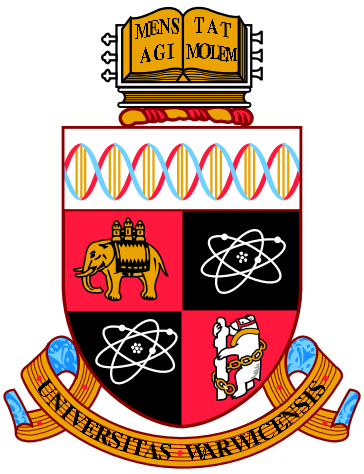


Uniscale and Multiscale Gait Recognition in Realistic Scenario

by

Sruti Das Choudhury

Thesis

Submitted to the University of Warwick in partial fulfilment of

the requirements for the degree of

Doctor of Philosophy

School of Engineering

October 2013

THE UNIVERSITY OF WARWICK

Contents

List of Tables

List of Figures

Acknowledgments

I would like to take the opportunity to thank my research supervisor Dr. Tardi Tjahjadi for his invaluable academic support towards this research. I am also very grateful to my annual progress panel members of the School of Engineering of the University of Warwick: Dr. Mark Leeson and Prof. Evor Hines. Their constructive advice, guidance and encouragement throughout the process were of enormous help to me.

I would like to acknowledge the financial, academic and technical assistance of the University of Warwick and its staff members, particularly in the award of Warwick Postgraduate Research Scholarship and Engineering bursary that provided the necessary financial support for this research.

I would like to sincerely thank Dr. Amlan Chakrabarti, the Coordinator of A.K. Choudhury School of Information Technology, University of Calcutta, and my M.Tech project supervisor, Dr. Debotosh Bhattacharjee for their invaluable contributions in shaping my educational career. A special thanks to my beloved friend Miss Ting Chen whose supportive presence has made my days brighter in abroad. Last, but by no means least, I would like to express my sincerest gratitude to my Late father, Mr. Gopal Das Choudhury, and my mother Mrs. Aditi Das Choudhury, whose blessings, I firmly believe, are my guiding light, and my colleague and friend, Mr. Rana Pratap Pal who stood by me at difficult times of my life with a source of inspiration- it is to them my thesis is dedicated.

Declarations

This thesis is submitted in partial fulfilment for the degree of Doctor of Philosophy under the regulations set out by the Graduate School at the University of Warwick. This thesis is solely composed of research undertaken by Sruti Das Choudhury under the supervision of Dr. Tardi Tjahjadi. The research materials have not been submitted in any previous application for a higher degree. All sources of information are specifically acknowledged in the content.

Abstract

The performance of a gait recognition method is affected by numerous challenging factors that degrade its reliability as a behavioural biometrics for subject identification in realistic scenario. Thus for effective visual surveillance, this thesis presents five gait recognition methods that address various challenging factors to reliably identify a subject in realistic scenario with low computational complexity. It presents a gait recognition method that analyses spatio-temporal motion of a subject with statistical and physical parameters using Procrustes shape analysis and elliptic Fourier descriptors (EFD). It introduces a part-based EFD analysis to achieve invariance to carrying conditions, and the use of physical parameters enables it to achieve invariance to across-day gait variation. Although spatio-temporal deformation of a subject's shape in gait sequences provides better discriminative power than its kinematics, inclusion of dynamical motion characteristics improves the identification rate. Therefore, the thesis presents a gait recognition method which combines spatio-temporal shape and dynamic motion characteristics of a subject to achieve robustness against the maximum number of challenging factors compared to related state-of-the-art methods. A region-based gait recognition method that analyses a subject's shape in image and feature spaces is presented to achieve invariance to clothing variation and carrying conditions. To take into account of arbitrary moving directions of a subject in realistic scenario, a gait recognition method must be robust against variation in view. Hence, the thesis presents a robust view-invariant multiscale gait recognition method. Finally, the thesis proposes a gait recognition method based on low spatial and low temporal resolution video sequences captured by a CCTV. The computational complexity of each method is analysed. Experimental analyses on public datasets demonstrate the efficacy of the proposed methods.

Abbreviations

1D	1-dimensional
2D	2-dimensional
3D	3-dimensional
ARPoLK	angular rotation pattern of leading knee
ASM	active shape model based gait recognition in [35]
BDHM	Bhattacharyya distance histogram matching
CCR	correct classification rate
CCTV	closed circuit television
CMC	cumulative match characteristic
COM-SC	centre of mass of the subject's silhouette contour
CSI	concatenated silhouette image
CVMR	correct view matching rate
CASIA	silhouette analysis-based gait recognition in [29]
CMU	Carnegie Mellon University
CMU2	gait recognition based on shape estimation in [89]
CMU1	silhouette-based human identification from body shape and gait in [19]

CHLAC	cubic higher-order local auto-correlation
CGI	chrono gait image based gait recognition in [28]
DFT	discrete Fourier transform
DT-CWT	dual-tree complex wavelet transform
DTW	dynamic time warping
DWT	discrete wavelet transform
DNGR	dynamics normalisation based gait recognition method in [26]
EFD	elliptic Fourier descriptor
ER-TSR	gait recognition method using example-based and reconstruction-based temporal super resolution in [69]
FD	Fourier descriptor
FED-HMM	frame to exemplar distance and hidden Markov model based gait recognition in [37]
FDA	Fisher discriminant analysis
FFT	fast Fourier transform
GEI	gait energy image and gait energy image based gait recognition method in [25]
GFI	gait flow image based gait recognition in [49]
GKI	gait keyframe image
GPDF	gait recognition method using Gabor wavelet and patch distribution features in [47]
GR-LSTR	gait recognition method based on low spatial and low temporal resolution silhouette sequences
GTDA	general tensor discriminant analysis
GTDA-GF	general tensor discriminant analysis and Gabor features based gait recognition in [50]

HMM	hidden Markov model
Hp-Gf	high-pass Gaussian filter
IDT-CWT	inverse dual-tree complex wavelet transform
LDA	linear discriminant analysis
LGSR	locality-constrained group sparse representation
Lp-Gf	low-pass Gaussian filter
MGI	multiscale gait image
Morph	level-set morphing based gait recognition method in [123]
MSL	multiple subspace learning
MMFA	gait recognition using matrix-based marginal Fisher analysis in [111]
NNC	nearest neighbour classifier
NAC	normalized auto correlation
OPA	ordinary Procrustes analysis
PCA	principal component analysis
PSA	Procrustes shape analysis
P-TSR	periodic temporal super resolution based gait recognition method in [68]
PWMS	phase-weighted magnitude spectra
RALDA	radon transforms and linear discriminant analysis based gait recognition in [119]
ReG-IF	region-based gait recognition method in image and feature spaces
R-GEI	reflected gait energy image
ROC	receiver operating characteristic

ROI	region-of-interest
Rf-ROI	reference region-of-interest
RCK-G	radial integration transform, circular integration transform and weighted Krawtchouk moments with genetic algorithm based gait recognition in [65]
SDIS	shape descriptors in the image space
SDFS	shape descriptors in the feature space
SGEI	segmented gait energy image
SGEnI	segmented gait entropy image
STM-SPP	spatio-temporal motion characteristics, statistical and physical parameters based gait recognition method in [7]
STS-DM	spatio-temporal shape and dynamic motion characteristics based gait recognition method in [8]
SI-PSA	speed-invariant gait recognition method based on Procrustes shape analysis in [42]
ST-WS	silhouette transformation based walking speed invariant gait recognition in [43]
SSM	shape sequence matching based method in [23]
SVB frieze	gait recognition using shape variation-based frieze pattern in [40]
SSP	image self-similarity plot based gait recognition in [21]
Tr-ROI	target region-of-interest
VI-MGR	view-invariant multiscale gait recognition method in [9]
WAvgI	weighted average identification rate
WAvgV	weighted average verification rate
WRSL	weighted random subspace learning

Chapter 1

Introduction

1.1 Introduction

The necessity of automatic human identification is becoming indispensable to provide adequate security measures for the welfare of the mankind. The traditional systems for ascertaining a human subject's identity are either knowledge-based, i.e, based on something that the subject knows, or possession-based, i.e, based on something that the subject posseses. Passwords, pin codes and pass-phrases, as used in knowledge-based systems can be forgotten, stolen or used surreptitiously. Tokens and ID cards, as used in possession-based systems, are also exposed to these vulnerabilities, with an additional possibility of being misplaced. While traditional security systems fail to cope with the growing importance of reliable human identification due to these disadvantages, biometrics has emerged as a reliable means of identifying a human subject in the presence of real-life challenges.

1.2 Basics of Biometrics

The word 'biometrics' is derived from two Greek roots: 'bios' meaning life; and 'metron' meaning measurement. Hence, biometrics deals with the science and technology to uniquely identify a human subject based on the distinguishing biological features. The biometrics must satisfy the following properties [1]:

1. Universality: The feature must be possessed by every human subject in a normal population.
2. Uniqueness: The feature of a subject should be unique and must be sufficiently distinguishable from every other subject.

3. Permanence: The feature must remain relatively unaffected with ageing over the period of interest.

4. Collectibility: The feature must be captured in real time without any intrusions on the subject's privacy.

5. Measurability: The feature must be quantitatively measurable with ease.

In addition to these properties, a biometrics system must take into account of the following practical considerations to be effectively used in the real world applications:

1. Throughput: A biometrics system must be able to identify a subject in real time.

2. Usability: A biometrics system should have the capability to be intutively used with ease to provide user satisfaction.

3. Acceptability: A biometrics system must be acceptable irrespective of age, sex, race and cultural concerns.

1.3 Types of Biometrics

Biometrics can be broadly classified in two categories: physiological biometrics and behavioural biometrics. Physiological biometrics analyses the physical attributes of a subject such as face, fingerprint, iris, palm-print, earlobe geometry etc., and requires cooperation from the subject for a particular view in a controlled-lighting environment [2; 3]. Behavioural biometrics analyses human behaviour or action performed over a defined time interval for identifying a subject. The most promising example of behavioural biometrics is gait, which is defined as the "manner of moving on foot" [4].

1.4 Advantages of Gait Recognition

Human identification based on gait analysis has drawn the attention of computer vision researchers due to following advantages: (a) each person has a sufficiently distinctive way of walking; (b) gait features can be extracted secretly from a distance without interfering with the subject's activity; (c) gait characteristics can be analysed from low resolution video sequences; and (d) it is very difficult to conceal and disguise gait characteristics.

1.5 Gait Recognition System

A gait recognition system consists of the following components as shown in Fig. 1.1:

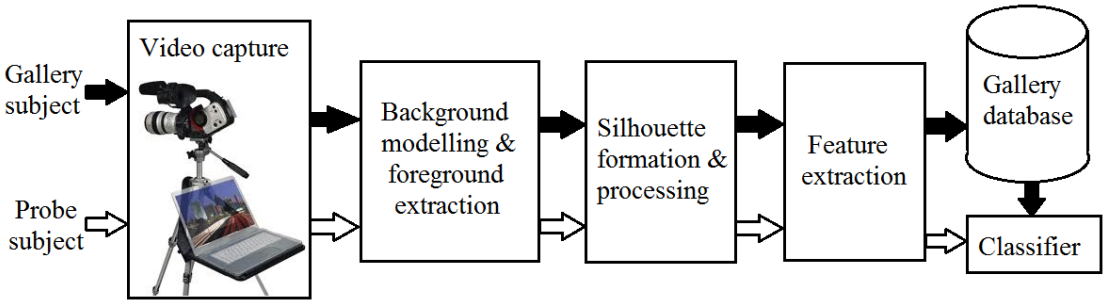


Figure 1.1: Gait recognition system.

- Video capture: This subsystem is responsible for capturing raw video sequences of the walking gallery subjects (the subjects with whom an unknown subject is matched to determine his/her identity) and the probe subject (the unknown subject who is matched with the gallery subjects for identification).
- Background modelling and foreground extraction: This subsystem involves the isolation of background, i.e., the part of the scene which remains static over the period of interest, from the moving foreground we are interested in, i.e., the walking subject.
- Silhouette formation and processing: The extracted foreground is binarised to form a silhouette. The silhouette is cropped according to the bounding rectangle enclosing it to remove camera depth variations, and resized to a fixed height.
- Feature extraction: Feature extraction involves transformation of the input video sequences into a reduced set of features which represent their relevant information in order to perform subject identification based on gait.
- Data storage: The extracted features of the gallery subjects are stored in a database either centrally or locally.
- Classifier: The features of a probe subject are compared with the features stored in the gallery database using appropriate classification rules for identifying the subject.

1.6 Challenges and Applications of Gait Recognition

There are various challenging factors that adversely affect the performance of a gait recognition system. A gait recognition system is mainly affected due to presence of a carried item, and variations in view and clothing type. In addition to these challenges, there are other covariate factors which also affect a gait recognition system, e.g., varying weather conditions, change in hair style, footwear and ground surface, shadows under feet, self-occlusions,

physical injury, pregnancy, drunkenness, fatigue and illness. The variations in walking speed induced by the subject's change in mood also bring a challenge to a gait recognition system. The performance of a gait recognition method is affected due to presence of occluding objects in the scene. Like most other biometrics, a subject's gait characteristics also change with ageing.

Despite these challenges, gait recognition has made significant contributions in the field of visual surveillance, access control, biometric authentication and criminology (e.g.,[2; 5]). Its usefulness in forensics is also affirmed [6]. Recently gait recognition has contributed to providing evidences to several criminal convictions, e.g, the murder of Swedish Foreign Minister Anna Lindh, bank robbery in Noerager (Denmark) and a burglary in Lancashire [2].

1.7 Contributions and Thesis Structure

The principal contributions of this thesis are as follows:

1. It introduces the application of traditional shape analysis in gait recognition to effectively combine spatiotemporal gait characteristics, statistical and physical parameters of a human body for identification with reduced computational complexity using simplified feature space;

2. It introduces a gait recognition method with low computational complexity which analyses spatio-temporal shape and dynamic motion characteristics of a subject's contour to achieve robustness against the maximum number of challenging factors that affect an existing gait recognition system;

3. It introduces a region-based gait recognition method in image and feature spaces which is robust against variations in clothing and carrying conditions;

4. It introduces a view-invariant multiscale gait recognition method which is robust to variations in clothing and presence of a carried item in addition to variations in view;

5. It introduces a gait recognition method using low frame-rate and low spatial-resolution video sequences captured by a closed circuit television (CCTV) camera.

This thesis presents five novel model-based and model-free gait recognition methods to identify human subjects in realistic scenarios by addressing various real-life challenges of gait recognition for visual surveillance and public security. The thesis is organised in eight Chapters of which Chapters 3 to 7 present the novel contributions of this research.

In each chapter, a review of related methods are presented. The individual chapters of this thesis are structured as follows:

Chapter 2 provides a detailed discussion of various public gait databases. A review of model-based and model-free gait recognition methods is also presented in this chapter.

Chapter 3 presents a gait recognition method which combines spatio-temporal motion characteristics, statistical and physical parameters (STM-SPP) of a subject for its classification by analysing shape of the subject's silhouette contours using Procrustes shape analysis (PSA) and elliptic Fourier descriptors (EFDs). A part-based shape analysis using EFDs is also introduced to achieve robustness against carrying conditions. The classification results by PSA and EFDs are combined, resolving tie in ranking using contour matching based on Hu moments. The results of experimental analysis on public datasets are provided and a comparison is made against related gait recognition methods. The computational complexity of the method is also analysed. The results from this chapter have been published in [7].

Chapter 4 presents a three-phase gait recognition method that analyses the spatio-temporal shape and dynamic motion (STS-DM) characteristics of a subject's silhouettes to identify the subject in the presence of most of the challenging factors that affect existing gait recognition systems. The match scores generated in the three phases as a result of analysing the spatio-temporal shape, and local and global dynamic motion characteristics of a subject are fused using weight-based score-level fusion for robust identification. The results of experimental analysis on public datasets are provided and the computational complexity of the method is analysed. The results from this chapter have been published in [8].

Chapter 5 introduces a method for region-based gait recognition in image and feature spaces (ReG-IF), which is robust to variations in the subject's clothing, carrying conditions, walking speed and occlusions. The shape descriptors in the image space uses lowpass and highpass Gaussian filters each at five cut-off frequencies, and the shape descriptors in the feature space uses weighted Krawtchouk moments to analyse the shape of silhouettes. Extensive experiments on publicly available datasets demonstrate the efficacy of ReG-IF. The computational complexity of the method is also analysed.

Chapter 6 proposes a robust two-phase view-invariant multiscale gait recognition method (VI-MGR) which is not only robust to variations in view, but also robust to variations in clothing types and presence of a small carried item. The method determines which of the available views in the gallery match most closely with the unknown view of the probe subject, and the probe subject is compared with the matched view of the gallery subjects based on multiscale shape analysis for subject identification. Experimental analysis on publicly available datasets shows that VI-MGR outperforms the state-of-the-art gait recognition methods. The results from this chapter have been submitted for publication in [9].

Due to limited transmission bandwidth and storage capacity, the gait sequences captured by a CCTV camera are recorded at low frame-rates with low spatial resolution, which significantly affect the performance of a gait recognition method. Thus, Chapter 7 presents a gait recognition method using low spatial and low temporal resolution gait sequences (GR-LSTR) captured by a CCTV camera. Experimental analysis of this method on a public dataset is provided, and the computational complexity of the method is analysed. The results from this chapter have been submitted for publication in [10]. Finally, Chapter 8 concludes the thesis and provides directions for future works.

1.8 List of publications

1.8.1 Published Papers

1. S. D. Choudhury and T. Tjahjadi, "Silhouette-based gait recognition using Procrustes shape analysis and elliptic Fourier descriptors," Pattern Recognit., vol. 45, pp. 3414-3426, 2012.

2. S. D. Choudhury, T. Tjahjadi, "Gait recognition based on shape and motion analysis of silhouette contours," Comput. Vis. Image Und., vol. 117, no. 12, pp. 1770-1785, 2013.

3. Y. Guan, C.-T. Li and S. D. Choudhury, "Robust Gait Recognition from Extremely Low Frame-Rate Videos," in: Proc. Int. Workshop on Biometrics and Forensics, Lisbon, Portugal, April, 2013.

1.8.2 Submitted Papers

1. S. D. Choudhury and T. Tjahjadi, "Robust View-Invariant Multiscale Gait Recognition," Pattern Recognit., 2013.

2. S. D. Choudhury, Y. Guan, C.-T. Li, "Gait Recognition using Low Spatial and Temporal Resolution Videos," in: Proc. Int. Workshop on Biometrics and Forensics, Valletta, Malta, 2014.

Chapter 2

Background Knowledge

2.1 Introduction

This chapter presents a review of the state-of-the-art gait recognition methods in order to facilitate the appreciation of the novel contributions of the gait recognition methods introduced in the thesis. The performance of these gait recognition methods are evaluated on the silhouettes provided by the various publicly available gait datasets. Therefore, this chapter also presents a discussion on the early and current gait datasets used in the gait recognition research.

2.2 Literature Review

The moving light display experiment of Johansson [11] pioneered the study of human identification based on gait analysis. In this experiment, an image sequence is reduced to a set of moving light dots attached at the different joints of the human subject. It was successfully demonstrated that human activities such as walking, running or stair climbing can easily be recognised from the sequence of relative movements of these light dots. Cutting et al. [12] showed that this ability can also be extended to recognise friends and gender of the subjects. Since then, gait recognition gained growing interest from the computer vision researchers, and numerous methods have been proposed.

Markerless gait recognition methods can be classified as model-based and model-free. Model-based methods (e.g., [13; 14; 15; 16; 17; 18]) use a structural model to measure time-varying gait parameters, e.g., gait period, stance width and stride length, and a motion model to analyse the kinematical and dynamical motion parameters of the subject, e.g., rotation patterns of hip and thigh, and joint angle trajectories, to obtain gait signatures. Model-free methods (e.g., [7; 19; 20; 21; 22; 23; 24; 25; 26; 27; 28]) analyse the spatio-

temporal shape and motion characteristics of a silhouette in a gait sequence to achieve better recognition accuracy at the expense of viewpoint-dependency and susceptibility to across-day variations. The model-free gait recognition methods in [7; 29] analyse the dynamic and/or static gait characteristics of silhouettes or the extreme outer boundary of the silhouettes, i.e., contours of a gait sequence for identifying a human subject. But the performance of these methods largely depends on the efficiency of the background segmentation techniques, presence of occluding objects in the scene and shadows under feet, as these factors considerably determine the quality of the silhouettes and the extracted contours. A gait energy image (GEI) introduced in [25] is formed by averaging the silhouettes of a gait period to capture spatio-temporal gait characteristics. Since then many promising model-free gait recognition methods have been proposed based on a GEI, as it is robust to segmentation imperfections, and facilitates noise-resilient gait feature extraction in reduced space and time complexity. The model-based and model-free gait recognition methods analyse the static shape and dynamic motion characteristics of gait sequences to achieve robustness against the various covariate factors, i.e., intra-class variations in view, walking speed and clothing types, presence of a carried item as well as other covariate factors, e.g., segmentation noise, occlusions, changes in ground surface and shadows under feet.

2.2.1 Model-based Methods

The method in [30] models the limb as two articulated inter-connected pendula and used dynamic Hough transform to extract lines representing the thigh in each frame. The least squares analysis method is used to deal with the missing lines due to self-occlusions, and the phase-weighted magnitude spectra (PWMS) of the thigh rotation pattern are used as gait signatures for subject identification. The method in [31] also models the thighs as two thick lines joined at a single point, and describes the thigh rotation pattern as a pendulum motion for evidence gathering based gait recognition. However, the use of velocity Hough transform in [31] enables articulated thigh extraction which is robust to self-occlusions, and hence does not require least squares analysis as in [30]. A genetic algorithm is used in this method to reduce the dimensionality of the feature space. The method in [32] achieves higher robustness against noise and self-occlusions compared to the methods in [30; 31] by using an evidence gathering process based on velocity Hough transform. The superiority of velocity Hough transform to traditional Hough transform tracking for noise-resilient spatio-temporal gait feature extraction is demonstrated. The method computes both the structural and motion model of the thigh of a walking subject, and models the thigh rotation pattern using a Fourier series for feature extraction. Like the method in [31], this method also uses a genetic algorithm to reduce the computational complexity.

The method in [33] uses Euclidean distance to measure similarity between the time-

normalised joint angle trajectories of the subjects' lower limbs in the sagittal plane and uses nearest neighbour classifier (NNC) to identify the subjects. The full-body layered deformable model in [15] consists of ten segments corresponding to the head, torso, left/right upper arms, left/right lower arms, left/right upper legs and left/right lower legs modelled by a circle, a semi-ellipse on top of a rectangle, rectangles and quadrangles. The static and dynamic parameters of the layered deformable model are determined by using simple geometric operations, filtering and mean-shift algorithm. The layered structure addresses self-occlusions and a subject is classified using dynamic time warping (DTW). The inclusion of upper body dynamics in addition to lower body dynamics enables the method to reliably identify a human subject in presence of shadows under feet, change in ground surface and carrying conditions.

The model-based method in [13] uses Bayesian belief network for tracking interacting subjects in a 3-dimensional (3D) unconstrained environment based on multi-camera image fusion using Kalman filter. The tracked subject's body is represented by a hierarchical, structural model which is a combination of bounding volume and stick component. The method analyses stride length, walking speed, cadence, arm-swing, stance width and stance time for identifying the subjects. In order to increase tracking accuracy and inter-subject variations, the hierarchical, structural body model is augmented with additional hard kinematic constraints based on physical measurements, e.g., joint angles, and soft kinematic constraints based on learned body movement in a stochastic probabilistic framework to account for the variability of structural model parameters. The method in [34] analyses the angular motion of hip, knee and ankle joints using phase-weighted Fourier magnitude spectra to obtain dynamic gait features. The dynamic gait features are fused with the static gait features, i.e., subject's height and stride length, and height of different parts of the body for subject identification. The adaptive sequential forward floating search algorithm is used to select discriminatory gait features for improved recognition rate.

The shape model in [16] segments the subject's body into three regions: head, torso and limb to obtain static parameters, e.g., height of the body, its centroid coordinates and gait period. The motion model consists of four segments, i.e., head, torso, legs and arms, and is used to estimate the dynamic parameters. Instead of using line Hough transform which suffers from the constraint of Hough space determination, this method uses active contour models to precisely determine the borders of each limb. Each limb is modelled as two sticks representing the thigh and lower leg which are interlinked at the knee joint, and their rotation patterns form the dynamic gait signature. Dynamic Hough transform is used to study the effect of arm swing on gait recognition using NNC.

The method in [17] uses models to obtain skeleton parameters by wavelet decomposition of a GEI and invariant moments for combining anatomical and behavioural char-

acteristics of gait. Thermal imaging is used to extract silhouettes that are invariant to carrying conditions and lighting variations. A hierarchical prediction-based active shape model (ASM) is used in [35] to extract model parameters for gait recognition. The feature extraction process of ASM consists of the following components: (a) silhouette modelling of a subject based on a priori shape information; (b) foreground extraction based on Mahalanobis distance; (c) rectification of geometric distortion based on homography estimation using a simple linear algorithm to achieve view invariance; and (d) gait parameter estimation using ASM. Kalman filter is used in this method to analyse the global motion characteristics of a subject which is invariant to changes in illumination, shadows and occlusions. Unlike the method in [22] which detects a gait period by counting the number of foreground pixels of silhouette sequences, the method uses a model-based gait period detection strategy using prediction-based ASM for improved performance.

The methods in [36; 37] use hidden Markov model (HMM) based framework for gait recognition. The static posture of different phases of gait are represented by different states of HMM in [36], while the transition between these states are modelled using a transition matrix to take into account of the dynamic characteristics of gait. The method based on frame to exemplar distance and HMM (FED-HMM) in [37] uses two image features, i.e., the width of the silhouette contour and the binary silhouette sequences over a gait period, for identifying a subject based on two approaches for feature extraction. In the indirect approach, it computes a frame to exemplar distance to transform the high dimensional image feature based on silhouette contour into low dimensional space, and uses HMM for subject identification. In the direct approach, it avoids learning high dimensional probability density functions by not employing frame to exemplar distance to obtain an observation vector. Instead, it directly uses the image features to train the HMM for subject identification.

The method in [38] analyses the spatio-temporal shape characteristics of a subject based on PSA and uses Procrustes mean shape distance to obtain a dissimilarity score. It also analyses dynamic gait characteristics, e.g., joint angle trajectories based on DTW and uses Euclidean distance to obtain a dissimilarity score. Hence, this method is a combination of model-based and model-free gait recognition methods, and considers both structural and transitional characteristics of gait independently to recognise subjects based on the nearest exemplar pattern classifier. The dissimilarity scores obtained by static and dynamic gait characteristics are also fused on a decision level based on score-summation-based classifier combination rule following transformation of the scores to the same scale using a logistic function to improve the identification rate.

The above-mentioned methods rely on 2-dimensional (2D) models, and are evaluated based on 2D gait datasets. In recent years, methods have been developed based on 3D models as in [39] for view-invariant gait recognition. The method in [39] uses four 3D

cylinders to model the shin and thigh of each leg connected at the knee and hip joints with 3D degrees of freedom. The three static gait features, namely height, stride length, and footprint pose of a subject are extracted from the voxel gait sequence. The dynamic gait characteristics, namely rotation pattern of centre of hip and gait kinematics trajectories are extracted by fitting the gait model onto the 3D voxel gait sequence based on a correlation algorithm for generating a correlation energy map. DTW is used to measure the similarity between the hip rotation pattern, and a subject is classified using NNC.

2.2.2 Model-free Methods

The correspondence-free method in [20] computes the gait period of a subject by analysing width of the bounding box which encloses the subject's moving silhouette and uses a Bayesian classifier to confirm the subject's identity. However, silhouette width is not effective for computing gait period of the frontal view of a moving subject. Hence, temporal change of silhouette height is also considered in [19] to achieve robustness against different views. This method performs template matching between key frames of the gallery and probe gait sequences using normalised correlation to obtain correlation scores, and uses NNC for gait recognition. The method in [29] converts a binary silhouette into a 1-dimensional (1D) normalised distance signal by contour unwrapping with respect to the silhouette centroid. Principal component analysis (PCA) is then used to reduce the dimensionality of the feature space and to obtain projection centroids corresponding to each gallery sequence in the eigenspace. Finally, NNC and NNC with respect to class exemplars are used for identification. The identification is validated based on the subject's physical parameters for increased accuracy. The HumanID Gait Challenge Problem introduced in [22] plays a significant role in the advancement of gait recognition research by presenting a Baseline algorithm, HumanID Gait Challenge dataset and a set of 12 experiments. The Baseline algorithm computes spatio-temporal correlation of silhouettes to measure similarity between two gait sequences. The HumanID Gait Challenge data set is a publicly available data set consisting of 1870 sequences of 122 subjects considering five covariates: viewpoint, shoe, carrying a briefcase, ground surface and time. The 12 gait challenge experiments are designed to analyse the effect of these covariates on gait recognition.

Different subjects have different walking speeds, and a subject's walking speed differs significantly in a gait sequence induced by the subject's mood changes. Since the performance of a gait recognition method is adversely affected by the variations in walking speed, the methods in [26; 40; 41; 42; 43; 44] address walking speed variations for improved gait recognition. The method in [40] computes shape variation-based frieze pattern (SVB frieze) of the difference frames obtained by subtracting the key frames, e.g., double support stance frames from the series of subsequent frames of a gait period. The method

in [41] replaces the centroid-based shape configuration of traditional PSA with high-order shape configuration to take into account of dynamic gait characteristics. The method introduces a differential composition model for increased inter subject discriminability and uses Procrustes distance for identifying a subject.

The dynamics normalisation based gait recognition (DNGR) method in [26] normalises gait dynamics using population HMM whose states represent specific gait stances over a gait period determined by Viterbi algorithm, and gait recognition is achieved by estimating the distances between two time-normalised gait signatures in linear discriminant analysis (LDA) space so as to maximise intra-class discrimination of subjects. DNGR uses an eigenstance reconstruction model to smooth silhouettes and achieves invariance to walking speed and changes in ground surface. Unlike DNGR which only takes into account of temporal changes, e.g., gait period due to walking speed variations, the silhouette transformation based walking speed invariant method (ST-WS) in [43] considers spatial changes, e.g., stride length in addition to temporal changes for robust walking speed-invariant gait recognition. Since variations in walking speed does not have any effect on static features, ST-WS separates static and dynamic features by fitting a 2D trapezoidal link human model and uses a factorization-based transformation model to transform the dynamic features from a reference speed to a target speed. Finally, the transformed dynamic and unaltered static features are combined to generate the silhouette sequence in the target speed. The speed-invariant gait recognition method based on PSA (SI-PSA) [42] uses PSA to compute high-order derivative shape configuration. The method in [44] uses cubic higher-order local auto-correlation (CHLAC) to extract static shape and dynamic motion characteristics of a gait sequence based on Fisher discriminant analysis (FDA) for better class separability. While FDA-based CHLAC is very useful to take into account of walking speed variations between gait sequences, HMM deals with walking speed variations within a gait sequence more effectively. Therefore, the features extracted by FDA-based CHLAC are trained with HMM in [44] for robust walking speed-invariant gait recognition.

Different approaches have been used to address variations in carrying condition. The method based on spatio-temporal motion characteristics, statistical and physical parameters (STM-SPP) [7] proposed in this thesis analyses the shape of a silhouette contour using PSA at the double support phase and elliptic Fourier descriptors (EFDs) at ten phases of a gait period. The classification results by PSA and EFDs are combined using rank-summation based classifier combination rule, and a tie in ranking is resolved using contour matching based on Hu moments. STM-SPP introduces a part-based EFD analysis to address shape distortion due to carrying conditions. STM-SPP is also robust to variations in walking speed and partial occlusions, but is adversely affected by clothing variations. The method in [45] uses an iterative local curve embedding algorithm to extract double helical

signatures and analyses symmetry changes in double helical signatures at the limb region to take into account of shape distortion due to a specific carrying condition, e.g., a briefcase by an upright subject.

The method in [46] uses appearance and dynamic traits of gait by analysing the parameters of ellipses fitted to seven regions of a subject's silhouette, i.e., centroid, aspect ratio and elongation along with the subject's height for identification which is invariant to limited clothing variations and segmentation imperfections. The GEI [25] which contains spatio-temporal motion information of a gait period, is computed from binary silhouettes. Real gait templates are computed from each gait period and distorted to generate synthetic gait templates. Component and discriminant analyses are then performed on the templates for dimensionality reduction. The real and synthetic gait features thus obtained are combined using a feature fusion strategy for improved identification performance. This approach is not only computationally efficient and consumes less storage space, but also robust against noise.

The method GEI [25] captures spatio-temporal motion information of a gait period in a single GEI and the method in [28] captures temporal information of a gait sequence in a single multichannel chrono-gait image (CGI). GEI and CGI manually compute synthetic gait templates by employing a cutting and fitting scheme based on anthropometry to take into account distortions of lower body part due to carrying a briefcase, and variations in ground surface, clothes and footwear, but not distortions of upper body-part due to variations in clothing and carrying condition. A GEI is noise-resilient, and its use enables a method to be computationally less expensive in terms of time and space. CGI uses a gait period detection technique that is robust to shadows under feet and carrying a briefcase. The method GPDF in [47] uses a set of local augmented Gabor features extracted from different scales and orientations to characterise a GEI, and uses a new patch distribution feature for subject identification based on NNC. A sophisticated classifier called locality-constrained group sparse representation (LGSR) is introduced to address the presence of different covariate factors, e.g., change in ground surface and carrying a briefcase for improved identification rate at the expense of high computational complexity. The method in [48] enhances the dynamic information content of GEI by computing gait entropy image for identifying a subject in unconstrained environment with limited variations in the covariates over different days, but performs poorly in the presence of changes in viewpoint. The method in [49] computes gait flow image (GFI) from binary silhouettes using optical flow field for identifying a subject. The method GTDA-GF [50] uses a general tensor discriminant analysis (GTDA) as a preprocessing step for LDA to address undersampling due to the high dimensionality of the feature space compared to the number of training samples. GTDA-GF uses the following image representations for feature extraction: GaborS, i.e.,

the magnitude of the results of convolution of a GEI with sum of Gabor functions over five scales keeping the direction fixed; GaborD, i.e.,the magnitude of the results of convolution of a GEI with sum of Gabor functions over eight directions keeping the scale fixed; and GaborSD, i.e., the magnitude of the results of convolution of a GEI with sum of Gabor functions over five scales and eight directions. The use of GaborS, GaborD and GaborSD based gait representations enable GTDA-GF to reduce the computational complexity of Gabor function based image decomposition, as they require fewer filters than the Gabor function based gait representations. GTDA-GF performs very well in the presence of 30° variations in viewpoint and different types of footwear. It is also considerably robust to briefcase carrying conditions. The method in [51] considers gait sequences as a third-order tensor to introduce a gait representation called EigenTensorGait followed by application of linear discriminant analysis for gait recognition using multilinear PCA based tensor object recognition framework. The method is robust against limited variations in view and footwear of the subjects.

Most gait recognition methods (e.g., [7; 28; 47; 48; 49; 52]) aim to address different covariate conditions that affect gait recognition, but assume same views, i.e., lateral views of the gallery and probe gait sequences, as this view contains most of the significant gait characteristics. Therefore, to effectively identify a human subject moving freely in unconstrained directions in the real world, several gait recognition methods have recently been proposed to address variations in views.

The multi-view gait recognition methods either depend on a) extraction of gait features which are invariant to change in view [2; 53; 54; 55]; b) learning mapping or projection relationship between the gait characteristics of one view to another based on view transformation [56; 57; 58; 59], and c) view synthesis based on 3D reconstruction using camera calibration [39; 60]. A statistical feature extraction strategy is used in [53] to extract view-invariant features from the parts of a GEI which overlap between different views of a probe and a gallery subject. However, the performance of the method is not satisfactory if the GEIs of different views have little overlapping regions due to extreme variations in viewpoint. The method in [55] computes and evaluates the view-normalised trajectories of two parts of a subject's body, i.e., head and feet from the monocular video sequences. The view normalisation process involves the decomposition of walking trajectory into piece-wise linear segments to transform the head and feet trajectories from different views into fronto-parallel view based on homography. Since the feet trajectory is affected by self-occlusions, the method is applicable only to a limited variations in view.

The method in [2] determines the markerless motion of a subject's lower limb based on anatomical positions of hip, shin and ankle for view-invariant gait recognition using a viewpoint rectification approach. The method uses adaptive sequential forward floating

selection search algorithm to select the minimum set of gait features that ensures the best inter-class separability. However, the ankle of a subject is most likely to be occluded by the presence of shadows under feet. Since it is impossible to estimate the positions of hip and shin in the case of a subject either wearing a skirt or a long coat, and carrying an item in upright position, the method is also not robust against variations in clothing and carrying conditions. The method in [54] projects a gait texture image formed by averaging the binary gait images of a gait period from a certain view onto the canonical view based on domain transformation technique using transform invariant low-rank textures. Following view normalisation, PSA is used for gait feature extraction, and a subject is identified based on Procrustes distance.

The view-invariant methods based on a view transformation model, e.g., [56; 57; 61], aim to transform the gait features of a probe viewing angle into that of gallery viewing angles before a distance measure is computed for matching. The method in [56] uses discrete Fourier transform (DFT) to obtain gait features from a spatio-temporal gait silhouette volume, and applies a view transformation model on the extracted gait features for identifying a subject based on a Euclidean distance matching measure. The method in [57] creates a view transformation model using support vector regression based on local dynamic feature extraction to transform gait characteristics of one view into the required probe view, and uses Euclidean distance to measure the similarity between the subjects. Although these methods can cope with large variations in viewing angles based on a mapping relationship and without relying on camera calibration, they suffer from degeneracies and singularities caused by the gait features which are perceived in one view but not in the transformed view in the case of a large view angle difference.

To overcome the problems associated with the view transformation model, the method in [58] uses a projection relationship of normalised gait characteristics across views into a shared subspace to compute a similarity score. Similar to the view transformation model based methods, this method also relies on a mapping relationship between gait characteristics of different views, but instead of reconstructing gait features in the required probe view, the method uses canonical correlation analysis to project the gait sequences of two different views on two different maximally correlated subspaces, and uses the correlation strength as a similarity measure between the two gait sequences. However, the method does not consider the effect of variations in clothing and carrying conditions on gait recognition. The method in [59] uses joint subspace learning technique to learn a subject's prototype of different views, and represents the subject as a linear combination of these prototypes for view-invariant gait feature extraction. Radon transform is used for classifying a subject. But the methods based on mapping and projection relationships as in [56; 57; 58] depend on supervised learning, i.e., require the availability of the gait characteristics of all views

to establish a relationship between them beforehand in the training process. Hence, the unpredictability of views that a probe subject might encounter in real-life brings obvious limitation to these methods, as it is unrealistic to determine a training process which covers all possible views of real-life.

The methods in [39; 60] construct a 3D model of a subject from the 2D images captured from different views using multiple calibrated cameras. 2D gait features of an arbitrary view same as the view of the probe subject is then obtained from the 3D model for view-invariant gait recognition. The method in [60] represents the 3D pose of a subject by using a tree structure of a human skeleton, where the joints are denoted as the nodes of the tree. It also uses a truncated conic model to represent the appearance of the subject. The gait of a subject is simulated by a stick model. The method combines static gait characteristics obtained by anthropometric measurements of different body parts with the dynamic gait characteristics obtained by analysing the joint angle trajectories of lower limbs for identifying a subject based on linear time normalisation technique. In addition to variations in view, the method is also robust to self-occlusions and change in ground surface. Since the methods based on 3D structure information are only suitable for fully-controlled environment set up with multiple calibrated cameras, they are not only costly, but complicated.

The performance of a gait recognition method deteriorates if: a) the captured gait sequences are of very low resolution either due to the low resolution of the camera or large distance between the subject and the camera; and b) the inter-subject discriminative information is reduced due to the projection of gait sequences onto non-optimal low-dimensional subspace in order to reduce the dimensions of the feature space. The method in [62] thus uses superresolution with manifold sampling and backprojection to transform low resolution gait sequences into high resolution, and incorporates non-parametric multilinear tensor-based dimensionality reduction technique for improved identification rate.

A significant drop in recognition performance on the well-known public datasets, i.e., HumanID gait challenge dataset, MIT dataset, is reported by the methods in [22] when time covariate is encountered. However, there was no restrictions on the clothing of the subjects in any of these studies. Therefore, the method in [63] analysed the effect of elapsed time on gait recognition in the absence of other covariates including clothing variations, and concluded that gait successfully meets the 'permanence' crieterion of a biometrics for reliably identifying a human subject at a distance over a considerable time interval. Since the time difference encountered in existing gait databases implicitly implies the variations in clothing and footwear of the subjects, a novel temporal gait database called multi-biometric tunnel gait database [64] was created by capturing gait sequences in the multi-biometric tunnel of the University of Southampton to facilitate the analysis of effect of time on gait recognition in isolation of other challenging factors, mainly, variations in clothing. The gait

signatures extracted from three different views, i.e., top, lateral and frontal, of a GEI are concatenated and the subject is identified using NNC. Extensive experimental anlyses also confirm that elapsed time is not related with the degradation of gait recognition performance due to decrease in resolution of image sequences, and variations in clothing is the most challenging factor for model-free gait recognition methods.

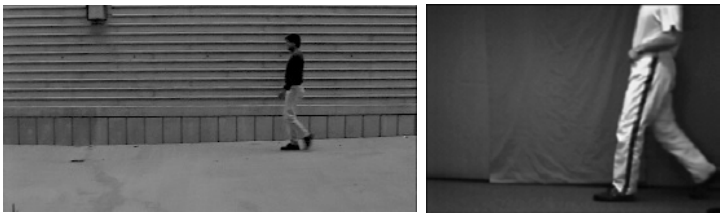
The discriminability of a subject decreases due to shape distortions caused by clothing variations of the same subject over different days. Therefore, the method in [52] decomposes a silhouette into eight different parts based on anatomical study of human subject including four overlapping parts to make a trade-off increased robustness against clothing variations and the discriminability of the subjects. DFT is used to extract gait signatures from each part, and PCA is used to reduce the dimensionality of the feature space. The subject is identified using a matching score which is computed by combining the weighted sum of z-normalised distances between each part of gallery and probe video sequences based on a probabilistic adaptive weight control strategy. The method based on radial integration transform, circular integration transform, Krawtchouk moments with genetic algorithm (RCK-G) in [65] assigns depth information captured by a calibrated stereo camera to binary silhouettes using 3D radial silhouette distribution transform and 3D geodesic silhouette distribution transform. Genetic algorithm fuses the 2D and 3D features extracted by radial integration transform, circular integration transform and weighted Krawtchouk moments. RCK-G is robust to very limited clothing variations, but not insensitive to carrying conditions. Motivated by the component-based object models for improved recognition, the method in [63] decomposes a silhouette into five parts, i.e., sub-gaits, namely upper-gait, mid-gait, lower-gait, left-gait and right-gait, and uses Baysian network to establish an intrinsic relationship between the characteristics of sub-gaits for robust gait recognition in the presence of occlusions, carrying conditions, clothing variations and change in ground surface. The correspondence-free, view-dependent method based on image self-similarity plot (SSP) in [21] obtains 2D gait signatures from 3D volumetric model that encloses the subject. The method computes SSP, i.e., the absolute correlation of each pair of images in two gait sequences. The SSPs are normalised following determination of frequency and phase of the gait period. The dissimilarity between two gait sequences are measured using pattern matching techniques, and the subject is identified based on NNC. SSP is robust to segmentation errors and limited variations in clothing, but lacks sufficient discriminatory power for robust gait recognition in realistic scenario.

The gait recognition methods discussed above are based on image sequences with standard spatial resolution, i.e., at least 128 pixel height recorded by a stationary camera with standard temporal resolution, i.e, 30 frames per second (fps). But the image sequences captured by a CCTV camera fitted in important public places are often recorded at a quite

low frame rate, e.g., 1fps-6fps due to limitations in transmission bandwidth and storage device. So the existing methods do not perform satisfactorily for the low frame rate videos captured by a CCTV camera due to sparsity of the different phases or stances in a gait period. Recently, few methods have attempted to address the challenging issue of using low frame rate videos in gait recognition. The method in [66] compares the low frame rate probe video sequences with the normal frame rate gallery video sequences based on gait period based phase synchronisation. Although the method has the potential to contribute in several criminal convictions, it fails to address the challenge of cross-camera human identification, when both the probe and gallery gait sequences are captured at low frame rates.

The method in [67] thus uses reconstruction-based temporal super resolution to construct a high frame-rate image sequence having a single gait period from a low frame-rate quasi periodic image sequence based on phase registration in energy minimization framework to identify a human subject using low frame-rate videos for both gallery and probe gait sequences. Since the method uses linear phase evolution prior for estimating initial phase, it is susceptible to wagon wheel effect (i.e., illusionary backward movement of subjects) for extremely low frame-rate video sequences, i.e., 1 or 2 fps. In addition to this, it is also susceptible to stroboscopic effect which causes the moving subject to appear stand still due coincidence of the sampling interval with the subject's gait period. The periodic temporal super resolution based method (P-TSR) in [68] successfully addresses wagon wheel effect by using a dynamic programming to directly match the low frame rate video sequence with an exemplar high frame rate video sequence for robust initial phase estimation. However, it is stll adversely affected by stroboscopic effect, and hence its performance is not satisfactory. Therefore, the method in [69] effectively combines example-based and reconstruction-based periodic temporal super resolution (ER-TSR) methods to overcome wagon wheel effect and stroboscopic effect. By introducing the standard manifold obtained from the normal frame rate training gait sequences based on a energy minimisation framework, the method successfully reconstructs the missing phases of a gait period from the nearly still probe video sequences due to stroboscopic effect.

The above-reviewed state-of-the-art model-based and model-free gait recognition methods attempt to address one or a few challenging factors of gait recognition, and hence cannot successfully identify a human subject in realistic scenario with numerous challenging factors. A few attempts have been made to identify a human subject based on low temporal resolution gait sequences, but no method has been proposed so far to take into account of low spatial as well as low temporal resolution gait sequences captured by a CCTV camera. The thesis is thus motivated to develop novel gait recognition methods to identify a human subject in realistic scenario by addressing various real-life challenges with reduced computational complexity. It introduces the use of dual-tree complex wavelet



(a) (b)

Figure 2.1: Samples from early gait datasets: (a) UCSD gait dataset; and (b) early SOTON gait dataset.

transform (DT-CWT) in gait recognition based on low spatial and low temporal resolution gait sequences. It introduces a part-based shape analysis strategy using anatomical studies of human body to achieve invariance to all common types of carrying conditions at the subject's back, with folded arms and in upright position based on a sound methodology. It exploits the higher discriminatory power of multiscale shape analysis compared to single scale by introducing a multiscale gait recognition method which is robust against most of the challenging factors of gait recognition that affect existing gait recognition systems. Thus the literature review provides a strong foundation for the development of the novel gait recognition methods presented in this thesis.

