages and walking speed. We show in Figure 2.18 examples of walking speed, clothing and carrying variations on GEI feature. From the figure we can see that, changes in walking speed has great effect on dynamic features (movement of lower limbs) while changes in clothing and carrying variations affect the body shape significantly. In later sections, we will introduce the existing methods that attempt to solve these two kinds of variations separately.

2.4.1 Walking Speed Variation

Among the previously mentioned factors that can affect human's gait, walking speed is a most common one because people often change their walking speeds in real life and as a result, gait characteristics such as period, stride as well as arm swing will also change accordingly.

To quantitatively estimate the effects of speed variation on the gait characteristics

mentioned above during walking, the authors in [83] firstly collected video data in which the subjects were asked to walk at different speeds, then they proposed a stride normalization algorithm to alleviate the effect of speed variation on stride. In the proposed method, a statistical relation between the stride and the walking speed are learned [90] based on only five silhouettes (which show the minimum or maximum stride in a cycle). Obviously, other informative images were discarded which may also be helpful for recognition.

To cope with the variations in walking speed, Liu *et al.* [55] showed that improved gait recognition can be achieved by focusing on the shape information after normalization of dynamics. To normalize the gait dynamics, a population HMM (pHMM) defined for a set of individuals was applied to generate a generic walking model. After normalization, the shape information of each gait sequence is obtained by Viterbi decoding of the gait dynamics to get an averaged gait cycle of fixed length. The method only considered the temporal change as a result of speed variation and ignored the spatial changes (e.g. stride) which are also important for recognition. Moreover, the method is applicable to a small range of speed changes.

In [89], the authors proposed a SVD-based speed transformation model based on the assumption that dynamic features (e.g. stride) are more sensitive to speed variations than static features (e.g. thigh and shin lengths). Therefore, they proposed to separate the static and dynamic features first by fitting a human model to each of the silhouette sequence, then apply SVD-based factorization (Eq. 2.16) only to the dynamic features. Based on the factorization, any dynamic feature can be transformed from a reference speed to a target speed. Finally, the transformed dynamic features are combined with the previously separated static features to form the silhouette sequence in the target speed. Obviously, both the reference and the target speeds should be known in advance which was almost impossible in real surveillance system.

Motivated by the view-invariant feature extraction scheme proposed for gait recognition across view variations, Aqmar *et al.* [82] proposed to extract speed-invariant features against speed variations. Features were first extracted by Fisher discriminant analysis (FDA)-based cubic higher-order local auto-correlation (CHLAC) [82] and then HMM was applied for statistical learning the relations among the extracted features under different speeds. Thus, the proposed method is applicable to speed variations not only within a gait sequence but also across different sequences.

2.4.2 Walking with Carrying or Wearing Condition Variation

Both carriages and clothing variations have significant effect on the recognition performance because the gait appearance features change a lot under such walking condition variations [86, 92]. Several methods have been proposed to remove such effects and we will give a brief introduction to them in the following paragraphs.

Region-based gait representation was first introduced in gait recognition by Foster *et al.* [80]. They used predefined masking functions to select certain image area and regarded these areas as a time varying signal which was used as a signature for recognition. In [91], a subject-dependent mask function was proposed to select the gait dynamics from GEI. It further demonstrated that dynamic information is effective under carrying or clothing variations. In [81], the authors also proposed to construct the GEI with sway alignment instead of upper body alignment to alleviate the effect due to carrying and clothing variations.

Attempting to remove the effect as a pre-processing step is also a possible solution. In [87], the authors proposed a backpack removal method based on recursive principal component analysis (PCA) reconstructions and error compensation, in which GEI was chosen as the gait feature representation. Given the principal components obtained from the training GEIs without a backpack, the backpack effect in a new GEI can be alleviated by applying these backpack-free principal components.

In [84], the authors considered a special condition when only one cycle was available for training while the testing sequences may contain variations such as wearing a coat or carrying objects. Considering these practical scenarios, they proposed a method for dimensionality reduction based on artificial subjects and Linear Discriminant Analysis (LDA). The artificial subjects were generated based on the rotation of original images and the introduction of artificial clothing.

In fact, multiple walking condition variations usually happen at the same time in real video surveillance systems. However, the above proposed methods only aim at solving a particular kind of walking condition variation, thus the corresponding solutions will not work under multiple variations.

2.5 Conclusion

In this chapter, we first provided a brief introduction to the background knowledge of gait recognition, in which a generic gait recognition system as well as some popular gait databases was described. Secondly, we reviewed related work in the literature according to the problems we will study in this thesis, including gait feature representation, gait recognition across varying views and varying walking conditions. For gait feature descriptors, existing proposals can be divided into two categories: model-based and appearance-based. In comparison, the appearance-based descriptors are more popular because they are simple and effective. For gait recognition across varying views, existing solutions were reviewed according to the number of views available. However, a common problem is that most of these methods need to know the exact view of the testing sequences before recognition. For gait recognition across walking condition variations, existing work mainly focuses on variations such as speed, clothing, or carriages. Moreover, they only consider certain kind of variation and thus can't handle multiple walking condition variations.

Chapter 3

Gait Recognition Based on Dynamic Feature Analysis

Walking with different carriages or clothing usually happens in real applications which could affect the gait signal greatly. Based on previous literature survey in chapter 2, we can conclude that such effect could be easily alleviated by conducting dynamic feature analysis on the extracted gait appearance. In this chapter, we also study the gait dynamics and base our dynamic analysis on the popular gait feature descriptor, Gait Energy Image (GEI), because it is simple and effective. To make the extracted gait feature robust to different walking condition variations, we conduct both spatial and temporal dynamic analysis on GEI. Details of the analysis will be provided in the subsequent sections.

3.1 Introduction

In early days of gait recognition research, most of the efforts were put on gait sequence representation because this step was regarded to be very important in a generic gait recognition system. As a result, lots of feature descriptors were proposed and we have reviewed them in Chapter 2. Among these feature descriptors, GEI was a most popular one. It represents a gait sequence or cycle by a single image which is obtained

Figure 3.1: Examples of relevant feature selection from GEI. (a) and (b) are from [91], (c) is from [108] and (d) is from [109].

through averaging all frames in the sequence/cycle. Thus it is very simple to calculate. Moreover, several studies have reported that GEI is also much effective in the recognition task. Hence, more attention were attracted to study GEI.

Many descriptors have been proposed based on GEI and they can be roughly grouped into two kinds, one reduces the information redundancy in GEI and the other adds temporal information into GEI. For information redundancy, Foster et al. [80] demonstrated for the first time that selecting relevant gait features through simple area masks could achieve satisfied recognition results. Later, several algorithms [91, 108, 109] were proposed to select relevant features (see Figure 3.1, detailed descriptions of these methods can be referred to Chapter 2) in GEI for recognition and the higher recognition results compared with GEI proved that not all pixels in GEI will contribute to the recognition. Due to silhouette alignment in the preprocessing step, it seems that some body parts (pixels with high intensity value) move little during walking. These body parts are defined as static part of gait and they, in fact, are redundant for recognition. Then how to alleviate such redundancy and extract valuable information from GEI to further improve the recognition performance interests many researchers in this area.

Another point is that by averaging, GEI ignores the order of frames in the gait

sequence which leads to loss of temporal information. Therefore, how to incorporate temporal information into a compact representation is also worth to study. Existing methods that attempt to alleviate either of the two limitations have been reviewed in Chapter 2. Although improvements can be achieved with these methods under walking condition variations, the performance still leaves much to be desired.

In view of this, we investigate in this chapter the problem of how to effectively extract both the spatial and temporal dynamic information to represent the gait sequence so as to enhance the gait recognition performance under different walking conditions. Firstly, spatial dynamic feature analysis on GEI is presented. A new feature extraction criterion is proposed. It is based on two different dynamic feature selection rules, one is the feature weight obtained from the learned Discriminative Common Vectors (DCV) which measures the degree of between-class dynamic variation and the other is the feature selection mask based on class mean sample which selects the within-class dynamic information. Thus, the combination of the two would lead to more discriminative power. Secondly, a spatial-temporal dynamic feature extraction scheme is proposed to incorporate the temporal information into dynamic GEI. Given the previously learned feature weight, the spatial dynamic information in each frame is first selected. Then, the selected dynamic feature in each frame is colored according to its gait pose. And at last, the proposed Spatial-Temporal Dynamic Gait Image (STDGI) is obtained by averaging all color images in a gait cycle, which contains both spatial and temporal dynamic information. Experimental results on two popular gait databases demonstrate the effectiveness and robustness of the proposed dynamic feature for recognition task, especially under walking condition variations.

The rest of this chapter is organized as follows: firstly, the proposed spatial dynamic feature analysis on GEI is illustrated; then, how to extract the proposed STDGI feature is provided; and finally, experimental results are listed with further discussion.

Figure 3.2: Flow chart of spatial dynamic feature extraction.

3.2 Spatial Dynamic Feature Analysis Based on GEI

We illustrate in Figure 3.2 the flow chart of the proposed spatial dynamic feature extraction scheme. Firstly, GEI is obtained for each gait sequence, then, given the obtained GEIs, two different dynamic feature extraction rules are applied. One is the dynamic feature weight obtained from the learned DCVs and the other is the dynamic feature selection mask gained by maintaining the pixels in the class mean sample whose intensity values are smaller than a predefined threshold. As we have mentioned previously, these two rules define dynamic information from different point of views, therefore, we propose to combine them to extract a more discriminative dynamic feature.

3.2.1 Discriminant Feature Extraction

To better describe the data samples within each class, discriminant common vector (DCV) is first calculated which has been reported to be able to extract both the discriminative information from the within-class scatter matrix and the common properties of each class. Common vectors are the vectors that span the null space of the within-class scatter matrix S_W which is defined as

$$S_W = \sum_{i=1}^C \sum_{n=1}^{N_i} (x_n^i - \mu_i)(x_n^i - \mu_i)^T \quad (\text{Eq. 3.1})$$

where x_n^i is a d -dimensional column vector which denotes the n th sample from the i th class ($n = 1, \dots, N_i$), μ_i is the mean of samples in the i th class, C is the number of classes in the training set and N_i ($i = 1, \dots, C$) is the number of samples in the i th class. Let R^d be the original d -dimensional sample space the range space and null space of S_W are defined as follows:

$$V = \text{span}\{\alpha_k | S_W \alpha_k \neq 0, k = 1, \dots, r\} \quad (\text{Eq. 3.2})$$

and

$$V^\perp = \text{span}\{\alpha_k | S_W \alpha_k = 0, k = r + 1, \dots, d\} \quad (\text{Eq. 3.3})$$

where $r < d$ is the rank of S_W , $\{\alpha_1, \dots, \alpha_d\}$ is an orthonormal set that spans the whole space of S_W , and $\{\alpha_1, \dots, \alpha_r\}$ is the set of orthonormal eigenvectors corresponding to the nonzero eigenvectors of S_W .

In fact, the task to directly get the common vectors is computationally intractable since the noise level in the null space is significant compared with the original signal [173]. However, using the orthogonal complement of the null space of S_W can be a more efficient way to achieve this goal.

Suppose matrices $Q = [\alpha_1, \dots, \alpha_r]$ and $\overline{Q} = [\alpha_{r+1}, \dots, \alpha_d]$. Since $R^d = V \oplus V^\perp$, every $x_n^i \in R^d$ has a unique decomposition of the form

$$x_n^i = QQ^T x_n^i + \overline{Q}\overline{Q}^T x_n^i = y_n^i + z_n^i \quad (\text{Eq. 3.4})$$

where $y_n^i \in V$, $z_n^i \in V^\perp$ and z_n^i is just the common vector of the i th class.

To calculate z_m^i , we first do eigenanalysis of S_W to find the eigenvectors α_k corresponding to the nonzero eigenvalues of S_W . Then the common vector can be computed as

$$x_{com}^i = z_n^i = x_n^i - QQ^T x_n^i (n = 1, \dots, N_i, i = 1, \dots, C) \quad (\text{Eq. 3.5})$$

Actually, it turns out that we obtain the same common vector for all samples of each class [135]. And the information in the null space is reported to be more discriminant than that in the feature space [175].

Figure 3.3: Feature weight comparison. For better visualization, histograms of each weight are calculated which are shown on the right side of each weight.

3.2.2 Dynamic Feature Selection

Given the normalized DCVs, a new feature weight is proposed to record the between-class dynamic variations. Mathematically, it is calculated as

$$F(l) = \sqrt{\frac{1}{C} \sum_{i=1}^C (x_{com}(l) - \mu_{com}(l))^2} \quad l = 1, \dots, d, \quad (\text{Eq. 3.6})$$

where μ_{com} is the mean of all common vectors and l represents the l th element in the d by 1 column vector. Obviously, this feature weight is sample independent.

Then $F(l)$ is normalized to the range of $[0, 1]$. We show in Figure 3.3(b) our proposed feature weight along with the feature weight proposed in [108] (Figure 3.3(a)) which is obtained by taking variance on GEI directly. From the figure we can see that the contour and the bottom part of the legs are both emphasized in the two feature weights. This means both the body shape and the dynamic movement are important for recognition. It is also reasonable to give smaller weight to the black area inside the body contour because this part moves little during walking. Nevertheless, the histograms of the two weights are shown on the right side of each figure separately. Through the histograms we can conclude that the proposed weight suppresses more pixels than the variance weight. Although the number of pixels selected by the

Figure 3.4: Illustration of feature selection mask and the new proposed feature selection weight.

proposed weight is smaller, the subsequent experimental results will show that these features are more discriminative for recognition.

Noted that the feature weight only records the between-class discriminative dynamic information, we also introduce the feature selection mask which is generated by maintaining the pixels whose intensity values in the mean GEI of the same class is smaller than a predefined threshold, that is, the pixel value of the mask is set to 1 only if its corresponding pixel value in GEI is lower than the threshold:

$$M(x, y) = \begin{cases} 0, & \text{where } G(x, y) \geq \theta \\ 1, & \text{otherwise} \end{cases} \quad (\text{Eq. 3.7})$$

where θ is the predetermined threshold and $G(x, y)$ is the mean GEI of the same class. Obviously, the feature selection mask is subject dependent and thus it only selects the within-class dynamic information. Meanwhile, in the recognition stage, the test sample would have its own mask and the final mask for each testing and training sample pair should be the combination of their corresponding masks.

The above two feature selection criteria select dynamic feature from different points of view, therefore, we propose to combine them to form a new feature which could ex-

Figure 3.5: Different features extracted from GEI

tract both the within-class and between-class dynamic information. The combination is defined as:

$$D(x, y) = M(x, y) \bullet (F(x, y))^\lambda \quad (\text{Eq. 3.8})$$

where $F(x, y)$ is obtained by reshaping 1D $F(l)$ to 2D $F(x, y)$ as the same dimension as $G(x, y)$ and the parameter λ is added to adjust the influence of the weight to the new feature.

Both the feature selection mask and the combined dynamic feature weight are shown in Figure 3.4. Obviously, with the combination of the feature selection mask, the effect from the shadow is alleviated.

3.2.3 Spatial Dynamic Feature Generation

By applying the three feature selection rules, binary mask $M(x, y)$, the feature weight $F(x, y)$ and their combination $D(x, y)$, to GEI, we can obtain different dynamic features as follows:

$$FGEI(x, y) = (F(x, y))^\lambda \bullet GEI(x, y), \quad (\text{Eq. 3.9})$$

$$MGEI(x, y) = M(x, y) \bullet GEI(x, y), \quad (\text{Eq. 3.10})$$

$$DGEI(x, y) = D(x, y) \bullet GEI(x, y), \quad (\text{Eq. 3.11})$$

Figure 3.6: Flow chart of spatial-temporal dynamic feature extraction.

The results $FGEI(x, y)$, $MGEI(x, y)$, $DGEI(x, y)$ are defined as FGEI, MGEI, DGEI respectively. All the three features are shown in Figure 3.5(b)-(d). Figure 3.5(e) is the dynamic feature generated by the feature weight in Figure 3.3(a).

The feature weight regards the dynamic parts in gait as the parts that have larger between-class variance, that is, the selected dynamic features are dynamic between classes and they are class independent. However, the dynamic feature selection mask regards the dynamic parts in a GEI image as the region in which the pixel's intensity value is smaller than a pre-defined threshold, that is, the selected dynamic parts are dynamic between frames in a walking sequence. Thus, it is sample dependent. These two rules select the dynamic features from different ways. Therefore, the combination of could achieve better performance because both the within-class dynamic information and the between-class dynamic information are selected.

3.3 Spatial-Temporal Dynamic Feature Analysis

We illustrate the flow chart of our proposed spatial-temporal dynamic feature extraction scheme in Figure 3.6. Given a gait sequence, its period is first detected. Then each gait frame is weighted by the learned feature weight and colored according to the gait pose in the cycle. At last, all the frames in a gait cycle are averaged to derive a compact representation, which is the proposed spatial (feature weight) -temporal (color) dynamic feature.

Figure 3.7: (a) width vector of a gait sequence (b) the corresponding autocorrelation.

3.3.1 Temporal Information by Color

To preserve the temporal information by color, we need to detect the period of the gait sequence first. We use a simple and common approach [12] to detect the gait cycle for all sequences. It regards the number of foreground pixels from the bottom half of each silhouette as a periodical vector and uses the autocorrelation of the number vector to calculate the period.

Suppose we obtain the number vector $W = (w_1, w_2, \dots, w_n)$ from a gait sequence, we calculate its autocorrelation $C = (c_1, c_2, \dots, c_r)$ as follows:

$$c_r = \frac{1}{n} \sum_{k=1}^{n-r} (w_k - \mu)(w_{r+k} - \mu) \quad (\text{Eq. 3.12})$$

where μ is the mean of W , and then assign twice the width between two nearest maximum values as the gait cycle for the sequence. Noise in the binary silhouette can easily affect the number vector, therefore, we choose the autocorrelation of the vector because the curve of autocorrelation is more smooth than the original one, see Figure 3.7.

Suppose we use the RGB color image, after we obtain a complete gait cycle, we also need to know what the stage of current frame is in a gait cycle. Therefore, we define $p_t = (w_t - w_{min}) / (w_{max} - w_{min})$ (w_t is the number of foreground pixels in bottom half rows of the t th frame, w_{min} and w_{max} are the extreme values of w within the $1/4$ period which the t th frame belongs to, please see Figure 3.7 for the change of w with time t) as the variable that indicates the gait pose and design a interpolation function to calculate the three color components for each of the frame in the gait period according to p_t . For efficiency, a liner interpolation function is applied whose three color components (R=Red, G=Green, B=Blue) are defined as follows:

$$R_{p_t}(x, y) = \begin{cases} 0, & p_t \leq 1/2 \\ (2p_t - 1), & p_t > 1/2 \end{cases} \quad (\text{Eq. 3.13})$$

$$G_{p_t}(x, y) = \begin{cases} 2p_t I, & p_t \leq 1/2 \\ (2 - 2p_t), & p_t > 1/2 \end{cases} \quad (\text{Eq. 3.14})$$

$$B_{p_t}(x, y) = \begin{cases} (1 - 2p_t), & p_t \leq 1/2 \\ 0, & p_t > 1/2 \end{cases} \quad (\text{Eq. 3.15})$$

where t represents the frame number, (x, y) are the image coordinates. We also visualize the above interpolation functions in Figure 3.8. For each frame $B_t(x, y)$ with gait pose (p_t), its three color components can be calculated by multiplying itself with the above interpolation functions.

3.3.2 Spatial-Temporal Dynamic Feature Generation

To apply color to visualize time-varying signals, several integration functions can be used. However, due to the serious overlap between gait frames, these functions can't be applied directly because the resulting images would lose much important information [43]. Therefore, we propose to first use the learned feature weight to extract spatial dynamic information from each frame to solve this problem. Then the colored gait

Figure 3.8: Visualization of the interpolation functions in Eq. 3.13 ~ Eq. 3.15.

silhouette image STD_t of the t th frame is calculated as follows:

$$STD_t(x, y) = \begin{pmatrix} F(x, y)^\lambda * R_{p_t}(x, y) \\ F(x, y)^\lambda * G_{p_t}(x, y) \\ F(x, y)^\lambda * B_{p_t}(x, y) \end{pmatrix} * B(x, y). \quad (\text{Eq. 3.16})$$

And the proposed spatial-temporal dynamic feature of a given gait sequence is generated as:

$$STDGI(x, y) = \frac{1}{n} \sum_{t=1}^n STD_t(x, y) \quad (\text{Eq. 3.17})$$

The results of each step to generate STDGI are shown in Figure 3.9. The first row shows seven silhouettes in a $1/4$ gait period. And the second row shows the results of applying the feature weight $F(x, y)$ on the silhouettes above, and the rightmost image in the same row is the average of all weighted frames in the gait cycle. In fact, this is equal to apply $F(x, y)$ on GEI directly (FGEI): $F(x, y)^\lambda * GEI(x, y) = F(x, y)^\lambda * \frac{1}{n} \sum_{t=1}^n B_t(x, y) = \frac{1}{n} \sum_{t=1}^n F(x, y)^\lambda * B_t(x, y)$. The third row shows the colored and weighted frames and rightmost image is the proposed STDGI which is also the mean of all processed frames in a gait cycle. It can be seen from the figure that STDGI has better visualization than FGEI.

Figure 3.9: Illustration of the spatial-temporal dynamic feature generation.

