Doctoral Dissertation (Shinshu University)

Covariate-invariant gait analysis for human identification

March 2018

YEOH TZE WEI

SUPERVISORS:
Kiyoshi Tanaka
Hernán Aguirre
Youhei Akimoto

Dedicated to my family.

ABSTRACT

Gait recognition has been an active research topic in recent years. It has attracted increasing attention as it is an unobtrusive biometric that can be captured from distance and without the need of interaction with the subject. As a result, gait has become one of the latest promising biometric. However, the practical deployment in real-time surveillance and security access control applications are still considered a challenging task because of the presence of covariate factors such as different clothing types, carrying conditions, walking surfaces, camera viewpoint and walking speed with respect to the subject. This is mainly because of the variations will alter an individual's appearance and increase the difficulty of gait recognition.

The goal of this thesis is addressing the problem of covariate factors that affect gait recognition performance. In particular, we aim to extract the gait characteristics of the human for recognition under variable covariate conditions. For this purpose, an efficient covariate factor mitigation framework is proposed. First, we propose a feature extraction method for robust gait recognition against different covariate factors. We then adapt feature selection methods to select a subset of the most relevant gait features which further reduces the effect of covariate factors of gait recognition, and hence improves the overall performance. Furthermore, we are also interested in exploring the use of deep learning techniques to extract gait signatures for recognition. Therefore, the contributions of this thesis are divided in three parts.

In the first part, we focus on modelling gait dynamics and eliminating the effect of the covariates by designing a simple yet efficient model-based gait recognition system. The proposed approach can handle occluded silhouette either from self-occluded or those by objects which usually present significant challenges for other conventional approaches. Hence, it shows robust performance against various covariate factors in SOTON covariate database and also outperforms other state-of-the-art algorithms.

However, the recognition task can become complicated due to the existence of challenging covariate factors. For instance, clothing type has been demonstrated to be the most challenging one. This is because it can occlude a significant amount of gait features and combined with the location of occlusion, which may differ for different covariate factors, relevant gait features may become irrelevant when the covariate factor changes, and exploiting them can hinder the recognition performance. Therefore, feature selection has become an important step to make gait analysis more manageable and to extract useful information for the classification task. In the second part of the

thesis, we present an empirical approach to evaluate the degree of consistency among the performance of different selection algorithms in the context of gait identification under the effect of various covariate factors. In addition, we systematically compare the feature subsets selected by six popular selection methods and the computational cost of different selection approaches. We present a performance measure method via statistical and mixed-model analysis to examine the feature selection across different classifiers and covariates. Then, we investigate the effectiveness of feature selection approaches through extensive experiments on two well-known benchmark databases: the SOTON covariate database and the CASIA-B dataset. Our experimental results show that feature selection approaches significantly improve the overall performance of gait recognition, also obtains results that are comparable to other state-of-the-art recognition approaches.

In recent years, deep learning has gained significant attention from the computer vision community. This is because deep learning models are capable of learning multiple layers of feature hierarchies by constructing high-level features from low-level features. Hence, they are more generic since the feature construction is fully automated. Specifically, many recent studies have shown promising results for applying deep learning approaches to a variety of applications (e.g. image classification, text classification, natural language processing, scene labeling, etc.). Therefore, in the last part of the thesis, we propose the use of convolutional neural network (CNN) for gait recognition task. Inspired by the great success of CNNs in image classification tasks, we feed in the most widely used gait representation, that is, the gait energy image (GEI), as the input to the CNN. More specifically, the network structure contains six layers, with three convolutional layers, two fully connected layers and followed by a softmax layer. We conduct our experiments on the CASIA-B dataset and OU-ISIR Treadmill B dataset which includes the largest variations of view angles and clothing of combinations types. The experiment results show that our method can achieve far better performance compared to hand-crafted features in conventional methods with minimal knowledge of the problem required. For clothing variation, we also employ a stacked progressive auto-encoders model to extract clothing-invariant gait feature for recognition. The proposed method is evaluated on the same clothing dataset, the results show that the proposed method can yield good recognition accuracy even when the combination of clothing types get more complicated.

The thesis concludes with a summary of our contributions. Last but not least, we also suggest promising directions for future work related to gait recognition system.

PUBLICATIONS AND AWARDS

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

Journal Papers

1. **T.W. Yeoh**, F. Daolio, H. Aguirre and K. Tanaka, "On the Effectiveness of Feature Selection Methods for Gait Classification under Different Covariate Factors," in *Applied Soft Computing*. Vol.61, pp. 42-57, 2017.

2. **T.W. Yeoh**, Y. Akimoto, H. Aguirre and K. Tanaka, "A Gait-Based Human Identification Method Under Various Covariate Factors (Special Issue on Computer Vision and Applications)," in *IIEEJ Transactions on Image Electronics and Visual Computing 3(2)*. Vol.3, No.2, pp. 193-205, 2015.

International Conference Papers

1. **T.W. Yeoh**, H. Aguirre and K. Tanaka, "Stacked Progressive Auto-encoders for Clothing-invariant Gait Recognition," in *International Conference on Computer Analysis of Images and Patterns (CAIP 2017)*. Ystad, Sweden, pp. 151-161, 2017.

2. **T.W. Yeoh**, H. Aguirre and K. Tanaka, "Multi-view Gait Recognition based on Convolutional Neural Network," in *Proc. IIEEJ Image Electronics and Visual Computing Workshop (IEVC2017)*. Da Nang, Vietnam, CD-ROM, Feb. 2017.

3. **T.W. Yeoh**, H. Aguirre and K. Tanaka, "Clothing-invariant Gait Recognition using Convolutional Neural Network," in *IEEE International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS 2016)*. Phuket, Thailand, pp. 1-5, Oct. 2016.

4. **T.W. Yeoh**, S. Zapotecas Martinez, Y. Akimoto, H. Aguirre and K. Tanaka, "Feature Selection in Gait Classification Using Geometric PSO Assisted by SVM," in *International Conference on Computer Analysis of Images and Patterns (CAIP 2015)*. Valleta, Malta, pp. 566-578, 2015.

Awards

1. Excellent paper award - IIEEJ, Image Electronics and Visual Computing Workshop, 2017

ACKNOWLEDGEMENTS

First of all, I would like to express my most sincere gratitude to my supervisor Prof. Kiyoshi Tanaka and co-supervisor Assoc. Prof. Hernan Aguirre and Asst. Prof. Youhei Akimoto for their invaluable guidance and assistance throughout the course of my Ph.D studies. They have always been encouraging, helpful and providing constructive advice whenever needed. This study would never have been completed without their support and dedicated involvement in every research stage.

Next, I want to thank Fabio Daolio for his help and guidance during his short-term stay at Shinshu University. My sincere thanks also goes to my laboratory group members. They offered insightful comments, suggestions and discussions that undoubtedly helped to facilitate my research progress.

I also want to take this opportunity to thanks to Ms. Keiko Nishizawa and Ms. Watanabe Masako for all their help for supporting me with all the administration and application works. Thanks to the Japanese language teachers of Matsumoto and Nagano campus for your dedication, patience and effort in teaching me Japanese language.

I am grateful for the MEXT Scholarship, received from Japan Government at Shinshu University for their financial support in funding my studies for these three wonderful years.

Last but not least, I wish to thank my family members for their love, encouragement and support.

CONTENTS

I INTRODUCTION TO GAIT RECOGNITION 1

1 INTRODUCTION 3
- 1.1 Biometrics 3
- 1.2 Human Identification Using Gait 3
- 1.3 Challenges in Gait Recognition 5
- 1.4 Objectives 8
- 1.5 Main Contributions 9
- 1.6 Outline of Thesis 9

2 LITERATURE REVIEW 11
- 2.1 General framework of a Gait Recognition System 11
- 2.2 Related works 11
  - 2.2.1 Gait Recognition Review 11
  - 2.2.2 Human Gait Recognition Using Feature Selection 13
  - 2.2.3 A Review on Deep Learning Techniques 15
- 2.3 Existing gait databases 17
  - 2.3.1 Soton database 17
  - 2.3.2 CASIA dataset 17
  - 2.3.3 OU-ISIR Treadmill database 18
  - 2.3.4 USF dataset 19
  - 2.3.5 TUM-GAID dataset 19
- 2.4 Summary 19

II SILHOUETTE ANALYSIS-BASED GAIT RECOGNITION FOR HUMAN IDENTIFICATION 21

3 BASELINE GAIT RECOGNITION ALGORITHM 23
- 3.1 Overview 23
- 3.2 Existing works 23
- 3.3 Methodology 25
  - 3.3.1 Extracting the Gait Features 25
  - 3.3.2 Feature Post-process 31
  - 3.3.3 Classification Process 34
- 3.4 Experiments 34
  - 3.4.1 Datasets 34
  - 3.4.2 Experimental Setup 35
- 3.5 Results and Discussions 36
  - 3.5.1 Performance Evaluation of SVM Classifier 36
  - 3.5.2 Gait Recognition Confusion Matrix 37
  - 3.5.3 The Effect of Covariate Factors 37
  - 3.5.4 The Effect of Occlusions 40
  - 3.5.5 Evaluation on Benchmark Methods 41
- 3.6 Summary 42

III IMPROVING HUMAN GAIT RECOGNITION USING FEATURE SELECTION 43

4 FEATURE SELECTION FOR GAIT RECOGNITION 45
4.1 Overview 45
4.2 Methodology 46
4.2.1 Gait Pattern Extraction and Representation 47
4.2.2 Statistical Methods for Gait Feature Selection 48
4.2.3 Search-Based Methods for Gait Feature Selection 49
4.2.4 Embedded Methods for Gait Feature Selection 50
4.2.5 Joint Mutual Information Maximisation (JMIM) 52
4.3 Experiments 52
4.3.1 Design of experiments 53
4.3.2 Database Description 53
4.4 Results and Discussions 54
4.4.1 Exploratory Analysis 54
4.4.2 Statistical Analysis 55
4.4.3 Analysis of Computational Time and Best Feature Subset Length 63
4.4.4 Further Comparisons with Conventional Methods 66
4.5 Summary 67

IV WHEN GAIT RECOGNITION MEETS WITH DEEP LEARNING 71

5 DEEP LEARNING FOR GAIT RECOGNITION 73
5.1 Overview 73
5.2 Existing work 73
5.3 Convolutional Neural Network 74
5.3.1 Methodology 75
5.3.2 Experiments 77
5.3.3 Results and Discussion 81
5.4 Stacked Progressive Auto-encoders (SPAE) 84
5.4.1 Methodology 84
5.4.2 Experiments 88
5.4.3 Results and Discussions 89
5.5 Summary 91

V CONCLUSIONS 93

6 CONCLUSIONS AND FUTURE WORK 95
6.1 Conclusions 95
6.2 Contributions and Discussions 95
6.3 Future research directions 97
6.3.1 Investigations in the short-term 97
6.3.2 Investigations in the long-term 97

VI APPENDICES 99

A DATASET 101

A.1 SOTON Covariate database 101

A.2 OU-ISIR Treadmill B Dataset 101

A.3 CASIA-B dataset 101

BIBLIOGRAPHY 105

Part I

INTRODUCTION TO GAIT RECOGNITION

In this part, we present a brief introduction to gait recognition. Then, it describes the background, challenges, objectives and contributions of this research. Finally, it outlines the content of this work.

1 INTRODUCTION

1.1 BIOMETRICS

Biometrics is the study of recognizing humans based on their physiological or behavioural traits. Physiological traits are biological patterns found on and in the human body, for example: face, iris, fingerprint, DNA, palm print, hand geometry, etc. On the other hand, behavioral traits patterns develop over time and become a consistent characteristic, for example: gait, typing rhythm, signature, etc. In Figure 1, fingerprint, iris, face, ear, hand geometry, palm print, finger vein geometry, gait, voice, signature, keyboard stroke pattern are common biometrics used in various systems.