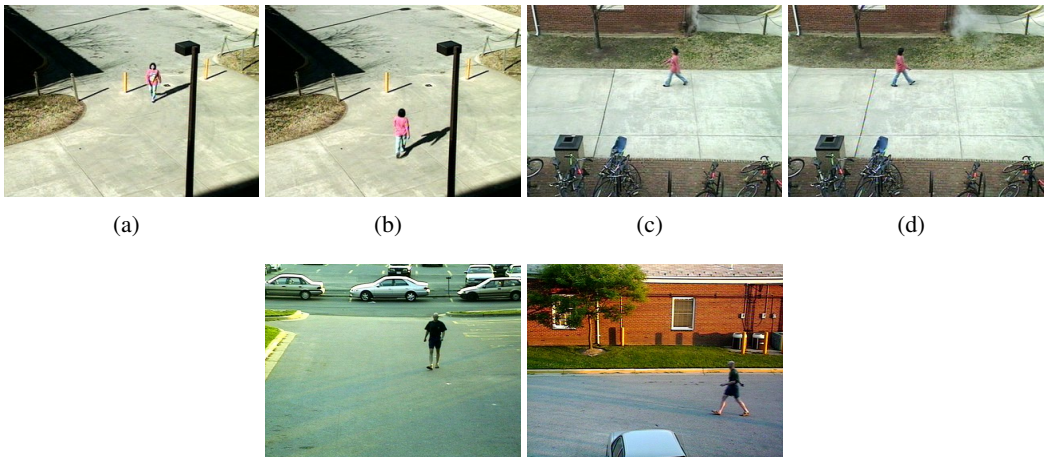
2.3 Gait Datasets

The public gait datasets are indispensable for the success and evolution of the gait recognition research, as they enable the evaluation of the performance of a gait recognition method for its uniform comparison with related state-of-the-art methods. The early gait datasets created before the year 2000 were constrained by the low but very expensive memory, and hence contain gait sequences of few subjects captured in controlled environments. The current datasets are comparatively more realistic and have benefited from the early datasets in their development [70].

2.3.1 Early Gait Datasets

UCSD Gait Dataset

The UCSD gait dataset [71] is the first public gait dataset which was formed by the Visual Computing Group of the University of California San Diego. The dataset consists of 7 sequences for each of 6 subjects walking in a circular path in an outdoor courtyard. The sequences are captured by a stationary Sony Hi8 video camera at 30 fps. Fig. 2.1(a) shows a sample subject from UCSD gait dataset.



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

(e) (f)

Figure 2.2: Row 1- sample images from UMD gait dataset-1 for four different views: (a) frontal view and walking forward; (b) frontal view and walking away; (c) fronto-parallel view and toward left; and (d) fronto-parallel view and toward right. Row 2- (e) and (f): sample images from UMD gait dataset-2.

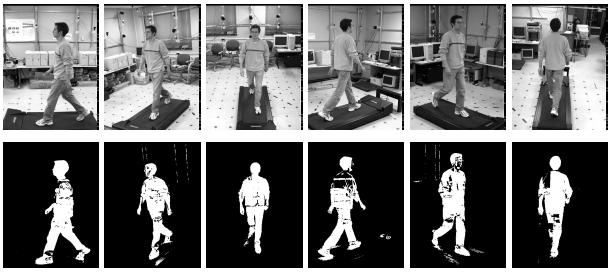
Early SOTON Gait Dataset

The early SOTON gait dataset [30] was created in 1996 at the University of Southampton with an aim to facilitate model based gait recognition by analysing the rotation patterns of limbs. Therefore, the subjects were instructed to walk against a static, cloth background wearing a special type of white trousers with a dark stripe in the middle of each leg facing outward, so that the leg closest to the camera is distinguishable from the other at all times (see Fig. 2.1(b)). The gait sequences were recorded on a video recorder using a CCD camera fixed perpendicularly to the walking path of the subjects in an indoor background with controlled illumination and later digitized.

2.3.2 Current Gait Datasets

UMD Gait Dataset

The UMD gait dataset [37] was created at the University of Maryland. It consists of two subsets: dataset-1 and dataset-2. Dataset-1 comprises gait sequences of 25 subjects captured by a Phillips G3 EnviroDome camera in an outdoor environment between the months of February and May, 2001 walking along four paths: (a) frontal view and walking forward; (b) frontal view and walking away; (c) fronto-parallel view and toward left; and (d) fronto-parallel view and toward right. Fig. 2.2(a)-(d) (row 1) shows sample images of dataset-1 for the four views obtained from http://www.umiacs.umd.edu/labs/pirl/hid/data.html.



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

Figure 2.3: Samples from CMU Mobo data set: (row 1) original sample images and (row 2) the corresponding silhouettes for six views (a) to (f).

Dataset-2 comprises gait sequences of 55 subjects walking along a "T"-shaped path of an outdoor environment recorded in two sessions in June-July, 2001. The gait sequences were captured using two Phillips G3 EnviroDome cameras placed orthogonal to each other and mounted at a height of 1.5 meters from the ground to resemble real-world visual surveillance scenario. The spatial resolution of each camera was 640×480, and they operated at 25 fps. The sample images of dataset-2 are shown in Fig. 2.2(e)-(f) (row 2), which are obtained from http://www.umiacs.umd.edu/labs/pirl/hid/data.html.

CMU MoBo Gait Dataset

The Carnegie Mellon University (CMU) Motion of Body (MoBo) dataset [72] comprises gait sequences of 25 subjects performing four types of walk: slow walk (walking speed: 2.06 mph); fast walk (walking speed: 2.82 mph); walk holding a ball (walking speed: 2.04 mph); and walk on an inclined plane of a treadmill (walking speed: 1.96 mph). Each sequence is of approximately 11 seconds duration and is recorded at 30 fps from six different views. The sequences were captured on a same day using six high resolution calibrated cameras evenly distributed around the treadmill. Row 1 of Fig. 2.3(a)-(f) shows the original images of a subject from CMU MoBo dataset at six views, while row 2 shows the extracted silhouettes at the corresponding views.

USF HumanID Gait Challenge Dataset

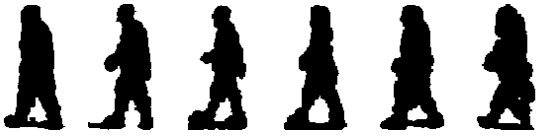
The USF HumanID gait challenge dataset [22] has two versions: (a) the small version consisting of 452 gait sequences of 74 subjects acquired in May only; and (b) the full version consisting of 1870 sequences of 122 subjects acquired in May and November. The data set comprises sequences of subjects walking along elliptical paths in an outdoor environment in front of two cameras with the following five covariates: walking surface (grass (G) or concrete (C)); shoe type (A or B); viewpoint (right (R) or left (L)); carrying conditions

Table 2.1: The specification of the gait challenge experiments for the small version of USF HumanID gait challenge dataset [22].

Exp.	Probe	No. of subjects	Covariate
A	(G, A, L)	71	View
B	(G, B, R)	41	Shoe
C	(G, B, L)	41	Shoe, View
D	(C, A, R)	70	Surface
E	(C, B, R)	44	Surface, Shoe
F	(C, A, L)	70	Surface, View
G	(C, B, L)	44	Surface, Shoe, View

Table 2.2: The specification of the gait challenge experiments for the full version of USF HumanID gait challenge dataset [22].

Exp.	Probe	No. of subjects	Covariate
A	(G, A, L, NB, M/N)	122	View
B	(G, B, R, NB, M/N)	54	Shoe
C	(G, B, L, NB, M/N)	54	Shoe, View
D	(C, A, R, NB, M/N)	121	Surface
E	(C, B, R, NB, M/N)	60	Surface, Shoe
F	(C, A, L, NB, M/N)	121	Surface, View
G	(C, B, L, NB, M/N)	60	Surface, Shoe, View
H	(G, A, R, BF, M/N)	120	Briefcase
I	(G, B, R, BF, M/N)	60	Shoe, Briefcase
J	(G, A, L, BF, M/N)	120	View, Briefcase
K	(G, A/B, R, NB, N)	33	Time, Shoe, Clothing
L	(C, A/B, R, NB, N)	33	Surface, Time, Shoe, Clothing



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

Figure 2.4: Sample silhouettes of different subjects from USF Gait Data set: (a) and (b) walk on grass with different shoe types and viewpoints; (c) and (d) walk on concrete with different shoe types and viewpoints; and (e) and (f) walk on grass carrying a briefcase.

(a) (b) (c)

Figure 2.5: Samples from CASIA gait dataset A: row 1- original images of different subjects for views (a) 0°, (b) 90° and (c) 45°; row 2- silhouettes of a subject for the three views.

(carrying a briefcase (BF) or not carrying a briefcase (NB)); and elapsed time between the acquisition of the sequences (May (M) or November (N)) [22]. Twelve experiments of increasing difficulty are designed as shown in Table 2.2 to investigate the robustness of a gait recognition method against the five covariates. The gait sequences are captured at 30 fps, and the spatial resolution of each silhouette is 128×88. The thirty three subjects that are common in the May and November data sets account for time covariate. There are no common sequences between the gallery and the probe sets, and all subjects did not participate in all experiments [25; 22]. Table 2.1 shows the probe sets for each of the gait challenge experiments of the small version of USF HumanID gait challenge dataset. It also includes the number of participating subjects in each of these experiments, and the associated covariate factors. Similarly, Table 2.2 shows the probe sets for each of the gait challenge experiments of the large version of USF HumanID gait challenge dataset, number of participating subjects in each of these experiments and the associated covariate factors. Fig. 2.4 shows the sample silhouettes of different subjects from USF HumanID gait challenge dataset.

CASIA Gait Dataset

The CASIA gait dataset consists of the four subsets, i.e., dataset A (created in 2001), dataset B (created in 2005), dataset C (created in 2005) and dataset D (created in 2009).

- Dataset A (standard dataset) [29]: This dataset consists of video sequences of 20 subjects captured by a digital camera (Panasonic NV-DX100EN) fitted on a tripod in an outdoor background on two different days. The video sequences of the subjects walking along a straight-line path are recorded at a rate of 25 fps from three different views, i.e., lateral (0°), frontal (90°) and oblique (45°) w.r.t. the image plane. There are 4 sequences corresponding to each of the three views of a subject resulting in a total of 240 gait sequences. The number of frames of a gait sequence varies from 37 to 127 based on the subject's walking speed and the resolution of each frame (24-bit full colour) is 352×240. The dataset includes 19139 frames and its size on disk is approximately 2.2GB. Fig. 2.5 shows sample images and silhouettes of CASIA gait dataset A.

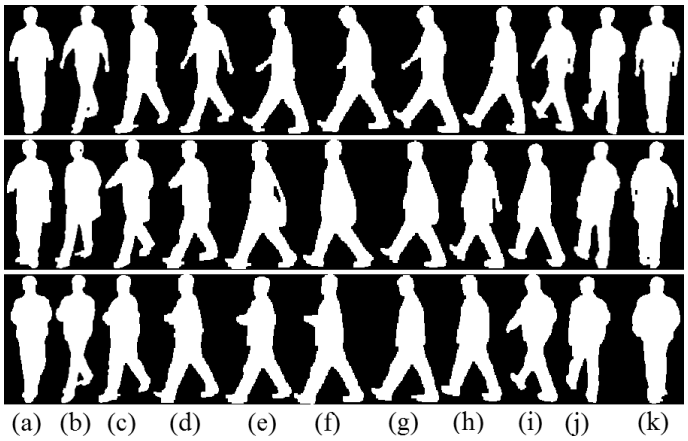
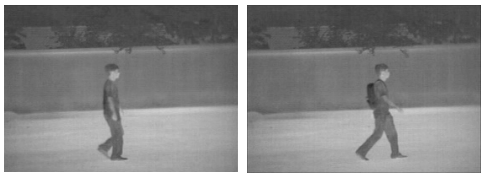


Figure 2.6: Sample silhouettes of a subject from CASIA gait dataset B for normal walking (row 1), walking with a bag (row 2) and walking wearing a coat (row 3) for 11 views: (a) 0°; (b) 18°; (c) 36°; (d) 54°; (e) 72°; (f) 90°; (g) 108°; (h) 126°; (i) 144°; (j) 162°; (k) 180°.

- Dataset B (multiview gait dataset)[73]: This dataset consists of video sequences of 124 subjects (31 females and 93 males) captured by tripod-fitted 11 USB cameras (Fametech 318SC) from 11 views in the range [0°,180°] with a difference of 18° between two adjacent views. There are 10 video sequences for each view of a subject: 6 sequences for normal walking, i.e., without wearing a coat or carrying a bag; 2 sequences wearing a coat; and 2 sequences with carrying either a knapsack, a satchel, or a handbag. The video sequences are recorded indoor at a rate of 25 fps and the resolution of each frame is 320×240. The dataset includes 13640 video sequences and its size on disk is approximately 17 GB. Fig. 2.6 shows sample silhouettes of a subject from CASIA gait dataset B for normal walking, walking with a bag and walking wearing a coat for 11 views.



(a) (b)

Figure 2.7: Sample images of a subject from CASIA gait dataset C: (a) walking without a bag at normal speed; and (b) walking with a bag at normal speed.

- Dataset C (infrared gait dataset) [74]: This dataset consists of lateral-view video sequences of 153 subjects (23 females and 130 males) captured by a thermal infrared camera in an outdoor environment at night. Each subject has 10 video sequences: 4 sequences for normal walking; 2 sequences for walking with a bag at normal speed; 2 sequences for slow walk; and 2 sequences for fast walk. The dataset includes 1530 video sequences and its size on disk is approximately 17 GB. Sample images of a subject from this dataset walking at normal speed with and without carrying a bag are shown in Fig. 2.7.

- Dataset D (gait and its corresponding footprint dataset) [75]: This dataset consists of cumulative foot pressure images of subjects walking on a floor equipped with a RScan USB foot pressure sensing device to test an algorithm's efficiency against variations in walking speed and shoes. To take into account of walking speed variations in the dataset, foot pressure images of 88 subjects (20 females, 66 males) aged between 20 and 60 are recorded for two different types of walk: normal walk and fast walk. Each subject walked 5 times on the foot pressure sensing device in an indoor background without wearing shoes for each type of walk resulting in $10 \times 88 = 880$ foot pressure recordings consisting of $3 \times 880 = 2640$ images. To take into account of shoe variations in the dataset, foot pressure images of 30 young Chinese subjects (6 females and 24 males) aged between 20 and 40 are recorded. Each subject walked on the foot pressure sensing device twice without wearing shoes and 4 times wearing three different types of shoes: running shoes, Chinese cloth shoes and leather shoes, resulting in $30 \times 6 \times 3 = 540$ images. Fig. 2.8 which shows the set-up for the dataset and the accumulated footprints of a subject, is obtained from http://www.cbsr.ia.ac.cn/.

OU-ISIR Treadmill Gait Dataset

The OU-ISIR gait dataset [76] captured at the Institute of Scientific and Industrial Research (ISIR), Osaka University (OU) consists of gait sequences of human subjects walking on

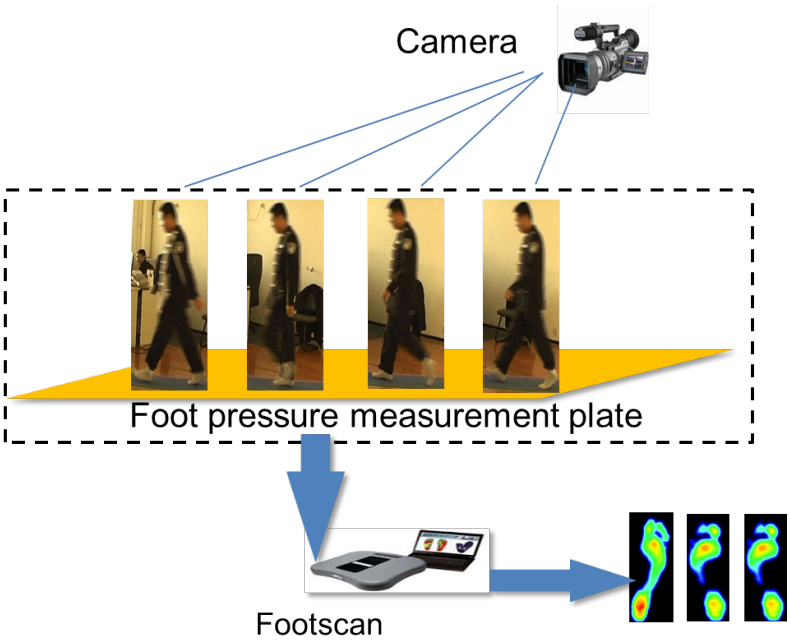


Figure 2.8: Set-up for CASIA gait dataset D.

(a) (b) (c)

Figure 2.9: Sample images of a subject from OU-ISIR gait dataset A walking at speeds: (a) 2 km/h; (b) 4 km/h; and (c) 7km/h.

a treadmill. The dataset consists of four subsets, i.e., dataset A, dataset B, dataset C and dataset D to respectively facilitate the evaluation of gait recognition methods in the presence of speed variations, clothing variations, view variations and gait fluctuation. Since the dataset considers large varieties of ages of both sexes, it is also suitable for gait based age group and gender classifications.

- Dataset A (speed variation): It consists of gait sequences of 34 subjects (26 males and 8 females) with 9 variations in walking speed in the range 2 km/h and 10 km/h with an interval of 1 km/h [43]. The subjects are video recorded from lateral view while walking (for speeds of 2km/h to 7km/h) and jogging (for speeds of 8km/h to 10km/h) on a speed-controllable treadmill. The spatial and temporal resolution of each image are respectively 640×480 and 60 fps. Fig. 2.9 shows sample images of

Table 2.3: List of clothes used in OU-ISIR gait dataset B [52] (Abbreviation: description).

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

Table 2.4: Different clothing combinations[52] (#: Clothing combination type, S_i: i th clothes slot).

#	S_1	S_2	S_3	#	S_1	S_2	#	S_1	S_2
2	RP	HS	-	A	RP	Pk	T	Sk	FS
3	RP	HS	Ht	B	RP	Dj	U	Sk	Pk
4	RP	HS	Cs	I	BP	Hs	V	Sk	Dj
9	RP	FS	-	K	BP	FS	D	CP	HS
X	RP	FS	Ht	J	BP	LC	F	CP	FS
Y	RP	FS	Cs	L	BP	Pk	E	CP	LC
5	RP	LC	-	M	BP	DJ	G	CP	Pk
6	RP	LC	Mf	N	SP	HS	H	CP	DJ
7	RP	LC	Ht	Z	SP	FS	O	CP	CW
8	RP	LC	Cs	P	SP	Pk	R	RC	RC
C	RP	DJ	Mf	S	Sk	HS	-	-	-

a subject from this dataset walking at varying speeds.

- Dataset B (clothing variation): It consists of gait sequences of 68 subjects with up to 32 different types of clothing combinations. The different types of clothes used in this dataset are listed in Table 2.3, while Table 2.4 shows the different clothing combinations used in constructing the dataset [52].

  The dataset is divided into three parts, i.e., a training set consisting of 446 sequences of 20 subjects with all types of clothes; a gallery set consisting of gait sequences of remaining 48 subjects with standard clothes, i.e., type 9/ regular pants+full shirt (see Table 2.4); and a probe set consisting of 856 sequences for these 48 subjects with other types of clothes excluding the standard clothes. The sample images for all combination of clothing types used in this dataset are shown in Fig. 2.10.

- Dataset C (view variation): It consists of gait sequences of 168 subjects comprising 88 males and 80 females of ages 4 to 75 years old from 25 views (12 azimuth

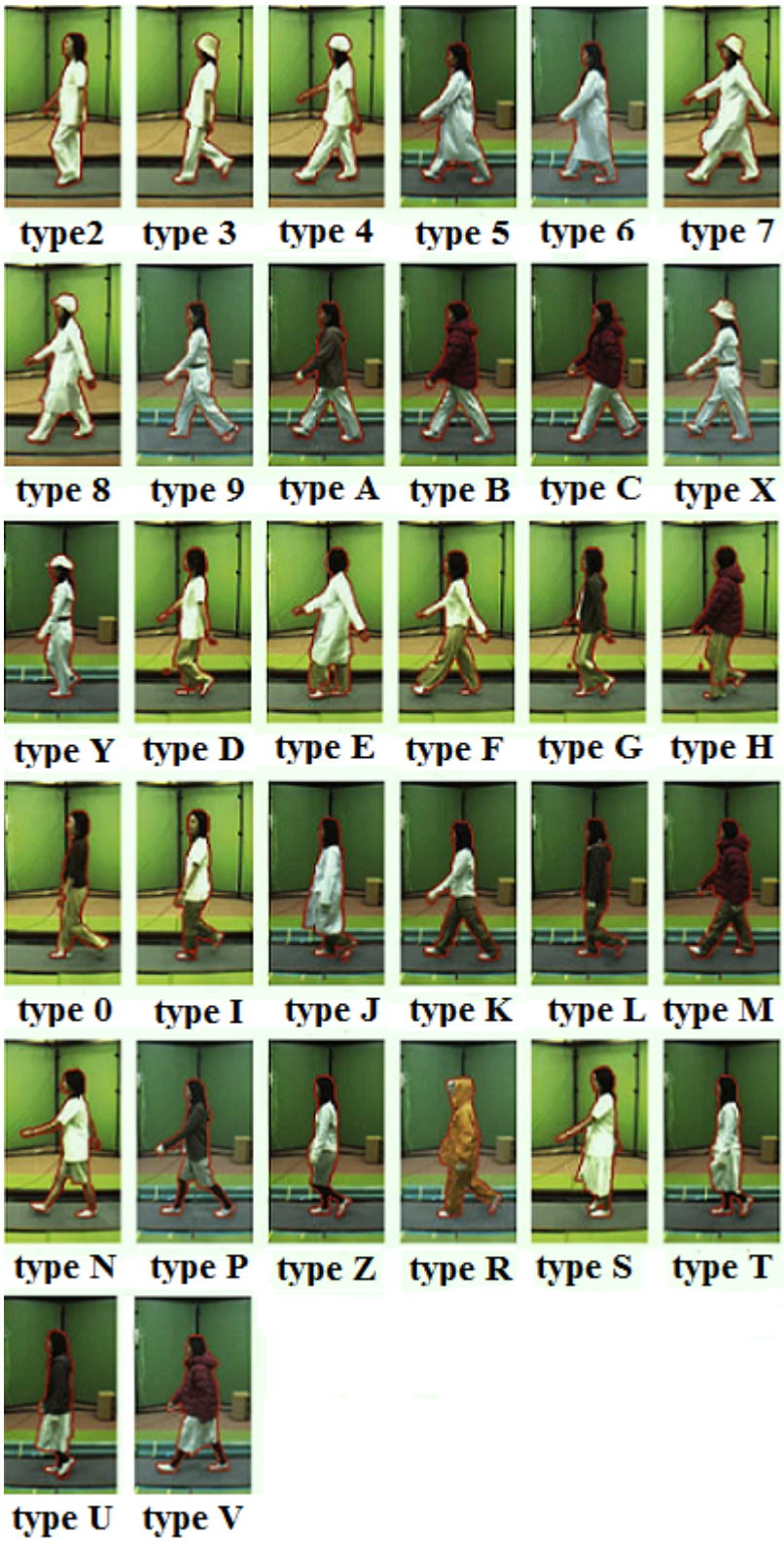


Figure 2.10: Sample images for all types of clothing combinations used in OU-ISIR gait dataset B [52].

directions × 2 tilt angles and one top view) [77]. The gait sequences are captured using 25 calibrated, synchronised cameras at 60 fps of VGA resolution placed around a treadmill. The treadmill is fitted with six screens to facilitate chroma-key background subtraction.

- Dataset D (gait fluctuation): The dataset consists of 370 lateral-viewed gait sequences of 185 subjects captured at 60 fps [76]. This dataset enables the evaluation of the robustness of gait recognition methods against gait fluctuations, i.e., the variations in the silhouette shape at similar phases of different gait periods in a gait sequence due to change in frame-rate. The gait fluctuations are measured in terms of normalized auto correlation (NAC) of size-normalised silhouettes on temporal axis. The NAC value is inversely proportional to the gait fluctuations. Hence, the dataset is divided into two subsets based on NAC values: (a) DB_{high} consisting of 100 subjects with high NAC values denoting stable gait; and (b) DB_{low} consisting of 100 subjects with low NAC values denoting fluctuated gait. The spatial resolution of the original silhouettes are 128×88, and the recording time of each sequence is approximately 6 secs.

University of Southampton multi-biometric tunnel 3D gait dataset

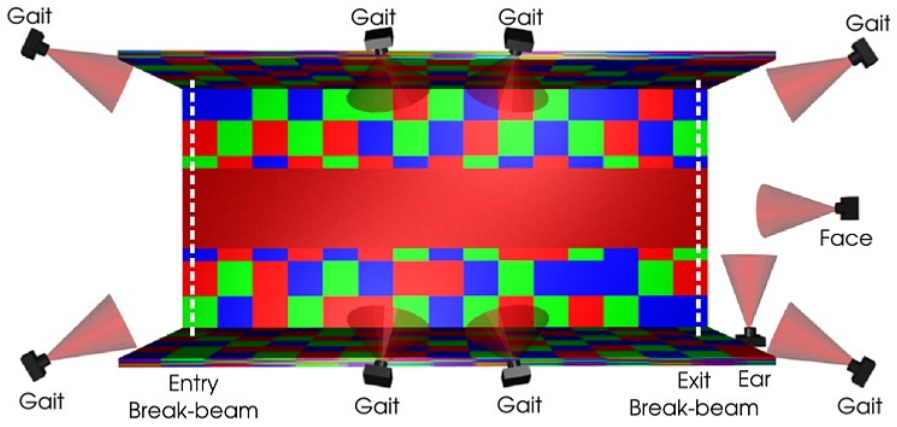


Figure 2.11: Placement of cameras and the break-beam sensors in the multi-biometric tunnel of University of Southampton [64].

The multi-biometric tunnel [64] at the University of Southampton provides a constrained environment similar to an airport for allowing subjects walking through it, while capturing their multiple biometrics, i.e., gait, face and ear in an unobtrusive way for automatic human identification. The system has the potential to create a large dataset consisting of multiple biometrics of a subject efficiently in a very short time period for real-life practical applications [64].

Figure 2.12: Gait images captured by the synchronized cameras in the multi-biometric tunnel of University of Southampton [64].

As the subject walks through the tunnel, his/her video is recorded at 30 fps by eight synchronized Point Grey Dragonfly cameras of resolution 640×480 connected by an IEEE1394 network for 3D gait recognition. The face and ear biometrics are captured by two very high resolution, i.e., 1600×1200 IEEE1394 cameras, which are placed at the exit of the tunnel. The walls of the tunnel are painted with a non-repeating rectangular grids of different colours to aid automatic camera calibration.

A subject starts walking as the status light indicator mounted at the entry of the tunnel shows a signal indicating that the system is ready for the capturing process, and upon walking through the tunnel, the indicator will show a busy signal until it is ready for the next subject. As the subject enters into the tunnel, he/she walks through a break-beam sensor commencing the video capturing, and there is another break-beam sensor located at the exit of the tunnel which terminates the capturing process. Fig. 2.11 shows the placement of the cameras and the break-beam sensors inside the tunnel, and Fig. 2.12 shows the gait images captured by the eight synchronised cameras.

OU-ISIR large population dataset

The OU-ISIR dataset is the largest gait dataset which consists of gait sequences of 4007 subjects (1872 females and 2135 males) of ages ranging from 1 to 94 years [78]. The dataset consists of two subsets: OULP-C1V1-A and OULP-C1V1-B. OULP-C1V1-A consists of two gait sequences for each subject, and facilitates performance evaluation of gait recognition methods in terms of variations in clothing and view, but not walking speed. OULP-C1V1-B consists of one gait sequence for each subject, and facilitates the investigation of gender classification and age estimation based on gait analysis.

OULP-C1V1-A and OULP-C1V1-B are also further subdivided into five subsets, i.e., A/B-55, A/B-65, A/B-75, A/B-85, and A/B-ALL based on viewpoints of 55°, 65°, 75°, 85° and all four angles, respectively, with each subject belonging to at least one of these

subsets. The gait sequences of the subjects walking along a 10 m long straight path are captured using Flea2 camera with HF3.5M-2 lens at 30 fps in an indoor environment. The size of each image is 640×480 pixels.

The early gait datasets made significant contributions to give a direction to gait recognition research in its infancy, as they sufficed to establish that a human subject can be identified by his/her gait. However, they are largely superceeded by the current datasets recently. This is because the current datasets include much larger number of subjects in presence of various covariate factors to cope up with the real world challenges for the maturity of this field [70].

Chapter 3

Shape Analysis of Silhouette Contours using Statistical and Physical Parameters

3.1 Introduction

The accuracy of a gait recognition method increases with the number of appropriate gait signatures considered, and the spatio-temporal deformation of a subject's shape provides more significant gait characteristics than its kinematics based on model fitting[23]. Therefore, this chapter introduces a gait recognition method that effectively combines three types of model-free methods, i.e., spatio-temporal motion based, statistical and physical parameter based (referred to as STM-SPP) for improved classification rate with reduced computational complexity using a simplified feature space. Spatio-temporal motion-based methods (e.g., [22; 25; 29; 79; 80; 81; 82]) analyse both the spatial structural and temporal transitional characteristics of gait [5]. These methods are easy to implement with low computational complexity using a simplified feature space. However, they are susceptible to variation in camera view and walking speed. Statistical methods (e.g., [38; 83; 84; 85]) usually describe silhouettes using shape and motion descriptors such as velocity moments [83], Zernike velocity moments [84] and Procrustes mean shape distance [38]. Eigenspace transformation and canonical space analysis are widely used in these methods to reduce the dimensionality of input feature space and optimise class discrimination. Statistical methods are resilient to noise [5]. The physical parameter based methods (e.g., [19; 20]) estimate the subject's geometrical and structural properties, e.g., step length, cadence and height. These methods are robust against lighting variation and segmentation imperfections. However, they require camera calibration, body-part labelling and depth compensation [5].

Shape analysis by comparing landmarks has been applied in diverse fields [86], e.g., study of shape differences of brains to identify schizophrenic patients, handwriting recognition, fish recognition, robotic harvesting of mushrooms, and study of shape and size variability of microfossils, using traditional or geometrical methods. Traditional shape analysis methods either examine the ratios of distances or the angles between landmarks, whereas the geometrical methods analyse coordinates of the landmarks. One of the major motivations for STM-SPP (proposed in this thesis) is to demonstrate the potential of traditional shape analysis for identifying human subjects based on their gait signature. With regards to a subject's identity, the temporal changes of the subject's shape in a gait sequence provides better discriminative power than the discrete snapshots of images, but it increases the computational complexity. The performance of any shape-based gait recognition method degrades with variation in hair style, clothing and footwear over different days, as these factors distort the silhouette shape. The shape of a silhouette is also significantly altered due to carrying conditions. If either of the gallery or probe subject carries any item, certain parts of the silhouette shape are likely to change and the discriminative ability of the shape-based gait recognition algorithm decreases with respect to these parts.

STM-SPP extends the application of a traditional shape analysis method in gait recognition by comparing distances of specific landmarks from the centre of mass of the subject's silhouette contour (COM-SC) using PSA. The purpose of using these distances rather than coordinates of the landmarks is to reduce the dimensionality of the feature space from 2D to 1D. STM-SPP also analyses the spatio-temporal deformation of contours' shape using EFDs [87] and provides an efficient means of obtaining dissimilarity score between gallery and probe sequences using PSA and EFDs for subject classification. The method validates similar dissimilarity scores obtained by PSA using spatio-temporal gait characteristics and physical parameters of human body to enhance the classification rate in the presence of across-day gait variation (e.g., walking speed, different types of clothes and footwear, and change of hair style). The method also provides an experimentally supported insight into the detection procedure of small carried items based on anatomical studies of human body and introduces a part-based EFD analysis to achieve robustness against shape variation due to carrying conditions among the probe and gallery subjects (as detailed in Section 3.3.3). To utilise the benefit of shape sequence processing with reduced processing time, STM-SPP characterises the subject's contours using EFDs at specific phases of a gait period to considerably reduce computational complexity as well as to achieve robustness against walking speed variation and missing or distorted frames (as detailed in Section 3.3.3). The output of the two classifiers (PSA and EFDs) combined by using rank-summation based combination rule is used for identifying the subject, with the application of hierarchical contour matching based on Hu moments [88] to resolve the case of two

classes with the same combined rank. The method is defined on the lateral (i.e., profile) view of silhouettes of a subject walking parallel to the image plane, as most of the significant gait characteristics are captured in this view. Instead of analysing the silhouette, only its boundary (i.e., contour) is considered to reduce the computational complexity of STM-SPP.

STM-SPP has the following novelties: (1) it effectively combines spatio-temporal motion based, statistical and physical parameter based gait recognition methods with low computational complexity for improved subject identification; (2) it introduces the application of traditional shape analysis in gait recognition using 1D feature space based on PSA; (3) it validates the similar dissimilarity scores obtained by PSA using spatio-temporal gait characteristics and physical parameters of a subject to achieve robustness against walking speed variation, limited clothing variation and change in hair style; (4) it provides an in-depth analysis of different phases of a gait period to introduce a process of extracting its ten phases that capture most of the significant gait characteristics; (5) it analyses the spatio-temporal deformation of the shape of a subject's contour uisng EFDs at ten phases of a gait period to achieve robustness against variation in walking speed and missing/distorted frames; (6) it introduces a part-based EFD analysis to achieve invariance to shape distortions due to small carrying conditions with folded hands and in upright position; and (7) experimental analyses on public datasets demonstrate the efficacy of STM-SPP.

3.2 Related Work

STM-SPP is compared with the model-free gait recognition methods in [19; 21; 22; 26; 29; 89]. The method SSP [21] uses pattern classification techniques for subject identification. Spectral partitioning is performed in CMU2 [89] and human identification is achieved using weighted correlation and median weighted distance. The method CASIA in [29] converts a binary silhouette into 1D normalised distance signal and uses PCA for reducing dimensionality of the feature space. The correspondence-free method in [20] computes the gait period of a subject by analysing width of the bounding box which encloses the subject's moving silhouette and uses Bayesian classifier to confirm the subject's identity. However, silhouette width is not effective for computing gait period of the frontal view of a moving subject. Hence, temporal change of silhouette height is also considered in [19] to achieve robustness against different views. This method performs template matching between key frames of the gallery and probe gait sequences using normalised correlation to obtain correlation scores, and uses NNC for gait recognition.

The DNGR method in [26] normalises gait dynamics using population HMM and computes shape distances between stance phases using LDA to maximise inter-class and

minimise intra-class variation of the subjects. The method in [90] recognises the periodic movements of a human subject using motion power spectral analysis of the Fourier coefficients of unstructured feature-point kinematic data acquired from a marker-based 3D optical motion capture system. The feature extraction process in [82] involves the computation of angular distance between the foreground pixels and the centre of the silhouette. The method uses linear time normalisation to determine the subject's identity. The method in [91] represents deforming shape sequences of human subjects over a gait period by 2D discrete Fourier series and uses the resulting magnitude spectra to form the gait signatures. Motivated by the encouraging identification rate of this method, STM-SPP analyses the shape of silhouette contours at ten phases of a gait period with EFDs to form gait signatures for subject classification.

STM-SPP compares favourably with the shape sequence matching based method (referred to as SSM) in [23] and the method in [38] that also use PSA for gait recognition. SSM uses both parametric and nonparametric methods to compare deforming silhouettes for human identification and activity recognition. Human shapes are described as k-dimensional complex vectors, and Procrustes distance metrics are used to compute distances between the shape sequences. Although we consider the distances of specific landmarks from COM-SC to compare human shapes at specific phases of a gait period to simplify the feature space, a specific frame-wise comparison may not always produce satisfactory identification due to distortions of silhouettes, e.g., caused by partial occlusions. We thus characterise the contour of a silhouette at specific phases of a gait period by translation-rotation-scale invariant EFDs and perform subject classification based on a dissimilarity score to exploit the benefit of shape sequence processing without compromising the simplicity in implementation and simplification of feature space.

The PSA in [38] describes the boundary points of the silhouette as a vector of k complex numbers in 2D shape space called a configuration and compares two such configurations to measure similarity using the Procrustes mean shape distance. To ensure the same set of boundary points in different images are used for comparison, an interpolation technique with point correspondence analysis is used. We consider twenty eight landmarks based on anatomical and geometric properties of human body on the silhouette contour. Instead of considering the coordinates of landmarks for comparing shapes, we use the distances of the landmarks from COM-SC which correspond to the different rows of the uni-columnar configuration matrix. Identification is achieved by analysing the average dissimilarity score obtained by comparing the probe configuration matrix with the gallery configuration matrices of multiple sequences of same subjects from the gallery dataset.

The method in [38] employs classifier combination rules to combine the static and dynamic gait characteristics obtained respectively by analysing silhouettes' shapes and joint

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

Figure 3.1: Extraction of a moving silhouette: (a) background image; (b) original image; (c) segmented regions; (d) smoothed segmented region; (e) binary silhouette; (f) silhouette contour; and (g) contour after polygon approximation.

angle trajectories of lower limbs. STM-SPP combines the classification results obtained by PSA and EFDs of silhouette contours using rank-summation based classifier combination rule for its simplicity and effectiveness.

3.3 Proposed method: STM-SPP

The proposed method, STM-SPP, comprises three modules: module 1 extracts and postprocesses subject's silhouettes; module 2 classifies the subject using PSA (phase 1) and using EFDs (phase 2); and module 3 combines the two classification results.

3.3.1 Module 1: Silhouette extraction and postprocessing

Silhouette extraction involves segmenting region/s that correspond to a walking subject in a cluttered scene. STM-SPP employs background modelling and moving object segmentation in [92; 93], where background is considered to be any static or periodically moving parts of a scene that remains static or periodic over the period of interest. The segmented regions are smoothed using Gaussian filter and subjected to connected-component analysis involving morphological operation of dilation to remove noisy pixels and followed by erosion to fill up any small holes inside the silhouette to give a single connected region. The smoothed segmented region is then tracked based on the overlap of the centroid of the bounding rectangle which encloses the region in the subsequent frames as in [93]. The process is illustrated in Fig. 3.1(a)-(d). The tracked segmented region is binarised using 2D Otsu automatic thresholding technique [94], which utilises both the grey level information of each pixel and its spatial correlation information within the 2D neighbourhood to outperform the Otsu method [95] in the presence of noise for extracting the subject's silhouette as illustrated in Fig. 3.1(e). The extreme outer boundary of the largest connected component, i.e., the silhouette contour as shown in Fig. 3.1(f), is obtained using the sequence of vertices traversal algorithm based on connectivity [93].

3.3.2 Phase 1 of Module 2: PSA

The successful identification of a subject should not depend on how far the subject is from the camera and the direction of walking. Thus, the shape feature vector used for identification must be invariant to scale, translation and rotation. One means of achieving this is through PSA, which involves analysing the distribution of a set of shapes by matching configurations (where each configuration is a set of geometric locations of landmarks of a shape) to calculate the best shape-preserving Euclidean transformations, using least squares techniques. The 2D Cartesian moment of order p and q of a contour $I(x, y)$ is

$$m_{p,q} = \sum_{i=1}^{N} I(x, y) x^p y^q, \tag{3.1}$$

where N is the total number of pixels in the contour. The coordinates of COM-SC, (x_c, y_c), are given by the ratio of first-order to zero-order contour moments [96], i.e.,

$$x_c = \frac{m_{10}}{m_{00}} \quad , \quad y_c = \frac{m_{01}}{m_{00}}. \tag{3.2}$$

We assign anatomical landmarks by considering the different positions of the human body joints as a fraction of subject's height (H), obtained by measuring the height of the bounding rectangle which encloses the silhouette contour. The vertical positions of ankle (A), knee (K), hip (HI), chest (C), shoulder (SH) and head (HD) are then estimated as a fraction of the body height following anatomical studies in [97] as $0.039H$, $0.285H$, $0.530H$, $0.720H$, $0.818H$ and $0.870H$ measured from the bottom of the bounding rectangle, respectively. The boundary points of the contour that correspond to A, K, HI, C, SH and HD are located (labelled as 11, 12, 17, 19, 9, 14, 15, 21, 7, 23, 5, 25, 4, 26, 3 and 27 in Fig. 3.2) and are treated as anatomical landmarks.

Two mathematical landmarks are also defined as the end points of the contour diameter (Diam) joining two farthest boundary points (labelled as 1 and 18 in Fig. 3.2), i.e., [98]

$$\text{Diam(Contour)} = \max[\text{Dist}(q_i, q_j)], \tag{3.3}$$

where Dist(.,.) computes the distance between two boundary points q_i and q_j. For better results, ten additional pseudo landmarks are also considered (labelled as 2, 28, 6, 24, 8, 22, 10, 13, 16 and 20 in Fig. 3.2), each of which is equi-distantly spaced between the anatomical landmarks. The twenty eight landmarks are traversed in anticlockwise direction starting from landmark 1 with respect to COM-SC, in the double support phase of the gait period when both feet are almost flat on the ground and farthest from each other as shown in Fig. 3.2, resulting in the maximum width of the bounding rectangle of the contour.

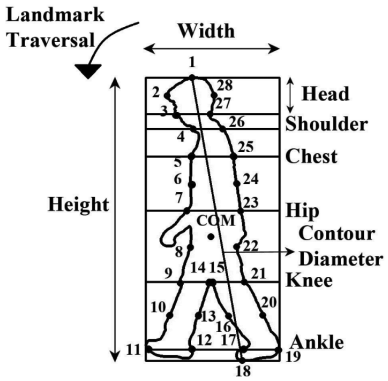


Figure 3.2: Vertical positions of ankle (A), knee (K), hip (HI), chest (C), shoulder (SH) and head (HD) as a fraction of body height. Positions of COM-SC, landmarks 1 to 28, contour diameter and direction of landmark traversal for one subject from CMU MoBo dataset.

Therefore, the width of the bounding rectangle in each frame of a gait period is measured in terms of pixel units and the frame which corresponds to the maximum width of the bounding rectangle is considered as the subject's double support phase.

The presence of shadows under feet distort the contours near the feet and thus bring challenges to the estimation of landmarks at the toe and ankle. Therefore, in the case of datasets containing shadows under feet, STM-SPP encloses the silhouettes using a bounding rectangle and estimate the region-of-interest (ROI) from the bounding rectangle having identical width but slightly reduced height αH to discard the shadows under feet, where H is the height of the bounding rectangle and α is a fraction. The value of α is experimentally set to be 0.9375 for USF dataset [27] and $0.990H$ for CMU MoBo dataset. The estimated ROI is then copied to a destination image of fixed height, i.e., height-normalised, for all the subjects of CMU dataset to remove camera depth variations. However, the silhouettes provided by the USF HumanID dataset are already cropped, centre-aligned and normalised to a fixed size 128×88. Thus, for the USF dataset, we do not need to perform any normalisation of a silhouette after cropping its height to remove the shadows.

The distance dl_i between each landmark (x_i, y_i) and COM-SC is given by

$$dl_i = [(x - x_i)^2 + (y - y_i)^2]^{\frac{1}{2}}. \tag{3.4}$$

These distance values as a function of equally-spaced monotonically increasing positions along contours form the gallery and probe shape signals, and are labelled with integers that correspond to landmarks (denoted by solid circles) in Fig. 3.3(b) and (c).

The dimension of the configuration matrices corresponding to gallery and probe sequences is $k \times m$, where k=28 is the number of landmarks and m=1 is the number of dimensions of the landmarks. PSA is used to measure the dissimilarity between two such

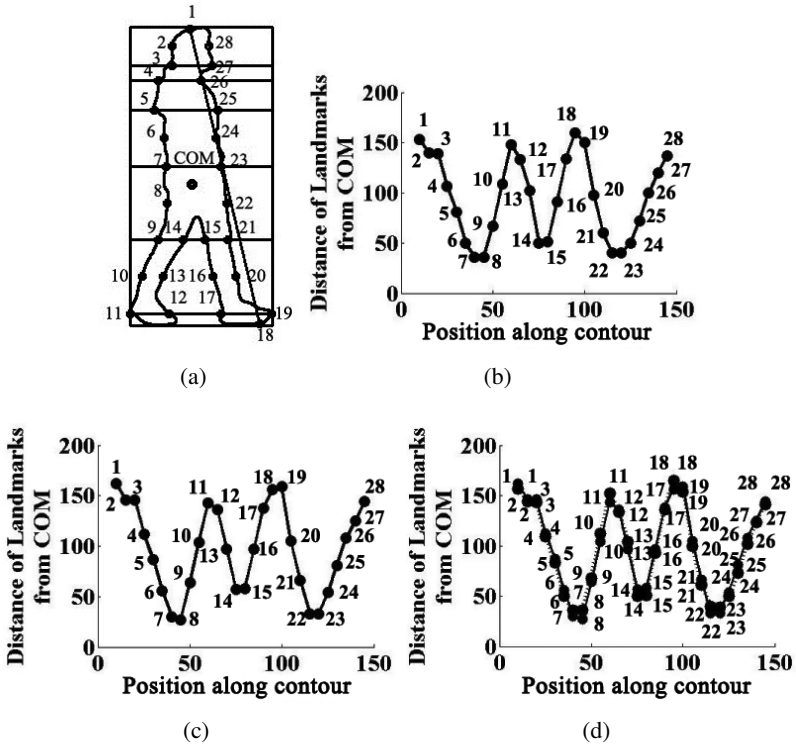


Figure 3.3: Silhouette representation: (a) Positions of COM-SC and landmarks 1 to 28; (b) and (c) are respectively probe and gallery shape signals consisting of distances of landmarks from COM-SC in anti-clockwise direction; (d) Superimposition of the transformed probe shape signal (dashed line) on the gallery shape signal (bold line) for visualizing differences between landmark distances from COM-SCs of probe and gallery sequences of the same subject.

configuration matrices to achieve view, rotation and scale invariance. Firstly, the gallery and probe configuration matrices ($\mathbf{S}_1$ and $\mathbf{S}_2$, respectively) are centred using [86]

$$\mathbf{S}_{1c} = \mathbf{C}\mathbf{S}_1 \qquad , \qquad \mathbf{S}_{2c} = \mathbf{C}\mathbf{S}_2, \tag{3.5}$$

where centring matrix $\mathbf{C} = \mathbf{I}_k - \frac{1}{k}\mathbf{1}_k\mathbf{1}^{\mathrm{T}}{}_k$, $\mathbf{I}_k$ is a $k \times k$ identity matrix, $\mathbf{1}_k$ is a k-dimensional vector of ones, and T is the transpose operator. The centred probe configuration matrix $\mathbf{S}_{2c}$ is then subjected to PSA to be transformed using a combination of translation, scaling and rotation operations to give the transformed probe configuration matrix, i.e., [86]

$$\mathbf{Y} = \{\beta\mathbf{S}_{2c}\Gamma + 1_k\gamma^T : \beta \in \Re^+, \Gamma \in \mathbf{SO}(m), \gamma \in \Re^m\}, \tag{3.6}$$

where $\beta \in \Re^+$ is scale, Γ is an $(m \times m)$ rotation matrix, $\mathbf{SO}(m)$ is the special orthogonal group of $(m \times m)$ rotation matrices and γ is an $(m \times 1)$ translation m-vector. The similarity parameters γ, Γ and β are estimated by minimizing the squared Euclidean distance [86], i.e.,

$$D^2_{\mathrm{OPA}}(\mathbf{S}_{1c}, \mathbf{S}_{2c}) = \|\mathbf{S}_{1c} - \beta\mathbf{S}_{2c}\Gamma - \mathbf{1}_k\gamma^{\mathrm{T}}\|^2, \tag{3.7}$$

where OPA stands for ordinary Procrustes analysis, $\|\mathbf{S}\| = [\mathrm{trace}(\mathbf{S}^{\mathbf{T}}\mathbf{S})]^{1/2}$ is the Euclidean norm. The rows of the transformed matrix contain the transformed values of landmark distances from COM-SC. These distance values are used to form the transformed probe shape signal (represented by dashed line in Fig. 3.3(d)) and is superimposed on the gallery shape signal to visualise the differences between corresponding positions of landmark-distances from COM-SC in the two shape signals. The transformed probe configuration matrix $\mathbf{Y}$ is then compared with the centred gallery configuration matrix $\mathbf{S}_{1c}$ to obtain dissimilarity score D_{PSA} between them using

$$D_{\mathrm{PSA}} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{m}(\mathbf{S}_{1ci,j} - \mathbf{Y}_{i,j})^2}{\sum_{i=1}^{k}\sum_{j=1}^{m}(\mathbf{S}_{1ci,j} - \mathbf{A}_j)^2}, \tag{3.8}$$

where $\mathbf{A}$ is a row vector whose each element is the mean value of the elements of the corresponding columns of $\mathbf{S}_{1c}$. The range of D_{PSA} is [0,1] and denotes the difference between the gallery shape and probe shape, the larger the value the more dissimilar are the two shapes.

