Robust gait recognition under variable covariate conditions

Bashir, Khalid

Robust Gait Recognition Under Variable Covariate Conditions

Khalid Bashir

Submitted to the University of London in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Queen Mary University of London

2010

Robust Gait Recognition Under Variable Covariate Conditions

Khalid Bashir

Abstract

Gait is a weak biometric when compared to face, fingerprint or iris because it can be easily affected by various conditions. These are known as the covariate conditions and include clothing, carrying, speed, shoes and view among others. In the presence of variable covariate conditions gait recognition is a hard problem yet to be solved with no working system reported.

In this thesis, a novel gait representation, the Gait Flow Image (GFI), is proposed to extract more discriminative information from a gait sequence. GFI extracts the relative motion of body parts in different directions in separate motion descriptors. Compared to the existing model-free gait representations, GFI is more discriminative and robust to changes in covariate conditions.

In this thesis, gait recognition approaches are evaluated without the assumption on cooperative subjects, i.e. both the gallery and the probe sets consist of gait sequences under different and unknown covariate conditions. The results indicate that the performance of the existing approaches drops drastically under this more realistic set-up. It is argued that selecting the gait features which are invariant to changes in covariate conditions is the key to developing a gait recognition system without subject cooperation. To this end, the Gait Entropy Image (GEnI) is proposed to perform automatic feature selection on each pair of gallery and probe gait sequences. Moreover, an Adaptive Component and Discriminant Analysis is formulated which seamlessly integrates the feature selection method with subspace analysis for fast and robust recognition.

Among various factors that affect the performance of gait recognition, change in viewpoint poses the biggest problem and is treated separately. A novel approach to address this problem is proposed in this thesis by using Gait Flow Image in a cross view gait recognition framework with the view angle of a probe gait sequence unknown. A Gaussian Process classification technique is formulated to estimate the view angle of each probe gait sequence. To measure the similarity of gait sequences across view angles, the correlation of gait sequences from different views is modelled using Canonical Correlation Analysis and the correlation strength is used as a similarity measure. This differs from existing approaches, which reconstruct gait features in different views through 2D view transformation or 3D calibration. Without explicit reconstruction, the proposed method can cope with feature mis-match across view and is more robust against feature noise.

Submitted to the University of London in partial fulfilment of the requirements for

the degree of Doctor of Philosophy

Queen Mary University of London

2010

Declaration

I hereby declare that this thesis has been composed by myself and that it describes my own work. It has not been submitted, either in the same or different form, to this or any other university for a degree. All verbatim extracts are distinguished by quotation marks, and all sources of information have been acknowledged.

Some parts of the work have previously been published as:

- K. Bashir, T. Xiang and S. Gong. Gait recognition without subject cooperation. Pattern Recognition Letters, Vol. 31, No. 13, pp. 2052-2060, October 2010.
- K. Bashir, T. Xiang and S. Gong. Cross view gait recognition using correlation strength. In British Machine Vision Conference, 2010.
- K. Bashir, T. Xiang and S. Gong. Gait representation using flow fields. In British Machine Vision Conference, 2009.
- K. Bashir, T. Xiang and S. Gong. Gait recognition using gait entropy image. In 3rd International Conference on Crime Detection and Prevention, 2009.
- K. Bashir, T. Xiang and S. Gong. Feature selection for gait recognition without subject cooperation. In British Machine Vision Conference, 2008.

Khalid Bashir.
London, October 2010.

Acknowledgements

It is the grace of Allah and the love of the Prophet Muhammad (P.B.U.H) that have lead me through all those years. It has been the prayers of my parents, sisters and brother, the support of my wife (Ayesha Snover) and the enjoyable side of life provided by my sons (Muhammad Ahmad Bajwa and Muhammad Ibrahim Bajwa) that I have heavily relied on during my research at Queen Mary.

Nothing and absolutely nothing could have replaced the endless support and encouragement of my supervisor Dr. Tao Xiang. I am greatly thankful to Dr. Tao Xiang for all this and even more with rewarding discussions and the necessary criticism. I am also thankful to Professor Shaogang Gong for his valuable inputs at different times throughout my research.

I would like to acknowledge my friends who have provided me with financial and moral support whenever it has been required. In reverse alphabetical order of their names: Shahid Tanveer, Hassan Sher, Abid Rafique, Zeeshan Murtaza, Irfan Muhammad and Farrukh Khurshid.

I would also like to thank the members of staff and PhD students for the rewarding discussions and help in many ways. In no particular order: Fabrizio Smeraldi, Lourdes Agapito, Samuel Pachoud, Chen Change Loy, Bryan Prosser, Parthipan Siva, Tim Kay, Lukasz Zalewski, Matt Bernstein, Carly Wheeler, Julie Macdonald, Sue White, Rupal Vaja, Melissa Yeo and Matteo Bregenzio.

I am very grateful to the University of Engineering and Technology Taxila for supporting me during my studies at Queen Mary University of London.

Finally, I would like to once again thank Allah Almighty for the grace that I always look up to and putting in me the courage, determination and a little bit of intellect to go the distance.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Scope of the Thesis

Biometrics identify humans based on their physical or behavioural traits. Anderson in his book [1] describes biometrics as

> Biometrics identify people by measuring some aspect of individual anatomy or physiology (such as your hand geometry or fingerprint), some deeply ingrained skill, or other behavioral characteristic (such as your handwritten signature), or something that is a combination of the two (such as your voice).

Biometrics are mainly divided into two groups: physiological and behavioural biometrics. Examples of physiological biometrics include face, fingerprint, iris, hand geometry etc. whereas typing rhythm, signature, gait etc. fall under the behavioural biometrics. Biometrics have been used since long as a means of identity recognition. The most obvious and the oldest one is facial recognition dating back to the very first humans. Handwritten signatures have been used in classical China and Europe [1]. Fingerprints have been reportedly used in Japan as early as the eighth century [1]. Biometrics thus have played an important role in identity recognition over time and with the evolution of technology new dimensions are being opened up. Examples of the common biometrics used now are shown in Fig. 1.1.

According to Jain *et al*. [45] a biometric needs to satisfy the following properties

1. Universality: everybody should have the characteristic

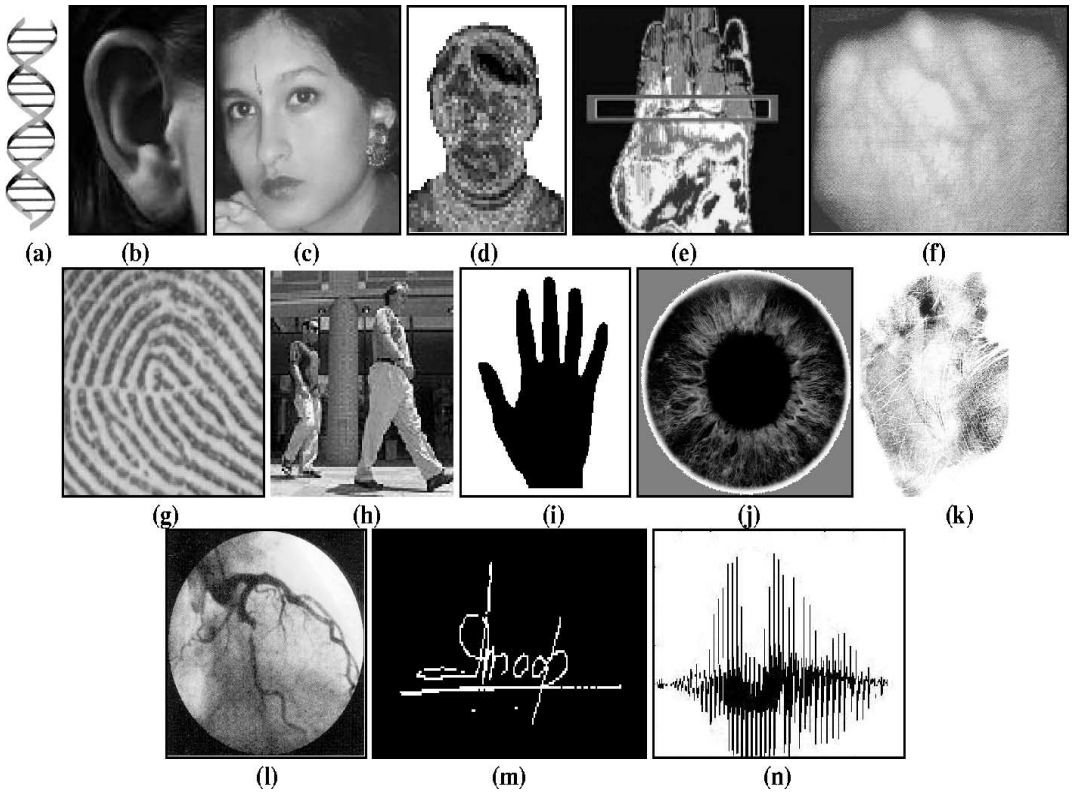


Figure 1.1: Biometrics: (a) DNA, (b) Ear, (c) Face, (d) Facial thermogram, (e) Hand thermogram, (f) Hand vein, (g) Fingerprint, (h) Gait, (i) Hand geometry, (j) Iris, (k) Palmprint, (l) Retina, (m) Signature, and (n) Voice. (from [45]).

2. Distinctiveness: individuals are well separated by the characteristic

3. Permanence: sufficient invariance over time

4. Collectability: quantitatively measurable

It turns out that on the yardstick of these properties physiological biometrics tend to be placed better than the behavioural ones. For instance, physiological biometrics such as fingerprints and iris satisfy more strongly the properties of distinctiveness and permanence than for that matter the behavioural biometric gait. This makes the physiological biometrics a suitable candidate for automated recognition systems.

With the advancement in technology and the increase in the amount of biometric data available, methods based on computer vision are researched in order to build automatic recognition systems based on biometrics. Systems based on physiological biometrics such as iris, fingerprints, face and hand geometry have already been developed and are in use in real world applications [45]. Fig. 1.2 shows biometric systems using iris and fingerprints at Heathrow airport London in December 2006.

(a) Iris recognition system (b) Fingerprint recognition system

Figure 1.2: Automated biometric systems using Iris and Fingerprint recognition (from [80])

However, systems based on physiological biometrics suffer from their inherent limitations, i.e., they require close proximity, good quality imagery and subject cooperation and therefore, can only be installed and used in well controlled locations. Automated systems based on computer vision techniques for behavioural biometrics such as gait have the potential to address the limitations of above mentioned systems. This is because gait can operate at a distance with relatively low quality imagery. In addition gait does not require subject cooperation, and for these reasons has gained significant research interest in recent years. However, gait is a weak biometric and changes for the same person under different conditions, which makes gait recognition a hard problem. Research is in study for a system that can benefit from the useful attributes of gait in being able to operate at a distance without subject cooperation while achieving high accuracy in recognition. To this end, novel work is presented in this thesis to address the limitations of contemporary methods and improve the performance of gait recognition under different conditions.

1.2 Human Identification Using Gait

Jain *et al*. [45] observe that:

> Gait is the peculiar way one walks and is a complex spatio-temporal biometric. Gait is not supposed to be very distinctive, but is sufficiently discriminatory to allow verification in some low-security applications. Gait is a behavioral biometric and may not remain invariant, especially over a long period of time, due to fluctuations in body weight, major injuries involving joints or brain, or due to inebriety.

(a) Normal

(b) Bare feet

(c) Carrying a bag

(d) Wearing a coat

(e) Different view

Figure 1.3: Frames from gait video sequences of the same subject under different conditions.

Gait is a weak biometric and is known to vary for the same person under different circumstantial and physical conditions. Common examples of these conditions are the carrying condition,

clothing condition, view, time, mood, injuries, shoes and speed among others. The conditions affecting gait herein will be called the covariate conditions as multiple conditions could simultaneously affect gait. Example frames from gait video sequences of commonly occurring conditions are shown in Fig. 1.3.

Although gait is vulnerable to covariate condition changes it has certain advantages over the stronger physiological biometrics such as face, fingerprint and iris. Specifically, all of the above mentioned physiological biometrics require close contact and good quality imagery to be effective. Gait on the other hand can operate at a distance given low quality imagery such as those obtained from Closed Circuit Television (CCTV). Consequently, gait can operate on non-cooperative subjects without interfering with their activity. These attributes have been the main driving force in gait research and make it an ideal candidate in situations where direct contact or cooperation is not feasible or desired, e.g. medium to long distance security and surveillance applications in public space as shown in Fig. 1.4.

Figure 1.4: Potential application scenario for gait recognition (from [74])

1.3 Challenges in Gait Recognition

Gait recognition as with many computer vision systems suffers from the problems caused by image noise and changing lighting conditions. Specifically, extracting features from a gait video sequence involves segmentation of the moving person from the background. Image noise and

changing lighting conditions directly affect the ability of algorithms to segment correctly the moving person from the background thus causing missing body parts and the inclusion of background (e.g. shadows) as shown in Fig. 1.5. To reduce the affect of image noise and changing lighting conditions a pre-processing stage is normally required in a gait recognition framework. Pre-processing helps remove some artifacts but inevitably some of the problems caused by image noise and lighting conditions filter through the pre-processing stage undetected or uncorrected to later stages of a gait recognition framework. It can be seen from the pre-processed silhouettes shown in Fig. 1.5 that even after pre-processing the extracted silhouettes are still imperfect which means a gait recognition system has to deal with large amounts of noise in the features.

(a) Missing body parts (b) Noise and shadows

Figure 1.5: Pre-processed silhouettes with missing body parts, noise and shadows

The problem is further compounded with occlusions which are of two types, self occlusion and occlusions from other objects. Self occlusion is caused by the cross over of the legs and the swing of the hands as the person moves. Self occlusion is unavoidable because of how people move and incurs minor loss of information in gait feature extraction. On the other hand, occlusions from other objects, caused as they appear in between the moving person and the camera, pose major problems. A significant loss in performance is expected as features disappear/reappear because of occluding objects.

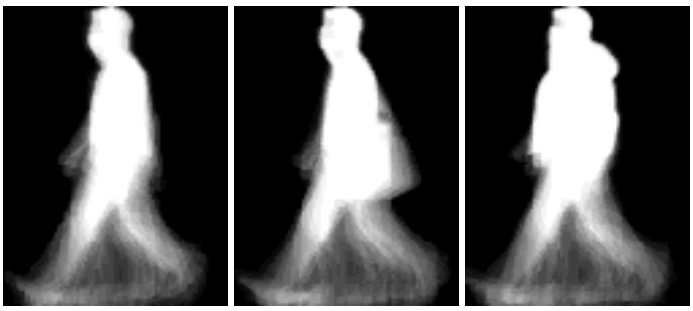
Gait Recognition under Variable and Unknown Covariate Conditions

In addition to image noise, lighting condition changes and occlusions, gait is affected by variable covariate conditions. The presence of variable covariate conditions in gait changes the available features which eventually affects the way gait is represented. Example gait sequences for the same person represented using state-of-the-art Gait Energy Images proposed by Han and Bhanu [37] are shown in Fig. 1.6 to highlight this aspect.

In gait recognition the gallery[1] set consists of people walking under normal (see Fig. 1.6(a))

[1]In object identification, a gallery set corresponds to a set of images of objects whose identities are

covariate conditions (i.e. people walking under similar and common, known conditions) and if the probe set also consists of gait sequences under conditions similar to the gallery set this brings out handsome results (see Table 3.1, Chapter 3). However, when the covariate conditions in the probe set are variable (see Fig. 1.6(b,c)) this incurs significant drop in recognition performance (see Table 3.1, Chapter 3). This is because in the presence of variable covariate conditions gait representation for the same person changes because of a change in available features and therefore, may become more similar to someone else's.



(a) Normal (b) Carrying a bag (c) Wearing a Coat

Figure 1.6: Gait Energy Images of a subject under variable covariate conditions.

The effect of covariate conditions on the performance of gait recognition is even more severe when the covariate conditions are both variable and unknown. This means they are allowed to be variable in the probe set as well as the gallery set, which more closely represents a practical scenario. In a practical gait recognition set-up operating at a distance without subject cooperation, the assumption that the gallery set consists of gait sequences captured under similar and known covariate conditions is invalid.

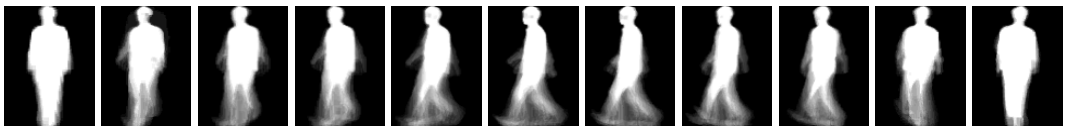
It is well known that, given cooperative subjects, gait cannot compete with physiological biometrics in terms of recognition accuracy. It is thus necessary and crucial to evaluate the performance of existing gait recognition approaches without the assumption on cooperative subjects, i.e., the gallery set is composed of a mixture of gait sequences under variable and unknown covariate conditions. To the best of the author's understanding, none of the existing work has done such an evaluation, let alone proposing methods to address it.

Cross View Gait Recognition

Amongst the covariate conditions affecting gait, view poses one of the biggest problems and is treated separately. This is because a change in the view angle causes some body parts to

known, whilst a probe set contains images of objects whose identities are unknown and need to matched against the gallery set for identification.

become visible/invisible across views. Change in view is a standard problem affecting pattern recognition. The problem in the case of gait is more severe because of the movement happening in space and time which causes the view to change even within a single gait sequence as the person moves in front of the camera. The Gait Energy Images in Fig. 1.7 for the same person walking under normal conditions captured simultaneously at different view angles highlight the change of available features across views.



(a) 0° (b) 18° (c) 36° (d) 54° (e) 72° (f) 90° (g) 108° (h) 126° (i) 144° (j) 162° (k) 180°

Figure 1.7: Gait Energy Images for different views from the multi-view CASIA dataset.

1.4 Objectives

The overall objective of this work is to research and define methods for robust gait recognition under variable and unknown covariate conditions and across view angle changes.

1. **Robust Gait Recognition Under Variable and Unknown Covariate Conditions:** Changes in covariate conditions adversely affect the performance of gait recognition methods. One of the major goals of this work is to improve the performance of existing gait recognition methods in the presence of variable covariate conditions in the probe set. In particular, this thesis aims to investigate and improve the performance of gait recognition under the challenging experimental set-up of gait recognition without subject cooperation. This set-up more closely represents the situation prevailing in practical applications and truly tests the effectiveness of gait as a biometric in unconstrained environments.

2. **Cross View Gait Recognition:** Gait is a spatio-temporal activity and spans movement over a period of time. This inevitably brings out the problem of view in gait recognition. Recognizing gait sequences across views is non-trivial and needs to be addressed in order for gait to be effective as a biometric for recognition at a distance. An objective of this thesis is to address the limitations of existing cross view gait recognition methods in order to improve recognition performance across view angle changes.

1.5 Approaches

1.5.1 Gait Flow Image Representation Extracted From Optical Flow Field Computation

A representation of gait the Gait Flow Image, based on extracting features from optical flow field computations is proposed in this thesis to perform robust gait recognition under variable covariate conditions present in the probe set. Gait Flow Image extracts more discriminative information for a gait video sequence by considering the relative motion of body parts in different directions. A histogram based feature extraction method is adopted to extract information from the flow field in separate descriptors and reduce flow field noise.

1.5.2 Adaptive Feature Selection for Gait Recognition Under Variable and Unknown Covariate Conditions

In previous work, gait recognition approaches are evaluated using a gallery set consisting of gait sequences of people under similar covariate conditions. This evaluation procedure, however, implies that the gait data are collected in a cooperative manner so that the covariate conditions are known a priori. In this thesis, gait recognition approaches are evaluated without the assumption on cooperative subjects and an adaptive feature selection method is proposed to automatically select covariate condition invariant features. The Gait Entropy Image is proposed to measure the relevance of gait features extracted from the gait representation. Constructed by computing the Shannon entropy for the silhouettes extracted from a gait sequence, a Gait Entropy Image can be readily used to distinguish dynamic (motion) and static (appearance) information contained in a gait representation with the former being selected as features that are relevant to gait recognition under variable and unknown covariate conditions. Since in a realistic experimental set-up, the covariate conditions for both the gallery and probe sets are unknown, it is proposed to select a set of features that are unique to each pair of gallery and probe sequences.

1.5.3 Cross View Gait Recognition Using Correlation Strength and Gaussian Process Classification

In this thesis a cross view gait recognition approach is presented by using Gait Flow Image in a cross view gait recognition framework. The proposed framework uses the correlation of gait sequences across views computed from a Canonical Correlation Analysis model as a similarity measure. In addition, view angle is estimated using Gaussian Process classification and inte-

grated into the cross view recognition framework in order to perform automated recognition.

1.6 Contributions

The major contributions of this thesis are as follows.

1. **The Gait Flow Image:** A representation of gait the Gait Flow Image based on optical flow field computations is proposed [6]. An important feature of this representation is the capturing of more discriminative information in separate gait descriptors. This consequently makes the Gait Flow Image robust to variable covariate conditions.

2. **Gait Recognition Without Subject Cooperation:** For the first time an experiment is conducted to evaluate the performance of gait recognition methods in a true sense of the phrase 'Without Subject Cooperation'. In this challenging experimental set-up both the gallery and probe sets consist of gait sequences under variable and unknown covariate conditions [3].

3. **Adaptive Feature Selection:** An adaptive pairwise feature selection method on the Gait Energy image is developed by using the Gait Entropy Image [5]. To integrate the adaptive feature selection process into subspace analysis an Adaptive Component and Discriminant Analysis is also proposed [5]. This provides the necessary speed up by enabling pair wise feature selection in the component domain and eliminating the need of retraining the subspace.

4. **Cross View Gait Recognition Using Correlation Strength:** A novel approach to cross view gait recognition using correlation strength is proposed [4] to address the limitations of contemporary methods. The method is robust to large changes in view angle. In addition, estimation of view angle of a gait sequence which is considered to be known a priori by existing methods is also performed using Gaussian Process classification. Gaussian Process classification along with a soft decision based algorithm is proposed [4] to incorporate the view classification into the recognition framework.

1.7 Thesis Structure

The rest of this thesis is organized as follows.

- **Chapter 2** reviews the relevant literature on gait recognition with emphasis on robust recognition under variable covariate conditions and provides a critical analysis of existing methods.

- **Chapter 3** investigates the performance of existing gait recognition approaches under variable covariate conditions. A gait representation the Gait Flow Image based on computing optical flow fields is proposed to address the problem of variable covariates in gait recognition. Experiments are carried out on two of the largest public gait datasets available (CASIA dataset [116], SOTON dataset [93]) to demonstrate the effectiveness of the Gait Flow Image as a representation robust to changes in covariate conditions.

- **Chapter 4** introduces the problem of gait recognition under variable and unknown covariate conditions and investigates the performance of state-of-the-art methods under this realistic assumption of no subject cooperation. To overcome the significant degradation in performance of existing methods an adaptive feature selection method is developed and integrated into subspace analysis to make the process computationally feasible using the novel Adaptive Component and Discriminant Analysis. The experimental set-up without subject cooperation is tested on the largest publicly available datasets (CASIA dataset [116], SOTON Small dataset [93]) and a significant improvement in performance over existing methods is demonstrated.

- **Chapter 5** builds a cross view gait recognition system robust to variable covariate conditions by using the Gait Flow Image representation proposed in this thesis in a cross view framework. Not only a method to perform cross view recognition is proposed using correlation strengths but also gait view angle estimation is done using Gaussian Process classification to identify the view angle of a probe sequence. Experiments on the largest multi-view gait dataset (CASIA [116]) reveal an improvement in performance over state-of-the-art cross view gait recognition methods which assume the probe view angle to be already known.

- **Chapter 6** concludes the thesis with a brief summary of ground covered and suggests directions for future research.

Chapter 2

Literature Review

Like many problems in computer vision gait recognition does not come as easily to machines as it does to humans. A seemingly easy problem is so hard that not a single gait recognition system has been reported to be working in challenging real world conditions. In an effort to make it feasible a wide variety of techniques have been proposed to tackle various aspects of gait recognition ranging from the segmentation and pre-processing, gait representation to the matching algorithms used for recognition.

A survey of gait as a biometric can be found in the work of Boyd and Little [19]. General surveys of gait from a computer vision perspective are to be found in Nixon and Carter [83, 84] and Liu *et al*. [66]. A more approach centred survey of gait from a model based perspective is provided by Yam and Nixon [109]. The work of Gafurov [35] goes beyond the mere boundaries of visual approaches and along with vision based methods also surveys the use of other sensors in gait literature.

In the following a review of the representative works in gait recognition is presented by following the flow of information through a gait recognition system from segmentation and pre-processing, feature representation to recognition as shown in Fig. 2.1.

2.1 Segmentation and Pre-processing

Segmentation for gait which is a spatio-temporal activity is to separate the moving person from the background both spatially and temporally. Spatial segmentation in gait literature is typically synonymous with silhouette extraction because of gait being independent of color or texture.

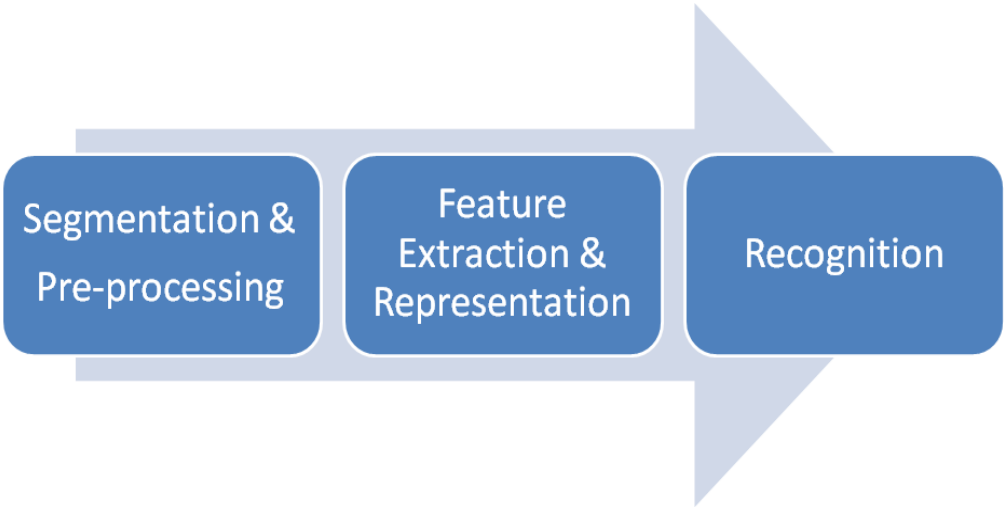


Figure 2.1: Typical data flow in a gait recognition system.

After spatial segmentation the silhouettes are pre-processed in order to remove noise and global motion. Pre-processed silhouettes are then used to temporally segment the gait sequences by estimating the gait cycle which is essential in many state-of-the-art methods [7, 27, 56, 88, 102].

2.1.1 Spatial Segmentation

Spatial segmentation is accomplished by extracting silhouettes from gait video sequences. Silhouettes are featureless black and white images and are at the heart of most gait recognition techniques. Given an input gait video sequence silhouettes are extracted from the individual frames of the gait video sequences by foreground background segmentation. The person in question is the foreground whilst everything else has to be classified as background.

The simplest method of foreground/background segmentation is the background subtraction technique used by Yu *et al*. [116] and Bodor *et al*. [11]. A background model in this case consists of an image of the background captured without any foreground objects. Frames of the gait video sequence are subtracted from the background to segment the person and generate the silhouettes.

Chroma key is another simple method of foreground/background segmentation which separates humans from a uniform color background [111] but is only applicable in controlled conditions. Chroma key has been used by Shutler *et al*. [93] to extract silhouettes from gait video sequences captured on a green background. Examples of silhouettes extracted from video sequences in the lab conditions using background subtraction and chroma key are shown in Fig. 2.2. Note that the silhouettes are clean and also no shadows are visible because of controlled lighting.