3.4 Feature Space Dimension Reduction

We choose linear discriminant analysis (LDA) [110] for feature space dimension reduction. Given a sample set $\mathbf{X} = [\mathbf{x}_1^1, \mathbf{x}_1^2, \dots, \mathbf{x}_1^{N_1}, \dots, \mathbf{x}_i^j, \dots, \mathbf{x}_C^{N_C}]$ which contains $N(= N_1 + N_2 + \dots + N_C)$ samples from C classes, where $\mathbf{x}_i^j$ is the j th sample in the i th class, N_i is the total number of samples in the i th class. LDA aims to find a low dimensional feature space defined by the projection matrix ϕ , in which the inter-class distances are maximized and intra-class distances are minimized. Mathematically, it leads to maximize the following:

$$J(\phi) = \frac{\phi^T \mathbf{S}_b \phi}{\phi^T \mathbf{S}_w \phi} \quad (\text{Eq. 3.18})$$

where $\mathbf{S}_b$ is the between-class scatter matrix and $\mathbf{S}_w$ is the within-class scatter matrix, which are defined as follows:

$$\mathbf{S}_b = \sum_{i=1}^C n_i (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^T \quad (\text{Eq. 3.19})$$

and

$$\mathbf{S}_w = \sum_{i=1}^C \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)^T \quad (\text{Eq. 3.20})$$

where $\bar{\mathbf{x}}_i$ is the mean samples of the i th class and $\bar{\mathbf{x}}$ is the mean sample of all the training samples. The optimization problem defined in Eq. 3.18 can be deduced to a generalized eigenvalue problem as follows:

$$\mathbf{S}_b \phi = \Sigma \mathbf{S}_w \phi \quad (\text{Eq. 3.21})$$

Noted that for LDA, the maximal number of discriminative vectors(columns of ϕ) is at most $C - 1$ due to the rank limit of $\mathbf{S}_b$.

3.5 Classification

In the space defined by LDA, Euclidian distance is applied as the similarity measure metric. Given the similarities among all the testing and training samples, nearest neighbor classifier is first used to assign identity to each testing sample. Then, especially for experiments on the USF database, voting is further utilized to obtain the identity of each gait sequence because in the USF database, each gait sequence is divided into several gait cycles according to the gait period.

3.6 Experiments

3.6.1 Database

We conducted our experiments on two gait databases (the CASIA-B gait database [17] and the USF gait database (version 2.1)[12]). We illustrate in Figure 3.10 some example GEIs from the two databases. The CASIA database was obtained in an indoor environment with simple background, thus preprocessing noises was not significant. It contained different walking conditions and it was used for feature descriptor comparison. While the USF HumanID database was a large enough outdoor database taken

Figure 3.10: Illustration of example GEIs obtained from the two chosen databases.

under different walking conditions as well as background, hence preprocessing noises were difficult to remove and we can see from the figure that shadows remain in GEI. As a result, we will not consider the last few rows in period detection on this database. Moreover, it is appropriate to validate the robustness of different feature descriptors.

3.6.2 Parameter Optimization

For our proposed feature descriptors, there are two parameters that need to be initialized, one is the parameter θ in the feature selection mask (Eq. 3.7) and the other is the parameter λ in the new feature weight (Eq. 3.8). In our experiments, we tuned them empirically. Figure 3.11 shows the effect of different parameter values on the final recognition results (mean results from the three views in the CASIA-A database). From the figure we can see that both parameters affect the final recognition results slightly. In our experiments, we used the optimal values as shown in the figure, that is, $\theta = 237$ and $\lambda = 0.6$ respectively.

3.6.3 Results on CASIA-B Database

In the experiments, we chose 4 out of 6 normal walking sequences from view 90° of each subject as training set. We had 3 testing sets named set A to C whose views were the same with that of the training set. Set A contained the left 2 normal walking

Figure 3.11: Illustration of the effect from the design parameters on the final recognition results. (a) θ in Eq. 3.7 and (b) λ in Eq. 3.8.

sequences from each subject, set B contained all the 2 walking sequences carrying a bag and set C was formed by all the 2 walking sequences wearing a coat.

Several gait feature descriptors were chosen for comparison. Gait History Image (GHI) [42], Active Energy Image (AEI) [136] and Chrono-Gait Image (CGI) [43] are the descriptors that incorporate temporal information into the representation; EGEI [108] and MGEI [109] extract dynamic feature from GEI to reduce information redundancy. FGEI, DGEI and STDGI are our proposed new feature descriptors among which FGEI and DGEI are spatial dynamic features and STDGI is spatial-temporal dynamic feature. We show in Figure 3.12 the examples of these feature descriptors on the two databases respectively. Clearly, the descriptors that retain the temporal information have much difference from those without such information.

Table 3.1: Experimental results (%) using original feature descriptors on CASIA-B database.

Method	GEI	GHI	AEI	CGI	EGEI	MGEI	FGEI	DGEI	STDGI
Set A	93.95	80.08	84.55	91.87	95.97	95.56	95.97	95.97	96.37
Set B	38.71	41.87	52.44	54.44	41.53	39.92	42.74	46.37	57.26
Set C	20.97	30.08	32.49	45.12	24.60	23.39	28.23	31.05	47.18

Figure 3.12: Illustration of the selected feature descriptors on example sequence of the two databases. Samples in the first row are from CASIA-B database and those in the second row are from USF database.

Table 3.2: Experimental results (%) using original feature descriptors and LDA on CASIA-B database.

Method	GEI	GHI	AEI	CGI	EGEI	MGEI	FGEI	DGEI	STDGI
Set A	98.79	94.72	96.75	96.34	99.19	98.39	99.19	98.79	99.19
Set B	55.65	57.72	63.82	69.51	58.87	56.85	64.11	65.73	71.37
Set C	44.35	41.46	50.41	58.13	48.39	46.77	50.81	54.03	60.89

The gait recognition results using these feature descriptors on CASIA-B database are shown in Table 3.1. Obviously, walking condition variations would degrade the recognition results greatly. As we have mentioned before, the popular feature descriptor GEI has two limitations, information redundancy and loss of temporal information. Results in Table 3.1 clearly demonstrate that efforts made to alleviate either of the two limitations are effective especially under walking condition variations. We can see that AEI and CGI outperform GEI by nearly 15% under the walking condition of carrying a bag, CGI outperforms GEI by more than 20% under the walking condition of wearing a coat. While the improvements achieved by reducing information redundancy are not so much. The best one (DGEI) outperforms GEI by nearly 10% under both walking condition variations. This indicates that the temporal information is very important

in the recognition tasks. Moreover, retaining the temporal information by color is the most effective way because CGI performs better than both GHI and AEI. Our proposed spatial-temporal dynamic feature which attempts to alleviate both limitations simultaneously performs the best among all the descriptors, its recognition rates are higher than those of GEI by nearly 20% under the two walking condition variations.

Results in Table 3.2 are achieved by using the above feature descriptors as well as LDA. By introducing the discriminative subspace learning techniques, we can see that the recognition results are enhanced significantly and the gap of recognition power between the two kinds of descriptors no longer exists. The superiority of our proposed descriptors is clear compared with corresponding descriptors. That is, for FGEI which adopts the proposed discriminative feature weight, it outperforms EGEI whose weight is based on unsupervised learning; by combining the feature weight with feature selection mask, the proposed DGEI performs the best among all spatial dynamic features. At last, the spatial-temporal dynamic feature STDGI which takes the advantages of both kinds of descriptors performs the best among all features considered.

3.6.4 Results on USF Database

As we have introduced in Chapter 2, the USF HumanID outdoor gait database (version 2.1) is divided into one gallery set and 12 probes labeled from “A” to “L”. As the walking conditions and the recognition performance are different among the 12 probes, we also show the mean recognition rate at the last column to evaluate the efficiency of the feature descriptors. Table 3.3 shows the experimental results which are divided into two parts: the upper part shows the results of different feature descriptors that aim at alleviating the previously mentioned limitations from GEI while the bottom part are the results cited from relevant papers which propose some popular algorithms to address the problem of varying walking conditions.

Table 3.3: Experimental results (%) using original feature descriptors and LDA on USF database.

Probe	A	B	C	D	E	F	G	H	I	J	K	L	Mean
GEI	89	87	78	31	33	19	24	50	52	51	6	12	43.67
GHI	88	89	74	27	38	21	23	76	62	62	3	6	47.42
AEI	88	87	76	32	30	23	23	83	65	61	6	12	48.83
CGI	90	89	81	29	32	21	23	82	76	63	9	9	50.33
EGEI	88	87	83	33	40	25	26	56	53	60	6	9	47.25
MGEI	88	86	81	33	37	20	26	67	60	63	6	6	47.75
FGEI	90	87	74	36	41	27	28	71	70	62	9	15	50.83
DGEI	91	89	83	37	40	23	27	86	78	68	12	15	54.0
STDGI	91	91	85	43	48	26	27	88	82	67	12	15	56.25
GEI_{fusion} [29]	90	91	81	56	64	25	36	64	60	60	6	15	54.0
$STDGI_{fusion}$	90	91	86	48	55	28	32	87	83	65	18	15	57.67
$P_{GaussianIm}$ [179]	82	78	76	48	54	56	55	-	-	-	-	-	-
pHMM[55]	85	89	72	57	66	46	41	83	79	52	15	24	59.08

From the results shown in the upper part of Table 3.3 we can conclude that, firstly, all the selected feature descriptors which are developed based on GEI achieve a higher recognition rate than the original GEI. The proposed STDGI achieves the highest mean recognition rate which is higher than that of GEI by more than 10%. Secondly, for performance on the probe sets D, E, F and G which have different walking surfaces compared with the gallery set, the spatial dynamic features (EGEI, MGEI, FGEI, DGEI) which aim to reduce the information redundancy in GEI, perform better than the temporal descriptors (GHI, AEI, CGI) which attempt to retain the temporal information ignored by GEI. However, they perform a little worse on probe sets H and I in which the subjects carried a briefcase. These demonstrate that the spatial dynamic features are robust to walking surface variations while temporal dynamic features are robust to carrying conditions.

When we compare the proposed method with those shown in the bottom part of Table 3.3, we realized that there are several ways to improve the performance of proposed method. Firstly, due to training sample limitation, Han et al. [29] proposed

to generate synthesis training sequences in which different number of the bottom rows are eliminated. We can clearly see from the table that the fusion of real and synthesis images greatly improves the recognition rate by almost 10% on average, especially for the probe sets D and E which have variations in walking surfaces and shoes. According to their methods, we also generated similar synthetic STDGI features and obtained the corresponding recognition results shown after their method. Although the overall performance after fusion is enhanced, the improvements under walking surface variation are not as significant as those in [29]. This may be due to the fact that the resize of image has a large effect on the alignment of RGB images. Moreover, such synthetic generation is not generic and database dependent. Secondly, the rate-invariant recognition method $P_{GaussianIm}$ proposed in [179] also achieves comparable results with ours for probe sets A to E. However, it outperforms STDGI by almost 20% for probe sets F and G. The reason may be that it uses a leave-one-out training process which increases the variations among the training sets and thus makes the algorithm robust to different kinds of variations within the database. We have also adapted such leave-one-out training process to our proposed STDGI feature but the performance has not been satisfying. This demonstrates that the distribution of samples containing multiple variations is not typical Gaussian (LDA works under the assumption that the distribution of samples from the same class is Gaussian). Last but not least, the popular model-based method pHMM[55] achieves the best average recognition results among all the algorithms listed. However, it also performs a different training process from ours. It selects a subset for training from both the gallery set and all the probe sets which makes it robust to variations within the database. Despite of this, we note that the last two methods show us a potential direction for solving walking condition variations, i.e., increasing the variations among the training samples. Further study is required to develop new subspace learning method that can adapt to the variations in training samples.

3.6.5 Further Discussion

It has been proved that both the body shapes as well as the gait dynamics play an important role in gait recognition. Therefore, walking condition variations that affect these two factors will degrade the recognition performance. We can see from the experimental results shown previously that carrying a bag or wearing a short coat can reduce the recognition rates nearly by half compared to those achieved under normal walking. Even the proposed descriptor (which performs the best among comparable descriptors) recognizes about less 30% subjects than that in normal walking condition. Hence, it is reasonable to assume that wearing a long coat that hides the legs would further degrade the recognition results because both the body shapes and the gait dynamics are almost covered. This should be one of the extreme walking condition variations that are difficult to solve.

Due to database limitation, we can't provide the recognition results under the condition of wearing a long coat. However, we design a similar experiment to show the effect of wearing coat on the recognition performance. We only use the upper body (upper half of each frame) to perform the recognition under two conditions: without coat and with coat (short coat) and the results are shown in Table 3.4.

Table 3.4: Experimental results (%) using the upper half frames in CASIA-B database.

Features	GEI	GHI	AEI	CGI	STDGI
Upper body without coat	91	85	84	93	95
Upper body with coat	17	16	14	19	21

From the results shown in Table 3.4 we can see that all the selected descriptors nearly fail in the designed experiments with the recognition rates at only 15% $\sim$ 20%. This means that new technologies are in demand to solve such extreme walking condition variations.

3.6.6 Summary

From the results showed above, we can draw the following conclusions:

- (1) There are two directions to improve the GEI descriptor, incorporating temporal information and reducing information redundancy. Among the spatial dynamic features EGEI, MGEI and FGEI, which aim to reduce the information redundancy in GEI, our proposed FGEI performs the best. This demonstrates that common vectors contain more important discriminative information than GEI for the recognition task. Moreover, the combination of the feature selection mask as well as the feature weight, which results in the proposed new feature DGEI, performs better than the above three features because it contains both between-class dynamic information and within-class dynamic information.
- (2) The temporal templates, GHI, AEI and CGI, belong to the other direction which attempts to incorporate temporal information in the feature representation. Among the three templates, CGI achieves the highest recognition rates under different walking conditions. This proves that applying color to retain temporal information is very effective.
- (3) Both kinds of descriptors have their own advantages. For spatial dynamic features, they are robust when the walking surface changes, while for the temporal templates, they perform better under wearing or carrying condition variation.
- (4) By combining the two kinds of descriptors, that is, using the discriminative feature weight as well as color to obtain both spatial and temporal dynamic information, the proposed STDGI feature performs the best among the selected descriptors. It is also robust under walking condition variations.

3.7 Conclusion

In this chapter, we investigate the problem of how to effectively represent a gait sequence so that it is robust under walking condition variations. Noted that the popular GEI feature is sensitive to walking condition variation because it has two main limitations: information redundancy and loss of temporal information, we propose two dynamic feature extraction schemes to alleviate these limitations. To reduce the information redundancy, we propose a new spatial dynamic feature selection scheme which is the combination of our proposed feature weight based on DCVs and the feature selection mask. As the two select dynamic feature from different point of views, their combination can record both within-class and between-class dynamic information. To incorporate temporal information into the representation, we also propose a new spatial-temporal dynamic feature which is obtained by first applying the learned feature weight to each frame of the gait sequence and then coloring it according to its gait pose. At last, the proposed feature is obtained by averaging all frames in a gait cycle. Experimental results on two popular gait databases have demonstrated the effectiveness of the proposed dynamic features and showed that they are robustness to walking condition variations.

Chapter 4

Gait Recognition Across Varying Views

Most of the existing motion-based gait recognition methods are view-dependent which means that the capturing views of the testing sequences must be the same with that of the training sequences. However, in real applications, the subjects can walk through any view direction in front of the camera and it is impossible for us to store gait samples in all possible directions. Therefore, how to deal with view variation between training and testing sequences becomes an important research problem. The spatial-temporal dynamic feature analysis proposed in previous chapter is not suitable for the view variation problem because gait dynamics are quite different under different views. Therefore in this chapter, we would address this problem from another point of view and propose new solutions to solve it.

4.1 Introduction

Gait recognition across varying views is a challenging problem and has been a bottleneck in automatic human gait analysis [11, 16, 70, 72, 73, 74, 75, 168]. In recent years, several approaches have been proposed to address this problem and they can be classified into two categories: model-based [70, 72] and motion-based [73, 74, 168]. In the

model-based methods, a statistic or generic model is usually applied to characterize the dynamic of human movements to attain view-invariance to some extent. For example, Shakhnarovich *et al.*[70] proposed to synthesize virtual images in arbitrary views from a visual hull for gait recognition. Bodor *et al.*[72] learned a three-dimensional (3D) model of the walking person from samples acquired in four different views and used image-based rendering techniques to reconstruct gait to arbitrary target views. Although these methods can achieve relatively high recognition rates under large view variation, they usually require a well-calibrated multi-camera system and are computationally expensive. Therefore, many researchers in this area have also resorted to motion-based methods recently.

For motion-based approaches, view-invariant features are usually sought by some image processing or statistical learning techniques for recognition. For example, Goffredo *et al.* [73] proposed a markerless joint motion system based on 2D silhouettes of human gait sequences, and rectified these features across different views to achieve view-invariance. While reasonably good recognition rates can be attained for large view variation, its performance degrades under a small view variation compared with some other algorithms. Moreover, it can't handle extreme views. Lu and Tan [74] proposed to project gait features extracted from two different views into a low-dimensional subspace, so as to preserve the intra-class geometrical structures and maximize the interclass distances simultaneously. However, the performance still leaves much to be desired when the view variation is relatively large.

Recently, the view transformation approaches have been proposed for the view-invariant recognition task [64, 79, 166] and we have reviewed them in Chapter 2. The basic idea is to transform a gait feature from one view to another by learning a view relationship between the two corresponding views, and subsequently the transformed virtual feature is used for recognition. However, no matter it is in the pixel level or

the entire gait level, the transformation of a feature to another view could generate large errors and thus compromise the discriminative information, both of which would degrade the recognition performance.

Different from the above mentioned methods, we propose in this chapter a new view-invariant feature for gait recognition across different views. Our idea is inspired by the fact that if a three-dimensional (3D) object can be well represented by a linear combination of a small number of prototypes from the same view, then the representation coefficients with the same prototypes remain fairly similar across different views [165]. Therefore, we propose to use these representation coefficients as view-invariant features for recognition. To obtain the coefficients, we first conduct joint subspace learning (JSL) using Canonical Correlation Analysis (CCA) and Principle Component Analysis (PCA) to obtain the prototypes of different views. Then we represent the samples in both the gallery set and the probe set acquired from different views as a linear combination of these prototypes in the corresponding views and extract the coefficients for recognition. In addition, we also propose a new gait feature descriptor, Radon transform based Energy Image (REI), and divide it into patches to further enhance the performance. Experimental results on the widely used CASIA-B gait database are presented to demonstrate the effectiveness of the proposed method.

The rest of this chapter is organized as follows. Firstly, the proposed new gait feature descriptor is introduced, secondly, we illustrate in detail the proposed joint subspace learning (JSL) approach for view-invariant gait recognition, thirdly, extensive experimental results are presented to demonstrate the effectiveness of proposed approach, at last, conclusion of this chapter is provided.

Figure 4.1: Illustration of the line defined by (ρ, θ) (left), the ρ - s coordinate system (middle) and the result of the transform (right).

4.2 Gait Feature Descriptor

Motivated by the fact that radon transform is a competent descriptor for view-invariant applications [113, 114, 115], we propose a new gait feature descriptor based on radon transform [112] for the task of gait recognition across varying views. Radon transform has many different forms and a most popular one is defined as follows:

$$R(\rho, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \delta(\rho - x \cos \theta - y \sin \theta) dx dy \quad (\text{Eq. 4.1})$$

where $f(x, y)$ is a 2D function, and $R(\rho, \theta)$ is the integral of f along the line $\rho = x \cos \theta + y \sin \theta$ defined by two parameters (ρ, θ) . The illustration of the line and the new coordinate system is shown in Figure 4.1(a) and 4.1(b) separately. In our work, the 2D function is the binary silhouette $B(x, y)$ and its corresponding radon transform is in the discrete form which is the summation of pixel intensities along the lines of different directions. The center of the silhouette is defined as the reference point. Given a specific value θ_i which determines the direction of ρ -axis, the radon transform records the summation of pixel intensities along the lines that parallel to the s -axis separately according to different ρ_i . Obviously, radon transform goes through the entire silhouette

Figure 4.2: The flowchart of the proposed approach.

for each value of θ , that is, Radon transform depicts the binary silhouette delicately from different directions. The result radon transform of Figure 4.1(b) is shown in Figure 4.1(c).

We normalize the result from Eq. 4.1 based on the shape area, thus to alleviate the effects caused by view variation to some extent. The normalization step goes mathematically as follows:

$$R_N(\rho, \theta) = \frac{R(\rho, \theta)}{S} \quad (\text{Eq. 4.2})$$

where $S = \int_{\rho} R(\rho, \theta_i) d\rho$ is the shape area. Then given a gait sequence of M frames, the new feature descriptor can be calculated as

$$REI(\rho, \theta) = \sum_{t=1}^M R_{Nt}(\rho, \theta) \quad (\text{Eq. 4.3})$$

We call it Radon transform-based Energy Image (REI). It is robust to noise because each of its pixel value is obtained through averaging in both spatial and temporal domain. Moreover, by defining the center of the silhouette as the reference point, REI is translation and scale invariant. The effectiveness of the proposed REI feature in gait recognition across view variations will also be proved in later experiments.

4.3 Joint Subspace Learning (JSL)

Figure 4.2 shows the flowchart of our proposed approach, which mainly consists of two phases: training and recognition. In the training phase, we learn the prototypes in different views by applying JSL on an independent training set. In the recognition phase, we represent each sample in both the gallery and probe sets by linearly combining these prototypes under the corresponding views and use the representation coefficients as the view-invariant features for recognition.