We obtain dissimilarity scores by comparing the transformed probe configuration matrix with the centred gallery configuration matrix of every sequence of a particular subject at the double support phase of a gait period, and computing the average of the dissimilarity scores. This average dissimilarity score is then used for classification. The gallery class with whose sequences the probe sequence obtains the lowest average dissimilarity

score is assigned rank-1, the second lowest average dissimilarity score is assigned rank 2, and so on. In this way each class receives a ranking based on the dissimilarity score obtained by PSA. In cases where a probe subject generates similar dissimilarity scores with two different gallery subjects, STM-SPP validates the two scores with the spatio-temporal gait characteristics (gait period and step length) and physical parameters of human body (build and compactness) to determine the class with the lower rank for classifying the subject (see Section 3.3.2, Section 3.3.2 and Section 3.3.2).

Estimation of gait period and step length

A step is the motion between successive heel strikes of a subject's opposite feet and a gait period consists of two steps. The method in [19] computes the gait period by considering the change in width and height of the bounding rectangle which encloses a moving silhouette contour in the lateral and frontal views, respectively, and the method in [20] performs autocorrelation of the bounding box width of a series of consecutive silhouettes for view-invariant gait recognition. We use the method similar to that used in [27] to compute gait period as it is robust to shadows under feet and carrying conditions, thereby outperforming the gait period estimation method used in the baseline method [22] for near fronto-parallel view of subjects for USF HumanID gait challenge dataset. The average width W of the leg region enclosed between aH and bH in frame I is given by

$$W = \frac{1}{aH - bH} \sum_{i=aH}^{bH} (d_i), \quad 0 \leq a \leq b \leq 1, \tag{3.9}$$

where, the values of a and b are respectively chosen to be 0.750 and 0.9375 for USF dataset to reduce the effects of shadows under feet and carrying condition, and d_i is the Euclidean distance between the leftmost foreground pixel (l_x, l_y) and the rightmost foreground pixel (r_x, r_y) on the ith line, i.e.,

$$d_i = [(l_x - r_x)^2 + (l_y - r_y)^2]^{\frac{1}{2}}. \tag{3.10}$$

It results in a periodic signal with distinct peaks and valleys that respectively correspond to the expansion and contraction of the bounding rectangle as the subject's legs extend and come back together during a gait period. The gait period is estimated as the average distance (in terms of number of frames) between each pair of consecutive valleys or peaks.

Differences in walking speed of the same subject in different gait sequences result in variations in the gait period, i.e., if a subject walks slowly in a certain situation and more quickly in another, the gait period will comprise different number of constituent frames. Thus, to detect similarities in walking patterns of the same subject in different

video sequences collected over different days, DTW [99; 70] is applied to account for variation in human movement. DTW uses dynamic programming to compute a warping function that optimally aligns two time-dependent sequences of variable lengths for measuring similarity under certain restrictions. Given two sequences of the same subject $E = (W_{E1}, W_{E2}, ..., W_{EM})$ and $F = (W_{F1}, W_{F2}, ..., W_{FN})$ of respective lengths $M \in \mathbb{N}$ and $N \in \mathbb{N}$, and W_{Ei} and W_{Fj} are the respective average width of leg region (as given by Eq. (3.9)) of their elements, DTW constructs an $M \times N$ matrix of Euclidean distances between corresponding widths, i.e.,

$$d(W_{Ei}, W_{Fj}) = (W_{Ei} - W_{Fj})^2. \tag{3.11}$$

An $M \times N$ warping path is a sequence $p = (p_1, p_2, ..., p_L)$ with $p_l = (m_l, n_l) \in [1 : M] \times [1 : N]$ for $l \in [1 : L]$ for mapping two sequences which satisfies the followings: (a) boundary condition: $p_1 = (1, 1)$ and $p_L = (M, N)$; (b) monotonicity condition: $m_1 \leq m_2 \leq m_3 \leq .... \leq m_L$ and $n_1 \leq n_2 \leq n_3 \leq .... \leq n_L$; and (c) step size condition: $p_{l+1} - p_l \in (1, 0), (0, 1), (1, 1)$ for $l \in [1 : L - 1]$. DTW minimises the cost of warping E and F together, i.e.,

$$\text{DTW}(E, F) = \min \left(\frac{(\sum_{l=1}^{L} p_l)^{\frac{1}{2}}}{L} \right). \tag{3.12}$$

The similarity between sequences E and F is measured using Eq. (3.12) to determine gait period of the same subject.

The step length (SL) is the longitudinal distance between two feet when they are maximally apart in a gait period. It is measured as the width of the bounding rectangle enclosing the silhouette contour at the double support phase of a gait period only for the lateral view of the silhouettes to essentially remove the foreshortening effects due to different views.

Estimation of physical parameters

We estimate the physical parameters of the height-normalised and centre-aligned silhouettes of the CMU MoBo [72] and USF HumanID [22] datasets in lateral views and at the double support phase of the gait period. The compactness (CM) of the contour is estimated in terms of its *perimeter* and *area* of the silhouette as $(perimeter)^2/area$. We use CM because it is dimensionless and is thus scale invariant. It is also invariant to orientation and thus acts as a good region descriptor [98]. Build (B), being a ratio of subject's chest width to subject's height is used to differentiate between thin and fat subjects, and between tall and short subjects. These parameters are combined with the gait period (G) and subject's step length (SL) to form a 4-dimensional vector $< G, SL, CM, B >$ for each sequence of the same

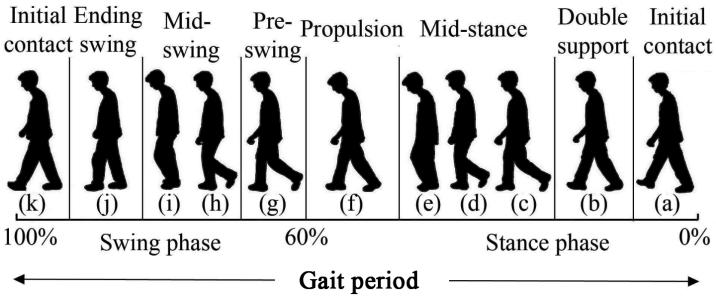


Figure 3.4: Ten phases of a gait period (a)-(k) of a subject from CMU MoBo dataset: stance phase (a)-(f) comprising about 60% of the gait period; and swing phase (g)-(j), comprising remaining 40% of the gait period.

subject. Note that this is unlike the method in [29] which uses $< G, SL, H, B >$ to validate the similarity scores between projection centroids of two gait sequences obtained by using normalised Euclidean distance. The measurement values of the physical parameters and spatio-temporal gait characteristics for different sequences of all subjects are stored to form the gallery database.

Validation using NNC

We use NNC due to its simplicity and ease of implementation [96] to validate similar dissimilarity scores generated by a probe sequence with two different gallery sequences to determine the lower rank of class to classify the probe sequence. We compute the sum of the Euclidean distances for all four measurements of $< G, SL, CM, B >$ between the probe sequence (K_i) and each of the two gallery sequences (T_i), i.e.,

$$d = (\sum_{i=1}^{4} (T_i - K_i)^2)^{\frac{1}{2}}. \tag{3.13}$$

The gallery class whose sequence is the nearest neighbour of the probe, i.e., gives the smaller d is selected as the correct class for the probe and is assigned the lower of the two rankings determined by PSA.

3.3.3 Phase 2 of Module 2: Shape characterisation using EFDs

We characterise the subject's contours at specific phases of its gait period using EFDs, as EFDs are capable of describing complex contours (i.e., straight lines emanating from the geometrical centre of the contour intersect the contour more than once) with any degree of irregularity more accurately but using fewer coefficients than the classical Fourier de-

scriptors (FDs) [100]. Thus, EFDs serve as a data compression tool which reduces space complexity.

A gait period comprises periodic alternating movement of the lower limbs resulting in a forward movement of the body. It starts with the heel strike of either foot and continues until the heel of the same foot touches the ground again. It consists of two steps, where a step is the time period between successive heel strikes of opposite feet. Each foot in a gait period alternates between two phases, a stance phase and a swing phase which respectively constitute 60% and 40% of the gait period as illustrated in Fig. 3.4. In the stance phase the foot remains in contact with the ground, while in the swing phase the foot is in the air. The stance phase begins with initial contact of the heel with the ground and ends with the toe lift off the same foot from the ground. This phase has the following components: (a) initial contact when the heel of the forward foot (i.e., the foot making a forward movement) touches the ground; (b) mid-stance when the foot is positioned flat on the ground carrying the weight of the body while the other foot is in swing phase; and (c) propulsion which begins with lifting of the heel from the ground and ends with the toe lifting off the same foot indicating the termination of the stance phase. The swing phase begins with lifting of the foot from the ground and continues until the heel of the same foot touches the ground. This phase has the following three components: (a) pre-swing which begins with the toe off the ground and continues until the occurrence of maximum knee flexion; (b) mid-swing, i.e., the motion between maximum knee flexion and when the tibia is vertical to the ground; and (c) ending swing which starts from the vertical position of the tibia and continues until just prior to the forward foot making initial contact with the ground.

We captured the video sequence of a subject (Fig. 3.1(a)) walking laterally to the image plane in a stationary indoor background using a digital camera (Nikon Coolpix S3000) fixed on a tripod at a rate of 30 fps. After estimating the subject's gait period, we obtain its ten specific phases by visually analysing the constituent frames of the gait period and extract the corresponding contours. The criteria for selecting the ten phases as shown in Fig. 3.4 is to choose at least one frame from each component of stance and swing phases to ensure that distinct and significant shape of a subject over a gait period are considered. Since mid-swing and mid-stance components persist for a longer duration and provide more than one distinct phases of a subject preserving distinguishable shape characteristics, we choose two and three phases from mid-swing and mid-stance components, respectively. The portion of the contours enclosed in the region between the bottom of the bounding rectangle and up to the anatomical position of just before the hand of an upright human subject (i.e., $0.377H$ measured from the bottom of the bounding rectangle [97]) at the ten specific phases are set as the reference region-of-interests (Rf-ROIs), as this portion of contour remain unaffected by the self-occlusions caused by arm-swing. To obtain the ten specific phases of any

gait sequence automatically, the Rf-ROIs are compared one at a time with the same portion of contours of all the frames of a subject's gait period (referred to as the target region-of-interests (Tr-ROIs)) using contour matching based on Hu moments to obtain similarity scores (S_{score}) using [93]

$$S_{score} = \sum_{i=1}^{7} \left| \frac{1}{m_i^{Rf}} - \frac{1}{m_i^{Tr}} \right|, \tag{3.14}$$

where, $m_i^{Rf} = \text{sign}(h_i^{Rf}).\log|h_i^{Rf}|$ and $m_i^{Tr} = \text{sign}(h_i^{Tr}).\log|h_i^{Tr}|$. h_i^{Rf} and h_i^{Tr} denote the Hu moments of the Rf-ROI and Tr-ROI, respectively. The frame whose Tr-ROI results in the lowest S_{score} with the Rf-ROI, is extracted as one of the ten specific phases of the gait period and the process is continued by comparing the next Rf-ROI with the remaining Tr-ROIs until all the ten specific phases are obtained.

Analysing a subject's shape at the specific phases of a gait period enables STM-SPP not only to reduce computational complexity, but also to considerably overcome the adverse effect of walking speed variations under different circumstances, e.g., due to the subject's mood changes. Furthermore, the extracted contours can be considerably distorted by the presence of occlusions in the scene, severe shadows under feet and extreme lighting variations. If these distorted contours are not part of any of the ten specific phases of the gait period, they can be discarded in the case of shape analysis using EFDs, without having any effect on the classification rate. However, if any of the ten specific phases is missing including the double support phase, its immediate adjacent frame is considered. In this way, STM-SPP achieves robustness to missing or distorted frames to some extent.

Since a silhouette contour is a closed curve, it can be expressed by a periodic signal $c(t)$ of period T, i.e.,

$$c(t + T) = c(t), \tag{3.15}$$

where T is the perimeter of the contour. We consider ten contours corresponding to the ten specific phases of the gait period as shown in Fig. 3.4 to obtain EFDs for the gait signatures. To ensure similar set of equal number of points along the selected ten contours, each contour is approximated by $m = 2^7$ i.e., 128 points using interpolation based on point correspondence analysis [91].

Each contour with points of coordinates $(x(t), y(t))$ is defined in a complex plane as

$$c(t) = x(t) + jy(t). \tag{3.16}$$

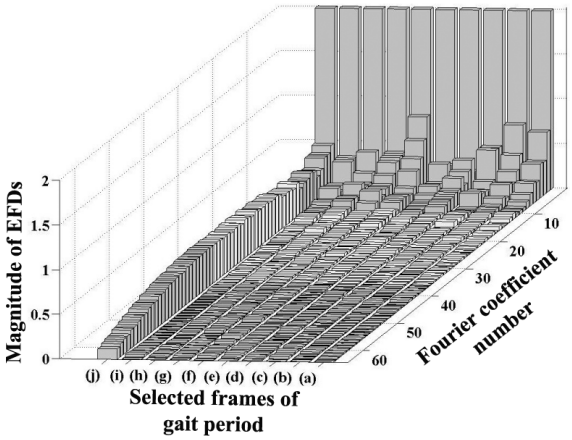


Figure 3.5: The 3D bar graph with bars representing the magnitude of EFDs corresponding to 64 Fourier coefficients grouped together for each of ten specific phases of gait period.

The elliptical Fourier representation of a contour is [96]

$$c(t) = \frac{a_{x_0}}{2} + \sum_{k=1}^{m/2}(a_{x_k}cos(k\omega t) + b_{x_k}sin(k2\omega t)) + j[\frac{a_{y_0}}{2} + \sum_{k=1}^{m/2}(a_{y_k}cos(k2\omega t) + b_{y_k}sin(k2\omega t))], \tag{3.17}$$

where

$$a_{x_k} = \frac{2}{m}\sum_{i=1}^{m} x_i cos(k\omega i\tau) \quad , \quad b_{x_k} = \frac{2}{m}\sum_{i=1}^{m} x_i sin(k\omega i\tau), \tag{3.18}$$

$$a_{y_k} = \frac{2}{m}\sum_{i=1}^{m} y_i cos(k\omega i\tau) \quad , \quad b_{y_k} = \frac{2}{m}\sum_{i=1}^{m} y_i sin(k\omega i\tau), \tag{3.19}$$

fundamental frequency $\omega = T/2\pi$ and sampled period $\tau = T/m$. The contour represented in matrix form is [96]

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \frac{1}{2}\begin{bmatrix} a_{x_0} \\ a_{y_0} \end{bmatrix} + \sum_{k=1}^{\infty}\begin{bmatrix} a_{x_k} & b_{x_k} \\ a_{y_k} & b_{y_k} \end{bmatrix}\begin{bmatrix} cos(k\omega t) \\ sin(k\omega t) \end{bmatrix}. \tag{3.20}$$

This matrix resembles that of an ellipse, with a_{x_k} and b_{y_k} representing its major and minor axes, respectively. The scale, rotation and translation invariant EFD is

$$\text{EFD} = \frac{\sqrt{a_{x_k}^2 + a_{y_k}^2}}{\sqrt{a_{x_1}^2 + a_{y_1}^2}} + \frac{\sqrt{b_{x_k}^2 + b_{y_k}^2}}{\sqrt{b_{x_1}^2 + b_{y_1}^2}}. \tag{3.21}$$

We represent EFDs in the form of a matrix of dimension $a_r \times b_c$ (with a_r representing the ten specific phases of the gait period and b_c representing the number of elliptic

Fourier coefficients considered). Since the number of elliptic Fourier coefficients equaling one half of the total number of contour points are capable of reconstructing a very good approximation of the original contour [96], we use $b_c = m/2$, i.e., 64, for STM-SPP. Each EFD is represented by a bar in the 3D bar graph shown in Fig. 3.5, where elements of the rows of the matrix are grouped together. Let P and Q be two such matrices for a gallery and a probe gait sequence, respectively. The dissimilarity score between them is

$$D_{\text{EFD}} = \frac{\sum_{i=1}^{a_r} \sum_{j=1}^{b_c} (\mathbf{P}_{i,j} - \mathbf{Q}_{i,j})^2}{\sum_{j=1}^{b_c} (\mathbf{P}_j - \text{mean}(\mathbf{P}_j))^2}, \tag{3.22}$$

where $\mathbf{P}_j$ is jth column of $\mathbf{P}_{i,j}$ and mean(.) computes average. The range of D_{EFD} is [0,1], the larger the value the more dissimilar are the two shapes. We obtain dissimilarity score by comparing the EFDs of a probe subject with EFDs of each of the gallery subjects for a gait period. In a similar manner to phase 1, this dissimilarity score is used to classify the subject.

The above shape analysis gives excellent classification results in the absence of shape variation between the gallery and probe sequences of the same subject. The classification performance decreases significantly if the shape of the gallery and probe subjects differ due to different activities (e.g., slow or fast walk vs walking with ball) performed by the same subject for the CMU MoBo dataset, and carrying conditions (briefcase vs no briefcase) and shadows under feet for the USF HumanID dataset. To reduce the effects of shape variations on the classification rate, a part-based shape analysis using EFDs is invoked when carrying condition is detected in a sequence.

The shape of an upright silhouette above the wrist is not affected by shape variations when the subject's hand carries a briefcase or a small bag. According to anthropometry, the position of the wrist as a fraction of body height is estimated to be $0.485H$ [97] measured from the bottom of the bounding rectangle. Thus, an analysis of the part of silhouettes enclosed in the region (1-0.485H), i.e., $0.515H$ of the bounding rectangle measured from the top using EFDs remove the shape variations due to carrying conditions (briefcase vs no briefcase) among the gallery and probe sequences. The leg region of a silhouette enclosed between aH and bH (where $a = 0.750$ and $b = 0.9375$ [27]) removes the effect of shape distortion due to the presence of briefcase and shadows under feet. Thus, STM-SPP detects the presence of briefcase or a small bag automatically by examining the difference in the number of contour points enclosed in the region between $0.515H$ and $0.750H$. A substantial increase in the number of contour points (e.g., for USF dataset an increase of at least twenty contour points without applying polygon approximation and for the same phase of gait period between gallery and probe sequences) confirms the presence of briefcase for most of the cases. STM-SPP analyses the EFDs of the subjects carrying briefcase with shadows

Figure 3.6: Illustration of part-based shape analysis: (a)-(c) upper segments; and (d)-(f) lower segments of three subjects from CMU MoBo dataset to exclude the ball.

under feet in two parts, so as to avoid the variations in shape among gallery and probe sequences. We verified the effectiveness of this part-based shape analysis on all subjects carrying a briefcase for the USF dataset.

If a subject carries a small item, e.g., ball, package, tiffin box, etc., with folded arms, it is unlikely that the shape of silhouettes above the shoulder and below the position of wrist, i.e., $0.515H$, from the top of the bounding rectangle will be affected. Experimental analysis for all the subjects holding a ball in CMU MoBo dataset verifies the appropriateness of the assumption. We found that the segment of the silhouette enclosed between $0.225H$ from the top of the bounding rectangle, and the lower segment enclosed between $0.500H$ and the bottom of the bounding rectangle exclude the ball for all the twenty four subjects walking with ball in the CMU MoBo dataset (Fig. 3.6 (a)-(f)). Experimental analysis reveals that an increase in the number of contour points enclosed in the region between the anatomical position of wrist and the top of the bounding rectangle by at least fifty confirms the presence of a ball for all the subjects in the CMU MoBo dataset without applying any morphological operation and polygon approximation technique. Thus the effect of carrying small items on silhouette shape can be removed by segmenting it in two parts: (1) the upper segment spanning from the top of the bounding rectangle up to the shoulder; and (2) the lower segment spanning from the anatomical position of wrist to the bottom of the bounding rectangle. For part-based EFD analysis, we obtain a dissimilarity score corresponding to each part and compute the average dissimilarity score.

3.3.4 Module 3: Combining classifications

We use rank-summation based classifier combination rule [101; 102] to combine the outputs of the two classifiers for improved reliability in human identification. We choose this combination rule due to its appropriateness in STM-SPP, as it enables to effectively combine the results of a small number of classifiers with a relatively large number of classes. It is also easier to implement than the score-based fusion strategy [103], as the latter requires transformation of the scores to a same scale in order to be comparable before being combined.

Let $R^{(j)}(\theta)$ be the rank assigned to class θ by classifier $j \in J$, where J represents the set of classifiers consisting of two elements for STM-SPP. The sum of all ranks assigned to each class by all classifiers is

$$S(\theta) = \sum_{j}^{J} (R^{(j)}\theta). \tag{3.23}$$

The class with the lowest sum rank is chosen as the correct class for the probe sequence. Noting that Hu moments are linear combinations of normalised central moments that are invariant to changes in rotation, reflection and scale, to resolve cases where two classes have the same sum rank, STM-SPP performs hierarchical contour matching [93] based on the Hu moments [88] between each image of the probe gait sequence and the corresponding images of the gallery sequences.

The translational invariant 2D $(p + q)^{th}$ order central moments $(\mu_{p,q})$ of a contour $I(x, y)$ is

$$\mu_{p,q} = \sum_{i=0}^{N} I(x, y)(x - x_{\text{avg}})^p (y - y_{\text{avg}})^q, \tag{3.24}$$

where $x_{\text{avg}} = m_{10}/m_{00}$, $y_{\text{avg}} = m_{01}/m_{00}$ and N is the total number of pixels in the contour. To ensure objects of the same shape but dissimilar sizes give similar values, we use the normalised central moment

$$\eta_{p,q} = \frac{\mu_{p,q}}{m_{00}^{(p+q)/2+1}}. \tag{3.25}$$

The hierarchical contour matching technique involves the formation of contour trees [104] prior to the contour comparisons based on Hu moments. Since the resultant contour trees are susceptible to minor variations in the contours, all contours are approximated by a polygon using Douglas-Peucker approximation algorithm [93; 105] having fewer vertices (as shown in Fig. 3.1(g)), for better comparison.

3.4 Experiments

We use the silhouettes from CMU MoBo dataset [72] and USF HumanID gait challenge dataset [22] to evaluate the performance of STM-SPP with variation in terms of walking speed, carrying condition, clothing, footwear and view. Experimental analyses on these two datasets enable us to make uniform comparisons with several related methods.

3.4.1 Experiments on CMU MoBo dataset

The performance of STM-SPP is evaluated on the profile view of the silhouettes of this data set using different gallery and probe sequences for slow walk, fast walk and walk

Table 3.1: Top-rank identification rates of STM-SPP (in percentage) on MoBo dataset for lateral view with rates of SSP from [21] enclosed in parentheses.

Probe	Gallery		
Walk type	Slow walk	Fast walk	Walk with ball
Slow walk	100(100)	94(54)	93
Fast walk	91(32)	100(100)	84
Walk with ball	82	82	100

Table 3.2: Top-rank identification rates (in percentage) of CMU1 (from [19]) and STM-SPP on CMU MoBo dataset for profile views only.

Gallery	Probe	CMU1	STM-SPP
profile, slow	profile, fast	76	96
profile, slow	profile, ball	92	93

with ball. STM-SPP is compared with the following silhouette based approaches: SSP [21], CMU1 [19] and SSM [23]. For uniform comparison with SSP, STM-SPP also uses holdout cross validation technique, in which gallery and probe sets correspond to different combination of walking speeds for each of twenty five subjects. Since STM-SPP is defined only on lateral view of the silhouettes, we consider lateral view of two sequences per subject (total 50 sequences) walking at slow pace (2.06 miles/h) and fast pace (2.82 miles/h), for performance comparison with SSP. Table 3.1 shows that the performance of SSP for profile view (obtained from [21]) degrades significantly when the probe and gallery sequences differ in walking speed, whereas STM-SPP which analyses shape at ten specific phases of gait period and uses DTW for gait period estimation overcomes the effects of variations in walking speed, and outperforms SSP for the profile view.

Table 3.2 shows that STM-SPP outperforms CMU1 for profile view. A comparison with SSM is presented in Table 3.3 for subjects walking parallel to the image plane. The shape based approach of SSM which uses stance correlation, shows encouraging results for subjects performing same activities with varying speed. Although speed variations are also accounted for in STM-SPP, like SSM the performance of STM-SPP degrades when shape of the subject's silhouettes change due to different activities (e.g., slow walk vs walk with ball). The better performance of STM-SPP over SSM is attributed to the validation based on physical parameters (i.e., gait period and step length) which are independent of body shape, and part-based shape analysis using EFDs.

Table 3.3: Top-rank identification rates (in percentage) of STM-SPP on MoBo dataset for across-activities with rates of SSM from [23] enclosed in parentheses.

Activity	Slow walk	Fast walk	Walk with ball
Slow walk	100(100)	95(80)	93(48)
Fast walk	96(84)	100(100)	84(48)
Walk with ball	82(68)	82(48)	100(92)

3.4.2 Experiments on USF HumanID gait challenge dataset

STM-SPP is evaluated on both the small version (452 sequences from 74 subjects, data acquired in May only) and the full version (1870 sequences from 122 subjects, data acquired in May and November) of USF HumanID gait challenge dataset.

Table 3.4 shows the results on the full version of the dataset in terms of identification rate (P_I) at ranks 1 and 5, to enable a comparison with that of the baseline [22], GEI [25], RCK-G [65] and GFI [49]. Table 3.5 shows the results on the full version of the dataset in terms of verification rate (P_V) at the false alarm rates of 1 percent and 10 percent, to enable a comparison with baseline [22] and DNGR [25]. We report the verification rates for baseline method obtained by using z-normalised similarity scores.

Since the number of probe subjects in the gait challenge experiments varies, we compute the weighted average identification rate (WAvgI) [47], i.e.,

$$\text{WAvgI} = \frac{\sum_{i=1}^{g} w_i x_i}{\sum_{i=1}^{g} w_i}, \tag{3.26}$$

and the weighted average verification rate (WAvgV) [47], i.e.,

$$\text{WAvgV} = \frac{\sum_{i=1}^{g} w_i x_v}{\sum_{i=1}^{g} w_i}, \tag{3.27}$$

where g denotes the number of challenge experiments whose values are respectively 12 (i.e., Exp. A-L) and 7 (i.e., Exp. A-G) for the full and small versions of HumanID gait challenge dataset. x_i and x_v are respectively the ith element of the set of identification rates (%) and verification rates (%) for all challenge experiments, and w_i is the ith element of the set of number of probe subjects participating in the respective experiments.

Table 3.4 reports the identification rates of GEI for the twelve gait challenge experiments which are obtained by combining the real and synthetic gait features. All methods perform satisfactorily for experiments A-J but poorly for experiments K and L, with STM-SPP achieving the best performances in all experiments, followed by GEI. The better

Table 3.4: Top-rank identification rates (in percentage) for different methods on the full version of USF HumanID gait challenge dataset using the gallery set (G, A, R, NB, M/N) of 122 subjects. The rates of GEI from [25] are enclosed in parentheses.

Exp.	Identification Rate (P_I)%					
	Baseline(GEI)		RCK-G(GFI)		STM-SPP	
	Rank1	Rank5	Rank1	Rank5	Rank1	Rank5
A	73(90)	88(94)	83(89)	96(98)	92	96
B	78(91)	93(94)	86(93)	94(94)	95	98
C	48(81)	78(93)	78(70)	88(93)	84	95
D	32(56)	66(78)	39(19)	66(40)	72	80
E	22(64)	55(81)	34(23)	63(47)	68	84
F	17(25)	42(56)	20(7)	51(26)	29	59
G	17(36)	38(53)	21(8)	46 (25)	40	61
H	61(64)	85(90)	43(78)	66(94)	69	92
I	57(60)	78(83)	40(67)	68(85)	60	84
J	36(60)	62(82)	40(48)	65(74)	64	85
K	3(6)	12(27)	16(3)	44(24)	20	30
L	3(15)	15(21)	5(9)	22(24)	18	27
WAvgI	61(58)	78(76)	44(46)	67(64)	63	79

Table 3.5: Verification rates at a false alarm rate (P_F) of 1% and 10% for Baseline from [22], DNGR from [26] and STM-SPP on full version of USF HumanID gait challenge dataset using the gallery set (G, A, R, NB, M/N) of 122 subjects.

Exp.	Verification Rate (P_V)% at		
	Baseline	DNGR	STM-SPP
	P_F:1(10)%	P_F:1(10)%	P_F:1(10)%
A	82(94)	93(98)	88(100)
B	87(94)	94(98)	94(100)
C	65(94)	80(94)	86(98)
D	44(80)	68(96)	80(94)
E	35(76)	62(90)	74(84)
F	20(60)	53(86)	50(82)
G	28(55)	43(79)	52(76)
H	72(91)	91(99)	83(95)
I	67(85)	86(97)	76(93)
J	48(76)	58(92)	65(92)
K	6(24)	27(61)	21(58)
L	6(24)	24(46)	19(52)
WAvgV	51(76)	70(91)	70(89)

Table 3.6: Top-rank identification rates (in percentage) for different methods on the small version of USF HumanID gait challenge dataset using the Gallery set (G, A, R) of 71 subjects. The rates for Baseline, CASIA, CMU2, CMU1 and GEI are from [22], [29], [89], [89] and [25], respectively.

Exp.	Baseline	CASIA	CMU2	CMU1	GEI	STM-SPP
A	87	70.42	85	87	100	100
B	81	58.54	81	81	90	94
C	54	51.22	60	66	85	89
D	39	34.33	23	21	47	73
E	33	21.43	17	19	57	69
F	29	27.27	25	27	32	40
G	26	14.29	21	23	31	36
WAvgI	50.62	40.83	44.93	46.44	62.83	71.74

performance of STM-SPP in experiment G than in experiment F inspite of an additional covariate, namely shoe type, is attributed to the fewer subjects participating in experiment G and thus the smaller likelihood of including subjects from the class which is difficult to identify across all the experiments for STM-SPP [22]. Also, the validation based on physical parameters contributes to making STM-SPP robust against across-day gait variations, e.g., the same subject wearing different shoes.

Table 3.6 shows the results on the small version of the dataset (No-Briefcase data) to enable a comparison with Baseline, CASIA [29], CMU2 [89], CMU1 and GEI. Note that we present the identification rates at rank-1 of CMU2 obtained by weighted correlation similarity measure, as its performance is better than the identification rates at rank-1 obtained by median weighted distance similarity measure presented in [89]. We compare the performance of STM-SPP with the identification rates of GEI obtained by fusing real and synthetic gait templates, as it shows higher performance than the identification rates obtained separately by using real and synthetic gait templates. Table 3.6 shows that STM-SPP outperforms the other methods for all experiments.

The performance of STM-SPP for the twelve challenge experiments of the dataset is measured by two different modes of experimental analysis, namely identification mode and verification mode, using cumulative match characteristic (CMC) and receiver operating characteristic (ROC) curves respectively, following [106].

Identification refers to an attempt to determine the identity of an unknown subject by comparing a subject's probe sequence to all the gallery sequences in the database. In contrast to open-set identification, a closed-set identification always guarantees the existence of the subject in the database. We analyse the closed set identification performance of

STM-SPP on the profile view by taking out one subject as the probe sequence and train it on all the subjects of the dataset including the probe sequence. The identification result is represented by a CMC curve, i.e., the probability of correct matches versus ranks. According to this curve, the probability of correct identification at rank r implies that the probability of correct match is among the top r similarity scores, and the performance at rank-1 represents the correct classification rate (CCR), i.e., the identification rate. The percentage of CCR is

$$\mathrm{CCR}(\%) = s_c / s_t * 100, \tag{3.28}$$

where s_c and s_t are respectively the number of correctly identified subjects and the total number of subjects participating in the classification. Fig. 3.7(a) and (c) show that the identification rates at rank-1 range from 3% to 84% for PSA, which are increased to a range of 18% to 92% by using the classifier combination rule.

Verification refers to an attempt to confirm a subject's claimed identity by a one-to-one comparison of the probe sequence to one or more gallery sequences corresponding to the subject of the claimed identity in terms of false alarm rate (the probability that the method incorrectly matches the probe sequence to a nonmatching gallery sequence) and verification rate (the probability that the method succeeds to correctly detect a match between the probe and gallery sequences). An ROC curve is a graphical representation of the relationship between false alarm rate and verification rate of the classifier as its discrimination threshold is varied. Fig. 3.7(b) and (d) show that the verification rates using PSA range from 9% to 86% at a false alarm rate of 1%, and increase to 19% to 94% at a false alarm rate of 1% by combining the classifiers, for all the twelve challenge experiments. It is evident from Table 3.4 and Fig. 3.7, that STM-SPP is least affected by variation in shoe types, followed by about 30 degrees change in viewpoint. However, time (i.e., when the dataset was generated) has the most impact on the performance of STM-SPP, as it implicitly means the same subjects wearing different clothes and shoes.

3.4.3 Computational complexity analysis

The time for recognising a subject depends on the size of the datasets as well as on their characteristics, i.e., the number of cases in which similar dissimilarity scores are obtained by PSA and the number of subjects carrying items in different sequences. The processing time (measured using the computer system clock) for comparing all the ten Rf-ROIs one at a time with the Tr-ROIs obtained from the silhouette images of a subject's gait period based on Hu moments and determining the minimum S_{score} in each case for extracting the ten specific phases of the gait period is 5 secs using OpenCV 2.1 in Microsoft Visual Studio 2008 Express Edition environment on an Intel (R) Core (TM) i7 processor working

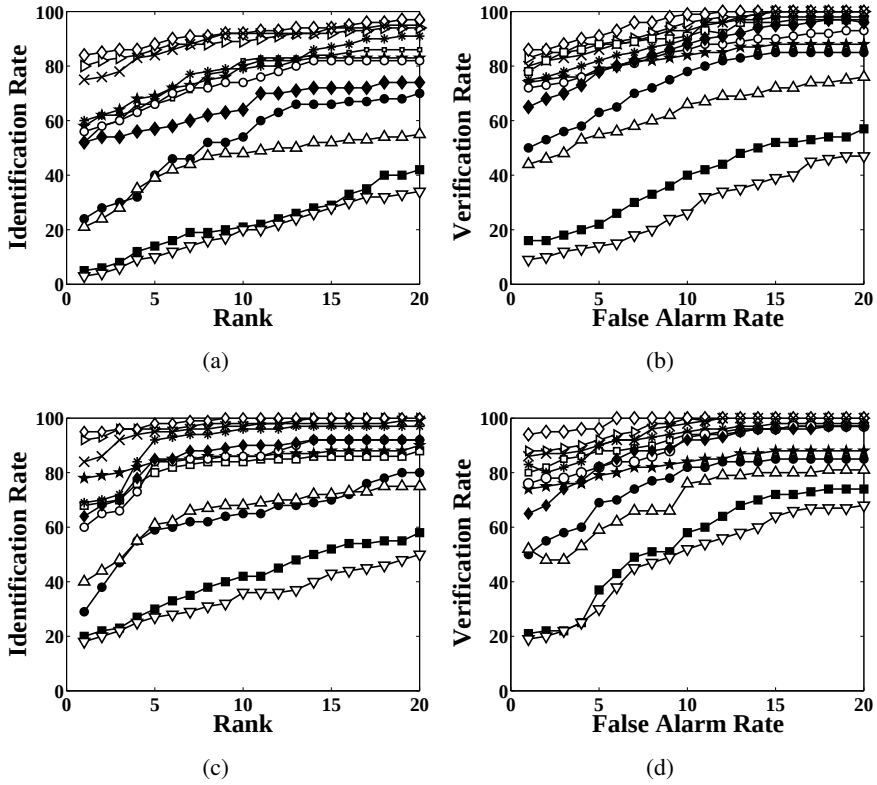


Figure 3.7: Performance on twelve challenge experiments of USF dataset. Identification mode (CMC): (a) PSA and (c) Combined. Verification mode (ROC): (b) PSA and (d) Combined. Keys: '▷'- Exp. A (Probe: G, A, L, NB, M/N); '◇'- Exp. B (Probe: G, B, R, NB, M/N); '×'- Exp. C (Probe: G, B, L, NB, M/N); '□'- Exp. D (Probe: C, A, R, NB, M/N); '★'- Exp. E (Probe: C, B, R, NB, M/N); '●'- Exp. F (Probe: C, A, L, NB, M/N); '△'- Exp. G (Probe: C, B, L, NB, M/N); '∗'- Exp. H (Probe: G, A, R, BF, M/N); '○'- Exp. I (Probe: G, B, R, BF, M/N); '◆'- Exp. J (Probe: G, A, L, BF, M/N); '■'- Exp. K (Probe: G, A/B, R, NB, N); and '▽'- Exp. L (Probe: C, A/B, R, NB, N).

at 2.93 GHz with 4 GB RAM and 500 GB HDD running Windows 7 operating system. The combined processing time to obtain the dissimilarity scores between a gallery subject and a probe subject using both PSA and EFDs along with estimation of the subject's physical parameters is 45 sec/gait period.

The baseline method has a very high computational complexity as it performs repeated intersequence spatio-temporal correlation between gallery and probe silhouette sequences to obtain the similarity measure [70; 29]. Instead of processing an entire gait sequence, the real-time method in [79] identifies a previously known subject by analysing its silhouettes over a gait period spanning up to 25 frames to reduce computational complexity. STM-SPP further reduces it by analysing the shape of contours instead of silhouettes at the double support phase and ten specific phases of the gait period by using PSA and EFDs, respectively. Furthermore, it converts the contour at the double support phase into 1D shape signal based on 28 landmarks before applying PSA. Since the configuration matrices formed by the shape signals have one column only, the space and time complexity of PSA is linear, i.e., O(k), where k=28 for STM-SPP. Despite the effectiveness of DTW, it has a quadratic time and space complexity, i.e., O(MN) (where M and N denote the lengths of the two time-varying sequences being compared) which limits its usefulness to small sequences. However, the number of frames of a gait period does not usually exceed 35 and this ensures the suitability of using DTW in STM-SPP. The ten specific phases of a gait period are obtained by ROI of contour matching based on Hu moments. The use of ROI helps to speed up execution time, as it enables processing of subregion of an image.

The NNC used for validating similar dissimilarity scores requires storage of entire gallery database, thus requiring much memory space and increased execution time. However, its use is limited only to resolving similar dissimilarity scores obtained by PSA, and therefore it does not increase the overall computational complexity of STM-SPP significantly. Since the computational complexity of EFDs is quadratic, i.e., O($b_c m$) (where b_c denotes the number of elliptic Fourier coefficients considered and m denotes the number of points of the polygon-approximated contour), it increases the processing time of STM-SPP. To address this, instead of analysisng all frames of a gait period, we analyse the contours by EFDs only at its ten specific phases (i.e., keeping $a_r = 10$) while capturing most of the significant gait characteristics. The contours are approximated with reduced number of points, i.e., $m = 128$, which reduces processing time.

3.5 Conclusion

The proposed two-phase gait recognition method, STM-SPP, analyses the shape of silhouette contours of a human subject in a video sequence. In the first phase, STM-SPP performs

subject classification based on a dissimilarity score by comparing distances of landmarks (anatomical, mathematical and pseudo) from COM-SC in the double support phase of a gait period using PSA. This classification performance is substantially enhanced by validating similar dissimilarity scores based on spatio-temporal gait characteristics and physical parameters of human body in the presence of limited variations in view, clothing and footwear. In the second phase, STM-SPP characterises the silhouette contours by EFDs at the ten specific phases of a gait period to obtain gait signatures and uses a dissimilarity score to classify the subjects. A part-based shape analysis using EFDs is applied to reduce the impact of shape variation between gallery and probe silhouette contours on the classification rate when across-day variations due to carrying conditions are detected. The outputs of the two classifiers are combined effectively by rank-summation based classifier combination rule, where a tie in ranking is resolved by contour matching based on Hu moments.

STM-SPP has several desirable advantages, which make it suitable for real-world applications. The shape analysis at the ten specific phases of a gait period and gait period detection by the application of DTW aids STM-SPP to deal with varying walking speeds of the same subject under different circumstances. STM-SPP is also robust to subjects carrying small items and limited across-day gait variations, but not significant change of styles, e.g., pants versus skirts or long coats, massive leg injury, variations of camera viewpoints, etc. It is also robust against missing or distorted frames to some extent mainly due to partial occlusions and segmentation imperfections. It is insensitive to colour and texture of the subject's clothing, as it analyses the shape of the contours. Since its feature space is simplified, it does not require any dimensionality reduction technique like PCA and multiple discriminant analysis as in [25; 29]. The attractiveness of the STM-SPP is the ease of implementation with low computational complexity. Experimental analyses on two publicly available datasets show that STM-SPP significantly outperforms several related silhouette-based gait recognition methods. However, it suffers from the following limitations which require further attention for its advancement:

- Walking direction: STM-SPP is designed to identify human subjects only for lateral views of the gallery and probe sequences. Although, lateral views of the walking subjects capture most of the significant gait characteristics, it is not always possible to capture image frames from the side of a subject, especially in hallways [79]. Hence, future developments are required to enable STM-SPP to address unconstrained human movements especially in cluttered scenes.

- Clothing invariance: STM-SPP is not robust against significant clothing variations between gallery and probe sequences, such as pants vs long skirts, shorts vs down jackets, trench coats, etc. Also, the extraction of landmarks in the limb region is

impossible in the case of a subject wearing long skirt or long coat as the clothing keeps the subject's limbs covered. Therefore, future work will involve improvements of STM-SPP using part-based clothing categorization to achieve substantial clothing invariance.

- Dynamic gait characteristics: STM-SPP analyses sequences of deforming shape of contours over a gait period, but does not incorporate the dynamic motion characteristics of gait which also play an important role in human identification. Since the more appropriate gait signatures are utilised the better is the performance of any gait recognition algorithm, we will consider oscillatory trajectories of joints in future work, giving arm-swing a consideration.

Chapter 4

Static Shape and Dynamic Motion Analysis of Silhouette Contours

4.1 Introduction

The model-free gait recognition method presented in Chapter 3, i.e., STM-SPP analyses the spatio-temporal deformation of a silhouette's shape at ten phases of a gait period to identify a subject. Although the spatio-temporal analysis of a subject's shape in a gait sequence provides better discriminative power than its kinematics, inclusion of dynamical motion characteristics improves the identification rate in the presence of numerous across-day gait variation. Thus, this chapter introduces a gait recognition method, i.e., STS-DM, which combines the spatio-temporal shape (STS) features of a subject's silhouettes with the subject's dynamic motion (DM) characteristics over a gait period using both model-free and model-based approaches to achieve robustness against the maximum number of challenging factors of gait recognition when compared to state-of-the-art gait recognition methods, namely robustness against small carried items, walking speed variation, shadows under feet, limited variation in clothing, segmentation noise, change in ground surface, missing body parts and distorted or missing frames due to presence of occluding objects in the scene. STS-DM operates on the lateral (i.e., profile) view of a subject since this view contains most of the significant gait characteristics.

Most gait recognition methods do not consider the subject's arm-swing and the self-occlusion caused by it. Thus, STS-DM introduces a novel analysis of angular rotation pattern of leading knee (ARPoLK) of silhouette contours for subject identification in the presence of across-day variation, e.g., clothing, footwear, hair style and ground surface, with a consideration of the subject's arm-swing. STS-DM analyses the shape of the silhouette contours at ten phases of a gait period via their low-pass filtered FDs to only retain their

global shape information. STS-DM uses ellipses fitted to body segments at ten phases of a gait period for full-body shape and motion analysis which is invariant to boundary shape distortions due to segmentation errors and missing or distorted body parts. Contour shape analysis at the ten phases that reveal most of the distinguishable shape characteristics also enables STS-DM to benefit from speed-invariant shape sequence processing with reduced processing time and achieve robustness against missing or distorted frames due to occlusions. Since the dynamic motion characteristics of gait manifest over a gait period more than in discrete phases, ARPoLK analysis is performed over a gait period.

The proposed STS-DM is thus motivated by the need for a gait recognition method that addresses a wide variety of challenging factors that limit the success of gait as a behavioural biometrics to reliably identify a subject in practical situations. The novelties of STS-DM are: (a) it effectively combines static shape characteristics with the local and global dynamic gait characteristics to achieve robustness against the maximum number of challenging factors; (b) it analyses the subject's shape by FDs, and uses PWMS to generate a match score; (c) it introduces an experimentally supported procedure for detecting carried items and a component-based FD analysis based on anatomical studies to achieve invariance to all common types of small carrying conditions, and this level of invariance has not been addressed before; (d) it introduces ARPoLK analysis which is invariant to self-occlusions of the limbs of a walking subject, and hence captures the local dynamic gait signature very efficiently; (e) the use of ARPoLK analysis enables STS-DM to implicitly address subject's arm-swing, and the use of DTW to obtain a match score which is invariant to walking speed; (f) it analyses the full-body shape and motion characteristics based on ellipse-fitting to body segments and uses Bhattacharyya distance histogram matching (BDHM) to obtain a match score; (g) the match scores obtained by PWMS, DTW and BDHM are combined using weighted sum rule of score level fusion for robust identification; (h) the robustness of STS-DM against missing frames is demonstrated; and (i) STS-DM provides competitive identification rates with reduced computational complexity.

The rest of the chapter is organized as follows. Section 4.2 discusses related work and Section 4.3 presents STS-DM. Experimental results are analysed in Section 4.4 and Section 4.5 concludes the chapter.

4.2 Related Work

Various existing gait recognition methods analyse static shape and dynamic motion characteristics of gait sequences to address variation in walking speed, carrying condition and clothing, as well as other covariate factors, e.g., segmentation noise, occlusions, low resolution, changes in ground surface and shadows under feet. The variation in walking speed

has been effectively addressed in the methods [40; 41; 26; 42; 43; 44]. The method SVB frieze in [40] computes SVB frieze pattern of the difference frames obtained by subtracting the key frames, e.g., double support stance frames from the series of subsequent frames of a gait period. The method in [41] replaces the centroid-based shape configuration of traditional PSA with high-order shape configuration to take into account of dynamic gait characteristics. The method introduces a differential composition model for increased inter subject discriminability and uses Procrustes distance for identifying a subject. The DNGR method in [26] normalises gait dynamics using population HMM whose states represent specific gait stances over a gait period, and a subject is identified based on LDA. DNGR is robust against limited variation in walking speed and change in ground surface. The method SI-PSA in [42] uses PSA to compute high-order derivative shape configuration to achieve invariance to walking speed. The walking speed invariant method using silhouette transformation, i.e., ST-WS in [43] separates static and dynamic gait features by fitting a human model and uses a factorization-based transformation model to transform the dynamic features from a reference speed to a target speed. The method in [44] uses the features extracted by FDA based CHLAC of the gait sequences to train a HMM to achieve invariance to walking speed.

The methods in [7; 17; 45] address variation in carrying conditions. The method STM-SPP [7] analyses the shape of a silhouette contour using PSA and EFDs. A part-based EFD analysis is used to address shape distortion due to carrying conditions. The method in [17] uses models to obtain skeleton parameters by wavelet decomposition of a GEI and invariant moments for combining anatomical and behavioural characteristics of gait. Thermal imaging is used to extract silhouettes that are invariant to carrying conditions and lighting variation. An iterative local curve embedding algorithm is used in [45] to extract double helical signatures.

The discriminability of a subject decreases due to shape distortions caused by clothing variation over different days. Therefore, the method in [52] uses an adaptive mechanism for combining part-based features to achieve robustness against clothing variation. The method RCK-G in [65] uses radial integration transform, circular integration transform and weighted Krawtchouk moments for feature extraction. RCK-G is robust to very limited clothing variation, but not insensitive to carrying conditions.