A biometric trait needs to satisfy the following properties [58]:

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

A biometric system based on a certain trait may have either a verification or and identification mode, depending on the application context [58]. In the verification mode, the system validates the claimed identity (of a subject) by comparing the query biometric trait with his/her own reference trait stored in the system's database. The system conducts a one-to-one comparison to determine whether the claim is true or not. Biometric verification has widely been used in commercial applications such as access control. In the identification mode, the system recognises a subject by searching all the biometric templates of all subjects in the database for a match. The system conducts a one-to-many comparison for a subject's identity. Biometric identification is used more frequently in the applications of law enforcement, e.g., latent fingerprint identification, human gait identification. The flowcharts of both modes are shown in Figure 2, and this thesis falls into the category of gait biometric identification.

1.2 HUMAN IDENTIFICATION USING GAIT

Human identity recognition is fundamental to human life, and the technology of human identification and tracking from a distance may

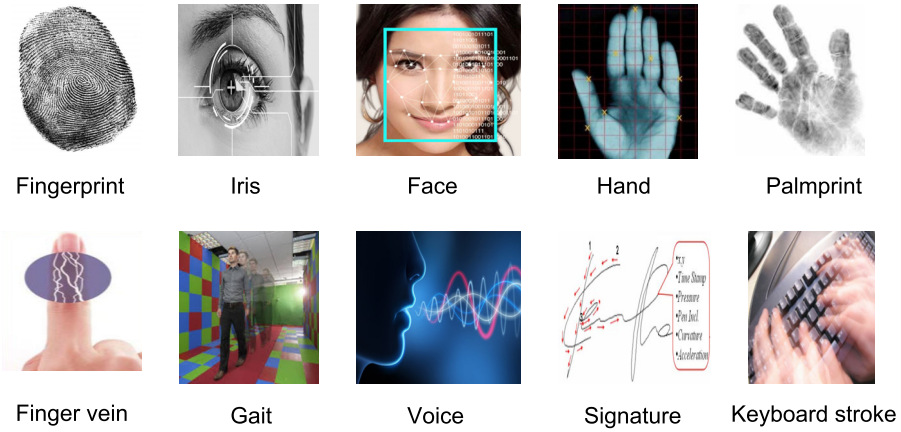


Figure 1: Examples of biometrics.

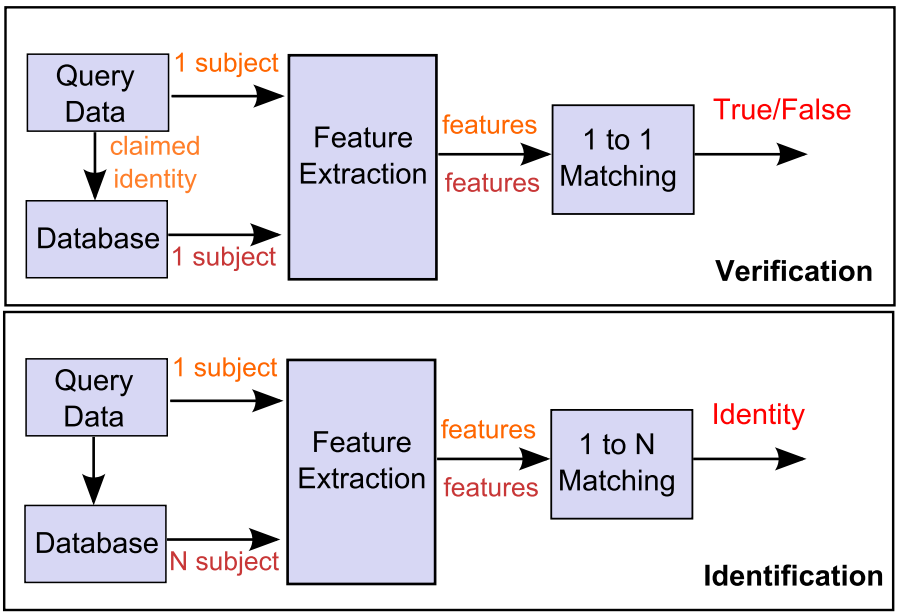


Figure 2: Biometric systems.

play important role in crime prevention, law enforcement, search for missing people (e.g. missing children or people with dementia), etc. Nowadays, CCTV cameras are widely installed in public places such as airports, government buildings, streets and shopping malls for the afore-mentioned purposes. In 2013, the British security industry authority (BSIA) estimated there are up to 5.9 million CCTV cameras nationwide, and that is around 1 every 11 people [6]. Because of the need for sufficient manpower to supervise such a large number of CCTVs, the need for automatic human identification systems is acute.

Recently, a number of reports (e.g., [70, 18]) suggested that behavioural biometrics, gait recognition, can be used for human identification from CCTV footage. In [70], based on a checklist for forensic gait analysis, Larsen et al. managed to identify a bank robber in Denmark by matching surveillance footage, as illustrated in Figure 3a.

Figure 3b shows a gait recognition scenario in UK where a burglar was identified through gait analysis from a podiatrist [18]. These pieces of gait-based evidences proved their usefulness by providing incriminating evidence, leading to convictions in a court of law.

However, automated recognition of humans by their gait is a very challenging research problem. Although solutions to many challenges have been proposed, some key issues remain. In addition, only a small number of approaches have been validated in real world environments. An operational automated gait recognition system does not yet exist. The work presented in this thesis bring gait recognition closer to real-world deployment. This is achieved by examining the factors (also known as covariates) that are known to affect recognition in a more principled manner and propose solutions to some of the fundamental problems.

(a)

(b)

Figure 3: CCTV images for the robbery case in Denmark [70], top: the perpetrator, right: the suspect; (b) CCTV images for the burglary case in UK [18], bottom: the perpetrator, right: the suspect.

1.3 CHALLENGES IN GAIT RECOGNITION

Gait recognition is one of the most active research topics in the interdisciplinary areas of biometrics, pattern recognition, computer vision and machine learning. However, in real-world scenarios, gait recognition became a difficult problem of computer vision because it often

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

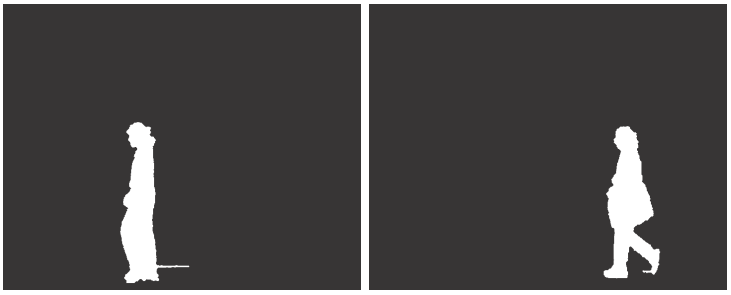
Figure 4: Pre-processed silhouettes with missing body parts, noise and shadows.

suffers from obstacles caused by image noise and changing lighting conditions. Specifically, extracting features from a gait video sequence involves segmentation of the moving person from the background. Image noise and changing lighting conditions directly affect the ability of algorithms to segment correctly the moving person from the background thus causing missing body parts and the inclusion of background (e.g. shadows) as shown in Figure 4. To reduce the effect of image noise and changing lighting conditions a pre-processing stage is normally required in a gait recognition framework.

The complexity of the problem is further compounded with occlusions as shown in Figure 5. There are two different types of occlusions that normally occur in the scene, such as self occlusion and occlusions from other objects. Self occlusion is caused by the cross over of the legs and the swing of the hands as the person moves. Self occlusion is unavoidable because a subject's legs are self-occluded during motion. However, it will incur minor losses of information in gait feature extraction. On the other hand, occlusions from other objects, obscures a significant amount of gait features and make human recognition difficult. Since the location of occlusion may differ for different objects, relevant gait features may become irrelevant when the object changes, and exploiting occluded gait features can hinder the recognition performance.

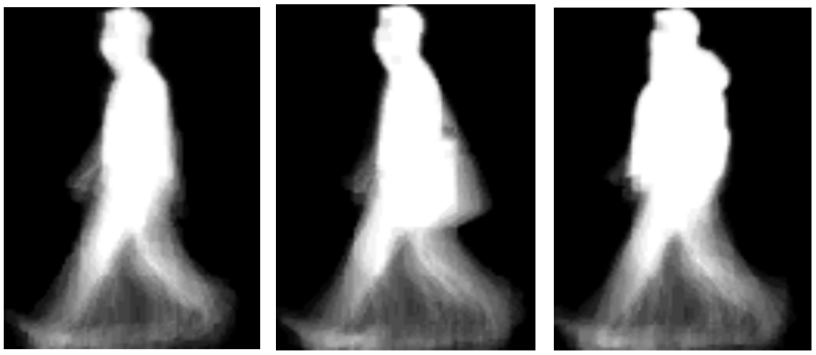
In addition to image noise, lighting condition changes and occlusions, gait is affected by variable covariate conditions. The presence of variable covariate conditions in gait changes the available features which eventually affects the way gait is represented. Example gait sequences for the same person represented using state-of-the-art Gait Energy Images proposed by Han and Bhanu [81] are shown in Figure 6 to highlight this aspect.

In gait recognition, the gallery set consists of people walking under normal (see Figure 6a) covariate conditions (i.e. people walking under similar and common, known conditions) and if the probe set also



(a) Self-occluded silhouettes (b) Occluded with handbag

Figure 5: Samples images of occluded gait silhouette frames taken from SOTON covariate database [97]: (a) Self-occluded silhouettes (b) Occluded with the object .



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

Figure 6: Gait Energy Images of a subject under variable covariate conditions.

consists of gait sequences under conditions similar to the gallery set this brings out good results. However, when the covariate conditions in the probe set are variable (see Figure 6b, Figure 6c) this incurs significant drop in recognition performance. This is because the same person's appearance can be very different due to these covariate condition differences between them.

Amongst the covariate conditions affecting gait, clothing poses one of the most challenging issues in this area [53], as illustrated in Figure 7. This is because variations in clothing alter an individual's appearance, making the problem of gait identification much more difficult [84]. It will drastically alter the individual's appearance with the variation of different clothing types, such as baggy pants, skirt, down jacket, and coats. In addition, variation in viewing angles will drastically affect the performance of gait recognition [126], as illustrated in Figure 8. The aforementioned issues has recently gained considerable attention from all the gait researchers.

Figure 7: A samples images of clothes variation from OU-ISIR Treadmill B dataset [53]

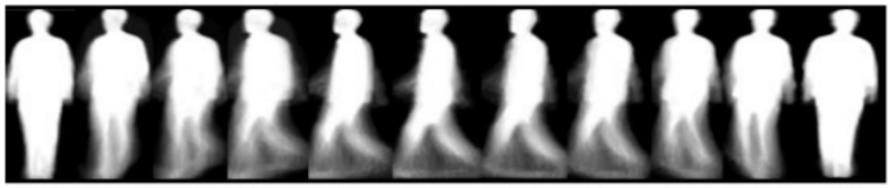


Figure 8: GEI examples from CASIA-B dataset [126] of a subject from view 0°to 18°, with an interval of 18°

1.4 OBJECTIVES

The overall objective of this work is to reduce the effect of covariate factors in gait recognition.

1. **Robust gait representation against various covariate factors**
   Changes in covariate conditions adversely affect the performance of gait recognition methods. One of the major goals of this work is to improve the performance of existing gait recognition methods in the presence of variable covariate conditions in the probe set. In particular, this work aims to investigate and improve the performance of gait recognition under variable covariate conditions.