(a) CASIA Dataset [116] (b) SOTON Dataset [93]

Figure 2.2: Examples of silhouettes extracted in controlled indoor conditions. (a) Background subtraction (b) Chroma key

The above mentioned methods fail in uncontrolled outdoor conditions since a static background is no longer applicable because of changing background and lighting conditions. Statistical methods of modelling the background have to be applied under these conditions. Sarkar *et al*. [88] propose a solution by using a statistical background model for the three color channels. A model of the background over the whole gait video sequence is generated by manually putting a bounding box around the person (foreground object) in the first frame of the video sequence which is then interpolated in future frames using simple linear interpolation as the person walks with uniform speed. This divides each frame into two parts, one containing the person within the bounding box and everything less interesting outside the bounding box. Background is now statistically modelled by using the information from regions outside of the bounding box over all frames of the input video sequence. A profile of the background consists of the mean and variance of different color channels at each pixel location. Mahalanobis distance is calculated for the pixels in the bounding box containing the person and the background and based on a threshold the decision is made on whether the pixel in question is foreground or background. This generates a human silhouette which will have some spurious noisy pixels and may have some holes. Morphological techniques are used to keep the largest connected region. Example silhouette generated using this procedure from a gait video sequence captured under outdoor conditions are shown in Fig. 2.3. It can be seen from Fig. 2.3 that the generated silhouette is cluttered with noise and shadows. A possible improvement on this method which uses a single Gaussian to build a background profile is the use of a Gaussian Mixture Model (GMM) proposed by Stauffer and Grimson [94].

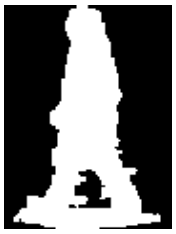


Figure 2.3: Example of an extracted silhouette using a statistical background model [88].

2.1.2 Pre-processing

In general silhouettes extracted in both indoor and outdoor conditions suffer from image noise, shadows and the problem of missing body parts. Therefore, after a silhouette is obtained steps need to be taken in order to improve silhouette quality. Heuristic methods have been developed to improve silhouette quality in the work of Bodor *et al*. [11] by using a combination of background subtraction, chromaticity analysis, and morphological operations. Boulgouris *et al*. [17] propose to address the specific problem of missing body parts in the extracted silhouettes by using temporal processing in the preprocessing stage. A temporal line which consists of all the pixels at the same position in sequence is used to correct pixel classification errors during foreground/background segmentation.

Liu and Sarkar [71] attempt to remove shadows from the extracted silhouettes using the so called eigenstance reconstruction. Eigenstances much like eigenfaces are the principal components obtained from the principal component analysis of stances generated from clean silhouettes. An input silhouette is projected into the clean eigenstance space and then reconstructed from the projections in order to produce clean silhouettes.

More recently the problem of shadow removal from silhouettes has been tackled by the use of infra-red cameras [52, 54, 79, 95]. Fig. 2.4 shows silhouettes extracted using an infra-red sensor. Shadows in this case are easily removed as they translate to lower temperature regions then the person which shows up as higher temperature regions in captured images.

Figure 2.4: Silhouette extracted using an infra red camera [79].

Removal of noise and shadows is only one part of the problem. Other problems such as scale changes and global motion exist and need to be addressed before the silhouettes can be used for feature extraction. Scale changes are caused by the change of distance between the person and the camera as the movement happens and are taken care off by normalizing the silhouettes. The global motion again caused by the movement in the camera frame is removed by centring the normalized silhouette. Sarkar *et al*. [88] propose to normalize and centre the silhouettes by normalizing the height of the box bounding the silhouette and then centring the silhouettes with top of the head in the middle.

2.1.3 Temporal Segmentation

Gait is a periodic activity and repeats at a more or less stable frequency which is determined by gait cycle length. Gait cycle is the functional unit of gait and consists of a stance phase and a swing phase as identified by Cuccurullo [28]. Fig. 2.5 shows different phases of the gait cycle from [28].

> A single sequence of functions of one limb is called a gait cycle. It is essentially the functional unit of gait. The gait cycle has two basic components, the swing phase and the stance phase.
>
> Stance: phase in which the limb is in contact with the ground
>
> Swing: phase in which the foot is in the air for limb advancement.

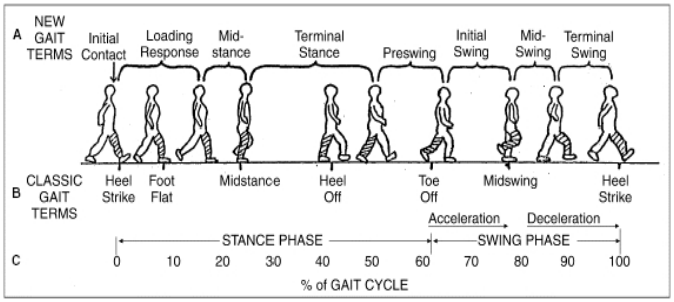


Figure 2.5: Phases of a gait cycle (from [28])

Computation of gait cycle in effect temporally segments a gait video sequence into different gait cycles and is used by almost all gait recognition methods.

A simple method of segmenting gait cycles is proposed by Sarkar *et al*. [88] by applying a maximum entropy estimation method in the lower half of the silhouettes. Boulgouris *et al*. [18]

compute gait cycle using the sum of foreground pixels. Ekinci [34] uses the projection of the silhouette on the bottom edge of the box bounding the silhouette to compute gait cycles. A few other approaches for computing gait cycles are to be found in [7, 27, 56, 102]. The main aim of all these temporal segmentation techniques is a precise estimate of gait cycle to be used at the representation stage where periodicity of gait is used to construct a representation of gait.

Segmentation both spatial and temporal along with pre-processing makes available the necessary information to construct a representation of gait.

2.2 Gait Representation

Gait representation has been the most active area of gait research. A good representation of gait should be able to discriminate i.e., well separate gait of different people, be robust to noise and changing covariate conditions, be space efficient and be easy to compute and manipulate. Representative works in gait representation are divided into different categories as listed in Table 2.1. In the following the advantages and limitations of each of these approaches are examined in the context of robust gait recognition under variable covariate conditions.

2.2.1 Static Body Parameters

Static body parameters during a gait cycle have been used to represent gait by Johnson and Bobick [47]. Each silhouette is transformed into a 4D vector with elements representing the height of the bounding box, the distance between head and pelvis, maximum value of distance between pelvis and left or right foot and the length of the stride. However, the stride length cannot be considered as a static body parameter as it changes for the same person under different runs. Keeping this in view BenAbdelkader *et al*. [7] extend the idea of Johnson and Bobick [47] and also include stride cadence which is simply the rate of steps or more precisely the number of steps in a given period of time, in the representation. Nevertheless, representations of gait based on these parameters cannot amply represent the rich spatio-temporal information content in a gait video sequence and thus lack the discriminative power required for an application like gait recognition.

2.2.2 Model Based Methods

Model based methods initialize a model of the human body whose parameters are updated over time as the person moves in the camera frame to represent the spatio temporal patterns in gait. A

Approach	Representative Works
Static Body Parameters	Static Shape Representations [7, 47]
Model Based Methods	2D Structure Models [30, 62, 75, 100] 3D Structure Models [112, 122] Analytical Models [13, 110] Active Shape Models [26, 54, 55]
Average Silhouettes	Average Silhouettes [70] Gait Energy Image [37] Motion Silhouette Image [58]
Silhouette Projection	Frieze Patterns [34, 68] Shape Variation Based Frieze Patterns [63]
Transform Based Methods	Fourier Transform [82, 103, 117] Gabor Filters [96] Wavelets [76, 86, 121] Radon Transform [15, 119]
Optical Flow Based Methods	Parameters Extracted from Optical Flow [39, 65] Flow Field Magnitude [42]
Feature Selection	GEI Based Methods [67, 113] Analysis of Features [16, 69, 85, 99] Feature Selection for Clothing and Carrying Condition [40, 61]
Cross View Representations	View Invariant Features [8, 38, 46] View Synthesis Based on 2D and 3D [11, 12, 36, 47, 50, 91, 122] View Transformations [56, 57, 79]
Generative Models	Hidden Markov Models [23–25, 51, 72]
Separate Spatio-Temporal Features	Spatio-Temporal Features [9, 21, 59]

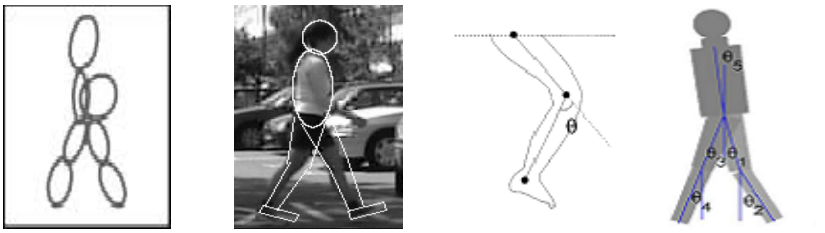
Table 2.1: Taxonomy of Gait Representation approaches

model can either be a 2D or a 3D structural model, an analytical model or a statistical model and is a mathematical construct with a set of parameters usually initialized at the full stride stance and updated over time using temporal tracking. The main advantages of model based methods are that they can handle self occlusion, rotation and scaling.

Cunado *et al*. [29] propose one of the earliest methods of model based gait recognition. Research has come a long way since then and models of human gait, both 2D and 3D, [10, 30, 62, 75, 100, 101, 120, 122] have been proposed. The focus has been more towards 2D models because of the computational complexity and the difficulty of a cooperative multi camera set-up required for 3D models.

Bobbick and Johnson [10] use a 2D structural model to recover body and stride parameters, Lee and Grimson [62] (see Fig. 2.6(a)) also use a 2D structural model and fit ellipses to regions of the human body whose parameters are then estimated from the data. Wagg and Nixon [100]

(see Fig. 2.6(b)) use a 2D model consisting of an ellipse for the torso and the head, four line segments for the legs and a rectangle for each foot. Layered Deformable Models by Lu and Venetsanopoulos [75] is another example of 2D structure model and describes the body part widths and lengths, the joint position and angles using 22 parameters. Representing the human body using a stick model has been explored in the work of Cunado *et al*. [30] (see Fig. 2.6(c)) which uses the angular motion of hip and thigh as the model parameters. Similarly Zhang *et al*. [120] (see Fig. 2.6(d)) use a five link biped model and the angles between them as the parameters of the model to represent gait. Model based approaches extracting model parameters from the whole body suffer from appearance changes of the human silhouette generally caused by clothing and carrying condition. Under these conditions the model parameters are error prone and can potentially be wrongly estimated.



(a) Lee and Grimson [62] (b) Wagg and Nixon [100] (c) Cunado *et al*. [30] (d) Zhang *et al*. [120]

Figure 2.6: 2D structural models of gait

Analytical models of human motion have also been proposed to represent gait. Yam *et al*. [110] use angular motion of thigh and knee for both walking and running with the gait signature being the magnitude of the phase weighted Fourier description of the angular motion. Bouchrika *et al*. [13] represent the joint displacements using the parametric elliptic Fourier descriptors. For computational simplicity Bouchrika *et al*. [13] incorporate heel strike data which is automatically extracted using a corner detector. Analytical models are robust to appearance changes as they generally focus on the moving body parts. However, using the analytical models of human motion for gait representation completely ignores the shape information content in a gait signal which has known to contribute positively towards recognition performance [69, 99].

Examples of 3D structure models are found in the works of Zhao *et al*. [122] and Yang and Lee [112]. Zhao *et al*. [122] propose a 3D human model computed from video sequences captured by multiple cameras and use the lengths of key segments and motion trajectories of lower limbs as features for recognition. Yang and Lee [112] estimate the human body pose by linear combination of prototype 2D silhouettes and the corresponding 3D models. 3D sturctural mod-

els are able to extract more discriminative information from a gait video sequence. Nonetheless, they are computationally intensive and require a cooperative multi-camera set-up.

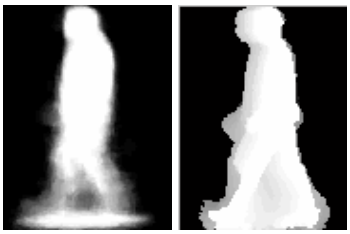
Methods based on statistical modelling examples of which are the Active Shape Models (ASM) have been used recently for gait recognition [26, 54, 55]. ASM has been successfully applied to many vision applications involving tracking and recognition of non-rigid objects. ASM are trained a priori on an object's shape and deform iteratively to fit the object in a new image. ASM are used by [26, 55] to extract relative model parameters from a gait video sequence. The reconstructed silhouette from the model parameters have been reported to be robust to noise and shadows and also gives a smooth prediction of gait cycle. But there are problems if the initial silhouettes are noisy and broken, which leads to performance degradation. Kim *et al*. [54] proposes a solution to this by using infra-red cameras to get rid of broken silhouettes and noise and report better performance.

2.2.3 Average Silhouettes

The average silhouette proposed by Liu and Sarkar [70] has been used as a choice of representation for gait by many researchers. A more popular version of average silhouettes is the Gait Energy Image (GEI) proposed by Han and Bhanu [37]. GEI and average silhouettes as the Motion Energy Image by Davis and Bobick [32] cannot indicate both the order and direction of motion. GEI represents gait over a complete cycle in a single grey scale image by averaging the preprocessed silhouettes (normalized and centred) extracted over a complete gait cycle. This simple representation of gait brings about some useful properties. 1) No silhouette alignment is required i.e. no search for key frames is required, starting from any frame within the gait cycle, number of frames equal to the gait cycle length is averaged to generate the GEI. 2) Compact representation of gait (space efficient). 3) Resistant to noise because of the averaging procedure.

GEI representation captures explicitly the shape of the subject in question and the dynamics implicitly. Pixels with high intensity values in a GEI correspond to body parts that move little during a walking cycle (e.g. head, torso), while pixels with low intensity values correspond to body parts that move constantly (e.g. lower parts of legs and arms). This explicit representation of shape makes the average silhouette (GEI) representation of gait vulnerable to appearance changes of the human silhouette caused by common conditions such as clothing and carrying. Recently feature selection methods on the GEI [67, 113] have shown some promise in dealing with covariate condition changes but still there is a significant margin of improvement.

Motion history has also been used for gait recognition. An example is the Motion Silhouette Image (MSI) [58] which embeds spatio temporal information in a grey level image. MSI represents the intensity of a pixel as a function of motion of that pixel in the temporal dimension. MSI as other average silhouette representation suffers from appearance changes of the human silhouette. Examples of GEI and MSI are shown in Fig. 2.7.



(a) GEI (b) MSI

Figure 2.7: Examples of GEI and MSI.

2.2.4 Silhouette Projections

The projection of silhouettes along rows and columns has been used to represent a gait signature by Liu *et al*. [68]. The projections are stacked over time to form 'Frieze patterns' that are mathematical entities and the mathematical theory of symmetry groups can be applied to analyse these patterns. Frieze patterns are used in gait recognition, view angle estimation, gait cycle estimation and model based body part analysis [68].

Based on the idea of Frieze Patterns another gait representation is proposed by Ekinci [34]. A silhouette extracted from a frame is represented using four 1-D distance vectors. These four distance vectors are obtained from the projections of the silhouette on the four sides of the box bounding the silhouette by calculating the number of pixels between the side of the box and the silhouette boundary. This gives a compact representation of silhouette capturing all the information in four 1-D vectors. Each of the four 1-D vectors is evolved into 2-D matrices over time to capture the complete gait video sequence. This generates 4 features for gait, one for each projection which are fused together at the score level to generate the final score. It has been reported that fusion indeed does improve the results [34].

Silhouette projections as with average silhouettes also suffer from appearance changes of the human silhouette. To address this problem a gait representation based on Shape Variation Based (SVB) Frieze Patterns is proposed by Lee *et al*. [63] which aims at mitigating the affects of appearance changes to the silhouette. Frieze patterns are generated by projecting the silhouette

images along the horizontal and vertical directions. SVB Frieze Patterns extend the idea and use key frames as references for computing self similarity. Double stance frames are chosen as the key frames and are selected by projecting on to the horizontal axis and detecting the local maxima for width. The difference from the key frames is computed to cater of appearance changes. The resulting silhouettes are then projected on to the horizontal and vertical axis to extract SVB Frieze Patterns. The results reported show that using SVB Frieze Patterns is a trade-off as they outperform other methods when there are significant appearance changes but under-perform in normal conditions as compared to state-of-the-art methods.

2.2.5 Transform Based Methods

Changing from spatial domain to transform domain a different set of features emerge. These features have also been studied for representing gait. From a frequency domain perspective Fourier descriptors have been used to represent periodic gait signals by Mowbray and Nixon [82]. Key Fourier descriptors was used by Yu *et al*. [117] and the outer contour of the extracted silhouettes is transformed to frequency domain by using simple Discrete Fourier Transform. It is seen that most of the energy in the amplitude spectrum lies in some of the Fourier descriptors; only these descriptor are used for recognition and are called Key Fourier Descriptors. Key Fourier Descriptors are also used in the study by Wang *et al*. [103] to fuse multi view gait sequences. Nevertheless, methods using the outer body contour for gait representation are prone to noise and changing covariate conditions affecting the outer body contour. Yu *et al*. [115] extend the idea and try to improve the noise robustness of the method using outer body contour to some extent by dynamic time warping but still are unable to deal with changing covariate conditions.

One of the most notable achievements in frequency domain gait representations is the work of Tao *et al*. [96]. Frequency selective filters (Gabor Filters) are used for gait representation. It is argued that they represent closely the model of the human cortex. Gabor gait representation is generated from the GEI representation of gait by convolution with Gabor filters. Tao *et al*. [96] use Gabor filters with different scales and orientations which when convolved with a 2D GEI produces a 4th order tensor representation of gait $R^{N_1 x N_2 x S x O}$, where N_1 and N_2 are the image dimensions and S, O are the number of scale and orientations respectively. Gabor gait representation extracts more discriminative information form a GEI by independently looking at different frequencies and orientations and hence performs better than contemporary methods. In addition to the frequency selective filters transform domain methods such as Discrete Cosine

Transform [87] and wavelets [76, 86, 121] have also been used for representing gait with reasonable success.

The Radon transform has been used as an effective tool in image processing and has been readily applied to represent gait [15,119]. In its simplest terms the Radon transform is the integral of a function over straight lines. When applied to binary silhouette images in x, y coordinates the radon transform does a mapping from the Cartesian (x, y) to Polar coordinates (ρ, θ) with origin being the centroid of the silhouette, and calculates a line integral by summing the pixels in the silhouette for a specific (ρ_i, θ_i). This generates a new gait representation which represents the movement of the arms and legs in a more efficient way by concentrating much of the energy in specific coefficients which vary over time due to the change of the angle of the legs and arms as the person moves. In addition, the radon transform also reduces noise because of the averaging procedure. Results of Boulgouris and Chi [15] show an improvement on standard databases. A sample silhouette sequence and their radon transform from [15] are shown in Fig. 2.8.

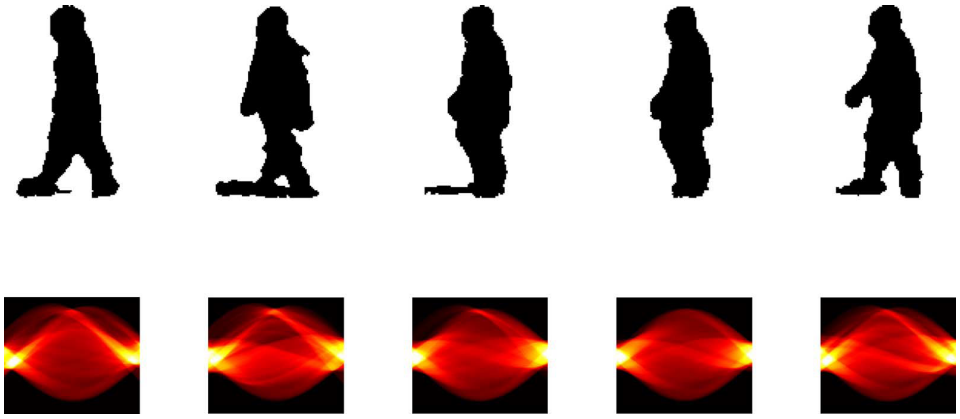


Figure 2.8: Silhouettes and their Radon Transforms (from [15]).

2.2.6 Optical Flow Based Methods

During gait, as the whole body moves in one direction multiple independent motions exist between different body parts in different directions. Apart from the model based representations of gait other approaches are unable to capture this useful and discriminative information. Optical flow based methods have the potential to extract this information of different body parts in different directions from the flow field. However, none of the existing representations of gait based on optical flow field explicitly capture this information.

Optical flow has been used by Acquah *et al.* [39] to derive a gait signature. Symmetry operator is applied on the flow field computed from successive frames and the results are averaged

over a complete gait cycle to represent gait. Before this the idea of flow field for gait features has been explored in the late 90's [42,65]. Little and Boyd [65] represent the distribution of flow field using moments and the periodic structure of the flow distribution features is then used for gait representation. Treating the optical flow fields as spatial patterns, the extracted holistic features are weak in discriminative power for an intra-class object recognition task such as gait recognition, although they have been shown to be very useful for inter-class recognitions (e.g. action recognition [33,98]). Instead of extracting holistic features, Huang *et al*. [42] proposed to use the magnitude of flow vectors directly as templates. These templates are then projected to a low dimensional canonical space for recognition. However, the useful motion information is neglected. In addition, relying on the exact value of the flow vector magnitude makes their representation sensitive to noise.

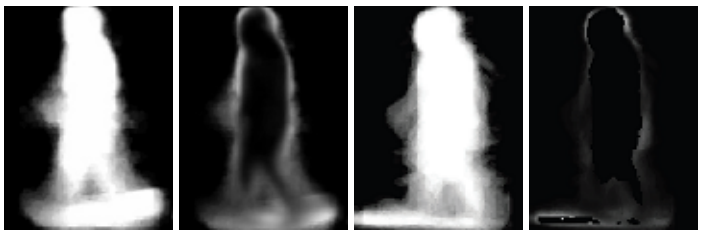
In this thesis a representation of gait based on flow field computations is proposed to explicitly capture the relative motion of different body parts in different directions. A histogram based approach is formulated to extract motion direction information from the flow field in different descriptors. The proposed representation of gait is found to outperform state-of-the-art methods under both normal and changing covariate conditions (see Table 3.1, Chapter 3).

2.2.7 Feature Selection

A gait representation could contain a lot of information. Some of it is useful, some is redundant and some of it may even be irrelevant. This is often the case in the presence of covariate condition changes. Therefore, feature selection methods have been employed in literature [31,40,61,67,69, 85,99,113] in order to select the most relevant gait features that reflect the unique characteristics of gait as a behavioural biometric.

Since gait is concerned with how people walk, selecting these (dynamic) features seems to be the most obvious direction for feature selection. Dynamic (motion) features from the Gait Energy Image (GEI) have been selected by Yang *et al*. [113] and Liu *et al*. [67]. Yang *et al*. [113] use variation analysis to identify regions from a GEI capturing dynamic information. Based on these regions a dynamic weight mask is constructed. The mask is used for feature selection and the Enhanced Gait Energy Image (EGEI) is formed. Liu *et al*. [67] also applies feature selection methods on the GEI to select motion features by applying a binary mask containing pixels with high discriminative power. Binary mask which is generated from statistical learning from the gallery set selects features from the GEI to produce the Dynamic Gait Energy Image (DGEI).

Fig. 2.9 shows the GEIs and the features selected by the feature selection methods [67, 113]. It can be seen from Fig. 2.9 that both methods extract the dynamic features. The results reported show an improvement in performance over the standard GEI for both methods which suggests that feature selection methods can be used effectively to improve recognition performance.



(a) GEI (b) EGEI [113] (c) GEI (d) DGEI [67]

Figure 2.9: Examples of EGEI [113] and DGEI [67].

Clothing and carrying condition are among the most common covariates effecting the performance of gait recognition algorithms. Hossain *et al*. [40] propose a feature selection method to improve recognition performance under clothing conditions by using a part based methodology. In the presence of clothing condition parts of the silhouette change. A weighted approach with different weights of body parts is used in order to mitigate the effects of clothing condition. The results indicate an improvement when compared to a whole body based representation. Lee *et al*. [61] also use feature selection to solve a specific carrying condition problem, the backpack. In the presence of backpack the recognition rate from a whole body representation of gait significantly drops. A backpack removal algorithm is applied by Lee *et al*. [61] which results in an eventual improvement in performance.

Feature selection methods have also been applied to study the contribution of different types of features towards recognition performance. For instance, Liu *et al*. [69] use feature selection to study the effect of different body parts separately on recognition performance. It is found that using features only from the legs matches the recognition performance of whole body features while the combination of legs and arms which constitute the dynamic features outperforms whole body. The use of only head and torso which represents static (appearance) content also has significant discriminative power which suggests that static cues cannot be completely discarded. In general this depends on the covariates involved, e.g. under carrying and clothing condition which mostly effect the torso region the recognition rate drops for these regions when compared to regions containing dynamic information such as hands and legs which have minimal effect because of clothing and carrying conditions. A similar study by Boulgouris and Chi [16] discov-

ered that different body components have different contribution towards recognition performance and calculated optimal weights for getting the best recognition performance. Furthermore, feature selection methods using statistical tools such as the analysis of variance have been applied by Veres *et al*. [99] in order to measure the contribution of specific types of features towards recognition performance with the result that both the appearance and motion features contribute towards recognition performance. Nixon and Guo [85] also use statistical methods to perform feature selection by using mutual information to remove irrelevant and redundant information from a gait representation and identify the underlying features that increase the discriminative power of gait recognition.

However, none of the feature selection methods are designed for gait recognition without subject cooperation. A more practical set-up involves no subject cooperation and hence both the gallery and probe sequences could consist of gait sequences captured under variable and unknown covariate conditions. In this thesis an experimental set-up assuming no subject cooperation is proposed and a Gait Entropy Image (GEnI) based adaptive feature selection method is developed to select covariate condition independent features. The features selection method is adaptive and selects features using information from both the gallery and probe sequence at hand to cater for changes in covariate conditions in both the gallery and the probe set.

2.2.8 Cross View Gait Representation

Most gait representations are designed with the assumption that the view angle is constant or varies slightly and for that reason are unable to cope with large variations in view angle between the gallery and probe gait sequences. Cross view recognition methods thus need to be developed to cater for large changes in view angle.

Early works on cross view gait recognition falls into two categories: 1) extracting view invariant features, and 2) view synthesis based on 2D and 3D geometry. Approaches in the first category aim to extract gait features that are invariant to view change. Self Similarity Plots (SSP) is one such feature that has been examined for both action recognition [49] (inter-class) and gait recognition [8] (intra-class). SSP uses the similarity between pairs of frames in a gait sequence to represent gait. It is claimed that the self similarity representation of gait encodes a projection of gait dynamics and is resistant to noise. This representation of gait is also resistant to appearance changes because of its inherent construction. A self similarity plot encodes a lot of useful information about gait and can be used to efficiently measure frequency (cycle length) and the

phase of gait. Nevertheless, a drawback of this representation is that the similarity plots need to be normalized and aligned which incurs extra overhead and inaccuracies. In addition, the self similarity representation loses shape information altogether which is deemed useful for gait recognition [69, 99].

Alternatively, a statistical method for extracting view invariant features from Gait Energy Images (GEI) is proposed by Han *et al*. [38] by which only parts of gait sequences that overlap between views are selected for constructing a representation for gait matching across views. The approach cannot cope with large view angle changes under which gait sequences of different views can have little or no overlap. Extracting normalized trajectories of body parts is another view invariant feature based approach proposed by Jean *et al*. [46]. However, tracking of body parts is unreliable due to self-occlusion. In addition, the problem of body parts becoming invisible given large view angle change remains.

Approaches in the second category either use a single camera and assume the subjects to be far away from the camera and perform a view synthesis for an arbitrary view using planar imaging geometry [12, 36, 47, 50], or use cooperative multi-camera set-up to extract 3D structure information via camera calibration [11, 91, 122]. Methods based on the planar view assumption have a disadvantage in that these methods cannot cope with large variations in view angle. On the other hand techniques based on 3D reconstruction are only suitable for a fully controlled and cooperative multi-camera environment such as a biometric tunnel [89].

Recently a number of approaches [56, 57, 79] based on view transformation have been presented which have the potential to cope with large view angle changes and do not rely on camera calibration. These approaches aim to learn a mapping relationship between gait features of the same subject observed across views. Frequency domain feature are used by Makihara *et al*. [79], optimized GEI is used as a feature by Kusakunniran *et al*. [56] and a Region of Interest (ROI) based method is used by Kusakunniran *et al*. [57] to select features from the GEI for transformation across views. When matching gait sequences from different views, the gait features are mapped/reconstructed into the same view before a distance measure is computed for matching this is done using Singular Value Decomposition in [56, 79] and Support Vector Regression in [57]. An advantage of these methods is that they have better ability to cope with large view angle change compared to earlier works. However, a view transformation based method also has a number of drawbacks. 1) It suffers from degeneracies and singularities caused by features visi-

ble in one view but not in the other when the view angle difference is large. 2) The reconstruction process propagates the noise present in the gait features in one view to another thus decreasing recognition performance.