The methods in [25; 46; 15; 21; 47; 50; 49; 28; 48] achieve limited invariance to a few covariate factors for improved identification rate. The method in [46] uses appearance and dynamic traits of gait by analysing parameters of ellipses fitted to seven regions of a subject's silhouette, i.e., centroid, aspect ratio and elongation along with the subject's height for identification which is invariant to limited clothing variation and segmentation imperfections. The full-body layered deformable model based method in [15] addresses

self-occlusions, and the incorporation of upper body dynamics in addition to limbs enables the method to achieve robust identification in the presence of variation in footwear, clothing, ground surface and time.

The use of synthetic gait templates in the methods GEI [25] and CGI [28] enable them to achieve robustness against lower body-part distortions due to carrying a briefcase, and variation in ground surface, clothes and footwear, but not distortions of upper body-part due to variation in clothing and a carried item especially with folded arms. The method CGI uses a gait period detection technique that is robust to shadows under feet and carrying a briefcase. The method ASM in [35] uses prediction-based hierarchical active shape model and Kalman filtering to achieve invariance to illumination variation, shadows and considerable occlusions.

While the trend of the state-of-the-art gait recognition methods is to address only one or a few covariate factors, STS-DM first attempts to identify a subject in the presence of a wide range of challenging factors with low computational complexity for practical deployment. It fuses the local and global gait characteristics obtained by analysing static shape and dynamic motion of silhouette contours to address the maximum number of covariate factors so as to achieve combined invariance to carrying conditions, walking speed, shadows under feet, limited variation in clothing, segmentation noise, changes in ground surface, missing body parts and occlusions. Like the method in [91], STS-DM also characterises a subject's shape using FDs but introduces a novel component-based FD analysis to achieve invariance to all common types of small carrying conditions, and this level of invariance has not been addressed before. STM-SPP [7] and the method in [29] only analyse the static shape characteristics of a subject, but STS-DM analyses the local and global dynamic motion characteristics with a consideration of arm-swing in addition to static shape characteristics to achieve robustness against more across-day gait variation. The ARPoLK analysis is used to capture the local dynamic gait signature. Since the upper body dynamics also play a significant role in gait recognition [15], similar to the method in [46] STS-DM uses the orientation angle, aspect ratio, area and eccentricity of the ellipses fitted to five segments of a subject's silhouette contour to analyse the shape and motion characteristics of the entire body. The advantage of contour-based ellipse-fitting over region-based ellipse-fitting as in [46] is low computational complexity. Following the attempt in [38] which combines static and dynamic gait signatures, STS-DM uses weight-based sum rule of score-level fusion to fuse the match scores obtained by different classifiers for subject identification. To demonstrate the efficacy of STS-DM in terms of robustness against most of the challenging factors that affect existing gait recognition systems, it is compared with several related state-of-the-art gait recognition methods.

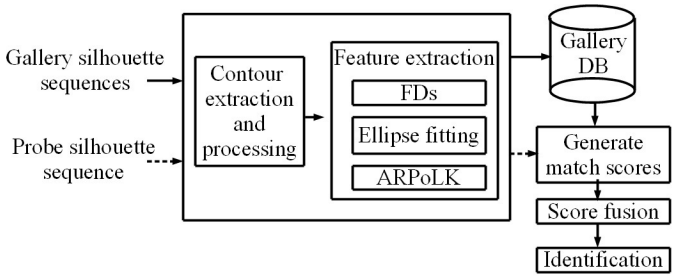


Figure 4.1: Overview of STS-DM.

4.3 Proposed method: STS-DM

STS-DM comprises three modules as shown in Fig. 4.1. Module 1 extracts and postprocesses silhouette contours. Module 2 extracts gait features in three phases. Phase 1 analyses spatio-temporal changes of a subject's shape based on PWMS of FDs of the silhouette contours to generate a dissimilarity score. Phase 2 performs full-body shape and motion analysis, and compares probe and gallery gait signatures by BDHM. Phase 3 uses DTW to measure similarity between ARPoLKs of the probe and gallery subjects. The match scores generated in three phases are fused using weight-based score-level fusion in module 3 for subject identification.

4.3.1 Module 1: Extract and postprocess silhouette contours

The performance of a contour-based method can be substantially enhanced if the contours are extracted from high quality silhouettes, i.e., silhouettes without shadows, missing body parts and parts of the background [107]. Thus, the silhouettes from the datasets used to evaluate the performance of STS-DM are improved using eigenstance reconstruction model [26; 107]. The silhouette is then subjected to vertices traversal algorithm based on connectivity [93] to extract its extreme outer boundary, i.e., contour. To remove camera depth variation, the image is cropped according to the perimeter of the bounding rectangle enclosing the contour and resized to a fixed height while retaining the aspect ratio (i.e., ratio of silhouette width to its height) using bilinear interpolation. The retainment of aspect ratio ensures the maintenance of the proportional relationship between the width and height of a silhouette to preserve actual silhouette shape characteristics, which is a very important factor in shape based subject classification. The resized contour is then copied to a destination image of fixed size by coinciding its centre-of-mass with the centre of the destination image to make it translation invariant.

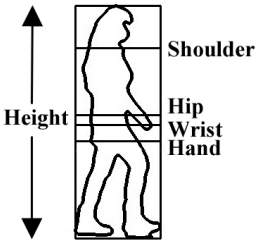


Figure 4.2: Anatomical positions of shoulder, hip, wrist and hand as a fraction of the subject's height are denoted by horizontal lines on a lateral-view of a walking subject's contour.

4.3.2 Phase 1 of Module 2: Analyse shape using FDs

The global shape of a subject is analysed using FDs at ten phases of a gait period as illustrated in Fig. 3.4. The pictorial illustration and detailed description of stanse and swing phases of a gait period, and the criteria for selecting ten specific phases from the stanse and swing phases are provided in STM-SPP (and presented in Chapter 3). Similar to the method in [22], a gait period is determined by the number of frames between two frames of a gait sequence with the most foreground pixels enclosed in the region bounded by bottom of the bounding rectangle and the anatomical position of just before the subject's hand measured from the bottom (i.e., $0.377H$ where H is height of the bounding rectangle) because this foreground region, i.e., the bottom segment of the bounding rectangle is not distorted by self-occlusions due to arm-swing (see Fig. 4.2). The anatomical positions are determined when the subject is standing erect and at rest, with feet together and arms at the side, and the head and the palms of the hands facing forward. In Fig. 4.2, horizontal lines are used to denote the anatomical positions of shoulder, hip, wrist and hand as the fractions of a subject's height, i.e., $0.818H$, $0.530H$, $0.485H$ and $0.337H$, respectively, measured from the bottom of the bounding rectangle [97]. Note that these positions, which are based on anthropometry, might slightly deviate from the actual positions of the shoulder, hip, wrist and hand of a subject especially when the subject is performing an activity, e.g., walking as illustrated in Fig. 4.2.

The Krawtchouk moments of order $(n+m)$ of a $N \times M$ silhouette with intensity function $f(x, y)$ are computed using the sets of weighted Krawtchouk polynomials $\overline{K}_n(x; p, N)$ and $\overline{K}_m(x; p, M)$ as [65; 108]

$$Q_{nm} = \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} \overline{K}_n(x; p1, N-1).\overline{K}_m(y; p2, M-1).f(x, y), \tag{4.1}$$

where $n = 0, 1, ..., N$ and $m = 0, 1, 2, ..., M$. The set of weighted Krawtchouk polynomials,

i.e., $\overline{K}_n(x; p, N)$, is defined as

$$\overline{K_n}(x; p, N) = K_n(x; p, N)\sqrt{\frac{w(x; p, N)}{\rho(n; p, N)}}, \text{where} \quad p \in (0, 1). \tag{4.2}$$

$w(x; p, N)$ and $\rho(n; p, N)$ in Eq. 4.2 are respectively defined as

$$w(x; p, N) = \binom{N}{x} p^x (1-p)^{N-x} \tag{4.3}$$

and

$$\rho(n; p, N) = (-1)^n \left(\frac{1-p}{p}\right)^n \frac{n!}{(-N)_n}, \quad n = 1, 2, ..., N. \tag{4.4}$$

Krawtchouk moments have better image reconstruction capability than the Zernike and Hu moments in both noisy and noise-free conditions, and the orthogonal property of weighted Krawtchouk moments ensures the minimal information redundancy [65; 108]. The scale and rotation dependency of Krawtchouk transform do not affect the extracted features as STS-DM considers only lateral views of silhouettes to achieve rotation invariance, and the silhouettes are pre-scaled and centre-aligned to achieve scale invariance. The Krawtchouk moments are also useful when dealing with partially distorted frames of a gait period, as they have the ability to extract local features from any ROI of an image by varying the parameters N and M.

The silhouettes of the ten phases as shown in Fig. 3.4(a)-(j) are manually extracted. The bottom segment of the bounding rectangle is set as the Rf-ROI and the same silhouette segments of all frames of a subject's gait period are each referred to as a Tr-ROI. Unlike STM-SPP [7] which uses contour matching based on Hu moments for the detection of ten phases, STS-DM computes weighted Krawtchouk moments of each of the Rf-ROIs and Tr-ROIs using Eq.(4.1) by suitably choosing the values of N (say, c) and M (say d) (such that they respectively denote the width and height of the bottom segment of the bounding rectangle) of order (c+d) using $p = 0.5$.

To obtain the ten phases of a gait period of any gait sequence automatically, the Rf-ROIs are compared with the Tr-ROIs using silhouette comparison based on weighted Krawtchouk moments to obtain similarity scores [93]

$$S_{score} = \left[(\text{Rf-ROI}_{k_{nm}} - \text{Tr-ROI}_{k_{nm}})^2\right]^{\frac{1}{2}}, \tag{4.5}$$

where $\text{Rf-ROI}_{k_{nm}}$ and $\text{Tr-ROI}_{k_{nm}}$ respectively denote the (c+d) order weighted Krawtchouk moments of the Rf-ROI and Tr-ROI. The frame whose Tr-ROI results in the lowest S_{score} with the corresponding Rf-ROI is extracted as one of the ten phases, and the process con-

tinues by comparing the next Rf-ROI with the remaining Tr-ROIs until all ten phases are obtained.

The DFT of a contour results in a set of complex numbers, i.e., FDs which represent the shape of the contour in the frequency domain. FDs can be used to reconstruct the shape of the contour and are thus useful boundary shape descriptors for object recognition. Since the low-frequency (i.e., low-order) FDs contain global shape characteristics and the higher frequency (i.e., higher order) FDs increasingly contain finer shape details, a subset of FDs substantiates the discrimination between different shapes. Hence, we characterise a subject's contour using FDs to take into account of spatio-temporal change in the subject's shape over a gait period.

The silhouette contour points are traversed anticlockwise and each point with coordinates (x, y) is represented by a complex number $c(t) = x(t) + jy(t)$, where $t = 0,1,2,...,T-1$ and T is the number of contour points. The FDs are

$$a(u) = \frac{1}{T}\sum_{t=0}^{T-1} c(t)e^{-i2\pi ux/T}, \quad \text{for} \quad u = 0, 1, 2, ..., T-1, \tag{4.6}$$

where u is frequency variable. The original contour is restored by its inverse discrete Fourier transform, i.e.,

$$c(t) = \sum_{u=0}^{T-1} a(u)e^{i2\pi ux/T}, \quad \text{for} \quad t = 0, 1, 2, ..., T-1. \tag{4.7}$$

To ensure that all ten contours of a gait period are represented by a similar set of equal number of points, each contour is approximated by $T = 2^8$, i.e., 256 points using interpolation based on point correspondence analysis [91].

The magnitude and phase of FDs are respectively

$$|a| = [R_a^2(u) + I_a^2(u)]^2 \text{ and } \phi(a) = \tan^{-1}\left[\frac{I_a(u)}{R_a(u)}\right], \tag{4.8}$$

where $R_a(u)$ and $I_a(u)$ are the real and imaginary components of $a(u)$, respectively. The dynamic range of the magnitude spectrum is compressed using log operation and the resulting spectrum is translated to the centre of the Fourier plane to enhance its display in Fig. 4.3(e)-(g).

Fig. 4.3(b)-(e) respectively show the reconstruction of contours using 2^4, 2^5, 2^6 and 2^7 FDs of subject 1's original contour in Fig. 4.3(a), and Fig. 4.3(k)-(n) respectively show the reconstruction of contours using 2^4, 2^5, 2^6 and 2^7 FDs of subject 2's original contour in Fig. 4.3(j). Note that the use of just a few low-frequency FDs, e.g., 2^4 FDs results in very similar contours without any inter-subject discriminatory shape characteristics. However, as the number of FDs increases, the contour gradually regains its original shape characteristics,

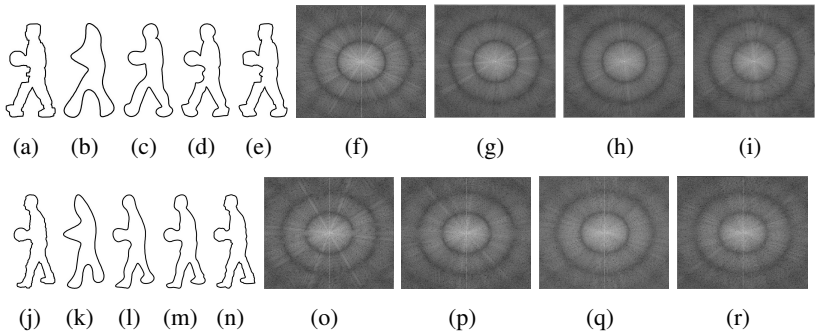


Figure 4.3: Reconstruction of contours using different number of FDs for subject 1(row 1) and subject 2 (row 2) from CMU MoBo dataset: (a) and (j) Original contours with 2^8 points. Reconstructed contours using: (b) and (k) 2^4 FDs; (c) and (l) 2^5 FDs; (d) and (m) 2^6 FDs; (e) and (n) 2^7 FDs. Magnitude spectra of the contours with: (f) and (o) 2^4 FDs; (g) and (p) 2^5 FDs; (h) and (q) 2^6 FDs; (i) and (r) 2^7 FDs.

and hence discriminability between different subjects also increases. Since global shape information is preserved in the low-frequency FDs, the use of the first half of the FDs for reconstruction results in the reconstructed contour which is almost the same as the original. We thus use an ideal lowpass filter to retain the first $T/2$, i.e., 2^7 FDs as they contain adequate subject-specific shape characteristics, while removing the higher-frequency FDs which contain the finer shape details. Their removal reduces flickering noise at the contour and smoothes contours from clothing curvatures. It also reduces the number of contour points to process.

The magnitude spectrum is multiplied by the corresponding phase for generating PWMS, the first gait signature. With initial contact (Fig. 7.3(a)) as the first phase of a gait period, PWMS provides the greatest variability between subjects, as it conveys additional information about temporal deformation of the sequence of shapes together with its frequency contents [32]. PWMS are represented as a $o \times k$ matrix, where o represent the ten phases and $k = T/2$, i.e., 128. Let $\mathbf{A}$ and $\mathbf{B}$ be two such matrices for a gallery and a probe gait sequences, respectively. The dissimilarity score between them is

$$\mathrm{d}_{\mathrm{PWMS}} = \frac{\sum_{i=1}^{o}\sum_{j=1}^{k}(\mathbf{A}_{i,j} - \mathbf{B}_{i,j})^2}{\sum_{j=1}^{k}(\sum_{i=1}^{o}(\mathbf{A}_{i,j} - \mathrm{mean}(\mathbf{A}_j))^2)}, \tag{4.9}$$

where $\mathbf{A}_j$ is jth column vector of $\mathbf{A}$, and mean(.) computes the average of the column vectors of $\mathbf{A}$. The range of $\mathrm{d}_{\mathrm{PWMS}}$ is [0,1], the smaller the value the more similar are the two shapes. We obtain dissimilarity scores by comparing the sequence of a probe subject with each sequence of the gallery subjects for a gait period, and the average dissimilarity score is used for classification.

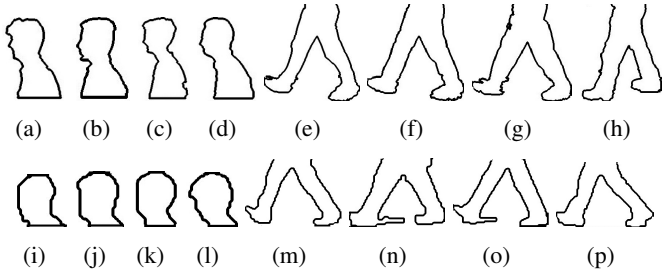


Figure 4.4: Row 1: (a)-(d) upper segments, and (e)-(h) lower segments of four subjects from CMU MoBo dataset to exclude the ball. Row 2: (i)-(l) upper segments, and (m)-(p) lower segments of four subjects from CASIA gait dataset B to exclude the backpack.

When either a gallery or a probe subject carries an item, certain parts of the silhouette are altered and the discriminability of the gait recognition algorithm decreases with respect to the affected parts of the silhouette. Therefore, we introduce a component-based FD analysis based on anatomical studies of human body to make STS-DM robust to carrying conditions. If a subject carries a small item with folded arms or a backpack, the shape of the silhouette above the anatomical position of shoulder, i.e., $0.182H$, and below the anatomical position of wrist, i.e., $0.515H$, measured from top of the bounding rectangle are not affected. The validity of this assumption is experimentally verified on the CMU MoBo dataset for the subjects carrying a ball with folded arms and on CASIA gait dataset B [109] for the subjects carrying a backpack. Thus, carrying a backpack or a small item with folded arms by a subject can be detected by analysing the difference in the number of contour points enclosed in the region bounded by the top of bounding rectangle and the anatomical position of wrist. It is experimentally shown that the upper segment of the silhouette enclosed between $0.225H$ from top of the bounding rectangle, and the lower segment enclosed between $0.500H$ and bottom of the bounding rectangle are not affected by the presence of a ball for all twenty four subjects walking with ball in the CMU MoBo dataset (Fig. 4.4(a)-(h)). Thus for component-based FD analysis, the shape of a silhouette is segmented into an upper segment spanning from top of the bounding rectangle up to the shoulder, and a lower segment spanning from the anatomical position of wrist to bottom of the bounding rectangle. The average of the dissimilarity scores of these components is used for subject classification.

The shape of a silhouette above the wrist is not altered when the subject's hand carries a small bag. According to anthropometry, the position of the wrist is estimated to be $0.485H$ [97] measured from bottom of the bounding rectangle. Thus, an analysis of silhouettes enclosed in the region (1-$0.485H$), i.e., $0.515H$, measured from the top of the bounding rectangle using FDs, remove the shape variation due to carrying briefcase among

the gallery and probe sequences. Analysing the leg region, i.e., between aH and bH (where a = 0.750 and b =0.9375 [27]), removes the effect of shape distortion due to the presence of a briefcase and shadows under feet. Thus, STS-DM detects the presence of a small bag by examining the difference in the number of contour points enclosed in the region between $0.515H$ and $0.750H$. A substantial increase in the number of contour points (e.g., for USF dataset an increase of at least twenty points of non approximated contour and for the same phase of a gait period between gallery and probe sequences) confirms the presence of a briefcase.

It is to be noted that the method in [45] analyses symmetry changes in double helical signatures at the limb region to take into account of shape distortion due to a specific carrying condition, e.g., a briefcase by an upright subject. The use of synthetic gait templates in GEI and CGI, manually computed by simulating distortions in the lower body part of the silhouettes, enables these methods to achieve invariance to the distortions in the lower part of the body, but not in the upper part, e.g., due to carrying conditions with folded arms. The component-based FD analysis in STS-DM and part-based EFD analysis in STM-SPP [7] both rely on anatomical studies of human body to achieve invariance to carrying conditions. However, STM-SPP takes into account of carrying conditions either with folded arms or in upright position, but does not consider subjects carrying a backpack. Hence, STS-DM provides the most in-depth analysis of invariance to carrying condition by taking into account of all common types of small items carried by a subject on the back, with folded ams and also in upright position.

4.3.3 Phase 2 of Module 2: Analyse full-body shape and motion

Undoubtedly, lower body dynamics capture the most distinguishable gait characteristics, but consideration of shape and motion characteristics of upper body enhances it. Therefore, the shape and motion characteristics of the full-body contour is analysed by parts at ten phases of a gait period for extracting global gait signatures. To take into account of change in appearance of different parts of a contour due to walking, the contour is divided into four regions with each region fitted with an ellipse. An ellipse is preferred to a circle and a rectangle as it has more useful parameters to describe shape characteristics (i.e., aspect ratio, area and eccentricity) and motion characteristics (i.e., orientation angle, the angle of the semi-major axis of the ellipse measured anti-clockwise from the positive horizontal axis). Also, ellipse fitting approach is robust to limited distortions at the contour due to poor segmentation, and enables STS-DM to take into account of subject-specific characteristics, i.e, fat vs slim and long hair vs bald.

The height H of the bounding rectangle enclosing a silhouette contour is used as the subject's height. Following anatomical studies of the human body [97], the vertical

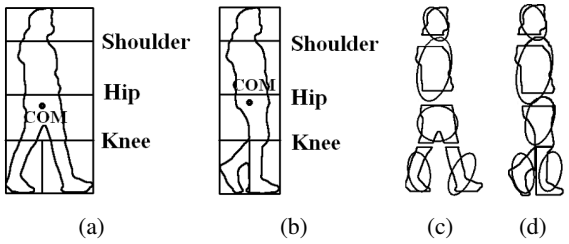


Figure 4.5: (a)-(b) Partitioning a subject's contour into five segments (COM denotes centre-of-mass); (c)-(d) ellipses fitted to each of the five segments.

positions of shoulder, hip and knee measured from the bottom of the bounding rectangle are estimated respectively to be $0.818H$, $0.530H$ and $0.285H$. The bounding rectangle is then subdivided into the following four regions (as shown in Fig. 4.5(a)) by drawing horizontal lines at the anatomical positions of shoulder, hip and knee joints: uppermost region enclosing head and neck; region enclosed between shoulder and hip; region enclosed between hip and knee; and bottommost region enclosing the legs. The bottommost region is subdivided by a vertical line into two regions, each enclosing one leg. The vertical line if extended, passes through the centre-of-mass of the contour. The process is illustrated in Fig. 4.5(a)-(b), where centre-of-mass is abbreviated as COM. The 2D Cartesian moment of order u and v of a contour $I(x, y)$ is

$$m_{u,v} = \sum_{i=1}^{T} I(x, y) x^u y^v. \tag{4.10}$$

The centre-of-mass of the silhouette contour, (x_c, y_c), is given by $x_c = \frac{m_{10}}{m_{00}}$ and $y_c = \frac{m_{01}}{m_{00}}$ [96].

The contour points enclosing each of the five regions representing a body part are best fitted by an ellipse using a non-linear least squares technique as illustrated in Fig. 4.5(c)-(d). The following four parameters of each of the fifty ellipses for the ten phases form the gait signature: aspect ratio; area; eccentricity; and orientation angle.

We compute twenty 1D histograms, each representing the distribution of one parameter of one ellipse (i.e., one segmented region) for ten phases of a gait period. The histograms are normalised to [0,1] as shown in Fig. 4.6. The normalised histograms of the probe gait sequences (Hist-p_n, n=1,..., 20) are compared with the corresponding histograms of the gallery sequences (Hist-g_n) using Bhattacharyya distance metric to obtain the dissimilarity score [93]

$$d_n(\text{Hist-p}_n, \text{Hist-g}_n) = \left(1 - \sum_{i}^{B} \frac{\sqrt{\text{Hist-p}_n(i).\text{Hist-g}_n(i)}}{\sqrt{\sum_i \text{Hist-p}_n(i). \sum_i \text{Hist-g}_n(i)}}\right)^{\frac{1}{2}}, \tag{4.11}$$

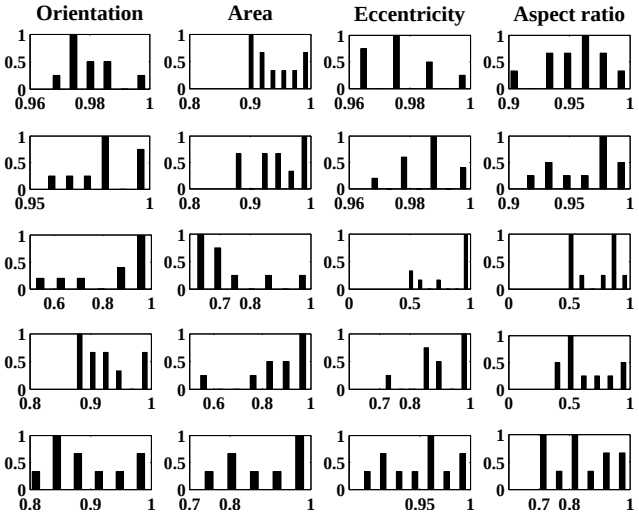


Figure 4.6: The histogram matrix with each column representing histograms of ellipse parameters (orientation angle, area, eccentricity and aspect ratio), each for ten phases, corresponding respectively from top to bottom to the following regions: head-shoulder, shoulder-hip, hip-knee, right leg and left leg.

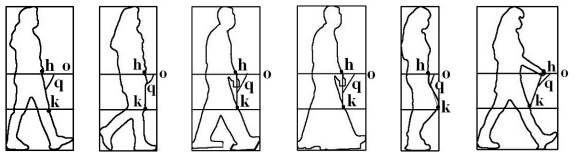


Figure 4.7: Illustration of ARPoLK: (a)-(b) subject 1 at two different phases of a gait period of two different sequences with different hair style and shoe type on different days; (c)-(d) subject 2 with and without shadows under feet in two different gait sequences; and (e)-(f) subject 3 with higher arm-swing.

where B is the number of bins in each histogram. Further information on histogram matching techniques are provided in Appendix A. The second gait signature is the average dissimilarity score

$$\mathrm{d_{BDHM}} = \frac{1}{20}\sum_{n=1}^{20} d_n. \tag{4.12}$$

The range of $\mathrm{d_{BDHM}}$ is [0,1], and the low values of $\mathrm{d_{BDHM}}$ indicate good matches. Hence, ideally a probe subject would result in the lowest $\mathrm{d_{BDHM}}$ for its correct match in the gallery.

4.3.4 Phase 3 of Module 2: Analyse ARPoLK

We analyse ARPoLK to take into account of a subject's lower body dynamics which is robust to the problems associated with self-occlusions. Points h and k in Fig. 4.7 respectively correspond to where horizontal lines drawn across the bounding rectangle at heights of $0.530H$ (denoting hip position) and $0.285H$ (denoting knee position), measured from bottom of the bounding rectangle, meet with the contour facing the direction of walking. The angle of the leading knee θ_{LK} (i.e., q in Fig. 4.7) is subtended by the horizontal line through h and the line joining h and k. Measuring this angle over a gait period gives ARPoLK. The main motivation in using ARPoLK analysis is its ability to capture lower body gait dynamics which remain unaffected by self-occlusion, i.e., occlusion of one knee by another. ARPoLK analysis always takes into account of only the leading knee, i.e., the knee which is at the front in the direction of walking, and does not consider the other knee which might be occluded by the leading knee at times during walking. Fig. 8(b) and (e) demonstrate the successful ARPoLK analysis in the case of an occluded knee by the leading knee. ARPoLK analysis is also invariant to across-day gait variation that affect the subject's shape above the hip and below the knee, e.g., change of hair style and shadows under feet, and takes into account of limited changes in clothing style, such as pant vs skirt or shorts across different days, but not subtle style change like tight vs loose clothing. Fig. 4.7(a)-(b) illustrate that ARPoLK analysis remains unaffected by change of hair style and footwear including the use of high heels, while Fig. 4.7(c)-(d) illustrate similar values of θ_{LK} for the same subject's silhouette with and without shadows under feet. ARPoLK analysis is also invariant to carrying of small items with folded arms as long as the subject's hip is not occluded.

Most but not all gait recognition methods that consider dynamic gait characteristics only focus on the motion of lower limb region, but ignore arm-swing although it is an unavoidable and integral part of gait. In normal walking, the contralateral arm is automatically swung forward with the swinging lower limb at a rate proportional to the walking speed, and different subjects have varying arm swings. Therefore, arm-swing is integrally related to the motion of lower limbs, and contributes to inter-subject discrimination. The method in [16] uses linear Hough transform to model arm-swing as a pendulum motion. But arm-swing can be arguably modelled as a pendulum motion because it is considerably influenced by neuromuscular forces. Although the method uses hip and shin angles to constrain the Hough space, it models the limb motion and arm-swing separately, and therefore, does not consider the integral relationship between them. The method in [65] indirectly addresses the integral relationship between arm-swing and motion of lower limb in gait by holistic image analysis. STS-DM also considers this integral relationship by analysing the lower limb motion in conjunction with the arm-swing using ARPoLK analysis. ARPoLK analysis implicitly addresses the effects of arm-swing as evident from Fig. 4.7(e)-(f).

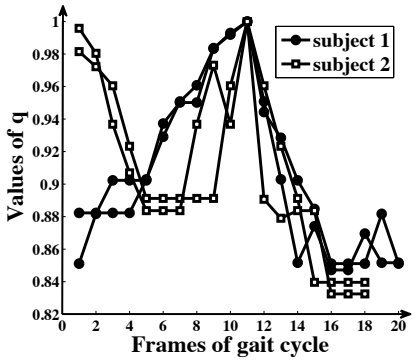


Figure 4.8: The discrete signals representing the ARPoLK for a gait period of each subject: subject 1 and subject 2.

However, the effect of arm-swing is not considered in ARPoLK analysis if the hands of a subject are engaged due to carrying conditions either with folded arms or in upright position. Hence, the subject identification performance of ARPoLK analysis is significantly affected if a subject has a higher arm-swing as in Fig. 4.7(f) in a gallery sequence but his/her hands are engaged by carrying conditions in a probe sequence, and vice versa.

Fig. 4.8 shows the discrete signals obtained by ARPoLK analysis of two different gait sequences corresponding to subject 1 (Fig. 4.7(a)-(b)) and subject 2 (Fig. 4.7(c)-(d)). The discrete signals are formed by plotting the different values of θ_{LK} for a gait period against N equally spaced monotonically increasing values, where N is the number of frames in a gait period. The signals are normalised in the range [0,1] by dividing each θ_{LK} with the maximum value of θ_{LK} to remove spatial scale variation for different subjects for uniform comparison. It is evident from Fig. 4.8 that discrete signals representing ARPoLK of two different gait sequences of the same subject resemble each other, while different subjects have quite dissimilar signals. Thus it is verified that ARPoLK analysis has a very good inter-subject discriminability in the presence of across-day gait variation and shadows under feet.

Different subjects have different walking speeds which result in varying number of frames in their gait period. Depending on the subject's state of mind, the number of frames in a gait period of the same subject also varies due to the walking speed variation in different situations. Therefore, we use DTW to account for such variation in classifying a probe subject based on its similarity of ARPoLK with that of a gallery subject over a gait period. DTW uses dynamic programming to compute a warping function that optimally aligns two time-dependent sequences of varying lengths for measuring similarity. Let $a_g = \theta_{LKg1}, \theta_{LKg2}, ..., \theta_{LKgm}$ and $a_p = \theta_{LKp1}, \theta_{LKp2}, ..., \theta_{LKpn}$ be the discrete signals representing ARPoLK for gait periods of lengths (i.e., number of frames of the gait period) $m \in \mathbb{N}$ and $n \in \mathbb{N}$, respectively of a gallery and a probe subject, where θ_{LKgi} and θ_{LKpj} are angles of the leading knee. DTW constructs an $m \times n$ matrix whose each element corresponds to the

Euclidean distance $d(\theta_{LKgi}, \theta_{LKpj}) = (\theta_{LKgi} - \theta_{LKpj})^2$.

An $m \times n$ warping path is a sequence $p = (p_1, p_2, ..., p_L)$ with $p_l = (\theta_{LKgm_l}, \theta_{LKpn_l}) \in [1:m] \times [1:n]$ for $l \in [1:L]$ and $\max(m,n) \leq L < (m+n)$, for mapping two sequences a_g and a_p which satisfies the followings: (a) boundary condition: $p_1 = (1,1)$ and $p_L = (m,n)$; (b) monotonicity condition: $\theta_{LKgm_1} \leq \theta_{LKgm_2} \leq .... \leq \theta_{LKgm_L}$ and $\theta_{LKpn_1} \leq \theta_{LKpn_2} \leq .... \leq \theta_{LKpn_L}$; and (c) step size condition: $p_{l+1} - p_l \in (1,0), (0,1), (1,1)$ for $l \in [1:L-1]$. DTW minimises the cost of warping a_g and a_p together to form the third gait signature,

$$\mathrm{d_{DTW}} = \min\left(\frac{(\sum_{l=1}^{L} p_l)^{\frac{1}{2}}}{L}\right). \tag{4.13}$$

4.3.5 Module 3: Identify subject

Each of the gallery and probe gait sequences respectively with N_g and N_p frames is partitioned into consecutive subsequences with gallery gait period (G_g) and probe gait period (G_p). The distance metric between the kth probe gait period and a gallery sequence for match score S, where S is either $\mathrm{d_{PWMS}}$, $\mathrm{d_{BDHM}}$ or $\mathrm{d_{DTW}}$, is

$$Dist_S(k) = \min_i(S), \tag{4.14}$$

where i = 1,2,...,n_g and $n_g = N_g/G_g$ is the number of gallery gait periods in a gallery sequence. The median of the distances

$$D_S = median(Dist_S(1), Dist_S(2), ..., Dist_S(m_p)), \tag{4.15}$$

is considered as the match score between the probe sequence and gallery sequence to be used in the score-level fusion for subject identification, where $m_p = N_p/G_p$ is the number of probe gait periods in a probe sequence.

Unlike STM-SPP, which uses a rank-based classifier combination rule to combine the classification results by Procrustes shape analysis and EFDs for identifying a subject, STS-DM uses score-level fusion to fuse the match scores obtained by the PWMS, BDHM and DTW. Since score-level fusion combines the match scores obtained by different classifiers, it is more informative than rank-level fusion. Rank-level fusion is also computationally more expensive and suffers from the drawback of a tie in ranking, which requires further processing to get resolved, e.g., STM-SPP uses Hu moments to resolve a tie in ranking. However, score-level fusion requires the inhomogenous scores obtained by different classifiers to be transformed into a common numerical range before being compared using score normalisation technique. The linear score normalisation techniques, e.g., min-max normalisation and z-score normalisation have similar computational complexities, but

z-score normalisation is preferred in STS-DM as it is less sensitive to outliers than min-max normalisation. Although non-linear score normalisation techniques based on double sigmoid function and hyperbolic tangent are more robust to outliers, they introduce complexity due to the use of many parameters, and the performance of these techniques are highly dependent on the chosen parameter values. Hence, to make a trade-off between the performance and computation complexity, STS-DM obtains z-scores of each of the three match score sets using

$$Z_{classifier} = \frac{\mathrm{d}_{classifier} - \mu_{classifier}}{\sigma_{classifier}}, \tag{4.16}$$

where μ is mean of the set of scores, d is the individual score, σ is standard deviation, and $classifier$ is either PWMS, BDHM or DTW. The three $classifiers$ do not perform equally well as evident from Fig. 4.10 (see Section 4.4 for CMC curve) which shows that PWMS is the best feature, while BDHM is the worst. A weight-based sum rule of score-level fusion [122] is thus used in STS-DM for improved identification rate where the weights are determined based on the contribution of each component $classifier$ to the final subject identification. The fused score is thus obtained using

$$S_f = \frac{\mathrm{I_{PWMS}} \times Z_{\mathrm{PWMS}} + \mathrm{I_{BDHM}} \times Z_{\mathrm{BDHM}} + \mathrm{I_{DTW}} \times Z_{\mathrm{DTW}}}{\mathrm{I_{PWMS}} + \mathrm{I_{BDHM}} + \mathrm{I_{DTW}}}, \tag{4.17}$$

where $\mathrm{I_{PWMS}}$, $\mathrm{I_{BDHM}}$ and $\mathrm{I_{DTW}}$ are the weights that respectively correspond to the CCR (see Section 4.4 for CCR) obtained using the match scores $\mathrm{d_{PWMS}}$, $\mathrm{d_{BDHM}}$ and $\mathrm{d_{DTW}}$ for a particular testing condition. The probe subject is identified based on the lowest S_f it measures with the member of a gallery class.

4.4 Experiments

Since the aim of STS-DM is to demonstrate its combined robustness against most of the challenging factors of gait recognition, it is extensively compared with several related methods that individually address one or more covariate factors. Therefore, to make uniform comparison with several related methods, STS-DM is evaluated using different experimental setup based on the reported available results of those methods on two public datasets: CMU MoBo dataset and USF HumanID gait challenge dataset.

4.4.1 Experiments on CMU MoBo dataset

A closed-set identification guarantees the existence of the subject in the database. We analyse the closed set identification performance of STS-DM on the profile view silhouettes of MoBo dataset by taking one subject as the probe sample and train it on all the subjects of the

dataset including the probe sample. The CCR(%) is obtained using Eq. (3.28). The identification is best interpreted by a CMC curve which shows CCR at different ranks. Since the smaller the values of the match scores the more similar are the two subjects, the CCR at rank r implies that the probability of correct match is among the lowest r match scores.

We use 3D scatter plots as shown in Fig. 4.9 to show the distribution of match scores (plotted along the vertical axis) obtained by PWMS, DTW, BDHM and the fused classifier as a result of comparing each of the fast walking probe subjects (plotted along the horizontal right axis) with all 25 slow walking gallery subjects (plotted along the horizontal left axis) of CMU MoBo dataset. Note that the ith probe subject along the horizontal right axis corresponds to its matching gallery subject i along the horizontal left axis, where i=1,2,3,...,25. The match scores obtained by comparing one probe with all the gallery subjects are represented by circles of the same sizes in the plots, while different circle sizes are used for different probe subjects. Since ideally a probe subject will result in the lowest match score for its matching gallery subject, very few circles are present in the bottom horizontal planes. A probe subject will generate higher match scores for all the non-matching gallery subjects, which explains why the circles of different sizes are cluttered around the higher horizontal planes of the plots.

Note that the number of probe subjects that results in the lowest match scores for their matching gallery subjects using PWMS, DTW, BDHM and fused classifier are respectively 23, 22, 21 and 24 for fast walk vs slow walk. Hence, the rank-1 CCR for PWMS, DTW, BDHM and fused classifier are respectively 23/25*100, i.e., 92%, 22/25*100, i.e., 88%, 21/25*100, i.e., 84% and 24/25*100, i.e., 96% which are verified in Fig. 4.10(a), where Fig. 4.10 shows the CMC curves of CCR obtained using PWMS, DTW and BDHM for three different walking conditions of CMU MoBo dataset, namely fast walk vs slow walk, slow walk vs fast walk, and fast walk vs walking with ball. It is clear that the performance of STS-DM is the best for fast walk vs slow walk using individual classifiers as well as the fused classifier. The rank-1 CCR of PWMS, DTW and BDHM are respectively 92%, 88% and 84% for fast walk vs slow walk; 88%, 84% and 84% for slow walk vs fast walk; and 87%, 83% and 79% for fast walk vs walking with ball. Since PWMS outperforms DTW and BDHM, it is shown that the shape of a subject provides better inter-subject discriminative characteristics than its kinematics in the case of very limited across-day gait variation. Fig. 4.10(d) shows that CCR is significantly improved, i.e., 96%, 96% and 92% respectively for fast walk vs slow walk, slow walk vs fast walk and fast walk vs walking with ball by fusing the results of individual classifiers using weight-based sum rule of score-level fusion.

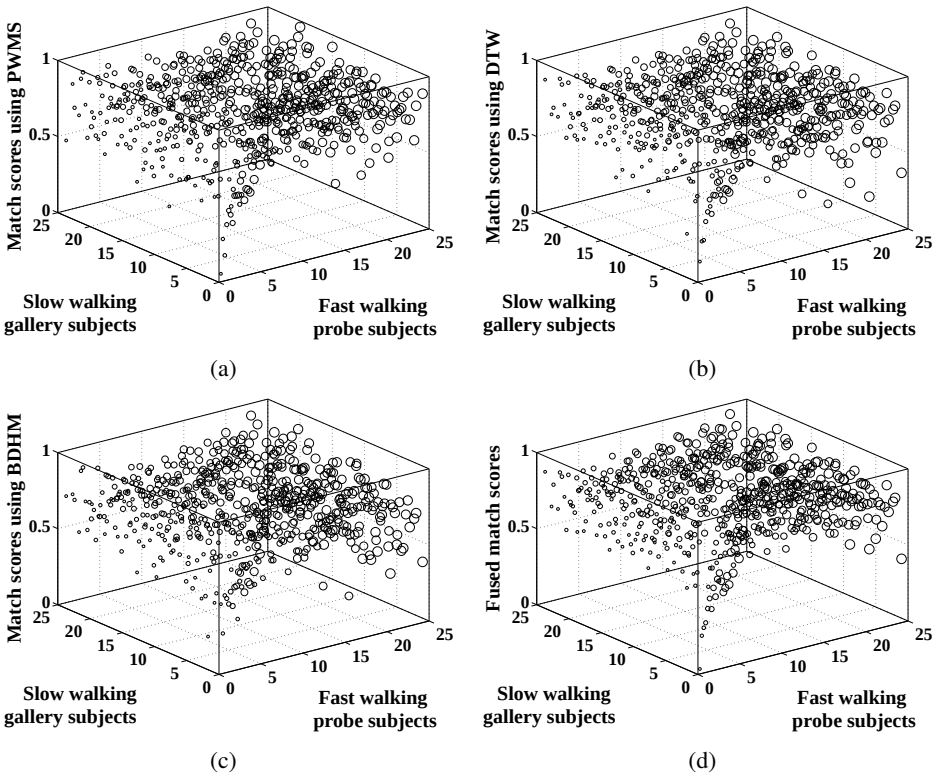


Figure 4.9: Distribution of match scores obtained by (a) PWMS, (b) DTW, (c) BDHM and (d) fused classifier for fast walk vs slow walk of lateral-view silhouettes of CMU MoBo dataset.

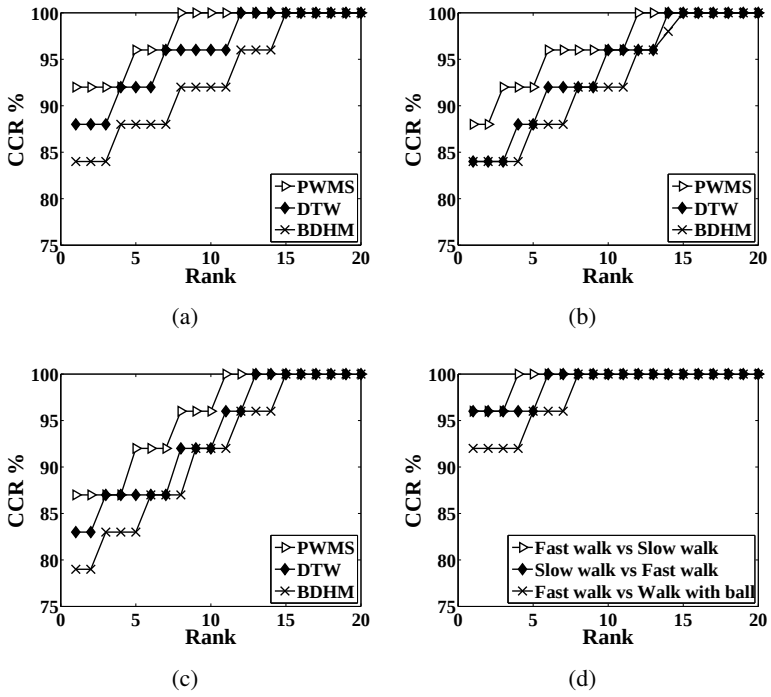


Figure 4.10: CMC curves of classification rates obtained using PWMS, BDHM and DTW of the lateral-view silhouettes from CMU MoBo dataset for (a) fast walk vs slow walk; (b) slow walk vs fast walk; and (c) fast walk vs walking with ball. (d): CMC curves of combined classification rates of (a)-(c) using weight-based sum rule of score-level fusion.

Table 4.1: Top-rank identification rates (in percentage) on CMU MoBo dataset with the rates of SSM from [23], Baseline from [22], CMU1 from [19], SSP from [21], STM-SPP from [7] and SVB frieze from [40] for the lateral view. Keys: 'G' - Gallery sequence; 'P' - Probe sequence; 'S' - Slow walk; 'F' - Fast walk; 'B' - Walk with ball.

G/P	SSM [23]	Baseline [22]	CMU1 [19]	SSP [21]	STM-SPP [7]	SVB frieze [40]	STS-DM
S/S	100	92		100	100	100	100
F/F	100	-		100	100	100	100
B/B	92	-		-	100	100	100
S/F	80	72	76	54	94	82	96
F/S	84	-		32	91	80	96
S/B	48	88	92	-	93	77	92
B/S	68	-		-	82	89	92
F/B	48	-		-	84	61	92
B/F	48	-		-	82	73	87

Comparisons

The performance of STS-DM on the lateral view of silhouettes of the CMU MoBo dataset is compared with SSM [23], SSP [21], STM-SPP [7] and SVB frieze [40]. Table 4.1 shows that the shape based approach in SSM using stance correlation for the subjects walking parallel to the image plane is robust to variation in walking speed, but its performance degrades significantly when the shape of the silhouettes change due to different activities (e.g., fast walk vs walking with ball). Since part-based shape analysis using EFDs and component-based shape analysis using FDs respectively aid STM-SPP and STS-DM to achieve invariance to carrying small items, they significantly outperform SSM and SVB frieze. The superiority of STS-DM over STM-SPP is attributed to the analyses of dynamic motion characteristics of silhouettes using ellipses fitted to various body parts and ARPoLK that enable it to achieve robustness against limited variation in clothing.

The method in [21] computes image SSP, i.e., correlation of corresponding pairs of images in two gait sequences of a subject. To make uniform comparison with SSP which is robust to segmentation noise, STS-DM also uses split-sample cross validation technique like SSP, where gallery and probe sets correspond to different combination of walking types for each of the twenty-five subjects. Since STS-DM is defined only on profile view of the silhouettes, we consider profile view of two sequences per subject (total 50 sequences) walking at slow pace (2.06 miles/h) and fast pace (2.82 miles/h). Table 4.1 shows that the performance of SSP for profile view degrades significantly when the probe and gallery samples differ in walking speed. STS-DM outperforms SSP by analysing shape and motion

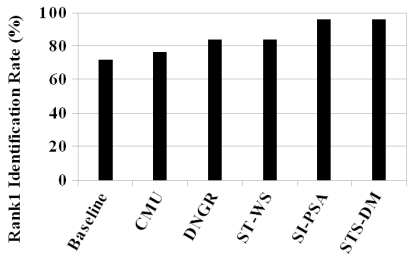


Figure 4.11: Comparison with related works. Baseline [22], CMU1, DNGR [26] and STS-DM are evaluated on CMU1 MoBo gait dataset (experiment 2 of CMU1) with walking speed variation of 3.3 km/h and 4.5 km/h, while ST-WS [43] and SI-PSA [42] are evaluated on OU-ISIR treadmill gait dataset A [76] with walking speed variation of 3 km/h and 4 km/h between gallery and probe gait sequences.

characteristics of ten phases of a gait period and using DTW for ARPoLK analysis so as to overcome the effects of walking speed variation.

To demonstrate robustness against speed variation by comparative experimental analysis with the related speed-invariant methods, STS-DM is evaluated using the experiment 2 defined by CMU1 as it enables evaluation of a gait recognition method across different speeds. The rank-1 identification rates of STS-DM, speed-invariant method DNGR, CMU1 [19] and Baseline [22] are respectively 96%, 84%, 76% and 72% (see Fig. 4.11) for the slow (3.3km/h) vs fast (4.5km/h) walking gait sequences of the profile view silhouettes of CMU MoBo dataset, where the rates of DNGR, CMU1 and Baseline are based on experiment 2 of CMU1 entitled "Across gaits condition". Since speed variation in this experiment is almost 1km/h, we compare STS-DM with speed-invariant methods ST-WS [43] and SI-PSA [42], that are evaluated on OU-ISIR treadmill gait dataset A [76] with walking speed variation of 3 km/h and 4 km/h between gallery and probe gait sequences. It is clear from Fig. 4.11 that STS-DM outperforms all other methods, and provides equal rank-1 identification rate as SI-PSA.