2. **Reduction of the effect of gait covariate factors using feature selection**
   Most of the existing approaches in this area have been evaluated without explicitly considering the most relevant gait features, which might have compromised the gait recognition performance. An objective of this work is to select the most relevant and optimal gait features that are invariant to changes in gait covariate conditions, on improving the performance of a gait recognition system.

3. **Improvements to gait recognition using deep learning techniques**
   Many efforts have been made in recent years to develop recognition algorithms for gait. However, previous approaches are mostly used hand-crafted methods for representing human gait. This research aims to investigate ways to incorporate the deep

learning algorithms to solve the problem of gait recognition under the presence of challenging covariate factors.

1.5 MAIN CONTRIBUTIONS

The main contributions of this thesis are listed as below:

1. **Automatic human silhouette body joint identification**
   We propose a simple model-based gait recognition focusing on modeling gait dynamics and eliminating the effect of the covariates on gait recognition. In particular, our method shows advantages when handling the occluded silhouette either from self-occlusion or those occluded by apparels, which are normally disastrous for other gait feature approaches. The experimental results demonstrate our method outperform other approaches on SOTON covariate database under various covariate factors.

2. **Feature selection for gait recognition**
   We present and compare feature selection techniques, designed to maximise the final gait classification accuracy. The overall approach constitutes a general framework for different machine learning algorithms, which we applied to the problem of feature selection under the effect of various covariate factors in a model-based approach. The implemented method addresses the problem of feature selection for gait recognition on two well-known benchmark databases: the SOTON covariate database and the CASIA-B dataset, respectively. The investigated approach is able to select the discriminative input gait features and achieve an improved classification accuracy on par with other state-of-the-art methods.

3. **Deep learning for gait recognition**
   We explore the use of deep learning approaches and show how to adapt it for gait recognition under challenging covariate conditions such as view and clothing. The proposed method is evaluated on the challenging view and clothing dataset and the results from our experiments demonstrate that our methods yield better recognition accuracy in the case of large intra-class variation such as the view and clothing variations.

1.6 OUTLINE OF THESIS

The remaining chapters of this thesis are structured as follows:

- Chapter 2 - **Literature Review**
  This chapter provides the relevant literature on gait recognition framework with emphasis on feature representation, databases and various algorithms against different covariate factors. We

also make emphasis on some fundamental knowledge on feature templates, feature selection, and the concept of deep learning for gait recognition problem.

- Chapter 3 - **Gait Recognition Framework**
  This chapter describes the implementation of the gait recognition framework used in this thesis. A simple yet efficient gait representation based on model-based approach is propose to address the problem of variable covariate conditions in gait recognition.

- Chapter 4 - **Feature Selection for Gait Recognition**
  This chapter investigates the problem of selecting a subset of the most relevant gait features for improving gait recognition. This is achieved by using a proper selection technique to find the relevant features and discard redundant and irrelevant gait features. We present an empirical approach to evaluate the degree of consistency among the performance of different selection algorithms in the context of gait identification under the effect of various covariate factors.

- Chapter 5 - **Deep Learning for Gait Recognition**
  This chapter proposes a method of gait recognition using a convolutional neural network (CNN) for learning and extracting higher-level view-invariant and clothing-invariant gait features that suitable for representing the human gait. However, recognizing subjects with variations caused by different types of clothes is one of the most challenging tasks in gait recognition, since the differences in appearances due to clothing variations may be even larger than the difference due to personal identity. Thus, we also propose to learn clothing-invariant gait features for gait recognition in a progressive way by stacking multi-layer of auto-encoders.

- Chapter 6 - **Conclusions and Future Work**
  This chapter summarises the achievements of this thesis and suggests directions for future research.

2 LITERATURE REVIEW

2.1 GENERAL FRAMEWORK OF A GAIT RECOGNITION SYSTEM

There are three main components in a general gait recognition system, as illustrated in Figure 9. Initially, segmentation processes are applied to extract the human silhouettes from the video sequences. A number of post-processing tasks including normalization of the silhouettes are applied to the extracted silhouettes to register them on a common platform. Then gait features are extracted and modelled to encode the individual-specific gait information. Finally, classification is performed to compute the similarity distance that will be used to claim individual's identity. Within all of these processes in the framework, the most important step is extracting appropriate gait feature that can uniquely distinguish subjects. Existing feature extraction methods can generally be classified as appearance-based methods and model-based methods. Both appearance-based or model-based methods can be applied to 2D or 3D data sources. This section discusses the research pertaining to the major components in the gait recognition framework: silhouette segmentation, feature extraction, feature modeling and classification.

2.2 RELATED WORKS

2.2.1 *Gait Recognition Review*

In the specialized literature, a wide variety of gait recognition systems have been proposed (cf. [73] for a recent review). Gait recognition approaches can be classified into two main categories: model-based and model-free methods [79]. There exists a considerable amount of work in the context of both approaches for gait recognition. In this section, we present conventional methods from each category with emphasis on feature extraction only.

Model-based methods analyze the human body structures underlying the gait data, and extract measurable parameters, such as the static and dynamic body parameters. Normally, these methods perform model fitting for matching in each frame of a walking sequence so that the extracted kinematic parameters, such as joint trajectories, can be measured. These approaches are normally robust to viewpoint change and scale [110]. Bobick et al. [16] used two sets of activity-specific static body parameters by considering the distances between different human body parts. Yoo et al. [124] constructed a two-dimensional

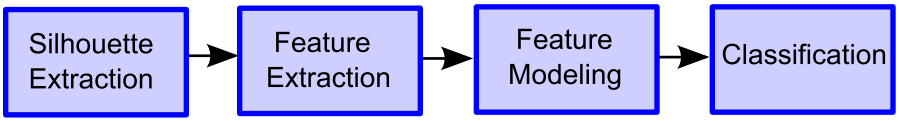


Figure 9: Major components of gait recognition algorithms. At first, silhouettes are extracted and pre-processed using background subtraction algorithms. Then, gait features are computed from the segmented silhouettes and these gait features are modelled to encode better distinguishable elements. Finally, these modelled-features of the test subject are compared on the pre-enrolled database to claim the identity.

(2-D) stick figure from the extraction of nine body points derived from skeleton data of each body segment based on human anatomical knowledge. Tanawongsuwan et al. [102] reconstructed the human structure by focusing on trajectories of lower body joint angles derived from motion capture data. Wang et al. [108] presented a method based on positioning body joints according to the geometrical characteristics shown during walking. Cunado et al. [32] proposed a moving feature extraction analysis that automatically extracts and describes human gait for recognition. The algorithm extracted the moving model from a sequence of images by using Fourier series to describe the motion of the thigh and apply temporal evidence gathering techniques.

Model-free methods generally employ either shape of binary silhouettes or the whole motion pattern of the human body. Conversely, these approaches are insensitive to the quality of silhouettes and noise and have the advantage of low computation costs compared to model-based approaches [113]. Collins et al. [28] established a method based on template matching of body silhouettes in key frames. In [72], Lee et al. described a moment-based representation of gait appearance by using the features extracted from orthogonal view video silhouettes of human walking motion. Sarkar et al. [94] proposed a baseline algorithm for human identification using spatio-temporal correlation of silhouette images. Wang et al. [109] developed a Chrono-Gait Image (CGI) where the contour in each of gait frame was first extracted and then encoded those in the same gait sequence to a multichannel image, and showed its robustness to a complex surrounding environment in their experimental results. Han et al. [81] proposed a spatio-temporal gait representation called Gait Energy Image (GEI) to distinguish human gait patterns for individual recognition. However, the main drawback of model-free gait recognition is dealing with various intra-class variations caused by the presence of covariate factors such as clothing variations, shadows and carrying conditions [93]. To overcome the limitations of GEI as the gait representation, several approaches have been proposed. Bashir et al. [8] proposed a feature selection method named Gait Entropy Image (GEnI). It measures

Shannon entropy at each pixel to distinguish static and dynamic pixels of GEI. The GEnI represents a measure of the relevance of the gait features extracted from the Gait Energy Image (GEI). Besides, Bashir et al. [9] proposed a gait representation based on optical flow fields computed from normalized and centred person images over a complete gait cycle for gait recognition. Recently, Rida et al. [93, 91] deployed a supervised feature extraction which uses the Modified Phase-Only Correlation matching method to improve gait recognition performance.

Our proposed method belongs to the class of model-based approaches instead. In the model-based approach, static and dynamic features are extracted from the silhouettes. In general, these features represent the position and pose of various human body parts with respect to each other as a person walks in the scene. The main advantage of forming a gait signature in this manner is that it is view as well as scale invariant. However, whether model-based or model-free, the extracted features could be affected by variations in different types of covariate factors, which will complicate the human recognition process.

2.2.2 *Human Gait Recognition Using Feature Selection*

The selection of an optimal subset of features is an important step in pattern recognition; often a large number of features are extracted to better represent the target concept. Given a set of d features, the problem is to select a subset of size m that maximizes a scoring function. This is essentially an optimization problem that involves searching the space of possible feature subsets to find one that is optimal with respect to a certain criterion [101].

Three main kind of approaches to features subset selection have been commonly adopted [61, 64]: wrapper approaches, filter approaches and embedded approaches. A wrapper approach uses machine learning algorithms in the search process, that is, it gauges the relevance of a potential feature subset by directly evaluating the accuracy of a model that has been trained on the selected features. In contrast, a filter approach filters undesirable data features before the classification process. Filter approaches are then more efficient, in terms of computational complexity, because they use statistical testing or heuristics based on general characteristics of the data, instead of training and evaluating a learning model on the data set as wrapper methods do. Finally, in an embedded approach [45], the search for an optimal subset of features is built into the learning stage. These methods are thus learning-algorithm specific, which could be a benefit in that they are tailored to the classification model, while at the same time being far less computationally intensive and less prone to over-fitting than wrapper methods. However, they are limited in terms of generality.

Recent research in automated human gait recognition has mainly focused on developing robust features representations and matching algorithms. To the best of our knowledge, feature selection has been mainly used as a method to pre-select discriminative human body parts in most of the conventional approaches [93, 37]. Rida et al. [93] proposed a method to explore the most discriminative human body part based on group Lasso of motion to reduce the intra-class variations so as to improve the gait recognition performance. In [37], Dupuis et al. proposed a feature selection method based on Random Forest algorithm to rank features importance. But then the respective gait analysis have mostly been evaluated without explicitly considering the most relevant gait features, which might have compromised the classification performance. Besides, some of these approaches [124, 81] have considered conventional dimensionality reduction or statistical tools, such as: Principal Component Analysis (PCA) and Analysis of Variance (ANOVA). In [44], Guo and Nixon used Mutual Information (MI) to measure the utility of selected features for recognition. Bashir et al. [7] developed a supervised approach based on cross-validation to explore features subsets from the Gait Energy Image (GEI) in order to optimise recognition performance. That required prior knowledge about the GEI feature characteristics to overcome the problems associated with searching exhaustively through a high-dimensional gait feature space. Moreover, the information contained in GEI can be erroneous when in the presence of various covariate factors. To reduce the effect of covariate factors in gait recognition, Guan et al. [43] employed a classifier ensemble method based on Random Subspace Method (RSM) and Majority Voting (MV) as feature selection and classification method to address this problem. Recently, Semwal et al. [96] introduced an optimized feature selection method based on incremental feature selection strategy for biometric gait data classification. Since gait features extracted from segmented video sequences are frequently interspersed with background noise or covariate factors, the classification could be misguided. Hence, without employing a simple yet effective gait feature selection, gait features being extracted could be redundant or irrelevant to the gait recognition task. Also, feature selection could provide valuable clues in terms of understanding the underlying distinctness among human gait patterns.