4.3.1 Motivation

We show in Figure 4.3 the gait appearance features (GEI [29]) of two different subjects under different views. Apparently, intra-subject differences under large view variation are much larger than inter-subject differences of the same view. That’s why gait recognition across large view variation is still challenging and remains to be solved. In this chapter, we attempt to solve this problem by seeking a common representation other than the appearance feature for the same subject across different views and such representation should be discriminative among different subjects. The feasibility is evidenced from the property of linear object class (LOC) [165], which shows that if a 3-D view of an object can be represented as the weighted sum of views of other objects, then its linearly rotated view is a linear combination of the linearly rotated views of the other objects with the same weights. Moreover, Vetter and Poggio showed that many 2D shape images satisfy such property. Let $\mathbf{x}^1$ and $\mathbf{x}^2$ be the 2D representations of the same gait sequence under two different views. Accordingly, we should have the following derivation:

$$\mathbf{x}^1 = \sum_{i=1}^m \alpha_i \mathbf{u}_i^1 \implies \mathbf{x}^2 = \sum_{i=1}^m \alpha_i \mathbf{u}_i^2 \quad (\text{Eq. 4.4})$$

where $\mathbf{u}_i^1$ and $\mathbf{u}_i^2$ are the representations of the same prototype under the two different views (Here, the prototype is referred to the representative basis of certain LOC. That

Figure 4.3: Illustration of gait appearances of two different subjects under different views.

is to say, any sample in the same LOC can be represented as a linear combination of the prototypes). The reconstruction coefficients (α_i) are the common representations that we seek—representations that are consistent across intra-subject view variations and discriminative across different subjects. We illustrate in Figure 4.4 the idea indicated by LOC. According to LOC, we would have $\alpha^1 \simeq \alpha^2$ if the gait images on the right side of the equation are prototypes under corresponding views respectively. Clearly, the known prototypes under different views are crucial for such representation. However, in real application, the prototype U is difficult to obtain. Therefore, we need to first learn the prototypes of different views, and then extract the representation coefficients for recognition.

4.3.2 Training

We need an independent training set to learn the prototypes under different views. One advantage of our proposed algorithm is that it allows samples from multiple different views. Here, we first use gait sequences from two different views as an example and then extend it to multiple views. Given sequence representation matrices $\mathbf{X}^1 = [\mathbf{x}_1^1, \mathbf{x}_2^1, \dots, \mathbf{x}_n^1]$ and $\mathbf{X}^2 = [\mathbf{x}_1^2, \mathbf{x}_2^2, \dots, \mathbf{x}_n^2]$ acquired in two different views, where $\mathbf{x}_i^1$ and

Figure 4.4: Illustration of the idea indicated by LOC. For each equation, the left hand side is the GEI under a certain view and the right hand side are prototypes under the same view (GEI is used for illustration here because it is better than REI for visualization of view variation).

$\mathbf{x}_i^2 (i = 1, 2, \dots, n) \in \mathbb{R}^d$ are the respective REI of the i th gait sequence under the two different views, and n is the total number of training samples. Our objective of extracting prototype matrices U^1, U^2 in the two views for common representations is formulated as follows:

$$\begin{aligned} (U^1, U^2) &= \arg_{(U^1, U^2)} \min \frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i^1 - U^1 \alpha_i\|_2^2 \\ &\quad + \frac{\lambda}{n} \sum_{i=1}^n \|\mathbf{x}_i^2 - U^2 \alpha_i\|_2^2 \\ &= \arg_{(U^1, U^2)} \min \frac{1}{n} \|\mathbf{X}^1 - U^1 F\|_2^2 + \frac{\lambda}{n} \|\mathbf{X}^2 - U^2 F\|_2^2 \end{aligned} \quad (\text{Eq. 4.5})$$

where $\alpha_i \in \mathbb{R}^m$ is the representation coefficients for both $\mathbf{x}_i^1$ and $\mathbf{x}_i^2$, $F = [\alpha_1, \alpha_2, \dots, \alpha_n]$ is the coefficients matrix, $m (< n)$ is the number of prototypes in each view. Before deducing the solution for Eq. 4.5, we propose to maximize the correlation between the sample pairs in $\mathbf{X}^1$ and $\mathbf{X}^2$ first because the data acquired usually have some distortion due to noises, which may cause variation in the representation. Specifically, we apply CCA [171] to find the projection matrices W_{cca}^1 and W_{cca}^2 for $\mathbf{X}^1$ and $\mathbf{X}^2$, respectively, so that the projections of the two matrices in the new subspace are maximally correlated.

Mathematically, we need to solve the following optimization problem:

$$(W_{cca}^1, W_{cca}^2) = \arg_{(W_{cca}^1, W_{cca}^2)} \max \frac{E[(W_{cca}^1)^T \mathbf{X}^1 (\mathbf{X}^2)^T W_{cca}^2]}{\sqrt{E[(W_{cca}^1)^T \mathbf{X}^1 (\mathbf{X}^1)^T W_{cca}^1] E[(W_{cca}^2)^T \mathbf{X}^2 (\mathbf{X}^2)^T W_{cca}^2]}} \quad (\text{Eq. 4.6})$$

which can also be written in terms of covariance matrices as:

$$(W_{cca}^1, W_{cca}^2) = \arg_{(W_{cca}^1, W_{cca}^2)} \max \frac{(W_{cca}^1)^T \mathbf{C}_{X^1 X^2} W_{cca}^2}{\sqrt{(W_{cca}^1)^T \mathbf{C}_{X^1 X^1} W_{cca}^1 E[(W_{cca}^2)^T \mathbf{C}_{X^2 X^2} W_{cca}^2]}} \quad (\text{Eq. 4.7})$$

Subsequently, Eq. 4.5 becomes:

$$\begin{aligned} (U^1, U^2) = \arg_{(U^1, U^2)} \min \frac{1}{n} \|(W_{cca}^1)^T \mathbf{X}^1 - (W_{cca}^1)^T U^1 F\|_2^2 \\ + \frac{\lambda}{n} \|(W_{cca}^2)^T \mathbf{X}^2 - (W_{cca}^2)^T U^2 F\|_2^2 \end{aligned} \quad (\text{Eq. 4.8})$$

Let $\mathbf{D} = \begin{bmatrix} (W_{cca}^1)^T \mathbf{X}^1 \\ \sqrt{\lambda} (W_{cca}^2)^T \mathbf{X}^2 \end{bmatrix}$ and $V = \begin{bmatrix} (W_{cca}^1)^T U^1 \\ \sqrt{\lambda} (W_{cca}^2)^T U^2 \end{bmatrix}$. The objective function can be further expressed as:

$$V = \arg_V \min \frac{1}{n} \|\mathbf{D} - VF\|_2^2 \quad (\text{Eq. 4.9})$$

To ensure linear independency, we assume that the new basis $v_i \in \mathbb{R}^{2d}$ ($i = 1, 2, \dots, m$) are orthonormal to each other. The objective of Eq. 4.9 is to find a linear combination of these bases to approximate $\mathbf{D}$ while minimizing the mean square error. This implies that the coefficients in F should be the projection of $\mathbf{D}$ onto corresponding basis vectors. Thus, based on the orthonormality we have:

$$V = \arg_V \min \frac{1}{n} \|\mathbf{D} - VV^T \mathbf{D}\|_2^2 = \arg_V \min \frac{1}{n} (\mathbf{D}\mathbf{D}^T - V^T \mathbf{D}\mathbf{D}^T V) \quad (\text{Eq. 4.10})$$

This is equivalent to the objective of PCA, $V = \arg_V \max V^T \mathbf{D}\mathbf{D}^T V$. That is, the optimal basis matrix V is formed by the first m eigenvectors of $\mathbf{D}\mathbf{D}^T$ with largest eigenvalues. However, it has a little difference from PCA as the feature vector in the data matrix $\mathbf{D}$ is formed by the concatenation of the projected sample pairs rather than the original samples. Therefore, we give it a new name Joint PCA (JPCA)

and term the whole training process as Joint Subspace Learning (JSL). We split the optimal solution V to obtain the solution for prototypes U^1 and U^2 in the original objective function (Eq. 4.5).

Let $V = W_{jpca} = \begin{bmatrix} W_{jpca}^1 \\ W_{jpca}^2 \end{bmatrix}$. Because W_{cca}^1 and W_{cca}^2 are obtained through SVD, we can conclude that they are orthogonal matrices and $W_{cca}^1(W_{cca}^1)^T = I$ and $W_{cca}^2(W_{cca}^2)^T = I$. We substitute these into the original expression of V before Eq. 4.9 and obtain:

$$U^1 = W_{cca}^1 W_{jpca}^1; \quad U^2 = \frac{1}{\sqrt{\lambda}} W_{cca}^2 W_{jpca}^2 \quad (\text{Eq. 4.11})$$

It is not difficult to extend the above training process to multiple views. Suppose the gallery samples are from n_1 different views and the probe samples are from n_2 different views, we should have $n_1 + n_2$ views for each class in the training process. To maximize the within-class sample correlation, we first form two joint matrices as follows:

$$\mathbf{X}^1 = \begin{bmatrix} \mathbf{x}_1^{g1} & \mathbf{x}_2^{g1} & \dots & \mathbf{x}_n^{g1} \\ \mathbf{x}_1^{g2} & \mathbf{x}_2^{g2} & \dots & \mathbf{x}_n^{g2} \\ \vdots & & \ddots & \vdots \\ \mathbf{x}_1^{g_{n_1}} & \mathbf{x}_2^{g_{n_1}} & \dots & \mathbf{x}_n^{g_{n_1}} \end{bmatrix}; \quad \mathbf{X}^2 = \begin{bmatrix} \mathbf{x}_1^{p1} & \mathbf{x}_2^{p1} & \dots & \mathbf{x}_n^{p1} \\ \mathbf{x}_1^{p2} & \mathbf{x}_2^{p2} & \dots & \mathbf{x}_n^{p2} \\ \vdots & & \ddots & \vdots \\ \mathbf{x}_1^{p_{n_2}} & \mathbf{x}_2^{p_{n_2}} & \dots & \mathbf{x}_n^{p_{n_2}} \end{bmatrix} \quad (\text{Eq. 4.12})$$

that is, we concatenate all samples from the gallery (probe) views of each class to generate $\mathbf{X}^1$ ($\mathbf{X}^2$) as used in the previous training process. Then, CCA in Eq. 4.6 and Eq. 4.7 becomes Joint CCA (JCCA). Conventional method, generalized CCA (gCCA) [172], is not applied here because the proposed JCCA is easier to implement.

4.3.3 Recognition

Given a probe sample $\mathbf{t}$ and all the gallery samples $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_s]$ taken under two different views and represented by REI, along with the learned prototypes $U^g = [u_1^g, u_2^g, \dots, u_m^g]$ and $U^t = [u_1^t, u_2^t, \dots, u_m^t]$ in the two views, we firstly represent each sample as a linear combination of its corresponding prototypes in the sense of minimum

mean square error and then extract the representation coefficients as the view-invariant feature. These features are used for similarity measure. That is,

$$d(\mathbf{t}, \mathbf{g}_i) = d(\beta^t, \beta_i^g) = \|\beta^t - \beta_i^g\|_2^2 \quad (\text{Eq. 4.13})$$

where $\beta^t = \arg \min_{\beta} \|\mathbf{t} - U^t \beta\|_2^2$, $\beta_i^g = \arg \min_{\beta} \|\mathbf{g}_i - U^g \beta\|_2^2$.

To calculate the coefficients β^t , let $L = \|\mathbf{t} - U^t \beta^t\|_2^2$. It follows that

$$\begin{aligned} \frac{\partial L}{\partial \beta^t} &= (U^t)^T U^t \beta^t - (U^t)^T \mathbf{t} = 0 \\ \Rightarrow \beta^t &= ((U^t)^T U^t)^{-1} (U^t)^T \mathbf{t} \end{aligned} \quad (\text{Eq. 4.14})$$

Similar procedure can obtain the coefficients β_i^g . In this way, each sample is represented by an $m \times 1$ vector ($m \ll d$) which can reduce the storage space and computational time. At last, we apply the nearest neighbor classifier to obtain the identity of the probe sample $\mathbf{t}$.

When samples from multiple views are available, certain score level fusion methods [173] can be applied to combine the above sample-to-sample distances to achieve the sample-to-class or class-to-class distances.

4.3.4 Patch-based JSL

It has been shown that local information is robust to pose variation in face recognition [170]. Therefore, we further propose to divide the original sample images into patches and conduct patch-based JSL for view-invariant gait recognition. The training and recognition procedures are very much the same as previously illustrated for the whole sample images and the difference mainly lies on the calculation of distances between samples. Suppose each whole image sample is divided into k different patches (denoted as $\mathbf{f}$) and we have obtained the representation coefficients for each patch by patch-based JSL, we compute the distance between the probe sample $\mathbf{t}$ and one of the gallery samples $\mathbf{g}$ through the following two steps:

(1) For each patch of the probe image $\mathbf{f}_i^t$, we define its distance to the gallery image $\mathbf{g}$ as the minimum distance between $\mathbf{f}_i^t$ and all the patches of $\mathbf{g}$, i.e., $d(\mathbf{f}_i^t, \mathbf{g}) = \min_j d(\mathbf{f}_i^t, \mathbf{f}_j^g)$, $j = 1, 2, \dots, k$. We term this as the patch-to-image distance.

(2) We then add up the first l ($l \leq k$) smallest patch-to-image distances as the final image-to-image distance.

These two steps are reasonable because view variation has great effect on the appearance of gait (see Figure 4.3) and it is possible that patches containing the same part of the body from two different views do not necessarily appear at the same position in the images. Besides, not all the probe patches could find their corresponding patches in the gallery image due to view variation.

In the case that multiple gallery views are available, the patch-to-class distance is first calculated by replacing the sample $\mathbf{g}$ in (1) with all samples from different views of the same class. Then, the image-to-class distance is obtained by using the same strategy as in (2).

4.4 Experimental Results

4.4.1 Database

We chose the CASIA-B multi-view gait database [167] in our experiments. It was appropriate because it contained samples from 11 different views that spanned the space of the left hand side of the walking subject. That is, from left to right, the viewing angles were 0° , 18° , 36° , 54° , 72° , 90° , 108° , 126° , 144° , 162° and 180° , respectively. Moreover, it also contained samples from a large number of subjects (124) each of which had 6 sequences walking under a normal condition.

4.4.2 Experimental Settings & Parameter Tuning

As shown in the flowchart in Figure 4.2, we need three sets of data samples for each experiment: training set, gallery set, and probe set. In each of our experiments, we

randomly selected n_c out of the 124 subjects as training subjects and the rest were used for testing. Given both the gallery views and probe views, in the training phase, we used all 6 sequences of the training subjects from these views for learning. In the recognition phase, 4 out of the 6 sequences of the testing subjects in the gallery views formed the gallery set and the other two sequences in probe views formed the probe set. To alleviate the effect caused by different training sets, we conducted each experiment 20 times and took the average recognition rate as the final result.

Table 4.1: First rank recognition rates (%) under different views. The gallery view is 90° .

Patch size	0°	18°	36°	54°	72°	108°	126°	144°	162°	180°	Mean
holistic	14.5	25.5	41	68.5	96	94	64	38	26	15.5	49.1
40*60	15.5	26	43	75.5	97	95.5	71	41	28	17.5	50.8
30*45	19.5	37	52	79	93.5	95.5	87	43.5	33	15	55.5
20*30	20.5	35.5	56.5	81.5	96.5	96	89.5	50	34.5	21.5	58.2
10*15	12.5	21	37.5	75.5	97	95.5	69.5	34.5	19.5	10.5	47.3

We empirically tuned the free parameters in the proposed method to achieve the best performance. The size of our gait representation feature REI was set to 41×90 pixels and the patch size was chosen as 20×30 pixels as it achieves a best performance (see Table 4.1). The overlap between neighboring patches was 10 pixels along either row or column direction. We also chose 24 training subjects ($n_c = 24$, the same as in [79]) for fair comparison with other methods.

4.4.3 Gait Feature Comparison

We conducted a toy experiment to demonstrate the robustness of proposed REI in view-invariant gait recognition task compared with the popular GEI. In this experiment, we chose 3 normal walking sequences from two different views (90° and 72°) of all 124 subjects for training and the left 3 for testing. That is, there were 6 samples for each subject in both sets. In the training stage, we applied LDA to obtain

Figure 4.5: Illustration of projections in the learned subspace from different feature representations. (a) GEI, (b) REI. For each class, samples come from two different views but share the same color.

the subspace that maintained the minimum within-class scatter as well as maximum between-class scatter. Then, we projected the test samples onto the learned subspace and showed part of them (samples from the first 5 subjects) in Figure 4.5. Obviously, samples represented by REI can be better classified than those by GEI. The subsequent experiments on gait recognition across varying views will also demonstrate this.

4.4.4 View-Invariant Gait Recognition Results

In real applications, multiple cameras are usually available. In our experiments, we limited the extension only to the case that multiple gallery views are available. According to the number of available gallery views, we divided the experimental results in this section into two parts: single gallery view and multiple gallery views.

4.4.4.1 Single gallery view ($n_1 = 1$, $n_2 = 1$)

In Figure 4.6, we show the contribution from different steps of our proposed JSL method. As the results indicated, the information recorded by the Radon transform

Table 4.2: First rank recognition rates (%) achieved by different methods. The gallery view is 90° .

Probe Views	0°	18°	36°	54°	72°	108°	126°	144°	162°	180°
PCA	2.8	2.8	4	8.8	50.4	72.2	12.9	4	1.5	2.5
CCA	3.5	6	9	27.5	73	76	29.5	8	5.5	3
Goffredo et al. [73]	-	-	56.3	60.4	64.5	64.5	68.8	-	-	-
UDSA [74]	2.2	3.8	9.7	27.1	100	98	35.7	10.9	2.4	2.8
SVR [79]	16.5	22.2	35.1	63.3	94.3	94.3	65.7	39.1	20.2	12.9
TFA [170]	15.5	18.5	34	73	88.5	87	70	36.5	27	10.5
Proposed JSL	20.5	35.5	56.5	81.5	96.5	96	89.5	50	34.5	21.5

is robust to large view-variation and the patch-based JSL further enhances the performance greatly in most of the views. For example, the recognition rate is enhanced by nearly 10% in view 18° and more than 10% in view 54° respectively. Besides, the applied CCA is also an important step as the performance can be enhanced by more than 10% in most views. Moreover, both larger number of training subjects and effectively selected patches can help achieve higher recognition rates. Compared with other parts, the part that using selected patches improves the results the least.

To further demonstrate the proposed coefficient representation is suitable for view-invariant gait recognition task, we also conducted another experiment with the GEI feature and JPCA training process, the results are shown in Figure 4.6(f) in which results of other two methods are also shown for comparison. As we can see from the figure that the recognition rates of GEI+PCA across large view variations ($> 36^\circ$) are lower than 10%. However, if JSL (even if only JPCA) is applied, the enhancement of recognition rates for large view variations are significant (more than 20%). This demonstrates that the proposed coefficient representation contributes more in the task of view-invariant gait recognition.

We also compared the proposed method JSL with state-of-the-art methods and the results are shown in Table 4.2. The results of Goffredo *et al.* [73] and SVR [79] were as reported in the corresponding papers and we implemented the other methods by

ourselves. As we can see, PCA, CCA and UDSA didn't perform well when the view variation is large, and only $2\% \sim 3.5\%$ recognition rates were achieved for extreme views 0° and 180° . The results of Goffredo *et al.* [73] were not as competitive and it couldn't deal with extreme views as the joint detection is difficult. For SVR and TFA, they improved the performance greatly especially under extreme views. However, the proposed JSL performed the best for all the views considered. Specifically, the recognition rate is enhanced by 9% in view 180° , 13% in view 18° , and more than 20% in view 36° . In addition, the feature dimension of the proposed JSL method is very low (less than $24 \times 6 = 144$) and the computation is limited to linear subspace, both of which make the implementation of our proposed method efficient.

In previous experiments, we considered the gallery view to be the side view which is common in current gait recognition research. To further show that our proposed algorithm is feasible in full range of view invariance, we also conducted several experiments using the other 10 views in the database as the gallery view, respectively. The results are shown in Figure 4.7(a)-(e). The average of each curve in Figure 4.7(a)-(e) is shown in Figure 4.7(f). Clearly, different views have different ability to handle view variations and we can see that view 54° performs the best among all the 11 views.

4.4.4.2 Multiple gallery views ($n_1 > 1$, $n_1 = 1$)

When multiple gallery views are available, there should be numbers of different combinations ($C_{11}^{m_1}$) among these views. Here, we only show results of some example combinations in Table 4.3. Obviously, the more gallery views are available, the higher overall recognition rate could be achieved. The mean recognition rates achieved by two different gallery views are higher than that from the single view by more than 30%, while those from three different gallery views are only higher than those from two views by 2~3%. Moreover, when multiple gallery views are available, the highest

Table 4.3: First rank recognition rates (%) achieved by different combinations of the gallery views which are indicated as in the first column.

Probe Views($^{\circ}$)	0	18	36	54	72	90	108	126	144	162	180	Mean
90°	21	36	57	82	97	98	96	90	50	35	22	62.2
$0^{\circ}, 90^{\circ}$	96	91	83	92	96	97	95	92	82	77	80	89.2
$18^{\circ}, 90^{\circ}$	90	97	96	93	95	95	96	92	84	81	76	90.5
$36^{\circ}, 90^{\circ}$	67	94	96	97	97	96	96	93	88	78	64	87.7
$54^{\circ}, 90^{\circ}$	47	77	94	97	97	96	96	96	87	71	2	83.7
$72^{\circ}, 90^{\circ}$	41	59	87	97	98	97	96	93	85	69	56	79.7
$0^{\circ}, 90^{\circ}, 108^{\circ}$	96	92	87	95	96	96	96	96	89	81	79	91.0
$18^{\circ}, 90^{\circ}, 126^{\circ}$	88	97	96	97	97	97	97	97	96	87	79	93.2
$36^{\circ}, 90^{\circ}, 144^{\circ}$	72	94	96	97	97	96	96	96	95	90	85	92.1
$54^{\circ}, 90^{\circ}, 162^{\circ}$	67	82	96	97	98	96	96	95	94	96	90	91.4
$72^{\circ}, 90^{\circ}, 180^{\circ}$	69	75	85	97	96	96	97	95	94	90	95	89.9

overall recognition rates are achieved when the available views are not so close to each other. For example, we can conclude from the table that the combinations ($18^{\circ}, 90^{\circ}$) and ($18^{\circ}, 90^{\circ}, 126^{\circ}$) perform the best among all their comparable combinations listed.