4.4.2 Experiments on USF HumanID gait challenge dataset

STS-DM is evaluated on both the small version (452 sequences from 74 subjects, data acquired in May only) and the full version (1870 sequences from 122 subjects, data acquired in May and November) of USF HumanID gait challenge dataset [22].

Comparisons

Table 4.2 shows the results on the full version of USF dataset in terms of identification rate (P_I) at ranks 1 and 5, to enable a comparison with the state-of-the-art methods, i.e.,

Table 4.2: Identification rates (in percentage) on full version of USF HumanID gait challenge dataset using the gallery set (G, A, R, NB, M/N) of 122 subjects. The rates for GTDA-GF, GEI, RCK-G, CGI, STM-SPP, DNGR and MMFA are from [50], [25], [65], [28], [7], [26] and [111], respectively.

Exp	Identification Rate (P_I)%							
	GTDA-GF(GEI)		RCK-G(MMFA)		CGI(STM-SPP)		DNGR(STS-DM)	
	Rank1	Rank5	Rank1	Rank5	Rank1	Rank5	Rank1	Rank5
A	91(90)	98(94)	83(89)	96(98)	91(92)	97(96)	85(93)	96(97)
B	93(91)	99(94)	86(94)	94(98)	93(95)	96(98)	89(96)	94(98)
C	86(81)	97(93)	78(80)	88(94)	78(84)	94(95)	72(86)	89(96)
D	32(56)	68(78)	39(44)	66(76)	51(72)	77(80)	57(70)	85(82)
E	47(64)	68(81)	34(47)	63(76)	53(68)	77(84)	66(69)	81(83)
F	21(25)	50(56)	20(25)	51(57)	35(29)	56(59)	46(39)	68(61)
G	32(36)	56(53)	21(33)	46 (60)	38(40)	58(61)	41(37)	69(60)
H	95(64)	95(90)	43(85)	66(95)	84(69)	98(92)	83(78)	96(95)
I	90(60)	99(83)	40(83)	68(93)	78(60)	97(84)	79(71)	95(89)
J	68(60)	84(82)	40(60)	65(84)	64(64)	86(85)	52(66)	79(83)
K	16(6)	40(27)	16(27)	44(48)	3(20)	27(30)	15(27)	46(39)
L	19(15)	40(21)	5(21)	22(39)	9(18)	24(27)	24(22)	39(28)
WAvgI	61(58)	78(76)	44(60)	67(80)	62(63)	79(79)	63(67)	82(80)

GTDA-GF [50], GEI [25], RCK-G [65], CGI [28], STM-SPP [7], DNGR [26] and MMFA [111]. The method using matrix-based marginal Fisher analysis (MMFA) in [111] applies marginal Fisher analysis on GEIs for gait representation to reduce the dimensionality of the feature space and extends marginal Fisher analysis to marginal based analysis for content-based image retrieval. Table 4.3 shows the results on the full version of USF dataset in terms of verification rate (the probability that the method successfully detects the correct match between the probe and gallery sequences, i.e., P_V) at false alarm rates (the probability that the method incorrectly classifies a probe sequence to a non-matching gallery sequence) 1% and 10% for Baseline, DNGR and STS-DM. Since the number of probe subjects in the gait challenge experiments A-L varies, we compute the weighted average identification rate, i.e., WAvgI, and the weighted average verification rate, i.e., WAvgV, by using Eq.(3.26) and Eq.(3.27), respectively.

The identification rates achieved by GEI for the twelve challenge experiments after combining the real and synthetic gait features are presented in Table 4.2. GTDA-GF reports the identification rates obtained by applying GTDA as a preprocessing step of linear discriminant analysis on the magnitude of the result of convolving a GEI with sum of Gabor functions over scales with direction fixed. Table 4.2 shows that STS-DM outperforms all

Table 4.3: Verification rates at a false alarm rate (P_F) of 1% and 10% for Baseline from [22], DNGR from [26], STM-SPP from [7] and STS-DM on full version of USF HumanID gait challenge dataset using the gallery set (G, A, R, NB, M/N) of 122 subjects.

Exp.	Verification Rate (P_V)% at			
	Baseline	DNGR	STM-SPP	STS-DM
	P_F:1(10)%	P_F:1(10)%	P_F:1(10)%	P_F:1(10)%
A	82(94)	93(98)	88(100)	94(100)
B	87(94)	94(98)	94(100)	97(100)
C	65(94)	80(94)	86(98)	88(98)
D	44(80)	68(96)	80(94)	79(94)
E	35(76)	62(90)	74(84)	76(84)
F	20(60)	53(86)	50(82)	66(82)
G	28(55)	43(79)	52(76)	62(76)
H	72(91)	91(99)	83(95)	85(95)
I	67(85)	86(97)	76(93)	76(93)
J	48(76)	58(92)	65(92)	68(92)
K	6(24)	27(61)	21(58)	29(58)
L	6(24)	24(46)	19(52)	25(52)
WAvgV	51(76)	70(91)	70(89)	75(90)

other methods for experiment A with a variation in view, and performs reasonably better than GTDA-GF, GEI, RCK-G, CGI and STM-SPP for experiments K and L with a variation in clothing. However, STS-DM is outperformed by GTDA-GF, CGI, DNGR and MMFA for the gait challenge experiments H and I. This is because ARPoLK analysis with a consideration of subject's arm-swing is particularly affected by the briefcase carrying condition as it prevents normal arm-swing and distorts the shape of a silhouette between hip and knee. Since the gait challenge experiments H and I take into account of briefcase covariate, the performance of STS-DM is degraded in these experiments. The aim of STS-DM is to achieve combined invariance to most of the challenging factors of gait recognition with low computational complexity, and not to achieve the best identification rates for every gait challenge experiment among the state-of-the-art gait recognition methods. The superiority of STS-DM to other methods in terms of WAvgI and WAvgV is demonstrated in Table 4.2 and Table 4.3. Table 4.2 shows that STS-DM achieves the highest WAvgI at rank-1, followed by DNGR, STM-SPP and CGI, and is only second to DNGR in terms of WAvgI at rank-5. It is clear from Table 4.3 that in terms of WAvgV STS-DM outperforms other methods at the false alarm rate of 1%.

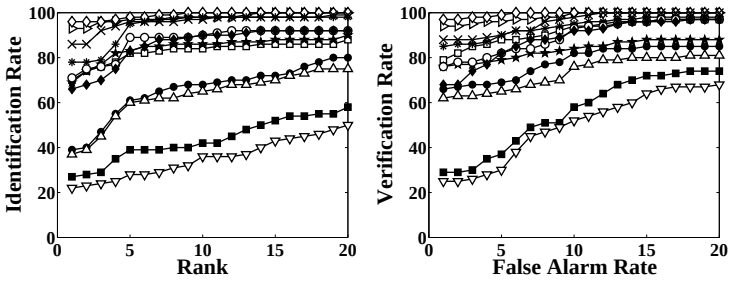
Table 4.4 shows the results on the small version of the dataset (No-Briefcase data) to enable a comparison with Baseline, CASIA [29], CMU2 [89], CMU1 [19], GEI [25], ASM

Table 4.4: Top-rank identification rates (in percentage) on the small version of USF HumanID gait challenge dataset (data acquired in May only) using the Gallery Set (G, A, R) of 71 subjects. The rates for Baseline, CASIA, CMU2, CMU2, RCK-G, GEI, ASM and STM-SPP are from [22], [29], [89], [89], [65], [25], [35] and [7], respectively. Unlike others, identification rates with '$*$' are not based on silhouettes provided by USF HumanID gait challenge dataset.

Method	Rank-1 Identification Rate (P_I) %							WAvgI
Baseline	87	81	54	39	33	29	26	50.62
CASIA	70.42	58.54	51.22	34.33	21.43	27.27	14.29	40.83
CMU2	85	81	60	23	17	25	21	44.93
CMU1	87	81	66	21	19	27	23	46.44
RCK-G	97	89	83	41	34	30	28	57.53
GEI	100	90	85	47	57	32	31	62.83
ASM	97*	95*	91*	92*	86*	85*	78*	89.66
STM-SPP	100	94	89	73	69	40	36	71.74
STS-DM	100	98	91	76	70	47	42	74.99

[35] and STM-SPP [7]. We present the identification rates at rank-1 of CMU2 obtained by weighted correlation similarity measure, and the identification rates of GEI obtained by fusing real and synthetic gait templates. Table 4.4 shows that STS-DM achieves the second highest WAvgI following ASM. All methods listed in Table 4.4 except ASM use the silhouettes provided by the USF HumanID gait challenge dataset for uniform comparison. Since these silhouettes are significantly affected by strong shadows under feet (mainly due to the subjects walking on a concrete surface as in the gait challenge experiments D, E, F and G) the methods that directly use the silhouettes provided by the USF HumanID gait challenge dataset do not provide satisfactory recognition rates for these experiments. ASM employs hierarchical prediction-based ASM framework with Kalman filter to extract the foreground which is unaffected by shadows, and analyses its model parameters for gait recognition. Hence, the superiority of ASM for the gait challenge experiments D, E, F and G is attributed to the use of shadow-free good quality silhouettes for feature extraction. Also, the feature extraction and classification processes involved in ASM are much more computationally expensive compared to STS-DM. Disregarding the performance of ASM, Table 4.4 shows that STS-DM outperforms all the methods for all the gait challenge experiments.

The performance of STS-DM for the twelve challenge experiments of the dataset is measured by identification mode and verification mode, using CMC and ROC curves respectively, following [106]. Fig. 4.12(a) shows that the identification rates of STS-DM range from 22% to 96% at rank-1, and 28% to 98% at rank-5. Fig. 4.12(b) shows that the



(a) (b)

Figure 4.12: Performance on twelve challenge experiments of USF dataset. (a) Identification mode (CMC) and (b) Verification mode (ROC). Keys: '▷'- Exp. A (Probe: G, A, L, NB, M/N); '◇'- Exp. B (Probe: G, B, R, NB, M/N); '×'- Exp. C (Probe: G, B, L, NB, M/N); '□'- Exp. D (Probe: C, A, R, NB, M/N); '⋆'- Exp. E (Probe: C, B, R, NB, M/N); '●'- Exp. F (Probe: C, A, L, NB, M/N); '△'- Exp. G (Probe: C, B, L, NB, M/N); '∗'- Exp. H (Probe: G, A, R, BF, M/N); '○'- Exp. I (Probe: G, B, R, BF, M/N); '◆'- Exp. J (Probe: G, A, L, BF, M/N); '■'- Exp. K (Probe: G, A/B, R, NB, N); and '▽'- Exp. L (Probe: C, A/B, R, NB, N).

verification rates of STS-DM range from 25% to 97% at a false alarm rate of 1%, and 52% to 100% at a false alarm rate of 10%. Table 4.2 and Fig. 4.12 show that STS-DM is least affected by variation in shoe types, followed by about 30° change in view. However, time (i.e., when the dataset was generated) has the most impact on the performance of STS-DM, as it implies variation in clothing and footwear of the same subject.

4.4.3 Effect of missing frames

Occlusions in the scene, large shadows under feet and extreme lighting variation can severely distort the extracted contours. If these distorted contours are not part of any of the ten phases of a gait period, they do not affect the classifications using FDs and ellipsoidal fits. If the distortion causes any frame of the ten phases to be missing, its immediate adjacent frame is considered. ARPoLK analysis is not affected if the portion of the contour enclosed in the region between hip and knee remain undistorted. It is also not affected by any missing or discarded frame due to excessive distortions resulting in different lengths of gait sequences. This is because the use of DTW in ARPoLK analysis enables detection of similarity between two sequences of varying lengths. Hence, STS-DM is robust to severely distorted and missing frames.

To support the claim that STS-DM is robust to missing frames by experimental results, we create probe gait sequences of shorter lengths from CMU MoBo dataset by discarding frames at a specified interval in order to stimulate a situation where probe frames

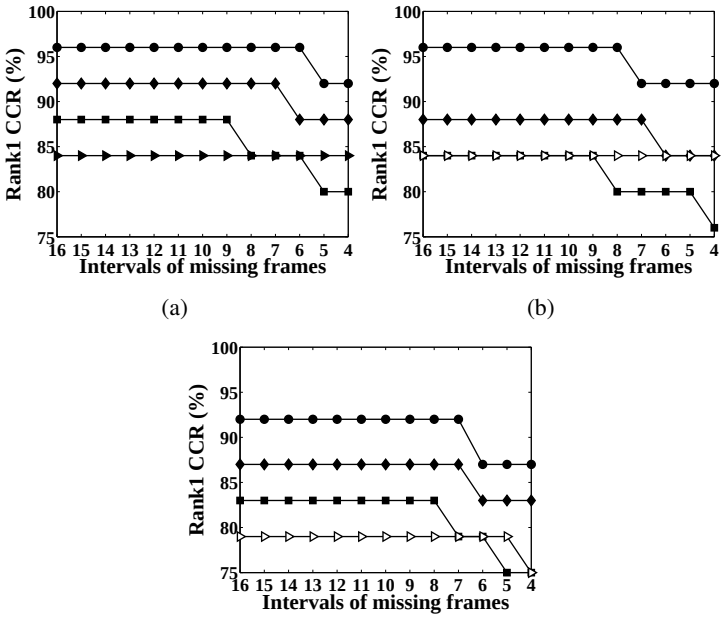


Figure 4.13: Effects of missing frames on performance using PWMS, DTW, BDHM and fused classifier on CMU MoBo dataset: (a) fast walk vs slow walk; (b) slow walk vs fast walk; and (c) fast walk vs walking with ball. Keys: '◆'- PWMS; '■'- DTW; '▹'- BDHM; '●'- Fused Classifier.

are missing. In Fig. 4.13, the rank-1 CCR is plotted along vertical axis, while the horizontal axis shows the intervals of missing frames in terms of number of frames, i.e., 6 at the horizontal axis denotes that every 7th frame is missing from the entire probe sequence. Fig. 4.13(a)-(c) respectively show the effect of missing frames on rank-1 CCR of STS-DM using fused classifier, PWMS, DTW and BDHM for three testing conditions of CMU MoBo dataset, namely fast walk vs slow walk, slow walk vs fast walk and fast walk vs walking with ball. It is evident from the three plots that the rank-1 CCR of STS-DM is not affected for at least every 8th frame is missing from the probe sequence for any of the component classifiers and the fused classifier for three testing conditions. Note that DTW is less robust against missing frames than PWMS and BDHM.

4.4.4 Computational complexity analysis

The computational time of STS-DM is measured using the computer system clock and OpenCV 2.1 in Microsoft Visual Studio 2008 Express Edition environment on an Intel (R) Core (TM) i7 processor working at 2.93 GHz with 4 GB RAM running Windows 7 operating system. For the silhouettes of the full version of USF HumanID gait challenge dataset,

the processing time for comparing all ten Rf-ROIs one at a time with the tr-ROIs for extracting ten phases of a gait period based on the lowest S_{score} using weighted Krawtchouk moments is 5 sec. The processing time to compute d_{PWMS}, d_{BDHM} and d_{DTW} between a probe and a gallery subject is approximately 0.77 fps. Since the Baseline method is characterised by unlimited spatio-temporal correlation of silhouettes, it has very high computational complexity. The hierarchical prediction-based ASM framework with Kalman filter used in ASM to analyse static and dynamic gait characteristics is also computationally very expensive. Unlike most gait recognition methods which process sequences of silhouettes, the real-time method in [79] analyses the set of the largest rectangles fitted onto silhouettes over a gait period spanning up to 25 frames to reduce the computational complexity. STS-DM further reduces this by analysing the shape of contour instead of silhouette at the ten phases of a gait period in computing d_{PWMS} and d_{BDHM}. Since ARPoLK analysis over a gait period uses a1D signal, it does not significantly increase computational complexity. It takes about 5 sec/gait-period to obtain the ten phases by comparing small subregions of an image, i.e., Rf-ROIs with Tr-ROIs, thus reducing time and space complexity.

The use of Cooley-Tukey 2D fast Fourier transform (FFT) algorithm [93; 110] reduces the quadratic time complexity of DFT and its inverse to O($MN \log MN$), where $M \times N$ is the size of the silhouette image. d_{PWMS} is obtained by analysing the low-frequency FDs of the contour points to reduce the computational complexity. DTW has a quadratic time and space complexity, i.e., O(mn), where m and n denote the length of the sequences being compared. However, it is used to compare short sequences, as the number of constituent frames of a gait period usually range between 18-35. Since STS-DM uses a simplified feature space, it does not require any dimensionality reduction technique like PCA and multiple discriminant analysis as in [25; 29]. Since sum rule of score-level fusion and z-score normalisation only require subtraction by mean, division by standard deviation and summation of normalised scores, it has less computational complexity, i.e., O(N), than the rank-based classifier combination rule which requires sorting score of computational complexity O(N log N), where N is the gallery size, followed by post processing to resolve tie in ranking.

4.5 Conclusion

Unlike existing systems which only address one or more challenging factors of gait recognition, STS-DM combines spatio-temporal shape and dynamic motion characteristics of silhouette contours to identify a human subject in the presence of most challenging factors of gait recognition with low computational complexity. It analyses the shape of a subject by FDs at ten phases of a gait period and introduces a component-based FD analysis to achieve robustness against shape distortion due to all common types of small carrying conditions

with folded hands, at the subject's back and in upright position. ARPoLK analysis with consideration of the integral relationship between the motion of limbs and arm-swing enables STS-DM to achieve robustness against gait variation over different days, e.g., limited clothing variation, hair style, shadows under feet and missing body parts. The similarity between the ARPoLK of two subjects is measured using DTW to achieve invariance to walking speed. STS-DM uses BDHM to analyse the full-body shape and motion characteristics by fitting ellipses to five different parts of the human body which is invariant to boundary shape distortions due to segmentation imperfections and missing body parts. The match scores obtained by analysing the local and global gait characteristics using the three feature extractors are combined using weight-based sum rule of score-level fusion for subject identification.

STS-DM analyses the shape of contours, hence it is insensitive to colour and texture of subject's clothing. The feature space of STS-DM does not require any dimensionality reduction. The excellent identification rates in the presence of various challenging factors demonstrate the efficacy of STS-DM. Being a contour based method, STS-DM has a low computational complexity, but it is sensitive to segmentation imperfections, and its performance largely depends on preprocessing. Also, STS-DM is designed for lateral views of gait sequences, thus future developments are required to enable STS-DM to address unconstrained human movements especially in cluttered scenes.

Chapter 5

Region-Based Gait Analysis in Image and Feature Spaces

5.1 Introduction

The gait recognition methods introduced in Chapter 3 and Chapter 4, i.e., STM-SPP and STS-DM analyse the extreme outer boundary, i.e., the contour of the silhouettes. Since these methods use only the pixels at the contours for analysing the static shape and dynamic motion characteristics of gait, they have low computational complexity. However, their discriminatory power decreases in the presence of noise and disjoint holes at the contour resulting from poor segmentation and partial occlusion. Hence the performance of these methods is highly dependent on the background segmentation and morphological image transformation techniques, as these factors determine the quality of the extracted contours. Therefore, in this chapter, a method for region-based gait recognition in the image and feature spaces (referred to as ReG-IF) is introduced, which uses the pixels comprising the silhouette to analyse its region-based shape characteristics as opposed to boundary-based shape descriptors, e.g., FDs [91; 120] and wavelet descriptors [112].

The region-based shape descriptor in the image space (SDIS) involves a lowpass Gaussian filter (Lp-Gf) and a highpass Gaussian filter (Hp-Gf) to filter the silhouettes at ten phases of a gait period in the frequency domain using different cut-off frequencies. Filtering with Lp-Gf causes smoothing or blurring of a silhouette and thus reduces noise. As the cut-off frequency of the Lp-Gf decreases, there is a gradual loss of boundary and exterior region. Thus, the application of Lp-Gf with decreasing cut-off frequencies gradually highlights the characteristics of inner part of a silhouette towards its central region more than its boundary, and enables ReG-IF to achieve robustness against tight vs loose clothing, and clothing type variation. It also reduces the effect of shape distortions due to small carried items at the

boundary. The use of Hp-Gf at the same cut-off frequencies focuses on the boundary and the exterior parts of a silhouette more than the central part, thus highlighting the boundary characteristics of the silhouettes to enable ReG-IF to achieve robustness to small holes in the silhouettes caused by segmentation imperfections. The cut-off frequencies of the Gaussian filters for optimal performance are determined experimentally based on an analysis of the focus values of the silhouettes.

The statistical moments, which provide a compact, geometric invariant and noise-resilient global shape information, are the region-based shape descriptors in the feature space [113; 98]. Orthogonal moments [84], e.g., Zernike velocity moments and Krawtchouk moments, enable extraction of independent image features with no information redundancy, and simplify reconstruction of the original image. However, the highly correlated features of non-orthogonal moments, e.g., Hu moments [88], complicate the image reconstruction. ReG-IF uses weighted Krawtchouk moments to obtain shape descriptors in the feature space (SDFS). A component-based shape analysis using weighted Krawtchouk moments is introduced to achieve robustness to shape distortions due to carried items.

If the gallery subjects for training are recorded under similar physical conditions, the learned features in the presence of varying covariate factors are likely to cause overfitting that decreases the subject identification rate. Thus, the methods in [25; 28] manually compute synthetic gait templates following a distortion model based on anthropometry to take into account of lower body part distortions due to variation in walking surface, footwear and clothes. The use of the templates enables these methods to be insensitive to the lower body part distortions, but not upper body-part distortions due to clothing variation and carrying conditions. The DNGR method in [26] uses eigenstance reconstruction model to significantly reduce the effect of shadows and segmentation errors by manually improving the silhouettes. Since there are numerous types of covariate factors that can affect a probe sequence, it is difficult to create the appropriate synthetic gallery gait template for robust gait recognition. Since variation in clothing of the same subject in gallery and probe gait sequences reduce the identification rate, the method in [52] applies part-based strategy to adaptively assign more weight to the unaffected body parts and less weight to the affected body parts to achieve insensitiveness to clothing variation. However, with numerous and unpredictable variation in clothing types, it is unrealistic to train the model with all known clothing types as in [52].

ReG-IF avoids overfitting without manually computing synthetic gait templates as in [25; 28] and achieves invariance to clothing variation by using multiple subspace learning (MSL) based classification for SDIS by introducing a new image representation called concatenated silhouette image (CSI). MSL is an ensemble classifier which constructs randomly selected multiple subspaces from the feature space with each subspace associated with a

classifier, and the final classification is achieved by majority voting. While most classification techniques suffer from high dimensionality, MSL exploits the high dimensionality of SDIS to address the trade-off between accuracy and over-adaptation [121]. Euclidean distance is used to classify SDFS. To achieve robustness against carrying conditions, ReG-IF divides a subject's body into four components and analyses each component using weighted Krawtchouk moments to obtain a dissimilarity score. The dissimilarity scores corresponding to each component thus obtained, are combined using weighted sum rule of score-level fusion, where the weights are computed based on each component's discriminatory power influenced by the shape distortions due to carried items at the back, with folded arms and in upright position.

The changes in a subject's shape over time in a gait period have better discriminatory power in identifying the subject than the discrete snapshots of images, but it increases the computational complexity. Furthermore, the variation in the subject's walking speed, e.g., induced by changes in the subject's mood brings challenges to gait recognition, as it causes variation in the length of a gait period. Thus, we analyse the shape of the silhouette at ten phases of a gait period to exploit the benefit of shape sequence processing with reduced processing time and achieve invariance to walking speed.

ReG-IF has the following novelties: (1) it introduces a new image representation called CSI which is formed by analysing the shape of a silhouette using Lp-Gf and Hp-Gf at different cut-off frequencies to obtain SDIS; (2) it uses weighted Krawtchouk moments to obtain SDFS; (3) it introduces a component-based shape analysis based on weighted Krawtchouk moments to achieve robustness to all common types of carrying conditions by considering different discriminability power of each part due to presence of a carried item at the back, with folded arms and in upright position; (4) the classification based on MSL exploits the high dimensionality of SDIS for improved identification by avoiding overfitting; (5) Lp-Gf blurs the silhouette highlighting its inner shape characteristics, thus enable ReG-IF to achieve robustness against the challenging factors that mainly cause distortions at the boundary, e.g., clothing variation, presence of a carried item, change in hair style, disjoint holes at the boundary and shadows under feet; (6) Hp-Gf highlights the boundary shape characteristics which plays an important role in inter-subject discrimination in absence of change in covariate factors; (7) the cut-off frequencies for Lp-Gf and Hp-Gf for optimal performance are selected based on focus value analysis of the silhouettes; (8) ReG-IF is robust against a wide variety of challenging factors, namely, variation in clothing and walking speed, presence of a carried item, segmentation noise, missing and distorted frames, change in ground surface and hair style, shadows under feet and occlusions; (9) ReG-IF achieves the highest rank-1 identification rate on USF HumanID gait dataset compared to the state-of-the-art gait recognition methods.

The rest of the chapter is organized as follows. Section 5.2 discusses related work and Section 5.3 presents ReG-IF. Experimental results are analysed in Section 5.4 and Section 5.5 concludes the chapter.

5.2 Related Work

Few gait recognition methods have attempted to address invariance to carrying conditions. The method in [45] uses an iterative local curve embedding algorithm to extract double helical signatures for limited invariance to carrying condition. The use of synthetic gait templates enables the methods GEI [25] and CGI [28] to achieve invariance to carrying a briefcase. The method STM-SPP [7] (and presented in Chapter 3) introduces a part-based shape analysis using EFDs to achieve robustness to shape distortions due to small carried items with folded arms and a briefcase in upright position. The method STS-DM [8] (and presented in Chapter 4) provides an analysis for increased robustness against carrying conditions compared to STM-SPP by introducing a component-based FD analysis based on anatomical studies of human body to achieve invariance to all common types of small carrying conditions with folded arms, at the back and in upright position. Although the methods STM-SPP and STS-DM are able to achieve substantial invariance to all common types of small carrying conditions compared to the state-of-the-art methods, they do not take into account of a subject's multiple carrying conditions, e.g., a bag at the back and also a briefcase in upright position.

The invariance to the subject's walking speed has been addressed in different ways. The method DNGR in [26] measures the dissimilarity between two dynamics-normalised gait signatures based on population HMM and uses LDA to maximise inter-class discriminability. The method in [40] computes SVB frieze pattern of the selected frames of a gait period. The method using silhouette transformation based on walking speed, i.e., ST-WS [43] separates static and dynamic features by fitting a human model and uses a factorization-based transformation model to transform the dynamic features from a reference speed to a target speed. Speed invariance is achieved in the method SI-PSA [42] by using PSA to compute high-order derivative shape configuration. The use of DTW in SSM [23], which uses both parametric models and nonparametric methods for comparing shape sequences, enables it to achieve limited invariance to walking speed.

ReG-IF is compared with the other model-free methods in [22; 47; 52; 65; 49; 111; 37; 50; 119] to demonstrate its efficacy. The use of Gabor wavelets for gait feature extraction in the methods in GTDA-GF [50] and GPDF [47] has been successful to achieve improved identification rates compared to other shape based gait recognition methods due to its following impressive properties: (a) the kernels of Gabor wavelets are similar to the

receptive field profiles of the simple cells of the mammalian visual cortex, which add to its suitability for biology related pattern recognition applications; and (b) Gabor wavelets are optimally localized in time and frequency domains, and exhibits useful characteristics of frequency and orientation selectivity; and (c) it facilitates desired local feature extraction due to its spatial localization property. Using Gabor functions incurs high computational complexity, thus GTDA-GF uses sum of Gabor filter responses over directions, sum of Gabor filter responses over scales and sum of Gabor filter responses over scales and directions for analysing GEIs. The use of LGSR classifier enables GPDF in [47] to achieve the highest identification rate on USF HumanID gait dataset.

5.3 Proposed method: ReG-IF

ReG-IF comprises three modules: 1) silhouette processing; 2) region-based gait feature extraction in image and feature spaces; and 3) classification of subjects.

5.3.1 Module 1: Silhouette processing

Silhouettes provided by publicly available datasets are used as the input gait sequences of ReG-IF. To remove camera depth variation, the silhouette is cropped to the perimeter of the bounding rectangle enclosing the silhouette, and the bounding rectangle is resized to a fixed height keeping the same aspect ratio using bilinear interpolation. The resized silhouette is then copied to a destination image of fixed size by coinciding its centre-of-mass with the centre of the destination image for centralising it, so as to make it translation invariant.

5.3.2 Module 2, phase 1: SDIS

Similar to [22], the gait period of a subject in ReG-IF corresponds to the number of frames between two frames of a gait sequence with the maximum number of contour points enclosed by the bottom of the bounding rectangle and the anatomical position of the subject's hand measured from the bottom of the bounding rectangle, i.e., $0.377H$ where H is the height of the bounding rectangle, as this foreground segment is not distorted by self-occlusion due to arm-swing. After estimating the gait period of the subject from the video sequence of lateral view of the subject, the ten phases of the gait period [7] (i.e., initial contact, ending swing, 2 mid-swings, pre-swing, propulsion, 3 midstances and double support) which capture most of the significant gait characteristics, are extracted using region-of-interest based contour matching following STM-SPP [7]. A detailed description and pictorial illustration of the ten phases of a gait period are provided in STM-SPP [7] (and presented in Chapter 3).

We analyse the silhouettes at the ten phases using Lp-Gf and Hp-Gf in the frequency domain at different cut-off frequencies to obtain the SDIS for a subject as follows. The DFT of an $M \times N$ silhouette image $I(x, y)$ is

$$DFT(u, v) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} I(x, y) e^{-j2\pi(ux/M)+(vy/N)}, \tag{5.1}$$

where $u = 0, 1, 2, \ldots, M-1$ and $v = 0, 1, 2, \ldots, N-1$ are frequency variables. The Fourier transformed silhouette, i.e., $DFT(u, v)$ is translation invariant, but since it retains rotation, it is subjected to shift operation to ensure that the zero-frequency components are at the centre. To represent the inner part of a silhouette gradually towards the centre more than its boundary, Lp-Gf is applied to the Fourier transformed image using selected cut-off frequencies, i.e.,

$$DFT_L(u, v) = DFT(u, v) e^{(-(u^2+v^2)/2D^2)}, \tag{5.2}$$

where $e^{(-(u^2+v^2)/2D^2)}$ is the transfer function of Lp-Gf in the frequency domain [98], and $DFT_L(u, v)$ denotes the image filtered using Lp-Gf at the cut-off frequency D. The filtered silhouette at cut-off frequency D in the image space is obtained by applying inverse DFT, i.e.,

$$I(x, y) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} DFT_L(u, v) e^{j2\pi(ux/M)+(vy/N)}. \tag{5.3}$$

Fig. 5.1(a)-(k) show the silhouettes filtered by Lp-Gf with decreasing cut-off frequency. Since Lp-Gf attenuates high frequency components, it blurs the silhouette and thus smooths detailed clothing curvatures at the silhouette boundary. As the cut-off frequency decreases, it results in a greater loss of boundary and exterior regions due to increase in blurring to gradually highlight the inner shape characteristics. Note that Gaussian functions in the spatial and frequency domain behave reciprocally with each other, hence an decrease in standard deviation of Lp-Gf in the frequency domain results in more blurring and vice versa, while the reverse is true in the spatial domain [98].

To represent the boundary and exterior regions of a silhouette more than its central part, Hp-Gf is applied to the silhouette at the same cut-off frequencies [98], i.e.,

$$DFT_H(u, v) = DFT(u, v)(1 - e^{(-(u^2+v^2)/2D^2)}), \tag{5.4}$$

where $1 - e^{(-(u^2+v^2)/2D^2)}$ is the transfer function of Hp-Gf with cut-off frequency D [98]. The filtered silhouette is similarly obtained using Eq. (5.3), and Fig. 5.1(l)-(v) shows the silhouettes filtered by Hp-Gf with decreasing cut-off frequencies. Hp-Gf emphasises the high frequency components but retains limited low frequency components, thus making

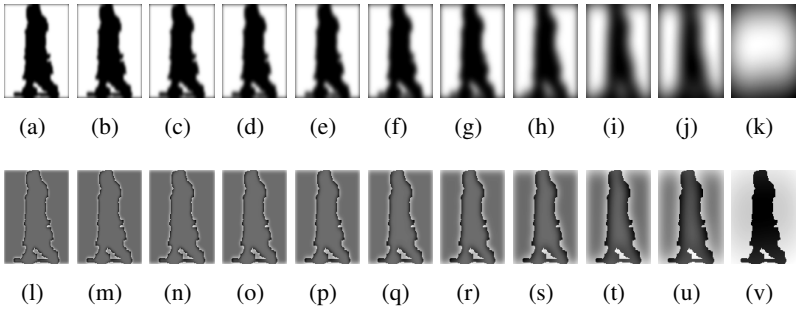


Figure 5.1: Application of Lp-Gf (Row 1) and Hp-Gf (Row 2) to a subject's silhouette from USF 2.1 dataset with decreasing cut-off frequency: (a) & (l) $D_1 = 20$; (b) & (m) $D_2 = 18$; (c) & (n) $D_3 = 16$; (d) & (o) $D_4 = 14$; (e) & (p) $D_5 = 12$; (f) & (q) $D_6 = 10$; (g) & (r) $D_7 = 8$; (h) & (s) $D_8 = 6$; (i) & (t) $D_9 = 4$; (j) & (u) $D_{10} = 3$; and (k) & (v) $D_{11} = 1$.

the boundary characteristics of a silhouette more prominent, and its application represents the exterior regions of a silhouette as the cut-off frequency decreases. Note that region-based shape descriptors using Lp-Gf and Hp-Gf are in the image space. Also, since ReG-IF is defined on lateral view of silhouettes, and all silhouettes are scale-normalised and translation invariant, it is unnecessary to reapply DFT to the polar transformed silhouettes to achieve rotation invariance as in [113].

The focus value which is used to measure the degree of sharpness of an image is the maximum for the most focused, i.e., the original silhouette. It is inversely proportional to the image blurriness caused by the Gaussian filtering at different cut-off frequencies. Common methods for computing focus values of an image include spatial domain based methods, e.g., Tenengrad [114] and sum modified Laplacian [115], and wavelet based methods [116]. The first level 2D Daubechies-6 wavelet decomposition of a silhouette image $f(x, y)$ of size $M \times N$ results in four subband images, W_{LL}, W_{HL}, W_{LH} and W_{HH}, where L and H respectively denote lowpass filtered and highpass filtered, and their order denotes the order of the filtering applied, e.g., W_{HL} is a subband image obtained by highpass filtering followed by lowpass filtering. The focus value of a silhouette is measured using [116]

$$FV = \frac{1}{MN} \sum_{y=0}^{N} \sum_{x=0}^{M} (|W_{HL}(x, y)| + |W_{LH}(x, y)| + |W_{HH}(x, y)|). \tag{5.5}$$

Since this wavelet based method provides the sharpest focus measure profile and higher depth resolution than the spatial domain based methods due to the localised support property of wavelet basis [116], it is used to compute the focus value of low resolution silhouettes.

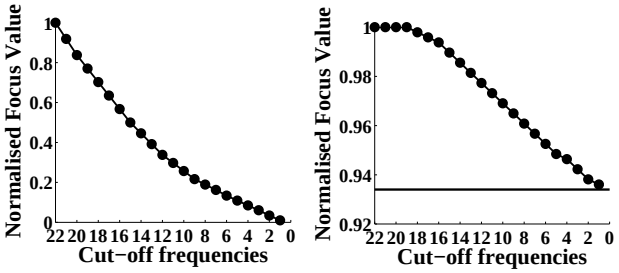
The focus value of the original silhouette always reduces to below 50% if it is filtered by Lp-Gf at cut-off frequency $D = 20$, and decreases linearly as the blurriness in-

creases with decreasing cut-off frequencies. If the cut-off frequency is decreased further to below $D = 8$, the focus value decreases abruptly. The focus value becomes infinitesimally small if $D < 4$, resulting in excessively blurred silhouette without any discriminating information (e.g., Fig. 5.1(j)-(k)).

The boundary of a silhouette is obtained by the application of Hp-Gf using D approximately in the range [18,22] for most of the silhouettes of the USF dataset [22]. Since the silhouette boundary corresponds to the sharpest image, e.g., Fig. 5.1(l)-(m), the focus value of a silhouette filtered by Hp-Gf using D in this range remains the maximum which is considerably higher than the focus value of the original silhouette (i.e., Fig. 5.2(b)). With further decrease in cut-off frequency, the focus value decreases linearly with the decrease in the sharpness of the image as the silhouette is reconstructed by regaining its central region. The focus value is nearly identical to that of the focus value of the original silhouette in the range $1 \leq D < 4$ due to almost reconstruction of the original silhouette. These silhouettes hardly contribute to gait recognition as their shape is considerably affected by the covariates. Since the boundary as well as central shape characteristics of a silhouette are considered separately by using Lp-Gf and Hp-Gf, it is not necessary to process all silhouettes which will increase the computational complexity. Thus, D in [4,20] is considered to be the ideal range of cut-off frequencies for the Gaussian filters to obtain SDIS.

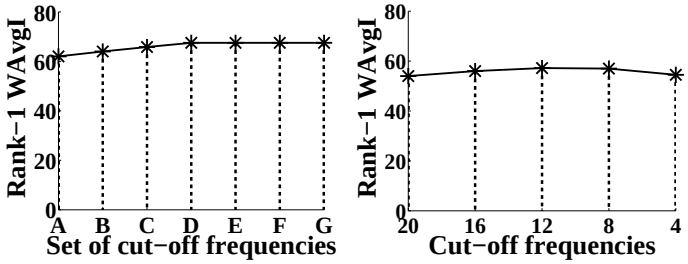
Fig. 5.2(a) and (b) respectively show the normalised focus value w.r.t. decreasing cut-off frequencies in the range [22, 0] of a silhouette filtered by Lp-Gf and Hp-Gf, where normalised focus values are obtained by dividing each of the focus values with the maximum focus value in the range [22, 0]. Since the focus value of a filtered silhouette maintains a linear relationship with the cut-off frequencies of Gaussian filters as shown in Fig. 5.2, a set of the smallest number of cut-off frequencies to obtain the best performance for SDIS is determined by analysing the WAvgI at rank-1 (see Section 5.4.2) of SDIS on the USF2.1 dataset vs sets of cut-off frequencies, where each set consists of increasing number of equidistant cut-off frequencies in the range [4,20]. Fig. 5.3(a) shows the results for sets of cut-off frequencies A = {4, 20}, B = {4, 12, 20}, C = {4, 9, 15, 20}, D = {4, 8, 12, 16, 20}, E = {4, 7, 10, 14, 17, 20}, F = {4, 7, 9, 12, 15, 17, 20} and G = {4, 6, 9, 11, 13, 15, 18, 20} on USF2.1 dataset. It shows that set D consists of the minimum number of cut-off frequencies for SDIS to achieve the highest WAvgI at rank-1 which is significantly higher than the WAvgIs obtained by using only the individual cut-off frequencies comprising the set D (Fig. 5.3(b)). Any increase in the number of cut-off frequencies used in D, e.g., to form sets E, F and G, only increases computational complexity without having any effect on the identification rate. Thus, the $N \times M$ images obtained by applying Lp-Gf and Hp-Gf each at cut-off frequencies comprising the set D (i.e., a total of 10 cut-off frequencies) on the silhouettes of ten phases of a gait period for a subject are concatenated by arranging the

images at ten phases of a subject row-wise, and the corresponding filtered images by Lp-Gf and Hp-Gf column-wise to form a new 2D $10N \times 10M$ concatenated silhouette image (CSI). The resulting feature image corresponding to SDIS of a gallery subject is used to form part of the gallery database.



(a) (b)

Figure 5.2: Normalised focus value w.r.t. decreasing cut-off frequencies of a silhouette from USF 2.1 dataset filtered using (a) Lp-Gf; and (b) Hp-Gf. '-' denotes focus value of the original silhouette.



(a) (b)

Figure 5.3: WAvgI at rank-1 for SDIS on USF 2.1 dataset vs sets of cut-off frequencies, i.e., A, B, C, D, E, F and G; (b) WAvgI at rank-1 for SDIS on USF 2.1 dataset vs individual cut-off frequencies comprising set D.

5.3.3 Module 2, phase 2: SDFS

Similar to variation in clothing, the presence of a carried item significantly distort the shape of different parts of a silhouette, and the discriminability between the subjects decreases with respect to the affected parts. While it is unrealistic to predict which part(s) of a silhouette will be affected due to clothing variation by assuming known clothing types in the training phase as in [52] due to the availability of numerous different types and combinations of clothing in real scenarios, the part(s) of a silhouette which is likely to be affected due to all small types of carrying conditions either at the subject's back, with folded arms

or in an upright position can be convincingly identified based on anthropometry. The underlying principles are: (a) If a subject carries a backpack or a small item such as a ball or a package with folded arms, it is unlikely that the shape of silhouettes above the shoulder and below the anatomical position of hand of an upright subject, i.e., respectively $0.182H$ from the top of the bounding rectangle and $0.377H$ from bottom of the bounding rectangle [97] will be affected [7]. Experimental analyses for all subjects holding a ball in CMU MoBo dataset [19] and carrying a backpack in CASIA gait dataset B verify the appropriateness of the assumption. (b) The shape of an upright silhouette above the wrist remains unchanged when the subject's hand carries a briefcase or a small bag. The position of the wrist is about $0.515H$ [97] measured from the top of the bounding rectangle. Thus, the segment of silhouettes enclosed in the region $0.515H$ of the bounding rectangle measured from the top is not affected by such carrying condition. The leg region of a silhouette enclosed between aH (where $a = 0.750$ measured from the top of bounding rectangle and the bottom of bounding rectangle assuming a standard size of a briefcase [27] removes any shape distortion due to carrying a briefcase. (c) It is not unusual that a subject carries a bagpack at the back, a small item in one hand and a briefcase in the other hand. In this case, the shape of silhouettes above the shoulder and below $0.750H$ remain undistorted.

Based on the above considerations, we divide a silhouette into four components, i.e., C1, C2, C3, and C4 as shown in Fig. 5.4 for component-based shape analysis using weighted Krawtchouk moments so as to reduce the adverse impact of carrying conditions on identification rate. The four components are enclosed by: (a) the top of bounding rectangle and up to the anatomical position of the shoulder, i.e., $0.182H$ (component C1); (b) the top of the bounding rectangle and up to the anatomical position of the wrist, i.e., $0.515H$ (component C3); (c) the anatomical position of hand of an upright subject and the bottom of the bounding rectangle (component C2); and (d) the leg region to exclude briefcase (component C4). It is to be noted that the first and fourth components remain unaffected by the carrying conditions considered in this method, whereas the second and third components are most likely to get distorted by the carrying conditions with folded arms and in upright position. It might be argued that the second and third components will contribute more to the discrimination between the gallery and probe subjects in case of absence of a carried item, as they are larger body parts than the first and fourth components. However, these larger body components are affected by the arm-swing (an integral gait characteristic) in absence of carried items which changes due to variation in walking speed induced by change in mood, haste, etc. Since the two smaller components are not affected by the carrying conditions and they are also included in the larger body components which are most likely to be affected either by carrying conditions or by arm-swing, the smaller components, i.e., C1 and C4, are considered to contribute more to the discrimination of the subjects in ReG-IF

than the larger components, i.e., C2 and C3.

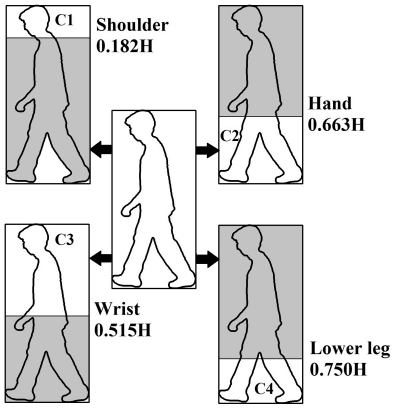


Figure 5.4: The silhouette components.

Following the unequal contribution of each body component to the subject identification due to carrying conditions, ReG-IF determines the pixel range of the silhouette image (i.e., the values of M and N) of Eq.(4.1), such that they represent the four components of the silhouette, and computes the weighted Krawtchouk moments corresponding to each of these segments, e.g., for USF dataset, the range of values of M and N used to extract local information of the silhouette from top of the bounding rectangle to the wrist are respectively [0,W], where W is the width of the bounding rectangle (i.e., [0,88]), and [0,0.515H] (i.e., [0,65]). The range of values of N and M for region between aH and bH are respectively [0,W], and [96,120]. The weighted Krawtchouk moments of the ten phases of a gait period of a gallery subject form part of the gallery database. ReG-IF computes Euclidean distances between the weighted Krawtchouk moments corresponding to each of these body components of the gallery and probe subjects, and subsequently combine the component-wise distances using weighted sum rule in order to determine the final similarity score between the subjects, where the weights are determined based on the discriminatory power of each component.

The Krawtchouk moments [65; 108] are adopted because their image reconstruction capability is better than Zernike and Hu moments. Since they can be used to extract features from any region of an image, they can also address partially distorted frames of the ten phases of a gait period. The orthogonal property of weighted Krawtchouk moments ensures minimal information redundancy. The scale and rotation dependency of Krawtchouk transform do not affect the extracted features as ReG-IF considers only lateral views of silhouettes to achieve rotation invariance, and the silhouettes are pre-scaled and centre-aligned to achieve scale invariance. The weighted Krawtchouk moments of order $(n+m)$ of a $N \times M$ silhouette with intensity function $f(x,y)$ are obtained using Eq. (4.1), Eq. (4.2), Eq. (4.3)

and Eq. (4.4).

The performance of any shape-based gait recognition method degrades due to variation in the subject's clothing and carrying conditions, as these variation distort the silhouette shape. The use of synthetic gait templates enables GEI and CGI to be invariant to distortions in the lower body-part but not in the upper body-part, e.g., due to carrying conditions with folded arms and at the back. The method in [45] analyses symmetry changes in double helical signatures at the limb region to take into account of shape distortion due to carrying condition, e.g., a briefcase carried by an upright subject. Although STM-SPP analyses different parts of a silhouette using EFDs to take into account of shape distortions due to carrying a briefcase and small items with folded arms, it does not consider subjects carrying a backpack. ReG-IF uses a new approach for a general analysis of invariance to carrying condition by taking into account common small items carried by a subject on his/her back, with folded ams and in upright position via component-based reconstruction using weighted Krawtchouk moments.

5.3.4 Module 3: classification of subjects

SDIS has a higher dimensionality than SDFS. MSL is thus used in SDIS to take the advantage of the high dimensionality for improved identification rate. Since MSL is not useful for SDFS, SDFS uses NNC due to its simplicity and lower computational complexity than MSL.

Case 1 - MSL classifier for SDIS

ReG-IF uses 2D PCA as a preprocessing step for MSL to reduce the dimensionality of the feature space without losing discriminant information via data decorrelation. Since 2D PCA is directly based on 2D image matrices rather than 1D vectors, it is superior to PCA in terms of accurate estimation of covariance matrices and computational complexity. Given n number of CSIs, i.e., { $\mathbf{Fm}_1$, $\mathbf{Fm}_2$,..., $\mathbf{Fm}_n$} in the gallery, the scatter matrix $\mathbf{S}$ is obtained using

$$\mathbf{S} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{Fm}_i - \mathbf{M})^T(\mathbf{Fm}_i - \mathbf{M}), \tag{5.6}$$

where $\mathbf{M} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{Fm}_i$. Since there are at most $n-1$ eigenvectors of $\mathbf{S}$ with nonzero eigenvalues, N eigenvectors (where $N < n-1$) are randomly chosen from the set of $n-1$ eigenvectors, i.e., $\{\mathbf{e_1}, \mathbf{e_2}, ..., \mathbf{e_{n-1}}\}$, with the largest eigenvalues to construct L subspaces $\{R_k\}_{k=1}^{L}$. The n-th eigenvector with zero eigenvalue is discarded in order to reduce the dimensionality of the feature space without losing discriminant information.