In this thesis a novel approach to cross view gait recognition is proposed by addressing the limitations of existing methods based on view transformation [56,57,79]. Specifically, the correlation of gait sequences from different views is modelled using Canonical Correlation Analysis (CCA). A CCA model projects gait sequences from two views into two different subspaces such that they are maximally correlated. Similar to the existing view transformation methods, the CCA model also captures the mapping relationship between gait features of different views, albeit implicitly. However, rather than reconstructing gait features in the same view and matching them using a distance measure, correlation strengths are used directly to match two gait sequences. This brings out two key advantages: 1) By projecting the gait features into the two subspaces with maximal correlation, features that become invisible across views are automatically identified and removed. 2) Without reconstruction in the original gait feature space, this approach is more robust against feature noise.

2.2.9 Generative Models

Generative models such as Hidden Markov Models (HMM) have also been used to represent the spatio-temporal patterns in gait. Examples of HMM used in gait can be found in [23–25, 51, 72]. HMM is a suitable candidate for gait because of a Markovian dependence in gait from one stance to another. The hidden states in this process are the transitions between stances and the observations are the stances themselves.

Liu and Sarkar [72] use HMM to normalize the gait dynamics by a population HMM (pHMM) generated from the training dataset. The states of pHMM are the gait stances over one complete gait cycle and the observations are the silhouettes extracted from the gait video sequence. The method selects silhouettes that belong to key stances to represent gait while discarding frames that belong to stances that have least discriminative power. This improves the overall discriminative power of the representation. Liu and Sarkar [72] report that silhouettes near the mid-stance have the least discriminative power while silhouettes near the full stride stance have maximum discriminative power.

Kale *et al*. [51] use two gait features; the width of the outer contour of the silhouette and the entire silhouette, and train HMMs for each person whose parameters are used to represent

gait. A probe gait sequence is then assigned to gallery sequence whose HMM has the highest likelihood of producing the observations in the probe sequence. Chen *et al*. [23] use a multi layered Factorial HMM to combine several features (Frieze, Wavelets) without collapsing them into single feature. It has been reported that FHMM approach is particularly useful if the features at different layers are unrelated.

Techniques based on HMM have been reported to be robust to noise and changing speed. However, HMM based methods show a degradation in performance in the presence of covariate conditions causing appearance changes and a substantial change in view angle [24].

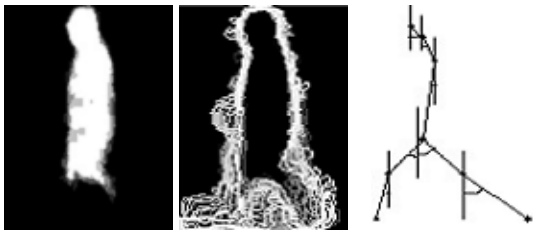
2.2.10 Separating Appearance and Motion Features

Separating the static (appearance) and dynamic (motion) features helps extract more discriminative information from a gait video sequence. In addition, this also enables individual analysis of the two types of features present in gait. One of the earliest methods using the separate static and dynamic features is proposed by Bhanu and Han [9]. Dynamic features are extracted from a 3D model fitted to binary 2D silhouettes while the static features are extracted from the four key frames of a walking cycle. Different strategies of combining the static and dynamic features are explored and an improvement in performance has been reported.

A few methods have been proposed recently to separate the appearance and motion information in gait representation. For instance, Lam *et al*. [59] formulate the Motion Silhouette Contour Templates (MSCT) and Static Silhouette Templates (SST) for representing the dynamic and static information respectively. MSCT are generated from the outer contour of the human silhouette while SST are generated from the static regions of the human silhouette. Fig. 2.10(a) gives an example of the SST generated while Fig. 2.10(b) shows the MSCT. Representing gait in SST and MSCT separates the appearance and motion information content of the gait signal but the appearance features represented using SST lack discriminative power while the motion content captured by MSCT is not descriptive enough and could be further enhanced by using motion direction information.

Chai *et al*. [21] also propose to separate the static and dynamic features into different descriptors. Static features are extracted from the silhouettes by looking at the number of foreground pixels along with the height to width ratio while the dynamics of gait is extracted from a 2D stick figure model based on anatomical studies. Joint positions in the skeleton generated from the binary silhouette are used for recording the joint angles and their coordinates. Fig. 2.10(c)

shows an example of the dynamic features.



(a) SST (b) MSCT (c) Kinematic Features

Figure 2.10: Examples of SST and MSCT from [59] and the dynamic gait features from [21].

Separating the spatio temporal features enables the individual analysis of these features. Tsuji *et al*. [97] analyses the static and dynamic feature separately to learn a transformation of silhouettes based on walking speed. As gait speed changes the static features do not exhibit significant changes but the dynamic features do. Keeping this in view a factorization based speed transformation model is proposed only for the dynamic features while the static features remain as they are.

To harness the potential of information present in a gait video sequence, in this thesis a representation of gait based on optical flow computations is proposed. In addition to extracting more discriminative information of relative body parts in different directions the proposed representation of gait also separates the motion intensity and motion directions into different descriptors. The motion intensity descriptor mainly represents the static information while the motion direction descriptors represent dynamic information.

2.3 Gait Recognition

The next step after the computation of gait representation in a gait recognition framework is to perform recognition. Given gait sequences represented as feature vectors gait recognition can be treated as a pattern classification problem. However, the feature vectors for gait representations usually have high dimensionality and the number of examples in each class are very few limiting the classifiers that can be used to perform recognition. In particular, the Support Vector Machines (SVM) classifier which tries to minimize the expectation of test error on the trained data, cannot be adequately trained given the smaller number of training examples for each class. Consequently, apart from a few exceptions [78, 118] SVM has not been used as a choice of classifier for gait recognition.

Alternatively, the Nearest Neighbour (NN) and the k Nearest Neighbour (kNN) have been

extensively used for classification in gait literature [8,12,14,22,37,88,95,102,116]. NN classifier matches a feature vector to the closest one in the gallery based on a selected distance measure. kNN is an extension of the NN classifier and instead of matching to the closest one, classifies based on a majority vote of the k closest feature vectors in the gallery. The feature vectors (gait representations in this case) used by the classifiers can be used directly or represented in subspace using learning methods. This is described next from a gait recognition perspective.

2.3.1 Direct Template Matching

This is the simplest method of matching probe features to gallery. This method involves no dimensionality reduction on the features and computes the closest match in the gallery to a probe sequence based on a distance measure (Euclidean distance is the most commonly used). Direct matching is rarely used in practice because of the problems of noise, speed and the availability of better approaches described next. Direct matching has been mainly used as a baseline [88,95, 116].

2.3.2 Matching in Subspace

The use of learning methods before using a classifier in order to better represent given features has been extensively studied in gait literature [8,15,37,43,67,72,96,99,102,107,108]. Learning methods are divided into unsupervised and supervised methods and are described below in the context of their applications to gait recognition.

Unsupervised Methods

Unsupervised subspace learning methods are employed to reduce the dimensionality of the feature space (if high dimensional) and learn a useful set of features. Matching can be performed in the new space using any available classifier. For gait recognition though, mostly the NN classifier is used.

Principal Component Analysis (PCA) [48] is a method to learn an orthogonal subspace from the training data so that the given data is best represented in a least square sense in a reduced dimensional space. PCA is a linear transformation and is obtained by performing an eigenvalue decomposition of the data covariance matrix and has been used by many gait recognition approaches [8, 37, 43, 72, 99, 102] to reduce the dimensionality of the data and perform implicit feature selection.

Independent Component Analysis (ICA) is another unsupervised method which has also been applied to gait recognition. ICA is a technique for revealing the underlying set of random variables that represent the input data based on the assumption that the random variables are as statistically independent of each other as possible. ICA is used to find the underlying factors or sources which generated samples of the data. State-of-the-art methods using ICA for gait recognition include [76, 77].

In both of the methods mentioned above the input data (gait representations) are converted to a long vector in a very high dimensional space before a subspace can be learned. The number of training samples available is very small as compared to the dimensionality of the vector space, this gives rise to Under Sampling Problem (USP). In addition to USP another drawback of these approaches is that the structure of the original input data is lost. Xu *et al*. [108] propose a matrix based unsupervised algorithm called the Coupled Subspace Analysis (CSA) to remove noise and retain the most representative information. Promising results for gait recognition have been reported using CSA in [108].

Supervised Methods

Linear Discriminant Analysis (LDA) or simply discriminant analysis is a supervised subspace learning method that best separates the input data in least square sense. LDA classifies a set of observations into a set of predefined classes. Together PCA and LDA have been used for gait recognition in order to produce the best data representation and class separability [37, 43, 72]. LDA has also been used independently to perform gait recognition [15, 67]. Promising results on some of the largest available gait datasets indicate the usefulness of LDA as a learning method.

LDA like PCA and ICA operates on vectors and thus looses the underlying structure in gait representations which are usually two dimensional. To this end matrix based techniques for discriminant analysis have been proposed for gait recognition [96, 107]. Structure preserving supervised learning methods have also been used in face recognition by Ye *et al*. [114] and have been reported to perform better than vector based methods. These methods have the following advantages over vectors based methods: 1) Reduction of USP and 2) Preservation of structural information. Promising results on the largest gait datasets have been reported using learning methods preserving the underlying structure of gait representations [96, 107].

2.3.3 Multi-view and Multi-modal Fusion for Gait Recognition

With the availability of multi-view gait data a new research dimension has come to existence in using fusion of multi-view gait sequences to improve recognition performance of gait. Wang *et al*. [103] studied the gain in performance on fusion of gait from multiple views. Their results suggest that fusion achieves consistently better results and the performance gain is greater, when the gait sequences that are more than 90° apart are fused, compared to small angle differences. Similar results are obtained when gait sequences from different views are fused for recognition by Huang *et al*. [44]. Weights are calculated for the fusion of scores from different views and it is reported that the optimal combination gives most weight to the front view and the front to parallel view. This is quite understandable as both these views export different set of features and should be given higher weight towards final score calculation. Feature fusion from multiple views has also been reported to improve performance when compared to a single view by Ming *et al*. [81].

The use of multiple views for improving the performance of gait recognition algorithms is appealing and is useful when multi-view gait data is available. This is because different views expose different features and thus increase the discriminative power. Optimizing the weights for different views for fusion leads to improved recognition performance of gait.

Gait has been also used in conjunction with other biometrics to increase the performance of recognition systems. Gait has been combined with face [73, 90, 123] to build multi modal recognition systems. For multi modal fusion the independent scores from different modalities are transformed into a common domain. The scores are then fused together using a fusion mechanism. The fusion schemes used by [73,90,123] include sum, Bayesian decision rule, confidence weighted scores, rank sum, product rule, max, min and majority rule.

2.4 Summary

The preceding review has covered essential techniques and works in the literature regarding segmentation, gait recognition and in particular gait representation. Most of the literature in gait revolves around the representation aspect but methods for segmentation and pre-processing and recognition also hold an important place.

Towards developing a robust cross view gait recognition system working under variable covariate conditions the work presented in this thesis addresses the following limitations of existing approaches:

1. Existing silhouette-based representations of gait are unable to capture the relative motion of body parts in different directions and therefore, lose important discriminative information. A representation of gait the Gait Flow Image is developed in this thesis to capture relative motion of body parts in different directions. Gait Flow Image is computed from optical flow field and acquires more discriminative information from a gait sequence in separate motion intensity and motion direction descriptors.

2. To the best of knowledge, all work in gait recognition up to date, uses a gallery set captured under similar and known covariate conditions and thus assumes subject cooperation. An experimental set-up consisting of both the gallery and probe sets consisting of sequences under variable and unknown covariate conditions is proposed in this thesis to evaluate the performance of gait recognition without subject cooperation. In addition to investigating the performance of existing methods an adaptive feature selection method based on the Gait Entropy Image (GEnI) is proposed in this thesis to select a different set of features depending on the probe and gallery sequences at hand. This is different from existing feature selection methods in gait recognition which select features for the whole set and are not suitable if the covariate conditions are unknown.

3. Very few methods in gait literature deal with gait recognition across large changes in view angle. The few exceptions to this are the view transformation based approaches, but these methods suffer from the problems of feature mis-match across view and noise propagation through the transformation process. In this thesis a mapping of features across views is established using Canonical Correlation Analysis in order to address the limitations of existing view transformation based methods. Instead of transformation to a different view the correlation of gait features across views is used to measure the similarity score. Furthermore, Gaussian Process classification is proposed to identify the view angle of gait sequences to perform automated cross view gait recognition. This is different from existing approaches where the view angle is assumed to be known a priori.

Chapter 3

Flow Field Representation of Gait

Gait is concerned with how people walk and can be characterised by the relative motions between different body parts during walking. However, most recently proposed gait representation approaches [8,37,63,68] focus solely on the motion intensity information. This is mainly because that, although the relative motion direction is useful for representing gait, it is also more difficult to capture and less reliable. In particular, during a walking cycle, whilst the whole body is moving towards one direction, multiple independent motions also co-exist across different body parts. Furthermore, these independent motions constantly change in both direction and magnitude. One possible solution to capturing relative motion information is to adopt a model based approach [53, 64, 75, 122] which models the human body configuration (e.g. 2D/3D skeletons) and the model parameters estimated over time encode the detailed relative motion information. Nonetheless, model-based methods are computationally intensive and require good quality imagery which leads to inferior performance on public gait datasets [37, 59, 63].

Based on the above observations in this chapter a novel gait representation the Gait Flow Image (GFI) is proposed which captures both the intensity and direction information about the relative motions between body parts. More specifically, optical flow fields are computed for centred and normalised human figures extracted from a complete gait cycle. In order to extract robust gait features from the noisy optical flow field, exact values of the magnitude and direction of the flow vectors are not used. Instead, the flow direction is discretised and a histogram based direction representation is formulated, which gives rise to a set of spatio-temporal motion descriptors. In particular, the gait representation consists of a Motion Intensity Image (MII),

which measures the intensity of relative motion at each pixel location, and four Motion Direction Images (MDIs), each of which represents the likelihood of the direction of motion being along one specific motion direction during a complete gait cycle. These motion images are then fused together during recognition.

Extensive experiments are performed using the CASIA [116] and SOTON dataset [93] which contain both indoor and outdoor gait sequences to validate the effectiveness of GFI against state-of-the-art. The results indicate that flow field based gait representation outperforms existing alternatives, especially when there are changes in covariate conditions.

3.1 Flow Field Gait Descriptors

Given a gait sequence, a figure-centric optical flow field is computed for each frame. This is followed by gait cycle segmentation, and computation of gait motion descriptors for each cycle.

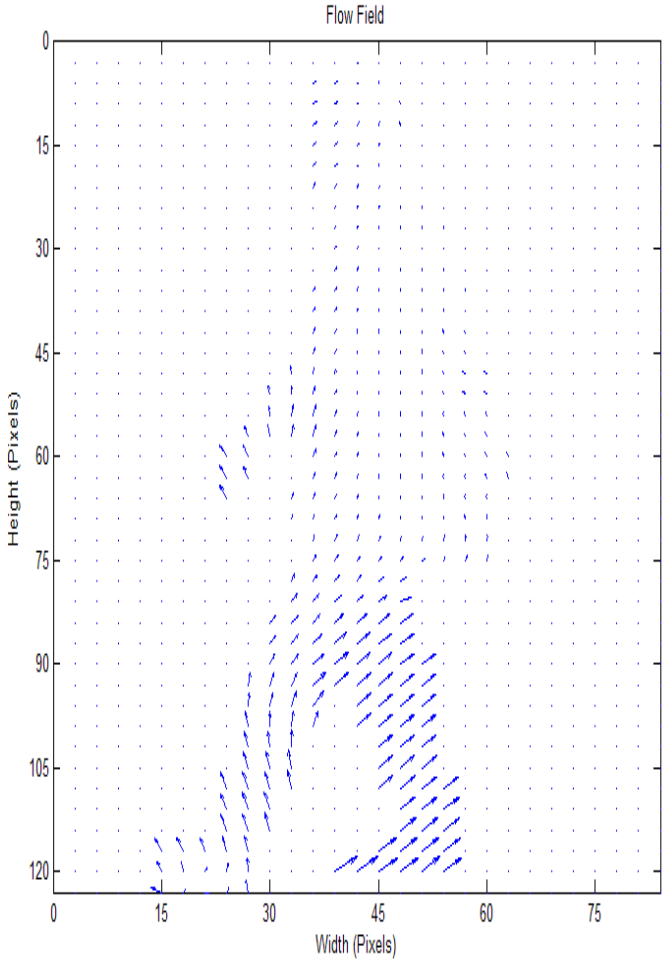
3.1.1 Computing Figure-centric Optical Flow Fields

A figure-centric spatio-temporal volume for each walking person is first extracted. To this end the walking person needs to be segmented from each image frame. This is achieved by background subtraction. Connected component algorithm is then applied to the segmented foreground regions and a silhouette image of the person is obtained. The silhouettes extracted from the sequence are used for two purposes: 1) They are used as the foreground masks to extract the original images of the person, which are then centred and normalised to a standard height to generate the figure-centric spatio-temporal volume. 2) The silhouettes are also used to extract gait cycles via the maximum entropy estimation [88] in the lower half of the silhouettes. An example of normalized silhouette and the corresponding figure-centric original image of the walking person from a carrying-bag sequence in the CASIA dataset [116] are shown in Fig. 3.1(a) and (b) respectively.

It can be seen from the plot of flow field in Fig. 3.1(d) that the magnitude of optical flow has minimal values at the torso and the regions containing the bag. This is because there is no pronounced motion between two successive frames in these areas. In contrast the magnitude of optical flow is indicated by larger arrows in the regions containing the legs with arrows pointing in the direction of motion. This enables separating motion intensity and direction into different descriptors at each location in the image.

Given a complete gait cycle consisting of T figure centric images $\{\mathbf{I}_1,..,\mathbf{I}_t,..,\mathbf{I}_T\}$, optical

(a) Silhouette (b) $\mathbf{I}_{t-1}$ (c) $\mathbf{I}_t$



(d) $\mathbf{F}_t$

Figure 3.1: Computing figure-centric optical flow field. (a) shows the silhouette at frame $t-1$; (b)&(c) are the extracted figure-centric images of the walking person at $t-1$ and t respectively; (d) is the computed optical flow field.

flow fields are estimated for each frame. The flow field for the tth frame is denoted as $\mathbf{F}_t$ and computed using two consecutive frames $\mathbf{I}_{t-1}$ and $\mathbf{I}_t$. Note that no alignment for the start of the gait cycle or key frames are required as all the frames in the gait cycle are used to construct the motion descriptors i.e. starting from any frame in the gait cycle number of frames equal to the

gait cycle length are used to generate the Gait Flow Image representation. A recently proposed optical flow estimation algorithm by Brox *et al.* [20] is used. In order to reduce the affect of noise, the flow fields are smoothed by applying a 3x3 Gaussian filter. An example of the figure centric flow field is shown in Fig. 3.1(d).

3.1.2 Computing Motion Descriptors

Optical flow fields are inevitably noisy and error-prone due to image noise and self-occlusions. For intra-class object recognition, extracting robust motion descriptors from noisy flow field input is crucial. To this end, instead of using the exact values of the flow vectors as in previous methods [33, 42, 98], a very coarse yet informative motion intensity and direction information is extracted by taking the following steps:

1. The flow field $\mathbf{F}$ (the subscript t is omitted here for notational simplicity) is decomposed into two scalar fields $\mathbf{F}_x$ and $\mathbf{F}_y$ corresponding to the horizontal and vertical components of the flow. The two scalar fields are further decomposed into four non-negative fields denoted as $\mathbf{F}_x^+$, $\mathbf{F}_x^-$, $\mathbf{F}_y^+$, and $\mathbf{F}_y^-$. In particular, at pixel location (i, j)

$$F_x^+(i,j) = \begin{cases} F_x(i,j), & \text{if } F_x(i,j) > 0 \\ 0, & \text{Otherwise} \end{cases} \tag{3.1}$$

$$F_x^-(i,j) = \begin{cases} 0, & \text{if } F_x(i,j) > 0 \\ -F_x(i,j), & \text{Otherwise} \end{cases} \tag{3.2}$$

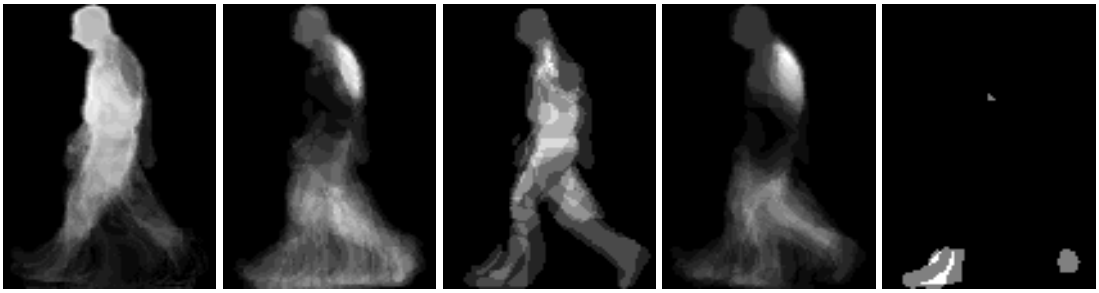
   Similarly $\mathbf{F}_y^+$, and $\mathbf{F}_y^-$ are obtained from $\mathbf{F}_y$. Note that the direction of flow vector will become extremely unreliable when the flow magnitude is small. Therefore, if the magnitude $\|\mathbf{F}(i,j)\|$ is less than 1 pixel, all four non-negative fields at that pixel location are set to zero.

2. The four non-negative flow fields are discretized into binary images and denoted as $\hat{\mathbf{F}}_x^+$, $\hat{\mathbf{F}}_x^-$, $\hat{\mathbf{F}}_y^+$, and $\hat{\mathbf{F}}_y^-$. This is because the exact values of these non-negative fields are still too noisy to be trustworthy.

3. Given the 4 binary images obtained from the original flow field for each frame of the complete gait cycle of T frames, how the motion intensity and direction are distributed during the gait cycle is computed. To achieve this, a 5-bin histogram is computed at each

pixel location $\mathbf{B}(i,j) = \{B_I(i,j), B_x^+(i,j), B_x^-(i,j), B_y^+(i,j), B_y^-(i,j)\}$. More specifically, at each frame t, if all of the four flow fields ($\hat{\mathbf{F}}_x^+$, $\hat{\mathbf{F}}_x^-$, $\hat{\mathbf{F}}_y^+$, $\hat{\mathbf{F}}_y^-$) are zero, $B_I(i,j)$ is incremented by one. $B_I(i,j)$ therefore counts the total number of frames in the gait cycle where there is no relative motion at pixel (i,j). The count in the bin $B_x^+(i,j)$ will be incremented if the corresponding binary flow images has a non-zero value. Therefore $B_x^+(i,j) = \sum_1^T \hat{\mathbf{F}}_x^+$. Similarly the values of the other three histogram bins are obtained.

4. The 5 histogram bins at each pixel location are normalised by the length of the gait cycle T, which give rise to 5 motion descriptors used in this approach for gait representation. Specifically, from $B_I(i,j)$ the Motion Intensity Image (MII) is obtained and is denoted as $\mathbf{M}$. MII measures the motion intensity distribution over a complete gait cycle. The other four bins contribute to four Motion Direction Images (MDIs) which measure the distribution of the motion direction along four directions: right, left, up, and down, and are denoted as $\mathbf{M}_x^+$, $\mathbf{M}_x^-$,$\mathbf{M}_y^+$, and $\mathbf{M}_y^-$ respectively. The size of the 5 descriptors is identical to that of the figure-centric person image.



(a) $\mathbf{M}$ (b)$\mathbf{M}_x^+$ (c) $\mathbf{M}_x^-$ (d) $\mathbf{M}_y^+$ (e) $\mathbf{M}_y^-$

Figure 3.2: Example of the 5 motion descriptors.

An example of the 5 motion descriptors illustrated as grey-level images is shown in Fig. 3.2. The Motion Intensity Image $\mathbf{M}$ is a time-normalised accumulative motion intensity image with lower value meaning that relative motion occurs more frequently during the gait cycle. As can be seen in Fig. 3.2, low values are observed at the legs and arms area, whereas the static areas such as head and torso are represented as high values. Note that the MII is similar in spirit to the widely adopted Gait Energy Image (GEI) [37], which aims to capture the same information but termed differently as motion energy. However, since a GEI is computed using the binary silhouette as opposed to flow fields, the measurement of motion intensity is indirect and less accurate. The motion intensity image though differs from the GEI in that the intensity of motion

captured by MII is mainly for the static regions during the gait cycle with the dynamic areas appearing with minute values, this is made possible by the novel use of information extracted from the flow field in the construction of MII as opposed to binary silhouettes in GEI. In other words the MII represents the overall shape of the human during a gait cycle. The four Motion Direction Images on the other hand concentrate on the dynamic aspect of information available. This is because the MDIs are only formed once it has been established that motion exists by looking at the magnitude of flow field. MDIs give more detailed information as for how likely relative motions of different directions can take place at each pixel location during a gait cycle. It can be seen from Fig. 3.2(b)-(e) that, as expected, motions of different directions have different distributions at different locations. It is also noted that compared with the other three MDIs, all the high intensity areas in $\mathbf{M}_y^-$ are concentrated around the two feet. This is not surprising as downwards motion is rare for people walking on a flat surface. Nevertheless, this suggests that the discriminative power of $\mathbf{M}_y^-$ is much weaker than that of the other three MDIs. Including $\mathbf{M}_y^-$ in the final representation has adverse affects on the performance of Gait Flow Image representation and consequently, in this chapter only three MDIs are used for gait representation together with the MII. Note that this analysis is only made possible by separating the dynamic content into its constituent motion directions where only the best performing could be chosen to construct a final representation.

Fig. 3.3 shows how different motion descriptors are affected by different covariate condition changes. In the analysis below normal condition shown in the first row of Fig. 3.3 is used as a reference. It can be seen clearly that the appearance related covariate conditions such as carrying and clothing have a visible effect on the MII (see second and third row, Fig. 3.3(a)). This is because that the shape of the low motion-intensity areas in the MII are affected by the appearance changes caused by variations in these conditions. On the other hand covariates affecting gait itself such as shoes and speed have minimal effect on the MII as seen from fourth and fifth row Fig. 3.3(a). In contrast, second and third rows Fig. 3.3(b)-(d) indicate that the effect on the MDIs is minimal for the clothing and carrying conditions. Minimal affect on the MDIs for covariate conditions such as clothing and carrying indicates that they are more robust against shape appearance changes. On the other hand, the MDIs would be more sensitive to covariate condition changes such as shoes and speed that affect gait itself as is shown in the last two rows of Fig. 3.3(b)-(d). Overall, these two types of motion descriptors capture complementary

(a) $\mathbf{M}$ (b)$\mathbf{M}_x^+$ (c) $\mathbf{M}_x^-$ (d) $\mathbf{M}_y^+$

Figure 3.3: Effect of different covariate conditions on the motion descriptors. First row: normal condition; Second row: the same person carrying a bag; Third row: the same person wearing a bulky coat; Fourth row: the same person in different shoes; Last row: the same person walking fast.

information about gait and have different levels of sensitivity towards different types of covariate conditions. They are therefore fused together for recognition as described next.

3.2 Gait Recognition

The four motion descriptors ($\mathbf{M}$, $\mathbf{M}_x^+$, $\mathbf{M}_x^-$, $\mathbf{M}_y^+$) are used independently for computing the dis-similarity between the gait sequences of a probe subject and a gallery one. The dis-similarity scores are then fused for matching the two subjects. Instead of using the descriptors directly as templates, they are projected to a subspace for dimensionality reduction. This is achieved using Component and Discriminant Analysis (CDA) based on Principal Component Analysis and Multiple Discriminant Analysis to simultaneously achieve a good data representation and class separability [43]. The Euclidean distance of the descriptors in the CDA subspace is used to measure the dis-similarity between two subjects.