The idea of using an evolutionary computational approach for feature selection in gait recognition was firstly explored in our previous work [123, 120]. In [123, 120], we proposed wrapper approaches based on evolutionary algorithms, namely Genetic Algorithm (GA) and Geometric Particle Swarm Optimization (GPSO) assisted by Support Vector Machine (SVM). These methods have been been used to reduce the effect of the covariates: the results showed a slight improvement in gait recognition performances on the SOTON covari-

ate database [97]. In addition, previous methods involved data preprocessing steps such as feature outlier removal. In [2], Altilio et al. introduced a feature selection method based on genetic algorithm for stereophotogrammetric analysis to extract useful information from medical data. Moreover, these methods have a high computational cost (i.e. computational time) and therefore are not applicable for real-time gait recognition systems. In our work, we extend our methodology on deploying optimal feature selection approaches for selecting the relevant set of gait features to mitigate the effect of changes in different covariate factors while reducing the computational cost. To the best our knowledge, a comprehensive comparative analysis on feature selection is seldom addressed in gait classification literature, which mostly focuses on presenting novel approaches. Therefore, we aim to evaluate the degree of consistency among the performance of different selection algorithms in the context of gait identification under the effect of various covariate factors.

2.2.3 *A Review on Deep Learning Techniques*

The convolutional neural network (CNN) model is an important type of feed-forward neural with special success on applications where the target information can be represented by hierarchy of local features (see ref. [41]). A CNN is defined as the composition of several convolutional layers and several fully connected layers. Each convolutional layer is, in general, the composition of a non-linear layer and a pooling or sub-sampling layer to get some spatial invariance. For images, the non-lineal layer of the CNN takes advantage, through local connections and weight sharing, of the 2D structure present in the data. These two conditions impose a very strong regularization on the total number of weights in the model, which allows a successful training of the model by using back-propagation. In our approach, although we do not feed the model directly with the RGB image pixels, the CNN approach remains relevant since the Gait Energy Image (GEI) information also shares the local dependency property as the pixels do.

In the last years, CNN models are achieving state-of-the-art results on many different complex applications (e.g. object detection, text classification, natural language processing, scene labeling, etc.) [29, 66, 39, 128]. However, to the extent of our knowledge, CNN has not been applied to the problem of gait recognition yet. The great success of the CNN model is in part due to its use on data where the target can be represented through a feature hierarchy of increasing semantic complexity. When a CNN is successfully trained, the output of the last hidden layer can be seen as the coordinates of the target in a high level representation space. The fully connected layers, on top of

the convolutional ones, allow us to reduce the dimensionality of such representation and, therefore, to improve the classification accuracy.

Traditionally, deep learning approaches based in Convolutional Neural Networks (CNN) have been used in image-based tasks with great success [66, 99, 127]. In the last years, deep architectures for video have appeared, specially focused on action recognition, where the inputs of the CNN are subsequences of stacked frames. In [98], Simonyan and Zisserman proposed to use as input to a CNN a volume obtained as the concatenation of frames with two channels that contain the optical flow in the x-axis and y-axis respectively. To normalize the size of the inputs, they split the original sequence in subsequences of 10 frames, considering each subsample independently. Donahue et al. [36] propose another point of view in deep learning using a novel architecture called "Long-term Recurrent Convolutional Networks". This new architecture combines CNN (specialized in spatial learning) with Recurrent Neural Networks (specialized in temporal learning) to obtain a new model able to deal with visual and temporal features at the same time. Recently, Wang et al. [112] combined dense trajectories with deep learning. The idea is to obtain a powerful model that combines the deep-learnt features with the temporal information of the trajectories. They train a traditional CNN and use dense trajectories to extract the deep features to build a final descriptor that combines the deep information over time. On the other hand, Perronnin et al. [88] propose a more traditional approach using Fisher Vectors as input to a Deep Neural Network instead of using other classifiers like SVM.

Although several papers can be found for the task of human action recognition using deep learning techniques, it is hard to find such type of approaches applied to the problem of gait recognition. However, there are a few gait recognition studies that use a deep learning framework because deep learning requires a large number of training samples and it is difficult to collect a large number of training gait samples. To the best of our knowledge, deep learning-based gait recognition has been performed only by Hossain et al. [52] and Wu et al. [116]. Hossain et al. propose the use of Restricted Boltzmann Machines to extract gait features from binary silhouettes, but a very small probe set (i.e. only ten different subjects) were used for validating their approach. is completely insufficient for a deep learning framework. On the other hand, Wu et al. employed CNN for gait recognition, and reported better recognition accuracy than those of benchmarks. One concern of their work is that they represented gait by randomly selected silhouette image set, and used it as an input to the CNN. This representation is not appropriate for gait recognition, because useful dynamic information for gait feature cannot be considered in the representation. Therefore, we propose a method of gait recognition using deep learning approaches with appropri-

Figure 10: Gait images with different covariates from the SOTON covariate dataset

ate gait representation and also demonstrate its effectiveness in gait recognition tasks using the available gait database.

2.3 EXISTING GAIT DATABASES

There are several gait databases that have been produced by the members of the computer vision and biometrics community. This section will introduce various gait databases.

2.3.1 *Soton database*

In [97], it contains two types datasets: a large dataset with more than 100 subjects and a small dataset with only 10 subjects. The large dataset has two viewpoints (frontal and oblique) and contains subjects in both outdoor and indoor environments and on a treadmill. The small dataset is extensive with respect to conditions such as the type of footwear, clothes and surface. Hence, it is also used for exploratory factor analysis of gait recognition [20], as illustrated in Figure 10.

2.3.2 *CASIA dataset*

The database [126] contains the largest azimuth view variations and hence, it is useful for the analysis and modeling of the impact of view on gait recognition. It is available in three datasets. CASIA-A consists of 20 subjects. Each subject has 12 image sequences (length of 37 to 127 frames), four sequences for each of three directions (parallel, 45°and 90°to the image plane). CASIA-B (large multi view gait database) has 124 subjects for 11 view angles (from 0°to 180°stepped by 18°) and for each view angles with four sets of normal condition and two sets of subjects with bag and subjects wearing coat, as illus-

Figure 11: Gait images from CASIA-B dataset of a subject from view 0°to 180°, with an interval of 18°

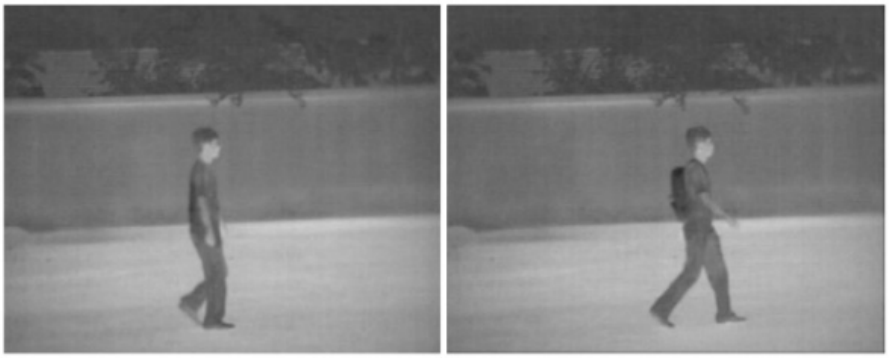


Figure 12: Gait images from CASIA-C dataset of a subject collected at night environment using infrared cameras two different walking conditions: normal walking and carrying condition, from left to right

trated in Figure 11. CASIA-C was collected by an infra-red (thermal) camera and contains 153 subjects and four walking conditions: normal walking, slow walking, fast walking, and normal walking with a bag, as illustrated in Figure 12.

2.3.3 *OU-ISIR Treadmill database*

It is available in four datasets [78]. The dataset focuses on variations in walking conditions and includes 34 subjects with 9 speed variations from 2 km/h to 10 km/h with a 1 km/h interval (OU-ISIR-A), as illustrated in Figure 13. The OU-ISIR-B dataset was constructed by Hossain et al. [53] for studying the effect of clothing on gait recognition. It includes 68 subjects walking on a treadmill with up to 32 clothes variations, as illustrated in Figure 14. The OU-ISIR-C database contains images of 200 subjects from 25 views. The OU-ISIR-D database consists of gait silhouette sequences of 185 subjects from side view with various gait fluctuations among periods. Gait fluctuations are mea-

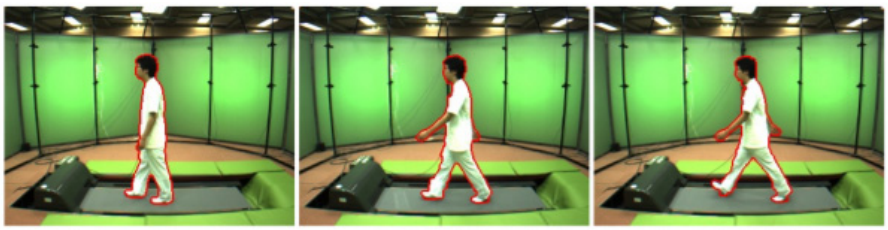


Figure 13: Gait images with different speeds from the OU-ISIR-A dataset

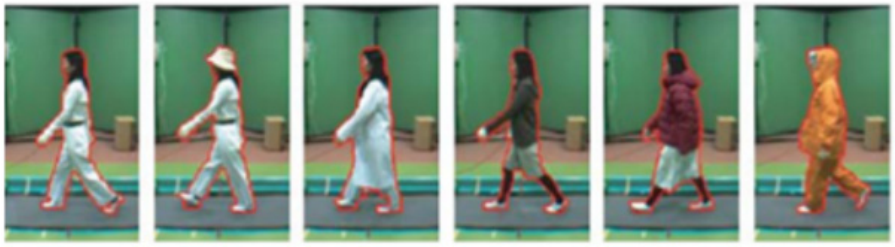


Figure 14: Gait images with several different clothes types from the OU-ISIR-B dataset

sured by Normalized AutoCorrelation (NAC) for temporal axis of size-normalized gait silhouette images. DB_{high} comprising 100 subjects with the highest NAC, and DB_{low} comprising 100 subjects with lowest NAC.

2.3.4 *USF dataset*

The USF dataset [94] is a large outdoor gait database consisting of 122 subjects. A number of covariate factors are considered: camera viewpoints, shoes, surface types, carrying conditions, elapsed time, and clothing. Several gait images from this dataset are shown in Figure 15.

2.3.5 *TUM-GAID dataset*

The TUM-GAID dataset [50] simultaneously contains RGB images, depth images, and audio of 305 subjects in total. In [50], Hofmann et al. designed an experimental protocol (based on 155 subjects) to evaluate the robustness of algorithms against covariate factors like shoe, carrying condition (5kg backpack), elapsed time (January/April) which also potentially includes changes in clothing, lighting condition, etc. Several gait images from this dataset are shown in Figure 16.

2.4 SUMMARY

In this chapter, we first introduced an overview of a typical gait recognition system. We then discussed the related works of gait recognition and further introduced the problems caused by the covariate factors

Figure 15: Gait images in the outdoor environment from the USF dataset

Figure 16: Gait images from TUM-GAID dataset of a subject with six different walking variations: normal walking (in Jan.), with a backpack (in Jan.), with a different pair of coating shoes (in Jan.), normal walking (in April), with a backpack (in April), with a different pair of coating shoes (in April), from left to right

which is the main theme of this thesis. We have introduced the approaches for gait recognition and summarised the current challenges. Finally, we introduced the popular databases used for the research in gait recognition. We will present our proposed methods for the gait recognition in the following chapters.

Part II

SILHOUETTE ANALYSIS-BASED GAIT RECOGNITION FOR HUMAN IDENTIFICATION

In this part, a simple but efficient gait recognition algorithm focusing on modelling gait dynamics and eliminating the effect of the covariates on recognition is proposed.