For each subspace R_k, k = 1,..., L, the projected image

$$\mathbf{Y}_i = \mathbf{M}_{\mathrm{2DPCA}}\mathbf{F}\mathbf{m}_i = [\mathbf{e_1}, \mathbf{e_2}, ..., \mathbf{e_N}]^T\mathbf{F}\mathbf{m}_i, \qquad i = 1, ..., n. \tag{5.7}$$

where $\{\mathbf{Y}_1, ..., \mathbf{Y}_n\}$ belong to C gallery classes. To achieve optimal class separability and overcome singularity problem associated with 1D LDA, 2D LDA is used to seek a transformation matrix W that maximizes the ratio of the between-class scatter matrix $\mathbf{S_B}$ to the within-class scatter matrix $\mathbf{S_W}$, i.e.,

$$J(W) = \frac{|W^T S_B W|}{|W^T S_W W|}. \tag{5.8}$$

where

$$\mathbf{S_W} = \sum_{i=1}^{C} \sum_{\mathbf{Y}\in D_i} (\mathbf{Y} - \mathbf{m}_i)(\mathbf{Y} - \mathbf{m}_i)^T, \tag{5.9}$$

$$\mathbf{S_B} = \sum_{i=1}^{C} n_i(\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T, \tag{5.10}$$

$\mathbf{m_i} = \frac{1}{n_i}\sum_{\mathbf{Y}\in D_i}\mathbf{Y}$, D_i is the training template set that belongs to the i-th class and n_i is the number of templates in D_i. J(W) is maximised when the columns of W are the generalised eigenvectors that correspond to the largest eigenvalues in

$$\mathbf{S_B}\mathbf{w}_i = \lambda_i\mathbf{S_W}\mathbf{w}_i, \qquad i = 1, ..., C. \tag{5.11}$$

Since there are at most C-1 nonzero eigenvalues, the corresponding v_1,..., v_{C-1} eigenvectors are used to form the training gait feature matrix $\mathbf{Z}_i$ for each of L subspaces using

$$\mathbf{Z}_i = \mathbf{M}_{\mathrm{2DLDA}}\mathbf{Y}_i = [\mathbf{v_1}, ..., \mathbf{v_{C-1}}]^T\mathbf{Y}_i, \qquad i = 1, ..., n. \tag{5.12}$$

Let $\{\mathbf{r}\}$ be the set of gallery gait feature matrices belonging to C classes. Therefore, each of C classes of the gallery with n_j feature matrices is represented by its centroid, i.e., $G_j = \frac{1}{n_j}\sum_{r\in R_j}\mathbf{r}$, where j = 1,..., C and R_j is the set of feature matrices belonging to the j-th class.

The distance between a probe sequence P with m_p gait periods and a gallery class centroid G_j is [25]

$$D(P, G_j) = \frac{1}{m_p}\sum_{i=1}^{m_p} \|\mathbf{s} - G_j\|, \qquad j = 1, ..., C. \tag{5.13}$$

where **s** is the set of a probe gait feature matrices. We assign

$$P \in w_k, \quad \text{if } D(P, G_k) = \min_{j=1}^{C} D(P, G_j), \tag{5.14}$$

where w_k is the k-th gallery class. A probe sequence P is assigned a class label by each component classifier $A_{\text{LH-GF}} \in \{A_{\text{LH-GF}}\}_{k=1}^{L}$, where $\{A_{\text{LH-GF}}\}_{k=1}^{L}$ is the set of classifiers constructed from L subspaces. The final class of P is determined based on majority voting and the classification at rank-r implies that the number of votes received by a probe belong to the top r ranks for correct classification.

MSL is run ten times for each experiment with L $\in \{100, 300, 500\}$ and N fixed for SDIS. It is observed that a few pairs of common values for classification rates at rank-1 and rank-5 are repeatedly produced in different runs. The maximum and minimum values thus obtained are discarded as outliers to remove occasional bad sampling and the average of the remaining eight values are used as the classification rates at both ranks.

Case 2 - 1-NNC for SDFS

The Euclidean distance between the (n+m)th order weighted Krawtchouk moments of a pair of silhouettes with respect to a particular body segment, i.e., σ from a gallery and probe sequence respectively denoted by $(Q_g\sigma)$ and $(Q_p\sigma)$ is

$$\text{AKRAT}_{\text{ED}\sigma} = \left[(Q_{g\sigma} - Q_{p\sigma})^2\right]^{\frac{1}{2}}, \sigma = \text{C1, C2, C3, C4.} \tag{5.15}$$

The range of $\text{AKRAT}_{\text{ED}\sigma}$ is [0,1], the smaller the value the more similar are the two silhouette segments.

The individual Euclidean distances between the corresponding body segments are used to form a distance vector, i.e.,

$$\mathbf{D} = \begin{bmatrix} \text{AKRAT}_{\text{EDC1}} \\ \text{AKRAT}_{\text{EDC2}} \\ \text{AKRAT}_{\text{EDC3}} \\ \text{AKRAT}_{\text{EDC4}} \end{bmatrix} \tag{5.16}$$

The individual Euclidean distances between the corresponding body segments are fused using weighted sum rule to quantify a single distance between the pair of gallery and probe silhouettes using

$$\overline{\mathbf{D}} = \mathbf{W}^{\mathbf{T}}\mathbf{D}, \tag{5.17}$$

where $\mathbf{W} = [w_{C1}, w_{C2}, w_{C3}, w_{C4}]^{\text{T}}$ is a weight vector, and w_{C1}, w_{C2}, w_{C3} and w_{C4} re-

spectively denote the weights associated with the silhouette segments C1, C2, C3 and C4. w_{C1}, w_{C2}, w_{C3} and w_{C4} are assigned fixed values in accordance with the contributions of C1, C2, C3 and C4 to the classification based on SDFS (explained in Section 5.3.3), i.e., w_{C1}=w_{C4}=0.4 and w_{C2}= w_{C3}=0.1.

The average Euclidean distance between the (n+m)th order weighted Krawtchouk moments of the silhouettes of a gallery and the probe gait period for all ten phases is

$$\text{AKRAT}_{\text{ED}} = \frac{1}{10}\sum_{i=1}^{10}\overline{\mathbf{D}}. \tag{5.18}$$

Each of the gallery and probe gait sequence, denoted by N_g and N_p, is respectively partitioned into consecutive subsequences corresponding to a gallery gait period (G_g) and a probe gait period (G_p). The distance metric between the kth probe gait period and the gallery sequence, for a match score S, i.e., AKRAT_{ED}, is $Dist_S(k) = min_i(S)$, where i = 1,2,...,n_g and $n_g = N_g/G_g$, i.e., the number of gallery gait periods in a gallery sequence. The dissimilarity score used for classification is the median of the distances, i.e.,

$$Dism_{SDFS} = median(Dist_S(1), Dist_S(2), ..., Dist_S(m_p)), \tag{5.19}$$

where $m_p = N_p/G_p$, i.e., the number of probe gait periods in a probe sequence. The classification at rank-r implies that the probe results in the top r D_S for correct classification.

5.3.5 Combining classifiers

In order to demonstrate the effectiveness of combining two classifiers using SDIS and SDFS to obtain better identification rates than using each of the classifiers separately, a weighted sum rule of score level fusion [103] is used due to its simplicity, as it is more informative than the rank based fusion, and does not require any post processing technique to resolve a tie in ranking. In addition to a tie resolving technique, rank-based classifier combination requires a sorting algorithm with computational complexity O($N \log N$), where N is the number of gallery classes, and hence it is computationally more expensive than the score based classifier fusion. However, score based classifier combination rule requires transformation of the non-homogeneous scores obtained by different classifiers to the same scale using score normalisation technique prior to the combination.

The linear score normalisation techniques, e.g., min-max normalisation and z-score normalisation have similar computational complexities, but z-score normalisation is preferred in ReG-IF as it is less sensitive to outliers than min-max normalisation. Although non-linear score normalisation techniques based on double sigmoid function and hyperbolic tangent are more robust to outliers, they have higher computational complexity as they use

many parameters, and the performance of these techniques are highly dependent on the chosen parameter values. Hence, to make a trade-off between the performance and computation complexity, ReG-IF obtains z-scores of the two sets of dissimilarity scores produced by two classifiers for SDIS and SDFS using

$$Z_{classifier} = \frac{\mathrm{d}_{classifier} - \mu_{classifier}}{D_{classifier}}, \tag{5.20}$$

where μ is mean of the set of dissimilarity scores, d is the individual dissimilarity score and $classifier$ is either SDIS or SDFS. A probe sequence P is assigned L dissimilarity scores, i.e., $D(P, G_j)$ for each $j = 1, ..., C$ (see Eq. (5.13)) by each component classifier $A_{\text{LH-GF}} \in \{A_{\text{LH-GF}}\}_{k=1}^{L}$, where $\{A_{\text{LH-GF}}\}_{k=1}^{L}$ is the set of classifiers constructed from L subspaces. The dissimilarity score for classifier based on SIDS for score-level fusion is

$$\text{Dism}_{\text{SDIS}} = \frac{1}{L}\sum_{1}^{L} D(P, G_j). \tag{5.21}$$

$\text{Dism}_{\text{SDIS}}$ and $\text{Dism}_{\text{SDFS}}$ are fused by weighted sum rule of score level fusion for classifier combination using

$$S_f = \frac{\text{I}_{\text{SDIS}} \times Z_{\text{SDIS}} + \text{I}_{\text{SDFS}} \times Z_{\text{SDFS}}}{\text{I}_{\text{SDIS}} + \text{I}_{\text{SDFS}}}, \tag{5.22}$$

where I_{SDIS} and I_{SDFS} respectively denote the CCR (%) obtained by SDIS and SDFS for a particular testing condition. The probe subject is classified as the member of the gallery class which gives the lowest S_f. The CCR(%) is the percentage of correctly identified subjects at rank-1 which is obtained using Eq. (3.28).

5.4 Experiments

ReG-IF is evaluated on three publicly available datasets: CMU MoBo dataset [19], USF HumanID gait challenge dataset [22] and OU-ISIR treadmill gait dataset B [76].

5.4.1 Experiments on CMU MoBo dataset

CMU MoBo dataset [19] comprises sequences of twenty-five subjects performing four types of walk: slow walk, fast walk, walking holding a ball, and walking on an inclined plane of a treadmill. Each sequence is recorded from six different views on a same day. We split the gait sequences of lateral view of the silhouettes for the first three walking types into non-overlapping subsequences corresponding to gait periods.

Comparisons

The performance of ReG-IF is compared with SSM [23], SSP [21], STM-SPP [7] and SVB frieze [40] on lateral views of the silhouettes using L = 300, as most of the significant gait characteristics are captured in this view. Table 5.1 shows that SSM using stance correlation achieves good performance for subjects walking with different speeds but no carrying conditions. Its performance degrades significantly when the silhouette shape is distorted due to different activities between the probe and gallery sequences, e.g., fast walk vs walking with ball. Since component-based shape analysis using weighted Krawtchouk moments and part-based shape analysis using EFDs respectively aid ReG-IF and STM-SPP to achieve invariance to small carrying items, they significantly outperform SSM and SVB frieze. But the superiority of ReG-IF is attributed to the use of Lp-Gf at different cut-off frequencies to emphasize the inner part of a silhouette, which enables it to be robust against shape distortions at the boundary due to small carried items and segmentation noise.

Table 5.1: Rank-1 identification rates (%) of various methods on CMU MoBo dataset for the lateral view. Keys: 'G' - Gallery sequence; 'P' - Probe sequence; 'S' - Slow walk; 'F' - Fast walk; 'B' - Walk with ball.

G/P	SSM [23]	SSP [21]	STM-SPP [7]	SVB frieze [40]	ReG-IF
S/S	100	100	100	100	100
F/F	100	100	100	100	100
B/B	92	-	100	100	100
S/F	80	54	94	82	100
F/S	84	32	91	80	100
S/B	48	-	93	77	96
B/S	68	-	82	89	96
F/B	48	-	84	61	92
B/F	48	-	82	73	92

To make a fair comparison with SSP which uses a normalised image self-similarity plots, ReG-IF also uses split-sample cross validation technique like SSP, where gallery and probe sets correspond to different combination of walking types for each of the twenty-five subjects. We consider lateral view of two sequences per subject (total 50 sequences) walking at slow pace (2.06 miles/h) and fast pace (2.82 miles/h). Although SSP employs frequency and phase normalization of gait sequences by computing self-similarity units to account for differences in sequence length and starting phase, Table 5.1 shows that SSP is not robust to variation in walking speed. It is to be noted that the effects of variation in walking speed can be classified as either as temporal and spatial. The temporal effect

is associated with the change in phase evolution speed affecting cadence and gait period. The methods in GEI [25], CGI [28] and GFI [49] detect gait period from a gait sequence, and averages all the silhouettes of a gait period to form a GEI for gait feature extraction. Hence, these methods are robust to change in gait period, and therefore, take into account of temporal aspect of the effect of walking speed variation. The spatial effect is associated with the change in kinematics, e.g., the change in stride length and variation in arm-swing of a subject as considered in [42; 76]. Since ReG-IF analyses silhouettes using SDIS and SDFS at ten phases of a gait period, it achieves robustness against temporal effect of walking speed variation. The component-based silhouette analysis based on weighted Krawtchouk moments takes into account of arm swing while assigning weights to each component, which enables ReG-IF to achieve limited invariance to spatial effect as well.

Fig. 5.5 shows that the rank-1 identification rates of Baseline, CMU, DNGR and ReG-IF are respectively 72%, 76%, 84% and 100% for the slow vs fast walking gait sequences of profile view silhouettes, i.e., using experiment 2 of CMU entitled Across gaits condition. Since speed variation in this experiment is almost 1 km/h, we compare ReG-IF with speed-invariant methods ST-WS [43] and SI-PSA [42], which are evaluated on the OU-ISIR treadmill gait dataset A [76] with walking speed variation of 3 km/h and 4 km/h between gallery and probe gait sequences. Note that ReG-IF is robust to the temporal as well as spatial effects of walking speed variation. Additionally, it is also robust to small carrying items, segmentation noise and shadows under feet. This explains the excellent identification rates achieved by ReG-IF on CMU MoBo dataset, outperforming all other methods.

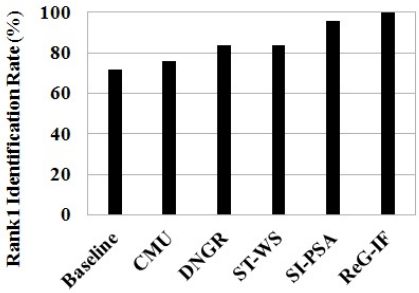


Figure 5.5: Comparison with related works. Baseline [22], CMU [19], DNGR [26] and ReG-IF are evaluated on CMU MoBo dataset (exp.2 of CMU) with walking speed variation of 3.3 km/h and 4.5 km/h, while ST-WS [43] and SI-PSA [42] are evaluated on OU-ISIR treadmill gait dataset A with walking speed variation of 3 km/h and 4 km/h between gallery and probe sequences.

5.4.2 Experiments on USF HumanID gait challenge dataset

ReG-IF is evaluated on the silhouettes of both the small version (452 sequences from 74 subjects, data acquired in May only) and the full version (1870 sequences from 122 subjects, data acquired in May and November) of the USF HumanID gait challenge dataset [22]. The silhouettes are cropped, centre-aligned and normalised to a fixed size 128×88.

Table 5.2 shows that the identification rates obatined by combining the classifiers SDIS and SDFS significantly improves the identification rates obtained by using SDIS only on the full version of USF dataset at ranks 1 and 5 using L=100, which demonstrates the effectiveness of using the combined classifier in ReG-IF.

Table 5.2: Identification rates (%) using SDIS and SDIS+SDFS at rank-1 and rank-5 on full version of USF dataset using the gallery set (G, A, R, NB, M/N) of 122 subjects using L=100.

Probe Set	A	B	C	D	E	F	G	H	I	J	K	L	WAvgI
Rank-1 Identification Rate													
SDIS	94	95	85	63	61	35	37	91	83	68	27	24	67.50
SDIS+SDFS	94	95	85	68	66	34	37	90	83	69	25	23	68.22
Rank-5 Identification Rate													
SDIS	96	98	94	81	80	61	60	95	90	84	43	39	80.46
SDIS+SDFS	97	98	95	82	81	65	66	95	90	86	44	39	82.17

Table 5.3 shows the identification rates of ReG-IF on the full version of USF dataset at ranks 1 and 5 using L=100, L=300 and L=500 to enable a comparison with methods that outperform Baseline, i.e., RCK-G [65], GFI [49], GEI [25], MMFA [111], GTDA-GF [50], CGI [28], DNGR [26], STM-SPP [7] and GPDF [47]. Since the number of probe subjects in the gait challenge experiments varies, the weighted average identification rate, i.e., WAvgI is used which is obtained by using Eq. (3.26).

ReG-IF outperforms all other methods using L=500 to achieve the best WAvgI at rank-1. However, ReG-IF is slightly inferior to GPDF using LGSR, i.e., GPDF-LGSR, in terms of WAvgI at rank-5 which uses a sophisticated classifier based on LGSR for improved WAvgI. Since the aim of the method is to demonstrate the efficacy of region-based shape analysis in gait recognition rather than achieve higher WAvgI through intensive parameter calibration, we fix N=16, as very high value of N might cause over learning. The WAvgIs for L=100, L=300 and L=500 show that the recognition accuracy does not decrease with the increase in the number of randomly chosen subspaces. Table 5.3 lists the identification rates of GTDA-GF obtained by applying GTDA as a preprocessing of LDA on the magnitude of the convolution of a GEI with sum of Gabor functions over scales with direction fixed.

Table 5.3: Identification rates (%) at rank-1 and rank-5 of the state-of-the-art gait recognition methods on full version of USF dataset using the gallery set (G, A, R, NB, M/N) of 122 subjects for comparison. Keys: 1- L=100, 2- L=300 and 3- L=500.

Probe Set	A	B	C	D	E	F	G	H	I	J	K	L	WAvgI
Rank-1 Identification Rate													
RCK-G	83	86	78	39	34	20	21	43	40	40	16	5	44.34
GFI	89	93	70	19	23	7	8	78	67	48	3	9	46.14
FED-HMM	89	88	68	35	28	15	21	85	80	58	17	15	53.54
GEI	90	91	81	56	64	25	36	64	60	60	6	15	57.66
MMFA	89	94	80	44	47	25	33	85	83	60	27	21	59.90
GTDA-GF	91	93	86	32	47	21	32	95	90	68	16	19	60.58
CGI	91	93	78	51	53	35	38	84	78	64	3	9	61.69
DNGR	85	89	72	57	66	46	41	83	79	52	15	24	62.81
STM-SPP	92	95	84	72	68	29	40	69	60	64	20	18	63.05
GPDF-NNC	90	91	85	53	52	32	28	92	86	64	12	15	62.99
GPDF-LGSR	95	93	89	62	62	39	38	94	91	78	21	21	70.07
ReG-IF[1]	94	95	85	68	66	34	37	90	83	69	25	23	68.22
ReG-IF[2]	95	96	89	55	58	34	37	92	90	79	30	28	68.77
ReG-IF[3]	95	96	90	58	60	35	37	92	91	80	32	29	70.32
Rank-5 Identification Rate													
RCK-G	96	94	88	66	63	51	46	66	68	65	44	22	67.03
GFI	98	94	93	40	47	26	25	94	85	74	24	24	63.89
FED-HMM	-	-	-	-	-	-	-	-	-	-	-	-	-
GEI	94	94	93	78	81	56	53	90	83	82	27	21	76.23
MMFA	98	98	94	76	76	57	60	95	93	84	48	39	79.90
GTDA-GF	98	99	97	68	68	50	56	95	99	84	40	40	77.58
CGI	97	96	94	77	77	56	58	98	97	86	27	24	79.12
DNGR	96	94	89	85	81	68	69	96	95	79	46	39	82.05
STM-SPP	96	98	95	80	84	59	61	92	84	85	30	27	79.13
GPDF-NNC	98	94	94	82	79	57	53	99	98	88	33	36	80.84
GPDF-LGSR	99	94	96	89	91	64	64	99	98	92	39	45	85.31
ReG-IF[1]	97	98	95	82	81	65	66	95	90	86	44	39	82.17
ReG-IF[2]	100	99	96	80	78	66	65	98	95	93	50	46	84.49
ReG-IF[3]	100	99	96	82	81	67	66	99	95	93	50	48	85.13

By using SDIS and SDFS, ReG-IF addresses shape distortions due to the presence of a briefcase, and outperforms most other methods for experiments H-J, except GTDA-GF and GPDF.

In addition to carried items, clothing variation also degrade recognition performance. Although validation based on physical parameters helps STM-SPP to achieve limited robustness to across-day gait characteristics, it is not robust to significant clothing variation, and thus fails to perform satisfactorily in experiments K- L. Since the use of Lp-Gf emphasizes the inner part of a silhouette, ReG-IF is robust against variation in clothing (which mainly cause distortions at the silhouette boundary) and achieves the best performance in terms of WAvgI at rank-1 and rank-5 for experiments K-L to ensure that ReG-IF is substantially insensitive to clothing variation. ReG-IF achieves reasonably better WAvgI than DNGR [26] despite DNGR uses eigenstance reconstruction model to create additional training data by manually annotating silhouettes for population HMM learning. ReG-IF is least affected by the variation in shoe type, which explains its better performance in experiment I than in J. Also, the number of subjects participating in experiment I is fewer than in J, with a possibility of containing few subjects from the class of subjects [22] which are difficult to identify across all the experiments for ReG-IF. Since STM-SPP crops the silhouettes to remove shadows and thus the effect of surface covariate of USF dataset, it outperforms ReG-IF for experiments D, E and G.

Table 5.4 shows the results on the smaller version of USF dataset to enable a comparison with GEI, RCK-G, RALDA and STM-SPP, which outperforms Baseline. The performance of ReG-IF is compared with that of GEI using fused classifier. Since the average identification rate of RALDA using 70 eigenvectors is better than that using 50 eigenvectors, we list the identification rates of RALDA for experiments A-G using 70 eigenvectors in Table 5.4. Table 5.4 shows that ReG-IF achieves the best WAvgI.

5.4.3 Experiments on OU-ISIR treadmill gait dataset B

ReG-IF is evaluated on the size-normalised, lateral-view silhouette sequences of OU-ISIR treadmill gait dataset B [52] using L=300 consisting of 68 subjects with up to 32 combinations of different types of clothing. The dataset is divided into three sets: a training set comprising 446 sequences of 20 subjects with all clothing combinations; a gallery set comprising gait sequences of the remaining 48 subjects with the standard clothing combination, i.e., type 9/regular pants+full shirt; and a probe set comprising 856 sequences for these 48 subjects with other clothing combinations. Since no experimental results of clothing invariant gait recognition methods, e.g., [52] are available based on OU-ISIR dataset B in terms of identification rates, we evaluated GEI on OU-ISIR dataset B using gallery and probe sets each comprising 48 subjects to compare with ReG-IF. Fig. 5.6 shows that

Table 5.4: Rank-1 identification rates (%) on the smaller version of USF dataset using the Gallery Set (G, A, R) of 71 subjects. The rates for GEI, RCK-G, RALDA and STM-SPP are from [25], [65], [119] and [7], respectively.

Exp.	GEI	RCK-G	RALDA	STM-SPP	ReG-IF
A	100	97	100	100	100
B	90	89	85	94	99
C	85	83	83	89	90
D	47	41	35	73	75
E	57	34	29	69	72
F	32	30	29	40	48
G	31	28	26	36	42
WAvgI	62	57	55	71	75

ReG-IF significantly outperforms GEI at rank-1 identification rate for all 32 probe items of varying clothing combinations.

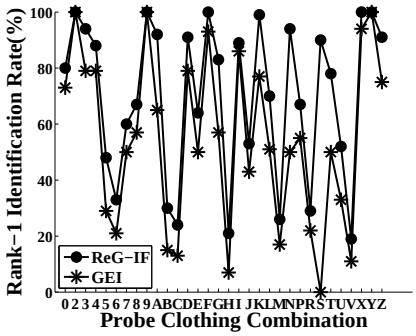


Figure 5.6: Identification rate at rank-1 for 32 probe items of OU-ISIR dataset B with different clothing combinations using gallery set of subjects with RP+FS.

5.4.4 Computational complexity analysis

The processing time to obtain SDIS and SDFS at ten phases of a gait period on Matlab 7.11.0 (R2010b) on an Intel (R) Core (TM) i7 processor working at 2.93-GHz with 4 GB RAM is 10.87 secs, whereas the average running time to obtain the reconstruction coefficient for one probe video in GPDF- LGSR using λ = 1/16 and σ = 1/2 is 9.22 secs with Matlab code on an IBM workstation (3.33-GHz CPU with 16-GB RAM) [47]. The use of ROI for detecting ten phases of a gait period enables ReG-IF to process only a small subregion of an image, thus reducing time and space complexity [7]. The computational complexity of Baseline method is very high, as it is characterised by unlimited spatio-temporal

correlation of silhouettes. The real-time method in [79] only analyses the silhouettes via the set of largest rectangles fitted to them over a gait period spanning up to 25 frames to reduce the computational complexity. But ReG-IF analyses the shape of the silhouettes at ten phases of a gait period to obtain SDIS which further reduces the computational complexity. However the analysis of every phase at five different cut-off frequencies each for Lp-Gf and Hp-Gf increases the complexity. It is to be noted that as the number of cut-off frequencies increases, the computational complexity of ReG-IF also increases. Thus ReG-IF has higher computational complexity than STM-SPP, as STM-SPP analyses the shape of contours instead of silhouettes at the double support phase using PSA and at ten phases of a gait period using EFDs.

The computational complexity of 2D FFT used to compute DFT of an $M \times N$ silhouette $I(x, y)$ in Equation 6.8 is O($MN \log MN$), where MN is the number of pixels of $I(x, y)$. The use of separable kernel reduces the computational complexity of applying Gaussian filters to an image of height h and width w to O($w_k wh$)+O($h_k wh$) as opposed to O($w_k w_h wh$) for a non-separable kernel, where w_k and w_h respectively denote the width and height of the kernel. The weighted Krawtchouk polynomials are computed using recurrence relation to avoid numerical instability [108].

5.5 Conclusion

The chapter provides a new approach to gait recognition by introducing ReG-IF, which is based on region-based analysis of a silhouette shape at ten phases of a gait period in image and feature spaces. ReG-IF uses SDIS to analyse a silhouette using Lp-Gf and Hp-Gf separately at five cut-off frequencies, where the cut-off frequencies are determined based on focus value analysis of the silhouettes for optimal performance. ReG-IF introduces a new image representation called CSI, and employs MSL to compute a dissimilarity score for SDIS. The use of MSL enables to avoid overfitting by exploiting high dimensionality of the feature space. SDFS analyses silhouettes by weighted Krawtchouk moments to compute a dissimilarity score for subject classification based on Euclidean distance classifier. A component-based shape analysis based on weighted Krawtchouk moments is introduced to achieve invariance to carrying small items. The dissimilarity scores thus obtained by SDIS and SDFS are combined using weighted sum rule of score level fusion for identifying a subject.

ReG-IF is robust to variation in subject's clothing, carrying conditions, segmentation noise, walking speed and occlusions. However, ReG-IF suffers from the following two limitations which require further investigation. First, large carried items, e.g., the rucksack carried by a mountaineer is large enough to distort the shape of the silhouette above the

shoulder, thus limiting the performance of SDFS. Also, ReG-IF does not consider shape distortion due to a subject pulling a trolley bag. Second, ReG-IF is based on lateral view of the sbjects as this view captures the most important gait characteristics. Since ReG-IF does not consider the movement of the subjects in unconstrained directions, it is not robust to variation in view.

Chapter 6

Robust View-Invariant Multiscale Gait Recognition

6.1 Introduction

The gait recognition methods STM-SPP, STS-DM and ReG-IF respectively presented in Chapter 3, Chapter 4 and Chapter 5 analyse either the contour or the silhouette of a subject for identification in the presence of a wide range of covariate factors, e.g., variation in clothing, presence of a carried item, segmentation noise, occlusion, shadows under feet and change in ground surface. Since these methods are designed to operate only on lateral views of the subjects, they are thus not robust against variation in views. Also, in addition to variation in clothing and presence of a carried item, variation in views brings challenge to a gait recognition method that limits its success to recognise a subject in an unconstrained environment. Therefore, a gait recognition method must achieve robustness against these challenges to successfully identify an unknown subject in real world for visual surveillance and public security. This chapter introduces a two-phase view-invariant multiscale gait recognition method (referred to as VI-MGR) which is not only invariant to unconstrained variation in view, but also achieves robustness against most other challenging factors of gait recognition including variation in clothing types and presence of small carried items for identifying a subject in an unconstrained environment.

GEIs are widely used in model-free methods, as it facilitates noise-resilient robust gait feature extraction with reduced storage space and computation time. The proposed VI-MGR is a model-free method based on GEIs. It comprises two phases: (1) to determine the matching gallery view with the probe; and (2) to identify the probe subject. The limb region of a GEI better captures the discriminative information due to variation in view compared to other parts of a subject's body, and is least affected by most carrying conditions and

clothing variation. Thus, phase 1 of VI-MGR computes entropy of the limb region of the GEIs based on anthropometry to determine the matching view of the probe in the gallery using 2D PCA and 2D LDA.

Although successful attempts have been made, e.g., [28; 49; 47; 111; 50; 48], in using GEI to outperform the original GEI method [25], multiscale GEI based gait recognition in the image space using Gaussian filter has not been exploited despite its high discriminatory power. Since spatio-temporal deformation of subject's shape provides better inter-subject discrimination than its kinematics in gait analysis [23], phase 2 of VI-MGR analyses shape characteristics of GEIs in the image space using Gaussian filters at three scales, as the Gaussian filter is the only filter which can generate scale-space of an image parameterized by the size of a smoothing kernel. The motivation for analysing scale-space representation of a GEI originates from the fact that a subject in the real world shows different discriminatory shape characteristics at different scales. Motivated by the superiority of Gaussian filters than the Zernike moments, EFDs and wavelet transforms for shape classifications [113], to identify the probe subject after its view detection, VI-MGR applies Gaussian filter to a GEI at three scales to generate scale-space representation of a GEI. Gaussian filter causes smoothing or blurring of a GEI, and as the scale, i.e., standard deviation of the Gaussian filter increases, the blurriness increases with the gradual loss of boundary and exterior region highlighting the inner shape characteristics. Multiscale shape analysis is also more noise resilient as it can selectively utilise the dominant features that persist across scales. Since only dominant features persist across scales, the method is noise-resilient.

As discussed in Section 5.1 of Chapter 5, the methods in [25; 28] manually compute synthetic gait templates to achieve limited invariance to the lower body part distortions, while the method DNGR uses eigenstance reconstruction model to improve the quality of silhouettes by reducing the effect of shadows and segmentation errors. The training procedure using all predefined clothing types as in [52] is not suitable in a realistic scenario in the presence of numerous different types of clothing. Thus VI-MGR avoids overfitting without manually computing synthetic gait templates by simulating distortions and achieves invariance to clothing variation and carrying conditions by introducing multiscale gait image (MGI) with weighted random subspace learning (WRSL) for classification. Like MSL as presented in Chapter 5, WRSL is also an ensemble classifier which involves randomly selected multiple subspaces from the feature space, with each subspace associated with a classifier. However, WRSL takes into account of the weights of individual classifiers, and hence provides improved performance than MSL.

The proposed VI-MGR is motivated by the unavailability of a gait recognition method that addresses three main challenges of gait recognition, i.e., variation in view and clothing, and the presence of a carried item to identify a subject in an unconstrained sce-

nario. The novelties of VI-MGR are: (1) It first attempts to achieve robustness against three main challenges of gait recognition, as well as other covariates, e.g., segmentation noise, missing body parts, change in ground surface, shadows under feet and occlusions. (2) It introduces a reflected GEI (R-GEI) to create variation of the reference gallery views to address any unknown probe view in the range [0°, 360°]; (3) It provides a new approach to achieve view invariance by comparing the probe with all the reference gallery views based on the entropy of the limb region of a GEI; (4) It achieves invariance to clothing variation and carrying conditions by introducing MGI; (5) It uses the focus value as a measure of blurriness of the filtered GEIs to determine the ideal range of scales. A minimum number of three scales are selected from this range to make a trade-off between the computational complexity and the identification rate; (5) By introducing WRSL based classification method, VI-MGR exploits high dimensionality of the feature space to avoid overfitting and achieve high identification rate.

The rest of the chapter is organized as follows. Section 6.2 discusses related works and Section 6.3 presents VI-MGR. Experimental results are analysed in Section 6.4 and Section 6.5 concludes the chapter.

6.2 Related Work

Several gait recognition methods have been proposed in the literature to achieve robustness against variation in view between the probe and gallery subjects. The methods in [2; 54] obtain view invariant gait features for subject identification. The methods in [56; 61; 57; 58; 59] depend on either mapping or projection relationship between the gait characteristics obtained across views in a shared subspace based on supervised learning. The methods in [60; 39] require a fully cooperative environment fitted with multiple calibrated cameras to capture 2D images of a subject from different views. The 2D images captured by multiple cameras are used to create a 3D model of a subject. The 2D gait characteristics of unknown view of the probe are reconstructed from the 3D model for view invariant gait recognition.

The methods in [7; 8; 45; 17] aim to achieve invariance to carrying conditions. The method STM-SPP [7] (and presented in Chapter 3) analyses the shape of a silhouette contour using PSA at the double support phase and EFDs at ten phases of a gait period. The method STS-DM in [8] (and presented in Chapter 4) combines model-based and model-free approaches to analyse the spatio-temporal shape and dynamic motion characteristics of a subject's contour. A part-based EFD analysis and a component-based FD analysis based on anthropometry are respectively used in STM-SPP and STS-DM to address shape distortions due to a subject carrying small items. An iterative local curve embedding algorithm is used in [45] to extract double helical signatures from the subject's limb to take into account of

shape distortion due to a specific carrying condition, e.g., a briefcase in upright position. The method in [17] uses models to obtain skeleton parameters by wavelet decomposition of a GEI and extract invariant moments for combining anatomical and behavioural gait characteristics. The use of thermal imaging enables it to achieve invariance to carrying conditions and lighting variation.

Most gait recognition methods (e.g., [7; 48; 28; 49; 47; 52]) aim to address different covariate conditions that affect gait recognition, but assume same views, i.e., lateral views of the gallery and probe gait sequences, as this view contains most of the significant gait characteristics. Therefore, they cannot effectively identify a human subject moving freely in unconstrained directions of real life. The method in [48] enhances the dynamic information content of GEI by computing gait entropy image from a gait period to identify a subject with varying covariate conditions for gallery and probe sequences. The method in [28] captures temporal information of the gait period into a CGI and generates real and synthetic CGI templates to address lower body part distortions due to carrying a briefcase, variation in ground surface, clothing and footwear, but not the distortions in the upper body due to carrying conditions and variation in clothing. Since different clothes worn by the same subject in gallery and probe gait sequences reduce the identification rate, the method in [52] introduces part-based feature extraction strategy based on adaptive weight control to achieve clothing invariance.

While the trend of the existing gait recognition methods is to address one or a few covariates,VI-MGR first aims to address three main challenges of gait recognition, i.e., variation in view, clothing and carrying conditions in addition to other covariate factors that affect existing gait recognition systems, namely change in ground surface, missing body parts due to occlusions and segmentation errors.The aim is achieved by multiscale shape analysis and WRSL.

6.3 Proposed method: VI-MGR

6.3.1 Phase 1: probe view matching

To automatically determine the matching gallery view of the probe subject, all the probe views are required to be available in the gallery. However, since phase 2 of VI-MGR has been designed to be robust against slight variation in view, the subject identification rate is not significantly affected for the following cases: (1) if phase 1 incorrectly matches the probe view to a gallery view close to the matching view; and (2) if phase 1 matches the closest gallery view due to absence of the exact matching view in the gallery. In addition, to take into account of a probe subject moving freely in different directions between 0° to 360°, VI-MGR computes mirror reflection of the available views of the gallery GEIs to

create R-GEIs using

$$f(x, y) = f(-x, y), \tag{6.1}$$

where $f(x, y)$ is the original GEI. Fig. 6.1 shows reflected GEIs corresponding to the 11 different views of CASIAB gait dataset in [0°,180°] to create additional gallery views in [180°,360°]. Thus, VI-MGR does not require the availability of all probe views in the gallery as in the method in [58].

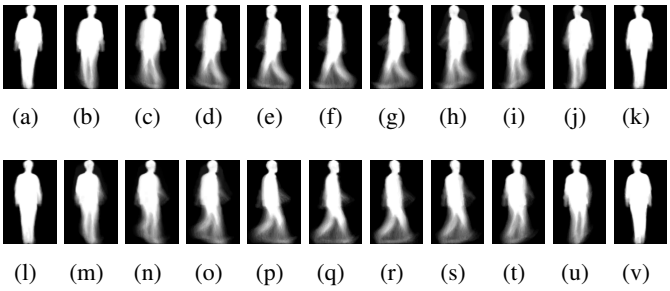


Figure 6.1: (a)-(k): GEIs of a subject from CASIA B gait dataset at 11 views; (l)-(v): the corresponding R-GEIs.

When there is a variation in view, better distinguishable shape variation is manifested in the limb region compared to the head and torso of the GEIs. Also, the shape of the head and/or torso of a GEI are significantly affected by the carried items on head and back, and using folded arms. It is also affected by carrying a briefcase in upright position, and most clothing types. Thus, VI-MGR obtains segmented GEIs, i.e., SGEIs, by cropping the region enclosed between the bottom of the GEIs and up to the anatomical positions of knee, i.e., $0.285H$, where H is the height of GEI. Fig. 6.2 (a)-(k) show the GEIs of a normal walking sequence from CASIA B gait dataset at 11 views and Fig. 6.2 (l)-(v) show the corresponding SGEIs. Fig. 6.3 (a)-(k) show the GEIs of a subject carrying a bag from CASIA B gait dataset at 11 different views and Fig. 6.3 (l)-(v) show the corresponding SGEIs. The similarity between SGEIs of the normal walking sequence and walking with a bag for a particular view shows that SGEIs are not usually affected by the carrying conditions. Although SGEIs are likely to be partially occluded by the presence of long skirts and long coats, they remain unaffected by the most clothing which supports the use of SGEIs to achieve invariance to view. VI-MGR computes 22 SGEIs corresponding to 22 views of a subject's GEI created by mirror reflection of its 11 views from CASIA gait dataset B walking without wearing coat or carrying a bag. These 22 SGEIs are set as the Reference SGEIs (Rf-SGEIs) for use as the matching probe view detection.

The entropy or average information of an image is a statistical measure of randomness of the image used to characterise the image texture. To automatically detect the view

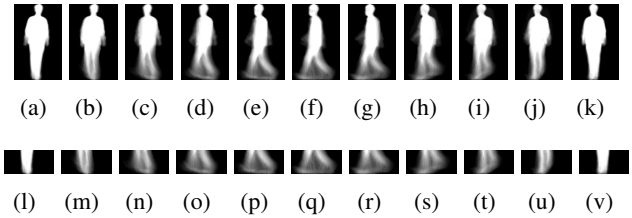


Figure 6.2: Row 1: (a)-(k) - GEIs of a same subject from CASIA B gait dataset for normal walking at 11 views. Row 2: (l)-(v) - the corresponding SGEIs.

of an unknown probe subject, the entropy of Rf-SGEIs and the SGEI of the unknown view of the probe, i..e., target SGEI (Tr-SGEI) are computed using [48]

$$\text{SGEnI} = -\sum_{b=1}^{B} P_b(x,y) log_2 P_b(x,y), \tag{6.2}$$

where $P_b(x,y)$ is the probability that the grey level of pixel (x,y) is b. Fig. 6.4 shows the segmented gait entropy images (SGEnI) obtained from the SGEIs shown in Fig. 6.3. It shows that the SGEIs are characterised by high intensity values in their corresponding SGEnIs, thus enhancing the dynamic characteristics of the limb region for probe view detection.

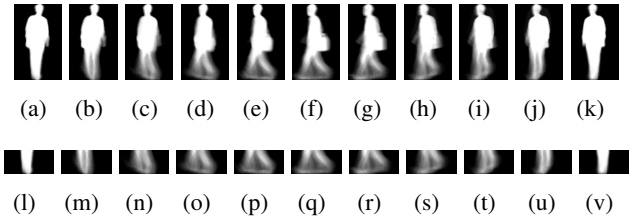


Figure 6.3: Row 1: (a)-(k)- GEIs of a same subject carrying a bag from CASIA B gait dataset at 11 views. Row 2: (l)-(v)- the corresponding SGEIs.

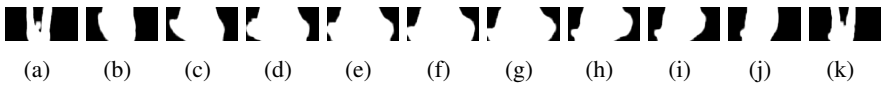


Figure 6.4: SGEnIs of a subject from CASIA B gait dataset for 11 views.

Let {$\mathbf{En}_1$, $\mathbf{En}_2$,..., $\mathbf{En}_M$} be the set of M number of $m \times n$ reference SGEnIs, i.e., Rf-SGEnIs, where $M = 22$ corresponding to 22 views. The scatter matrix is obtained using

$$\mathbf{S} = \frac{1}{M}\sum_{i=1}^{M}(\mathbf{En}_i - \mathbf{M})^T \times (\mathbf{En}_i - \mathbf{M}), \tag{6.3}$$

where $\overline{\mathbf{M}} = \frac{1}{M}\sum_{i=1}^{M}\mathbf{En}_i$. For a given Rf-SGEnI denoted by **En**, let

$$\mathbf{Y}_k = \mathbf{EnX}_k, \tag{6.4}$$

where $\mathbf{X_k}$ for k=1, 2,..., d are the orthonormal eigenvectors of **S** corresponding to the first d largest eigenvalues. Thus, $m \times d$ feature images are represented by $\mathbf{V}_i = [\mathbf{Y}_1^i, \mathbf{Y}_2^i, ..., \mathbf{Y}_d^i]$ for i=1,2,...,M. For detecting the view of a probe subject, we define

$$D(\mathbf{V}_i, \mathbf{V}_j) = \sum_{k=1}^{d} \|\mathbf{Y}_k^i - \mathbf{Y}_k^j\|_2. \tag{6.5}$$

The probe sample **V** is assigned to class W_k if

$$D(\mathbf{V}, \mathbf{V}_l) = \min_{j=1}^{c} D(\mathbf{V}, \mathbf{V}_j), \qquad j = 1, ..., c \tag{6.6}$$

and $\mathbf{V}_l \in W_k$.

6.3.2 Phase 2: subject identification

The probe subject and its matching views of the gallery subjects are subjected to multiscale analysis. The transfer function of 2D Gaussian distribution in spatial domain is given by [117]

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{\frac{-(x^2+y^2)}{2\sigma^2}}, \tag{6.7}$$

where σ denotes the standard deviation, i.e., the scale of the Gaussian distribution. The surface plot of 2D Gaussian function with $\sigma = 2$ is shown in Fig. 6.5.

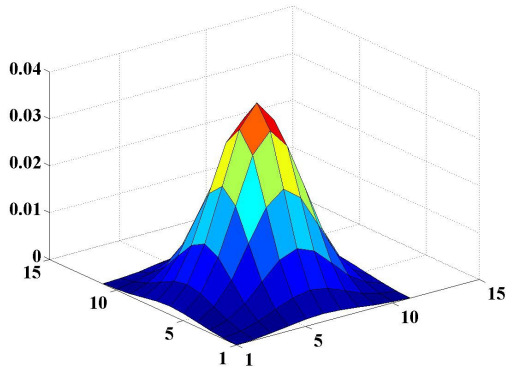


Figure 6.5: The 2D Gaussian kernel with σ=2.

The DFT of an $M \times N$ $GEI(x, y)$ is

$$GEI(u, v) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} GEI(x, y) e^{-j2\pi((ux/M)+(vy/N))}, \tag{6.8}$$

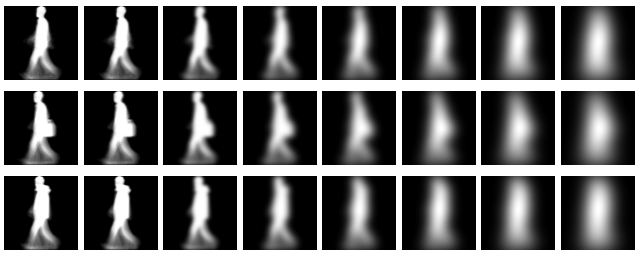
where $u = 0, 1, 2, \ldots, M-1$ and $v = 0, 1, 2, \ldots, N-1$ are frequency variables.

To highlight the characteristics of inner region of a GEI gradually towards the centre while losing its boundary shape characteristics which are more likely to be distorted due to variation in clothing types and carrying conditions, Gaussian filter is applied to the GEI using selected scales s, i.e.,

$$GEI_L(u, v)^{<s>} = GEI(u, v)\, G(u, v). \tag{6.9}$$

The Gaussian-blurred GEI at scale s in the image space is obtained by applying inverse DFT, i.e.,

$$GEI(x, y)^{<s>} = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} GEI_L(u, v)^{<s>} e^{j2\pi((ux/M)+(vy/N))}. \tag{6.10}$$



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

Figure 6.6: Original and Gaussian-blurred GEIs of CASIA B gait dataset for: row 1- normal walking gait sequences; row 2- gait sequences of subjects with clothing variation; and row 3- gait sequences of subjects with carrying conditions. Row 1, row 2 and row 3: (a) Original GEIs; and (b)-(h)- original GEI filtered by Gaussian filter using increasing scales: (b) σ_1 = 1; (c) σ_2 = 5; (d) σ_3 = 10; (e) σ_4 = 15; (f) σ_5 = 20; (g) σ_6 = 25; and (h) σ_7 = 30, respectively.

Fig. 6.6 shows the GEIs of a subject for three types of walking sequences, i.e., normal walking, variation in clothing and carrying conditions of CASIA B gait dataset filtered by Gaussian filter with respect to increasing scales, i.e., $\sigma_1 = 1$, $\sigma_2 = 5$, $\sigma_3 = 10$, $\sigma_4 = 15$, $\sigma_5 = 20$, $\sigma_6 = 25$, $\sigma_7 = 30$. Since Gaussian filter attenuates high frequency components, it blurs the GEIs. As the scales of the Gaussian filter increases, the blurriness increases resulting in gradual loss of exterior regions of a GEI with the inner shape characteristics being highlighted. The variation in clothing and carrying conditions mainly cause alterations

in the boundary of a GEI as evident from Fig. 6.6(a), (g) and (m). The shape distortions caused due to clothing variation and presence of a carried item are very prominent in the original GEIs, but are gradually reduced as the blurriness increases due to the application of Gaussian filter with decreasing scales as evident from Fig. 6.6, thus enabling VI-MGR to achieve robustness against these challenges. In addition to these challenges, VI-MGR is also insensitive to the other covariate factors that cause distortions in the shape of a boundary, i.e., variation in the footwear, change in hair style, shadows under feet and segmentation imperfections.