Once the dis-similarity scores have been computed for each of the four motion descriptors, they are fused together as follows. First, the dis-similarity scores need to be normalised as they fall into different value ranges. In particular, dis-similarity score for each descriptor is normalised by the averaged dis-similarity score between two different subjects in the training dataset. The averaged dis-similarity score for the MII descriptor is given as

$$D_{\mathbf{M}}^{Avg} = \frac{2\sum_{i=1}^{c}\sum_{j=1,j\neq i}^{c} D(\mathbf{M}_i,\mathbf{M}_j)}{c(c-1)} \tag{3.3}$$

where c is the total number of classes in the training set and $D(\mathbf{M}_i,\mathbf{M}_j)$ is the dis-similarity score between the class i and j. Since each class can have more than one examples present in the training set $\mathbf{M}_i$ is the class center of the i^{th} class computed by averaging the feature vectors for the i^{th} class. Same is true for M_j. Eqn. 3.3 is used to compute the average distance for the remaining three descriptors $\mathbf{M}_x^+$, $\mathbf{M}_x^-$ and $\mathbf{M}_y^+$. The dis-similarity scores are then normalised as follows

$$D_{\mathbf{M}} = \frac{D(\mathbf{M}_P,\mathbf{M}_X)}{D_{\mathbf{M}}^{Avg}} \tag{3.4}$$

where $\mathbf{M}_P$ is the class center of the probe MII and $\mathbf{M}_X$ is the class center of the MII from the training set which has the least dis-similarity score $D(\mathbf{M}_P,\mathbf{M}_X)$. Eqn. 3.4 is then used to normalise the scores for the other descriptors $\mathbf{M}_x^+$, $\mathbf{M}_x^-$ and $\mathbf{M}_y^+$. After normalisation, the final similarity

score used for template matching is obtained as:

$$D = \lambda D_{\mathbf{M}} + (1-\lambda)(D_{\mathbf{M}_x^+} + D_{\mathbf{M}_x^-} + D_{\mathbf{M}_y^+}) \tag{3.5}$$

where $D_{\mathbf{M}}, D_{\mathbf{M}_x^+}, D_{\mathbf{M}_x^-}, D_{\mathbf{M}_y^+}$ are the normalized dis-similarity score obtained using the gait descriptors, and λ is the fusion weight. λ dictates how much of the static (MII) and how much of the dynamic (MDIs) cues are to be used for final recognition.

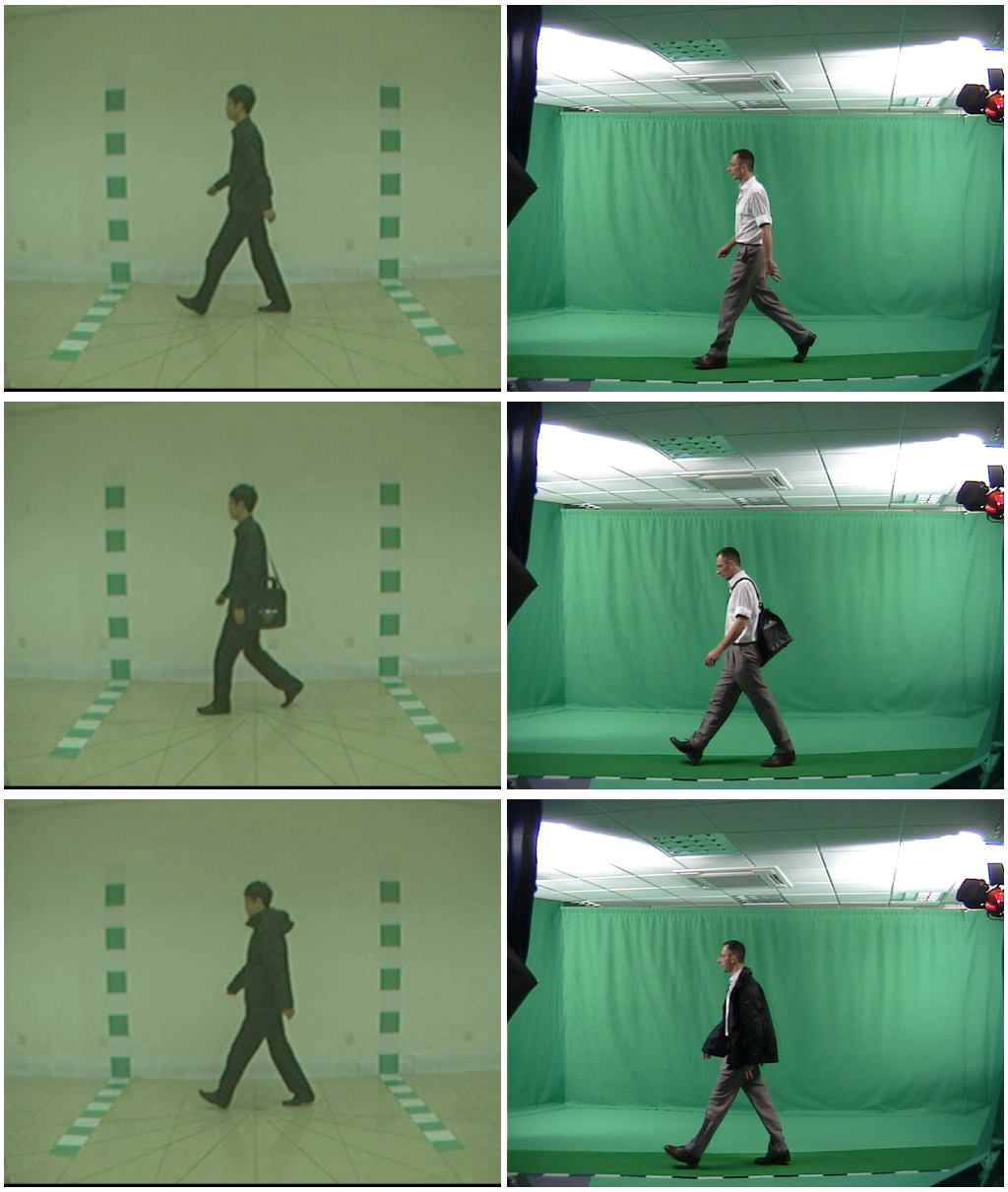
3.3 Experiments

3.3.1 Datasets and Settings

Two datasets, the CASIA and SOTON datasets, are used for experiments which cover both indoor and outdoor scenarios and are amongst the most comprehensive public gait datasets. Both the datasets capture gait sequences from multiple views. Only the front-to-parallel (90°) view is used in the following experiments.

The CASIA gait dataset [116] captures gait video sequences in an indoor environment. The dataset consists of 124 subjects. For each subject there are 10 walking sequences consisting of 6 normal walking sequences where the subject does not carry a bag or wear a bulky coat (CASIASetA), 2 carrying-bag sequences (CASIASetB) and 2 wearing-coat sequences (CASIASetC). Each sequence contains multiple gait cycles. The original image size of the database is 320x240. In the experiments 4 of the 6 normal sequences are used as the gallery set and the rest of the sequences in set CASIASetA (CASIASetA2) are used as the probe set together with CASIASetB and CASIASetC. Example frames from video sequences for a subject under different covariate conditions from the CASIA dataset are shown in Fig. 3.4(a).

The SOTON Small dataset consists of 11 subjects captured indoor under various covariate condition changes. Those covariate conditions include clothing condition (e.g. heavy/light clothes), carrying condition (e.g. carrying/without carrying bags), shoe (e.g. boots/trainers) and speed (quick/slow). In particular, for each subject there are two sequences under normal conditions (SOTONSmallSetA), three carrying-bag sequences (SOTONSmallSetB), two wearing-coat sequences (SOTONSmallSetC), five sequences with different shoes (SOTONSmallSetSh), and one sequence each for slower and quicker walking (SOTONSmallSetS). One of the normal sequences from SotonSmallSetA is used as the gallery set. The other normal sequence (SOTONSmallSetN) and all other subsets are used as the probe set. Example frames from video sequences of the



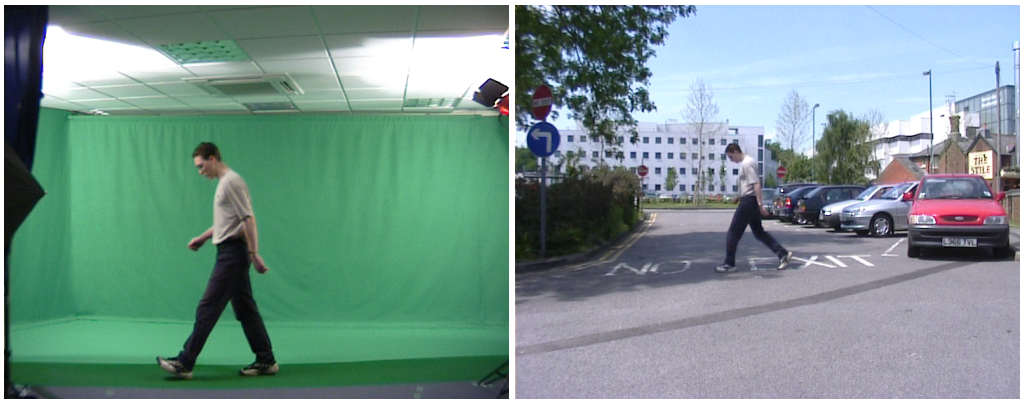
(a) CASIA Dataset [116] (b) SOTON Small Dataset [93]

Figure 3.4: Frames from video sequences of the same subject under different conditions

same subject under different covariate conditions from the SOTON small dataset are shown in Fig. 3.4(b).

The SOTON Large dataset is part of the SOTON database [93]. It contains 116 subjects captured in both indoor and outdoor environment. The dataset has 6 subsets, A to F. In experiments the most widely used Set A (SOTONLargeSetA) and Set E (SOTONLargeSetE) are used. Set A is captured in a controlled indoor environment whilst Set E is captured outdoor with objects moving in the background thus closely resembling a real-world application scenario. Both Set A and Set E feature people walking under unchanged normal conditions. For both Set A and Set E,

half of the sequences for each subject are used as the gallery set and the rest as probe set. Both the Large and Small SOTON datasets are captured at standard PAL resolution. Example frames from the SOTON Large dataset A and E are shown in Fig. 3.5.



(a) SOTON Large Dataset A [93] (b) SOTON Large Dataset E [93]

Figure 3.5: Frames from video sequences of SOTON large dataset A and E

The size of normalized figure-centric images for all datasets are set to 128x88. For all experiments, the results obtained using $\lambda = 0.5$ (see Eqn. 3.5) are reported. The effect of different values of λ is also investigated.

3.3.2 Comparative Evaluation

Probe Set	TM	GEI+CDA	EGEI+CDA	**M**+CDA	$\mathbf{M}_x^+$+CDA	$\mathbf{M}_x^-$+CDA	$\mathbf{M}_y^+$+CDA	Fusion
CASIASetA2	97.6%	99.4%	81.1%	**99.4%**	96.3%	84.9%	93.4%	97.5 %
CASIASetB	52.0%	60.2%	42.3%	56.6%	57.3%	32.8%	46.7%	**83.6%**
CASIASetC	32.7%	30.0%	21.3%	14.8	50.9%	29.1%	24.6%	48.8%
Overall	60.2%	62.8%	48.2%	56.7%	68.2%	48.9%	54.9%	**76.6%**

Table 3.1: Results on the CASIA dataset. The proposed descriptors are compared with a template matching method in [116], GEI [37] using CDA and EGEI [113] using CDA.

Results on CASIA dataset

Table 3.1 shows the recognition rates obtained using the 4 different motion descriptors and the fusion result. The results shows that when there is no change in the covariate conditions (CASIASetA2) Motion Intensity Image (MII **M**) gives slightly better results than the three Motion Direction Images (MDIs $\mathbf{M}_x^+$,$\mathbf{M}_x^-$,$\mathbf{M}_y^+$). When the shape appearance related covariate conditions change in the probe set (CASIASetB and CASIASetC), the opposite is observed. In particular, $\mathbf{M}_x^+$ achieves better result than the other three descriptors when used alone. It can also be seen in Table 3.1 that the fusion result is better than each descriptor alone or close to the best. Overall, considerable improvement can be obtained by fusing all 4 descriptors together. The results are

also compared with those obtained using existing methods in Table 3.1. Apart from the template matching method (TM) [116] all the other methods in Table 3.1 use the same recognition method (i.e., Component and Discriminant Analysis). This comparison method using the same recognition algorithm highlights the effectiveness of Gait Flow Image when compared with state-of-the-art gait representations. A single descriptor in the proposed approach can achieve comparable results whilst the fusion result is significantly higher than those of the alternative methods, particularly when there are changes in covariate conditions.

Probe Set	$\mathbf{M}$	$\mathbf{M}_x^+$	$\mathbf{M}_x^-$	$\mathbf{M}_y^+$	Fusion
SOTONLargeSetA	99.1%	96.5%	96.5%	93.1%	**99.1%**

Table 3.2: Results for Gait Flow Image descriptors on SOTON Large Set A.

Probe Set	MSI	Frieze	SVB Frieze	MSCT+SST	GEI	Fusion
SOTONLargeSetA	84.8%	96.0%	84.0%	84.0%	99.1%	**99.1%**

Table 3.3: Comparison with existing approaches including MSI [58], Frieze Patterns [68], SVB Frieze Patterns [63], MSCT+SST [59] and GEI on SOTON Large Set A.

Probe Set	Wagg and Nixon [100]	$\mathbf{M}$	$\mathbf{M}_x^+$	$\mathbf{M}_x^-$	$\mathbf{M}_y^+$	Fusion
SOTONLargeSetE	67.0%	**100%**	93.1%	93.9%	93.9%	97.4%

Table 3.4: Results on the outdoor SOTON Large Set E.

Results on SOTON Large dataset

The results obtained on the indoor SOTON Large Set A using the proposed motion descriptors are reported in Table 3.2. The result based on fusing the 4 motion descriptors is also compared with the results reported in the literature in Table 3.3. Again, under the same covariate condition, both the MII and the fusion give excellent result. Compared with alternative approaches, both the proposed approach and GEI yield near-perfect result and outperform other approaches. Table 3.4 shows the results on the outdoor dataset SOTON Large Set E. The result suggests that the Gait Flow Image motion descriptors are robust to lighting changes and moving background presented in the outdoor environment. In particular, the results are significantly higher than that of Wagg and Nixon [100].

Probe Set	GEI	$\mathbf{M}$	$\mathbf{M}_x^+$	$\mathbf{M}_x^-$	$\mathbf{M}_y^+$	Fusion
SOTONSmallSetN	100%	100%	100%	100%	100%	**100%**
SOTONSmallSetB	86.3%	76.2%	76.2%	66.7%	61.9%	**90.4%**
SOTONSmallSetC	72.7%	54.6%	81.8%	63.6%	72.7%	**90.9%**
SOTONSmallSetSh	100%	100%	92.3%	82.1%	87.2%	**100%**
SOTONSmallSetS	100%	100%	85.7%	76.1%	76.2%	**100%**
Overall	94.2%	90.3%	87.3%	77.6%	79.6%	**97.1%**

Table 3.5: Results on the Soton Small dataset.

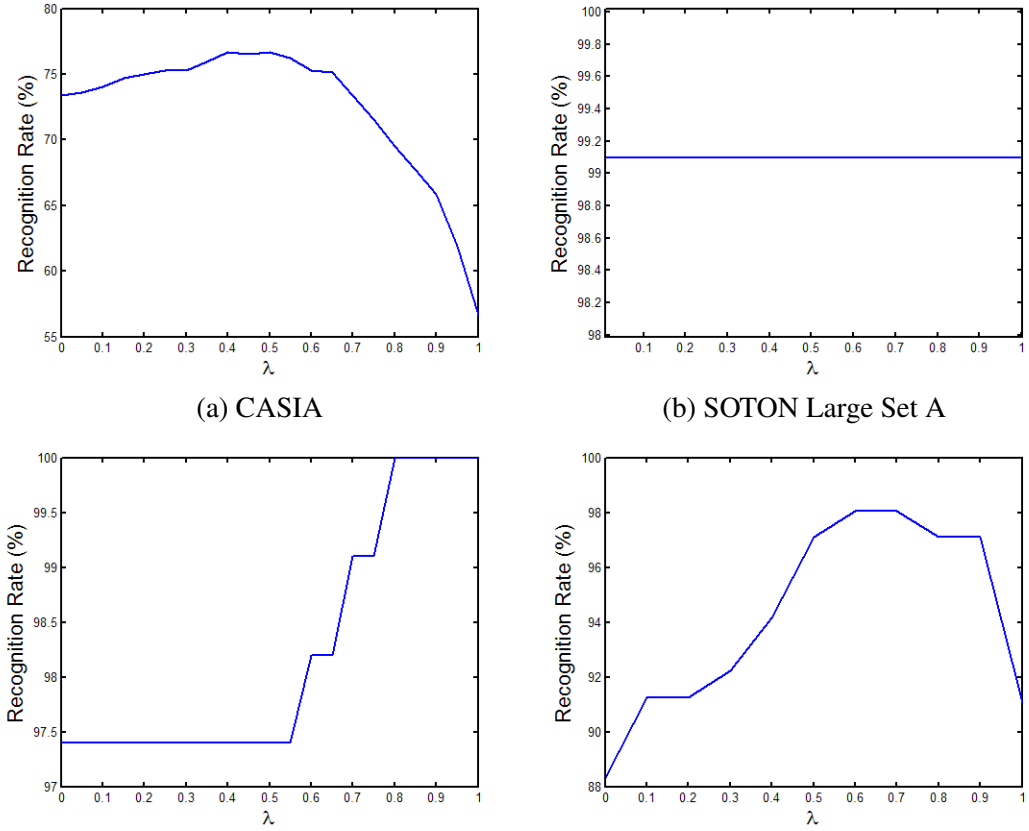
Results on SOTON Small dataset

Table 3.5 shows the results on the SOTON Small dataset which contains changes in shape-appearance related covariates (SOTONSmallSetB and SOTONSmallSetC) and covariates that affect gait itself (SOTONSmallSetSh and SOTONSmallSetS). It can be seen from Table 3.5 that under shape appearance related covariate changes, MDIs give better result than MII. However, when the covariate conditions affect gait itself MII achieves better performance. This result is consistent with the analysis on the characteristics of different descriptors in Sec. 3.1.2. It is also obvious from the results that the descriptor fusion result outperforms that of using each descriptor alone. The performance of the descriptors is compared with GEI and Table 3.5 shows that Gait Flow Image performance is better.

3.3.3 The Effect of the Fusion Weight λ

The results reported so far are obtained by setting the value of the fusion weight λ to 0.5, i.e. equal weight is given to the MII and the three MDIs. GFI representation uses the parameter λ to select contributions from the appearance and motion descriptors. The effect of λ is investigated and the result is shown in Fig. 3.6.

It is found out by experiments that under no covariate condition changes such as the SOTON Large Set A (Indoor) and SOTON Large Set E (Outdoor), (see Fig. 3.6(b,c)) MII plays a major role in recognition performance. MII thus is descriptive enough and can be used as a representation of gait alone when there are no covariate changes. Where as under variable covariate conditions CASIA and SOTON Small Set A (see Fig. 3.6(a,d)) the best result is obtained by incorporating contributions from both static and dynamic aspects of gait. It is found out that starting with only the MDIs i.e. $\lambda = 0$ and increasing the value of λ which in-turn means incorporating the MII, increases the recognition performance up to and around $\lambda = 0.5$ which corresponds to equal weighting of the MII and the three MDIs. As the value of λ continues to increase more and



(a) CASIA (b) SOTON Large Set A

(c) SOTON Large Set E (d) SOTON Small Set

Figure 3.6: Effect of the fusion weight λ. Larger value of λ gives more weight to MII.

more contribution is from the MII with the contribution from MDIs decreasing, the recognition performance degrades. Hence, setting the value of λ either too low or too high will lead to worse performance. This is because that the MII and MDIs contain complementary information and have different levels of sensitivity towards different types of covariates. Overall, a value of 0.5 seems to be a safe choice.

3.4 Discussions

The key findings of experiments are summarised and discussed as below:

1. The Gait Flow Image (GFI) significantly outperforms conventional methods under both normal and variable covariate conditions. The improvement in performance is particularly significant under covariate condition changes. This is indicated by the results of experiments on two of the largest public gait datasets available. This gain in performance of GFIs is mainly attributed to the fact that GFI captures more discriminative information from the relative motion different body parts in different descriptors.

2. GFI captures information about different aspects of gait in different descriptors. The two set of descriptors MII and the MDIs contain complementary information which produces promising results upon fusion.

3. GFI separates the information contained in a gait video sequence into different descriptors and thus enables the individual analysis of different aspects of gait information. For instance, the descriptor corresponding to downward motion direction is removed from the final GFI representation because of lesser discriminative power.

4. The proposed Gait Flow Image representation relies on the extraction of dense optical flow from texture information. In situations where the gait video sequences are captured at a lower resolution or the subjects occupy small regions in the image, the extraction of dense optical flow is not possible. This directly affects the Gait Flow Image descriptors which are extracted from optical flow field computations. The Motion Direction Descriptors are expected to be more severely affected as compared to the Motion Intensity descriptor. Nonetheless, the performance of Gait Flow Image is expected to drop in such circumstances.

5. The Gait Flow Image descriptors are fused at the score level using the fusion weight λ. In this thesis λ is computed experimentally. However, the fusion weight λ can be determined mathematically using a method as proposed by Huang *et al*. [44]. Huang *et al*. [44] propose to fuse the scores from multiple views to improve the performance of gait recognition in a single view. Fusion weights for individual views are computed by taking into consideration the between class and within class distances. A similar approach can be adapted to compute the fusion weight λ for combining the Motion Intensity and the Motion Direction descriptors. Computing the fusion weight mathematically also allows to explore individual weights for all the four descriptors i.e., the Motion Intensity and the three Motion Direction descriptors (λ_1, λ_2, λ_3 and λ_4). This enables analysis of the contributions of individual descriptors towards recognition performance, which experimentally is not feasible.

6. Motion direction descriptors in the Gait Flow Image representation are susceptible to noise in the flow field. Although, a histogram based approach in using the direction of the flow field is adopted to reduce noise, still the noise filters through and affects the recognition performance of motion direction descriptors $\mathbf{M}_x^+, \mathbf{M}_x^-$ and $\mathbf{M}_y^+$. This can be seen by par-

ticularly looking at the results of the normal probe sets (easy case) for $\mathbf{M}_x^+, \mathbf{M}_x^-$ and $\mathbf{M}_y^+$ in Tables 3.1, 3.3 and 3.4.

7. Although the Gait Flow Image is robust to variable covariate conditions it cannot cope with large view angle changes. However, the GFI can be integrated into a cross view recognition framework such as [56, 57, 79] as a choice of representation of gait. The use of Gait Flow Image in a cross-view setting is demonstrated in Chapter 5 where GFI is used in a cross-view recognition framework to perform robust gait recognition under variable covariate conditions.

Chapter 4

Selecting Robust Gait Features for Gait Recognition Without Subject Cooperation

---

The main advantage of gait compared to physiological biometrics is that it does not require subject cooperation and operates at a distance. However, the existing work on gait recognition uses a gallery set consisting of gait sequences of people under similar covariate conditions and evaluates the performance of the proposed methods on probe sets of possibly different covariate conditions [37, 72, 75, 88, 117, 120, 122]. This in effect assumes subject cooperation to acquire a gallery set under normal walking conditions. This chapter investigates the performance of gait recognition approaches when evaluated without the assumption on cooperative subjects, i.e. both the gallery and the probe sets consist of a mixture of gait sequences under different and unknown covariate conditions. Evaluation of the performance of existing gait recognition approaches under the aforementioned realistic experimental set-up indicates that the existing approaches yield very unsatisfactory performance (a nearly 3-fold decrease in recognition rate in some experiments) compared to the result obtained using gallery sequences of similar covariate conditions (see Table 4.3).

The reason for the poor performance is the appearance changes caused by variable and unknown covariate conditions. More specifically, most existing approaches represent gait using features extracted from silhouettes. By extracting silhouettes, many physical appearance features have been removed from the image representation of the human. Nevertheless, a silhouette still contains information about the shape of human body that is vulnerable to changes caused by

conditions such as clothing and carrying. This problem becomes more acute when gait recognition is performed without subject cooperation, i.e. the gallery sequences also have variable covariate conditions. To overcome the problem, it is crucial to select the most relevant gait features that reflect the unique characteristics of gait as a behavioural biometric, and importantly are invariant to appearance variations caused by changes of covariate conditions. To this end, the Gait Entropy Image (GEnI) is proposed to perform automatic feature selection on each pair of gallery and probe gait sequences. Moreover, an Adaptive Component and Discriminant Analysis (ACDA) is formulated which seamlessly integrates the proposed feature selection method with subspace analysis for robust recognition, and importantly is computationally much more efficient compared to the conventional Component and Discriminant Analysis.

The proposed approach is summarised as follows.

1. Gait recognition approaches are evaluated without assuming subject cooperation.
2. A novel adaptive feature selection method based on the Gait Entropy Image (GEnI) is proposed for selecting the most relevant and informative gait features that are invariant to various covariate conditions.
3. A novel Adaptive Component and Discriminant Analysis (ACDA) is developed for fast gait recognition.

To evaluate the effectiveness of the proposed approach, extensive experiments are carried out on two comprehensive benchmarking gait databases: the CASIA database [116] and the Southampton Human ID at a distance gait database (SOTON database) [93]. The results demonstrate that the proposed feature selection based gait recognition method significantly outperforms previous approaches, especially when the gallery set is composed of sequences under variable and unknown gait covariate conditions. Experiments also suggest that the proposed Adaptive Component and Discriminant Analysis (ACDA) is computationally much more efficient than the conventional Component and Discriminant Analysis (CDA) whilst being able to achieve very similar recognition accuracy.

4.1 Feature Selection using Gait Entropy Image

In the following feature selection method based on the Gait Entropy Image is described. The Gait Energy Image is used as the choice of representation to motivate the adaptive feature se-

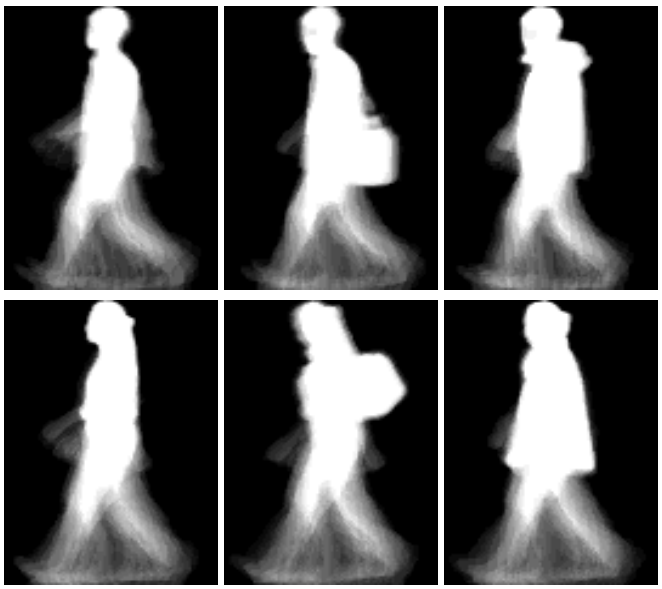
lection method. However, the proposed feature selection method can be readily applied to any image based representation of gait and a comparison of the performance of Gait Energy Image representation with the Gait Flow Image (proposed in this thesis) under variable and unknown covariate conditions is presented in experiments (see Sec. 4.3.5).

4.1.1 Gait Representation using Gait Energy Image

Given a human walking sequence, a human silhouette is extracted from each frame using the method of Sarkar *et al*. [88]. After applying size normalization and horizontal alignment to each extracted silhouette image, gait cycles are segmented by estimating gait frequency using a maximum entropy estimation technique of Sarkar *et al*. [88]. The Gait Energy Image (GEI) is then computed as

$$GEI = G(x,y) = \frac{1}{T}\sum_{t=1}^{T} I(x,y,t), \tag{4.1}$$

where T is the number of frames in a complete gait cycle, I is a silhouette image whose pixel coordinates are given by x and y, and t is the frame number in the gait cycle.



(a) Normal (b) Carrying a bag (c) Wearing a coat

Figure 4.1: Gait Energy Images of people under different carrying and clothing conditions. Top Row: a subject from the CASIA database [116]; bottom row: a subject from the SOTON database [93]. Compared to (b) and (c), the subjects in (a) did not carry a bag or wear a bulky coat.

Examples of GEIs are shown in Fig. 4.1. Note that pixels with high intensity values in a GEI correspond to body parts that move little during a walking cycle (e.g. head, torso), while pixels with low intensity values correspond to body parts that move constantly (e.g. lower parts of legs

and arms). The former mainly contain information about body shape and stance, whilst the later tells more about how people move during walking. The former are called the static areas of a GEI and the latter dynamic areas of a GEI. The dynamic areas are insensitive to human appearance changes caused by common covariate conditions such as carrying condition and clothing; they are thus the most informative part of the GEI representation for human identification given variable covariate conditions. The static areas of a GEI also contain useful information for identification. However, since they mainly contain body shape information, they are sensitive to changes in various covariate conditions. For instance, in each row of Fig. 4.1, three GEIs are computed from three sequences of the same person walking under different conditions. The dynamic areas of the GEI suggest that they are the same person but the static areas suggest otherwise. The above analysis suggests that parts of the information contained in a GEI are redundant and erroneous for human identification. It is well known that the inclusion of redundant and erroneous features in a pattern representation can hamper its recognition.