3

BASELINE GAIT RECOGNITION ALGORITHM

3.1 OVERVIEW

Robust human identification based on gait is challenging because of the presence of various types of covariate factors such as clothing, load carrying condition, footwear and walking speed. Variations in covariate factors have a strong impact on the recognition of gait. To address this problem, we presents a gait-based human identification method with automatic human silhouette body joint identification that is robust against various covariate factors at the same time. The proposed approach consists of three parts: (i) extraction of human gait features from enhanced human silhouette images, (ii) feature post-process on extracted human gait features to determine the significant gait features, and (iii) classification by Support Vector Machine (SVM) technique. This combination approach is designed to adapt the variations in covariate factors across and within a gait walking sequence. Hence, the proposed approach shows robust performance against various covariate factors in SOTON covariate database and also outperforms other state-of-the-art algorithms.

3.2 EXISTING WORKS

Human gait recognition has attracted increasing attention as it is unobtrusive biometrics that can be captured from a distance and without requiring any cooperation from the user [73]. As a result, gait has become one of the latest promising biometrics. However, the practical deployment in a real application is still considered a challenging task [17]. The performance of gait as biometric can be affected by variation in covariate factors such as clothing, load carrying condition, footwear and walking speed with respect to the subject [115, 43, 114]. This is mainly because of the variations will alter an individual's appearance and increasing the difficulty of gait recognition.

In recent years, various techniques for human gait recognition have been proposed in the literature [86]. Gait recognition methods can be mainly classified into 2 categories: model-based method and model-free method. A model-based method generally models the human body structure as blobs or rectangles and extracts the features to match them to the modeled components. It incorporates knowledge of the human shape and dynamics of human gait into an extraction process. For instance, Bobick et al. [16] used activity-specific static body parameters for gait recognition without directly analyzing gait

dynamics. Cunado et al. [32] used thigh joint trajectories as the gait features. The advantages of these methods are the ability to derive the gait signatures directly from model parameters rather than correlating with other measures (such as motions of other unrelated objects). Thus, the noise effects from the surrounding environment can be removed easily. However, it creates many parameters which require high computational cost due to complex matching and searching process.

On the contrary, a model-free method differentiates the whole motion pattern of the human body through a concise representation which does not consider the underlying structure. For example, Ben-Abdelkader et al. [12] proposed an eigengait method using image self-similarity plots. Collins et al. [28] established the method based on template matching of body silhouettes in key frames during a human's walking cycle. Pratheepan et al. [90] proposed individual identification using dynamic static silhouette template. Philips et al. [89] characterized the spatial-temporal distribution generated by gait motion in its continuum. Compared to the model-based method, this method is simple and thus incurs low computational requirements. However, the performance of this method is intensely affected by the background noise or covariate factors such as changes of the subject's apparel and load carrying condition. With regard to background noise, a couple of approaches to silhouette refinements for gait recognition [51, 74, 111] have been proposed. In method [51], they exploit alpha-matting process for better representation for foreground background matte. The methods [74, 111] exploits the gait information as shape prior for foreground segmentation in the form of eigen stance and standard gait model, respectively. With regards to covariate factors, several recent methods demonstrate to mitigate the effects of covariate factors. For example, the method [53] adaptively controls the part weights depending on the degree of affection by clothing changes, while the method [104, 33] explores the robust expression against carrying status. In recent work [82], they pursuit an approach to robust gait recognition against a certain type of covariate factors using rank SVM. Although these works improved gait recognition performance against a certain covariate type, they can still suffer from the effect of various covariate factor such as the presence of shadows, clothing variations and carrying conditions (backpack, briefcase, handbag, etc.). This motivated us to extend our previous work [119] to achieve a robust gait recognition performance under various covariate factors.

Therefore, in this work, we propose a simple model-based gait recognition system focusing on modelling gait dynamics and eliminating the effect of the covariates on recognition. In our approach, we extract the human gait features from original silhouette images. After that, the human silhouette is divided into eight segments, from

which the body joints are automatically identified and the joint trajectories are computed. This work extends our previous work [119] by employing well-structured combination of feature post-process techniques to achieve a better computationally efficient solution to alleviate the effect of variations in covariate factors for gait recognition. In addition, this work does not attempt to detect the legs separately. Therefore, it can handle occluded silhouette either from self-occluded or those by apparels such as subject apparel (long blouses or baggy trousers) or load carriage (hand bag held in hand or barrel bag slung over shoulder at hip height) which are usually present significant challenges for other conventional approaches. Moreover, the computation of joint trajectories is more straightforward than the Hough transform-based technique presented by Ng et al. [85]. Due to its simplicity, the proposed approach also executes more efficiently than model-based methods such as the linear regression approach by Yoo et al. [125] and temporal accumulation approach by Wagg et al. [107] and elliptic Fourier descriptors by Bouchrika et al. [19, 20].

3.3 METHODOLOGY

In this section, we describe the details of the proposed method. At first, we describe a method of feature extraction from gait images, and we introduce the human identification method that achieves higher performance compared with the previous method [119] and is robust to variation in covariate factors.

In this work, we propose a model-based approach by extracting the gait features from original human silhouette images with automatic human silhouette body joint identification. Then, effective feature post-processing techniques to apply on extracted gait features. To assess the performance of the proposed approach, SVM classification technique is used to classify subjects from the database.

3.3.1 *Extracting the Gait Features*

For the gait feature extraction, morphological opening is first applied to reduce background noise on the extracted human silhouette images. Each of the human silhouettes is then measured for its width and height. Next, each of the enhanced human silhouettes is segmented into eight body segments based on anatomical knowledge. The lower body joints that define the pivot points in human gait are then automatically identified and the joint trajectories are computed. After that, step-size and crotch height are measured.

3.3.1.1 *Silhouette Image Enhancement*

The original human silhouette images are obtained from the SOTON covariate database [97]. This database was used to evaluate the recognition rate of the walking subjects with different covariate factors. Figure 17a shows gait sample video images of an individual. In most of the human silhouette images, shadow is chronically found near the feet. It is usually merged to the subject body in the human silhouette image as shown in Figure 17b. This will hinder the gait feature extraction as it interferes with the body joint identification. The problem can be reduced by applying morphological opening and closing operations to smooth the objects boundaries and fills in small holes. Both morphological operations are using 7x7 structuring element. The enhanced human silhouette image is shown in Figure 17c.

(a) (b) (c)

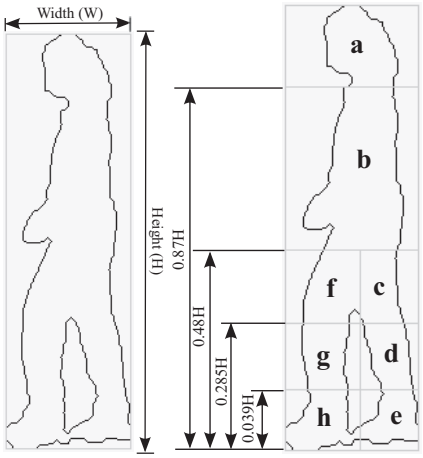
Figure 17: The original, extracted and enhanced image after morphological opening.(a) Original video image. (b) Extracted human silhouette image. (c) Enhanced human silhouette image.

3.3.1.2 *Width and Height Measurement*

Next, the width and height are measured from enhanced human silhouette. These two features will be used later for gait analysis in the later stage. Figure 18a shows the width and height of the silhouette.

3.3.1.3 *Human Silhouette Segmentation*

The enhanced human silhouette is then segmented into eight body segments based on a study of human body proportion by Dempster et al. [34]. Figure 18b shows the eight segments of the body, where a represents head and neck, b represents torso, c represents right upper thigh and hip, d represents right middle thigh, e represents right lower thigh and foot, f represents left upper thigh and hip, g represents left middle thigh and h represents left lower thigh and foot.



(a) (b)

Figure 18: The static gait features which extracted from sequences of images of walking subjects.(a) The width and height of a human silhouette. (b) The eight body segments.

3.3.1.4 *Automatic Silhouette Body Joint Identification*

To extract the body joints that define the human gait, the vertical position of hip, knees and ankles with respect to the body height is estimated by referring to a priori information of the human body proportion.

HIP POSITION With the height of the hip determined, the next step is to find its horizontal position. When a dashed horizontal line is drawn across the hip, there should be only two edges in a normal silhouette as highlighted in orange dots in Figure 19a. The horizontal hip position can then be determined by calculating the width between both edges using the following equation:

$$C_{pos} = C_{rise} + \frac{C_{fall} - C_{rise}}{2} \ ,$$

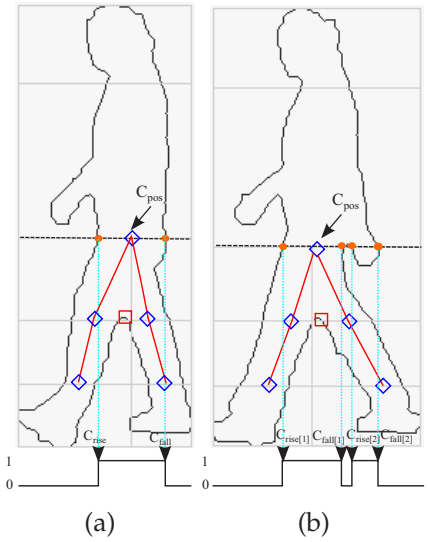
where C_{rise} and C_{fall} are the horizontal positions of the rising edge and falling edge respectively and C_{pos} is the horizontal position of the hip.

However, it happens that there are more than two edges. For example, there are four edges on the same dashed horizontal line as highlighted in orange dots in Figure 19b when one hand swing is detected at the hip height. The horizontal hip position can be determine by finding the width between each edges with respect to the horizontal position of the hip. Also, the width of the hip is always larger than the width of the hand in all cases. Therefore, the horizontal hip position can be determine by calculating the hip width only regardless

that whether one hand or both hands swings are detected at the hip height using the following equations:

$$C_{width}[i] = C_{fall}[i] - C_{rise}[i], \quad \text{for } i = 1, 2\ (,3)$$
$$j = \text{argmax}\, C_{width}[i]\ ,$$
$$C_{pos} = C_{rise}[j] + \frac{C_{fall}[j] - C_{rise}[j]}{2}\ .$$

where $C_{rise}[i]$ and $C_{fall}[i]$ are the no. of horizontal positions of the rising edge and falling edge on the hip respectively; $C_{width}[i]$ is the width of between the $C_{rise}[i]$ and $C_{fall}[i]$ respectively; j is set of points of the given argument for which the given function attains its maximum value; $C_{rise}[j]$ and $C_{fall}[j]$ are the chosen for the width of the hip computation; C_{pos} is the horizontal position of the hip.



(a) (b)

Figure 19: Two edges on the horizontal line (dashed black line) that passes through the hip. (a) Two edges are found across the hip height. (b) Four edges are found across the hip height.

KNEE POSITION By referring to a priori information of the human body proportion, the height of the knees can be found. To find the horizontal position of both knees, a horizontal line is drawn at knee height across human silhouette. For a normal silhouette with no self occlusion, there should be four edges on the image profile along the dashed horizontal line as shown in green dots in Figure 20a. The horizontal knee positions can then be determined by finding the width between two adjacent edges on each leg using the following the equations:

$$K_{leftPos} = K_{leftRise} + \frac{K_{leftFall} - K_{leftRise}}{2}\ ,$$
$$K_{rightPos} = K_{rightRise} + \frac{K_{rightFall} - K_{rightRise}}{2}\ ,$$

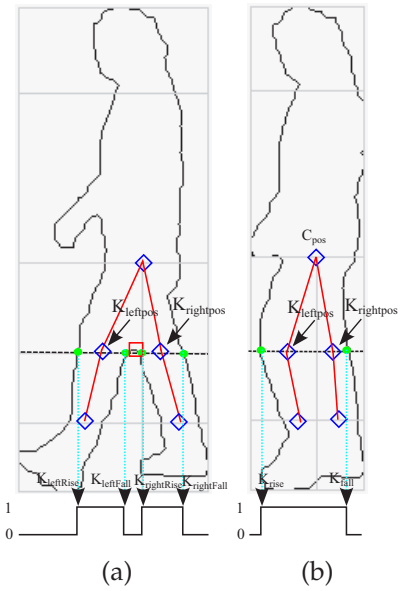
where $K_{leftPos}$ and $K_{rightPos}$ are the horizontal positions of the left and right knees respectively; $K_{leftRise}$ and $K_{rightRise}$ are the horizontal positions of the rising edge on the left and right knees respectively; $K_{leftFall}$ and $K_{rightFall}$ are the horizontal positions of the falling edge on the left and right knee respectively.