Scale selection

For multiscale shape analysis, inappropriately chosen scales decrease the shape recognition performance considerably, while the use of proper combination of scales enhances the recognition performance substantially. It is thus necessary to make an early judgment about which shape characteristics play an important role for a given application, and to choose the proper combination of scales responsible for these characteristics for an optimal performance [117]. VI-MGR computes the focus value of a silhouette of CASIA B gait dataset filtered by Gaussian filter using different scales to determine the effective range of analyzing scales. Like ReG-IF (presented in Chapter 5), VI-MGR also uses a wavelet based method to compute the focus value of a silhouette to measure the degree of sharpness of an image, as the wavelet based methods provide the sharpest focus measure profile and higher depth resolution due to the localised support property of wavelet basis [116]. But instead of using the magnitude of the subband images resulting from the first level 2D Daubechies-6 wavelet decomposition of a silhouette image as in ReG-IF, VI-MGR computes the variance of the subband images to measure the focus value (FV) of a silhouette as follows [116]

$$FV = (\frac{1}{MN}\sum_{y=0}^{N}\sum_{x=0}^{M}(\text{var}(|W_{HL}(x,y)|) + \text{var}(|W_{LH}(x,y)|) + \text{var}(|W_{HH}(x,y)|)))^{\frac{1}{2}}, \qquad (6.11)$$

where var(.) computes variance, and W_{LL}, W_{HL}, W_{LH} and W_{HH} are the four subband images generated by the first level 2D Daubechies-6 wavelet decomposition of a silhouette image (see Section 5.3.2 of Chapter 5).

Fig. 6.7 shows that normalised focus value decreases w.r.t. increasing scales in the range [1, 30], where focus values are normalised by dividing each of them with the maximum focus value in that range, i.e., the focus value of the silhouette filtered using σ=1. The normalised focus value decreases abruptly in the low scale range, i.e., up to scale 14, and then steadily decreases up to $\sigma = 24$. If the scale is increased above $\sigma = 24$, the GEIs fail to preserve any inter-subject discriminatory information due to excessive blurriness,

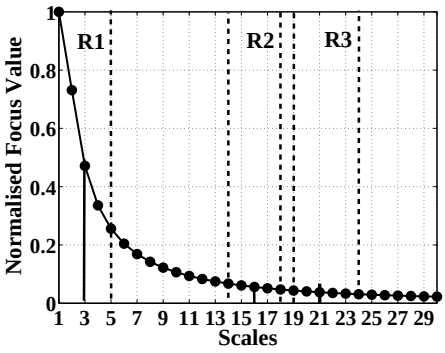
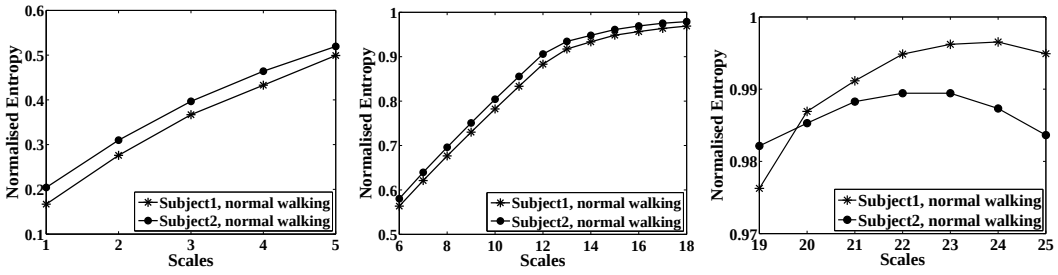


Figure 6.7: Normalised focus value w.r.t. increasing scales of the Gaussian filter.

and hence is not worthy for consideration. Therefore, the ideal range of scales is chosen as [1, 24]. For improved identification rate, it is very important to determine the analysing scales correctly. Also, the computational complexity increases as the number of scales increases. Hence, to make a trade-off between the identification rate and the computational complexity, a set of minimum number of scales are chosen based on analysing the entropy of the filtered GEIs by Gaussian filter using scales in the range [1, 24] for the following three cases:

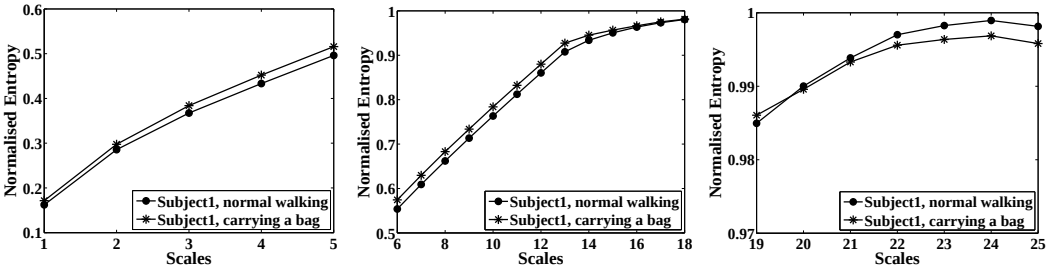
Case 1- Different subjects with no variation in clothing and carrying conditions. Fig. 6.8, Fig. 6.9 and Fig. 6.10 shows that the difference in entropy between two normal walking sequences of different subjects are higher at three scale ranges compared to same subject with clothing variation and with/without carrying conditions. However, the comparative study between Fig. 6.8 (a), (b) and (c) shows that the discriminability is the highest in the low range of scales, i.e., [1, 5]. Thus, the scales in the range [1, 5] are more informative for inter-subject discriminability without any variation in clothing and presence of a carried item.



(a) (b) (c)

Figure 6.8: Comparison between normalised entropy of filtered GEIs of normal walking sequences of two different subjects at three range of scales: (a) σ = [1, 5]; (b) σ = [6, 18] and (c) σ = [19, 25].

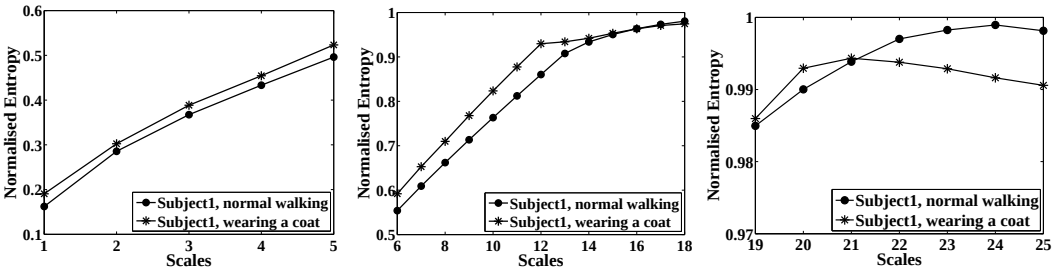
Case 2-Same subjects with variation in carrying conditions. The entropy difference between a normal walking subject and the same subject with a carried item is moderate in the low scale range, i.e., [1, 5], but decreases favourably when the entropy values of two sequences become almost similar in the intermediate range of scales, i.e., [14 18] (see Fig. 6.9).



(a) (b) (c)

Figure 6.9: Comparison between normalised entropy of filtered GEIs of a normal walking subject and the same subject with carrying a bag at three range of scales: (a) σ = [1, 5]; (b) σ = [6, 18] and (c) σ = [19, 25].

Case 3-Same subjects with variation in clothing types. The entropy difference between a normal walking subject and the same subject with clothing variation is high in the low scale range, i.e., [1, 5]. However, the difference is significantly low in the intermediate scale range, i.e., [14, 18] ensuring that Gaussian blur reduces the shape distortions due to clothing variation in this range (see Fig. 6.10). However, for excessive clothing variation and presence of more than one carried items (at the back, folded arms or is upright position), more blurriness resulted by the high scale range, i.e., [19, 24] is desirable.



(a) (b) (c)

Figure 6.10: Comparison between normalised entropy of filtered GEIs of a normal walking subject and the same subject with clothing variation at three range of scales: (a) σ = [1, 5]; (b) σ = [6, 18] and (c) σ = [19, 25].

VI-MGR thus considers three ranges of scales - R1: [1,5], R2: [14,18] and R3: [19,24] (see Fig. 6.7) for multiscale analysis of GEIs. For the sake of computational com-

plexity, the mid scale from each range, i.e., 3 (from R1), 16 (from R2) and 21 (from R3) are selected. The filtered GEIs using scales 3, 16 and 21 are concatenated to form MGI for use as a gait signature.

Classification using WRSL

2D PCA is used in VI-MGR as a preprocessing step for WRSL in order to reduce the dimensionality of the feature space via data decorrelation without losing discriminatory information. Since 2D PCA is directly based on 2D image matrices rather than 1D vectors, it is superior to PCA in terms of accurate estimation of covariance matrices and computational complexity for feature extraction [118]. Given n MGIs, i.e., $\{\mathbf{G}_1, \mathbf{G}_2, ..., \mathbf{G}_n\}$ in the gallery, the scatter matrix $\mathbf{S}$ is obtained using

$$\mathbf{S} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{G}_i - \mathbf{M})^T \times (\mathbf{G}_i - \mathbf{M}), \tag{6.12}$$

where $\mathbf{M} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{G}_i$. Since there are at most $n-1$ eigenvectors of $\mathbf{S}$ with nonzero eigenvalues, N eigenvectors (where $N < n-1$) are randomly chosen from the set of $n-1$ eigenvectors, i.e., $\{\mathbf{e_1}, \mathbf{e_2}, ..., \mathbf{e_{n-1}}\}$, with the largest eigenvalues to construct L subspaces $\{R_k\}_{k=1}^{L}$. The n-th eigenvector with zero eigenvalue is discarded in order to reduce the dimensionality of the feature space preserving discriminatory information.

For each subspace R_k, k = 1,..., L, the projected image is

$$\mathbf{Y}_i = \mathbf{M}_{2\mathrm{DPCA}}\mathbf{G}_i = [\mathbf{e_1}, \mathbf{e_2}, ..., \mathbf{e_N}]^T\mathbf{G}_i, \qquad i = 1, ..., n. \tag{6.13}$$

where $\{\mathbf{Y}_1, ..., \mathbf{Y}_n\}$ belong to C gallery classes. To achieve optimal class separability and overcome singularity problem associated with 1D LDA, 2D LDA is used to seek a transformation matrix W that maximizes the ratio of the between-class scatter matrix $\mathbf{S_B}$ to the within-class scatter matrix $\mathbf{S_W}$, i.e.,

$$J(W) = \frac{|W^T S_B W|}{|W^T S_W W|}. \tag{6.14}$$

where

$$\mathbf{S_W} = \sum_{i=1}^{C}\sum_{\mathbf{Y}\in D_i}(\mathbf{Y} - \mathbf{m}_i)(\mathbf{Y} - \mathbf{m}_i)^T, \tag{6.15}$$

$$\mathbf{S_B} = \sum_{i=1}^{C} n_i(\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T, \tag{6.16}$$

$\mathbf{m_i} = \frac{1}{n_i}\sum_{\mathbf{Y}\in D_i}\mathbf{Y}$, D_i is the training template set that belongs to the i-th class and n_i is the

number of templates in D_i. J(W) is maximised when the columns of W are the generalised eigenvectors that correspond to the largest eigenvalues in

$$\mathbf{S_B w}_i = \lambda_i \mathbf{S_W w}_i, \qquad i = 1, ..., C. \tag{6.17}$$

Since there are at most C-1 nonzero eigenvalues, the corresponding v_1,..., v_{C-1} eigenvectors are used to form the training gait feature matrix $\mathbf{Z}_i$ for each of L subspaces using

$$\mathbf{Z}_i = \mathbf{M}_{\text{2DLDA}} \mathbf{Y}_i = [\mathbf{v_1}, ..., \mathbf{v_{C-1}}]^T \mathbf{Y}_i, \qquad i = 1, ..., n. \tag{6.18}$$

Let $\{\mathbf{r}\}$ be the set of gallery gait feature matrices belonging to C classes. Therefore, each of C classes of the gallery with n_j feature matrices is represented by its centroid, i.e., $G_j = \frac{1}{n_j} \sum_{r \in R_j} \mathbf{r}$, where j = 1,..., C and R_j is the set of feature matrices belonging to the j-th class.

The distance between a probe sequence P with m_p gait periods and a gallery class centroid G_j is [25]

$$D(P, G_j) = \frac{1}{m_p} \sum_{i=1}^{m_p} \|\mathbf{s} - G_j\|, \qquad j = 1, ..., C. \tag{6.19}$$

where $\mathbf{s}$ is the set of a probe gait feature matrices. We assign

$$P \in w_k, \text{ if } D(P, G_k) = \min_{j=1}^{C} D(P, G_j), \tag{6.20}$$

where w_k is the k-th gallery class. A probe sequence P is assigned a class label by each component classifier $A_{\text{LH-GF}} \in \{A_{\text{LH-GF}}\}_{k=1}^{L}$, where $\{A_{\text{LH-GF}}\}_{k=1}^{L}$ is the set of classifiers constructed from L subspaces. The final class of P is determined based on majority voting and the classification at rank-r implies that the number of votes received by a probe belong to the top r ranks for correct classification.

WRSL is run ten times for each experiment with L $\in \{100, 300, 500\}$ and N fixed. It is observed that some pairs of common values for rank-1 and rank-5 identification rates are produced more than once. Therefore, VI-MGR computes weighted average of the rank-1 and rank-5 identification rates produced in 10 runs using

$$\frac{1}{10} \sum X_i W_i, \tag{6.21}$$

where W_i denotes the number of times the identification rate X_i is produced.

6.4 Experiments

VI-MGR is evaluated on three publicly available gait datasets to demonstrate its efficacy: CASIA B gait dataset, USF HumanID gait challenge dataset and OU-ISIR treadmill gait dataset B.

6.4.1 Experiments on CASIA B dataset

Table 6.1: Correct view matching rate (CVMR) of VI-MGR on CASIA B gait dataset.

Probe View	CVMR (%)			
	Normal	Bag	Coat	AvgA
0°	83	80	79	80.67
18°	94	87	85	88.67
36°	88	85	80	84.33
54°	92	90	89	90.33
72°	81	80	78	79.67
90°	89	79	72	80
108°	79	75	70	74.67
126°	90	88	85	87.67
144°	83	81	79	81
162°	89	86	84	86.33
180°	82	80	75	79
Mean	86.4	82.8	79.6	-

Table 6.1 shows the correct view matching rate (CVMR) of VI-MGR on CASIA B gait dataset for the subjects walking normally, walking with a bag and walking wearing a coat for 11 views. The mean CVMR% is the highest, i.e., 84.64% for the normal walking sequences, followed by walking with a bag, and walking wearing a coat as the SGEIs are partially occluded by the long coats for several subjects. The average CVMRs of a particular angle for all types of walking (AvgAs) are comparatively low for views 0° and 180°, as the shape characteristics of a subject remain almost same in these cases as evident from Fig. 6.3. Also, the degradation in performance for views 72° and 108° is attributed to the similar shape characteristics with the subjects at 90°. However, these degradation in performance does not affect the identification rate, as the subject identification rate is invariant to slight variation in view.

Table 6.2 shows the rank-1 and rank-5 identification rates produced by VI-MGR using L=300 on CASIA B gait dataset. The better subject identification rate even when the CVMR% is low is attributed to the robustness of the phase 2 of VI-MGR against limited

variation in view. The tables shows that VI-MGR provides excellent identification rates, i.e., 100% at rank-1 and rank-5 for most views of the normal walking sequences. However, the identification rate is decreased in the presence of carrying conditions and clothing variation.

Table 6.2: Rank-1 and rank-5 identification rates (%) of VI-MGR on CASIA B gait dataset.

View	Identification Rate (%)					
	Normal/Normal		Normal/Coat		Normal/Bag	
	Rank1	Rank5	Rank1	Rank5	Rank1	Rank5
0°	100	100	67	80	93	95
18°	99	100	56	77	89	94
36°	100	100	70	83	89	95
54°	99	99	80	95	90	97
72°	100	100	71	95	77	91
90°	100	100	75	93	80	94
108°	99	100	77	94	82	91
126°	99	100	75	88	84	91
144°	100	100	65	76	92	97
162°	100	100	64	75	93	96
180°	99	100	66	73	89	93
Mean	99.5	99.9	69.6	84.5	87.1	94

Since CGI is evaluated on CASIAB gait dataset using gait sequences captured from 90° only, Table 6.3 shows the rank-1 identification rates of VI-MGR using L=300 evaluated on 90° view of the gait sequences using same experimental set up as that of CGI for a fair comparison. Note that GEI performs better than CGI when the training and testing conditions are same, and vice versa [28]. This shows that CGI is more robust to covariates than GEI. However, VI-MGR significantly outperforms CGI and GEI for all experiments, thus showing that the method is substantially invariant to clothing variation and carrying conditions. Additionally, VI-MGR is also robust against variation in view.

Table 6.3: Rank-1 identification rates (%) of VI-MGR, GEI and CGI on CASIA gait dataset B, with rates of GEI and CGI obtained from [28] for lateral views.

Gallery/Probe	GEI [25]	CGI [28]	VI-MGR
Normal/Normal	91.57	88.06	100
Normal/Bag	31.71	43.67	89
Normal/Coat	24.07	42.98	76

6.4.2 Experiments on USF HumanID gait challenge dataset

We evaluate VI-MGR on the GEIs of the full version, i.e., version 2.1 of USF HumanID gait challenge dataset which are downloaded from http://www.GaitChallenge.org. The GEIs provided by the USF HumanID dataset are centre-aligned and normalised to a fixed size 128×88.

Comparisons

Table 6.4 shows the identification rates of methods that outperform Baseline, i.e., RCK-G, GFI, GEI, MMFA, GTDA-GF, CGI, DNGR, STM-SPP, GPDF using NNC, i.e., GPDF-NNC, GPDF using LGSR classifier, i.e., GPDF-LGSR and VI-MGR using L = 100, 300, 500 on the full version of the USF dataset at ranks 1 and 5. Since there are different number of probe subjects in the gait challenge experiments, the weighted average identification rate, i.e., WAvgI is used which is obtained using Eq. (3.26). The method MMFA in [111] applies marginal Fisher analysis on GEIs for gait representation to reduce the dimensionality of the feature space and extends marginal Fisher analysis to marginal based analysis for content-based image retrieval. Table 6.4 shows the identification rates of GEI obtained by fusing real and synthetic gait features. GTDA-GF reports the identification rates obtained by applying GTDA as a preprocessing step of LDA on the magnitude of convolving a GEI with sum of Gabor functions over scales with direction fixed. VI-MGR outperforms all other methods using L=100, 300 and 500 at rank-1 and rank-5 except GPDF-LGSR. GPDF-LGSR achieves higher identification rates at rank-1 and rank-5 than VI-MGR at the expense of very high computational complexity (see Section 6.4.4) due to the use of a sophisticated classier called LGSR. Since the aim of this Chapter is to demonstrate the efficacy of multiscale approach in gait recognition rather than achieve higher WAvgI through intensive parameter calibration, we fix N=16, as very high value of N might cause over learning. The WAvgIs for L=100, L=300 and L=500 show that the recognition accuracy does not decrease with the increase in the number of subspaces.

It is revealed from Table 6.5 that VI-MGR outperforms Baseline, ETGLDA, GEI, RALDA, RCK-G and STM-SPP in terms of WAvgI on the small version of USF dataset. W-MGA is compared with the identification rates of GEI obtained by using fused feature classifier, as it shows better performance than the individual real and synthetic feature classifiers for most of the gait challenge experiments.

The rank-1 identification rates obtained using single scale (σ=0, i.e., without blurring of the GEIs) and multiple scales (σ=3, 16, 21) for all 12 gait challenge experiments (exp. A-L) are shown in Fig. 6.11 to demonstrate the effectiveness of the proposed multiscale approach. The dashed line and the straight line in the figure respectively denote

Table 6.4: Identification rates (%) at rank-1 and rank-5 of the state-of-the-art gait recognition methods on full version of USF HumanID gait challenge dataset using the gallery set (G, A, R, NB, M/N) of 122 subjects. Keys: 1- L =100; 2-300; and 3-500.

Probe Set	A	B	C	D	E	F	G	H	I	J	K	L	WAvgI
Rank-1 Identification Rate													
RCK-G	83	86	78	39	34	20	21	43	40	40	16	5	44.34
GFI	89	93	70	19	23	7	8	78	67	48	3	9	46.14
GEI	90	91	81	56	64	25	36	64	60	60	6	15	57.66
MMFA	89	94	80	44	47	25	33	85	83	60	27	21	59.90
GTDA-GF	91	93	86	32	47	21	32	95	90	68	16	19	60.58
CGI	91	93	78	51	53	35	38	84	78	64	3	9	61.69
DNGR	85	89	72	57	66	46	41	83	79	52	15	24	62.81
STM-SPP	92	95	84	72	68	29	40	69	60	64	20	18	63.05
STS-DM	93	96	86	70	69	39	37	78	71	66	27	22	66.68
GPDF-NNC	90	91	85	53	52	32	28	92	86	64	12	15	62.99
GPDF-LGSR	95	93	89	62	62	39	38	94	91	78	21	21	70.07
VI-MGR[1]	94	95	85	50	51	29	37	91	86	79	29	26	66.18
VI-MGR[2]	95	96	88	51	52	33	37	92	90	80	30	27	67.79
VI-MGR[3]	95	96	89	55	58	35	38	92	90	80	32	29	69.07
Rank-5 Identification Rate													
RCK-G	96	94	88	66	63	51	46	66	68	65	44	22	67.03
GFI	98	94	93	40	47	26	25	94	85	74	24	24	63.89
GEI	94	94	93	78	81	56	53	90	83	82	27	21	76.23
MMFA	98	98	94	76	76	57	60	95	93	84	48	39	79.90
GTDA-GF	98	99	97	68	68	50	56	95	99	84	40	40	77.58
CGI	97	96	94	77	77	56	58	98	97	86	27	24	79.12
DNGR	96	94	89	85	81	68	69	96	95	79	46	39	82.05
STM-SPP	96	98	95	80	84	59	61	92	84	85	30	27	79.13
STS-DM	97	98	96	82	83	61	60	95	89	83	39	28	80.48
GPDF-NNC	98	94	94	82	79	57	53	99	98	88	33	36	80.84
GPDF-LGSR	99	94	96	89	91	64	64	99	98	92	39	45	85.31
VI-MGR[1]	97	98	95	81	80	63	60	95	90	86	44	39	81.36
VI-MGR[2]	100	99	96	80	78	66	64	98	95	92	46	43	83.87
VI-MGR[3]	100	99	96	81	80	67	65	99	95	92	50	48	84.75

Table 6.5: Top-rank identification rates (in percentage) for different methods on the small version of USF HumanID gait challenge dataset (data acquired in May only) using the gallery set (G, A, R) of 71 subjects. The rates for ETGLDA, FED-HMM, GEI, RALDA, RCK-G and STM-SPP are from [51], [37], [25], [119], [65] and [7], respectively.

Exp.	ETGLDA	FED-HMM	GEI	RALDA	RCK-G	STM-SPP	VI-MGR
A	99	99	100	100	97	100	100
B	88	89	90	85	89	94	99
C	83	78	85	83	83	89	90
D	36	35	47	35	41	73	75
E	29	29	57	29	34	69	72
F	21	18	32	29	30	40	48
G	21	24	31	26	28	36	42
WAvgI	53	52	62	55	57	71	75

the WAvgI using single scale, i.e., 62.90% and multiple scales, i.e., 67.80%. Note that the rank-1 identification rates obtained using multiple scales as shown in the figure assume L=300 for WRSL. The figure shows that the use of multiple scales enables to achieve better identification rates than using single scale for all gait challenge experiments.

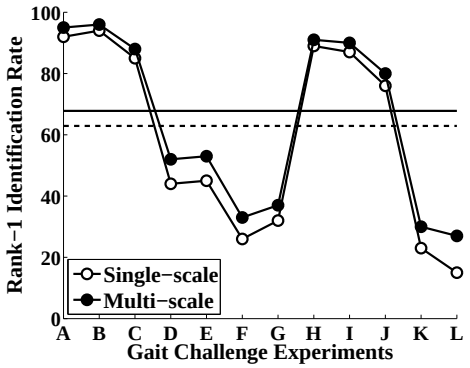


Figure 6.11: Rank-1 identification rates of VI-MGR using single scale (σ = 0) and multiple scales (σ=3, 16, 21) using L = 300 for 12 gait challenge experiments of HumanID gait challenge dataset. Dashed line: WAvgI using single scale and straight line: WAvgI using multiscale.

6.4.3 Experiments on OU-ISIR treadmill gait dataset B

VI-MGR is evaluated on the size-normalised, lateral-viewed GEIs of OU-ISIR treadmill gait dataset B [52] comprising 68 subjects with up to 32 combinations of different types of clothing. The dataset is divided into three subsets, i.e., a training set comprising 446

sequences of 20 subjects with all types of clothes; a gallery set comprising sequences of the remaining 48 subjects with standard clothes (type 9/ regular pants+full shirt); and a probe set comprising 856 sequences for these 48 subjects with other types of clothes excluding the standard clothes. Unlike the method in [52], we do not use the training dataset as it is unrealistic to train a gait recognition system with all possible types of clothing combinations. Since no experimental results of clothing invariant gait recognition methods, e.g., [52], are available based on OU-ISIR dataset B in terms of identification rates, we evaluated GEI on OU-ISIR dataset B using gallery and probe sets each comprising 48 subjects to compare with VI-MGR. Fig. 6.12 shows that VI-MGR significantly outperforms GEI at rank-1 identification rate for all 32 probe items of varying clothing types.

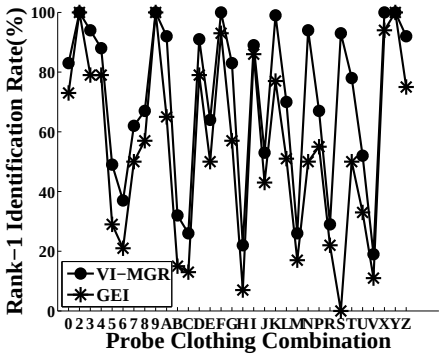


Figure 6.12: Identification rate at rank 1 for 32 probe items of OU-ISIR dataset B with different clothing combinations using gallery set of subjects with RP+FS.

6.4.4 Computational complexity analysis

The average processing time required to detect the matching gallery view with the probe, extract multiscale shape features of the probe subject and the gallery subjects of matching probe view, and to identify the probe subject using MRSL (L=300) is 5.56 secs using Matlab 7.11.0 (R2010b) on an Intel (R) Core (TM) i7 processor working at 2.93-GHz for CASIA B gait dataset. Note that the average processing time to obtain the reconstruction coefficient of a probe subject in GPDF- LGSR for $\lambda = 1/16$ and $\sigma = 1/2$ is 9.22 secs using Matlab on an IBM workstation (3.33-GHz CPU with 16-GB RAM) [47]. Hence, VI-MGR is significantly less computationally expensive than GPDF- LGSR. Like STS-DM and ReG-IF, VI-MGR also uses 2D FFT to compute the DFT of a GEI and the Gaussian filter to reduce the computational complexity. The computational complexity increases with the number of scales. Hence, to make a trade-off between identification rate and the computational comnplexity, a minimum number of three scales are selected for multiscale shape feature extraction.

6.5 Conclusion

A gait recognition method is mainly affected by the variation in view, clothing types, and presence of a carried item as well as other covariates, e.g., segmentation imperfection, shadow under feet, change in ground surface and occlusion. This Chapter proposes a two-phase view-invariant multiscale gait recognition method, i.e., VI-MGR, to achieve invariance to all these covariates for identifying a subject in an unconstrained environment. Phase 1 determines which of the available views of the gallery matches most closely with the probe view; and phase 2 compares the probe with the matching view of the gallery subjects for identification.

In phase 1, VI-MGR computes the entropy of the limb region of the GEIs to obtain SGEnIs to enhance dynamic gait characteristics for better discriminability. To determine the matching gallery view of the probe, the SGEnIs of the probe are compared with the SGEnIs of all available views in the gallery using 2D PCA and Euclidean distance classifier. It introduces a R-GEI to create synthetic gallery views to take into account of any unknown probe view in the range 0° and 360°. In phase 2, VI-MGR applies Gaussian filter to the GEIs at three scales to generate MGI as a gait feature. Focus value is used as a measure of blurriness of the filtered GEIs to determine the ideal range of scales. From this range, three scales are selected based on entropy value analysis of the filtered GEIs which are effective to provide improved identification rate in presence of clothing variation and carrying conditions with reduced computational complexity. A probe subject is classified using WMSL for overfitting avoidance exploiting high dimensionality of the feature space. Excellent identification rates on publicly available datasets demonstrate the efficacy of VI-MGR.

Chapter 7

Gait Recognition using Low Spatial and Temporal Resolution Videos

7.1 Introduction

CCTV cameras are widely installed in public places, e.g., airports, stations, government buildings, streets and shopping malls for surveillance against the threat of terrorism and to prevent criminal activities. A good quality digital video with high spatial resolution, low compression ratio and high frame-rate reproduces actual scene at the expense of increased cost, as it requires high transmission bandwidth and large disc space. To provide adequate security measures, the trend is to reduce the price of CCTV cameras by compromising the network bandwidth and storage space to install an increased number of CCTV cameras. Due to limited transmission bandwidth and storage capacity, the gait sequences captured by a CCTV camera are recorded at low frame-rates with low spatial resolution, which significantly affect the performance of a gait recognition method. Thus, this chapter introduces a gait recognition method based on low spatial and low temporal resolution silhouette sequences (referred to as GR-LSTR) captured by a CCTV camera.

GR-LSTR has the following novelties: (a) it first attempts to address the challenge of low spatial as well as low temporal resolution video sequences on gait recognition; (b) it introduces the use of DT-CWT for spatial resolution enhancement in gait recognition; (c) it introduces a new gait keyframe image (GKI) which combines the key frames of a gait sequence in a single image to preserve significant gait characteristics in reduced storage space; (d) GR-LSTR provides an experimental analysis to make a trade-off between the price of a CCTV camera and the frame-rate for a desired performance; and (e) GR-LSTR uses a simplified feature space and provides good identification rate in reduced computational complexity.

7.2 Related Work

The state-of-the-art gait period based gait recognition methods are evaluated using image sequences of standard spatial resolution, e.g., 128×88, 240×240 captured at standard temporal resolution, i.e., 30 fps. Hence, they do not perform satisfactorily on low frame-rate videos captured by a CCTV camera due to sparsity of the different phases of a gait period in the temporal domain resulting in false gait period detection. Recently, a few methods have attempted to address the challenging issue of using low frame-rate videos in gait recognition. The method in [66] compares the low frame-rate probe video sequences with the normal frame-rate gallery video sequences based on gait period based phase synchronisation. However, it fails to address the challenge when both the probe and gallery gait sequences are captured at low frame- rates. To address this challenge, the level-set morphing based method Morph in [123] uses a level-set approach to temporal interpolation and a morphing based gait cycle construction. PCA is used for dimensionality reduction and feature extraction. The method in [67] uses reconstruction-based temporal super resolution to construct a high frame-rate image sequence with a single gait period from a low frame-rate quasi periodic image sequence based on phase registration in energy minimization framework. Since the method uses linear phase evolution prior to estimation of initial phase, it is susceptible to aliasing for extremely low frame-rate video sequences, i.e., 1 or 2 fps. In addition to this, it is also susceptible to stroboscopic effect which causes the moving subject to appear stand still due to coincidence of the sampling interval with the subject's gait period.

The periodic temporal super resolution based method, i.e., P-TSR in [68] successfully addresses aliasing by using dynamic programming to directly match the low frame-rate video sequence with an exemplar high frame-rate video sequence for robust initial phase estimation. Also, it uses morphing to insert missing frames between estimated gait stances and overcomes ghosting effects. However, it is still adversely affected by stroboscopic effect. Thus, the method ER-TSR in [69] effectively combines reconstruction-based and example-based periodic temporal super resolution methods to overcome aliasing and stroboscopic effects for improved gait recognition using low frame-rate video sequences.

Although attempts have been made for gait recognition using low frame-rate videos, no gait recognition methods have addressed the challenge of low temporal resolution following spatial resolution enhancement of video sequences. The method in [91] addressed the variation in spatial resolution on gait recognition, and demonstrated that a drop in recognition rate is expected if the spatial resolution of images is decreased considerably. Thus, the aim of this chapter is to introduce a gait recognition method which is robust against low spatial and temporal resolution video sequences captured by a CCTV camera. GR-LSTR uses DT-CWT to enhance the spatial resolution of silhouettes of gait sequences. The

method provides a new direction to address the challenge of low frame-rate videos in gait recognition by introducing GKI by combining the key frames of a gait sequence. Unlike temporal interpolation based methods, GR-LSTR is unaffected by stroboscopic, aliasing and ghosting effects. The subjects are identified based on a similarity score obtained by Procrustes image distance between the GKIs.

The rest of the chapter is organized as follows. Section 7.3 presents GR-LSTR. Experimental results are analysed in Section 7.4, and Section 7.5 concludes the chapter.

7.3 Proposed method: GR-LSTR

GR-LSTR consists of four modules: 1) silhouette processing; 2) spatial resolution enhancement of silhouettes; 3) formation of GKIs; and 4) subject identification.

7.3.1 Module 1: Silhouette processing

The silhouettes are binarised after being extracted from the video sequences using background modelling and background subtraction techniques as in STM-SPP [7] (and presented in Chapter 3). The binary silhouettes are cropped according to the perimeter of the bounding rectangle enclosing the silhouette and resized to a fixed height while keeping the aspect ratio the same so as to remove camera depth variations. The resized silhouette is then copied to a destination image of fixed size by coinciding its centre-of-mass with the centre of the destination image to make it translation invariant. The silhouette is then scaled down to one-fourth of its original size to considerably decrease its spatial resolution.

7.3.2 Module 2: Spatial resolution enhancement

Spatial image resolution enhancement techniques are extensively used in satellite imaging. GR-LSTR extends its application in gait recognition by using the DT-CWT based image resolution enhancement in [124] to increase human identification rate from low spatial resolution images captured by a CCTV camera. DT-CWT is a combination of two real-valued decimated discrete wavelet transform (DWT). Compared to DWT, DT-CWT has improved directional selectivity. It is also shift invariant and has limited redundancy [124]. Fig. 7.1 shows the real and imaginary parts of DT-CWT and its magnitude in six directions [125].

DT-CWT is used to decompose a 2D low spatial resolution silhouette image $f(x, y)$ into low and high frequency subbands using a series of dilations and translations of a complex scaling function $\Phi_{j_0,l}$ and six complex wavelet functions $\psi^{\theta}_{j,l}$, [126] i.e.,

$$f(x,y) = \sum_{l \in Z^2} S_{j_0,l} \Phi_{j_0,l}(x,y) + \sum_{\theta \in \Theta} \sum_{j \geq j_0} \sum_{l \in Z^2} c^{\theta}_{j,l} \psi^{\theta}_{j,l}(x,y), \tag{7.1}$$

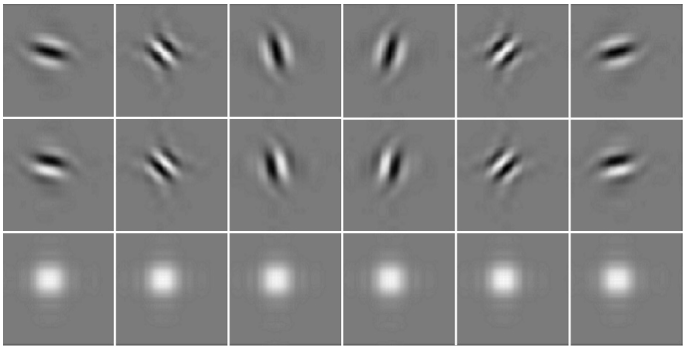


Figure 7.1: Row 1: Real part of DT-CWT in six directions; Row 2: Imaginary parts of DT-CWT in six directions; and Row 3: Magnitude of DT-CWT [125].

where Z is the set of natural numbers, j and l respectively refer to the index of shifts and dilations, $S_{j_0,l}$ is the scaling coefficient, $c_{j,l}$ is the complex wavelet coefficient and $\theta \in \Theta = \{-15°, -45°, -75°, +15°, +45°, +75°\}$ provides the directionality of the complex wavelet function. Hence, the decomposition of $f(x, y)$ by DT-CWT produces one complex-valued low-pass subband and six complex-valued high-pass subbands at each level of decomposition, where each high-pass subband corresponds to one unique direction θ.

Let one-level DT-CWT decomposition of a low spatial resolution silhouette image $\mathbf{X_L}$ of size $N \times M$ results in a matrix of DT-CWT($\mathbf{X_L}$) = [$\mathbf{LR_X}$ $\mathbf{HR_X}$], where the matrix $\mathbf{LR_X}$ of size $N/2 \times M/2$ denotes the complex-valued low-pass subband and the matrix $\mathbf{HR_X}$ of size $N/2 \times M/2 \times 6$ denotes the collection of all six complex-valued high-pass subbands. To enhance the spatial resolution of $\mathbf{X_L}$ with a scale factor of 4 in the horizontal and vertical directions, the high frequency subband images are interpolated using bicubic interpolation method by a factor α, where $\alpha = 4$ to produce the matrix $\mathbf{HR_Y}$ of size $2N \times 2M \times 6$. The original low resolution silhouette image, i.e., $\mathbf{X_L}$ is also interpolated using bicubic interpolation method by a factor $\alpha/2$ to produce $\mathbf{X}$ of size $2N \times 2M$. The interpolated high frequency subbands and the interpolated original low resolution silhouette image are combined to generate a new high-resolution silhouette image $\mathbf{X_H}$ by using inverse DT-CWT (IDT-CWT), i.e., $\mathbf{X_H}$ = IDT-CWT([$\mathbf{X}$ $\mathbf{HR_Y}$]).

Fig. 7.2(b) shows the input low spatial resolution silhouette image generated by subsampling (with a factor of 4) the groundtruth (original) high resolution silhouette image shown in Fig. 7.2(a).The reconstructed high resolution silhouette image is shown in Fig. 7.2(c). The marginal spectral distortion in the reconstructed image compared to the original image as evident from Fig. 7.2(a) and (c) demonstrate the efficacy of the DT-CWT for spatial resolution enhancement. Since low frequency subband images are the low resolution of the original image, they contain less information than the original input image. Hence, the original input image is used in the interpolation instead of the low frequency

subband images.

(a) (b) (c)

Figure 7.2: Silhouettes: (a) original resolution; (b) low spatial resolution obtained from the original resolution with a sampling factor of 4; and (c) reconstructed high spatial resolution.

The quality of the reconstructed high spatial resolution silhouette image for gait recognition is estimated using the root-mean-square error (rmse) between the groundtruth silhouette ($\mathbf{X_O}$) and the resolution-enhanced silhouette ($\mathbf{X_H}$), i.e.,

$$\text{rmse} = \sqrt{\frac{1}{NM}\sum_{x,y}(\mathbf{X_O}(x,y) - \mathbf{X_H}(x,y))^2}. \tag{7.2}$$

The closer the value of rmse to zero, the better is the quality. The rmse between the silhouettes in Fig. 7.2(a) and Fig. 7.2(c) is 0.2202 ensuring good quality of the reconstructed image. This low rmse value explains why the reconstructed image in Fig. 7.2(c) is almost same as the original image in Fig. 7.2(a) in terms of visual quality.

7.3.3 Module 3: GKI formation

A gait period starts with the heel strike of either foot and ends with the subsequent heel strike of the same foot comprising of two steps [7] (presented in Chapter 3 and summarised in this chapter). Each foot in a gait period transits between two phases: a stance phase when the foot remains in contact with the ground and a swing phase when the foot does not touch the ground. The stance phase has the following components: (a) initial contact; (b) double support; (c) mid-stance; and (d) propulsion. The swing phase has the following components: (a) pre-swing; (b) mid-swing; and (c) ending swing as illustrated in Fig. 7.3. In GR-LSTR, the seven key frames corresponding to each component of the stance and swing phases of a gait period as shown in Fig. 7.3 are extracted from a gait sequence. These key frames are characterized by distinguishable shape characteristics from each other in a gait period, and therefore effective for shape-based gait recognition.

Since it is not possible to determine gait period from a low frame-rate gait sequence due to unavailability of adequate number of frames, GR-LSTR extracts the seven key frames (i.e., initial contact, ending swing, mid-swing, pre-swing, propulsion, midstance and double support) from a gait sequence using ROI based silhouette matching by Krawtchouk mo-

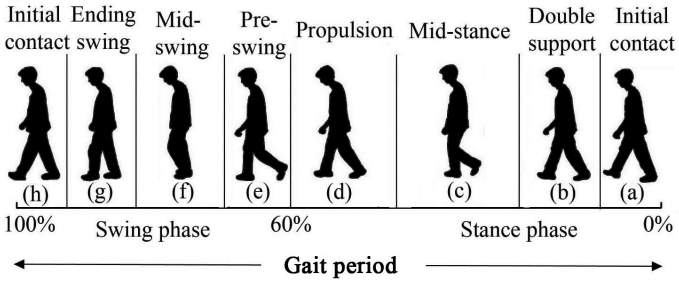


Figure 7.3: Selected key frames of a gait period (a)-(h) of a subject: stance phase (a)-(d); and swing phase (e)-(g).

ments. Krawtchouk moments have better image reconstruction capability than the Zernike and Hu moments in both noisy and noise-free conditions, and the orthogonal property of weighted Krawtchouk moments ensures minimal information redundancy [108]. They are also useful when dealing with partially distorted frames of a gait period, as they can be used to extract local features from any ROI of an image by varying the parameters N and M. The Krawtchouk moments of order $(n+m)$ of a $N \times M$ silhouette with intensity function $f(x, y)$ are computed using the sets of weighted Krawtchouk polynomials $\overline{K}_n(x; p, N)$ and $\overline{K}_n(x; p, M)$ as in Eq. (4.1), Eq. (4.2), Eq. (4.3) and Eq. (4.4). The silhouettes of the seven key frames in Fig. 7.3(a)-(g) are manually extracted. The bottom segment of the bounding rectangle, i.e., the region bounded by bottom of the bounding rectangle and the anatomical position of just before the subject's hand measured from bottom of the bounding rectangle (i.e., $0.377H$ where H is height of the bounding rectangle) is set as the Rf-ROI as this foreground region is not distorted by self-occlusions due to arm-swing. To obtain the key frames from any gait sequence automatically, the Rf-ROIs are compared with the same silhouette segments of all frames of a subject's gait sequence (each referred to as a (Tr-ROI)) using silhouette comparison based on weighted Krawtchouk moments to obtain similarity scores, i.e.,

$$S_{score} = \left[(\text{Rf-ROI}_{km_n} - \text{Tr-ROI}_{km_n})^2 \right]^{\frac{1}{2}}, \tag{7.3}$$

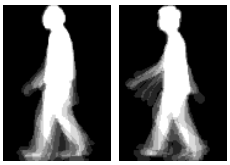
where Rf-ROI_{km_n} and Tr-ROI_{km_n} respectively denote the (n+m) order weighted Krawtchouk moments of the Rf-ROI and Tr-ROI. The frame whose Tr-ROI results in the lowest S_{score} with the corresponding Rf-ROI is extracted as one of the seven key frames, and stored in an array. The process continues by comparing the next Rf-ROI with the remaining Tr-ROIs to obtain all key frames one at a time based on the lowest S_{score}. As a result, an array of seven key frames are generated. Unlike the silhouette-based gait recognition method STM-SPP in [7] which uses contour matching based on Hu moments for the detection of ten phases of gait period, GR-LSTR computes (n+m) order weighted Krawtchouk moments of each

of the Rf-ROIs and Tr-ROIs using Eq. 4.1 by suitably choosing the values of M and N for detecting seven key frames from a gait sequence.

The preprocessed binary silhouette images $B_p(x, y)$ at the pth key frame are combined to form grey-level GKI using

$$\text{GKI} = \frac{1}{N} \sum_{p=1}^{N} B_p(x, y), \tag{7.4}$$

where N=7, and x, y denote the coordinates of a binary silhouette. Similar to GEI in [25] which is robust against silhouette noise in individual frames of a gait period, GKI is also less sensitive to silhouette noise in the constituent key frames of a gait sequence as GKI is formed by averaging the key frames of a gait sequence. Gait period detection is essential prior to GEI formation as GEI is formed by averaging all the frames of a gait period of a subject. Since it is impractical to determine the gait period of a subject from low frame-rate gait sequences, GEI is not effective for low temporal resolution gait recognition. GKI preserves significant gait characteristics of a gait sequence in reduced storage by averaging its key frames and hence, it effectively addresses the challenge of low frame-rate gait sequences. Fig. 7.4(a) and (b) respectively show the GKIs of two subjects of OU-ISIR gait dataset D.



(a) (b)

Figure 7.4: GKIs constructed from two gait sequences of OU-ISIR gait dataset: (a) subject from DB_{high} subset; and (b) subject from DB_{low} subset (see Section 7.4 for DB_{high} and DB_{low}).

7.3.4 Module 4: Subject identification

A GKI is formed to correspond to each gallery and probe gait sequence. Let S_1 ={G-GKI$_1$, G-GKI$_2$, ..., G-GKI$_g$} and S_2 = {P-GKI$_1$, P-GKI$_2$, ..., P-GKI$_p$} respectively denote the sets of gallery and probe GKIs, where g and p respectively denote the number of gallery and probe gait sequences. The Procrustes image distance between k th P-GKI and the G-GKIs of size $m \times n$ is

$$Dist_{GKI}(k) = \min_i \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} (\textbf{P-GKI}_{i,j} - \textbf{G-GKI}_{i,j})^2}{\sum_{j=1}^{n} (\sum_{i=1}^{m} (\textbf{P-GKI}_{i,j} - \text{mean}(\textbf{P-GKI}_j))^2)}, i = 1, 2, ..., g, \tag{7.5}$$

where $\mathbf{P\text{-}GKI}_j$ is jth column vector of **P-GKI**, and mean(.) computes the average of the column vectors of **P-GKI**. The range of $Dist_{GKI}$ is [0,1], the smaller the value the more similar are the two GKIs. The similarity score between the gallery and the probe subject is computed by

$$\text{Sim}_{\text{score}} = Median(Dist_{GKI}(1), ..., Dist_{GKI}(p)). \tag{7.6}$$

The subject's identity corresponding to the gallery sequence which gives the smallest score is selected as the subject's identify of the probe sequence. Since the GKIs are size-normalised, centralised and lateral-viewed, it is not necessary to perform scale, rotation and translation transformations prior to computing Procrustes image distance, which saves computation time.

7.4 Experiments

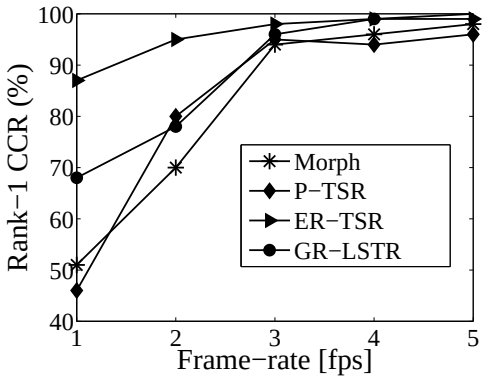
GR-LSTR is evaluated on OU-ISIR treadmill gait dataset D [76] for comparison with related methods.

7.4.1 Experiments on OU-ISIR treadmill gait dataset D

The low frame-rate gait sequences are constructed by selecting the silhouettes from the original video sequence at a specified interval, while the interval is determined by the desired low frame-rate to reduce temporal resolution. The spatial resolution of the original silhouettes and the recording time of each sequence of OU-ISIR gait dataset D are respectively 128×88 and 6 secs. Since GR-LSTR is evaluated using one-fourth of the original size of the silhouettes, the spatial resolution of the input silhouette sequence is 32×22. GR-LSTR is compared with the methods Morph [123], periodic temporal super resolution (P-TSR) [68], and example-based and reconstruction-based temporal super resolution (ER-TSR) [69].

The CCR(%) is obtained using Eq. (3.28). The CCR (%) at rank r implies that the correctly identified subject is at the top r similarity scores. Fig. 7.5 shows the rank-1 CCR(%) of Morph [123], P-TSR [68], ER-TSR [69] obtained from [69] and GR-LSTR on DB_{high} subset of OU-ISIR gait dataset D at increasing same frame-rates for probe and gallery sequences. The performance of GR-LSTR is not satisfactory at 1-2 fps as the key frames are not detected adequately based on a similarity score and not all key frames are detected due to extremely low frame rate. Note that for OU-ISIR gait dataset D, there are only six frames in a gait sequence at 1 fps. Hence, the process of forming GKI as described in Section 7.3.3 actually averages all the available image frames in this case. However, the performance is significantly improved even for the extremely low frame-rate gait sequences if the length of the gait sequences is increased to produce an adequate number of image

frames. The CCR(%) abruptly increases at 3 fps, and steadily reaches to 100% for 5 fps for DB_{high} subset. Thus GR-LSTR outperforms other methods for gait sequences captured using at least 5 fps frame-rate.