Based on this observation, an automatic feature selection method is developed to select the most informative gait features from a GEI. This is achieved using a novel Gait Entropy Image (GEnI) based approach formulated below.

4.1.2 Gait Entropy Image

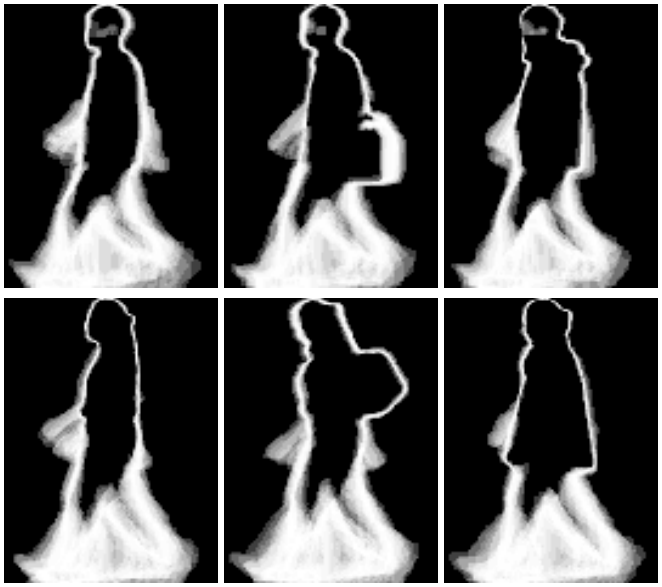
To distinguish the dynamic and static areas of a GEI Shannon entropy [92] is measured at each pixel location in the GEI. More specifically, a gait cycle consists of a sequence of human silhouettes (T silhouettes); consider the intensity value of the silhouettes at a fixed pixel location as a discrete random variable, Shannon entropy measures the uncertainty associated with the random variable over a complete gait cycle and can be computed as

$$GEnI = H(x,y) = -\sum_{k=1}^{K} p_k(x,y)\log_2 p_k(x,y) \tag{4.2}$$

where x,y are the pixel coordinates and $p_k(x,y)$ is the probability that the pixel takes on the k^{th} value. In this case the silhouettes are binary images and thus $K = 2$, $p_1(x,y) = \frac{1}{T}\sum_{t=1}^{T} I(x,y,t)$ (i.e. the GEI) and $p_0(x,y) = 1 - p_1(x,y)$. Note that there is a close link between GEnI and a GEI. Specifically, let $z = p_1(x,y)$ representing a GEI. GEnI can thus be written in terms of GEI by expanding Eqn. 4.2

$$GEnI = -z * \log_2 z - (1-z) * \log_2(1-z). \tag{4.3}$$

GEnI gives an insight into the information content of the gait sequence as the intensity value at pixel location (x,y) is proportional to its entropy value $H(x,y)$. Fig. 4.2 shows the Gait Entropy Images (GEnI) obtained from the GEIs depicted in Fig. 4.1. It is evident from Fig. 4.2 that the dynamic areas in a GEI are featured with high intensity values in its corresponding GEnI whilst the static areas have low values. This is not surprising because silhouette pixel values in dynamic areas are more uncertain and thus more informative leading to higher entropy values.



(a) Normal (b) Carrying a bag (c) Wearing a coat

Figure 4.2: Examples of Gait Entropy Images. Their corresponding Gait Energy Images are shown in Fig. 4.1.

GEnI can be used directly for selecting informative gait features from GEI. However, Fig. 4.2 also suggests that the body shape changes caused by varying covariate conditions such as carrying and clothing are still visible in the GEnI. Consequently, the selected features from a GEI will contain information that is irrelevant to gait. To overcome this problem, a feature selection method is proposed which selects a unique set of features for each pair of gallery and probe gait sequences.

4.1.3 Pairwise Feature Selection using Gait Entropy Image

For each pair of gallery and probe gait sequences and their corresponding GEIs, the aim is to select a set of gait features that are invariant to covariate condition changes and unique to the pair. To this end, first a binary feature selection mask [1] $M_G(x,y)$ is generated for each GEI using

[1] A feature selection mask determines whether features from a specific pixel location should be selected. It therefore has to be binary.

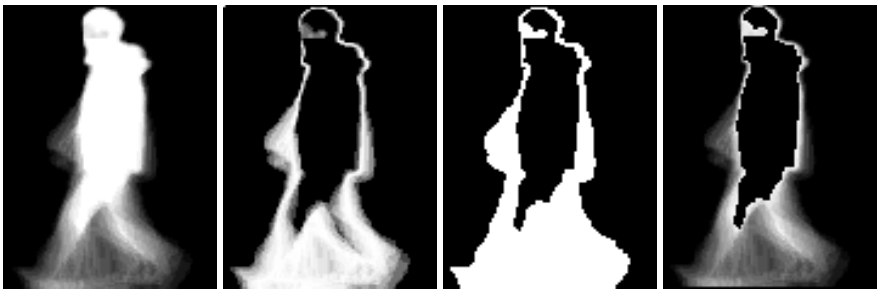
its corresponding GEnI $H(x,y)$.

$$M_G(x,y) = \begin{cases} 1, & \text{if } H(x,y) > \theta \\ 0, & \text{Otherwise} \end{cases} \tag{4.4}$$

where θ is a threshold. Note that instead of putting a threshold on the GEnI an upper and lower threshold can be obtained for the GEI (generating the same feature selection mask) by solving Eqn. 4.3 for z as follows.

$$\theta = -z * \log_2 z - (1-z) * \log_2 (1-z). \tag{4.5}$$

This solution gives z_1 and $z_2 = 1 - z_1$ as lower and upper thresholds respectively for the GEI. Examples of feature selection mask generated using GEnI or the GEI (both generate same feature selection mask) are shown in Fig. 4.3.



(a) GEI (b) GEnI (c) Mask Image (d) Masked GEI

Figure 4.3: An example of feature selection mask generated using GEnI.

Suppose the gallery set contains N with $n = 1,...,N$ GEIs belonging to C classes with $i = 1,...,C$ and there are multiple examples in each class. For the nth gallery GEI belonging to the ith class, a feature selection mask $M_G^n(x,y)$ is then generated using Eqn. 4.4. Similarly $M_G^l(x,y)$ is obtained for the lth probe GEI belonging to the jth probe class. The masks for individual examples belonging to the same class in the gallery set are combined to generate a single mask for that class using a simple 'AND' operation in order to select features common to all examples in the class. A similar strategy is used in case of multiple examples in the probe class. Let $M_G^i(x,y)$ and $M_G^j(x,y)$ denote the combined feature selection masks for the ith gallery class and the jth probe class respectively. Now to select features that are relevant for both classes, these two masks are to be combined. This is done using the binary 'AND' operation to select features specific to the probe/gallery pairs in question.

$$M_G^{ij}(x,y) = M_G^i(x,y) \&\& M_G^j(x,y) \tag{4.6}$$

where &&, is the binary 'AND' operator.

(a) (b) (c) (d) (e)

Figure 4.4: Example of feature selection for a probe/gallery pair. (a) gallery GEI; (b) probe GEI; (c) feature selection mask $M_G^{ij}(x,y)$; (d) gallery GEI with $M_G^{ij}(x,y)$ applied; (e) probe GEI with $M_G^{ij}(x,y)$ applied.

Fig. 4.4 shows an example of applying the feature selection method to a pair of gallery and probe GEIs under different covariate conditions. It is evident that after applying the pairwise feature selection mask, the effect of the changes in covariate conditions in the gallery and probe sequences is alleviated effectively. In particular, Fig. 4.4(d) and (e) give strong indication that the two images are captured from the same person, although both the carrying and clothing conditions are different as shown in Fig. 4.4(a) and (b).

4.2 Adaptive Component and Discriminant Analysis

After applying a feature selection mask $M_G^{ij}(x,y)$, gait recognition can be performed by matching the class centre of probe GEIs to the class centre of gallery GEIs that has the minimal distance between them. However, direct template matching has been shown to be sensitive to noise and small silhouette distortions [37, 73]. This is because that the dimensionality of the GEI feature space is high even after feature selection (typically in the order of thousands). To overcome this problem, subspace Component and Discriminant Analysis (CDA) based on Principal Component Analysis (PCA) and Multiple Discriminant Analysis (MDA) can be adopted which seeks to project the original features to a subspace of lower dimensionality so that the best data representation and class separability can be achieved simultaneously [43].

Suppose there are N d-dimensional gallery GEI templates $\{\mathbf{x}_1, ..., \mathbf{x}_n, ..., \mathbf{x}_N\}$ belonging to C different classes (individuals), where each template is a column vector obtained by concatenating

the rows of the corresponding GEI. To compute the distance between the ith gallery and the jth probe classes, $M_G^{ij}(x,y)$ is applied to each gallery GEI, which gives a new set of template $\{\mathbf{x}_1^{ij},...,\mathbf{x}_n^{ij},...,\mathbf{x}_N^{ij}\}$ of dimension d^{ij}. PCA is an orthogonal linear transformation that transforms the data to a subspace of dimensionality $\tilde{d^{ij}}$ (with $\tilde{d^{ij}} < d^{ij}$). The PCA subspace keeps the greatest variances by any projection of the data so that the reconstruction error defined below is minimized:

$$J_{\tilde{d^{ij}}} = \sum_{n=1}^{N} \left\| \left(\mathbf{m} + \sum_{k=1}^{\tilde{d^{ij}}} a_{nk} \mathbf{e}_k^{ij} \right) - \mathbf{x}_n^{ij} \right\|^2 \tag{4.7}$$

where $\mathbf{m}$ is the mean of the data, $\{\mathbf{e}_1^{ij},\mathbf{e}_2^{ij},...,\mathbf{e}_{\tilde{d^{ij}}}^{ij}\}$ are a set of orthogonal unit vectors representing the new coordinate system of the subspace, a_{nk} is the projection of the nth data to $\mathbf{e}_k^{ij}$. $J_{\tilde{d^{ij}}}$ is minimised when $\{\mathbf{e}_1^{ij},\mathbf{e}_2^{ij},...,\mathbf{e}_{\tilde{d^{ij}}}^{ij}\}$ are the $\tilde{d^{ij}}$ eigenvectors of the data covariance matrix with the largest eigenvalues (in decreasing order). Now the gallery template $\mathbf{x}_n^{ij}$ is represented as a $\tilde{d^{ij}}$-dimensional feature vector $\mathbf{y}_n^{ij}$ and hence

$$\mathbf{y}_n^{ij} = M_{pca}^{ij} \mathbf{x}_n^{ij} = [\mathbf{e}_1^{ij},...,\mathbf{e}_{\tilde{d^{ij}}}^{ij}]^T \mathbf{x}_n^{ij}. \tag{4.8}$$

PCA is followed by MDA which aims to find a subspace where data from different classes are best separated in a least square sense. Different from PCA, MDA is a supervised learning method which requires the gallery data to be labelled into classes. The MDA transformation matrix, W^{ij} maximizes

$$J(W^{ij}) = \frac{|W^{ij^T} S_B^{ij} W^{ij}|}{|W^{ij^T} S_W^{ij} W^{ij}|}$$

where S_B^{ij} is the between-class scatter matrix and S_W^{ij} the within-class scatter matrix of the gallery data in the PCA subspace $\{\mathbf{y}_1^{ij},...,\mathbf{y}_n^{ij},...,\mathbf{y}_N^{ij}\}$. $J(W^{ij})$ is maximized by setting the columns of W^{ij} to the generalized eigenvectors that correspond to the $C-1$ nonzero eigenvalues in

$$S_B^{ij} \mathbf{w}_k^{ij} = \lambda_i^j S_W^{ij} \mathbf{w}_k^{ij}$$

where $\mathbf{w}_k^{ij}$ is the kth column of W^{ij} and C is the number of classes in the gallery data. Denoting these generalised eigenvectors as $\{\mathbf{v}_1^{ij},\mathbf{v}_2^{ij},...,\mathbf{v}_{C-1}^{ij}\}$, a gallery template is represented in the MDA subspace as:

$$\mathbf{z}_n^{ij} = M_{mda}^{ij} \mathbf{y}_n^{ij} = [\mathbf{v}_1^{ij},...,\mathbf{v}_{C-1}^{ij}]^T \mathbf{y}_n^{ij}. \tag{4.9}$$

Note that the choice of $\tilde{d^{ij}}$ is affected by the dimensionality of the MDA subspace, i.e. $C-1$. In particular, S_W^{ij} becomes singular when $\tilde{d^{ij}} < C$ or $\tilde{d^{ij}} \gg C$. Therefore $\tilde{d^{ij}} = 2C$ in this thesis.

Now after three steps of dimensionality reduction (feature selection using $M_G^{ij}(x,y)$, PCA, and MDA), both the gallery and probe GEI feature vectors are represented in a $C-1$ dimensional subspace. This dimensionality reduction process is computationally expensive mainly due to the PCA step. This is because for each new gallery and probe class pair, a new mask $M_G^{ij}(x,y)$ is generated and PCA needs to be re-done which involves eigen-decomposition of a $N \times N$ matrix. To make this approach more computationally efficient, an Adaptive Component and Discriminant Analysis (ACDA) is developed. More specifically, instead of applying each $M_G^{ij}(x,y)$ to the gallery templates and re-do the PCA on $\{\mathbf{x}_1^{ij},...,\mathbf{x}_n^{ij},...,\mathbf{x}_N^{ij}\}$, PCA is only computed once for the original gallery templates $\{\mathbf{x}_1,...,\mathbf{x}_n,...,\mathbf{x}_N\}$, which results in a base PCA subspace. Base PCA subspace is adapted towards each gallery and probe sequence pair by applying $M_G^{ij}(x,y)$ directly to the base principal components. Specifically, let $\{\mathbf{e}_1,\mathbf{e}_2,...,\mathbf{e}_{\tilde{d}}\}$ be the base components, each component can be treated as an eigenGEI, similar to eigenface for face recognition. The adapted components $\{\mathbf{u}_1^{ij},\mathbf{u}_2^{ij},...,\mathbf{u}_{\tilde{d}}^{ij}\}$ are then obtained by applying $M_G^{ij}(x,y)$ to the eigenGEIs. Now Eqn. 4.8 can be re-written as

$$\mathbf{y}_n^{ij} = M_{pca}^{ij}\mathbf{x}_n^{ij} = [\mathbf{u}_1^{ij},...,\mathbf{u}_{\tilde{d^{ij}}}^{ij}]^T\mathbf{x}_n^{ij}. \tag{4.10}$$

The MDA step that follows will remain unchanged (see Eqn. 4.9) and recognition can now be performed by computing a distance measure between the class centres of the gallery and probe GEIs.

In Adaptive Component and Discriminant Analysis (ACDA) $\{\mathbf{e}_1^{ij},\mathbf{e}_2^{ij},...,\mathbf{e}_{d^{ij}}^{ij}\}$ are approximated using $\{\mathbf{u}_1^{ij},\mathbf{u}_2^{ij},...,\mathbf{u}_{\tilde{d}}^{ij}\}$ in order to reduce the computational cost. What price has to be paid for this improvement in computational efficiency will depend on the accuracy of the approximation. Intuitively, applying a binary mask $M_G^{ij}(x,y)$ to the gallery data collapses some of the original coordinate axes. $\{\mathbf{e}_1^{ij},\mathbf{e}_2^{ij},...,\mathbf{e}_{d^{ij}}^{ij}\}$, as the subspace expressed in the original coordinate system, should also have the corresponding axes collapsed, which is exactly how $\{\mathbf{u}_1^{ij},\mathbf{u}_2^{ij},...,\mathbf{u}_{\tilde{d}}^{ij}\}$ are generated. Theoretically, it can be readily proved that the projection of $\{\mathbf{x}_1^{ij},...,\mathbf{x}_n^{ij},...,\mathbf{x}_N^{ij}\}$ to $\{\mathbf{u}_1^{ij},\mathbf{u}_2^{ij},...,\mathbf{u}_{\tilde{d}}^{ij}\}$ will have an identical diagonalised covariance matrix as their projection on $\{\mathbf{e}_1^{ij},\mathbf{e}_2^{ij},...,\mathbf{e}_{d^{ij}}^{ij}\}$ [48]. More importantly, it is demonstrated through experiments in the next section that the approximation is extremely accurate in practice.

4.3 Experiments

Two experiments are carried out in this thesis. In the first experiment, the gallery set contains sequences of people walking under similar covariate conditions, i.e. the same experimental set-up as the existing work. In the second experiment, the gallery set is composed of a mixture of gait sequences collected under different unknown covariate conditions. The experimental setting for the second experiments is designed to reflect a real world scenario where no subject cooperation is required and therefore the covariate conditions for the subjects in both the gallery and probe sets are different and unknown.

4.3.1 Datasets

The CASIA Gait Database [116] and the Small dataset from Southampton Human ID at a distance gait database (SOTON database) [93] are used to evaluate the performance of the proposed approach. The datasets are described in detail in Sec. 3.3.1 of Chapter 3.

Both the CASIA and the SOTON Small datasets are designed for investigating the robustness of gait recognition techniques to imagery of the same subject in various common conditions (e.g. carrying items, clothing). In previous work where they are used, gait sequences in either probe or gallery sets have the same covariate condition, whilst in these experiments, both the gallery and the probe sets consist of a mixture of gait sequences under different and unknown covariate conditions. The two datasets are perfectly suitable for gait recognition without subject cooperation under an experimental setting that differs from those in previous work. It is thus more suitable for evaluating uncooperative gait recognition and is employed in experiments.

Sample GEI images and GEnI images from both databases are shown in Fig. 4.1 and Fig. 4.2 respectively. Note that the only free parameter in this approach is the threshold value θ used in Eqn. 4.4. In the rest of the section, reported results are obtained when θ is set to 0.75. The effect of θ on the recognition performance is analysed in Section 4.3.4.

4.3.2 Gallery Sequences under Similar Covariate Conditions

In this experiment, the same experimental set-up detailed in Sec. 3.3.1 of Chapter 3 is used. The results obtained using the proposed approach, termed as $M_G^{ij}(x,y)$+ACDA, are compared with the results published in [116] which are obtained using direct template matching on the same databases and the approach in [37] which is based on the standard CDA without any feature

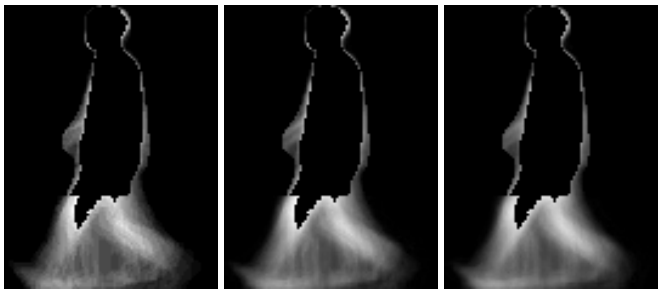
selection on the GEIs. The approach in [37] is widely regarded as one of best gait recognition approach. The performance is measured using recognition rates and is presented in Table 4.1.

Probe Set	TM	GEI+CDA	EGEI+CDA	M_G^j+CDA	M_G^j+ACDA	$M_G^{ij}(x,y)$+CDA	$M_G^{ij}(x,y)$+ACDA
CASIASetA2	97.6%	99.4%	81.1%	100%	99.1%	**100%**	100%
CASIASetB	52.0%	60.2%	42.3%	70.5%	70.0%	**78.3%**	77.8%
CASIASetC	32.7%	30.0%	21.3%	35.5%	35.1%	**44.0%**	43.1%
SOTONSmallSetN	100%	100%	—	100%	100%	**100%**	100%
SOTONSmallSetB	54.5%	86.3%	—	86.3%	**86.3%**	81.8%	81.8%
SOTONSmallSetC	45.4%	72.7%	—	81.8%	81.8%	**83.3%**	81.8%

Table 4.1: Comparing different approaches using a gallery set consisting of sequences under similar covariate conditions (without carrying a bag or wearing a coat). TM: direct GEI template matching; GEI+CDA: method in [37] based on CDA; EGEI+CDA: method in [113] followed by CDA; M_G^j+CDA: feature selection using masks generated only from the probe sequence followed by CDA; M_G^j+ACDA: feature selection using masks generated only from the probe sequence followed by ACDA; $M_G^{ij}(x,y)$+CDA: the proposed approach with feature selection using masks generated from each pair of gallery and probe sequences followed by CDA; $M_G^{ij}(x,y)$+ACDA: the proposed approach with feature selection using masks generated from each pair of gallery and probe sequences followed by ACDA.

It can be seen from Table 4.1 that direct template matching gives the worst results. CDA based approach improves on template matching for most probe sets but there is still much room for further improvement, especially when the probe sets have different covariate conditions from the gallery sets. Table 4.1 shows that the proposed method ($M_G^{ij}(x,y)$+ACDA) significantly outperforms both template matching and GEI+CDA, EGEI+CDA for all probe sets except for SOTONSmallSetB where people carry different bags (e.g. rucksack, laptop bag, and suitcases). The improvement is particularly substantial for the probe set with a different clothing condition (CASIASetC and SOTONSmallSetC), on which poor results are obtained without feature selection.

This approach is also compared with an alternative approach which also performs features selection using GEnIs but use feature selection mask generated from the probe GEnIs only (M_G^j+ACDA). These results will highlight the effect of having a pair-wise feature selection mask ($M_G^{ij}(x,y)$). Specifically, a feature selection mask using Eqn. 4.4 is generated for each probe class and applied to all gallery GEIs. Table 4.1 shows that using the mask generated from the probe only, the result is still much better compared to those of previous approaches without feature selection, albeit it is slightly worse than the result obtained using a mask generated for gallery-probe pair.



(a) (b) (c)

Figure 4.5: (a): A GEI with $M_G^{ij}(x,y)$ applied; (b) The reconstructed GEI using $\{\mathbf{e}_1^{ij},\mathbf{e}_2^{ij},...,\mathbf{e}_{d^{ij}}^{ij}\}$; (c) The reconstructed GEI using $\{\mathbf{u}_1^{ij},\mathbf{u}_2^{ij},...,\mathbf{u}_{\tilde{d}}^{ij}\}$. The root-mean-square errors, which are $J_{\tilde{d}^{ij}}$ (see Eqn. 4.7) normalised by the image size, is 0.0031 for (b) and 0.0060 for (c).

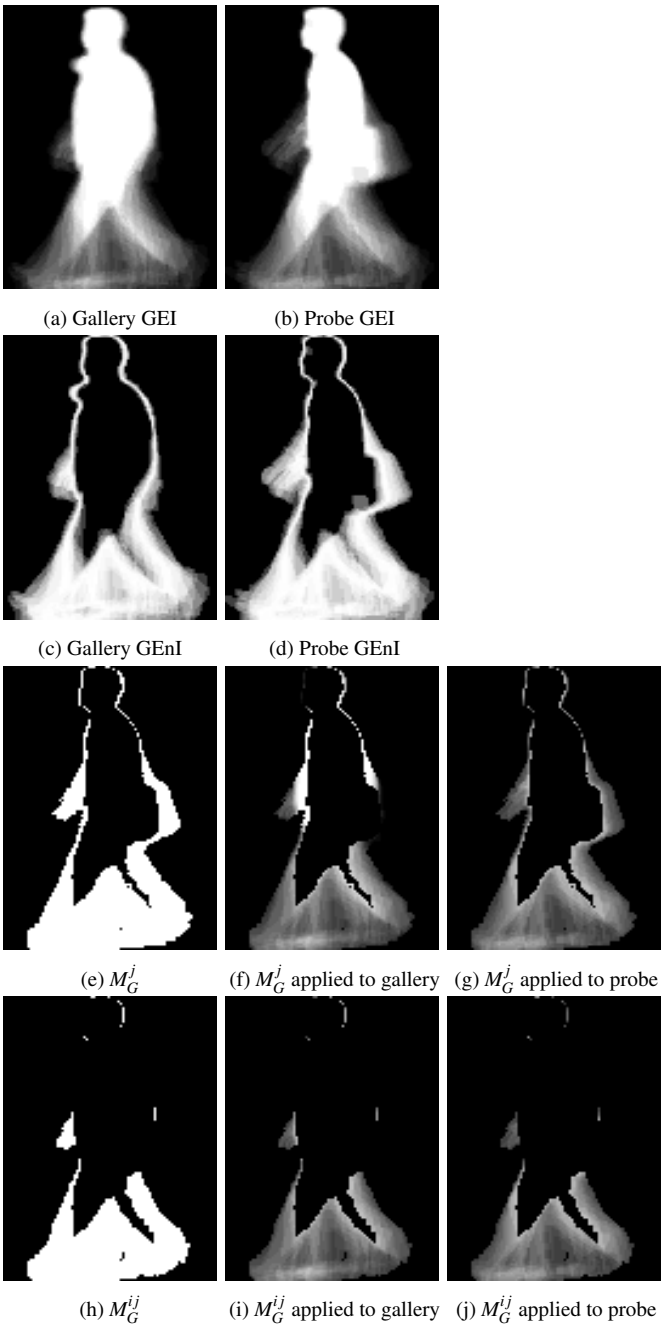
Table 4.1 also lists the results with feature selection using the mask generated from GEnIs followed by CDA. The difference in the respective results obtained by using CDA (M_G^j+CDA, M_G^{ij}+CDA) and ACDA (M_G^j+ACDA, M_G^{ij}+ACDA) will indicate how much sacrifice in recognition accuracy needs to be made in exchange for lower computational cost for the proposed ACDA. It can be seen that Adaptive Component and Discriminant Analysis (ACDA) method achieves almost identical results as the CDA approach. This suggests that approximation of $\{\mathbf{e}_1^{ij},\mathbf{e}_2^{ij},...,\mathbf{e}_{d^{ij}}^{ij}\}$ using $\{\mathbf{u}_1^{ij},\mathbf{u}_2^{ij},...,\mathbf{u}_{\tilde{d}}^{ij}\}$ is accurate. Fig. 4.5(b) and (c) show examples of reconstructed GEIs using $\{\mathbf{e}_1^{ij},\mathbf{e}_2^{ij},...,\mathbf{e}_{d^{ij}}^{ij}\}$ and $\{\mathbf{u}_1^{ij},\mathbf{u}_2^{ij},...,\mathbf{u}_{\tilde{d}}^{ij}\}$ respectively. Both of them gave extremely small reconstruction errors. As for computational cost, as indicated by Table 4.2, ACDA is much more computationally efficient than CDA. The result is obtained using a platform with an Intel Dual Core 1.86GHz CPU and 2GB memory.

	CDA	ACDA
Computational cost	36.70 sec	0.19 sec

Table 4.2: Comparison of the computational cost for performing a computation of ACDA and CDA once on the CASIA dataset.

4.3.3 Gallery Sequences Under Variable and Unknown Covariate Conditions

In this experiment, the gallery sets include a mixture of normal, carrying-bag, and wearing-coat sequences, which give a challenging experimental setting closely representing the condition for gait recognition without subject cooperation. More specifically, for the CASIA dataset the first one third of the sequences are selected from CASIASetC, the second one third from CASIASetB and the last one third from the first 4 sequences in CASIASetA. The probe sets consist of the rest of the dataset and are referred to as CASIASetA2, CASIASetB2 and CASIASetC2. For the SO-



(a) Gallery GEI (b) Probe GEI

(c) Gallery GEnI (d) Probe GEnI

(e) M_G^j (f) M_G^j applied to gallery (g) M_G^j applied to probe

(h) M_G^{ij} (i) M_G^{ij} applied to gallery (j) M_G^{ij} applied to probe

Figure 4.6: Comparing the effectiveness of the feature selection mask M_G^j generated using probe GEI only, and M_G^{ij} generated using both the gallery and probe GEIs. The subject wears a coat in the gallery sequence and carries a bag in the probe sequence.

TON dataset the mixed gallery set contains the first one third of the subjects from SOTONSmall-SetA the second one third from SOTONSmallSetB and last one third from SOTONSmallSetC. The probe sets include the rest of the sequences and are termed as SOTONSmallSetA3, SOTON-SmallSetB2 and SOTONSmallSetC2 respectively. The probe and the gallery sets are mutually exclusive.