For a silhouette with self-occlusion, there will be only two edges on the same dashed horizontal line as highlighted in green dots in Figure 20b. The horizontal knee positions can be determined by computing the width between each edge with respect to the horizontal position of the hip.

$$K_{leftPos} = K_{rise} + \frac{C_{pos} - K_{rise}}{2} \ ,$$

$$K_{rightPos} = C_{pos} + \frac{K_{fall} - C_{pos}}{2} \ ,$$

where $K_{leftPos}$ and $K_{rightPos}$ are the horizontal positions of the left and right knees respectively; C_{pos} is the horizontal position of the hip; K_{rise} and K_{fall} are the horizontal positions of the rising edge and falling edge respectively.



(a) (b)

Figure 20: Image profile along the dashed black horizontal line. (a) Four edges are found across the knees in a normal silhouette. (b) Two edges are found across the knees in a self-occluded silhouette.

ANKLE POSITION To determine the horizontal position of the ankles, the similar technique is employed. If a horizontal line is drawn at ankle height on a normal human silhouette with no self-occlusion, there should be four edges on the image profile along the dashed horizontal line as highlighted in yellow dots in Figure 21a. The hori-

zontal ankle positions can then be determined by using the following equations:

$$A_{leftPos} = A_{leftRise} + \frac{A_{leftFall} - A_{leftRise}}{2} \ ,$$

$$A_{rightPos} = A_{rightRise} + \frac{A_{rightFall} - A_{rightRise}}{2} \ ,$$

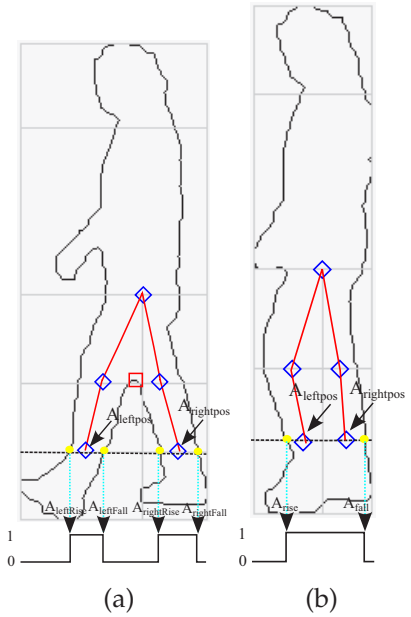
where $A_{leftPos}$ and $A_{rightPos}$ are the horizontal positions of the left and right ankles respectively; $A_{leftRise}$ and $A_{rightRise}$ are the horizontal positions of the rising edge on the left and right ankles respectively; $A_{leftFall}$ and $A_{rightFall}$ are the horizontal positions of the falling edge on the image profile respectively.

For human silhouette with self-occlusion, there will be only two edges on the dashed horizontal line as highlighted in yellow dots Figure 21b. The horizontal ankle positions can then be determined by finding the width between both edges using the following equations:

$$A_{leftPos} = A_{rise} + 0.25(A_{fall} - A_{rise}) \ ,$$

$$A_{rightPos} = A_{rise} + 0.75(A_{fall} - A_{rise}) \ ,$$

where $A_{leftPos}$ and $A_{rightPos}$ are the horizontal positions of the left and right ankles respectively; A_{rise} and A_{fall} are the horizontal positions of the rising edge and falling edge on the image profile respectively.



(a) (b)

Figure 21: Image profile along the dashed black horizontal line. (a) Four edges are found across the ankles in a normal silhouette. (b) Two edges are found across the ankles in a self-occluded silhouette.

3.3.1.5 *Joint Angular Trajectory Calculation*

Figure 22a shows how the joint trajectory is determined. All the joint trajectories are computed by using the following equation:

$$\phi_1 = \tan^{-1}\left(\frac{p2_x - p1_x}{p2_y - p1_y}\right) ,$$

$$\phi_2 = \tan^{-1}\left(\frac{p3_x - p1_x}{p3_y - p1_y}\right) ,$$

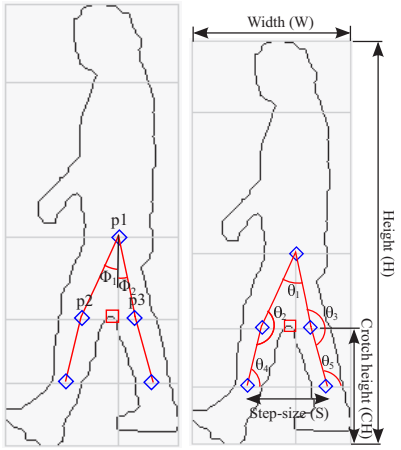
$$\theta = \phi_1 + \phi_2 ,$$

where $p1_x$, $p2_x$, $p3_x$ and $p1_y$, $p2_y$, $p3_y$ are the x-coordinates and y-coordinates of joint p1, p2 and p3 respectively.

3.3.1.6 *Step-size and Crotch Height Measurement*

The Euclidean distance between both ankles is determined to obtain subject's step-size (S). Crotch height (CH) is the Euclidean distance between subject's crotch and the floor is measured. If the crotch height is lower than the knee height, it is reduced to zero as the crotch is considered occluded. Figure 22b shows nine gait features extracted from a human silhouette.

In total, five joint angular trajectories have been extracted. For instance, these angular trajectories for walking sequence from right to left are hip angular trajectory (θ_1), left knee angular trajectory (θ_2), right knee angular trajectory (θ_3), left ankle angular trajectory (θ_4) and right ankle angular trajectory (θ_5).



(a) (b)

Figure 22: (a) Joint trajectory computation. (b) All the extracted gait features.

3.3.2 *Feature Post-process*

In this stage, we present an effective combination of post-processing techniques to process all the extracted gait features to improve the

robustness of gait recognition systems. All the extracted gait features from each walking sequences are first constructed into feature vector before going into feature normalization process. Performance of gait recognition system could be degraded by challenging covariate factors such as subject's apparels and load carrying condition. Therefore, we employ computationally efficient feature selection technique to the problem of extracted gait features selected before channeled into the classification process.

3.3.2.1 *Feature Vector Construction*

The gait features are constructed simultaneously using maximum angular trajectory components and average components are determined during the whole walking sequence. The gait features at each frame in one walking sequence is denoted as with $f_{i=1,2,\ldots,m}^{j=1,2,\ldots,n}$, where m is the number of gait features, and n is the number of frames per walking sequences. All the features $F = \{F_1, F_2, \ldots, F_{18}\}$ for each walking sequence are computed by using the following equation:

$$\begin{aligned} k &= \underset{j \in 1,\ldots,n}{\arg\max}\ \theta_1^j \ , \\ F = \{&\theta_1^k, \theta_2^k, \theta_3^k, \theta_4^k, \theta_5^k, W^k, H^k, S^k, CH^k, A^{\theta_1}, \\ &A^{\theta_2}, A^{\theta_3}, A^{\theta_4}, A^{\theta_5}, A^W, A^H, A^S, A^{CH}\} \ , \end{aligned} \tag{1}$$

where k represents the frame index where the hip trajectory attains its maximum value. When θ_1^k is identified, the corresponding left knee angular trajectory (θ_2^k), right knee angular trajectory (θ_3^k), left ankle trajectory (θ_4^k), right ankle trajectory (θ_5^k), step-size (S^k), width (W^k), height (H^k) and crotch height (CH^k) were also determined. In order to improve the the recognition rate, nine additional features are used. These features are the average of hip trajectory (A^{θ_1}), left knee angular trajectory (A^{θ_2}), right knee angular trajectory (A^{θ_3}), left ankle angular trajectory (A^{θ_4}), right ankle angular trajectory (A^{θ_5}), step-size (A^S), width (A^W), height (A^H) and crotch-height (A^{CH}). It can be computed by the following equation:

$$A^{f_i} = \frac{\sum_{j=1}^{j=n} f_i^j}{n} \ , \tag{2}$$

where A^{f_i} represents average value of i-th gait feature and n represents the total number of frames in the walking sequence.

3.3.2.2 *Feature Normalization*

Scaling data is very important in order to avoid attributes in greater numeric ranges dominating those in smaller numeric ranges. Before sending the feature vectors into classifiers for training and testing, it is required to normalize them in order to approximately equalize

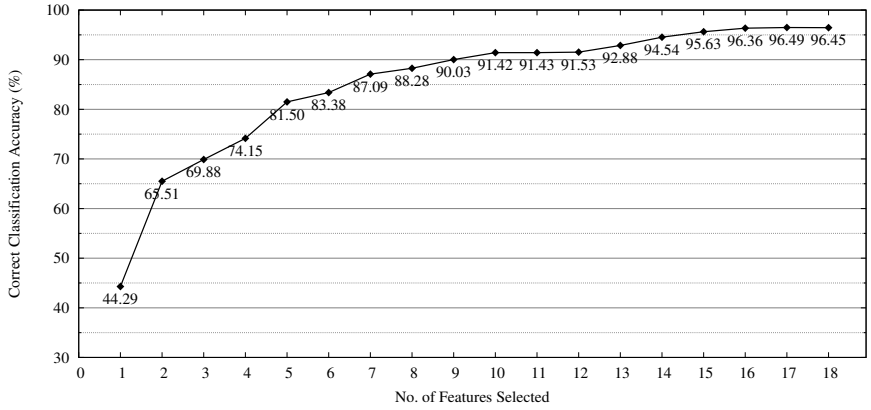


Figure 23: Cumulative recognition rates as the gait features are progressively selected. Average value obtained by ten-fold cross validation is shown. We have omitted the confidential intervals from the figure since they are sufficiently small less than 0.01 %

ranges of the features and make them have approximately the same effect in the computation of similarity. Therefore, in this work, we apply a very common normalization technique via linear scaling [1]. We normalize the training and test data with the same model. For the test data set, we perform the data scaling using the same scaling parameters. It is applied to normalize each feature component to the range between 0 and 1 by using the following equation:

$$\overline{F}_i = \frac{F_i - F_i^{min}}{F_i^{max} - F_i^{min}} \quad . \tag{3}$$

where $\overline{F}_i$ represents the normalized value for extracted gait features data, F_i represents the value of i-th component of the feature vector F to be normalized, F_i^{min} and F_i^{max} represent the minimum and maximum values among the extracted gait features, respectively.

3.3.2.3 *Feature Selection Analysis*

In our work, feature selection analysis is to choose the most informative and relevant extracted gait features and achieve the possible highest Correct Classfication Rate (CCR) on the reduced feature set. In the feature selection analysis, Ranker [47] is used to rank the features by their individual evaluations which helps to identify those extracted gait features that contribute positively in the recognition process. We use SVM classifer for feature ranking in this stage. Based on the average rank scores obtained in Table 1, it ranks attributes by their individual evaluations. The experiment was thus designed to assess the change of CCR as features are progressively selected is shown in Figure 23. The selection of seventeen or eighteen features

have exhibited the same positive contribution. Thus, only seventeen features are used in the this work as shown below:

$$F = \{\theta_1^k, \theta_2^k, \theta_3^k, \theta_4^k, W^k, H^k, S^k, CH^k, A^{\theta_1}, A^{\theta_2}, A^{\theta_3}, A^{\theta_4}, A^{\theta_5}, A^W, A^H, A^S, A^{CH}\} \tag{4}$$

3.3.3 *Classification Process*

After the feature post-process, multiclass SVM classification technique is employed for gait recognition. For the SVM technique used in this work, we refer to the description by C.J.C. Burges [24] and implement the SVM experiments by LIBSVM package [25], which implements the one-against-one approach [63] for multi-class classification.