(a)

Figure 7.5: The rank-1 CCR (%) of different methods at low frame-rates on OU-ISIR gait dataset D.

Table 7.1 shows the rank-1 CCR (%) of GR-LSTR for DB_{high} and DB_{low} subsets of OU-ISIR gait dataset D. The table shows that the method achieves 100% CCR at 5 fps (for both probe and gallery) of DB_{high} and at 6 fps (for both probe and gallery) of DB_{low} subsets. Thus, GR-LSTR is suitable for the gait sequences captured at 5-6 fps. This result is significant in making a trade-off between the frame-rate of gait sequences for a desired performance and the price of CCTV camera.

Table 7.1: Rank-1 CCR (%) of GR-LSTR using different frame-rates of DB_{high} and DB_{low} subsets of OU-ISIR gait dataset D.

Frame-rate [fps]	Rank-1 CCR%	
	DB_{high}	DB_{low}
1	68	67
2	78	71
3	96	96
4	99	99
5	100	99
6	100	100

7.4.2 Computational complexity analysis

Unlike the methods Morph [123], P-TSR [68] and ER-TSR [69], GR-LSTR does not generate the intermediate frames of a gait period based on temporal interpolation. Instead, GR-LSTR combines the key frames of a gait sequence into the single image GKI which significantly reduces the computational complexity. The use of ROI for extracting key frames of a gait sequence enables GR-LSTR to process only a small subregion of an image, thus reducing time and space complexity. The weighted Krawtchouk polynomials used in this method are computed using recurrence relation to avoid numerical instability [108]. Bicubic interpolation is used in GR-LSTR for its better performance despite its higher computational complexity than the nearest neighbour and bilinear interpolation methods.

GR-LSTR uses linear combination of two DWTs for implementing DT-CWT with J scales and six directions for spatial resolution enhancement. The total arithmetic operations required for multiplications/divisions and additions/subtractions are respectively [127]

$$12WH\left[\frac{2}{3}N\left[\sum_{j=1}^{J}\left(\frac{1}{2^J}+\frac{1}{2^{2j}}\right)\right]+\sum_{j=1}^{J}\frac{1}{2^{2j}}\right] \quad \text{and} \tag{7.7}$$

$$12WH\left[\frac{2}{3}(N-1)\left[\sum_{j=1}^{J}\left(\frac{1}{2^J}+\frac{1}{2^{2j}}\right)\right]+\sum_{j=1}^{J}\frac{1}{2^{2j}}\right], \tag{7.8}$$

where N is the number of coefficients in DT-CWT, H and W are respectively height and width of the silhouette image. GR-LSTR requires one-level decomposition of the silhouettes, hence J=1. Although Gabor filter also provides directional selectivity, DT-CWT is preferred as it is computationally less expensive than the Gabor wavelets [127]. A brief explanation on Gabor wavelets are provided in Appendix B. GR-LSTR uses a simplified feature space, and hence it does not require any dimensionality reduction technique, which reduces time complexity.

7.5 Conclusion

The chapter introduces a new direction to gait recognition which is robust against low spatial and temporal resolution gait sequences often captured by a CCTV camera due to limitations in transmission bandwidth and storage space. For spatial resolution enhancement, the proposed method, i.e., GR-LSTR decomposes the low spatial resolution input silhouette using DT-CWT to produce low and high frequency subbands. The high frequency subbands and the input low resolution silhouette are interpolated, and these interpolated images are combined using IDT-CWT to generate high spatial resolution silhouette sequences.

GR-LSTR determines seven key frames from a gait sequence using ROI based silhouette matching using weighted Krawtchouk moments to produce a noise-resilient GKI which preserves significant gait characteristics in reduced storage space to address the challenge of low frame-rate gait sequences. The subjects are identified based on a similarity score obtained by Procrustes image distance between the GKIs. The experimental results based on DB_{high} and DB_{low} subsets of OU-ISIR gait dataset D demonstrate that GR-LSTR is substantially invariant to low and high gait fluctuations.

Chapter 8

Conclusions and Further Work

8.1 Conclusions

Gait recognition has emerged to be an effective behavioural biometrics for identifying human subjects unobtrusively from a distance using low resolution video sequences when physiological biometrics such as face, fingerprint and iris cannot be used. However the variation in view and clothing of a subject, and the presence of a carried item bring the main challenges to a gait recognition method. In addition, the performance of a gait recognition method is adversely affected by a wide variety of other challenging factors, e.g., variation in walking speed, change in ground surface, shadows under feet; the subject's physical and mental conditions such as leg injury, drunkenness, illness, pregnancy [25]; and presence of occluding objects in the scene, which degrades its reliability for subject identification in a realistic scenario. To address the above-mentioned challenges of gait recognition, this thesis presents five novel methods in Chapters 3 to 7 for reliable human identification in the presence of various challenging factors for visual surveillance to ensure public security. To provide a strong foundation for this PhD research, Chapter 2 provides a thorough literature review of the state-of-the-art model-free and model-based gait recognition methods. The gait recognition methods presented in this thesis are evaluated using public datasets for uniform comparison with the related methods. Hence, Chapter 2 also provides a detailed description of the various public datasets.

Chapter 3 presents a gait recognition method, i.e., STM-SPP which effectively combines spatio-temporal motion characteristics, statistical and physical parameters of a subject for its classification by analysing shape of the subject's silhouette contours using PSA and EFDs with low computational complexity. The novel contribution of this method is to introduce the application of traditional shape analysis in gait recognition based on computing landmark distances from COM-SC, and a part-based shape analysis based on anthropom-

etry using EFDs to achieve robustness against small carrying conditions with folded arms and in upright position. The use of physical parameters to validate the similar dissimilarity scores obtained by PSA enables STM-SPP to achieve robustness against limited across-day gait variation. It also provides a detailed description and pictorial illustration of different phases of a gait period with a method for their extraction from a gait sequence to facilitate speed-invariant gait recognition. Experimental results show that STM-SPP outperforms several model-free gait recognition methods.

Chapter 4 presents STS-DM, a three-phase gait recognition method that effectively combines the subject's spatio-temporal shape, and local and global dynamic motion characteristics for identification to achieve robustness against the maximum number of challenging factors compared to the state-of-the-art gait recognition methods with low computational complexity. In phase 1, PWMS of the FDs of the silhouette contours at ten phases of a gait period are used to analyse the spatio-temporal changes of the subject's shape. In phase 2, a full-body shape and motion analysis is performed by fitting ellipses to contour segments of ten phases of a gait period and using a BDHM of parameters of the ellipses as dissimilarity scores. In phase 3, DTW is used to analyse the subject's ARPoLK with implicit consideration of arm-swing over a gait period to achieve identification that is insensitive to self-occlusions, walking speed and limited clothing variation, hair style changes and shadows under feet. The match scores generated in the three phases are fused using weight-based score-level fusion. The robustness of STS-DM against missing and distorted frames is demonstrated using a public dataset. Its superiority to STM-SPP in terms of robustness against the number of challenging factors and identification rates confirms that although shape analysis of a subject has a good discriminatory power, the inclusion of dynamic motion characteristics enhances it. Similar to part-based EFD analysis of STM-SPP, STS-DM also introduces a component-based FD analysis to achieve invariance to carrying conditions, but provides a more in-depth analysis to achieve robustness against shape variation caused by all common types of small carrying conditions with folded hands, at the subject's back and in upright position.

The methods STM-SPP (presented in Chapter 3) and STS-DM (presented in Chapter 4) analyse the contour of a silhouette at specific phases of a gait period, and thus benefit from low computational complexity. However, the performances of STM-SPP and STS-DM are considerably influenced by background modeling and segmentation techniques, as these factors determine the quality of contours. Therefore, Chapter 5 analyses a subject's silhouette using SDIS and SDFS to present a region-based gait recognition method, i.e, ReG-IF. SDIS analyses silhouettes using Lp-Gf and Hp-Gf, each at five cut-off frequencies to achieve robustness against clothing variation, shape distortions at the boundary due to small carried items and segmentation errors, and the region-based SDFS analyses

silhouettes using component-based weighted Krawtchouk moments to achieve robustness against carrying conditions. The optimal cut-off frequencies of Lp-Gf and Hp-Gf are determined based on the focus value of filtered silhouettes. SDIS and SDFS respectively use a MSL and a Euclidean distance based classifiers to obtain dissimilarity scores which are combined using weighted summation based score-level fusion. The use of MSL enables overfitting avoidance by exploiting high dimensionality of SDIS. In addition to carrying condition invariance that has been addressed in STM-SPP and STS-DM, ReG-IF also takes into account of multiple carrying conditions by a subject using a component-based shape analysis based on weighted Krawtchouk moments. ReG-IF provides a sound methodology to achieve invariance to all common types of carrying conditions which considers the different discriminablity of different parts of a subject's body affected by carried items, and this level of invariance has not hitherto been addressed. ReG-IF achieves the highest rank-1 identification rate on USF HumanID gait challenge dataset.

Like most other state-of-the-art gait recognition methods, STM-SPP, STS-DM and ReG-IF are also based on the the lateral view of the subjects, and hence they cannot cope with the subjects moving freely in the different directions of a realistic scenario. Therefore, a novel two-phase view-invariant method VI-MGR is proposed in Chapter 6, which is not only robust against variation in view, but also insensitive to other two main challenges of gait recognition, i.e., variation in clothing types and presence of a carried item. In phase 1, VI-MGR uses the entropy of the limb region of a GEI to determine the matching gallery view of the probe using 2D PCA and Euclidean distance classifier. In phase 2, the probe subject is compared with the matching view of the gallery subjects using multiscale shape analysis which has not been used in gait recognition despite its high discriminatory power. In this phase, VI-MGR applies Gaussian filter to a GEI to generate a new image representation called MGI for gradually highlighting the subject's inner shape characteristics to achieve insensitiveness to boundary shape alterations due to carrying conditions and clothing variation. The scales are selected automatically to make a trade-off between the identification rate in the presence of these covariate factors and the computational complexity based on computing entropy of the filtered GEIs. A WRSL based classification is used, which is an improvement over MSL, to exploit the high dimensionality of the feature space by avoiding overlearning. Experimental analyses on public datasets demonstrate the efficacy of VI-MGR.

The methods presented in Chapters 3 to 6 are based on standard spatial resolution and standard temporal resolution of the gait sequences, and hence they fail to perform satisfactorily in the case of low spatial and low temporal resolution gait sequences captured by a CCTV camera. Thus, Chapter 7 proposes GR-LSTR, a gait recognition method which uses low spatial and low temporal resolution silhouette sequences. It uses DT-CWT for

spatial resolution enhancement of the input low resolution silhouette sequences. To achieve invariance to low frame-rate gait sequences, the method determines seven key frames from a gait sequence using ROI based silhouette matching by weighted Krawtchouk moments. The key frames are combined to form a GKI which preserves significant gait characteristics with reduced storage. The subjects are identified using a similarity score obtained by Procrustes image distance between the gallery and probe GKIs. Experimental analysis on OU-ISIR gait dataset D demonstrates the efficacy of GR-LSTR.

8.2 Future Work

The 2D novel uniscale and multiscale gait recognition methods presented in this thesis have the potential to identify a subject in a realistic scenario with a wide variety of challenging factors. This research provides a strong foundation for the following new directions that are worthwhile to be considered in the future work:

- The gait recognition methods presented int the thesis either analyse the parts of a subject that are unlikely to be affected by different types of covariate factors (e.g., STM-SPP, STS-DM), or employ feature extraction techniques that have the capability to reduce the shape distortions caused by the varying covariate factors (e.g., ReG-IF and VI-MGR). Although these gait recognition methods provide excellent identification rates compared to the state-of-the-art gait recognition methods, they are somewhat constrained by the variation of predefined covariate factors. Hence, they cannot successfully deal with the unpredictable variation of numerous covariate factors encountered in an unconstrained scenario. Also, by the above two considerations for feature extraction, a situation has not been taken into account: if a gallery and its matching probe subject are subjected to the same variation of the covariate factors, e.g., both wearing same clothes and walking on a same surface, then the attempt of avoiding the affected parts and reducing the effect of covariate factors actually decrease the identification rate. This issue will be considered in the future work as a means of identifying a subject in an unconstrained realistic scenario with unpredictable covariate factors to establish the reliability of gait recognition as a standalone biometrics.

- The view-invariant method VI-MGR (presented in Chapter 6) is robust against a wide range of views but requires the availability of most of the matching or very closely matching probe views in the gallery. Therefore, the method could be augmented by including view-invariant gait feature extraction or construction of a view transformation model in the future work so as to provide improved identification rate even in

the presence of a fewer number of gallery views to identify a subject moving freely in unconstrained directions.

The three effective range of scales, i.e., R1, R2 and R3 used in VI-MGR to maximise inter-subject discriminability in the absence of carrying conditions and clothing variation (R1), and to achieve robustness against variation in clothing (R2) and carrying conditions (R3), are chosen based on graphical analysis of difference in entropy of the silhouettes of CASIA B gait dataset. Hence, a feature set selection methodology will be investigated in the future work to automatically select the effective range of scales independent of the datasets.

- Traditional biometrics systems rely on a single biometric, and hence they are inadequate to reliably provide a high level of accurate performance for high-security real world applications. The advantage of gait recognition is its potential to identify a remote human subject unobtrusively without interfering with the subject's activity using low resolution video sequences. It is to be noted that unlike other physiological biometrics, face data can be captured without requiring the subject to physically touch the sensor, and hence it is the only physiological biometrics that has been used for remote human identification. One future work is to develop a multimodal biometrics system by fusing face and gait characteristics extracted from low frame-rate and low-resolution video sequences captured by a CCTV camera to identify a human subject for visual surveillance.

- Another future work is to investigate 3D multimodal unobtrusive biometrics in forensics (i.e., scientific techniques used in the investigation of crimes) by developing a novel system, e.g., using Microsoft Kinect sensor, to be implemented in places that are prone to terrorist activities, and hence require very high security. Kinect is a motion sensing device which consists of a RGB camera and a 3D depth sensor, and thus has the capability of using 3D motion to reconstruct 3D scene information.

- The population of elderly citizens living alone is growing dramatically, leading to increased demands for care and medical services. Hence, future work will consider to develop a decision support system capable of integrating the results of gait abnormalities detector and facial expression detector for alarms and prognosis in order to provide the elderly people with a predictive and efficient health care paradigm in order to prevent serious illness while staying alone in their residences. Thus, the applications of gait recognition research can be extended to the practical needs of elderly people to technologically assist them in independent living.

Appendix A

Histogram matching techniques

Histogram matching refers to the process of comparing two histograms for measuring similarity. The different techniques for matching two histograms of same size, i.e., H_1 and H_2, are given below [93].

- Correlation method:

$$d_{correl}(H_1, H_2) = \frac{\sum_i H'_1(i).H'_2(i)}{\sqrt{\sum_i H'^2_1(i).H'^2_2(i)}}, \tag{A.1}$$

  where $H'_k(i) = H_k(i) - (1/N)(\sum_j H_k(j))$ and N is the number of histogram bins. The similarity score for perfect matching is 1 and total mismatching is -1. A high similarity score denotes a good match and vice versa.

- Chi-square method:

$$d_{chi-square}(H_1, H_2) = \sum_i \frac{(H_1(i) - H_2(i))^2}{H_1(i) + H_2(i)}. \tag{A.2}$$

  The similarity score for perfect matching is 0 and total mismatching is unbounded, i.e., a very large value. A low similarity score denotes a good match and vice versa.

- Intersection method:

$$d_{intersection}(H_1, H_2) = \sum_i min(H_1(i), H_2(i)). \tag{A.3}$$

  The similarity score for perfect matching is 1 and total mismatching is 0. A high similarity score denotes a good match and vice versa.

- Bhattacharyya distance method:

$$d_{Bhattacharyya}(H_1, H_2) = \sqrt{1 - \sum_i \frac{\sqrt{H_1(i).H_2(i)}}{\sqrt{\sum_i H_1(i). \sum_i H_2(i)}}}. \tag{A.4}$$

  The similarity score for perfect matching is 0 and total mismatching is 1. A low similarity score denotes a good match and vice versa.

Appendix B

Gabor filter

The Gabor wavelets, whose kernels resemble the receptive field profiles of the mammalian cortical simple cells, are extensively used in image processing applications, e.g., detection of edges, corners and blobs [128], texture analysis, face recognition [129] and gait recognition [50]. The 2D Gabor functions introduced by Daugman [130] exhibit the desirable properties of spatial locality, spatial frequency and orientation selectivity. The Gabor wavelet (kernel or filter), which is the product of an elliptical Gaussian envelop and a complex plane wave, is defined as [129; 130]

$$\psi_{\mu,\nu}(z) = \frac{||k_{\mu,\nu}||^2}{\sigma^2} e^{(-||k_{\mu,\nu}||^2||z||^2/2\sigma^2)}[e^{ik_{\mu,\nu}z} - e^{-\sigma^2/2}], \tag{B.1}$$

where μ and ν respectively denote the orientation and scale of the Gabor kernels, z = (x,y) is a variable in the spatial domain, ||.|| is the norm operator, and the wave vector is

$$k_{\mu,\nu} = k_\nu e^{i\phi_\mu}, \tag{B.2}$$

where $k_\nu = k_{\max}/f^\nu$, with $k_{\max}$ being the maximum frequency, $f = 1.2$ the spacing factor between kernels in the frequency domain, and $\phi_\mu = \pi\mu/8$. Fig. B.1 shows the real part of Gabor wavelets at five scales, i.e., $\nu = 0, 1, 2, 3, 4$, and eight directions, i.e., $\mu = 0, 1, 2, 3, 4, 5, 6, 7$, where $k_{\max} = 0.35$. Fig. B.2 shows the magnitude of Gabor wavelets at the same scales.

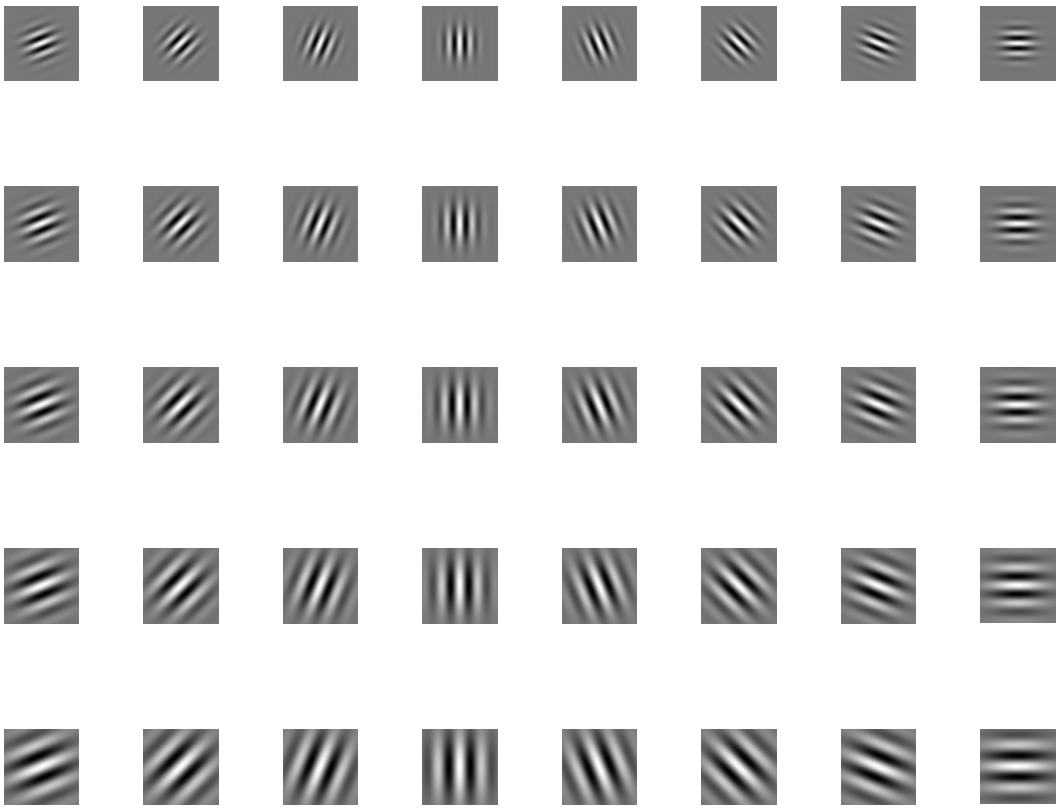


Figure B.1: The real part of Gabor wavelets at five scales and eight directions.

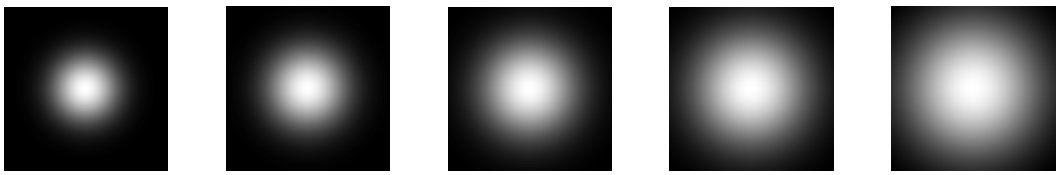


Figure B.2: The magnitude of Gabor wavelets at five scales.

Bibliography

[1] S. K. Modi, Biometrics in identity management: concepts to applications, Artech house, Boston, London, 2011.

[2] M. Goffredo, I. Bouchrika, J. N. Carter, and M. S. Nixon, "Self-calibrating view-invariant gait biometrics," IEEE Trans. Syst., Man, Cybern. B, vol. 40, no. 4, pp. 997-1008, 2010.

[3] B. Bhanu and J. Han, Human recognition at a distance in video, Springer Science+Business Media, London, 2011.

[4] The American Heritage Dictionary of the English Language, Fourth ed., Houghton Mifflin Company, Boston, USA, 2000.

[5] W. Hu, T. Tan, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviours," IEEE Trans. Syst., Man, Cybern. C, vol. 34, no. 3, pp. 334-352, 2004.

[6] M. S. Nixon, I. Bouchrika, B. Arbab-Zavar, and J. N. Carter, "On use of biometrics in forensics: gait and ear," in: Proc. Eur. Signal Process. Conf., Aalborg, Denmark, 2010, pp. 23-27.

[7] S. D. Choudhury and T. Tjahjadi, "Silhouette-based gait recognition using Procrustes shape analysis and elliptic Fourier descriptors," Pattern Recognit., vol. 45, pp. 3414-3426, 2012.

[8] S. D. Choudhury and T. Tjahjadi, "Gait recognition based on shape and motion analysis of silhouette contours," Comput. Vis. Image Und., vol. 117, no. 12, pp. 1770-1785, 2013.

[9] S. D. Choudhury and T. Tjahjadi, "Robust view-invariant multiscale gait recognition," Pattern Recognit., under review, 2013.

[10] S. D. Choudhury, Y. Guan, C.-T. Li, “Gait Recognition using Low Spatial and Temporal Resolution Videos,” in: Proc. Int. Workshop on Biometrics and Forensics, under review, 2014.

[11] G. Johansson, “Visual motion perception,” Scientific Am., vol. 232, pp. 75-88, 1975.

[12] J. E. Cutting and L. T. Kozlowski, “Recognizing friends by their walk: gait perception without familiarity cues,” Bull. Psychonomic Soc., vol. 9, no. 5, pp. 353-356, 1977.

[13] S. L. Dockstader, M. J. Berg, and A. M. Tekalp, “Stochastic kinematic modeling and feature extraction for gait analysis,” IEEE Trans. Image Process., vol. 12, no. 8, pp. 962-976, 2003.

[14] D. K. Wagg and M. S. Nixon, “On automated model-based extraction and analysis of gait,” in: Proc. IEEE Int. Conf. Autom. Face Gesture Recog., Seoul, Korea, 2004, pp. 11-16.

[15] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, “A full-body layered deformable model for automatic model-based gait recognition,” EURASIP J. Adv. Signal Process., vol. 2008, pp. 1-13.

[16] F. Tafazzoli and R. Safabakhsh, “Model-based human gait recognition using leg and arm movements,” Eng. Appl. Artif. Intell., vol. 23, pp. 1237-1246, 2010.

[17] Z. Xue, D. Ming, W. Song, B. Wan, and S. Jin, “Infrared gait recognition based on wavelet transform and support vector machine,” Pattern Recognit., vol. 43, pp. 2904-2910, 2010.

[18] J. Gu, X. Ding, S. Wang and Y. Wu, “Action and gait recognition from recovered 3-D human joints,” IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 40, no. 4, pp. 1021-1033, 2010.

[19] R. Collins, R. Gross, and J. Shi, “Silhouette-based human identification from body shape and gait,” in: Proc. IEEE Int. Conf. Autom. Face Gesture Recog., Washington, DC, USA, 2002, pp. 351-356.

[20] C. BenAbdelkader, R. Culter, and L. Davis, “Stride and cadence as a biometric in automatic person identification and verification,” in: Proc. IEEE Int. Conf. Autom. Face Gesture Recog., Washington, DC, USA, 2002, pp. 372 - 377.

[21] C. BenAbdelkader, R. G. Cutler, and L. S. Davis, “Gait recognition using image self-similarity,” EURASIP J. Applied Signal Process., no. 4, pp. 572-585, 2004.

[22] S. Sarkar, P. J. Philips, Z. Liu, I. Vega, P. Grother, and K. W. Bowyer, "The HumanID gait challenge problem: data sets, performance, and analysis," IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 2, pp. 162-177, 2006.

[23] A. Veraraghavan, A. K. Roy-Chowdhury, and R. Chellappa, "Matching shape sequences in video with applications in human movement analysis," IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 12, pp. 1896-1909, 2005.

[24] Z. Liu and S. Sarkar, "Simplest representation yet for gait recognition: averaged silhouette," in: Proc. Int. Conf. Pattern Recog., vol. 4, England, UK, 2004, pp. 211-214.

[25] J. Han and B. Bhanu, "Individual recognition using gait energy image," IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, no. 2, pp. 316-322, 2006.

[26] Z. Liu and S. Sarkar, "Improved gait recognition by gait dynamics normalization," IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, no. 6, pp. 863-876, 2006.

[27] C. Wang, J. Zhang, J. Pu, X. Yuan, and L. Wang, "Chrono-Gait Image: A novel temporal template for gait recognition," in: Proc. Eur. Conf. Comput. Vis., Part 1, LNCS 6311, Heraklion, Greece, 2010, pp. 257-270.

[28] C. Wang, J. Zhang, L. Wang, J. Pu, and X. Yuan, "Human identification using temporal information preserving gait template," IEEE Trans. Pattern Anal. Mach. Intell., vol.34, no.11, pp.2164-2176, 2012.

[29] L. Wang, T. Tan, H. Ning, and W. Hu, "Silhouette analysis-based gait recognition for human identification," IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 12, pp. 1505-1518, 2003.

[30] D. Cunado, M. Nixon, and J. Carter "Using gait as a biometric, via phase-weighted magnitude spectra," in: Proc. Int. Conf. Audio- and Video-Based Biometric Person Authentication, Crans-Montana, Switzerland, 1997, pp. 95-102.

[31] D. Cunado, J. M. Nash, M. S. Nixon, and J. N. Carter, "Gait extraction and description by evidence-gathering," in: Proc. Int. Conf. Audio- and Video-Based Biometric Person Authentication, Washington, DC, Mar. 22-23, 1999, pp. 43-48.

[32] D. Cunado, M. Nixon, and J. Carter, "Automatic extraction and description of human gait models for recognition purposes," Comput. Vis. Image Und., vol. 90, no. 1, pp. 1-41, 2003.

[33] R. Tanawongsuwan and A. Bobick, “Gait recognition from timenormalized joint-angle trajectories in the walking plane,” in: Proc. IEEE Int. Conf. Comput. Vis. Pattern Recog., 2001, Kauai, HI, USA, vol. 2, pp. 11-726-11-731.

[34] I. Bouchrika and M. S. Nixon, “Model-based feature extraction for gait analysis and recognition,” in: Proc. Int. Conf. Comput. Vis./Comput. Graph. Collaboration Techn. Appl., 2007, pp. 150-160.

[35] D. Kim, D. Kim, and J. Paik, “Gait recognition using active shape model and motion prediction,” IET Comput. Vis., vol. 4, no. 1, pp. 25-36, 2010.

[36] A. Sundaresan, A. Roy-Chowdhury, and R. Chellappa, “A hidden Markov model based framework for recognition of humans from gait sequences,” in: Proc. Int. Conf. Image Process., Barcelona, Spain, 2003, pp. 143-150.

[37] A. Kale, A. Sundaresan, A. N. Rajagopalan, N. P. Cuntoor, A. K. Roy Chowdhury, V. Kruger, and R. Chellappa, “Identification of humans using gait,” IEEE Trans. Image Process., vol. 13, no. 9, pp. 1163-1173, 2004.

[38] L. Wang, H. Ning, and T. Tan, “Fusion of static and dynamic body biometrics for gait recognition,” IEEE Trans. Circuits Syst. Video Technol., vol. 14, no. 2, pp. 149-158, 2004.

[39] G. Ariyanto and M.S. Nixon, “Model-based 3D gait biometrics,” in: Proc. Int. Joint Conf. Biometrics, Washington, D.C., United States, pp. 1-7, 2011.

[40] S. Lee, Y. Liu, and R. Collins, “Shape variation-based frieze pattern for robust gait recognition,” in: Proc. IEEE Int. Conf. Comput. Vis. Pattern Recog., Minneapolis, MN, 2007, pp. 1-8.

[41] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li, “Gait recognition across various walking speeds using higher order shape configuration based on a differential composition model,” IEEE Trans. Syst., Man, Cybern. B, Cybern., doi 10.1109/TSMCB.2012.2197823, 2012.

[42] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li, “Speed-invariant gait recognition based on Procrustes shape analysis using higher-order shape configuration,” in: Proc. IEEE Int. Conf. Image Process., Sydney, Australia, 2011, pp. 545-548.

[43] A. Tsuji, Y. Makihara, and Y. Yagi, “Silhouette transformation based on walking speed for gait identification,” in: Proc. IEEE Conf. Comput. Vis. Pattern Recog., San Francisco, CA, USA, 2010, pp. 717-722.

[44] M. R. Aqmar, K. Shinoda, and S. Furui, “Robust gait recognition against speed variation,” in: Proc. Int. Conf. Pattern Recog., Tokyo, Japan, 2010, pp. 2190-2193.

[45] Y. Ran, Q. Zheng, R. Chellappa, and T. M. Strat, “Applications of a simple characterization of human gait in surveillance,” IEEE Trans. Syst., Man, Cybern. B, Cybern., vol 40, no. 4, pp. 1009-1020, 2010.

[46] L. Lee and W. E. L. Grimson, “Gait analysis for recognition and classification,” in: Proc. IEEE 5th Int. Conf. Autom. Face Gesture Recog., Washington, DC, 2002, pp. 148-155.

[47] D. Xu, Y. Huang, Z. Zeng, and X. Xu, “Human gait recognition using patch distribution feature and locality-constrained group sparse representation,” IEEE Trans. Image Process., vol. 21, no. 1, pp. 316-326, 2012.

[48] K. Bashir, T. Xiang, and S. Gong, “Gait recognition without subject cooperation,” Pattern Recognit. Lett., vol. 31, pp. 2052-2060, 2010.

[49] T. H. W. Lam, K. H. Cheung, and J. N. K. Liu, “Gait flow image: A silhouette-based gait representation for human identification,” Pattern Recognit., vol. 44, pp. 973-987, 2011.

[50] D. Tau, X. Li, X. Wu, and S. J. Maybank, “General tensor discriminant analysis and Gabor features for gait recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 10, pp. 1700-1715, 2007.

[51] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, “MPCA: Multilinear principal component analysis of tensor objects,” IEEE Trans. Neural Netw., vol. 19, no. 1, pp. 18-39, 2008.

[52] M. A. Hossain, Y. Makihara, J. Wang, and Y. Yagi, “Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control,” Pattern Recognit., vol. 43, pp. 2281-2291, 2010.

[53] J. Han, B. Bhanu, and A. K. Roy-Chowdhury, “A study on view-insensitive gait recognition,” in: Proc. Int. Conf. Image Process., pp. 297-300, 2005.

[54] W. Kusakunniran, Q. Wu, J. Zhang, Y. Ma, and H. Li, “A New View-Invariant Feature for Cross-View Gait Recognition,” IEEE Trans. Inf. Forensics Security, doi:10.1109/TIFS.2013.2252342.

[55] F. Jean, R. Bergevin, and A.B. Albu, “Computing and evaluating view-normalized body part trajectories,” Image Vision Comput., vol. 27, no. 9, pp. 1272 -1284, 2009.

[56] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi, "Gait recognition using a view transformation model in the frequency domain," in: Proc. 9th Eur. Conf. Comput. Vis., Graz, Austria, 2006, pp. 151-163.

[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: Proc. IEEE Conf. Comput. Vis. Pattern Recog., United States, 2010, pp. 974 - 981.

[58] K. Bashir, T. Xiang, and S. Gong, "Cross-view gait recognition using correlation strength," in: Proc. 21st British Machine Vis. Conf. (BMVC), Aberystwyth, UK, 2010, pp. 109.1-109.11.

[59] N. Liu, J. Lu, and Y. Tan, "Joint subspace learning for view-invariant gait recognition," IEEE Signal Process. Lett., vol. 18, no.7, pp. 431- 434, 2011.

[60] G. Zhao, G. Liu, H. Li, and M. Pietikainen, "3D gait recognition using multiple cameras," in: Proc. IEEE Int. Conf. Autom. Face Gesture Recog., UK, 2006, pp. 529-534.

[61] W. Kusakunniran, Q. Wu, H. Li, and J. Zhang, "Multiple views gait recognition using view transformation model based on optimized gait energy image, in: Proc. IEEE Int. Conf. Computer Vision Workshops (ICCV Workshops), Kyoto, Japan, 2009, pp. 1058-1064.

[62] J. Zhang, J. Pu, C. Chen, and R. Fleischer, "Low-resolution gait recognition," IEEE Trans. Syst., Man, Cybern. B, vol. 40, no. 4, pp. 986-996, 2010.

[63] I. Venkat and P. D. Wilde, "Robust gait recognition by learning and exploiting sub-gait characteristics," Int. J. Comput. Vision, vol. 91, pp. 7-23, 2011.

[64] 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: Proc. IEEE 2nd Int. Conf. Biometrics: Theory, Applications and Systems, United States, 2008, pp. 1-6.

[65] D. Ioannidis, D. Tzovaras, I. G. Damousis, S. Argyropoulos, and K. Moustakas, "Gait recognition using compact feature extraction transforms and depth information," IEEE Trans. Inf. Forensics Security, vol. 2, no. 3, pp. 623-630, 2007.

[66] A. Mori, Y. Makihara, and Y. Yagi, "Gait recognition using period-based phase synchronization for low frame-rate videos," in: Proc. Int. Conf. Pattern Recog., Istanbul, Turkey, 2010, pp. 2194-2197.

[67] Y. Makihara, A. Mori, and Y. Yagi, "Temporal super resolution from a single quasi-periodic image sequence based on phase registration," in: Proc. 10th Asian Conf. Computer Vis., Queenstown, New Zealand, 2010, pp. 107-120.

[68] N. Akae, Y. Makihara, and Y. Yagi, "Gait recognition using periodic temporal super resolution for low frame-rate videos," in: Proc. Int. Joint Conf. Biometrics, Washington, D.C., pp. 1-7, 2011.

[69] N. Akae, A. Mansur, Y. Makihara, and Y. Yagi, "Video from nearly still: an application to low frame-rate gait recognition," in: Proc. IEEE Conf. Comput. Vis. Pattern Recog., Providence, RI, 2012, pp. 1537-1543.

[70] M. S. Nixon, T. N. Tan, and R. Chellappa, Human identification based on gait, Springer Science+Business Media, New York, USA, 2006.

[71] J. J. Little and J. E. Boyd, "Recognizing people by their gait: the shape of motion," Videre: J. Comput. Vis. Research, vol. 1, no. 2, pp. 1-32, 1998.

[72] R. Gross and J. Shi, "The CMU Motion of Body (MoBo) Database," Tech. Report CMU-RI-TR-01-18, Robotics Institute, CMU, 2001.

[73] S. Zheng, J. Zhang, K. Huang, R. He, and T. Tan, "Robust view transformation model for gait recognition," in: Proc. IEEE 18th Int. Conf. Image Process., 2011, pp. 2073-2076.

[74] D. Tan, K. Huang, S. Yu, and T. Tan, "Efficient night gait recognition based on template matching," in: Proc. 18th Int. Conf. Pattern Recog. (ICPR06), Hong Kong, China, 2006, pp. 1000-1003.

[75] S. Zheng, K. Huang, T. Tan, and D. Tao, "A Cascade Fusion Scheme for Gait and Cumulative Foot Pressure Image Recognition," Pattern Recognit., vol. 45, pp. 3603-3610, 2012.

[76] Y. Makihara, H. Mannami, A. Tsuji, M. A. Hossain, K. Sugiura, A. Mori, and Y. Yagi, "The OU-ISIR gait database comprising the treadmill dataset," IPSJ Trans. Comput. Vis. Appl., Tech. Note, vol. 4, 2012, pp. 53-62.

[77] Y. Makihara, H. Mannami, and Y. Yagi, "Gait analysis of gender and age using a large-scale multi-view gait database," in: Proc. 10th Asian Conf. Computer Vision (ACCV), Queenstown, New Zealand, 2010, pp 440-451.

[78] H. Iwama, M. Okumura, Y. Makihara, and Y. Yagi, "The OU-ISIR Gait database comprising the large population dataset and performance evaluation of gait recognition," IEEE Trans. Inf. Forensics Security, vol. 7, no. 5, pp. 1511-1521, 2012.

[79] O. Barnich, and M. V. Droogenbroeck, "Frontal-view gait recognition by intra- and inter-frame rectangle size distribution," Pattern Recognit. Lett., vol. 30, no. 9, pp. 893-901, 2009.

[80] L. Havasi, Z. Szlavik, and T. Sziranyi, "Higher order symmetry for non-linear classification of human walk detection," Pattern Recognit. Lett., vol. 27, pp. 822-829, 2006.

[81] Y. Ran, Q. Zheng, R. Chellappa, and T. M. Stat, "Applications of a simple characterization of human gait in surveillance," IEEE Trans. Syst., Man, Cybern. B, vol. 40, no. 4, 1009-1019, 2010.

[82] N. V. Boulgouris, K. N. Plataniotis, and D. Hatzinakos, "Gait recognition using linear time normalization," Pattern Recognit., vol. 39, pp. 969-979, 2006.

[83] J. D. Shutler, M. S. Nixon, and C. J. Harris, "Statistical gait description via temporal moments," in: Proc. IEEE Southwest Symp. Image Analysis and Interpretation, Austin, USA, pp. 291-295, 2000.

[84] J. D. Shutler and M. S. Nixon, "Zernike velocity moments for sequence-based description of moving features," Image Vis. Comput., vol. 24, pp. 343-356, 2006.

[85] L. Wang, T. Tan, W. Hu, and H. Ning, "Automatic gait recognition based on statistical shape analysis," IEEE Trans. Image Process., vol. 12, no. 9, pp. 1-13, 2003.

[86] I. L. Dryden and K. V. Mardia, Statistical shape analysis, 2nd ed. John Wiley & Sons Ltd., England, 1998.

[87] F. P. Kuhl and C. R. Giardina, "Elliptic Fourier features of a closed contour," Comput. Graph. Image Process., vol. 18, no. 3, pp. 236-258, 1982.

[88] M. K. Hu, "Pattern recognition by moment invariants," in: Proc. IEEE, vol. 49, no. 9, 1961, pp. 1428-1961.

[89] D. Tolliver and R. T. Collins, "Gait shape estimation for identification," in: Proc. Int. Conf. Audio- Video-Based Biometric Person Authentication (AVBPA), Guildford, UK, 2003, pp. 734-742.

[90] Q. Meng, B. Li, and H. Holstein, "Recognition of human periodic movements from unstructured information using a motion-based frequency domain approach," Image Vision Comput., vol. 24, pp. 795-809, 2006.

[91] S. D. Mowbray and M. S. Nixon, "Automatic gait recognition via Fourier descriptors of deformable objects", in: Proc. Audio- and Video-based Biometric Person Authentication (AVBPA), Guildford, UK, 2003, pp. 556-573.

[92] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Real-time foreground-background segmentation using codebook model," Real-Time Imaging, vol. 11, pp. 172-185, 2005.

[93] G. Bradski and A. Kaehler, Learning OpenCV Computer Vision with the OpenCV Library, O'Reilly Media, Sebastopol, 2008.

[94] L. Jianzhuang, L. Wenqing, and T. Yupeng, "Automatic thresholding of gray-level pictures using two-dimension Otsu method," in: Proc. Int. Conf. Circuits Systems, China, 1991.

[95] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," IEEE Trans. Syst., Man, Cybern., vol. 9, no. 1, pp. 62-66, 1979.

[96] M. Nixon and A. S. Aguado, Feature extraction and image processing, 2nd ed. Elsevier, London, 2006.

[97] D.A. Winter, Biomechanics and motor control of human movement, 3rd ed. John Wiley & Sons, New Jersey, 2004.

[98] R. C. Gonzalez and R. E. Woods, Digital image processing, 2nd ed. Addison-Wesley, USA, 2002.

[99] M. Müller, Information retrieval for music and motion, Springer, New York, 2007.

[100] J. M. L. Minor and M. Schmittbuhl, " Importance of elliptic Fourier methods for morphometry of complex outlines: application to the distal human femur," Tech. Report, Surgical and Radiologic Anatomy, vol. 21, no. 6, pp. 387-391, 1999.

[101] O. Melnik, Y. Vardi, and C. Zhang, "Mixed group ranks: preference and confidence in classifier combination," IEEE Trans. Pattern Anal. Mach. Intell., vol. 26, no. 8, pp. 973-981, 2004.

[102] K. Yu, X. Jiang, and H. Bunke, "Lipreading: A classifier combination approach," Pattern Recognit. Lett., vol. 18, pp. 1421-1426, 1997.

[103] B. Achermann and H. Bunke, “Combination of classifiers on the decision level for face recognition,” Tech. Report, IAM-96-002, University of Bern, 1996.

[104] H. Carr, J. Snoeyink, and M. van de Panne, “Progressive topological simplification using contour trees and local spatial measures,” in: Proc. 15th Western Computer Graphics Symposium, Big White, British Columbia, 2004.

[105] D. Douglas and T. Peucker, “Algorithms for the reduction of the number of points required to represent a digitized line or its caricature,” Cartographica: Int. J. Geographic Info. Geovisualization, vol.10, no. 2, pp.112-122, 1973.

[106] J. Phillips, H. Moon, S. Rizvi, and P. Rause, “The FERET evaluation methodology for face recognition algorithms,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 10, pp. 1090-1104, 2000.

[107] Z. Liu and S. Sarkar, “Effect of silhouette quality on hard problems in gait recognition,” IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 35, no. 2, pp. 170-183, 2005.

[108] P. Yap, R. Paramesran, and S. Ong, “Image Analysis by Krawtchouk Moments,” IEEE Trans. Image Process., vol. 12, no. 11, pp. 1367-1377, 2003.

[109] S. Yu, D. Tan, and T. Tan, “A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition,” in: Proc. Int. Conf. Pattern Recog., Hong Kong, China, 2006, pp. 441- 444.

[110] J. W. Cooley and O. W. Tukey, “An algorithm for the machine calculation of complex Fourier series,” Math. Comput., vol. 19, no. 90, pp. 297-301, 1965.

[111] D. Xu, S. Yan, D. Tao, S. Lin, and H. Zhang, “Marginal fisher analysis and its variants for human gait recognition and content-based image retrieval,” IEEE Trans. Image Process., vol. 16, no. 11, pp. 2811-2821, 2007.

[112] S. Rahati, R. Moravejian, and F. M. Kazemi, “Gait Recognition Using Wavelet Transform,” in: Pooc. 5th Int. Conf. Inf. Tech.: New Generations, Las Vegas, NV, pp. 932-936, 2008.

[113] C. Direkoglu and M.S. Nixon, “Shape classification via image-based multiscale description,” Pattern Recognit., vol. 44, pp. 2134-2146, 2011.

[114] J. M. Tenenbaum, Accommodation in computer vision, Ph.D. Thesis, Stanford University, 1970.

[115] S. K. Nayar and Y. Nakagawa, “Shape from Focus”, IEEE Trans. Pattern Anal. Mach. Intell., vol. 16, pp. 824-831, 1994.

[116] G. Yang and B.J. Nelson, “Wavelet-based autofocusing and unsupervised segmentation of microscopic images,” in: Proc. IEEE/RSJ Int. Conf. Intelligent Robots Systems, Las Vegas, Nevada, USA, 2003, pp. 2143-2148.

[117] T. Lindeberg, “Feature Detection with Automatic Scale Selection,” Int. J. Comput. Vis., vol. 30, no. 2, pp. 79-116, 1998.

[118] J. Yang, D. Zhang, A. F. Frangi, and J. Yang, “Two-dimensional PCA: a new approach to appearance-based face representation and recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 26, no. 1, pp.131-137, 2004.

[119] N. Boulgouris and Z. Chi, “Gait recognition using radon transform and linear discriminant analysis,” IEEE Trans. Image Process., vol. 16, no. 3, pp. 731-740, 2007.

[120] Z. Ling, C. Zhao, Q. Pan, Y. Wang, and Y. Cheng, “Analyzing human movements from silhouettes via Fourier descriptor,” in: Proc. IEEE Int. Conf. Automation and Logistics (ICAL), Jinan, China, 2007, pp. 231-236.

[121] X. Wang and X. Tang, “Using random subspace to combine multiple features for face recognition,” in: Proc. Int. Conf. Autom. Face Gesture recog., Seoul, Korea, 2004, pp. 284 -289.

[122] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, Handbook of fingerprint recognition, 2nd ed., Springer Science+Business Media, 2009.

[123] M. S. Al-Huseiny, S. Mahmoodi, and M.S. Nixon, “Gait learning-based regenerative model: a level set approach,” in: Proc. Int. Conf. Pattern Recog., Istanbul, 2010, pp. 2644-2647.

[124] H. Demirel and G. Anbarjafari, “Satellite image resolution enhancement using complex wavelet transform,” IEEE Geosci. Remote Sens. Lett., vol. 7, no. 1, pp. 123-126, 2010.

[125] I.W. Selesnick, R.G. Baraniuk, and N.G. Kingsbury, “The dual-tree complex wavelet transform,” IEEE Signal Process. Mag., vol. 22, pp. 123-151, 2005.

[126] T. Celik and T. Tjahjadi, “Image resolution enhancement using dual-tree complex wavelet transform,” IEEE Geosci. Remote Sens. Lett., vol. 7, no. 3, pp. 554-557, 2010.

[127] T. Celik, H. Özkaramanl, and H. Demirel, "Facial feature extraction using complex dual-tree wavelet transform," Comput. Vis. Image Und., vol. 111, pp. 229-246, 2008.

[128] A. Quddus and M. M. Fahmy, "Corner detection using Gabor-type filtering," in: Proc. IEEE Int. Symp. Circuits Systems, vol. 4, pp. 150-153, 1998.

[129] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," IEEE Trans. Image Process., vol. 11, no. 4, pp. 467-476, 2002.

[130] J. G. Daugman, "Two-dimensional spectral analysis of cortical receptive field profiles," Vis. Res., vol. 20, pp. 847-856, 1980.