4.5 Conclusion

We have proposed in this chapter a novel Joint Subspace Learning method for view-invariant gait recognition. As the appearance features vary significantly across different views, we attempt to seek such a representation that is stable across intra-subject view variations and discriminant among different subjects. This is evidenced by the properties of LOC. Therefore, we propose to first conduct joint learning through CCA and PCA to gain the prototypes of different views. Then, we represent each sample in both the gallery set and probe set acquired from different views as a linear combination of the prototypes in the corresponding views and extract the coefficients as the feature representation. Furthermore, we also develop a new feature descriptor REI and propose to divide it into patches to further enhance the recognition performance. Experimental results demonstrate that the proposed JSL method outperforms

the existing methods in the problem of gait recognition across varying views.

Figure 4.6: First rank recognition rates. The gallery view is 90°. The proposed approach contains 5 important parts which are denoted as “Radon + patch + JSL + ‘ $n_c = 24$ ’ + ‘ $l < k$ ’”. In each of the figures shown above, we replace the method of each part with the alternative method and show the results of the modified approach.

Figure 4.7: Experimental results of other 10 views except 90° as the gallery view for view-invariant gait recognition.

Chapter 5

View Recognition of Human Gait Sequences in Videos

The proposed JSL solution in the previous chapter achieves promising performance even under large view variations. However, its main drawback in real applications is that it needs to know the viewing angles of the gait sequences, which in fact is a common problem for most of the existing motion-based gait recognition algorithms. In this chapter, we aim to address the problem of recognizing the viewing angles of human gait sequences to enhance our contribution to automatic gait recognition under view variations.

5.1 Introduction

Human gait analysis has recently attracted much attention in the computer vision community because of the demand for automatic human identification at a distance in public places such as malls, airports and parking lots. In computer vision, several attempts on human gait analysis have been proposed in the literature [29, 44, 46, 137], and existing methods can be roughly grouped into two categories: model-based [8, 46, 138] and motion-based [11, 71, 74, 127, 139, 140].

Model-based methods usually model the human body or body movements by processing each frame of a gait sequence. Thus, the calculation is very complicated.

Moreover, sufficient samples are usually required to ensure acceptable recognition performance. For motion-based methods, the gait feature is characterized by different compact representations, such as optical flow [137], eigengait [139] and gait energy image (GEI) [28], which are simple and effective. Hence, motion-based methods are more popular and have been widely used in many existing human gait recognition systems.

Most existing motion-based gait recognition methods, however, assume that the view angles of the testing gait sequences are known as prior information. In many practical applications, this assumption may not hold. Hence, a natural question is that whether we can automatically recognize the view angles of human gait sequences in videos such that fully automatic view-invariant gait recognition can be achieved. To our knowledge, previous work that addressed this issue was based on image geometry [77] and therefore camera parameters such as focal length were needed which prevented the methods from practical use. In this chapter, we propose to address the problem using simple appearance features. In the previous work [146], the Gait Energy Image (GEI) [29] was used as gait sequence representation for view recognition. Although promising results have been reported, there is much to be desired on the robustness as well as the recognition performance because these are important for the subsequent identification step.

Motivated by the fact that in the problem of view recognition, the intra-class variation (differences among different subjects) may be diverse, we propose to use multiple feature descriptors to represent the gait sequences because it is reported that multiple feature representations can characterize the data discriminatively from various aspects. To better combine these multiple gait descriptors, we also develop an adaptive discriminant analysis with enhanced multiple kernel learning (ADA-EMKL) to extract low-dimensional features for view recognition. In our proposed method, we

not only generate high order kernels to model the non-linear relationship among different feature spaces but also impose pairwise data correlation to alleviate the confusion among samples that could be easily misclassified. Experimental results on a popular multi-view gait database taken under different walking conditions demonstrate the effectiveness as well as the robustness of the proposed approach. Meanwhile, we also investigate the influence of view recognition on the view-invariant gait recognition performance through experimental study.

The rest of the chapter is organized as follows. Firstly, we briefly review the motivation for studying the view recognition problem. Then, details of the proposed method for view recognition based on human gait sequences are presented. Finally, experimental results are shown with further conclusions.

5.2 Motivation

In early days, image geometry was applied to address the view recognition problem in gait analysis and Kale et al. [77] proposed to first estimate the angle between the walking direction and the image plane and then transform the frame in the estimated view into a target view through image geometry. In his method, focal length of the camera was needed which greatly prevented its practical applications because this parameter is usually very difficult to obtain. Although the focal length can also be estimated from the captured video sequences, the propagation of the estimation error remains uncertain. Later, the appearance feature descriptors [29, 111, 44, 32, 41, 42] became popular in gait recognition research because they were both simple and effective. In work [146], Gait Energy Image [29] was used for view recognition and promising results were reported. However, a single feature descriptor may be sensitive to walking condition variations. Like human faces, the gait appearance under different views should show different kinds of asymmetry (refer to Figure 5.2). Therefore, in

this chapter, we also introduce the asymmetry feature used in face recognition [155] as another feature descriptor.

Multi-modal or multiple features are now very popular in recognition or classification tasks because they usually characterize the data from various aspects and thus can extract different discriminative information for recognition. To fuse these multiple feature descriptors, several methods have been proposed which can be divided into feature level [101, 156, 157, 177] and decision level [132, 176] fusions. For decision level fusion, decision is first made based on each description independently and then decisions from different descriptions are fused according to certain rules [25]. Clearly, relationship between different descriptions is not fully exploited which may also be useful for the final task. For feature level fusion, subspace learning based methods [156, 157] are very effective because they not only can extract much discriminant information useful for the final task but also can reduce the feature dimension which greatly saves computing time.

Meanwhile, kernel-based methods [142, 143, 178] have gained much popularity due to their success. As the features adopted in most recognition or classification tasks are usually of high dimension, linearly separation among the features from different classes becomes difficult. Yet, the definition of the kernel space can easily address this. In recent studies [144, 145, 147, 148], it has been reported that the combination of multiple different kernels performs better than a single kernel in classification tasks. For kernel combination, the simplest way is to assign the same weight to all kernels but this may not be the most effective. A better strategy is to learn the weight for each kernel and it has been shown that learning SVMs [149, 150] with multiple kernels could result in better accuracy. Lanckriet *et al.* [147] formulated this as a semidefinite programming problem which permitted the finding of the combination weights and support vector coefficients simultaneously. It is well known that subspace learning

Figure 5.1: Flow chart of the proposed approach to view recognition of human gait sequences in videos.

methods also play an important role in recognition tasks [151, 152] and most of them benefit from kernel methods [153, 178]. Therefore, Lin *et al.* [154] combined the MKL problem with the subspace learning scheme and proposed to perform dimension reduction and multiple kernel learning simultaneously. However, they considered only linear relationship among the base kernels in their method. Nevertheless, confusion among neighboring samples were ignored which is an important issue in recognition.

5.3 Proposed Approach

To address the problem of view recognition, we should consider both the recognition accuracy which is very important for the view-invariant identification step and the computational efficiency which is necessary for real applications. Therefore, we propose to apply multiple feature descriptors because different features characterize the data from different aspects. Furthermore, to better exploit more discriminative information for recognition as well as to reduce the feature dimension for simplicity, we also develop an adaptive discriminant analysis algorithm with enhanced multiple kernel learning (ADA-EMKL). To better model the structure among feature spaces, we map each feature to a kernel space and generate multiple high order kernels to characterize their non-linear relationship. Meanwhile, the pairwise data correlation in local neighborhood is applied to impose large penalties on both interclass samples

Figure 5.2: Illustration of the two features GEI and IGEI under 11 different views and 3 different walking conditions.

with small differences and intra-class samples with large differences, thus alleviating the confusion between samples that could be easily misclassified. Figure 5.1 shows the flow chart of the proposed algorithm. In the training stage, ADA-EMKL is performed to obtain the subspace projection matrix Γ as well as the kernel weights η ; in the testing stage, the weighted high order kernels are projected on the learned subspace and the nearest neighbor classifier is applied to assign the specific view to the test sequence. The details of each main step of the proposed method will be discussed bellow.

Figure 5.3: Illustration of the mean absolute value of the IGEI feature under 11 different views and 3 different walking conditions.

5.3.1 Gait Sequence Representation

To enhance the recognition accuracy of the gait sequences, we apply two feature descriptors for representation which characterize the gait data from different points of view. One is the widely used feature, Gait Energy Image (GEI), which captures the gait appearance, and the other is the asymmetry feature [155] which shows the degree of the gait symmetry. It is calculated as the imagery part of the Fourier Transform of GEI (named as IGEI).

We illustrate the two features under different views and different walking conditions in Figure 5.2. For the IGEI feature, the absolute values are showed. From this figure we can see that the GEI feature changes with view variation and the difficulty should lie in distinguishing the neighboring view angles. Moreover, walking condition variations can affect the gait appearance especially from the profile views like 72°, 90° and 108°. While for the asymmetry feature IGEI, we show in Figure 5.3 the mean absolute value of corresponding IGEIs in Figure 5.2 (one IGEI image in Figure 5.2 corresponds to a point in Figure 5.3) to better visualize the effect of view variation on it. Obviously, the feature value changes with view variation. As for the view angles whose feature values are the same (symmetry of the feature line), they can still be distinguished

Figure 5.4: Illustration of the proposed idea for multiple kernel-based dimension reduction.

because they have different phases.

5.3.2 Adaptive Discriminant Analysis with Enhanced Multiple Kernel Learning (ADA-EMKL)

The main contributions of the proposed ADA-EMKL are three folds. Firstly, to better exploit the relation among different feature spaces, we project each of them to a high dimensional Hilbert space and apply multiple high order kernels to model their relationship. Secondly, to achieve computational efficiency, we introduce a subspace-based multiple kernel learning strategy which can project the multiple high-dimensional kernel features into a unified low dimensional subspace. Thirdly, to alleviate the confusion among neighboring samples, we employ pair wise data correlation to impose penalty on between-class pairs that have small differences and within-class pairs that have large differences.

Figure 5.4 shows the proposed idea. Specifically, given a set of samples $\{\mathbf{x}_{i,p}\} \in \mathbb{R}^{d_p}$, ($i = 1, 2, \dots, N$; $p = 1, 2, \dots, P$), where N is the total number of training samples, P is the number of descriptors for each sample and d_p is the feature dimension of the

p th descriptor, we first implicitly map them to a high dimensional Hilbert space as: $\mathbf{x}_{i,p} \mapsto \phi(\mathbf{x}_{i,p}) \in \mathfrak{H}$, where ϕ denotes the feature mapping. Then, we aim at learning a unified low-dimensional subspace in which projections of the same sample can be easily classified. Considering our third contribution mentioned above, the objective function for obtaining the unified subspace can be expressed as follows:

$$(\mathbf{V}, \eta) = \arg \min_{\mathbf{v}, \eta} \frac{\sum_{i,j=1}^N \sum_{p=1}^P \|\eta_p \mathbf{v}_p^T \phi(\mathbf{x}_{i,p}) - \eta_p \mathbf{v}_p^T \phi(\mathbf{x}_{j,p})\|^2 w_{ij}}{\sum_{i,j=1}^N \sum_{p=1}^P \|\eta_p \mathbf{v}_p^T \phi(\mathbf{x}_{i,p}) - \eta_p \mathbf{v}_p^T \phi(\mathbf{x}_{j,p})\|^2 w'_{i,j}} \quad (\text{Eq. 5.1})$$

$$\eta_p \geq 0, \quad p = 1, 2, \dots, P. \quad (\text{Eq. 5.2})$$

where $w'_{i,j}$ and $w_{i,j}$ are the between-class and within-class penalty functions defined by the pairwise correlation as follows:

$$w'_{ij} = \begin{cases} \frac{\langle \mathbf{x}_i, \mathbf{x}_j \rangle}{\|\mathbf{x}_i\|_2 \|\mathbf{x}_j\|_2}, & \text{if } C_i \neq C_j \text{ \& } i \in N_j \\ 0, & \text{otherwise} \end{cases}, \quad w_{ij} = \begin{cases} \frac{\|\mathbf{x}_i\|_2 \|\mathbf{x}_j\|_2}{\langle \mathbf{x}_i, \mathbf{x}_j \rangle}, & \text{if } C_i = C_j \text{ \& } i \in N_j \\ 0, & \text{otherwise} \end{cases} \quad (\text{Eq. 5.3})$$

where N_j denotes the neighbors of the j th sample. Obviously, kernel discriminant analysis (KDA) [178] is a special case of Eq. 5.1 when $P = 1$, w_{ij} and w'_{ij} are the graph edges indicating whether or not the two samples are from the same class.

According to KDA, if each feature space (descriptor) is treated separately, the optimal linear embedding should lie in the span of the training data. This means that the q th dimension $\mathbf{v}_p^q$ of the embedding subspace can be expressed as $\mathbf{v}_p^q = \sum_{n=1}^N \gamma_n^q \phi(\mathbf{x}_{n,p})$. Thus, the corresponding sample projections $(\mathbf{v}_p^q)^T \phi(\mathbf{x}_{i,p})$ in Eq. 5.1 can be evaluated through the popular kernel trick to avoid the explicit definition of the mapping function:

$$(\mathbf{v}_p^q)^T \phi(\mathbf{x}_{i,p}) = \sum_{n=1}^N \gamma_n^q \langle \phi(\mathbf{x}_{n,p}), \phi(\mathbf{x}_{i,p}) \rangle = \sum_{n=1}^N \gamma_n^q k_p(\mathbf{x}_n, \mathbf{x}_i) \quad (\text{Eq. 5.4})$$

Hence, the unified projection $\sum_{p=1}^P \eta_p (\mathbf{v}_p^q)^T \phi(\mathbf{x}_{i,p})$ becomes:

$$\sum_{p=1}^P \eta_p (\mathbf{v}_p^q)^T \phi(\mathbf{x}_{i,p}) = \sum_{p=1}^P \eta_p \sum_{n=1}^N \gamma_n^q k_p(\mathbf{x}_n, \mathbf{x}_i) = (\gamma^q)^T K^{(i)} \eta \quad (\text{Eq. 5.5})$$

where

$$\gamma^q = [\gamma_1^q, \gamma_2^q, \dots, \gamma_N^q]^T \in \mathbb{R}^N, \quad (\text{Eq. 5.6})$$

$$\eta = [\eta_1, \eta_2, \dots, \eta_P]^T \in \mathbb{R}^P, \quad (\text{Eq. 5.7})$$

$$K^{(i)} = \begin{bmatrix} k_1(x_1, x_i) & \dots & k_P(x_1, x_i) \\ \vdots & \ddots & \vdots \\ k_1(x_N, x_i) & \dots & k_P(x_N, x_i) \end{bmatrix} \in \mathbb{R}^{N \times P}. \quad (\text{Eq. 5.8})$$

In the above projection, we only consider one specific dimension (q th) of the embedding subspace. Suppose we consider a Q -dimensional subspace, for sample $\phi(\mathbf{x}_{i,p})$, its unified projection on to the subspace becomes:

$$[\gamma^1, \gamma^2, \dots, \gamma^Q]^T K^{(i)} \eta = \Gamma^T K^{(i)} \eta \in \mathbb{R}^Q, \quad (\text{Eq. 5.9})$$

Hence, the objective function in Eq. 5.1 and Eq. 5.2 can be simplified as:

$$(\Gamma, \eta) = \arg \min_{\Gamma, \eta} \frac{\sum_{i,j=1}^N \|\Gamma^T K^{(i)} \eta - \Gamma^T K^{(j)} \eta\|^2 w_{i,j}}{\sum_{i,j=1}^N \|\Gamma^T K^{(i)} \eta - \Gamma^T K^{(j)} \eta\|^2 w'_{i,j}} \eta_p \geq 0, p = 1, 2, \dots, P. \quad (\text{Eq. 5.10})$$

Before providing the solution to Eq. 5.10, we would like to further exploit the non-linear structure among base kernels rather than just use their linear combinations as in Eq. 5.8 because it has been reported in [147] that the learned kernel weights through linear combination usually does not consistently outperform either the uniform combination of base kernels or simply the best single base kernels. Nevertheless, polynomial combinations of higher degree (≥ 1) of the base kernels with non-negative coefficients [161] have been shown to obtain notable performance improvements. Hence, we employ high order kernels, which are defined as the product of base kernels with different power, to model the nonlinear structure between the two base kernels. Suppose we

have P base kernels ($P = 2$ in our case), our proposed high order kernels can be expressed as follows:

$$k_m(\mathbf{x}_i, \mathbf{x}_j) = k_1^{a_1}(\mathbf{x}_i, \mathbf{x}_j) \dots k_P^{a_P}(\mathbf{x}_i, \mathbf{x}_j), \quad \text{for } a_i \geq 0, \sum_{i=1}^P a_i = 1 \quad (\text{Eq. 5.11})$$

where m indexes the combinations. As we have mentioned before, for a positive definite symmetric (PDS) kernel function, it is implicitly defined as an inner product in a high-dimension Hilbert space: $k(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$. Thus, we can conclude that the positive power of PDS is still PDS. In addition, it has been proven that the products and sums of PDS kernels are still PDS [162]. Therefore, our proposed high order kernel is a PDS kernel function. Given the generated high order kernels, the optimization problem remains the same except that the dimension of η and $\mathbf{K}^{(i)}$ in Eq. 5.7 and Eq. 5.8 should become $\Re^M$ and $\Re^{N \times M}$, where M is the number of combinations defined from Eq. 5.11. Applying these high order kernels to characterize the non-linear relationship between base kernels is important and can further enhance the classification performance which will be demonstrated in the experiments.

Selecting the kernel function is an important step in kernel-based methods. Usually for data sample represented in the form of vector, the RBF kernel is used: $k_p(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\frac{d^2(\mathbf{x}_{i,p}, \mathbf{x}_{j,p})}{\sigma^2})$ and σ^2 is a positive constant.

At last, to solve the optimization problem in Eq. 5.10 with the generated high order kernels, we adopt an iterative, two-step strategy to alternately optimize Γ and η [154]. In our experiments, we initialize Γ first by assuming $\Gamma\Gamma^T = I$ as this could have a more stable performance than initializing η first [154]. Then, in the first step of each iteration, η is first optimized with fixed Γ . However, the additional constraints on η make the optimization problem no longer a generalized eigenvalue problem and therefore, a relaxation is imposed to make the problem solvable by semidefinite programming (SDP) [154]. In the second step of each iteration, Γ is optimized with the

Table 5.1: Training and testing stage of the proposed ADA-EMKL.

Training	
Input	The penalty function $w_{i,j}$ and $w'_{i,j}$ (Eq. 5.3) for adaptive discriminant analysis Data described by various visual features in form of base kernels K_m (Eq. 5.8, Eq. 5.11)
Output	Sample coefficients vectors $\Gamma = [\gamma^1, \gamma^2, \dots, \gamma^Q]$ for dimension reduction Kernel weight vector $\eta = [\sqrt{\eta_1}, \sqrt{\eta_2}, \dots, \sqrt{\eta_P}]^T$ for kernel ensemble
1.	Initialize Γ as $\Gamma\Gamma^T = I$.
2.	η and Γ are optimized in an iterative, two-step strategy [154].
3.	Return Γ and η .
Testing	
Input	Projections of all training samples onto the learned unified subspace $\Gamma^T K^{(i)} \eta \in \mathbb{R}^{N'}$. Test sample $\mathbf{x}_t$
Output	Class label of sample $\mathbf{x}_t$
1.	Build the higher order kernel matrix $K^{(t)}$ for sample $\mathbf{x}_t$ according to Eq. 5.12
2.	Calculate the projection of $K^{(t)}$ onto the learned unified subspace as $\Gamma^T K^{(t)} \eta$.
3.	Similarity measure (Euclidean distances) and classification (nearest neighbor classifier).

solution from the first step and the optimization problem can be easily reformulated as a generalized eigenvalue problem. The iterations are repeated until convergence or a maximum number of iterations is reached.

5.3.3 View Recognition with ADA-EMKL

In the training stage, firstly, the penalty functions w_{ij} and w'_{ij} for each feature descriptor are calculated according to Eq. 5.3 and averaged separately to derive a unified one; secondly, kernel matrices for each sample are built according to Eqs. 5.8 and Eq. 5.11; thirdly, the optimization problem in Eq. 5.10 is solved to obtain the basis matrix Γ for the learned unified subspace and the kernel combination weights η .

In the testing stage, given a testing sample $\mathbf{x}_t$, we can build the corresponding kernel matrix with high order as follows:

$$K^{(t)} = \begin{bmatrix} k_1(\mathbf{x}_1, \mathbf{x}_t) & \dots & k_M(\mathbf{x}_1, \mathbf{x}_t) \\ \vdots & \ddots & \vdots \\ k_1(\mathbf{x}_N, \mathbf{x}_t) & \dots & k_M(\mathbf{x}_N, \mathbf{x}_t) \end{bmatrix} \in \mathbb{R}^{N \times M}. \quad (\text{Eq. 5.12})$$

Then, the projection of the testing sample onto the learned unified subspace can be evaluated as: $\Gamma^T K^{(t)} \eta$. Having obtained all the projected training samples in the

Figure 5.5: Illustration of samples from three different walking conditions and 11 different views.

learned subspace, the nearest neighbor classifier is then applied to assign the specific view to the testing sequence.

Table 5.1 summarizes the training and testing stages of the proposed method separately. Compared with other kernel-based methods, the training process of the proposed method is indeed more complicated because of the introduced semidefinite programming. However, the training process is usually regarded as an off-line process and thus it would not affect the application of the proposed method. The testing stage is much simpler and easy to implement and its computational complexity is on the same level as that of other kernel-based methods.

5.4 Experiments

5.4.1 Database

Our gait-based view recognition was performed on the CASIA database B [163]. It contained data samples from 124 subjects walking under 11 different view angles and three walking conditions. Therefore, it is very appropriate for our experimental study and we show in Figure 5.5 the sample images from the 11 cameras under the three different walking conditions.

Table 5.2: Experimental settings for the three test sets.