For this work, experiments were carried out to examine the effects on kernel functions - Linear (Ln), Polynomial (Poly) and Radial Basis Function (RBF). The kernel's parameters such as c (the cost of parameter), γ (gamma in kernel function), coef0 (coefficient in kernel function), d (degree of kernel function) were trained and set for the SVM different kernel types performance evaluation experiments as shown in Table 2. These parameters were chosen based on several test evaluations of our proposed approach and should not be considered the optimal configurations.

3.4 EXPERIMENTS

Five experiments were carried out in this study. In the first experiment, we present an evaluation of gait recognition performance using SVM with different kernels. Here, we also examine classification performance of different kernels. In the second experiment, we evaluate the inter-subject variation on gait recognition performance. In the third experiment, we evaluate the impact of different covariates for gait analysis. In the fourth experiment, we examine the effect of occlusions on the performance of gait recognition. To verify the effectiveness of the proposed approach, we evaluate the performance of our proposed approach with respect to the conventional benchmark approaches in the fifth experiment.

3.4.1 *Datasets*

The experiment was carried out for eleven subjects walking in parallel towards a static camera with fifteen covariate factors provided by the SOTON covariate database [97]. Each subject was captured wearing a variety of clothes (rain coat, trench coat and normal), footwear (flip flops, bare feet, socks, boots, trainers and own shoes) and carrying various bags (barrel bag slung over the shoulder or carried by hand on shoulder, hand bag held in hand, and rucksack). They were also

Table 1: Feature selection analysis

No. of Features Selected	Feature Attribute	Average Rank
1	CH	1+-0
2	A^{θ_3}	2+-0
3	θ_3	3+-0
4	A^{CH}	4+-0
5	A^{θ_2}	5+-0
6	θ_2	6+-0
7	A^{S}	7.2+-0.4
8	S	7.8+-0.4
9	A^{θ_1}	9+-0
10	A^{θ_4}	10.1+-0.3
11	θ_1^{max}	11.3+-0.64
12	θ_4	11.6+-0.49
13	W	13+-0
14	A^{θ_5}	14.2+-0.4
15	A^{H}	14.9+-0.54
16	A^{W}	16.1+-0.54
17	H	16.8+-0.4
18	θ_5	18+-0

recorded walking at different speed (slow, fast, and normal speed). For each subject, there are approximately twenty sets of walking sequences from right to left and vice-versa way on a normal track. The dataset consists of 3178 walking sequences from 11 subjects spanning 15 covariates. In total, there are 3,178 walking sequences that are used for training and testing process.

3.4.2 *Experimental Setup*

The experiment was carried out on SVM classification technique with various optimization parameters that obtained during the training. Experiments are also conducted using leave-one-out strategy taking all 11 subjects. In this case, ten training subjects and one test subject for each division, which results in 11-fold cross validation. The results were averaged over all the training and cross-validation data. The performance evaluation was in terms of Correct Classification Rate (CCR).

Table 2: SVM Kernel's Parameter Setup

Kernel Type	SVM Kernel's Parameters			
	c	γ	coef0	d
Ln	48	-	-	-
Poly	32	1.5	1	3
RBF	45	2	-	-

Table 3: Performance evaluation of SVM classifier

	SVM Kernel Type		
	Ln	Poly	RBF
CCR (%)	91.1	95.7	96.5

Considering in real-world gait recognition scenarios, the covariate type in a query gait sequence is unknown, and it is unrealistic to train a model with all possible covariate types in real world. We used normal walking sequences for training and other covariate factors walking sequences for testing. Individuals are unique in the gallery and each probe set, and there are no common sequence among the gallery set and all probe sets. To observe its effectiveness, some examples of subjects with different apparels and the identified body joints are shown in Figure 24.

3.5 RESULTS AND DISCUSSIONS

We describe in this section the details of our experiments and analyze the results.

3.5.1 *Performance Evaluation of SVM Classifier*

From the experimental results shown in Table 4, we noticed that the non-linear SVM (RBF or Poly kernel) outperforms linear SVM (Ln kernel). By using a kernel function the non-linear separable vector is mapped into a high-dimensional feature space and thereby makes it becomes linearly separable [95]. Since SVM with RBF kernel gave the best recognition rate during the experiment on different SVM kernel types, the experiments for covariate and person-independent gait analysis were carried out using only SVM with RBF kernel.

Table 4: Performance evaluation of SVM classifier

	SVM Kernel Type		
	Ln	Poly	RBF
CCR (%)	91.1	95.7	96.5

Table 5: Confusion matrix of 11-person gait recognition using SVM (RBF)

	P.1 (%)	P.2 (%)	P.3 (%)	P.4 (%)	P.5 (%)	P.6 (%)	P.7 (%)	P.8 (%)	P.9 (%)	P.10 (%)	P.11 (%)
P.1	**98.5**	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	1.1
P.2	0.0	**91.8**	3.5	0.4	1.1	0.4	0.0	2.8	0.0	0.0	0.0
P.3	0.0	2.8	**92.9**	0.0	0.0	1.5	0.0	2.8	0.0	0.0	0.0
P.4	0.0	0.8	0.0	**97.4**	0.0	0.8	0.0	0.8	0.2	0.0	0.0
P.5	0.0	0.0	0.4	0.0	**97.2**	0.7	0.0	1.4	0.0	0.3	0.0
P.6	0.0	0.0	1.1	1.8	0.7	**91.1**	0.0	4.6	0.7	0.0	0.0
P.7	0.0	0.0	0.0	0.0	0.0	0.0	**100**	0.0	0.0	0.0	0.0
P.8	0.0	0.0	2.0	0.3	0.3	1.4	0.0	**96.0**	0.0	0.0	0.0
P.9	0.0	0.0	0.0	0.4	0.4	0.7	0.0	0.0	**98.5**	0.0	0.0
P.10	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	**99.0**	0.0
P.11	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	**99.6**

3.5.2 *Gait Recognition Confusion Matrix*

In this experiment, the feature vectors generated from individual subject's gait was divided into disjoint training and test sets for ten-fold cross validation. The confusion matrices of the average recognition rate are presented in Table 5. From this table, it can be observed that our proposed approach was able to achieve above 91.1% CCR for all subjects from person 1 until person 11 i.e., P.1, P.2, . . . , P.11. The CCR reached 100% for person 7(P. 7) which indicated that person correctly identified from the testing walking sequence with various covariate factors.

3.5.3 *The Effect of Covariate Factors*

To further study the effects of different clothing types, we divided the probe experiment set into 4 groups, i.e., footwear, clothing, carrying condition and walking speed. Table 6 lists the covariates used in the dataset. We used normal walking sequences for training and other covariate factors walking sequences for testing. Individuals are unique in the gallery and each probe set, and there are no common sequence among the gallery set and all probe sets is listed in Table 7. The classification performance is assessed by matching the probe set against the gallery set.

The performance of our approach can be seen in Figure 25. For the proposed method, we have the following observations:

Table 6: List of covariate factors used in SOTON database

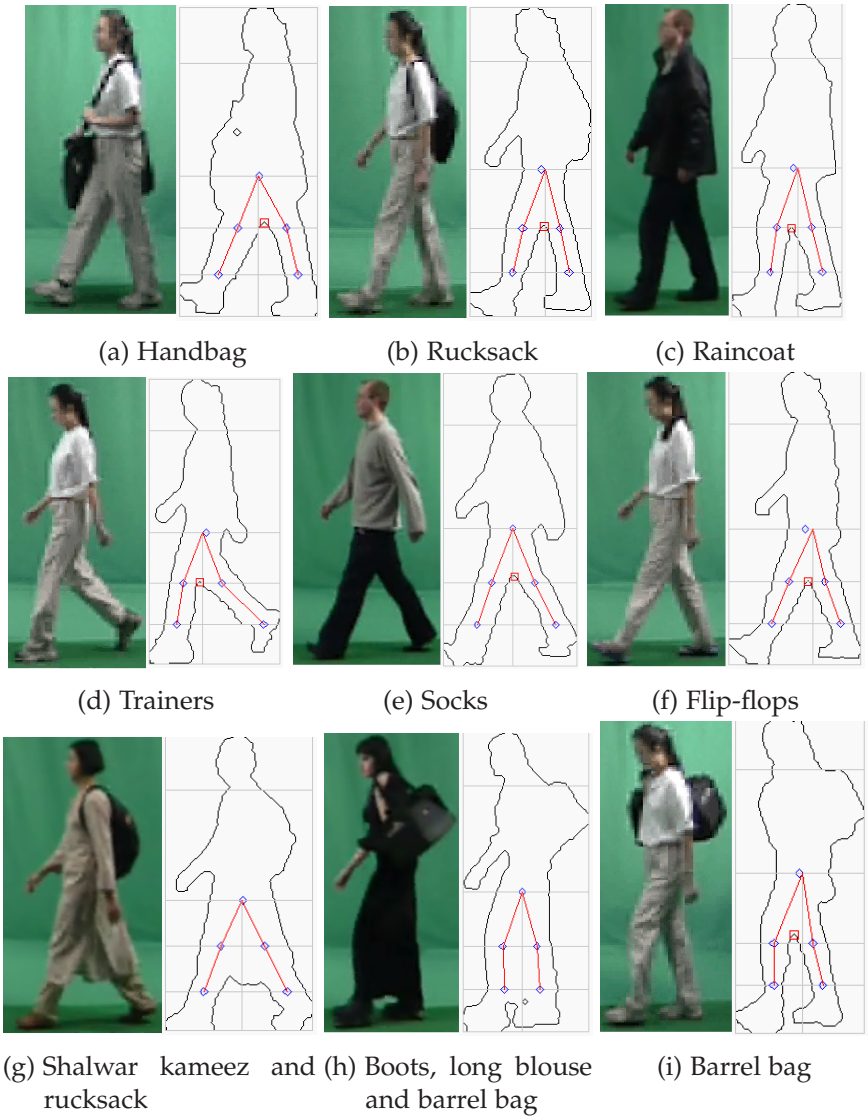
Covariates ID	Descriptions
000	subject walking in flip flops
001	subject walking with bare feet
002	subject walking in socks
003	subject walking in boots
004	subject walking in trainers
005	subject walking in own shoes
006	subject wearing rain coat
007	subject wearing trenchcoat
008	subject carrying hand bag (held in hand)
009	subject carrying barrel bag (slung over shoulder, bag at hip height)
010	subject carrying barrel bag (carried by hand on shoulder)
011	subject carrying rucksack
012	subject walking slowly
013	subject walking quickly
014	subject walking at normal speed

Table 7: The SOTON database

Dataset	Number of samples	Variations
Gallery	241	Normal condition
Probe A	448	Clothes
Probe B	886	Carrying conditions
Probe C	680	Walking speeds
Probe D	923	Footwear

For Group I which only involves the footwear covariates like flip-flops, bare feet, socks, boots, own shoes and trainers, the group average recognition rate is around 97.5%. Out of these footwear covariate type, the human gait is observed no impact on gait recognition when subject walk with bare feet as the recognition rate achieves 100.0%.

Variations in clothing alter an individual's appearance, making the problem of human gait identification much more difficult. In Group II which only involves the clothing covariates (e.g., rain coat and trench-coat), the recognition task becomes more difficult. However, based on the proposed method, the performance is still moderate, with an average recognition rate more than 86.0%.