The experimental results are presented in Table 4.3. The results indicate a drastic degra-

Probe Set	CDA	M_G^j+ACDA	$M_G^{ij}(x,y)$+ACDA
CASIASetA2	48.1%	58.2%	**69.1%**
CASIASetB2	31.9%	37.5%	**55.6%**
CASIASetC2	9.7%	23.6%	**34.7%**
TotalCASIA	32.6%	42.5%	**55.5%**
SOTONSmallSetA3	45.5%	45.5%	**63.6%**
SOTONSmallSetB2	31.82%	50.0%	**50.0%**
SOTONSmallSetC2	36.4%	36.3%	**54.6%**
TotalSOTON	36.3%	45.5%	**54.5%**

Table 4.3: Comparing different approaches using gallery sets consisting of sequences under different covariate conditions.

dation in performance for the CDA based method without feature selection [37] and the approach with feature selection using the probe GEnIs only. In comparison, the proposed method ($M_G^{ij}(x,y)$+ACDA) achieves much better result, especially for the probe sequences where people carry a bag or wear bulking clothes. This result suggests that under such a realistic experimental set-up, feature selection based on each pairs of gallery and probe gait sequences is critical for selecting the gait features that are invariant to covariate condition changes. This is evident from an example shown in Fig. 4.6. It can be seen from Fig. 4.6 that after applying the mask generated using both the gallery and probe GEnIs, the gallery and probe GEIs can be correctly matched, whilst the mask generated using the probe sequence alone cannot deal with the variations in GEIs caused by changes in covariate conditions resulting in an incorrect match.

4.3.4 The Effect of Feature Selection Threshold θ

The effect of the only free variable θ on the recognition rate for the CASIA database without subject cooperation is investigated. The result is shown in Fig. 4.7. The value of θ ranges from 0 to 1 with smaller values corresponding to less features being selected. Fig. 4.7 shows that similar recognition rate can be achieved when the value of θ is between 0 to 0.95. This suggests that the method is insensitive to the setting of θ.

4.3.5 Comparing Different Gait Representations

Thus far, the Gait Energy Image has been used to examine the affects of variable and unknown covariate conditions in order to perform gait recognition without subject cooperation. An experiment is conducted to investigate the performance of Gait Flow Image (GFI) proposed in Chapter 3 under this challenging experimental set-up. Using the set-up proposed in Sec. 4.3.3 results obtained are presented in Table 4.4. The GFI descriptors are treated individually and fu-

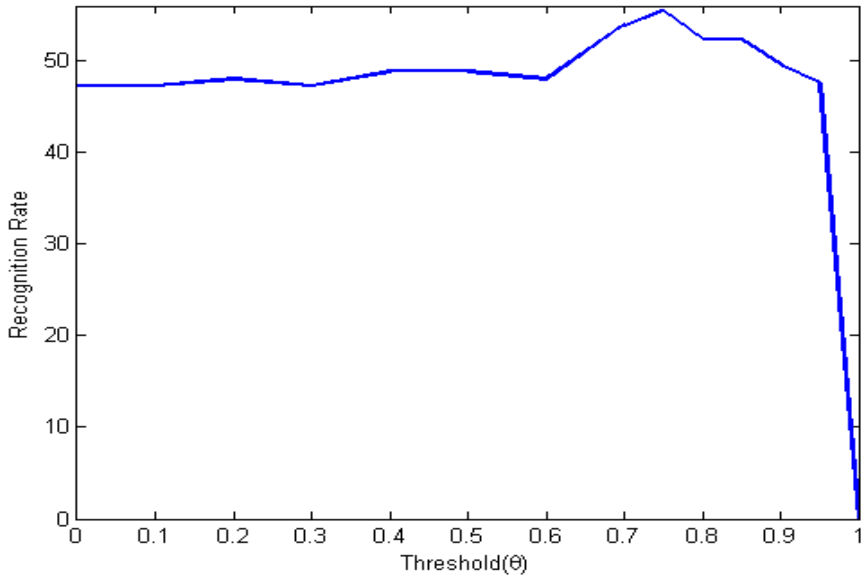


Figure 4.7: The effect of θ on the recognition rate.

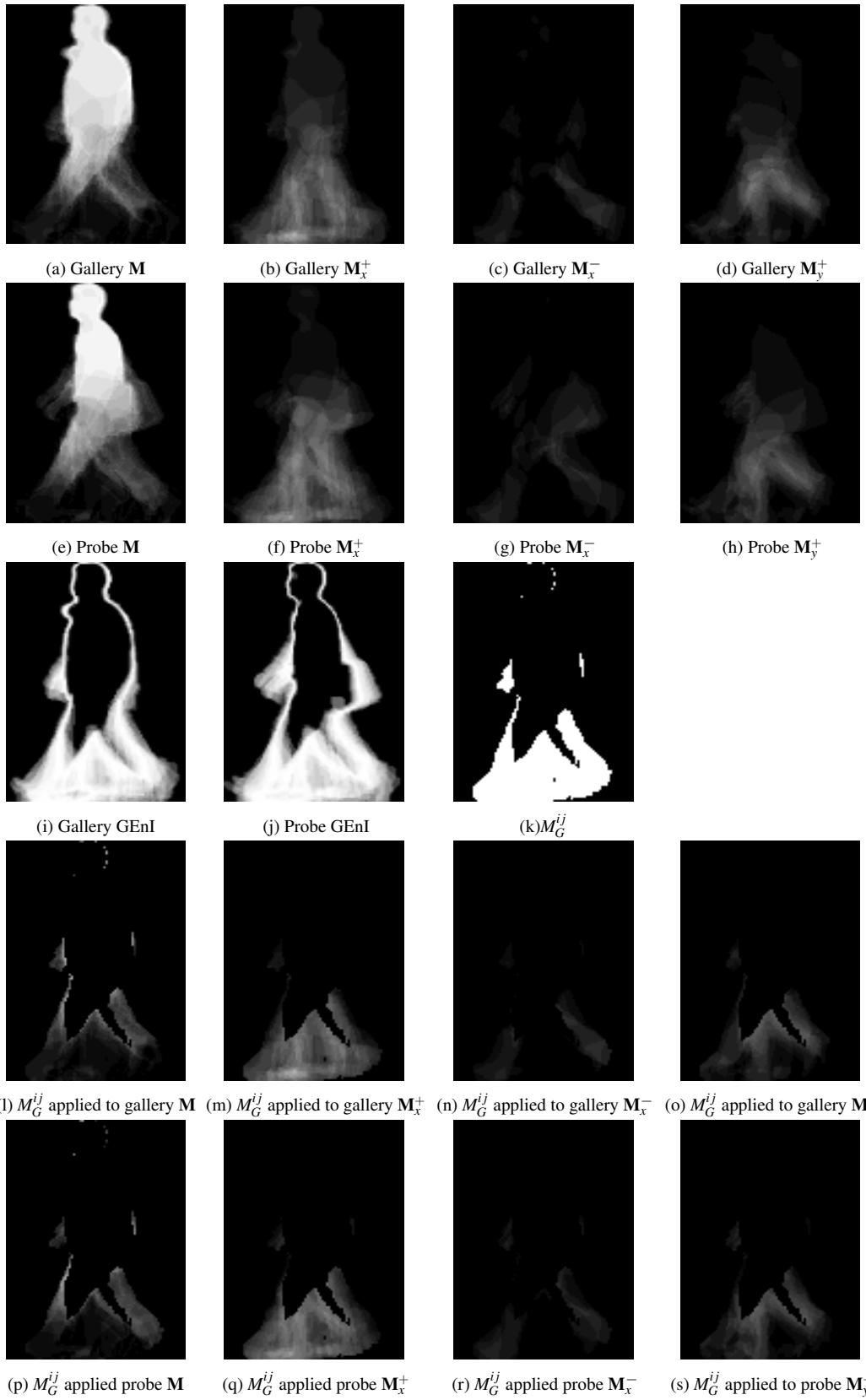
sion is done at the score level to reach at the final recognition score. The same value of $\lambda = 0.5$ is used as in Chapter 3 to fuse the MII and MDIs.

From Table 4.4 it can be seen that the presence of variable and unknown covariate conditions incurs a significant drop on the performance of GEI when adaptive feature selection methods are not used (see GEI+CDA, Table 4.4). The performance of GFI (GFI+CDA) is also affected. However, it drops more gracefully and performs better than GEI (GEI+CDA).

Probe Set	GEI+CDA	GFI+CDA	$M_G^{ij}(x,y)$+ACDA	$M_F^{ij}(x,y)$+ACDA
CASIASetA2	48.1%	51.8%	**69.1%**	66.3%
CASIASetB2	31.9%	48.6%	**55.6%**	52.7%
CASIASetC2	9.7%	19.4%	34.7%	**37.5%**
TotalCASIA	32.6%	41.7%	**55.1%**	54.3%
SOTONSmallSetA3	45.5%	45.5%	**63.6%**	54.5
SOTONSmallSetB2	31.82%	50.0%	50.0%	**54.5%**
SOTONSmallSetC2	36.4%	36.3%	54.6%	**63.6%**
TotalSOTON	36.3%	45.5%	54.5%	**56.8%**

Table 4.4: Comparing the performance of GEI and GFI using gallery set consisting of sequences under variable and unknown covariate conditions.

Using the proposed adaptive feature selection method the results for GFI, termed as $M_F^{ij}(x,y)$+ACDA, are much better than that of GFI without feature selection (see columns 3 and 5, Table 4.4). In Fig. 4.8 the same adaptive feature selection mask used in Fig. 4.6 is applied to the GFI. From Fig. 4.8 it is seen that before applying the feature selection mask the gallery and probe GFI mainly



(a) Gallery $\mathbf{M}$ (b) Gallery $\mathbf{M}_x^+$ (c) Gallery $\mathbf{M}_x^-$ (d) Gallery $\mathbf{M}_y^+$

(e) Probe $\mathbf{M}$ (f) Probe $\mathbf{M}_x^+$ (g) Probe $\mathbf{M}_x^-$ (h) Probe $\mathbf{M}_y^+$

(i) Gallery GEnI (j) Probe GEnI (k)M_G^{ij}

(l) M_G^{ij} applied to gallery $\mathbf{M}$ (m) M_G^{ij} applied to gallery $\mathbf{M}_x^+$ (n) M_G^{ij} applied to gallery $\mathbf{M}_x^-$ (o) M_G^{ij} applied to gallery $\mathbf{M}_y^+$

(p) M_G^{ij} applied probe $\mathbf{M}$ (q) M_G^{ij} applied probe $\mathbf{M}_x^+$ (r) M_G^{ij} applied probe $\mathbf{M}_x^-$ (s) M_G^{ij} applied to probe $\mathbf{M}_y^+$

Figure 4.8: Feature selection mask M_G^{ij} of Fig. 4.6 applied to the Gait Flow Images. The subject wears a coat in the gallery sequence and carries a bag in the probe sequence

differ in the Motion Intensity Images **M** (see Fig. 4.8(a & e)). However, after the application of the feature selection mask the gallery and probe Motion Intensity Images **M** (see Fig. 4.8(l & p)) are correctly matched. The same is true for the Motion Direction Images (MDIs) $\mathbf{M}_x^+$, $\mathbf{M}_x^-$ and $\mathbf{M}_y^+$.

The recognition performance of GFI and GEI with the proposed adaptive feature selection method are pretty similar under variable and unknown covariate conditions (see columns 4 and 5, Table 4.4). This suggests that after adaptive feature selection, the suggested GFI feature, although being able to deal with variable and unknown covariate conditions, loses part of its discriminative power.

4.4 Discussions

The key findings of experiments are summarised and discussed as below:

1. The proposed approach significantly outperforms direct template matching and Han and Bhanu's GEI+CDA approach [37] in both experimental conditions, i.e. the traditional experimental set-up and the new set-up proposed in this work assuming no subject cooperation. This is mainly due to the novel Gait Entropy Image (GEnI) based feature selection method proposed in this thesis. In particular, even with the same gait representation (GEI) and recognition algorithm (CDA), large improvement can be achieved by selecting features that are invariant to covariate condition changes.

2. When both the gallery and probe sets contain sequences of different and unknown covariate conditions, all gait recognition approaches, including this one, suffer from significant decrease in recognition performance. This is hardly surprising as human identification without cooperative subjects is the 'holy grail' of biometrics research and is widely regarded as the most challenging problem yet to be solved. But importantly results show that under this challenging and more realistic experimental setting, performing feature selection, particularly selecting a unique set of features for each pair of gallery and probe sequences, is crucial and much more promising than alternative approaches in solving the problem.

3. The improvement obtained by using the pair-wise feature selection strategy potentially comes with a price, that is, the computational cost can be very high which may hinder

the implementation of this approach for real time applications. Fortunately, the proposed Adaptive Component and Discriminant Analysis (ACDA) provides a solution to this problem. ACDA performs approximation in the PCA subspace rather than re-computing the subspace for each pair of gallery and probe sequences. The experimental results suggest that this approximation is accurate in practice and the gain in speed is much greater than the loss in recognition accuracy.

4. It is noted that the proposed method did not achieve the same amount of improvement for the SOTON carrying-bag sequences (SOTONSmallSetB2) compared with other probe sets (see Table 4.1). The reason is that for these sequences subjects often carry suitcases which occlude the dynamic areas around the leg region. Feature selection method can remove the static areas caused by the suitcases; but it will also remove some of the informative dynamic areas. Under this circumstance and with subject cooperation, the existing approach without feature selection may perform better as they utilise some of the shape information in the static areas located in the upper body region. However, because the objective is to develop a human identification method that is invariant to covariate condition changes and without relying on cooperative subjects, this is not considered as a drawback of the approach. Results in Tables 4.3 clearly show that without subject cooperation, the proposed approach performs much better than the alternatives even for the SOTON carrying-bag sequences.

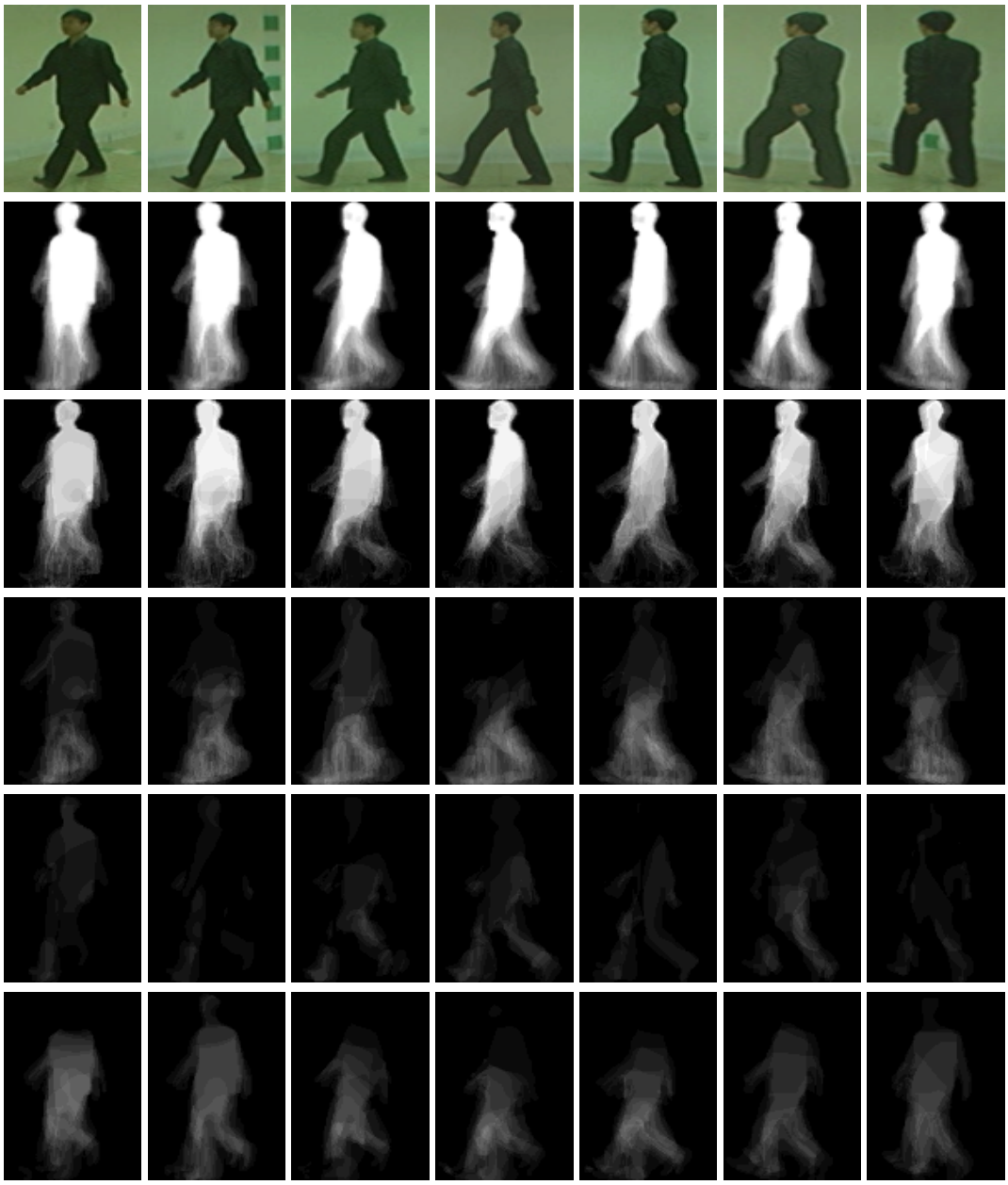
5. There exist a number of methods [2,60] which also learn a PCA space adaptively. However, there is a fundamental different between these methods and ACDA method. Specifically, the existing adaptive PCA methods are designed for incremental updating of PCA space when a new data point becomes available. Consequently, the updated PCA space is fairly similar to the old one as it is learned mostly based on the unchanged old data. In this problem, given a new probe sequence, it does not introduce a new data point. Instead, it changes the whole dataset for learning PCA as a new feature selection mask is generated. Therefore the existing adaptive PCA methods are not suitable here because it is designed for a completely different problem.

Chapter 5

Cross View Gait Recognition

Among various covariate conditions that can affect the performance of gait recognition, changes in viewpoint pose the biggest problem. This is because as the view angle changes a different set of features are exposed. These features tend to be completely different from the previous ones if the change in view angle is large. This change in feature space affects the representation of gait (see Fig. 5.1) which eventually degrades recognition performance.

Whilst a treatment of various covariate conditions affecting performance of gait recognition algorithms is given in the previous chapters, this chapter builds on the previous work to address the cross view gait recognition problem. To achieve robustness under variable covariate conditions in a cross view setting, Gait Flow Image (GFI) is used as a representation of gait. A novel approach to cross view gait recognition is developed with the view angle of a probe gait sequence unknown. The proposed approach is a two step process: (1) A Gaussian Process (GP) classification framework is formulated to estimate the view angle of each probe gait sequence (2) To measure the similarity of gait sequences at different view angles, the correlation of gait sequences from different views is modelled using Canonical Correlation Analysis (CCA) and correlation strength is used as a similarity measure. This procedure enables both view estimation and gait recognition in an arbitrary view and differs significantly from existing approaches, which reconstruct gait features in different views either through 2D view transformation [56, 57, 79] or 3D calibration [11, 91, 122]. Without explicit reconstruction, this approach can cope with feature miss-match across view and is more robust against feature noise. It is validated by the experiments that the proposed method significantly outperforms existing state-of-the-art methods.



(a) 36° (b) 54° (c) 72° (d) 90° (e) 108° (f) 126° (g) 144°

Figure 5.1: The affects of view angle on gait representations. Top row: the extracts from frames of video sequences under different view angles. Second row: the GEIs. Third row: the Motion Intensity Image (MII) from Gait Flow Image and the Motion Direction Images (MDIs) ($\mathbf{M}_x^+$, $\mathbf{M}_x^-$ and $\mathbf{M}_y^+$) from Gait Flow Image representation in subsequent rows.

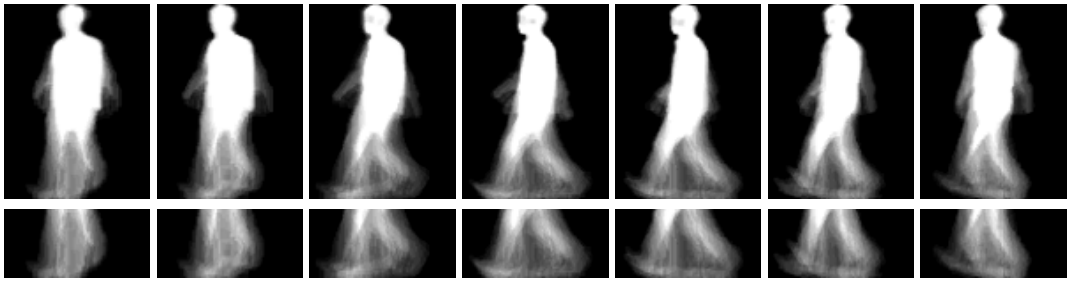
To learn the cross view correlation model it is assumed that a multi-view gait training dataset is available in which gait sequences of subjects are acquired in all views. In addition, the subjects appearing in the training dataset are independent from the subjects in the test dataset. This means that the system can be independently trained on an available dataset and then be tested on a different dataset.

The availability of a multi-view training dataset is a strong assumption and imposes restrictions on the applicability of the proposed approach to situations where such a training set can be captured. However, this provides the possibility of dealing with large changes in view angle between the probe and the gallery set as has been demonstrated in this thesis and existing methods based on view transformation model [56, 57, 79]. Approaches not requiring a multi-view training dataset in general are unable to deal with large changes in view angle and are only applicable when the change in view angle between the probe and gallery is small [12, 36, 38, 47, 49, 50].

5.1 Cross View Gait Representation

Given a human walking sequence, a human silhouette is extracted from each frame using the method of [88]. After applying size normalization and horizontal alignment to each extracted silhouette image, gait cycles are segmented by using the method in [56]. Two gait representations are then computed, one for view angle recognition and the other for cross view gait recognition. The use of separate representations for view angle estimation and recognition will become clear in the next section.

5.1.1 Truncated Gait Energy Image



(a) 36° (b) 54° (c) 72° (d) 90° (e) 108° (f) 126° (g) 144°

Figure 5.2: Top row: GEIs of the same subject for different views. Bottom row: TGEIs obtained from the GEIs in the top row.

For view recognition, gait sequences are represented using Truncated Gait Energy Images (TGEI). TGEI is simply Gait Energy Image (GEI) [37] without its top part (head & torso) and is generated by only taking the bottom one third of the GEI (see Fig. 5.2). In this work TGEI is used to learn the Gaussian Process classifier for view angle recognition. The advantage of using TGEI instead of GEI for view recognition becomes clear on a closer look at Fig. 5.2. It can be seen that the torso part of the GEI (which constitutes a major portion of it) for view angle 144° and 54° (Fig. 5.2(b) and (g)) is almost identical. This would make the classification process prone to

errors. The same can be said for view angles 72° and 126° (Fig. 5.2(c) and (f)). In contrast, from the bottom row of Fig. 5.2 it is seen that the TGEI for view angles 144° and 54° are completely different and also there are visible changes in TGEI for view angles 72° and 126°. Thus the TGEI is selected as a feature to learn the view classifier.

5.1.2 Gait Flow Image

Gait Flow Images (GFI) described in Chapter 3 are used as a gait feature for cross view gait recognition. GFIs provide more discriminative representation for identity recognition compared to GEI by looking at multiple independent motion of different body parts during a gait cycle and separating them into different descriptors. It is robust against variable covariate conditions and outperforms state-of-the-art approaches on the largest available public gait datasets as demonstrated in Chapter 3. Gait information in GFIs is captured in a set of motion descriptors including a motion intensity descriptor (representing shape information) $\mathbf{M}$, and 3 motion direction descriptor $\mathbf{M}_{\mathbf{x}}^{+}, \mathbf{M}_{\mathbf{x}}^{-}, \mathbf{M}_{\mathbf{y}}^{+}$ (representing motion information) corresponding to the right, left and up directions. Example of GFIs for a single subject from different viewing angles under normal walking conditions are shown in the bottom four rows of Fig. 5.1.

5.1.3 Separate Representations for View Classification and Recognition

In this thesis different representations are used for view classification and gait recognition. The reason for this becomes clear by comparing the TGEIs in Fig. 5.2 and GFIs in Fig. 5.1. In particular, for view recognition, which can be seen as an inter-class classification problem compared with the intra-class one in gait recognition, it is necessary to remove the top two third as most of the information there is invariant to view change (e.g. large part of torso). However, there are useful gait features that contribute towards better recognition performance of gait [21,59,99] although this information is largely concerned with appearance not kinematic or motion aspect of human gait. GEI or the TGEI are not used at the recognition stage because of the advantages of GFI over these representations. GFI is robust to variable covariate conditions as demonstrated in Chapter 3 and hence is a choice of representation for cross view gait recognition. GFI is not used for view angle recognition because of a similar argument as the GEI (see Sec. 5.1.1). In addition, since GFI captures much inter-subject variations, it is good for gait recognition but causes problem for view classification because these variations now become intra-class variations.

5.2 Learning a Gait View Classifier Using Gaussian Processes

Gaussian Processes (GP) have been used in regression [105] and classification problems [104]. These models are flexible as no selection of model complexity is required (as compared to e.g. Gaussian Mixture Models). GP are closely related to Support Vector Machines (SVM). A key advantage of GP compared to SVM is that they are probabilistic models that allow to incorporate prior information about data distribution. This often results in more robust and better models, a fact that is also reflected in the results of experiments in this chapter, where the GP classifier outperforms SVM.

Truncated Gait Energy Images are calculated from the available multi-view training data. But before they can be used to train the GP classifier, dimensionality reduction is required in order to make the computations feasible. Principal Component Analysis (PCA) is used for this purpose. Let $\{\mathbf{x_1},...,\mathbf{x_N}\}$ be $\tilde{d} < d$ dimensional component space representation of N, d-dimensional TGEI templates belonging to C views. The GP Classifier is trained using the reduced dimensional $\mathbf{x_i}$ where $i = 1,...,N$. The classification model uses a latent function $\mathbf{f}$ which is never observed. What is observed though are the $\mathbf{x_i}$ values and the class labels y. $\mathbf{y}$ is defined as a vector of the same length as the latent function $\mathbf{f}$ which for each $i = 1,...,N$ has an entry of 1 for the class which is the label for example i and 0 for other $C-1$ entries. The vector of latent function values at all N training points and C class labels is given by $\mathbf{f} = (f_1^1,...,f_N^1,f_1^2,...,f_N^2,...,f_1^C,...,f_N^C)^T$. The prior over $\mathbf{f}$ has the form $\mathbf{f} \sim \mathcal{N}(\mathbf{0},K)$ where K is the matrix of covariance function values. The covariance function expresses the correlation between the data values. The squared exponential covariance function is selected to encode the prior knowledge. This means that points lying close together are closely correlated or in other words gait features that resemble each other (belonging to same view) are highly correlated. The squared exponential covariance function is given as

$$K(x,x\prime) = \sigma^2 exp\left(-\frac{1}{2}(x-x\prime)^T \sum (x-x\prime)\right) \tag{5.1}$$

where $x\prime$ is the mean, σ defines the magnitude and $\sum = l^{-2}I$, l are the characteristic length scales (learn-able hyper parameters of the covariance function) and are associated with the relative importance of different inputs to prediction. The posterior over the training data defined by $p(\mathbf{f}|X,y)$ is not analytically tractable and a Laplace approximation is used instead. More details can be found in [104]. A test point x_* represented in the component space is then classified by computing the predictive distribution of $\mathbf{f}_* = \mathbf{f}(x_*) = (f_*^1,...,f_*^C)^T$ defined by $q(\mathbf{f}_*|X,\mathbf{x}_*,\mathbf{f})$ [104],

where X is the matrix of training data.

The learned GP gait view classifier is expected to make errors which will potentially introduce model mis-match in gait sequence correlation to be described in the next section.

To minimise such error propagation, instead of directly using the top label returned from the GP pose classifier, a soft decision is made and the top two candidates are considered. This is because the returned labels are Rank 2 correct at least 98% on the test dataset (see experiment section). Therefore, a weighted approach is used and the scores returned from the top two class labels suggested by the classifier are fused together. The weights are calculated by normalizing the confidence of the classifier in the top two labels.

Note that as a person moves in space, the view angle with respect to the camera is likely to change continuously. This could mean that two cycles from a gait sequence may be classified into two different views. The GP classifier caters for this and works on per cycle basis (i.e., a separate TGEI is generated for each cycle).

5.3 Learning Cross View Correlation Model

Canonical Correlation Analysis (CCA) [41] is used to perform recognition across views by using the correlation strength as a measure of similarity. CCA is a linear method[1] and measures the relationship between two sets of multidimensional variables . CCA finds two bases one for each set of variable in such a way that the two sets of variables are maximally correlated.