Exp.	Subjects	Condition	Sample/view	Total samples
1	124-n	Normal	$(124 - n) \times 6$	$(124 - n) \times 6 \times 11$
2	124-n	Bag	$(124 - n) \times 2$	$(124 - n) \times 2 \times 11$
3	124-n	Coat	$(124 - n) \times 2$	$(124 - n) \times 2 \times 11$

5.4.2 Experimental Settings

In the experiments, we randomly selected n ($n = 31$ for results of our experiments) subjects as training subjects whose normal walking sequences form the training set. That is, there were $6 \times n$ training samples for each view (class) and a total of $6 \times n \times 11$ samples for the whole training set. The left $124 - n$ subjects were used as testing subjects. Totally, we conducted three experiments according to the three different walking conditions, and their detailed settings are shown in Table. 5.2. To alleviate the effect caused by different training sets, we conducted each experiment 5 times and took the average recognition rate as the final result.

For the design of the high order kernels, we set the power of the base kernel from GEI feature as $a_1 = \{0, 0.1, 0.2, \dots, 1\}$, and according to the constraints in Eq. 5.11, the power of the other base kernel from the IGEI feature should be $a_2 = 1 - a_1$. Thus, we had a total of 11 kernels, i.e., $M = 11$.

5.4.3 Toy Example Illustration

To demonstrate the effectiveness of exploiting pairwise data correlation in local neighborhood as edge weights, we constructed a toy example using a synthetic dataset as in [146] and the comparison results are shown in Figure 5.6. Obviously, the main projection direction sought by ADA_b (only between-class pairwise correlations are used) [146] is much better than that of LDA. For ADA_wb (used in this chapter) which also considers the within-class pairwise data correlation, the projection direction should be

Figure 5.6: Illustration of a toy example to show the role of pairwise data correlation in subspace learning.

the best because the data after projection should have minimum within-class scatter and maximum between-class scatter.

Obviously, each feature descriptor would yield a set of penalty function and we simply average them separately to derive a unified one.

5.4.4 Results on Normal Gait Sequences

In the first experiment, we used normal walking sequences as test samples which was in accordance with the walking condition of the training sequences. The recognition rates for different methods are shown in Table 5.3. For better visualization of the results, some of the corresponding confusion matrices are shown in Figure 5.7 separately. As for the names of the methods, "G" denotes for GEI while I for IGEI. *Con.* and *Avg.* are two possible ways for kernel combination: Concatenation and Average.

We can conclude from the table that: firstly, the two features have different dis-

Table 5.3: Experimental results (%) of Exp. 1 with test samples from normal walking sequences.

Method	0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	Avg.
GEI	85.9	92.5	94.4	97.9	90.5	87.6	91.3	93.2	92.5	92.8	83.1	91.1
IGEI	85.7	94.9	86.1	92.4	92.0	87.9	89.6	87.1	89.9	94.09	89.6	89.9
ADA-G	95.7	96.9	96.5	97.9	97.1	93.1	93.4	95.6	94.3	94.73	95.2	95.5
ADA-I	95.8	95.5	93.7	96.5	96.9	93.3	93.3	94.9	95.0	93.48	95.3	94.9
KDA-G	96.0	96.1	94.7	96.5	96.8	95.1	95.6	95.7	93.5	94.37	97.7	95.6
KDA-I	95.9	95.8	94.8	97.1	97.5	92.6	94.4	95.6	95.1	92.26	94.4	95.0
KDA- <i>Con.</i>	96.3	95.9	94.3	96.8	97.0	93.9	95.3	96.7	93.2	95.06	97.4	95.6
KDA- <i>Avg.</i>	96.5	95.9	94.2	96.6	97.0	94.6	94.7	96.5	92.9	94.81	97.0	95.5
MKL	96.1	97.4	97.2	97.6	97.5	94.6	93.6	96.5	94.0	94.55	95.2	95.8
ADA-MKL	95.9	96.2	94.6	98.2	96.2	94.8	94.8	96.4	96.4	96.59	97.7	96.2
ADA-EMKL	97.5	96.4	95.7	98.9	99.1	95.9	97.5	98.2	94.3	96.95	98.9	97.2

Figure 5.7: Illustration of confusion matrices (%) of the results in Table 5.3. Only the values greater than 1 are shown in the figure.

criminative abilities in different views, for example, the recognition rates of view angles 36° and 54° using GEI feature are greater than those using IGEI feature by more than 5%, while for view angle 180°, the performance of IGEI is higher than that of GEI by more than 6%. Thus, the two feature descriptors could complement each other when combined appropriately. Secondly, the learning technique ADA can improve the performance for both features by nearly 5%. Similar results can be achieved with KDA. However, the improvements from the direct fusion methods (KDA-*Con.* and KDA-*Avg.*) are limited. This implies that simple kernel trick as well as linear combi-

Figure 5.8: The learned kernel weights for each method.

nation couldn't reflect the real structures of the data samples. Thirdly, ADA-EMKL performs the best, which suggests that both non-linearity as well as pairwise data correlation can't be ignored in recognizing view angles from gait sequences.

For the confusion matrices in Figure 5.7, it is obvious that the confusion mainly exists among neighboring views. Thus, the proposed ADA is reasonable. By imposing large penalty on samples of different classes that are similar to each other as well as samples of the same class that are far away from each other, ADA aims to alleviate the confusion of neighboring views. Results in Figure 5.7 also clearly show such improvement as we can see the number of mis-recognized neighboring views of the proposed method is greatly reduced compared with the baseline methods using original feature.

To assess the learned combination weights over the two base kernels for MKL and the high order kernels for EMKL, we plot in Figure 5.8 the averages of the learned η separately. For MKL, the learned weight for GEI is larger than that for IGEI. The reason is that in the normal walking condition, the feature GEI performs better than IGEI and such weights can better balance the performance among all the 11 views. For EMKL, we can see the nine weights in the middle (high-order combinations of the

Table 5.4: Experimental results (%) of Exp. 2 with test samples from walking sequences carrying a bag.

Method	0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	Avg.
GEI	73.7	82.8	87.1	89.3	86.0	64.0	69.4	81.7	84.4	82.8	64.0	78.7
IGEI	66.7	75.3	75.8	84.4	85.5	71.5	70.4	74.7	78.0	75.8	69.4	75.2
ADA-G	87.8	84.6	92.8	95.3	95.2	58.4	92.9	93.3	84.5	84.5	85.7	86.8
ADA-I	82.3	83.9	80.1	80.7	90.9	87.6	83.3	82.8	83.3	76.3	74.7	82.4
KDA-G	92.5	85.5	95.2	95.2	94.1	53.2	74.7	95.7	88.2	95.2	90.3	87.2
KDA-I	87.1	83.9	78.5	81.7	91.9	88.2	78.5	89.3	83.9	80.7	79.6	83.9
KDA- <i>Con.</i>	83.3	86.5	88.0	91.6	93.9	57.1	86.0	91.8	86.1	84.7	81.6	84.6
KDA- <i>Avg.</i>	85.4	87.1	89.3	91.8	95.6	63.7	85.4	92.8	87.5	84.2	82.8	86.0
MKL	86.6	87.6	87.1	92.5	91.4	78.0	74.7	93.6	91.4	86.6	89.3	87.2
ADA-MKL	89.3	85.5	93.6	94.6	93.6	61.8	83.3	90.9	87.6	94.6	87.6	87.5
ADA-EMKL	85.9	89.6	93.0	92.3	95.1	82.1	91.5	92.4	88.1	86.1	83.5	89.0

two base kernels) is much larger than those two on each side (the two base kernels themselves) in Figure 5.8(b). This means that the non-linear relationship dominates the relation among the two base kernels. As the weights depend on training set and we have only one training set in our experiments, the weights for the left two experiments are the same as illustrated in Figure 5.8.

5.4.5 Results on Gait Sequences Carrying a Bag

Figure 5.9: Illustration of confusion matrices (%) of the results in Table 5.4. Only the values greater than 1 are shown in the figure.

In the second experiment, we used gait sequences carrying a bag as the testing

Table 5.5: Experimental results (%) of Exp. 3 with test samples from walking sequences wearing a coat.

Method	0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	Avg.
GEI	81.1	93.8	79.4	84.0	73.3	45.1	67.0	69.8	89.3	78.7	75.4	76.1
IGEI	71.4	92.7	73.9	91.3	78.8	83.0	90.4	74.4	86.0	92.6	80.7	83.2
ADA-G	92.9	94.0	89.8	88.8	85.1	64.0	80.8	85.4	94.7	90.3	93.6	87.2
ADA-I	75.3	90.3	92.7	94.5	92.8	88.1	92.3	93.9	92.4	88.1	88.2	89.9
KDA-G	93.0	94.0	92.3	87.0	90.2	71.2	86.4	87.6	91.6	89.2	95.9	88.9
KDA-I	78.1	90.6	89.7	93.6	89.7	88.5	89.5	93.1	93.6	89.7	92.1	89.8
KDA- <i>Con.</i>	84.1	92.8	90.2	91.7	88.6	80.3	90.0	90.9	91.1	91.4	94.8	89.6
KDA- <i>Avg.</i>	83.8	92.8	90.4	91.3	89.4	83.3	89.5	91.1	92.5	91.1	94.1	89.6
MKL	91.9	93.3	91.3	89.2	90.3	73.7	87.1	89.3	92.5	90.9	95.0	89.5
ADA-MKL	94.6	96.8	88.2	88.7	91.9	81.7	85.5	83.3	95.2	90.9	92.5	89.9
ADA-EMKL	94.1	98.9	90.3	92.4	93.6	83.9	86.5	87.0	95.1	91.9	95.9	91.7

sequences. Similarly, the recognition rates and the selected confusion matrices are shown in Table. 5.4 and Figure 5.9 separately. We can conclude from these results that, firstly, the carrying condition can degrade the view recognition results by nearly 10% for all methods compared with those in normal walking condition; secondly, IGEI feature is more sensitive to carrying condition than GEI because the carried bag changes the symmetry of gait appearance greatly. Moreover, the effects on profile views (near 90°) is obvious especially for GEI and this can also be easily seen from Figure 5.2(c); thirdly, simple fusion methods (KDA-*Con.* and KDA-*Avg.*) can only balance the results but can not outperform the better one before fusion; lastly, MKL could better handle the multiple feature fusion problem and could achieve the results nearly to the best one before fusion. In addition, our proposed ADA-EMKL performs the best, with the average recognition rate enhanced by nearly 2% compared to MKL.

5.4.6 Results on Gait Sequences Wearing a Coat

In the third experiment, gait sequences wearing a coat were used as testing samples and the results are shown in Table. 5.5 and Figure 5.10 separately. Wearing a coat would also degrade the recognition results, however, IGEI is less sensitive to this variation than GEI, especially for view 90°, the recognition rate using GEI is even

Figure 5.10: Illustration of confusion matrices (%) of the results in Table 5.5. Only the values greater than 1 are shown in the figure.

less than 50%. Results in Figure 5.10(a) also show that the confusion caused by the wearing condition variation around view 90° is very obvious. We can also clearly see from Figure 5.2(e) that the coat covers the shape of the upper body. Learning based methods can enhance the recognition results significantly, yet the recognition rate for view 90° is still very low and only fusion of multiple features can enhance it. Our proposed method ADA-EMKL can achieve the best performance again which further demonstrates the necessary of exploring the non-linearity as well as adaptive pairwise correlation among training data samples.

5.4.7 Application of View Recognition from Human Gait Sequences

The most potential application of the proposed ADA-EMKL algorithm should be in a fully automatic view-invariant gait recognition system, in which the view angle of each testing gait sequence is first estimated and then view-invariant gait recognition (which usually assumes the view of the test sequence is known before hand) is conducted.

We show in Figure 5.11 the performance of such a system as mentioned above. The experiments were conducted on CASIA-B database, we randomly selected 24 subjects

Figure 5.11: Performance of a fully automatic view-invariant gait recognition system with the gallery view from 90°.

whose normal walking sequences from the 11 views are used as the training set. The gallery set consisted of 4 normal walking sequences of other 100 subjects from view 90° and the probe set had other 2 normal walking sequences from all the 11 views. The view-invariant gait recognition method we have adopted was our previous work in Chapter 4. Given a sample from the probe set, we first estimated its view based on the proposed ADA-EMKL, then with the estimated view, we obtained its identity through the approach proposed in Chapter 4. In the experiments, two different view estimation methods were chosen for comparison. From the performance we can see that the recognition rates with view information (either known or estimated) are much higher than those of the baseline algorithm (unknown view), especially for large view variations. Moreover, the performance with estimated views by proposed ADA-EMKL are better than that with estimated views by ADA, which indicates that large view estimation error can affect the subsequent view-invariant gait identification step and the efforts that can improve the view recognition accuracy is necessary for the development of a fully automatic view-invariant gait recognition system.

5.4.8 Summary

We summarize our findings here based on the experimental studies shown above:

- (1) High order kernels can better model the non-linear relationship among different feature spaces.
- (2) The proposed penalty function based on pairwise data correlation is reasonable because confusions mainly exist among neighboring views.
- (3) Different feature descriptors have different sensitivity to walking condition variations and the appropriate combination would achieve a good balance in performance.
- (4) The proposed algorithm has potential applications in automatic view-invariant gait recognition systems. Our effort attempting to enhance the view recognition accuracy is necessary because the view information is important for the subsequent identification step.

5.5 Conclusion

In this chapter, we have investigated the problem of view recognition from human gait sequences in videos. Gait analysis can play an important role in video surveillance system because gait is likely the most effective biometric that can be observed at a distance and without contact. However, existing motion-based gait recognition approaches usually assume that the walking direction of the testing gait sequence is known before recognition which in fact does not always hold in many real applications. Therefore, how to automatically recognize the view angle of a walking sequence becomes an important research problem. Moreover, with the success on view recognition, the fully automatic view-invariant gait recognition system could also be realized.

Inspired by the fact that multiple feature representations can usually characterize the data from different aspects and thus outperform a single feature descriptor in visual recognition, we propose to apply two important gait features for representation and fuse them by a new subspace learning method, ADA-EMKL. Moreover, to better exploit the structure between the two feature spaces, we also propose to generate high order kernels to model the non-linear relationship. In addition, pairwise data correlation is also applied to impose penalty on samples that could be easily misclassified. Experimental results on popular multi-view gait database taken under different walking conditions demonstrate the effectiveness as well as the robustness of the proposed approach.

Chapter 6

Gait Recognition Across Multiple Varying Walking Conditions

In previous chapters, we have addressed the problems of gait recognition under different walking condition variations separately, such as view variation, carrying or clothing variation. However, it frequently occurs in real applications that these variations may come into play simultaneously. To the best of our knowledge, little previous work has addressed the problem of gait recognition under multiple walking condition variations. Moreover, the existing solutions that aim to solve certain walking condition variation are not suitable for this problem because they mostly attend to one variation without the others. In this chapter, we will address this problem with a novel solution which can greatly enhance the recognition performance under multiple walking condition variations.

6.1 Introduction

In early years of gait recognition research, there was a common belief that the side view of gait provides the richest information of the walking subject [12]. As a result, most of the researchers limited their efforts only to the study on side-view based gait recognition. However, there has been experimental study in [125] which shows the

gait recognition performance from the frontal-parallel view is better than that from the side view. This suggests that the appearances of gait from different views can contain different discriminative information for recognition. Apparently, the recognition results can be improved if we appropriately combine the discriminative information from some additional views. In this chapter, we will address the problem of gait recognition across varying walking conditions through gait sequences from multiple views.

Till now, many researchers have addressed the gait recognition problem based on multiple views and proposed several solutions [11, 70, 126, 127]. Generally speaking, these methods can be divided into two kinds: model-based [70, 126] and image-based [11, 127, 164]. For model-based methods, a three-dimensional (3D) model or visual hull of the walking body was usually generated for recognition. In experimental research, the 3D methods indeed achieved good recognition results and had tolerance on certain intra-subject variations of views and walking conditions. However, they were complex in computation and time-consuming, and hence they were not appropriate for real applications. For image-based methods, data from different view angles were treated separately and the similarity or recognition results were obtained from each view through some view-dependent image-based approaches. Subsequently, some fusion strategies were applied to combine the results from multiple views. This kind of method was simple but it was view-dependent and sensitive to intra-subject variations. Therefore, they were not applicable to deal with walking condition variations.

A person's gait varies under different walking conditions, and such variations affect gait recognition. Researchers have studied gait recognition across many variations, including view variation [74, 79], clothing and carrying variation [87, 91, 109], walking speed variation [55, 82], and so on. However, the efforts usually focused on one variation at a time and thus the resulting methods were not directly applicable to

general cases where many variations may coexist. Moreover, it was necessary for these methods to know the specific view angles which can be hard to obtain in practical situations. In view of such limitations, a simple algorithm adaptive to different kinds of variations is in great demand to facilitate the gait recognition research in practical applications.

In this chapter, we propose a Multiview Subspace Representation (MSR) method to achieve this goal. Given gait data collected from different views of the same subject, MSR treats them as a feature set and uses the underlying subspace basis to represent the whole feature set. Although several approaches have been proposed for set-based representation [133, 153, 179, 180], the subspace-based representation is simple and robust to outliers. Due to the fact that linear subspaces characterize the data variance among the feature set, such representation would tolerate certain intra-subject variations. Moreover, it belongs to image-based approach and is simpler to compute than model-based methods. After applying MSR to the training data, the similarity between two subjects becomes the distance between two subspaces. In fact, it is a single-set based subspace similarity measure problem because we have only one set for each subject (class).

Measuring the distances between subspaces is not new in computer vision. A commonly used concept for such measurements is canonical correlations [171, 181], which has been shown to be effective and robust. The canonical correlations are obtained by rotating the subspaces so that they achieve the maximum correlation. Given the canonical correlations, several approaches [128, 130] used the canonical correlations or the corresponding canonical vectors for similarity measure directly. However, these approaches did not incorporate the discriminative learning [110, 182], which could improve classification performance. In contrast, the Constrained Mutual Subspace Method [183, 184] learned a new subspace in which the entire class exhibited

small variance. In [129], the authors proposed to learn a projection matrix so that the projected subspaces observed maximum between-class scatter as well as minimum within-class scatter. Although they achieved improvement in classification accuracy, they required each class to have more than one set which was not applicable when the number of samples in the training dataset was very limited.

To mitigate this problem, we develop a learning-based algorithm, Marginal Canonical Correlation Analysis (MCCA), to enhance the classification accuracy of these subspaces, each of which represents for a whole class. It is easy to imagine that misclassification mainly exists in subspaces that are very similar to each other; hence, MCCA attempts to find such a new subspace in which the distances between the neighboring subspaces are maximized. Such neighborhood-based discriminative learning has already been shown to be effective for vector-based classification [185, 186, 187, 188].

Compared with the existing methods, our proposed framework has the following four advantages:

- It belongs to image-based methods which makes it computationally inexpensive and suitable for practical applications.
- It takes into account the relationship between different views and avoids the error-prone view estimation step.
- It is tolerant to intra-subject variations between test and training samples as well as missing data.
- It robustly measures the similarity even with limited number of training samples of each subject.

The rest of the chapter is organized as follows. Firstly, we provide a detailed description of the proposed MSR. Secondly, we illustrate the learning based algorithm, MCCA. Finally, we discuss the experimental results and conclude the chapter.

Figure 6.1: Flowchart of the proposed approach MSR. The leftmost figures are the gait features ordered by view.

6.2 Proposed Approach

Our main contribution in this chapter is to propose a framework for set-to-set gait recognition based on multiview samples. The flow chart of the framework is shown in Figure 6.1. Two steps dominate this framework: Multiview Subspace Representation (MSR) and Marginal Canonical Correlation Analysis (MCCA). Given certain feature representations of original gait sequences (shown in upper left of Figure 6.1), MSR mixes all the samples from different views of the same class together and extracts the basis of the underlying linear subspace for representation. To better exploit discriminative information from these subspaces, MCCA learns a projection matrix $\mathbf{W}$ to transform these basis to a new subspace in which the neighborhood canonical correlations are maximized. At last, having the projections of both the test subspace and all the training subspaces, we define the largest canonical correlation as the distance

between subspaces and apply the nearest neighbor (NN) classifier for identification. Experimental results also demonstrate that the proposed methods are robust to different kinds of intra-subject variations. In the following sections, we will explain each of these steps in detail.

6.2.1 Feature Representation

Given a human walking video, we extract the silhouette from each frame using the method proposed in [12]. We then apply size normalization and horizontal alignment to each extracted silhouette image. After obtaining the aligned silhouette, we use the simple and effective feature, Gait Energy Image (GEI) [29], to represent each sequence as:

$$G(x, y) = \frac{1}{n} \sum_{t=1}^n B(x, y, t) \quad (\text{Eq. 6.1})$$

where n is the number of frames in the gait sequence, x and y are the image coordinates, and t is the frame number in the gait sequence.

We show in Figure 4.3 the GEI feature under different views. Obviously, the appearance changes greatly with view variation. Thus, it is easy to imagine that data from different views should contain different discriminative information for recognition. This is also the reason why recognition performance can be improved by multiview data.

6.2.2 Multiview Subspace Representation (MSR)

Given all the GEIs of the same subject, we mix them to form a feature matrix denoted as $I(\in \Re^{d \times N}) = [x_{\theta_1}^1, x_{\theta_1}^2, \dots, x_{\theta_1}^{n_{\theta_1}}, x_{\theta_2}^1, \dots, x_{\theta_k}^{n_{\theta_k}}]$, where $x(\in \Re^{d \times 1})$ represents the GEI vector and k denotes the number of views available, n_{θ_k} denotes the total number of GEI in view θ_k and $N(= n_{\theta_1} + n_{\theta_2} + \dots + n_{\theta_k})$ is the total number of all samples available. Set-based representations, especially linear subspace representations, tolerate certain

intra-subject variations. Therefore, we apply the corresponding subspace basis to represent the feature set. In fact, the sample x can be arranged at random order inside the feature matrix I as the subspace only takes into account the variance between samples.

To obtain the subspace basis, we apply SVD on I as follows:

$$I = U\Lambda V^T \quad (\text{Eq. 6.2})$$

Each column of U is a basis of the feature subspace and $\Lambda = \text{diag}([\lambda_1, \lambda_2, \dots, \lambda_N])$ is the diagonal matrix of singular values. As in many cases the obtained matrix U is not full rank ($N \ll d$), we only choose the first k_1 basis for representation. That is, we use $S = [u_1, u_2, \dots, u_{k_1}]$ to represent the feature subspace of I .

The subspace representation has been used in face recognition to solve intra-subject variations such as pose and illumination [129, 130], which is relatively close to the multiview gait recognition problem in this study. We illustrate in Figure 6.2(a) the original samples represented by GEI which contains intra-subject variations, such as view variation, carrying a bag and wearing a coat. All these variances affect the gait appearance significantly. For carrying and clothing variations, the influence on side views are more prominent than those on other views. We calculated the subspace representation under different walking conditions (row) based on MSR and the first six leading principle components are shown in Figure 6.2(b). Generally, the differences among MSR is much smaller than those among GEI, especially for the first few components. Moreover, we can also see from Figure 6.2(b) that the similarity among components in the same column becomes decreasing with the increase of the number of principle components chosen. Therefore, the optimal matching result should depend on the dimension of the subspace. We will further discuss such influence in the experimental part.

(a) Original samples represented by GEI

(b) Principle components

Figure 6.2: Comparison of original representation and subspace representation. (a) Original samples from 11 different views and 3 different walking conditions (normal (upper row), carrying a bag (middle row) and wearing a coat (bottom row)) represented by GEI. (b) The corresponding first six principle components computed from Eq. 6.2.

6.2.3 Marginal Canonical Correlation Analysis (MCCA)

In some real applications, it usually happens that very limited samples are obtained for training, especially for the research on gait because gait is a spatial-temporal signal and a large number of samples will cause problem not only in data storage but also in later processing. Therefore, we treat the limited number of samples available for the same subject as a single feature set and generate only one set per subject for set-based analysis. However, existing discriminative learning over linear subspaces often assume that each class has multiple sets for learning which is clearly not suitable for our task in this chapter. Hence, we propose a marginal canonical correlation analysis algorithm for single-set based discriminative learning. It aims at finding a new subspace in which the distances between the originally neighboring subspaces are maximized.

Mathematically, given the linear subspace representation of the training sets $\mathbf{S} = [\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_C]$, MCCA learns the projection matrix $\mathbf{W}$ according to the following optimization criterion:

$$\mathbf{W} = \arg \max_{\mathbf{W}} \sum_{i=1}^C \sum_{j \in N_{i(k)}} d(\mathbf{W}^T \mathbf{S}_i, \mathbf{W}^T \mathbf{S}_j) \quad (\text{Eq. 6.3})$$

where $N_{i(k)}$ is the set of k -nearest neighbor of subspace $\mathbf{S}_i$ and C is the total number of classes in the training set. Clearly, we need to calculate the distances $d(\mathbf{W}^T \mathbf{S}_i, \mathbf{W}^T \mathbf{S}_j)$ in Eq. 6.3 before optimizing $\mathbf{W}$. As we have reviewed previously, the concept of canonical correlations is usually used to calculate the distance between linear subspaces. However, the canonical correlations are only defined for orthonormal matrices. Although the subspace representation $\mathbf{S}$ is orthonormal (obtained through SVD), the projection $\mathbf{W}^T \mathbf{S}$ may not be orthonormal. Therefore, we need to normalize $\mathbf{W}^T \mathbf{S}$ first, which could be done by QR-decomposition as follows:

$$\mathbf{W}^T \mathbf{S}_i = \mathbf{Q}_i \mathbf{R}_i \quad (\text{Eq. 6.4})$$

According to the property of QR-decomposition, if we denote $\mathbf{S}'_i = \mathbf{S}_i \mathbf{R}_i^{-1}$, the result $\mathbf{W}^T \mathbf{S}'_i$ should be orthonormal. After normalization, the similarity between $\mathbf{W}^T \mathbf{S}_i$ and $\mathbf{W}^T \mathbf{S}_j$ can be measured by canonical correlations, which are calculated by applying SVD as follows [189]:

$$\mathbf{S}'_i{}^T \mathbf{W} \mathbf{W}^T \mathbf{S}'_j = \mathbf{P}_{ij} \Lambda \mathbf{P}_{ji} \quad (\text{Eq. 6.5})$$

where $\Lambda = \text{diag}([\lambda_1, \lambda_2, \dots, \lambda_N])$ is the singular value matrix, which is also known as canonical correlations, and $\mathbf{P}_{ij}, \mathbf{P}_{ji}$ are orthogonal rotation matrices. The larger the canonical correlations are, the more similar the projections of the linear subspaces. Thus, we define the distance $d(\mathbf{W}^T \mathbf{S}_i, \mathbf{W}^T \mathbf{S}_j)$ as follows:

$$d(\mathbf{W}^T \mathbf{S}_i, \mathbf{W}^T \mathbf{S}_j) = d(\mathbf{W}^T \mathbf{S}'_i, \mathbf{W}^T \mathbf{S}'_j) = N - \sum_{i=1}^N \lambda_i = N - \text{tr}(\mathbf{P}_{ij}^T \mathbf{S}'_i{}^T \mathbf{W} \mathbf{W}^T \mathbf{S}'_j \mathbf{P}_{ji}) \quad (\text{Eq. 6.6})$$

Substituting Eq. 6.6 into Eq. 6.3, the optimization problem becomes:

$$\mathbf{W} = \arg \max_{\mathbf{W}} \sum_{i=1}^C \sum_{j \in N_{i(k)}} (N - \text{tr}(\mathbf{P}_{ij}^T \mathbf{S}'_i{}^T \mathbf{W} \mathbf{W}^T \mathbf{S}'_j \mathbf{P}_{ji})) \quad (\text{Eq. 6.7})$$

Let $A = \mathbf{W}^T \mathbf{S}'_i \mathbf{P}_{ij}$, $B = \mathbf{W}^T \mathbf{S}'_j \mathbf{P}_{ji}$. It is noted that both A and B are orthonormal matrices. Therefore, it is easy to deduce the following result:

$$A^T B = \mathbf{I} - \frac{1}{2}(A - B)^T(A - B), \quad (\text{Eq. 6.8})$$

According to Eq. 6.8, the optimization problem in Eq. 6.7 can be generalized as follows:

$$\mathbf{W} = \arg \max_{\mathbf{W}} \text{tr}(\mathbf{W}^T \Phi_b \mathbf{W}) \quad (\text{Eq. 6.9})$$

where

$$\Phi_b = \sum_{i=1}^C \sum_{j \in N_{i(k)}} (\mathbf{S}'_j \mathbf{P}_{ji} - \mathbf{S}'_i \mathbf{P}_{ij})(\mathbf{S}'_j \mathbf{P}_{ji} - \mathbf{S}'_i \mathbf{P}_{ij})^T \quad (\text{Eq. 6.10})$$

It is noted that the solution $\mathbf{W}_{opt}$ should be the eigenvector of Φ_b . Also noted that the optimization process of $\mathbf{W}$ needs the value of $\mathbf{W}$ itself because of Eq. 6.5, and meanwhile, the other variables are not explicitly represented by $\mathbf{W}$. Hence, it is hard to find a closed form solution for Eq. 6.9 and an iterative learning framework is adopted here. In fact, the solution converges quickly in the experiments with the initial value of $\mathbf{W}$ equals to identity matrix [129].

6.2.4 Classification

For each training subject $i(i = 1, 2, \dots, C)$ in the database, we mix the data samples available as a feature set I_i^G and obtain the corresponding subspace representation S_i^G as previously illustrated. Then we apply the proposed MCCA to learn the projection matrix $\mathbf{W}$. Given the test samples of a subject, we denote its feature set as I^P and the corresponding subspace representation as S^P . Here, we propose a new distance definition which is different from what have been mentioned in Eq. 6.6:

$$d(I^P, I_i^G) = d(\mathbf{W}^T S'^P, \mathbf{W}^T S_i'^G) = k_2 - \sum_{j=1}^{k_2} \lambda_j \quad (\text{Eq. 6.11})$$

where S'^P and $S_i'^G$ are calculated based on Eq. 6.5, λ_j is the j th largest singular value obtained by applying SVD on $(S'^P)^T \mathbf{W} \mathbf{W}^T S_i'^G$. As we have illustrated in Figure 6.2, not all the principle components are useful for classification due to intra-subject variations. Similarly, not all canonical correlations will contribute to the identification. Therefore, only the first k_2 largest ones are used. We will further discuss the influence of value k_2 on the recognition results in the experimental part.

After obtaining the distances between the test subject and all the training subjects, the nearest neighbor classifier is applied for identity assignment:

$$c = \arg \min_i d(I^P, I_i^G), \quad (i = 1, 2, \dots, C) \quad (\text{Eq. 6.12})$$

6.3 Experimental Results

6.3.1 Databases

We conducted experiments on two popular multiview gait databases which contained several walking condition variations: CASIA-B database [17] and Motion of Body (MoBo) database [13] from the Carnegie Mellon University (CMU). In the CASIA-B database, each subject had 10 sequences, containing 6 normal walking, 2 carrying a bag and 2 wearing a coat. We numbered the cameras as 1 to 11 from left to right, see Figure 6.3. While for the CMU MoBo database, each subject was recorded performing four kinds of activities, i.e., fast walk, slow walk, carrying a basketball, walking on a cline surface.

6.3.2 Experiments and Results

We conducted five experiments. Firstly, we set the optimal value for free parameters empirically; then, we performed regular multiview gait recognition and compared the results of the proposed framework with that of alternative methods. We also conducted experiments to show the tolerance of the proposed framework on view variation, walking condition variation and missing data separately.

6.3.2.1 Parameter Optimization

There are two free parameters in the proposed methods, one is the feature dimension (k_1) of the linear subspace representation in MSR and the other is the number of canonical correlations (k_2) used for classification. In this part, we illustrate their influence on the final recognition results respectively with the other factor fixed.

We conducted the experiments on normal walking sequences. Training views were from cameras 2, 4, 6, 8, 10, and testing views were from cameras 1, 3, 5, 7, 9. In similarity measure, all the canonical correlations were used. The recognition results

Figure 6.3: Illustration of data capturing and sample frames captured by each camera.

with different feature dimensions are shown in Figure 6.4(a). From the figure we can see that the optimal results show at multiple values, and in our later experiments, we fixed the linear subspace dimension to one of these optimal values, that is, $k_1 = 13$.

Given the optimal feature dimension, we examined the optimal number of canonical correlations for similarity measure and the results are shown in Figure 6.4(b). From the results we can see that the recognition rate degrades with the increase of the number of canonical correlations considered. This is easy to imagine because intra-subject variations mainly reflect in the subordinate dimensions, see Figure 6.4(b). Therefore, in our later experiments, only the first canonical correlations are considered for similarity measure.

6.3.2.2 Multiview Gait Recognition

In this experiment, the test view angles were all the same with the corresponding training view angles. We used 3 out of 6 normal sequences from each of the available views as indicated in Table 1 for training and the left 3 for testing.

The recognition results are shown in Table 6.1. For feature fusion (F-fusion), we concatenated the feature vector in each view of the same subject to form a long vector

Figure 6.4: Experimental results to show the influence of the free parameters on the final recognition results. (a) Feature dimension of linear subspaces (k_1). (b) Number of canonical correlations for similarity measure (k_2).

Table 6.1: First rank recognition rates of different methods under conventional multiple view gait recognition. The available views are as indicated in the first column of each row, for which the number represents the camera in Figure 6.3.

Views	F-fusion	S-fusion	MSR+WSD	MSR+MCCA
2,6,9	0.992	0.968	0.984	0.992
2,6,9,11	0.992	0.984	0.984	1
2,4,6,9,11	1	1	0.992	1

for recognition. For score fusion (S-fusion), we chose the product rule as in [132]. We also chose different number of view angles for experiments, and the results are shown in each row of Table 6.1. From the results we can conclude that all these methods achieve very good recognition results (above 98% for first rank) on this database, and the proposed MSR+MCCA achieves the best results which suggests that it can be competent for conventional multiview gait recognition tasks. Moreover, the performance is better for data when more views are available. This shows that different views contain different discriminative information.

6.3.2.3 Tolerance of MSR on view variation

Only CASIA-B database was used in this experiment because it had gait sequences collected from 11 different view angles. Figure 6.5 is shown here to further illustrate the view variation problem in a multi-camera system. As shown in the figure, the CASIA-B database was captured by placing 11 cameras on the same side of the walking subject and the view angle between the nearest two cameras was 18° . To simulate view variation, we assume that the sequences captured by cameras that are symmetry with the original point (O in Figure 6.5) are left-right symmetry. Suppose in the experiments, we choose sequences from camera 1,3,5,7,9 as testing sequences. Consequently, if the walking direction changes 36° (angle between the two nearest testing views) counter-clockwise (dash line in Figure 6.5 with label “+1”) with respect to the original walking condition (solid line in Figure 6.5), then the sequences captured by camera 1 now are equal to the sequences captured by camera 3 in the database. Similarly for other cameras except for camera 9 which needs another step (left-right reverse) to make its captured sequences the same with those from camera 1 in the database. This implies that the testing sequence order is 3, 5, 7, 9, 1 now which is just like a cycle shift among the chosen cameras. The case when walking direction changes 72° (“+2”) is the same way as above.

The chosen training views and testing views are as indicated in Table 6.2. We can see that the testing views are totally different from the training views and this can be regarded as the walking direction changes by 18° in a multi-camera system shown in Figure 6.5. Besides, we also arrange the testing views in different order, where “+1” means shifting by 1 and “+2” means shifting by 2 (as explained above), to indicate different amount of view variations. For both training set and testing set, all the 6 sequences of each view were used in the experiments.

Figure 6.5: Illustration of the experimental setting for view variations in a multi-camera system.

As we have mentioned previously, both the data order and view correspondence have no effect on proposed method MSR. Therefore, we obtain only one significant result which is shown in Table 6.2. However, the alternative fusion method fails when the view or order of data is not available. See in Table 6.2, the first rank recognition rate of score fusion method significantly drops by 10% when the view variation is 18° . And if we do not know the specific views (order) of the testing views, the results are even worse: the first rank recognition rate further drops by 34% when we shift the order of testing view angles by 1 and it becomes only 23% when we shift the order further.

6.3.2.4 Tolerance of MSR on multiple walking condition variations

Both CASIA-B database and CMU MoBo database were used here to show the tolerance of MSR for different walking conditions: carrying a bag in Figure 6.6(a), wearing a coat in Figure 6.6(b), and walking speed in Table 6.3. For CASIA-B database, samples from normal walking sequences formed the training set and samples from different

Table 6.2: Cumulative recognition rates of different methods under view variation. Training views are from camera 2,4,6,8,10 and testing views are from 1,3,5,7,9 with different orders.

Rank	1	5	10	20
MSR	0.98	1	1	1
[132]-product	0.88	0.97	0.99	1
[132]-product,+1	0.64	0.71	0.79	0.87
[132]-product,+2	0.23	0.28	0.36	0.45

Figure 6.6: Experimental results to show MSR's tolerance on different walking condition variations. Both training views and testing views are from camera 2,4,6,8,10. (a) Carrying a bag. (b) Wearing a coat.

walking conditions formed the testing set. For CMU MoBo database, we used fast walking samples as training set and the corresponding slow walking samples as testing set.

Experimental results in Figure 6.6 show that the proposed MSR method outperforms alternative methods by 5% and 15% in first rank recognition rate for conditions of carrying a bag and wearing a coat, respectively. Compared with the condition of carrying a bag, the condition of wearing a coat can affect gait seriously because its first rank recognition rate drops by 35%. Moreover, the proposed MCCA performs better than WSD by more than 5% in both cases for first rank recognition rate, which demonstrates that the discriminative learning over linear subspaces also can enhance

Table 6.3: Experimental results on CMU MoBo database.

Rank	1	5	10
MSR	0.76	0.92	1
[127]	0.72	0.84	0.92

Figure 6.7: Experimental results to show MSR’s tolerance on both view variation and walking condition variation. Training views are from camera 2,4,6,8,10; testing views are from camera 1,3,5,7,9. (a) Carrying a bag. (b) Wearing a coat.

the classification results. Besides, the silhouette quality of the CMU MoBo database is very poor because there exists some preprocessing noises. However, our proposed MSR still achieves better performance than the alternative method which suggests that the subspace based representation is also robust to noise.

We also examined the performance under multiple intra-subject variations from both view and walking conditions. Experimental settings and corresponding results are shown in Figure 6.7. Compared with the results shown in Figure 6.6, we can conclude that multiple intra-subject variations degrade the performance of all methods. However, our proposed MSR still achieves better recognition results than the fusion methods by 10%, which shows that subspace representation can handle multiple intra-subject variations more effectively. Moreover, results from MCCA are also superior than those from WSD.

Table 6.4: First rank recognition rate of the proposed framework with missing data. Training views are from cameras 2, 4, 6, 8, 10 and testing views are from cameras as indicated in the table.

Views	4,6	3,7	5,7,9	3,7,11	3,5,7,9
Rates	1	1	0.95	1	1

6.3.2.5 Tolerance of MSR on missing data

In real applications, it is uncommon that one or some of the cameras in the multi-camera system may malfunction or there are occlusions from some viewing angles. In such cases, the number of testing view angles is less than that of the training view angles. As we have mentioned before, this would fail most of the conventional methods because they need to know the testing and training view correspondence. However, for our proposed representation method MSR, it has great tolerance on missing data. Along with the proposed learning-based algorithm MCCA which exploits more discriminative information for classification, the proposed framework is demonstrated (see Table 6.4) to be more effective under such condition compared with the results using WSD.

6.3.2.6 Tolerance of MSR on segmental gait cycle

Experiments in this section were conducted on CASIA-B database and only gait sequences from view angle 90° were used. Among the 6 normal walking sequences from each subject, 4 were chosen for training and the remaining 2 for testing. And for each training sequence, we extracted a complete cycle to learn the corresponding subspace. While for the testing sequences, we extracted different parts from the complete cycle for recognition. Because the recognition difficulty using different parts of a cycle were different, we conducted each experiment 15 times with different starting frames and regarded the mean as the final result. The period of gait was obtained through the algorithm presented in Chapter 3.

Table 6.5: Performance of methods using different lengths of a gait cycle.

Method	1 cycle	$\frac{3}{4}$ cycle	$\frac{1}{2}$ cycle	$\frac{1}{4}$ cycle	$\frac{1}{8}$ cycle
GEI	97.18	96.47	94.2	76.26	50.62
DTW	89.52	85.19	82.17	69.86	63.6
KLD [134]	91.13	89.67	88.84	85.9	81.31
EMD [133]	96.37	96.21	93.85	89.3	82.09
MSR	95.56	95.16	94.76	91.53	85.9

Experimental results are shown in Table 6.5. Clearly, the conventional methods (GEI and DTW) fail when the gait sequence is less than half a cycle. However, the methods proposed for set-to-set matching (KLD, EMD and MSR) can still work, the recognition rates are over 80% even with $\frac{1}{8}$ gait cycles (3-5 frames). Nevertheless, MSR performs the best and much faster than the other two. This suggests that the subspace representation is more robust to missing data as it can “fill-in” the missing information.

It can be seen that the recognition rate can reach above 80% with only a few number of frames. As gait dynamic would reduce greatly in such situation, the results prove that the static shape information contained in the gait pose is also very discriminative for recognition. Besides, as gait is a symmetry movement especially in the profile view, the recognition results from conventional methods is also promising with at least $\frac{1}{2}$ gait cycles.

6.3.3 Discussions

From the results mentioned above, we conclude as follows:

- (1) The proposed set-based subspace representation MSR achieves superior performance compared with conventional multiview gait recognition methods when the testing views are the same as the corresponding training views. Moreover, the proposed MSR not only can handle certain intra-subject variation, such as view

variation, walking with a coat or carrying a bag, but also can avoid the need of view estimation for each testing sample.

- (2) The proposed learning-based method MCCA can enhance the classification accuracy among linear subspaces. It is suitable for single-set based classification tasks which also make the proposed framework applicable when limited training samples are available for learning.
- (3) The proposed framework facilitates the applications of multiview gait recognition in practice. Suppose there is a surveillance system with multiple cameras, the MSR method can make it work even when the captured views are not the same with the registered views which usually happens, and can avoid the view estimation for each captured gait sequence. It can also handle some walking condition variations. Moreover, the proposed framework can still function well when some of the cameras break down or occlusions in some views occur.

6.4 Conclusions

In this chapter, we have proposed a framework for gait recognition with data from multiple views available. It contains two important steps: set-based feature representation MSR and single-set based discriminative learning MCCA. Different from conventional methods which treat each view separately, MSR collates all the samples from different views of the same subject and extracts the subspace bases for representation. Due to limited training samples in gait recognition, we generate only one subspace for each subject. Then, to better classify these linear subspaces, MCCA aims to learn a projection to differentiate the originally neighboring subspaces. At last, given the projections of both the training and testing subspaces, the largest canonical correlation is

defined as the distance metric and the nearest neighbor classifier is used for identification. Experimental results on the popular multiview gait databases have demonstrated that the proposed framework is not only superior than the existing methods in conventional recognition conditions but also effective in handling intra-subject variations and missing data. This makes it more suitable for practical applications.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

The implementation of intelligent video surveillance system relies greatly on human gait analysis because gait appears to be the most appropriate biometric for this application. Although gait recognition from side view and under normal walking condition has been well studied, it is not applicable as it ignores some challenges that usually happen in real scenario, including view variation, walking condition variation and so on. In this thesis, we have investigated such challenges and proposed new solutions to address these problems. In the following sections, we summarize the work of this thesis in three parts: dynamic feature analysis in gait recognition, gait recognition across varying views, and gait recognition across varying walking conditions.