To learn the cross view correlation model using CCA the GFI are computed for the multi-view training dataset. PCA is used to reduce the dimensionality of the GFI descriptors to make the computations feasible. Since the GFI are composed of more than one descriptor, in the following learning process is described for one of the gait flow descriptors $\mathbf{M}$. A similar method is used for the remaining descriptors $\mathbf{M}_{\mathbf{x}}^{+}, \mathbf{M}_{\mathbf{x}}^{-}, \mathbf{M}_{\mathbf{y}}^{+}$. Let $X = \{\mathbf{x}_1, ..., \mathbf{x}_N\}$ be matrix of $\tilde{d} < d$ dimensional component space representation of N d-dimensional $\mathbf{M}$ templates in view V_{θ_x} and $Y = \{\mathbf{y}_1, ..., \mathbf{y}_N\}$ be the component space representation in view V_{θ_y}. Let the linear combinations of canonical variables be $x = \mathbf{w}_x{}^T\mathbf{x}$ and $y = \mathbf{w}_y{}^T\mathbf{y}$. CCA can now be defined as

$$\rho = \frac{E[xy]}{\sqrt{E[x^2]E[y^2]}} = \frac{E[\mathbf{w}_x{}^T\mathbf{x}\mathbf{y}^T\mathbf{w}_y]}{\sqrt{E[\mathbf{w}_x{}^T\mathbf{x}\mathbf{x}^T\mathbf{w}_x]E[\mathbf{w}_y{}^T\mathbf{y}\mathbf{y}^T\mathbf{w}_y]}} \tag{5.2}$$

[1] A kernel version of CCA has also been tested and the results are presented in the experiments section. Kernel CCA is found to be inferior in performance, which suggests the linear assumption is valid.

which can also be written in terms of covariance matrices as

$$\rho = \frac{\mathbf{w}_x{}^T C_{\mathbf{xy}} \mathbf{w}_y}{\sqrt{\mathbf{w}_x{}^T C_{\mathbf{xx}} \mathbf{w}_x \mathbf{w}_y{}^T C_{\mathbf{yy}} \mathbf{w}_y}} \tag{5.3}$$

where $C_{\mathbf{xx}}, C_{\mathbf{yy}}$ are the within sets and $C_{\mathbf{xy}}$ the between sets covariance matrices. CCA maximizes ρ by solving for the derivative of Eqn. 5.3 and setting it to zero. This yields the following eigenvalue equations.

$$C_{\mathbf{xx}}^{-1} C_{\mathbf{xy}} C_{\mathbf{yy}}^{-1} C_{\mathbf{yx}} \mathbf{w}_x = \rho^2 \mathbf{w}_x \tag{5.4}$$
$$C_{\mathbf{yy}}^{-1} C_{\mathbf{yx}} C_{\mathbf{xx}}^{-1} C_{\mathbf{xy}} \mathbf{w}_y = \rho^2 \mathbf{w}_x \tag{5.5}$$

where eigenvalues ρ^2 are the square canonical correlations and $\mathbf{w}_x$ and $\mathbf{w}_y$ are the basis vectors.

Once $\mathbf{w}_x$ and $\mathbf{w}_y$ have been computed for V_{θ_x} and V_{θ_y} this is done for all the view combinations in the multi-view training dataset in order to complete the learning process. Cross view gait recognition using GP classification and CCA correlation strengths can now be performed.

5.4 Cross View Gait Recognition

To perform recognition across view gait templates of subjects in view V_{θ_g} are used as gallery data. Any probe sequence in an arbitrary view V_{θ_p} and variable covariate conditions for the subjects in the gallery can now be recognized. The essence is to use the correct model to compute correlation strengths for matching across views. This is done by using the learned GP classifier to identify the view angle of the probe sequence. Based on the output from the classifier the corresponding CCA models are then used for computing correlation strengths for matching across views.

Specifically, TGEI and GFI templates for the probe sequence are computed. GP classification is then applied using TGEI resulting in the predicted top two ranked class labels and the confidence in each. Let the top two views identified by the classifier be V_{θ_1} and V_{θ_2} with confidence ω_{θ_1} and ω_{θ_2} which are normalized so that they sum to 1.

After view classification the trained CCA models for $V_{\theta_1} \rightarrow V_{\theta_g}$ and $V_{\theta_2} \rightarrow V_{\theta_g}$ are used to compute correlation strength scores between the probe template and a gallery one. Since the GFI comprise of four descriptors for each gait cycle, this process is described for one of the descriptors, M. The same procedure can be applied to the remaining descriptors. Correlation strength for the two templates are computed using Eqn. 5.3 and are given as $\rho^{iM}_{\theta_1\theta_g}$ and $\rho^{iM}_{\theta_2\theta_g}$ where

$i = 1, ..., n$ and n is the number of templates in the gallery view V_{θ_g}. The correlation strength for M is then the weighted average of the correlation strength of the two models weighted by the normalized confidence scores and is given as

$$\rho^{iM} = \omega_{\theta_1} \rho^{iM}_{\theta_1 \theta_g} + \omega_{\theta_2} \rho^{iM}_{\theta_2 \theta_g} \tag{5.6}$$

Similarly correlation strength for other descriptors are obtained, the final score is then computed as follows

$$\rho^{i} = \rho^{iM} + \rho^{iM^{+}_{x}} + \rho^{iM^{-}_{x}} + \rho^{iM^{+}_{y}} \tag{5.7}$$

The gallery sequence with the largest ρ^i is then identified as the correct match. Since a test sequence can have multiple gait cycles generating multiple templates in this case a simple voting mechanism is used to generate the output label. In case of equal votes the output label with the larger correlation is selected.

5.5 Experiments

5.5.1 Dataset

The CASIA Gait Database [116] described in Chapter 3 is used for evaluating the performance of the proposed approach, which is the largest publicly available multi-view gait dataset. All the sequences are captured from 11 different views starting from 0° with 18° offset resulting in gait view sequences for 0° the front view, $18°, 36°, 54°, 72°$, 90° the front-to-parallel view, $108°, 126°, 144°, 162°$ and 180° the back view. Example frames from the CASIA dataset extracted from multiple views for the same subject are shown in Fig. 5.3.

Gait sequences from 7 views from $36° - 144°$ are used in the multi-view experiments for the reason that view angles close to the frontal and back views provide little gait information. The multi-view training set consists of 4 out of 6 normal sequences of 60% of the total subjects in the dataset from views $36° - 144°$. The remaining 40% of the subjects are used for testing, i.e. the subjects used in training and testing are mutually exclusive. For testing a gallery set and a probe set are formed. The gallery set consists of the first 4 normal sequences from views $36° - 144°$. The probe set include the last 2 normal sequences (Set A2), the 2 carrying bag sequences (Set B) and 2 wearing-coat sequences (Set C), all from views $36° - 144°$.

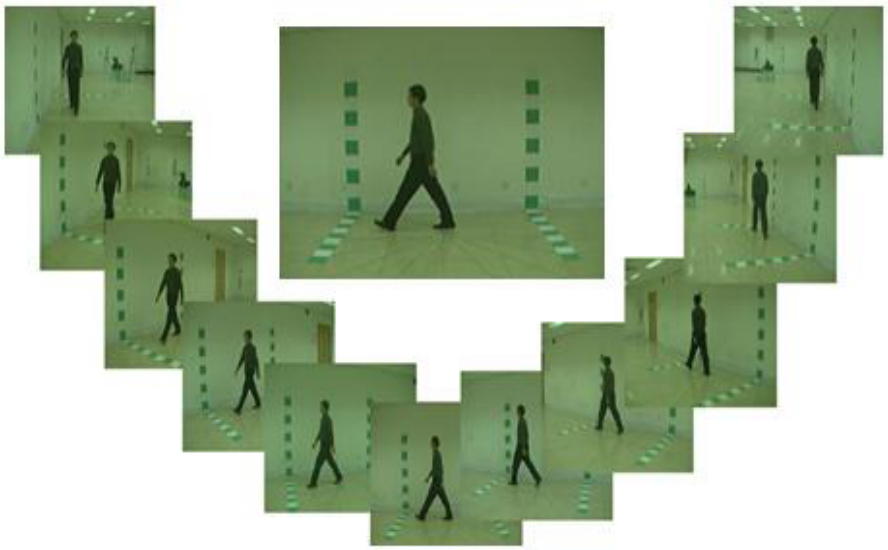


Figure 5.3: Sample Images from the multi-view CASIA dataset (from [116]).

5.5.2 View Classification with GP Classifier

GP classification is compared with Support Vector Machines (SVM) for view classification. It can be seen from Table 5.1 that GP Classification gives better results than SVM and also achieves satisfactory performance for the dataset over all covariate conditions and across all different views consistently. In contrast, although good result is obtained for some views, SVM achieved very poor results for others (e.g. 54° and 108°) even using identical feature representation as GP. Table 5.2 presents Rank 2 results for the classification experiment. From Table 5.2 it can be seen that the correct view is Rank 2 correct almost all of the time for GP classification whereas SVM lags in terms of performance and again is inconsistent across different views. The excellent Rank 2 results lead to the development of algorithm described in Sec. 5.4 in using the confidence score from the top two results returned from the GP Classification algorithm.

Probe Set	36°	54°	72°	90°	108°	126°	144°
Set A2 (GP)	84.0%	91.2%	85.3%	74.0%	86.0%	91.2%	93.5%
Set A2 (SVM)	94.9%	40.5%	85.4%	64.3%	24.0%	43.6%	98.0%
Set B (GP)	83.4%	88.7%	84.9%	68.6%	83.0%	92.7%	93.5%
Set B (SVM)	96.1%	41.8%	79.3%	62.6%	28.1%	50.6%	97.9%
Set C (GP)	84.0%	91.2%	85.3%	74.0%	86.0%	91.2%	93.5%
Set C (SVM)	93.7%	50.0%	81.0%	61.2%	22.5%	41.5%	96.6%

Table 5.1: Rank 1 view classification results for GP and SVM.

Probe Set	36°	54°	72°	90°	108°	126°	144°
Set A2 (GP)	98.0%	99.0%	98.8%	99.0%	99.0%	99.0%	99.0%
Set A2 (SVM)	98.7%	78.4%	92.4%	90.0%	82.9%	71.5%	100%
Set B (GP)	99.5%	99.5%	99.3%	98.8%	99.3%	99.2%	99.3%
Set B (SVM)	98.1%	79.0%	88.7%	88.6%	75.6%	79.2%	99.3%
Set C (GP)	98.1%	98.6%	99.5%	98.3%	98.6%	99.0%	97.9%
Set C (SVM)	97.5%	83.7%	86.0%	86.4%	66.2%	62.9%	97.9%

Table 5.2: Rank 2 view classification results for GP and SVM.

5.5.3 Cross View Gait Recognition using Correlation Strength

After view classification using GP Classification, correlation strength from CCA is used to perform gait recognition across multiple views. The results are reported in terms of recognition rate for the proposed method with GFI (GFI+CCA), the proposed method with known probe angle (GFI+CCA(P)), the proposed method with Gait Energy Image for gait representation (GEI+CCA). The proposed method is compared with state-of-the-art methods (FG+SVD) [79], (GEI+TSVD) [56] and the baseline [116] which simply matches GEI across views without any view transformation. Note that the results for gait energy image (GEI+CCA) using the proposed approach are also listed in order to provide a direct comparison with alternative approaches which also use GEI for gait representation. The results obtained using GEI and GFI also give insight on their ability to cope with different covariate conditions such as carrying and clothing.

Complete results under normal conditions are given in Table 5.5 under CCA and a comparison with state-of-art is shown in Fig. 5.4. It can be seen from Fig. 5.4 that the proposed method significantly outperforms existing methods over all views. It is interesting to note that the performance of FG+SVD [79] and GEI+TSVD [56] in some cases even falls below the Baseline [116] which does not compensate for view change at all. In contrast the proposed approach beats the baseline approach comfortably. It is also worth pointing out that all alternative methods assume the view angles are known a priori and the proposed approach does not make this assumption and recognises the view angle automatically. Fig. 5.4 also shows that when probe view angle is known (GFI+CCA(P)), the result is almost identical to the one with GP classification (GFI+CCA) which emphasizes that GP classification algorithm based on Rank 2 results effectively classifies unknown probe sequences. Comparing GFI+CCA with GEI+CCA it is clear that GFI is a better gait representation and justifies itself as the choice for representation of gait for doing robust cross view gait recognition under variable covariate conditions. This result is consistent with those reported in Chapter 3 for gait recognition under the same view angle.

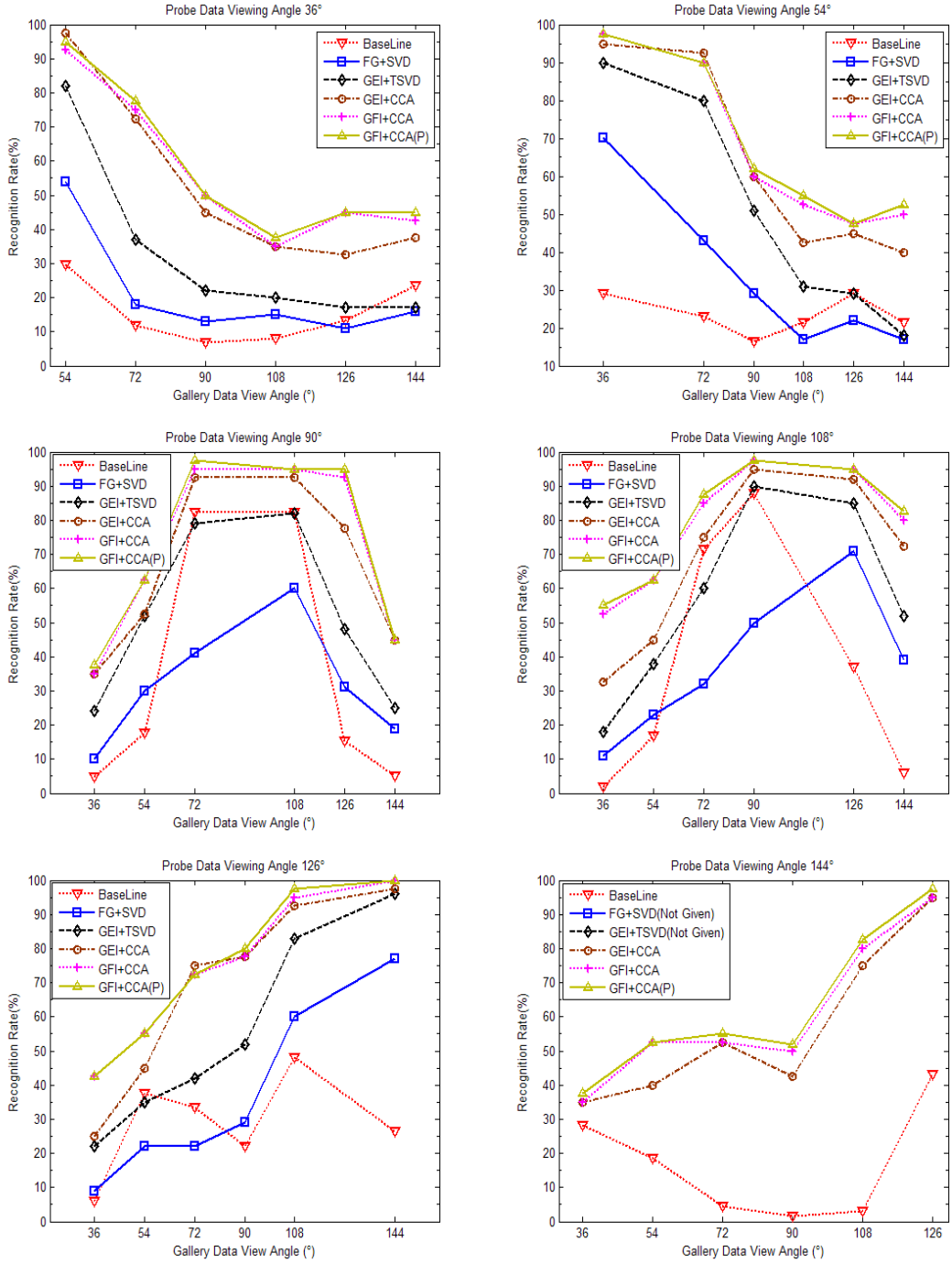


Figure 5.4: Comparison of cross view gait recognition performance under normal conditions. (For 144° the results of FG+SVD [79] and GEI+TSVD [56] are not available).

The performance of the proposed approach is also tested under variable covariate conditions present in the dataset i.e. the clothing and the carrying conditions. The results of the proposed method Gait Flow Image and correlation strength (GFI+CCA) are compared with Baseline [116] and with the proposed method using Gait Energy Image (GEI+CCA) only as [56, 79] did not present the results for these experiments. Complete results for the carrying and clothing condi-

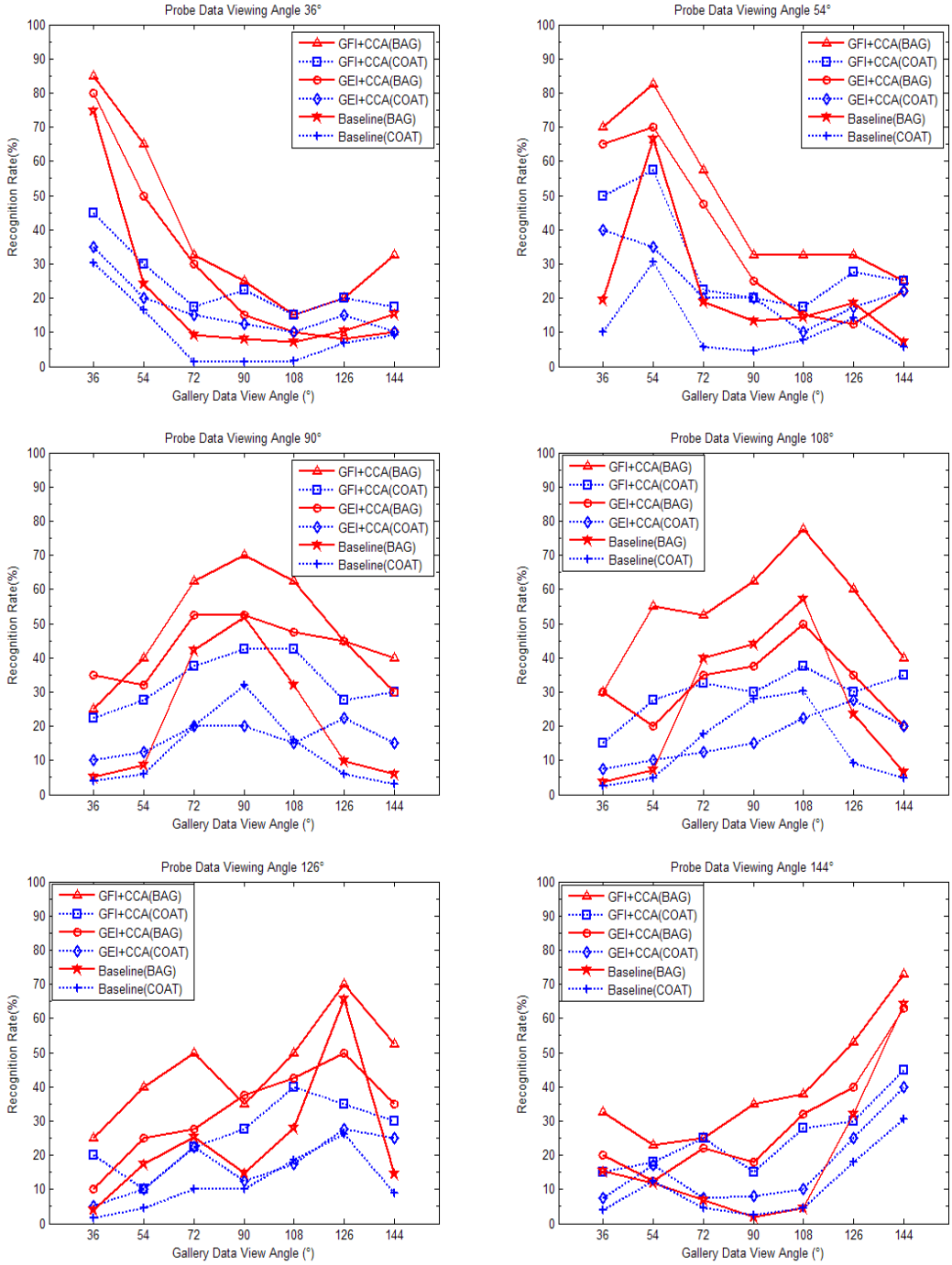


Figure 5.5: Comparison of cross view gait recognition performance under changing carrying and clothing conditions.

tions are given in Table 5.3 and Table 5.4 respectively and a comparison with other methods is shown in Fig. 5.5. Again, the proposed approach consistently outperforms the baseline methods. Fig. 5.5 also shows that the advantage of GFI over GEI becomes more apparent under changing carrying and clothing conditions as GFI is specifically designed for coping with different covariate conditions.

Probe Set	36°	54°	72°	90°	108°	126°	144°
36°	85.0%	65.0%	32.5%	25.0%	15.0%	20.0%	32.5%
54°	70.0%	82.5%	57.5%	32.5%	32.5%	32.5%	25.0%
72°	55.0%	67.5%	77.5%	52.5%	40.0%	42.5%	32.5%
90°	25.0%	40.0%	62.5%	70.0%	62.5%	45.0%	40.0%
108°	30.0%	55.0%	52.5%	62.5%	77.5%	60.0%	40.0%
126°	25.0%	40.0%	50.0%	35.0%	50.0%	70.0%	52.5%
144°	30.5%	35.0%	40.0%	32.5%	47.5%	62.5%	77.5%

Table 5.3: Cross view gait recognition under different carrying condition.

Probe Set	36°	54°	72°	90°	108°	126°	144°
36°	45.0%	30.0%	17.5%	22.5%	15.0%	20.0%	17.5%
54°	50.0%	57.5%	22.5%	20.0%	17.5%	27.5%	25.0%
72°	30.0%	27.5%	55.0%	37.5%	20.0%	22.5%	22.5%
90°	22.5%	27.5%	37.5%	42.5%	42.5%	27.5%	30.0%
108°	15.0%	27.5%	32.5%	30.0%	37.5%	30.0%	35.0%
126°	20.0%	10.0%	22.5%	27.5%	40.0%	35.0%	30.0%
144°	22.5%	15.0%	17.5%	25.0%	35.0%	32.5%	37.5%

Table 5.4: Cross view gait recognition under different clothing condition.

5.5.4 Comparison of CCA and Kernel CCA

In order to investigate the performance of CCA and Kernel CCA an experiment is constructed using the normal walking sequences of different views from the CASIA dataset represented using the Gait Flow Image. Cross view gait recognition is performed using CCA and Kernel CCA and the results are presented in Table 5.5. It is clear from Table 5.5 that CCA significantly outperforms Kernel CCA over all view combinations.

Probe	36°		54°		72°		90°		108°		126°		144°	
	CCA	KCCA	CCA	KCCA	CCA	KCCA	CCA	KCCA	CCA	KCCA	CCA	KCCA	CCA	KCCA
36°	100%	100%	92.5%	65.0%	75.0%	37.5%	50.0%	27.5%	35.0%	12.5%	45.0%	17.5%	42.5%	10.0%
54°	97.5%	72.5%	100%	100%	90.0%	65.0%	60.0%	35.0%	52.5%	20.0%	47.5%	20.0%	50.0%	17.5%
72°	77.5%	27.5%	95.0%	55.0%	100%	100%	100%	75.0%	95.0%	27.5%	80.0%	37.5%	55.0%	32.5%
90°	35.0%	25.0%	62.5%	32.5%	95.0%	82.5%	100%	97.5%	95.0%	77.5%	92.5%	55.0%	45.0%	27.5%
108°	52.5%	22.5%	62.5%	22.5%	85.0%	42.5%	97.5%	72.5%	100%	95.0%	95.0%	82.5%	80.0%	55.0%
126°	42.5%	20.0%	55.0%	22.5%	72.5%	40.0%	77.5%	57.5%	95.0%	87.5%	100%	100%	100%	85.0%
144°	40.0%	10.0%	52.5%	22.5%	52.5%	22.5%	47.5%	22.5%	80.0%	50.0%	95.0%	87.5%	100%	97.5%

Table 5.5: Comparison of Canonical Correlation Analysis(CCA) and Kernel Canonical Correlation Analysis(KCCA) for different view combinations using the GFI representation.

5.6 Discussions

The key findings of experiments are summarised and discussed as below:

1. Cross view gait recognition using correlation strength significantly outperforms state-of-the-art methods over all view combinations and also is effective when the view angle dif-

ference is large. The main reason for this improvement in performance is that the proposed method based on correlation strength does not transform representations across views. Transformation across views suffer from problems of features being visible in one view and not in an other and thus causes a many to one mapping, which leads to performance degradation.

2. View angle classification using Gaussian Process classification is extremely effective. The rank one result is promising but the rank two result is almost perfect. This leads to the development of a soft decision method using contributions from the scores of the top two candidate views weighted by the confidence of the classifier in that view. The method has shown to be pretty similar in performance as compared to when the probe view angle is known (see Fig. 5.4).

3. Comparison of performance of the proposed approach using correlation strength and simple GEI with state-of-the-art methods which use modified form of GEI as representation [56, 79] shows that the proposed approach outperforms existing methods even with simple GEI (see Fig. 5.4). This highlights the usefulness of the proposed approach as compared to transformation based methods which have been shown to be inferior in performance.

4. When there is no view change between the probe and the gallery view the results show (diagonal entries in Tables 5.5, 5.3 and 5.4) that Canonical Correlation Analysis performs very well as a recognition algorithm. The performance is comparable to that obtained using component and discriminant analysis used earlier (see Table 3.1, Chapter 3). This suggests that Canonical Correlation Analysis can also be used when there is no view change as a recognition algorithm to perform gait recognition.

5. The diagonal entries in all the three Tables 5.5, 5.3 and 5.4 present the results for gait sequences in same views and report high recognition rates as expected. It is noted though that the best recognition performance for gait under covariate condition changes (carrying and clothing) are not obtained in the front-to-parallel view rather they are found at views close to the front or the back view. This is because appearance changes such as carrying and clothing have the biggest affect in the front-to-parallel view and this affect is reduced as the view angle deviates away from it.

Chapter 6

Conclusion and Future Work

This thesis has set out to explore the problems caused by variable covariate conditions in gait recognition. In the presence of variable covariate conditions matching a probe gait sequence to a gallery sequence which may be captured under different conditions is non-trivial. The problem is even harder if the covariate conditions are completely unknown and allowed to be variable in the gallery set as well as the probe set. The problem of variable covariate conditions is further compounded if there are significant view changes between the gallery and the probe set.

The underlying motivation of this thesis is the performance enhancement of existing methods under variable covariate conditions across large view angle changes. Applications of this are in medium to long distance security and surveillance where direct contact and cooperation is not feasible or not desired. Further application areas include multi-modal biometric systems (e.g. face and gait) and human tagging across multiple cameras where gait can be used on its own or as a cue to enhance the performance of existing methods.

There is a considerable scope in performance improvement of gait under variable covariate conditions. For this purpose the key areas explored in this thesis are a more discriminative representation of gait robust to variable covariate conditions, feature selection for gait recognition under variable and unknown covariate conditions and cross view gait recognition. The investigations of each of these are summarized below.

6.1 A More Discriminative Representation of Gait

In this thesis a novel gait representation the Gait Flow Image (GFI), based on optical flow fields computed from normalized and centred person images over a complete gait cycle is proposed. In this representation, both the motion intensity and the motion direction information is captured in a set of motion descriptors. The formulated Motion Intensity Image (MII) and Motion Direction Images (MDIs) have different levels of sensitivity toward different types of covariate conditions. The fusion of them, thus gives a gait representation that is not only more discriminative, but also less sensitive to changes in various covariate conditions. In summary, the main advantages of Gait Flow Image representation are: 1) Capturing more discriminative information from a gait sequence by considering independent movement of body parts 2) The separation of appearance and motion content into different descriptors. Extensive experiments carried out on both indoor and outdoor public datasets demonstrate that Gait Flow Image representation outperforms state-of-the-art.

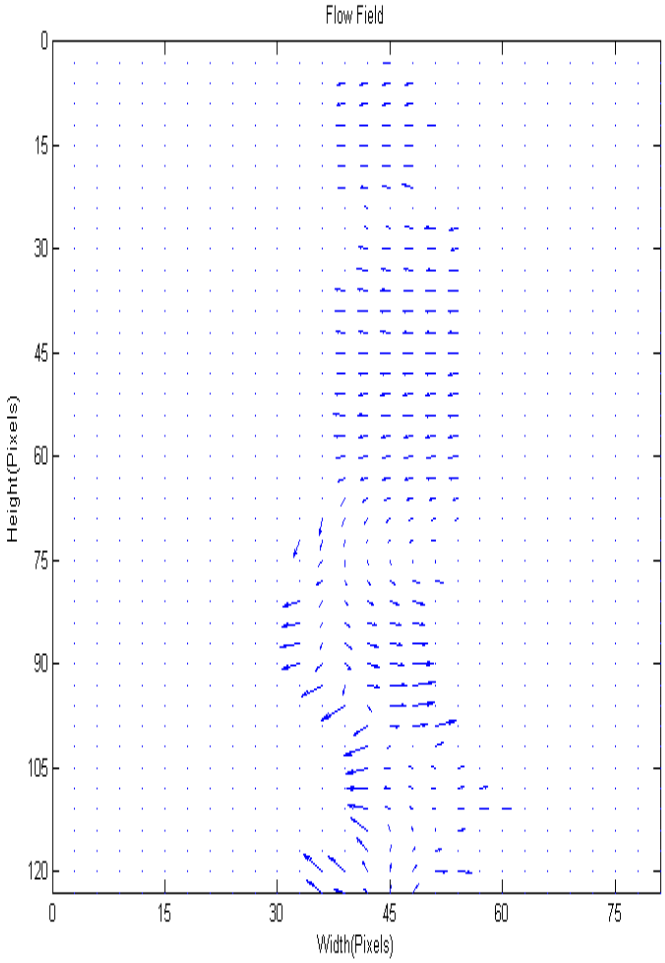
6.1.1 Future Work

The main problem of Gait Flow Image is the self occlusion at the mid-stance phase during a gait cycle as the legs cross over which generates errors in the flow field. Fig. 6.1(a,b and c) show the normalized stances at the mid-stance along with the flow field in Fig. 6.1(d). It can be seen from Fig. 6.1(d) that the self occlusion at the mid-stance produces errors in optical flow field. Specifically, as the back leg in Fig. 6.1(b) starts appearing behind the front leg in Fig. 6.1(c) the optical flow field computations become an ill-posed problem.

Although an accumulated histogram based method is used in the proposed approach to reduce the affects of noise and errors in the flow field. The error robustness of Gait Flow Image can be improved further by using an HMM based algorithm as proposed by Liu and Sarkar [72] to normalize gait dynamics. Normalizing the gait stances and eliminating the contributions from the mid-stance can potentially further improve the performance of Gait Flow Image representation.

Alternatively, information maximization criterion proposed by Xiang and Gong [106] can also be used to to identify frames at the mid-stance and remove them from the final computation of representation. This is because by computing the information content between successive frames the mid-stance phase can be identified automatically because of lesser variations (less information) between successive frames at the mid-stance as compared to the stances near the

(a) Silhouette (b) $\mathbf{I}_{t-1}$ (c) $\mathbf{I}_t$



(d) $\mathbf{F}_t$

Figure 6.1: Figure-centric optical flow field at the mid-stance. (a) shows the silhouette at frame $t-1$; (b)&(c) are the extracted figure-centric images of the walking person at $t-1$ and t respectively; (d) is the computed optical flow field.

full stride.

6.2 Gait Recognition Under Variable and Unknown Covariate Conditions

Gait is most suitable for recognition in unconstrained environments, examples of these are security and surveillance applications. This has been the main driving force in gait research. However, none of the contemporary methods put this to practice in a true sense. Specifically, the gallery set used by existing approaches consists of gait sequences captured under the similar and known, normal walking conditions. This in itself is a constraint and assumes cooperation. In a truly uncooperative set-up the gallery sequences should be captured on the fly with people appearing under variable covariate conditions.

In this thesis performance of state-of-the-art gait recognition approaches is investigated in a an experimental set-up with variable and unknown covariate conditions. The experimental results suggest that the existing approaches are unable to cope with changes in variable covariate conditions in the gallery set. Existing methods therefore, are unsuitable for challenges presented by gait recognition without subject cooperation. To overcome these problems, a novel gait recognition approach is proposed, which performs adaptive feature selection on each pair of gallery and probe gait sequences using Gait Entropy Image (GEnI), and seamlessly integrates feature selection into the subspace with an Adaptive Component and Discriminant Analysis (ACDA) for fast recognition. Experiments carried out on both the standard set-up (Gallery: normal conditions, Probe: variable conditions) and the proposed uncooperative set-up indicate that the proposed framework is able to cope effectively with variable covariate conditions changes in the probe set and outperforms existing methods. In addition, adaptive feature selection shows a vast improvement over existing methods for gait recognition without subject cooperation.

6.2.1 Future Work

The proposed adaptive feature selection method presents a generalized feature selection method to select features based on the probe and gallery sequence at hand without taking into account the covariate conditions present in the underlying sequences. A viable direction for future work is to tailor the feature selection methods based on the specific covariate conditions present in the gait sequences. The use of feature selection to address specific covariate conditions has been demonstrated by Hossain *et al*. [40] and Lee *et al*. [61] to deal with clothing and backpack conditions respectively. A similar strategy can be incorporated in the adaptive feature selection framework to select features based on the covariate conditions present in the probe and gallery

set in order to improve recognition performance. However, since the covariate conditions are unknown the recognition of the covariate conditions itself is critical and needs to be further investigated.

6.3 Cross View Gait Recognition

In this thesis a novel cross view gait recognition framework is presented to address the limitations of existing view transformation based approaches by using correlation strength as a measure of similarity across views. The cross view recognition framework uses Canonical Correlation Analysis to project the data into intermediate subspaces in such a way that the correlation between them is maximized. View angle which is an important ingredient of a cross view recognition framework and assumed to be known by existing methods is hereby estimated using Gaussian Process classification and integrated into the cross view recognition framework using a soft decision approach.

Results of the experiments carried out on the largest multi-view gait dataset demonstrate that the proposed cross view gait recognition frame work significantly outperforms view transformation methods even when the view angle is considered unknown and estimated using Gaussian Process classification. It is also found that the use of soft decision at the Gaussian Process classifier stage only shows minor performance degradation as compared to cases where the view angle is considered known before hand.

6.3.1 Future Work

The cross view gait recognition framework based on Canonical Correlation Analysis and Gaussian Process classification promises to improve the performance under variable conditions with no additional information using the whole Gait Flow Image representation. However, instead of using the representation as a whole, a part based methodology can further improve cross view gait recognition performance as demonstrated by Kusakunniran *et al*. [57]. It has been shown that using a Region of Interest and applying Support Vector Regression for view transformation only to those regions, the performance of view transformation based methods can be significantly improved. On similar lines, a viable direction of future work is to extend the proposed cross view gait recognition framework by computing correlation strength only in the Region Of Interest in the Gait Flow Image descriptors.

6.4 Summary

Gait research is in a quest for a robust system working in a realistic set-up without subject co-operation. This means that addressing the issues of variable covariate conditions, unconstrained training set and cross view gait recognition is critical. In response to the quest for bringing gait in to real applications this thesis has explored the affects of variable covariate conditions, investigated the performance of gait in a previously uncharted territory of recognition without subject cooperation and identified the problems caused by view. Novel methods have been developed to contribute to existing research in the above mentioned areas with significant gain in performance over contemporary methods.

Nonetheless, gait research is still in its early days and is faced with many hard problems. The issues of variable covariate conditions, unconstrained training set and view changes have been explored to some extent, but the problem of occlusion by other objects has not been looked at from the perspective of gait recognition. Occlusions by other objects are a common occurrence in surveillance (the largest candidate application for gait) and how they affect the performance of gait recognition has to be investigated in order to use gait recognition in real world applications. It may take many years of research to develop a gait recognition system using the full potential of gait as a biometric operating at a distance without subject cooperation in the presence of occlusions and view angle changes.

Bibliography

[1] R.J. Anderson. *Security Engineering: A Guide to Building Dependable Distributed Systems*. Wiley, 2008.

[2] C. Archer and T.K. Leen. Adaptive principal component analysis. Technical report, Oregon Health & Science University, 2002.

[3] K. Bashir, T. Xiang, and S. Gong. Feature selection for gait recognition without subject cooperation. In *British Machine Vision Conference*, 2008.

[4] K. Bashir, T. Xiang, and S. Gong. Cross-view gait recognition using correlation strength. In *British Machine Vision Conference*, 2010.

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

[6] K. Bashir, T. Xiang, and S.G. Gong. Gait representation using flow fields. In *British Machine Vision Conference*, 2009.

[7] C. BenAbdelkader, R. Cutler, and L. Davis. Stride and cadence as a biometric in automatic person identification and verification. In *Automatic Face and Gesture Recognition*, pages 357–362, 2002.

[8] C. BenAbdelkader, R. Cutler, and L. Davis. Gait recognition using image self-similarity. *EURASIP*, 2004(1):572–585, 2004.

[9] B. Bhanu and J. Han. Human recognition on combining kinematic and stationary features. In *Audio- and video-based biometric person authentication*, pages 600–608, Berlin, Heidelberg, 2003. Springer-Verlag.

[10] A.F. Bobick and A.Y. Johnson. Gait recognition using static, activity-specific parameters. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001.

[11] R. Bodor, A. Drenner, D. Fehr, O.T. Masoud, and N.P. Papanikolopoulos. View-independent human motion classification using image-based reconstruction. *Image and Vision Computing*, 27(8):1194–1206, July 2009.

[12] I. Bouchrika, M. Goffredo, J.N. Carter, and M.S. Nixon. Covariate analysis for view-point independent gait recognition. In *International Conference on Biometrics*, pages 990–999, 2009.

[13] I. Bouchrika and M.S. Nixon. Model-based feature extraction for gait analysis and recognition. *Mirage: Computer Vision / Computer Graphics Collaboration Techniques and Applications*, 2007.

[14] I. Bouchrika and M.S. Nixon. Exploratory factor analysis of gait recognition. In *Face and Gesture*, pages 1–6, 2008.

[15] N.V. Boulgouris and Z.X. Chi. Gait recognition using radon transform and linear discriminant analysis. *IEEE Transactions on Image Processing*, 16(3):731–740, 2007.

[16] N.V. Boulgouris and Z.X. Chi. Human gait recognition based on matching of body components. *Pattern Recognition*, 40(6):1763–1770, 2007.

[17] N.V. Boulgouris, K.N. Plataniotis, and D. Hatzinakos. An angular transform of gait sequences for gait assisted recognition. In *International Conference on Image Processing*, pages II: 857–860, 2004.

[18] N.V. Boulgouris, K.N. Plataniotis, and D. Hatzinakos. Gait recognition using dynamic time warping. In *6th Workshop on Multimedia Signal Processing*, pages 263–266, 2004.

[19] J.E. Boyd and J.J. Little. Biometric gait recognition. *Advanced Studies in Biometrics*, pages 19–42, 2005.

[20] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *European Conference in Computer Vision*, pages Vol IV: 25–36, 2004.

[21] Y. Chai, Q. Wang, J. Jia, and R. Zhao. A novel gait recognition method via fusing shape and kinematics features. In *International Symposium on Visual Computing*, pages 80–89, 2006.

[22] Y. Chai, Q. Wang, J.P. Jia, and R.C. Zhao. A novel human gait recognition method by segmenting and extracting the region variance feature. In *International Conference on Pattern Recognition*, pages IV: 425–428, 2006.

[23] C. Chen, J. Liang, H. Hu, L. Jiao, and X. Yang. Factorial hidden markov models for gait recognition. In *International Conference on Biometrics*, pages 124–133, 2007.

[24] C. Chen, J. Liang, H. Zhao, and H. Hu. Gait recognition using hidden markov model. In *International Conference on Neural Computation*, pages 399–407, 2006.

[25] M.H. Cheng, M.F. Ho, and C.L. Huang. Gait analysis for human identification through manifold learning and HMM. In *IEEE Workshop on Motion and Video Computing*, page 11, Washington, DC, USA, 2007. IEEE Computer Society.

[26] W. Cho, T. Kim, and J. K. Paik. Gait recognition using active shape models. In *Advanced Concepts for Intelligent Vision Systems*, pages 384–394, 2007.

[27] R. T. Collins, R. Gross, and J. Shi. Silhouette-based human identification from body shape and gait. In *International Conference on Automatic Face and Gesture Recognition*, page 366, Washington, DC, USA, 2002. IEEE Computer Society.

[28] S.J. Cuccurullo. *Physical Medicine and Rehabilitation Board Review*. Demos Medical Pub, 2009.

[29] D. Cunado, M.S. Nixon, and J.N. Carter. Using gait as a biometric, via phase-weighted magnitude spectra. In *Lecture Notes in Computer Science*, pages 95–102. Springer Verlag, 1998.

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

[31] N. Cuntoor, A. Kale, and R. Chellappa. Combining multiple evidences for gait recognition. In *International Conference on Multimedia and Expo*, pages 113–116, Washington, DC, USA, 2003. IEEE Computer Society.

[32] J.W. Davis and A.F. Bobick. The representation and recognition of human movement using temporal templates. *Computer Vision and Pattern Recognition*, 1997.

[33] A. Efros, C. Berg, G. Mori, and J. Malik. Recognizing action at a distance. In *International Conference in Computer Vision*, page 726, 2003.

[34] M. Ekinci. A new attempt to silhouette-based gait recognition for human identification. In *Canadian Conference on Artificial Intelligence*, pages 443–454, 2006.

[35] D. Gafurov. A survey of biometric gait recognition approaches, security and challenges. In *Nik Conference*, 2007.

[36] M. Goffredo, R.D. Seely, J.N. Carter, and M.S. Nixon. Markerless view independent gait analysis with self-camera calibration. In *Face and Gesture*, pages 1–6, 2008.

[37] J. Han and B. Bhanu. Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 28(2):316–322, Feb 2006.

[38] J. Han, B. Bhanu, and A.K.R. Chowdhury. A study on view-insensitive gait recognition. In *International Conference on Image Processing*, pages III: 297–300, 2005.

[39] J.B. Hayfron-Acquah, M.S. Nixon, and J.N. Carter. Automatic gait recognition by symmetry analysis. *Pattern Recognition Letters*, 24(13):2175–2183, 2003.

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

[41] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28(3/4):321–377, 1936.

[42] P.S. Huang, C.J. Harris, and M.S. Nixon. Human gait recognition in canonical space using temporal templates. *Visual servoing platform*, 146(2):93, April 1999.

[43] P.S. Huang, C.J. Harris, and M.S. Nixon. Recognizing humans by gait via parametric canonical space. *Artificial Intelligence in Engineering*, 13:359–366, 1999.

[44] X. Huang and N.V. Boulgouris. Human gait recognition based on multiview gait sequences. *EURASIP Journal on Advances in Signal Processing*, 2008.

[45] A.K. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 14:4–20, 2004.

[46] F. Jean, R. Bergevin, and A. Branzan. Trajectories normalization for viewpoint invariant gait recognition. In *International Conference in Pattern Recognition*, pages 1–4, 2008.

[47] A. Y. Johnson and A. F. Bobick. A multi-view method for gait recognition using static body parameters. In *Audio and Video-Based Biometric Person Authentication*, pages 301–311, London, UK, 2001. Springer-Verlag.

[48] I. T. Jolliffe. *Principal Component Analysis*. Springer-Verlag New York Inc., 2002.

[49] I.N. Junejo, E. Dexter, I. Laptev, and P. Perez. Cross-view action recognition from temporal self-similarities. In *European Conference in Computer Vision*, pages II: 293–306, 2008.

[50] A. Kale, R. Chowdhury, and R. Chellappa. Towards a view invariant gait recognition algorithm. In *Advanced Video and Signal Based Surveillance*, page 143, Washington, DC, USA, 2003.

[51] A. Kale, A. N. Rajagopalan, A. Sundaresan, A. Cuntoor, A. RoyChowdhury, V. Kruger, and R. Chellapa. Identification of humans using gait. *IEEE Transactions in Image Processing*, 2004.

[52] D. Kaziska and A. Srivastava. Cyclostationary processes on shape spaces for gait-based recognition. In *European Conference in Computer Vision*, pages II: 442–453, 2006.

[53] D. Kim and J. Paik. Model-based gait recognition using multiple feature detection. In *Advanced Concepts for Intelligent Vision Systems*, pages 1018–1029, Berlin, Heidelberg, 2008. Springer-Verlag.

[54] D.H. Kim, S. Lee, and J.K. Paik. Active shape model-based gait recognition using infrared images. In *Signal and Image Processing*, 2009.

[55] D.H. Kim and J.K. Paik. Gait recognition using active shape model and motion prediction. *IET Computer Vision*, 4(1):25–36, March 2010.

[56] W. Kusakunniran, Q. Wu, H. Li, and J. Zhang. Multiple views gait recognition using view transformation model based on optimized gait energy image. In *International Conference in Computer Vision Workshop*, 2009.

[57] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li. Support vector regression for multi-view gait recognition based on local motion feature selection. In *IEEE Conference on Computer Vision and Pattern Recognition*, June 2010.

[58] T.H.W. Lam and R.S.T. Lee. A new representation for human gait recognition: Motion silhouettes image (msi). In *International Conference on Biometrics*, pages 612–618, 2006.

[59] T.H.W. Lam, R.S.T. Lee, and D. Zhang. Human gait recognition by the fusion of motion and static spatio-temporal templates. *Pattern Recognition*, 40(9):2563–2573, September 2007.

[60] D.S. Lee, J.M. Park, and P. A. Vanrolleghem. Adaptive multiscale principal component analysis for on-line monitoring of a sequencing batch reactor. *Journal of Biotechnology*, 116(2):195 – 210, 2005.

[61] H. Lee, S. Hong, and E. Kim. An efficient gait recognition with backpack removal. *EURASIP Adv. Signal Process*, 2009:6–6, 2009.

[62] L. Lee and W.E.L. Grimson. Gait analysis for recognition and classification. In *Automatic Face and Gesture Recognition*, page 155, Washington, DC, USA, 2002. IEEE Computer Society.

[63] S.K. Lee, Y.X. Liu, and R.T. Collins. Shape variation-based frieze pattern for robust gait recognition. In *Computer Vision and Pattern Recognition*, pages 1–8, 2007.

[64] C.C. Lien, C.C. Tien, and J.M. Shih. Human gait recognition for arbitrary view angles. In *International Conference on Indicators and Concepts of Innovation*, page 303, 2007.

[65] J.J. Little and J.E. Boyd. Recognizing people by their gait: The shape of motion. *Videre*, 1(2), 1998.

[66] L.F. Liu, W. Jia, and Y.H. Zhu. Survey of gait recognition. In *International conference on Emerging intelligent computing technology and applications*, pages 652–659, Berlin, Heidelberg, 2009. Springer-Verlag.

[67] N. Liu, J.W. Lu, Y.P. Tan, and Z.Z. Chen. Enhanced gait recognition based on weighted dynamic feature. In *International Conference on Image Processing*, pages 3581–3584, 2009.

[68] Y. Liu, R. Collins, and Y. Tsin. Gait sequence analysis using frieze patterns. In *European Conference in Computer Vision*, May 2002.

[69] Z. Liu, L. Malave, A. Osuntogun, P. Sudhakar, and S. Sarkar. Toward understanding the limits of gait recognition. In *Proceedings of SPIE*, pages 195–205, 2004.

[70] Z. Liu and S. Sarkar. Simplest representation yet for gait recognition: Averaged silhouette. In *International Conference on Pattern Recognition*, pages 211–214, Washington, DC, USA, 2004. IEEE Computer Society.

[71] Z. Liu and S. Sarkar. Effect of silhouette quality on hard problems in gait recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 35(2):170–183, 2005.

[72] Z. Liu and S. Sarkar. Improved gait recognition by gait dynamics normalization. *IEEE Transactions in Pattern Analysis Machine Intelligence*, 28(6):863–876, 2006.

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

[74] A. G. Longjas. Silhouette extraction. `http://anthonylongjas.blogspot.com/2009/12/activity-3-silhouette-extraction.html`, Dec 2009.

[75] H. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos. A layered deformable model for gait analysis. In *International Conference on Automatic Face and Gesture Recognition*, pages 249–254, April 2006.

[76] J. Lu, E. Zhang, and C. Jing. Gait recognition using wavelet descriptors and independent component analysis. In *International Symposium on Neural Networks 2009*, pages 232–237, 2006.

[77] J. Lu, E. Zhang, Z. Zhang, and Y. Xue. Gait recognition using independent component analysis. In *International Symposium on Neural Networks*, pages 183–188, 2005.

[78] J.W. Lu and E. Zhang. Gait recognition for human identification based on ica and fuzzy svm through multiple views fusion. *Pattern Recognition Letters*, 28(16):2401–2411, December 2007.

[79] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi. Gait recognition using a view transformation model in the frequency domain. In *European Conference in Computer Vision*, pages 151–163, 2006.

[80] A. McCue. Heathrow testing biometric security checks. `http://news.cnet.com/Heathrow-testing-biometric-security-checks/2100-7348_3-6141368.html?tag=mncol`, Dec 2006.

[81] D. Ming, C. Zhang, Y. Bai, B. Wan, Y. Hu, and K. Luk. Gait recognition based on multiple views fusion of wavelet descriptor and human skeleton model. In *Virtual Environments, Human-Computer Interfaces and Measurement Systems*, pages 246–249, Piscataway, NJ, USA, 2009. IEEE Press.

[82] S.D. Mowbray and M.S. Nixon. Automatic gait recognition via fourier descriptors of deformable objects. In *Audio- and video-based biometric person authentication*, pages 566–573, 2003.

[83] M.S. Nixon and J.N. Carter. Advances in automatic gait recognition. *Automatic Face and Gesture Recognition*, 0:139, 2004.

[84] M.S. Nixon and J.N. Carter. Automatic recognition by gait. *Proceedings of the IEEE*, 94(11):2013–2024, November 2006.

[85] M.S. Nixon and B.F. Guo. Gait feature subset selection by mutual information. In *Biometrics: Theory, Applications and Systems*, pages 1–6, 2007.

[86] S. Rahati, R. Moravejian, and F.M. Kazemi. Gait recognition using wavelet transform. *Information Technology: New Generations*, 0:932–936, 2008.

[87] L. Rustagi, L. Kumar, and G.N. Pillai. Human gait recognition based on dynamic and static features using generalized regression neural network. *Machine Vision, International Conference on*, 0:64–68, 2009.

[88] S. Sarkar, P. Phillips, Z. Liu, I. Vega, P. Grother, and K. Bowyer. The HumanID gait challenge problem: Data sets, performance, and analysis. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 27(2):162–177, 2005.

[89] R.D. Seely, S. Samangooei, M. Lee, J.N. Carter, and M.S. Nixon. The university of southampton multi-biometric tunnel and introducing a novel 3D gait dataset. In *Biometrics: Theory, Applications and Systems*, pages 1–6, 2008.

[90] G. Shakhnarovich and T. Darrell. On probabilistic combination of face and gait cues for identification. In *IEEE International Conference on Automatic Face and Gesture Recognition*, page 176, Washington, DC, USA, 2002. IEEE Computer Society.

[91] G. Shakhnarovich, L. Lee, and T.J. Darrell. Integrated face and gait recognition from multiple views. In *Computer Vision and Pattern Recognition*, pages I:439–446, 2001.

[92] E.C. Shannon. Prediction and entropy of printed english. *The Bell System Technical Journal*, 30:50–64, January 1951.

[93] J. Shutler, M. Grant, M. Nixon, and J Carter. On a large sequence-based human gait database. In *International Conference on Recent Advances in Soft Computing*, pages 66–71, 2002.

[94] C. Stauffer and W.E.L Grimson. Adaptive background mixture models for real-time tracking. In *Computer Vision and Pattern Recognition*, 1999.

[95] D. Tan, K. Huang, S. Yu, and T. Tan. Efficient night gait recognition based on template matching. In *International Conference on Pattern Recognition*, pages 1000–1003, Washington, DC, USA, 2006. IEEE Computer Society.

[96] D. Tao, X. Li, X. Wu, and S.unhong Maybank. General tensor discriminant analysis and gabor features for gait recognition. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 29(10):1700–1715, 2007.

[97] A. Tsuji, Y. Makihara, and Y. Yagi. Silhouette transformation based on walking speed for gait identification. In *Computer Vision and Pattern Recognition*, 2010.

[98] H. Uemura, S. Ishikawa, and K. Mikolajczyk. Feature tracking and motion compensation for action recognition. In *British Machine Vision Conference*, 2008.

[99] G.V. Veres, L. Gordon, J.N. Carter, and M.S. Nixon. What image information is important in silhouette-based gait recognition? In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 776–782, June 2004.

[100] D.K. Wagg and M.S. Nixon. On automated model-based extraction and analysis of gait. In *Automatic Face and Gesture Recognition*, pages 11–16, 2004.

[101] L. Wang, H. Ning, T. Tan, and W. Hu. Fusion of static and dynamic body biometrics for gait recognition. In *International Conference in Computer Vision*, page 1449, Washington, DC, USA, 2003. IEEE Computer Society.

[102] L. Wang, T. Tan, H. Ning, and W. Hu. Silhouette analysis-based gait recognition for human identification. *IEEE Transactions Pattern Analysis Machine Intelligence*, 25(12):1505–1518, 2003.

[103] Y. Wang, S. Yu, Y. Wang, and T. Tan. Gait recognition based on fusion of multi-view gait sequences. In *International Conference on Biometrics*, pages 605–611, 2006.

[104] C. K. I. Williams and D. Barber. Bayesian classification with gaussian processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(12):1342–1351, 1998.

[105] C. K. I. Williams and C. E. Rasmussen. Gaussian processes for regression. In *Neural Information Processing Systems*, pages 514–520, 1995.

[106] T. Xiang and S. Gong. Model selection for unsupervised learning of visual context. *International Journal of Computer Vision*, 69:181–201, 2006.

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

[108] D. Xu, S. Yan, D. Tao, L. Zhang, X. Li, and H. Zhang. Human gait recognition with matrix representation. *IEEE Transactions on Circuits System Video Technologies*, 16(7):896–903, 2006.

[109] C. Yam and M.S. Nixon. Model-based gait recognition. In *Enclycopedia of Biometrics*, pages 633–639, 2009.

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

[111] A. Yamashita, T. Kaneko, S. Matsushita, and K.T. Miura. Region extraction with chromakey using stripe backgrounds. *Trans Electron (Inst Electron Inf Commun Eng)*, 2004.

[112] H. Yang and S. Lee. Reconstruction of 3D human body pose for gait recognition. *Advances in Biometrics*, 3832:619–625, 2005.

[113] X. Yang, Y. Zhou, T. Zhang, G. Shu, and J. Yang. Gait recognition based on dynamic region analysis. *Signal Processing*, 88(9):2350 – 2356, 2008.

[114] J. Ye, R. Janardan, and Qili. Two dimensional linear discriminant analysis. *Advances in neural information processing systems*, 17:1569–1576, 2005.

[115] S. Yu, D. Tan, K. Huang, and T. Tan. Reducing the effect of noise on human contour in gait recognition. In *International Conference on Biometrics*, pages 338–346, 2007.

[116] S. Yu, D. Tan, and T. Tan. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In *International Conference on Pattern Recognition*, pages 441–444, 2006.

[117] S. Yu, L. Wang, W. Hu, and T. Tan. Gait analysis for human identification in frequency domain. In *Image and Graphics*, pages 282–285, Dec 2004.

[118] E. Zhang, J. Lu, and G. Duan. Gait recognition via independent component analysis based on support vector machine and neural network. In *International Conference on Neural Computation*, pages 640–649, 2005.

[119] H. Zhang and Z. Liu. Gait representation and recognition using haar wavelet and radon transform. In *International Conference on Information Engineering*, pages 83–86, Washington, DC, USA, 2009. IEEE Computer Society.

[120] R. Zhang, C. Vogler, and D. Metaxas. Human gait recognition at sagittal plane. *Image Vision Computing*, 25(3):321–330, 2007.

[121] G. Zhao, L. Cui, H. Li, and M. Pietikainen. Gait recognition using fractal scale and wavelet moments. In *International Conference on Pattern Recognition*, pages 453–456, Washington, DC, USA, 2006. IEEE Computer Society.

[122] G. Zhao, G. Liu, H. Li, and Pietikainen. 3D gait recognition using multiple cameras. In *International Conference on Automatic Face and Gesture Recognition*, pages 529–534, April 2006.

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