7.1.1 Dynamic Feature Analysis in Gait Recognition

In this thesis, the first challenge we have investigated is the problem of how to effectively extract useful features from a gait sequence so that the representation is robust under clothing or carrying condition variations. It is reported that effects from these variations can be alleviated by conducting dynamic analysis on gait appearance feature. Therefore, we propose two dynamic feature extraction schemes based on the popular and widely used gait feature descriptor GEI. Firstly, a spatial dynamic feature

descriptor is proposed to select and weight the dynamic information by using feature selection mask and discriminative feature weight. Thus, the result feature records both within-class and between-class dynamic information. Then we propose a novel Spatial-Temporal Dynamic Gait Image (STDGI) which extracts spatial dynamic information by using the learned feature weight and temporal dynamic information through color interpolation. Experimental results on two popular gait databases demonstrate the effectiveness and robustness of the proposed dynamic features under varying clothing or carrying conditions.

7.1.2 Gait Recognition Across Varying Views

We have also addressed the view variation problem in gait recognition from two aspects: view-invariant gait recognition and view recognition of human gait sequences.

How to deal with view variation is a challenging research problem in gait recognition because the appearance of gait under different views could be totally different. In this thesis, we address this problem and propose a novel Joint Subspace Learning (JSL) method to solve it. Inspired by the fact that if a three-dimensional (3-D) object can be well represented by a linear combination of a small number of prototypes from the same view, then the representation coefficients with the same prototypes remain fairly similar across different views, we propose to use these representation coefficients as view-invariant features for recognition. To obtain the coefficients, we first conduct joint learning using Canonical Correlation Analysis (CCA) and Principle Component Analysis (PCA) to obtain the prototypes of different views. Then we represent the samples in both the gallery set and probe set acquired from different views as a linear combination of these prototypes in the corresponding views and extract the coefficients for recognition. In addition, we also develop a new descriptor, called Radon-transform-based Energy Image (REI), and propose to divide it into patches to further enhance

the performance. Experimental results demonstrate that the proposed JSL method outperforms the existing methods in the literature.

Although progress has been made in gait recognition across varying views, it still has limitations for applications. The reason is that most of the proposed view-invariant gait recognition methods assume the views of the testing gait sequences are known before recognition. However, this assumption may not hold in many practical applications. In this thesis, we address the problem of view recognition of human gait sequences and propose an effective algorithm for it. Motivated by the fact that in the problem of view recognition, the intra-class variation may be diverse due to identity variation, we propose to use multiple descriptors to represent the gait sequences because multiple feature representations usually characterize the data more precisely from various aspects. To better combine these multiple gait descriptors, we also propose a new subspace learning algorithm, adaptive discriminant analysis with enhanced multiple kernel learning (ADA-EMKL) which extracts low-dimensional features for view recognition. In our proposed method, high order kernels are generated to model the non-linear relationship between base kernels. Moreover, pairwise data correlation is applied to impose penalty to sample pairs that can be easily mis-classified. Experimental results on a popular multi-view gait database taken under different walking conditions demonstrate the efficacy as well as the robustness of the proposed approach.

7.1.3 Gait Recognition Across Varying Walking Conditions

In addition to view variation, other walking condition variations, such as walking speed, wearing a coat and carrying objects, also affect the gait appearance greatly. Moreover, usually more than one variation would happen at the same time in real applications. In this thesis, we propose a new multi-view subspace-based representation (MSR) method which would handle multiple walking condition variations. By assuming that the gait data from different views of the same subject lie in a low-dimensional

Figure 7.1: The diagram of a fully automatic view-invariant gait recognition system.

linear subspace, we treat them as a feature set and apply Singular Value Decomposition (SVD) on it to calculate the underlying subspaces. Then, the corresponding subspace basis is used for representation of the feature set. Thus, the distances between subjects become the distances between subspaces. We also proposed a new learning based method, marginal canonical correlation analysis (MCCA) to calculate the similarity. In addition, we also apply the proposed MSR to represent a gait sequence by regarding its frames as a feature set, thus handling different types of missing data. Experimental results on different gait databases demonstrate that the proposed subspace-based representation has tolerance on intra-subject variations, such as view variation, walking condition variations as well as missing data.

7.2 Future Work

Based on the research in this thesis and the existing problems in human gait analysis, we have identified some potential directions for future research in this section.

7.2.1 Fully Automatic View-Invariant Gait Recognition System

Studies in this thesis have shown that the performance of view recognition from human gait sequences could reach as high as 90% even under walking condition variations and confusion mainly exists within neighboring views. Moreover, good performance could

also be achieved in view-invariant gait recognition if the view of the testing sequence is known. Based on these studies, one possible direction for future work would be to implement a fully automatic view-invariant gait recognition system. We illustrate in Figure 7.1 the corresponding diagram of the system which mainly contains two steps: view recognition and then gait recognition. To alleviate the effects caused by mis-recognized views, prior information is added according to the corresponding confusion matrix from the view recognition step. Taken one of the confusion matrices we illustrate in chapter 5 for example (see Figure 7.2), it is not difficult to see from the figure that if the view of a testing gait sequence is recognized as 90° , then its real view may be 72° , 90° or 108° , and the probability is 12.6%, 82.0%, 4.5% respectively. By incorporating such information in the gait recognition step, one could expect to achieve a moderate final recognition result and alleviate the errors from the view recognition step.

7.2.2 Soft Biometrics From Gait

In addition to identity, the gait signal also carries other information, such as gender, age and race, which are also referred to soft biometrics. The possibility that we can obtain these soft biometrics from gait sequences has been shown in [116, 117, 118, 119, 120]. This would definitely expand the application of gait analysis to other research areas such as automatic collection of customer information at shopping malls or detection of abnormal behavior from elderly people in a care center.

It is reported that certain gait components, such as hip, shoulder and hair, contribute more for gender classification. Therefore, a potential future work may be the application of the proposed dynamic feature extraction method to extract and weight these particular gait components. For age estimation, it is shown in [122, 123] that we can obtain the age information of a walking subject with a small error. The method

Figure 7.2: A sample confusion matrix from view recognition of human gait sequences.

used is multi-label based subspace learning. In this thesis, we have proposed several subspace based learning methods for identification which could also be extended to age estimation.

However, it is not difficult to imagine that these soft biometrics can affect each other. Previous studies only focus on recognition of one of these soft biometrics and classification of multiple soft biometrics at the same time has rarely been touched which could be a potential direction for future gait recognition research. To facilitate such study, a large-scale database containing 168 participants has already been built [121]. In addition to the large number of subjects, the database also contains 88 males and 80 females between 4 to 75 years old to increase the diversity of gender and age.

Moreover, the samples are collected from 25 viewing angles.

7.2.3 Generic Database in Gait Analysis

Existing gait databases usually focus on specific recognition problem and thus only consider a small number of variables (conditions) for data collection. However, in real life applications, the condition could be much more complex than what has been studied in previous research. Therefore, a general gait database is in great demand for comprehensive research in gait analysis, which is also a potential direction for future research.

To make the database more convenient for view invariant gait recognition research, one could design the settings for data collection as follows: with a single camera fixed in a certain place (away from the walking subject such that the largest range of views could be captured); the walking subject is asked to walk along a circular path in front of the camera. Theoretically, we could have training data from different views in the range of $[0^\circ, 360^\circ]$. However, only a small number of frames are likely available for each view. In fact, in our study of this thesis, we have shown that the recognition performance could be as high as 80% with only 3 to 5 frames available. Therefore, such settings are feasible and we need to develop new algorithms to further enhance the performance.

Moreover, to have sufficient different walking condition variations, each selected subject should walk several times in front of the camera with different walking speeds, wearing conditions, carrying conditions and their combinations. To ensure the diversity in the database, the chosen subjects should have as many different races, gender, ages as possible.

7.2.4 Pre-Processing of Gait Video

In some real-world applications, the background is usually more complicated than that in the empirical study. Hence, pre-processing errors as we have illustrated in Figure 2.3 couldn't be avoided. Several studies have addressed this problem [93, 124] either through statistical modeling or techniques from other research areas such as image super-resolution. The study in this thesis has also shown that we can use subspace-based representation method to deal with certain kinds of missing data. However, all these efforts aim at alleviating the effects caused by the errors rather than removing them. Therefore, a potential future work could focus on how to reduce such errors in the pre-processing steps.

Publication

Journal Papers

- (i) N. Liu, J. Lu and Y.-P. Tan, “Joint subspace learning for view-invariant gait recognition,” *IEEE Signal Processing Letters*, vol. 18, no. 7, pp. 431-434, 2011.
- (ii) N. Liu, J. Lu and Y.-P. Tan, “Set-to-Set Gait Recognition Across Varying Views and Walking Conditions,” *Submitted to Journal of Vision Communication and Image Representation*, acceptable with minor revision, 2012.
- (iii) N. Liu, J. Lu and Y.-P. Tan, “View recognition of human gait sequences from videos”, *Submitted to IEEE Transactions on Information Forensics and Security*, require major revision, 2012.

Conference Papers

- (i) N. Liu, J. Lu, Y.-P. Tan, and Z. Chen, “Enhanced gait recognition based on weighted dynamic feature”, *IEEE International Conference on Image Processing*, pp. 3581-3584, 2009.
- (ii) N. Liu and Y.-P. Tan, “View invariant gait recognition”, *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1410-1413, 2010.
- (iii) N. Liu and Y.-P. Tan, “A novel study and analysis on segmental gait sequence recognition”, *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2011.

- (iv) N. Liu, J. Lu, Y.-P. Tan, and M. Li “Set-to-set gait recognition across varying views and walking conditions”, oral, *IEEE International Conference on Multimedia and Expo.*, 2011.

References

- [1] N. V. Boulgouris, D. Hatzinakos, and K. N. Plataniotis, “Gait recognition: a challenging signal processing technology for biometric identification,” *IEEE Signal Processing Magazine*, Vol. 22, No. 6, pp. 78-90, 2005.
- [2] G. Johansson, “Visual perception of biological motion and a model for its analysis,” *Percept. Psychophysics*, vol. 14, no. 2, pp. 201-211, 1973.
- [3] J.E. Cutting and L.T. Kozlowski, “Recognizing friends by their walk: gait perception without familiarity cues,” *Bulletin Psychonomic Society*, vol. 9, no. 5, pp. 353-356, 1977.
- [4] J. Cutting, D. Proffitt, and L. Kozlowski, “A biomechanical invariant for gait perception,” *Journal of Experimental Psychology: Human Perception and Performance*, vol. 4, no. 3, pp. 357-372, 1978.
- [5] G. Johansson, “Visual motion perception,” *Scientific American*, pp. 76-88, 1975.
- [6] S.A. Niyogi and E.H. Adelson, “Analyzing and recognizing walking figures in XYT,” *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 469-474, 1994.
- [7] H. Murase and R. Sakai, “Moving object recognition in eigenspace representation: gait analysis and lip reading,” *Pattern Recognition Letters*, vol. 17, no. 2, pp. 155-162, 1996.

- [8] D. Cunado, M.S. Nixon, and J.N. Carter, "Using gait as a biometric, via phaseweighted magnitude spectra," in *Proceedings of International Conference on Audio- and Video-Based Biometric Person Authentication*, pp. 95-102, 1997.
- [9] J. Little and J. Boyd, "Recognizing people by their gait: the shape of motion," *International Journal on Computer Vision*, vol. 14, no. 6, pp. 83-105, 1998.
- [10] P.S. Huang, C.J. Harris, and M.S. Nixon, "Visual surveillance and tracking of humans by face and gait recognition," in *Proceedings of 7th IFAC Symp. Artificial Intelligence in Real-Time Control*, pp. 43-44, 1998.
- [11] Jiwen Lu and Erhu Zhang, "Gait recognition for human identification based on ICA and fuzzy SVM through multiple views fusion," *Pattern Recognition Letters*, vol 28, no. 16, pp. 2401-2411, 2007.
- [12] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer, "The humanid gait challenge problem: data sets, performance, and analysis," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 27, no. 2, pp. 162-177, 2005.
- [13] R. Gross and J. Shi, "The CMU motion of body (mobo) database," *Technical report CMU-RI-TR-01-18*, 2001.
- [14] J. D. Shutler, M. G. Grant, M. S. Nixon, and J. N. Carter, "On a large sequence-based human gait database," In *Special Session on Biometrics, Proceedings of 4th International Conference on Recent Advances in Soft Computing*, pp. 66-71, 2002.
- [15] The Georgia Tech gait recognition database, <http://www.cc.gatech.edu/cpl/projects/hid/images.html>.

- [16] L. Wang, T. Tan, H. Ning, and W. Hu, "Silhouette analysis-based gait recognition for human identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1505-1518, 2003.
- [17] S. Yu, D. Tan, and T. Tan, "A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition," *Proceedings - International Conference on Pattern Recognition*, pp. 441-444, 2006.
- [18] HID-UMD. The UMD gait recognition database, <http://degas.umiacs.umd.edu/Hid/data.html>.
- [19] The MIT gait recognition database, <http://www.ai.mit.edu/people/llee/HID/data.htm>.
- [20] P. Kuchi, R. R. Hiremagalur, H. Huang, M. Carhart, J. He, and S. Panchanathan, "DRAG: a database for recognition and analysis of gait," *Proceedings of SPIE - The International Society for Optical Engineering*, pp. 115-124, 2003.
- [21] H. Sakoe and R. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech, Signal Processing*, vol. 26, no. 1, pp. 43-49, 1978.
- [22] R. Polana and R. Nelson, "Detection and recognition of periodic, non-rigid motion," *International Journal of Computer Vision*, vol. 23, no. 3, pp. 261-282, 1997.
- [23] R. Cutler and L. Davis, "Robust real-time periodic motion detection, analysis and applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 2, pp. 129-155, 2000.
- [24] I. Haritaoglu, R. Cutler, D. Harwood, and L. Davis, "Backpack: detection of people carrying objects using silhouettes," *Computer Vision and Image Understanding*, vol. 6, no. 2, pp. 385-397, 2000.

- [25] C. BenAbdelkader, R. Cutler, and L. Davis, "Stride and cadence as a biometric in automatic person identification and verification," *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 372-377, 2002.
- [26] S. Yu, L. Wang, W. Hu, and T. Tan, "Gait analysis for human identification in frequency domain," *In Proceedings of the 3rd International Conference on Image and Graphics*, pp. 282-285, 2004.
- [27] M. Ekinici, "Gait recognition using multiple projections," *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 517-522, 2006.
- [28] L. Wang, H. Ning, T. Tan, and W. Hu, "Fusion of static and dynamic body biometrics for gait recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 2, pp. 149-158, 2004.
- [29] J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 2, pp. 316-322, 2006.
- [30] Q. Ma, S. Wang, D. Nie, and J. Qiu, "Gait recognition at a distance based on energy deviation image," *IEEE International Conference on Bioinformatics and Biomedical Engineering*, pp. 621-624, 2007.
- [31] D. Tao, X. Li, X. Wu, and S.J. Mabank, "General tensor discriminant analysis and Gabor features for gait recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 10, pp. 1700-1715, 2007.
- [32] N.V. Boulgouris and Z.X. Chi, "Gait recognition using Radon transform and linear discriminant analysis," *IEEE Transactions on Image Processing*, vol. 16, no. 3, pp. 731-740, 2007.

- [33] S. Tabbone, L. Wendling, and J.P. Salmon, “A new shape descriptor defined on radon transform,” *Computer Vision and Image Understanding*, vol. 102, no. 1, pp. 42-51, 2006.
- [34] Y. Chai, J. Ren, R. Zhao, and J. Jia, “Automatic gait recognition using dynamic variance features,” *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 475-480, 2006.
- [35] L. Gu and F. Sun, “A novel gait recognition method based on standard deviation energy image,” *IEEE Proceedings of the Chinese Conference on Pattern Recognition*, pp. 293-298, 2008.
- [36] D. Tan, K. Huang, S. Yu, and T. Tan, “Uniprojective feature for gait recognition,” *In Proceedings of The 2nd International Conference on Biometrics*, pp. 673-682, 2007.
- [37] D. Tan, S. K. Huang, and T. Tan, “Walker recognition without gait cycle estimation,” *In Proceedings of The 2nd International Conference on Biometrics*, pp. 222-231, 2007.
- [38] D. Tan, K. Huang, S. Yu, and T. Tan, “Orthogonal diagonal projections for gait recognition,” *IEEE International Conference on Image Processing*, pp. 337-340, 2007.
- [39] M. Romero-Moreno, J. Fco. Martnez-Trinidad, and J.A. Carrasco-Ochoa, “Gait recognition based on silhouette, contour and classifier ensembles,” *Lecture Notes in Computer Science*, vol. 5197, pp. 527-534, 2008.
- [40] A. F. Bobick and J.W. Davis, “An appearance-based representation of action,” *IEEE International Conference on Pattern Recognition*, pp. 307-312, 1996.

- [41] A. Davis and A. Bobick, "The representation and recognition of action using temporal templates," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 928-934, 1997.
- [42] J. Liu and N. Zheng, "Gait history image: a novel temporal template for gait recognition," *IEEE International Conference on Multimedia and Expo*, pp. 663-666, 2007.
- [43] C. Wang, J. Zhang, J. Pu, X. Yuan, and L. Wang, "Chrono-gait image: a novel temporal template for gait recognition," *European Conference on Computer Vision*, pp. 257-270, 2010.
- [44] L. Wang, H. Ning, T. Tan, and W. Hu, "Fusion of static and dynamic body biometrics for gait recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.14, no.2, pp.149-158, 2004.
- [45] A. Y. Johnson and A. F. Bobick, "A multi-view method for gait recognition using static body parameters," *International Conference on Audio and Video Based Biometric Person Authentication*, pp. 301-311, 2001.
- [46] L. Lee and W. E. L. Grimson, "Gait analysis for recognition and classification," *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 155-162, 2002.
- [47] R. Urtasun and P. Fua, "3d tracking for gait characterization and recognition," *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 17-22, 2004.
- [48] D. K. Wagg and M. S. Nixon, "On automated model-based extraction and analysis of gait," *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 11-16, 2004.

- [49] C. Y. Yam, M. S. Nixon, and J. N. Carter, “On the relationship of human walking and running: automatic person identification by gait,” *IEEE International Conference on Pattern Recognition*, pp. 287-290, 2002.
- [50] R. Zhang, C. Vogler, and D. Metaxas, “Human gait recognition at sagittal plane,” *Image Vision Computing*, vol. 25, no. 3, pp. 321-330, 2007.
- [51] S. Wachter and H. Nagel, “Tracking persons in monocular image sequences,” *Computer Vision and Image Understanding*, vol. 74, no. 3, pp. 174-192, 1999.
- [52] Q. Delamarre and O. Faugeras, “3D articulated models and multi-view tracking with physical forces,” *Computer Vision and Image Understanding*, vol. 81, pp. 328-357, 2001.
- [53] C. Sminchisescu and B. Triggs, “Covariance scaled sampling for monocular 3D body tracking,” *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 447-454, 2001.
- [54] A. Kale, A. Sundaresan, A. N. Rajagopalan, N. P. Cuntoor, A. K. Roy-Chowdhury, V. Krüger, and R. Chellappa, “Identification of humans using gait,” *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1163-1173, 2004.
- [55] Z. Liu and S. Sarkar, “Improved gait recognition by gait dynamics normalization,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 6, pp. 863-876, 2006.
- [56] H. Suk and B. Sin, “HMM-based gait recognition with human profiles,” *Lecture Notes in Computer Science*, vol. 109, no. 4, pp. 596-603, 2006.
- [57] C. Chen, J. Liang, and X. Zhu, “Gait recognition based on improved dynamic Bayesian networks,” *Pattern Recognition*, vol. 44, no. 4, pp. 988-995, 2011.

- [58] J. Gu, X. Ding, S. Wang, and Y. Wu, "Action and gait recognition from recovered 3-D human joints," *IEEE Transactions on Systems, Man, and Cyberneticspart B: Cybernetics*, vol. 40, no. 4, pp. 1021-1033, 2010
- [59] C. Chen, J. Liang, H. Zhao, H. Hu, and J. Tian, "Factorial HMM and parallel HMM for gait recognition," *IEEE Transactions on Systems, Man, and Cyberneticspart C: applications and reviews*, vol. 39, no. 1, pp. 114-123, 2009.
- [60] A. Sundaresan, A. RoyChowdhury, and R. Chellappa, "A hidden markov model based framework for recognition of humans from gait sequences," *IEEE International Conference on Image Processing*, pp. 93-96, 2003.
- [61] A. Veeraraghavan, A. Srivastava, A.K. Roy-Chowdhury and R. Chellappa, "Rate-invariant recognition of humans and their activities," *IEEE Transactions on Image Processing*, vol. 18, no. 6, pp. 1326-1339, 2009.
- [62] N. Cuntoor, A. Kale, and R. Chellappa, "Combining multiple evidences for gait recognition," *IEEE International Conference on Accoustics, Speech, and Signal Processing*, pp. 33-36, 2003.
- [63] C.-S. Lee and A. Elgammal, "Towards scalable view-invariant gait recognition: multilinear analysis for gait," *International Conference on Audio - and Video-Based Biometric Person Authentication*, pp. 395-405, 2005.
- [64] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi, "Gait recognition using a view transformation model in the frequency domain," *European Conference on Computer Vision*, pp. 151-163, 2006.
- [65] A. Tyagi, J.W. Davis, and M. Keck, "Multiview fusion for canonical view generation based on homography constraints," *ACM International Multimedia Conference*, pp. 61-70, 2006.

- [66] N.V. Boulgouris and X.Huang, “Human gait recognition based on multiview gait sequences,” *Eurasip Journal on Advances in Signal Processing*, 2008.
- [67] W. Kusakunniran, Q. Wu, H. Li, and J. Zhang, “Multiple views gait recognition using view transformation model based on optimized gait energy image,” *IEEE International Conference on Computer Vision Workshops*, pp. 1058-1064, 2009.
- [68] E. Ribnick and N. Papanikolopoulos, “3D reconstruction of periodic motion from a single view,” *International Journal of Computer Vision*, vol. 90, no. 1, pp. 28-44, 2010.
- [69] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li, “Cross-view and multi-view gait recognitions based on view transformation model using multi-layer perceptron,” *Pattern Recognition Letters*, Article in press, 2011.
- [70] G. Shakhnarovich, L. Lee, and T. Darrell, “Integrated face and gait recognition from multiple views,” *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 439-446, 2001.
- [71] Y. Wang, S. Yu, Y. Wang, and T. Tan, “Gait recognition based on fusion of multi-view gait sequences,” *International Conference on Biometrics*, pp. 605-611, 2005.
- [72] R. Bodor, A. Drenner, D. Fehr, O. Masoud, and N. Papanikolopoulos, “View-independent human motion classification using image-based reconstruction,” *Image and Vision Computing*, vol. 27, no. 8, pp. 1194-1206, 2009.
- [73] M. Goffredo, I. Bouchrika, J.N. Carter, and M.S. Nixon, “Self-calibrating view-invariant gait biometrics,” *IEEE Transactions on Systems, Man, and Cybernetics, part B*, vol. 40, no. 4, pp. 997-1008, 2010.

- [74] J. Lu and Y.-P. Tan, "Uncorrelated discriminant simplex analysis for view-invariant gait signal computing," *Pattern Recognition Letters*, vol. 31, no. 5, pp. 382-393, 2010.
- [75] K. Bashir, T. Xiang, and S. Gong, "Cross-View gait recognition using correlation strength," *Proceedings of the British Machine Vision Conference*, pp. 190.1-190.11, 2010.
- [76] C. BenAbdelkader and R. Cutler, "View-invariant estimation of height and stride for gait recognition," *ECCV Workshop on Biometric Authentication*, pp. 155-167, 2002.
- [77] A. Kale, A. Chowdhury, and R. Chellappa, "Towards a view invariant gait recognition algorithm," *IEEE Conference on Advanced Video and Signal based Surveillance*, pp. 143-150, 2003.
- [78] S. Jeong, S. Kim, and B. Choi, "Canonical view synthesis for gait recognition," *IEEE Frontiers in the Convergence of Bioscience and Information Technologies*, pp. 734-739, 2007.
- [79] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li, "Support vector regression for multi-view gait recognition based on local motion feature selection," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 974-981, 2010.
- [80] J. P. Foster, M. S. Nixon, and A. Prgel-Bennet, "Automatic gait recognition using area-based metrics," *Pattern Recognition Letters*, vol. 24, no. 14, pp. 2489-2497, 2003.
- [81] S. Singh and K.K. Biswas, "Biometric gait recognition with carrying and clothing variants," *Lecture Notes in Computer Science*, vol. 5909, pp. 446-451, 2009.

- [82] M.R. Aqmar, K. Shinoda, and S. Furui, “Robust gait recognition against speed variation,” *IEEE International Conference on Pattern Recognition*, pp. 2190-2193, 2010.
- [83] R. Tanawongsuwan and A. Bobick, “Performance analysis of time-distance gait parameters under different speeds,” *Lecture Notes in Computer Science*, vol. 2688, pp. 715-724, 2003.
- [84] X. Huang and N.V. Boulgouris, “Gait recognition using linear discriminant analysis with artificial walking conditions,” *IEEE International Conference on Image Processing*, pp. 2461-2464, 2010.
- [85] X. Huang and N.V. Boulgouris, “Gait recognition for random walking patterns and variable body postures,” *IEEE International Conference on Acoustics, Speech and Signal Processin*, pp. 1726-1729, 2010.
- [86] Y. Ran, Q. Zheng, R. Chellappa, and T.M. Strat, “Applications of a simple characterization of human gait in surveillance,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 40, no. 4, pp. 1009-1020, 2010.
- [87] H. Lee, E. Kim, and S. Hong, “An efficient gait recognition with backpack removal,” *Eurasip Journal on Advances in Signal Processing*, vol. 2009, 2009.
- [88] D. Kim, D. Kim, and J. Paik, “Model-based gait recognition using multiple feature detection,” *Lecture Notes in Computer Science*, vol. 5259, pp. 1018-1029, 2008.
- [89] A. Tsuji, Y. Makihara, and Y. Yagi, “Silhouette transformation based on walking speed for gait identification,” *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 717-722, 2010.

- [90] R. Tanawongsuwan and A. Bobick, “Modelling the effects of walking speed on appearance-based gait recognition,” *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 783-790, 2004.
- [91] K. Bashir, T. Xiang, and S. Gong, “Feature selection on gait energy image for human identification,” *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 985-988, 2008.
- [92] D. Tao, X. Li, X. Wu, and S. J. Maybank, “Human carrying status in visual surveillance,” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1670-1677, 2006.
- [93] Z. Liu and S. Sarkar, “Effect of silhouette quality on hard problems in gait recognition,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 35, no. 2, pp. 170-183, 2005.
- [94] C. Chen, J. Liang, H. Zhao, H. Hua, and J. Tian, “Frame difference energy image for gait recognition with incomplete silhouettes,” *Pattern Recognition Letters*, vol. 30, no. 11, pp. 977-984, 2009.
- [95] C. BenAbdelkader, R.G. Cutler, and L.S. Davls, “Gait recognition using image self-similarity,” *Eurasip Journal on Applied Signal Processing*, vol. 2004, no. 4, pp. 572-585, 2004.
- [96] L. Cui and H. Li, “A new combined fractal scale descriptor for gait sequence,” *Lecture Notes in Computer Science*, vol. 4571, pp. 616-627, 2007.
- [97] S. Lee, Y. Liu, and R. Collins, “Shape variation-based frieze pattern for robust gait recognition,” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1-8, 2007.

- [98] R. Martn-Flez, R.A. Mollineda, and J. S. Sánchez, “Towards a more realistic appearance-based gait representation for gender recognition,” *IEEE International Conference on Pattern Recognition*, pp. 3810-3813, 2010.
- [99] M. Hu, Y. Wang, Z. Zhang, and Y. Wang, “Combining spatial and temporal information for gait based gender classification,” *IEEE International Conference on Pattern Recognition*, pp. 3679-3682, 2010.
- [100] D. Zhang, Y. Wang, and B. Bhanu, “Ethnicity classification based on gait using multi-view fusion,” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition- Workshops*, pp. 108-115, 2010.
- [101] X. Li, S.J. Maybank, S. Yan, D. Tao, and D. Xu, “Ethnicity classification based on gait using multi-view fusion,” *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, vol. 38, no. 2, pp. 145-155, 2008.
- [102] S. Yu, T. Tan, K. Huang, K. Jia, and X. Wu, “A study on gait-based gender classification,” *IEEE Transactions on Image Processing*, vol. 18, no. 8, pp. 1905-1910, 2009.
- [103] J. Lu and Y.-P. Tan, “Gait-based human age estimation,” *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 4, pp. 761-770, 2010.
- [104] X. Yang, Y. Zhou, T. Zhang, E. Zheng, and J. Yang, “Gabor phase based gait recognition,” *Electronic Letters*, vol. 44, no. 10, pp. 620-621, 2008.
- [105] Y. Huang, D. Xu, and T.-J. Cham, “Face and human gait recognition using image-to-class distance,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 3, pp. 431-438, 2010.
- [106] V. Kellokumpu, G. Zhao, S. Z. Li, and M. Pietikäinen, “Dynamic texture based gait recognition,” *Lecture Notes in Computer Science*, vol. 5558, pp. 1000-1009, 2009.

- [107] W. Kusakunniran, Q. Wu, H. Li, and J. Zhang, "Automatic gait recognition using weighted binary pattern on video," *IEEE International Conference on Advanced Video and Signal Based Surveillance*, pp. 49-54, 2009.
- [108] X. Yang, Y. Zhou, T. Zhang, G. Shu, and J. Yang, "Gait recognition based on dynamic region analysis," *Signal Processing*, vol. 88, no. 9, pp. 2350-2356, 2008.
- [109] K. Bashir, T. Xiang, and S. Gong,, "Featruue selection for gait recognition without subject cooperation," *Pattern Recognition Letters*, vol. 31, pp. 2052-2060, 2010.
- [110] P. N. Belhumenur, J. P. Hepanha, and D. J. Kriegman, "Eigenfaces vs. fisherface: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, 1997.
- [111] S. Hong, H. Lee, I. F. Nizami, and E. Kim, "A new gait representation for human identification: mass vector," *IEEE Conference on Industrial Electronics and Applications*, pp. 669-673, 2007.
- [112] S. Tabbone, L. Wendling, and J.-P. Salmon, "A new shape descriptor defined on the radon transform," *Computer Vision and Image Understanding*, vol. 102, no. 1, pp. 42-51, 2006.
- [113] A. Sharma, A. Dubey, P. Tripathi, and V. Kumar, "Pose invariant virtual classifiers from single training image using novel hybrid-eigenfaces," *Neurocomputing*, vol. 73, no. 10-12, pp. 1868-1880, 2010.
- [114] R. Souvenir and K. Parrigan, "Viewpoint manifolds for action recognition," *EURASIP Journal on Image and Video Processing*, 2009.
- [115] S. Srisuk, M. Petrou, W. Kurutach, and A. Kadyrov, "Face authentication using the trace transform," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 305-312, 2003.

- [116] R. Begg, "Support vector machines for automated gait classification," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 5, pp. 828-838, 2005.
- [117] J. Davis, "Visual categorization of children and adult walking styles," *In Proceedings of the International Conference on Audio-and Video-based Biometric Person Authentication*, pp. 295-300, 2001.
- [118] X. Li, S. Maybank, S. Yan, D. Tao, and D. Xu, "Gait components and their application to gender recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, vol. 38, no. 2, pp. 145-155, 2008.
- [119] S. Yu, T. Tan, K. Huang, K. Jia, and X. Wu, "A study on gait-based gender classification," *IEEE Transactions on Image Processing*, vol. 18, no. 8, pp. 1905-1910, 2009.
- [120] J. Yoo, D. Hwang, and M. Nixon, "Gender classification in human gait using support vector machine," *In Advanced Concepts For Intelligent Vision Systems*, pp. 138-145, 2006.
- [121] Y. Makihara, H. Mannami, and Y. Yagi, "Gait analysis of gender and age using a large-scale multi-view gait database," *Asian Conference on Computer Vision*, pp. 975-986, 2010.
- [122] J. Lu and Y.-P. Tan, "Gait-based human age estimation," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 4, pp. 761-770, 2010.
- [123] J. Lu and Y.-P. Tan, "Cost-sensitive subspace learning for human age estimation," *IEEE International Conference on Image Processing*, pp. 1593-1596, 2010.
- [124] J. Zhang, J. Pu, C. Chen, and R. Fleischer, "Low-resolution gait recognition," *IEEE Transactions on Systems, Man, and Cybernetics Part B: Cybernetics*, vol. 40, no. 4, pp. 986-996, 2010.

- [125] R.T. Collins, R. Gross, and J. Shi, "Silhouette-based human identification from body shape and gait," *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 351-356, 2002.
- [126] G. Zhao, G. Liu, H. Li, and M. Pietikainen, "3D gait recognition using multiple cameras," *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 529-534, 2006.
- [127] X. Huang and N.V. Boulgouris, "Gait recognition using multiple views," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1705-1708, 2008.
- [128] F. Li, Q. Dai, W. Xu, and G. Er, "Weighted subspace distance and its applications to object recognition and retrieval with image sets," *IEEE Signal Processing Letters*, vol. 16, no. 3, pp. 227-230, 2009.
- [129] T.-K. Kim, J. Kittler, and R. Cipolla, "Discriminative learning and recognition of image set classes using canonical correlations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1005-1018, 2007.
- [130] O. Yamaguchi, Ka. Fukui, and K. Maeda, "Face recognition using temporal image sequence," *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 318-323, 1998.
- [131] J.J. Gerbrands, "On the relationships between SVD, KLT and PCA," *Pattern Recognition*, vol. 14, no. 1-6, pp. 375-381, 1981.
- [132] J. Kittler, R.P.W. Duin, M. Hatef, and J. Matas, "On combining classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226-239, 1998.

- [133] J. Li, Y. Wang, and T. Tan, "Video-based face recognition using earth mover's distance," *Lecture Notes on Computer Science*, pp. 229-238, 2005.
- [134] G. Shakhnarovich, J. W. Fisher, and T. Darrel, "Face recognition from long-term observations," *Proceedings of European Conference on Computer Vision*, pp. 851-868, 2002.
- [135] H. Cevikalp, M. Neamtu, M. Wilkes, and A. Barkana, "Discriminative common vectors for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 1, pp. 4-13, 2005.
- [136] E. Zhang, Y. Zhao, and W. Xiong, "Active energy image plus 2DLPP for gait recognition," *Signal Processing*, vol. 90, pp. 2295-2302, 2010.
- [137] P. S. Huang, "Automatic gait recognition via statistical approaches for extended template features," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 31, no. 5, pp. 818-824, 2001.
- [138] A. Bobick and A. Johnson, "Gait recognition using static, activity-specific parameters," *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 423-430, 2001.
- [139] C. Benabdelkader, R. Culter, H. Nanda, and L. Davis, "Eigengait: motion-based recognition of people using image self-similarity," *Proceedings of International Conferene on Audio- and Video-Based Biometric Person Authentication*, pp. 284-294, 2001.
- [140] S. Yu, D. Tan, and T. Tan, "Modeling the effect of view angle variation on appearance-based gait recognition," *Proceedings of Asian Conference on Computer Vision*, pp. 807-816, 2006.

- [141] B. Schölkopf, A. Smola, and K.-R. Müller, “Nonlinear component analysis as a kernel eigenvalue problem,” *Neural Computation*, vol. 10, no. 5, pp. 1299-1319, 1998.
- [142] C. Cortes and V. Vapnik, “Support vector networks,” *Machine Learning*, vol. 20, pp. 273-297, 1995.
- [143] B. Scholkopf, S. Mika, C. J. C. Burges, P. Knirsch, K.-R. Muller, G. Ratsch, and A. J. Smola, “Input space vs. feature space in kernel-based methods,” *IEEE Transactions on Neural Networks*, vol. 10, pp. 1000-1017, 1999.
- [144] F. Bach, G. Lanckriet, and M. Jordan, “Multiple kernel learning, conic duality, and the SMO algorithm,” *International Conference on Machine Learning*, pp. 41-48, 2004.
- [145] M. Gönen and E. Alpaydin, “Localized multiple kernel learning,” *International Conference on Machine Learning*, pp. 352-359, 2008.
- [146] J. Lu and Y.-P. Tan, “View recognition of human gait sequences in videos,” *IEEE International Conference on Image Processing*, pp. 2457-2460, 2010.
- [147] G. Lanckriet, N. Cristianini, P. Bartlett, L. Ghaoui, and M. Jordan, “Learning the kernel matrix with semidefinite programming,” *Journal of Machine Learning Research*, vol. 5, pp. 27-72, 2004.
- [148] A. Vedaldi, V. Gulshan, M. Varma, and A. Zisserman, “Multiple kernels for object detection,” *In Proceedings of IEEE International Conference on Computer Vision*, pp. 606-613, 2009.
- [149] A. Rakotomamonjy, F. Bach, S. Canu, and Y. Grandvalet, “More efficiency in multiple kernel learning,” *In Proceedings of International Conference on Machine Learning*, pp. 775-782, 2007.

- [150] S. Sonnenburg, G. Rätsch, C. Schäfer, and B. Schölkopf, “Large scale multiple kernel learning,” *Journal on Machine Learning Research*, vol. 7, pp. 1531-1565, 2006.
- [151] M. Turk and A. Pentland, “Eigenfaces for recognition,” *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71-86, 1991.
- [152] X. He, S. Yan, Y. Hu, P. Niyogi, and H. J. Zhang, “Face recognition using Laplacianfaces,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 328-340, 2005.
- [153] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, “Graph embedding and extensions: A general framework for dimensionality reduction,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no.1, pp. 40-51, 2007.
- [154] Y.-Y. Lin, T.-L. Liu, and C.-S. Fuh, “Multiple kernel learning for dimensionality reduction,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 6, pp. 1147-1160, 2011.
- [155] S. Mitra, M. Savvides, and B. V. Kumar, “Face identification using novel frequency-domain representation of facial asymmetry,” *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 3, pp. 350-359, 2006.
- [156] M. Sargin, Y. Yemez, E. Erzin, and A. Tekalp, “Audiovisual synchronization and fusion using canonical correlation analysis,” *IEEE Transactions on Multimedia*, vol. 9, no. 7, pp. 1396-1403, 2007.
- [157] Y. Fu, L. Cao, G. Guo, and T. S. Huang, “Multiple feature fusion by subspace learning,” *Proceedings of the International Conference on Content-based Image and Video Retrieval*, pp. 127-134, 2008.

- [158] C. Saunders, A. Gammerman, and V. Vovk, “Ridge regression learning algorithm in dual variables,” *In International Conference on Machine Learning*, pp. 515-521, 1998.
- [159] H.-T. Chen, H.-W. Chang, and T.-L. Liu, “Local discriminant embedding and its variants,” *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 846-853, 2005.
- [160] S. Wang, S. Jiang, Q. Huang, and Q. Tian, “Multiple kernel learning with high order kernels,” *IEEE International conference on Pattern Recognition*, pp. 2138-2141, 2010.
- [161] C. Cortes, M. Mohri, and A. Rostamizadeh, “Learning non-linear combinations of kernels,” *Neural Information Processing Systems (NIPS)* 2009.
- [162] C. Berg, J. P. R. Christensen, and P. Ressel, “Harmonic analysis on semigroups,” *Springer-Verlag: Berlin-New York*, 1984.
- [163] The CASIA gait database. <Available at <http://www.sinobiometrics.com>>.
- [164] R.T. Collins, R. Gross, and J. Shi, “Silhouette-based human identification from body shape and gait,” *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 351-356, 2002.
- [165] T. Vetter and T. Poggio, “Linear object classes and image synthesis from a single example image,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 733-742, 1997.
- [166] A. Sharma, A. Dubey, P. Tripathi, and V. Kumar, “Pose invariant virtual classifiers from single training image using novel hybrid-eigenfaces,” *Neurocomputing*, vol. 73, no. 10-12, pp. 1868-1880, 2010.

- [167] S. Yu, D. Tan, and T. Tan, "Modelling the effect of view angle variation on appearance-based gait recognition," *IEEE Asian Conference on Computer Vision*, pp. 807-816, 2006.
- [168] F. Niu and M. Abdel-Mottaleb, "View-invariant human activity recognition based on shape and motion features," *IEEE International Symposium on Multimedia Software Engineering*, pp. 546-556, 2004.
- [169] N. Liu and Y.-P. Tan, "View invariant gait recognition," *IEEE International Conference on Acoustics Speech and Signal Processing*, pp. 1410-1413, 2010.
- [170] S. Prince, J. H. Elder, J. Warrell, and F. M. Felisberti, "Tied factor analysis for face recognition across large pose differences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 6, pp. 970-984, 2008.
- [171] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, pp. 321-377, 1936.
- [172] J.R. Kettenring, "Canonical analysis of several sets of variables," *Biometrika*, vol. 58, no. 3, pp. 433-451, 1971.
- [173] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 696-710, 1997.
- [174] J. Lu and Y.-P. Tan, "Regularized locality preserving projections and its extensions for face recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 40, no. 3, pp. 958-963, 2010.
- [175] X. Jiang, B. Mandal, and A. Kot, "Eigenfeature regularization and extraction in face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, pp. 383-394, 2008.

- [176] S. Prabhakar and A. K. Jain, “Decision-level fusion in fingerprint verification,” *Pattern Recognition*, vol. 35, no. 4, pp. 861-873, 2002.
- [177] W. Hwang, H. Wang, H. Kim, S.-C. Kee, and J. Kim, “Face recognition system using multiple face model of hybrid fourier feature under uncontrolled illumination variation,” *IEEE Transactions on Image Processing*, vol. 20, no. 4, pp. 1152-1165, 2011.
- [178] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, and K.-R. Müller, “Fisher discriminant analysis with kernels,” in *IEEE Neural Networks for Signal Processing*, pp. 41-48, 1999.
- [179] X. Liu and T. Chen, “Video-based face recognition using adaptive hidden markov models,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, pp. 340-345, 2003.
- [180] S. Satoh, “Comparative evaluation of face sequence matching for content-based video access,” *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 163-168, 2000.
- [181] T. Kailath, “A view of three decades of linear filtering theory,” *IEEE Transaction on Information Theory*, vol. 20, no. 2, pp. 146-181, 1974.
- [182] R. O. Duda, P. E. Hart, and D. G. Stork, “Pattern classification,” *second de.*, John Wily and Sons, 2000.
- [183] K. Fukui and O. Yamaguchi, “Face recognition using multi-viewpoint patterns for robot vision,” *Proceedings on Int’l Symp. Robotics Research*, pp. 192-201, 2003.
- [184] M. Nishiyama, O. Yamaguchi, and K. Fukui, “Face recognition with the multiple constrained mutual subspace method,” *Lecture Notes on Computer Sciene*, pp. 71-80, 2005.

- [185] H. Li, T. Jiang, and K. Zhang, “Efficient and robust feature extraction by maximum margin criterion,” *IEEE Transactions on Neural Networks*, vo. 17, no. 1, pp. 157-165, 2006.
- [186] F. Wang and C. Zhang, “Feature extraction by maximizing the average neighborhood margin,” *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2190-2193, 2007.
- [187] P. Han, A. Jin, and F. Abas, “Neighbourhood preserving discriminant embedding in face recognition,” *Journal of Visual Communication and Image Representation*, vol. 20, pp. 532-542, 2009.
- [188] P. Han, A. Jin, and T. Ann, “Kernel Discriminant Embedding in face recognition,” *Journal of Visual Communication and Image Representation*, vol. 22, pp. 634-642, 2011.
- [189] Å. Björck and G. H. Golub, “Numerical methods for computing angles between linear subspaces,” *Math. Computation*, vol. 27, no. 123, pp. 579-594, 1973.