(a) Handbag (b) Rucksack (c) Raincoat

(d) Trainers (e) Socks (f) Flip-flops

(g) Shalwar kameez and rucksack (h) Boots, long blouse and barrel bag (i) Barrel bag

Figure 24: Examples of body joint identification on human silhouettes with different apparels in the SOTON dataset. Cases (a) (b) and (c) present varying carrying conditions; cases (f) and (g) varying clothing conditions; cases (d) and (e) a combination of the former; cases (h) and (i) varying footwear. Notice the output of our model-based gait extraction on the side.

In Group III, four different covariate factors related to carrying conditions (e.g., subject carrying handbag, barrel bag slung over shoulder or carried by hand, and rucksack) are used in the experiment evaluation. For our proposed approach, gait features were extracted only from the lower body part. The achieved correct classification rate for the subject carrying barrel bag slung over shoulder or carried by hand and rucksack are almost the same with reported rates of 94.7%, 97.5% and 95.7% respectively. However, for the case of subject carrying handbag which is occluding the lower part of human body, the recognition rate drops slightly to 84.8%.

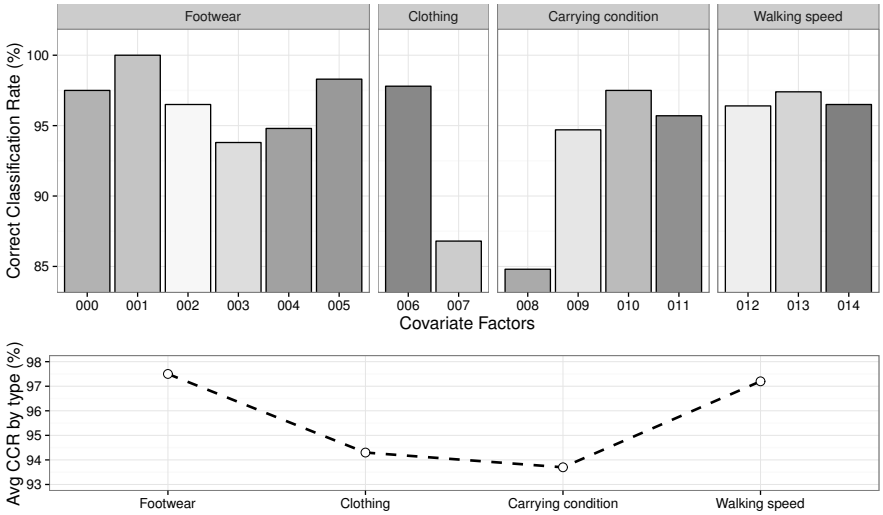


Figure 25: Covariate analysis experiment

Table 8: The effect of occlusions in gait recognition.

Experiment (Exp.)	Covariate Factors	No. Walking Sequences	Occlusion Ratio (%)	CCR (%)	
				w/o self-occlusion	w/ self occlusion
1	Clothes	689	39.9	96.4	94.3
2	Carrying conditions	1127	30.9	95.1	93.7
3	Walking speeds	921	17.2	96.6	97.2
4	Footwear	1164	17.0	97.1	97.5
5	All	3178	26.9	95.7	96.5

For Group IV, we investigated the impact of walking speed variations (e.g., slowly, normal and quickly) on gait recognition. The achieved average recognition rate for this group was 97.2%. This high accuracy is because in SOTON covariate database, the range of walking speed variations are significantly small and limited. We should evaluate our method for the performance of gait recognition under larger walking speed variation dataset.

Out of the covariate factors like footwear, clothing types, carrying condition and walking speed, it has been demonstrated that clothing is the most challenging covariate factor for our proposed approach. However, our proposed approach was still able to achieve at least 93.7% average CCR for all the experiments conducted. This shows the effectiveness of our proposed approach.

3.5.4 *The Effect of Occlusions*

Human gait recognition became more complicated when the existence of self-occlusion in the human body. This issue would seem to imply a critical importance of gait analysis in self-occlusion gait recognition. Therefore, we carried out the experiments to examine

Table 9: Correct classification rate (%) of different methods in the SOTON database [97].

Dataset	Method	Covariate conditions	#People	#Sequences	Classifier	CCR (%)
[20]	Method in Ref. [20]	11	10	440	KNN	73.4
	Our approach					78.0
[90]	Method in Ref. [90]	4	10	180	SVM-Ln	86.0
	Our approach					94.4
[85]	Method in Ref. [85]	13	11	2722	SVM-Ln	84.0
	Our approach					95.0
Full	**Our approach**	15	11	3178	SVM-RBF	96.5

the effect of occlusions in gait recognition as shown in Table 8. This experiment used to measure the overall performance of gait recognition by our proposed approach for knee and ankle positions under self-occlusion and without self-occlusion condition. In the case of clothes and carrying conditions, the knee and ankle positions are frequently occluded. Under this condition, identification of lower body joint points maybe complicated in the presence of the heavy occlusion region under lower body parts due to the subject apparels. Thus, this may seriously affect the performance of the trained classifier. Specifically, the results in Section 3.5.3 showed that clothes and carrying condition changes seem to affect gait more than walking speed and footwear changes. Hence, if we removed out the occluded frames in the walking sequences from the respective subject, we can see an improvement of gait recognition performance. In contrary, in the case of walking speed and footwear, the knee and ankle positions are not heavily occluded. Thus, it does not seriously degrade classification performance. However, if we removed out the occluded frames for these conditions, we can see a slight decrease of performance of gait recognition due to less number of training samples. In the case of evaluation using all covariate factors at the same time by removing the self-occluded frames from the walking sequences, it does not improve the overall performance of the gait recognition. This might be because our proposed approach could identify the lower trajectories of body points even under occluded silhouette either from self-occluded or occluded by apparels. It may remove out the important feature that can classify the respective subject in this database before going to the classification process.

3.5.5 *Evaluation on Benchmark Methods*

To evaluate this database, we apply recent three benchmark methods by Bouchrika et al. [20], Pratheepan et al. [90], and Ng et al. [85]. In comparison with them, we evaluate our approach under the same

experimental settings and data subset. The details can be found in the reference papers.

Table 20 summarizes the comparison with other conventional approaches that use SOTON covariate database, our approach outperforms the results obtained by Bouchrika et al. [20], Pratheepan et al. [90] and Ng et al. [85] after comparing the number of subjects, the number of covariate factors and the number of walking sequences that were used for training and testing. The inferior result by Bouchrika et al. is due to the requirement to manually label model template to describe the joints' motion. On contrary, our results are performed better than Pratheepan et al. as we do not incorporate the selection or estimation of the gait cycle. It is also obvious from the results that our approach result outperforms Ng et al. as their approach have difficulty in extracting gait feature of the silhouette and the erroneous extraction of the gait feature due to self-occluding silhouette model.

3.6 SUMMARY

We have presented a robust approach for extracting the human gait features from human silhouette images. The gait features are extracted from the human silhouette by determining the body joints from the body segments based on a priori knowledge of human body proportion. Once the body joints have been identified, the joint trajectories can be computed using a straightforward approach. This approach has shown to be more effective as it is capable to identify the body joints from self-occluded human silhouettes. Our main goal for proposed feature post-process approach is to mitigate the effect of variation in covariate factors which complicate the person's gait recognition process. Our experimental results demonstrate that the proposed approach significantly outperforms the existing conventional techniques. In addition, the higher recognition rate also showed that the proposed approach is robust and can perform well under different covariate factors.

Part III

IMPROVING HUMAN GAIT RECOGNITION USING FEATURE SELECTION

In this part, the problem of selecting a subset of the most relevant gait features for improving gait recognition performance is investigated. This is achieved by using proper feature selection technique to find the relevant features and discard redundant and irrelevant gait features. A comparative study on feature selection methods in the context of gait identification under the effect of various covariate factors is presented.

4

FEATURE SELECTION FOR GAIT RECOGNITION

4.1 OVERVIEW

Robust and reliable human identification for surveillance and access control has become highly sought these days. Recently, human gait identification (i.e., identifying individuals by the way they walk), has gained considerable attention due to its potential applicability in promising applications such as forensics, security, immigration, surveillance [17]. A person's gait can be measured unobtrusively at a distance and without the need for any subject cooperation or contact for data acquisition [73]. However, the presence of covariate factors such as different clothing types, carrying conditions, walking surfaces, etc., can seriously complicate the task. Clothing, for instance, can occlude a significant amount of gait features and make human recognition difficult. Since the location of occlusion may differ for different covariate factors, relevant gait features may become irrelevant when the covariate factor changes, and exploiting occluded gait features can hinder the recognition performance. Therefore, feature selection has become an important step to make the analysis more manageable and to extract useful information for the gait classification task. Nevertheless, although feature selection is often used in order to identify the relevant body parts, to the best of our knowledge, a comparative analysis of feature selection techniques in gait recognition is seldom addressed. In this work, we present an empirical approach to evaluate the degree of consistency among the performance of different selection algorithms in the context of gait identification under the effect of various covariate factors. First, a model-based framework for extracting informative gait features is introduced, then, an extensive comparative analysis of feature selection approaches in gait recognition is carried out. We perform a statistical study via ANOVA and mixed-effects models to examine the effect of six popular selection feature methods across classifiers and covariates. In addition, we systematically compare the selected feature subsets and the computational cost of the different selection approaches. The implemented method addresses the problem of feature selection for gait recognition on two well-known benchmark databases: the SOTON covariate database and the CASIA-B dataset, respectively. The investigated approach is able to select the discriminative input gait features and achieve an improved classification accuracy on par with other state-of-the-art methods.

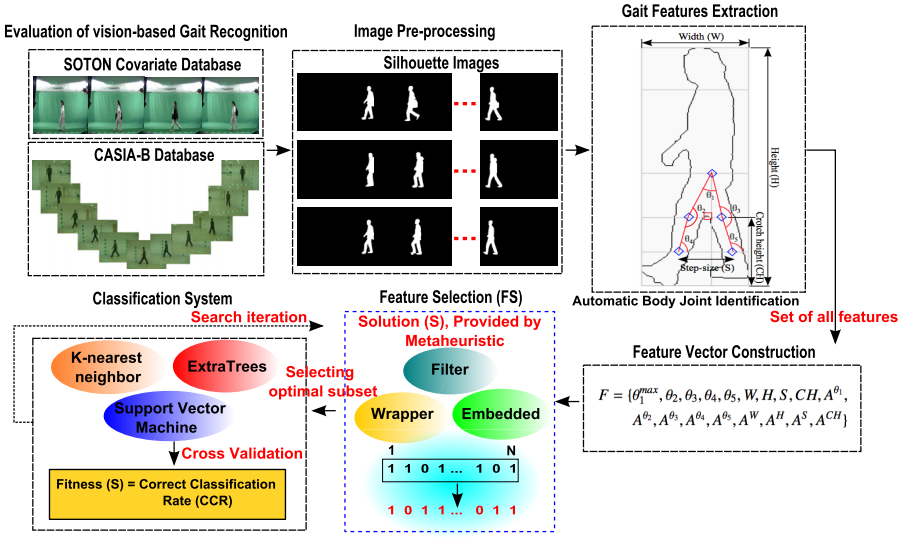


Figure 26: A simple scheme of how gait features are selected from the evaluation databases. Candidate subsets proposed by the metaheuristic are evaluated by means of a classifier and cross validation to obtain the corresponding average Correct Classification Rate (CCR).

4.2 METHODOLOGY

In this section, we present an empirical approach to evaluate the degree of consistency among the performance of different selection algorithms in the context of gait identification, which are designed to find the best gait feature subset in order to maximise the final accuracy of the classification process under the effect of various covariate factors. Fig. 26 shows the general operational idea; we can observe a simple framework of how gait features are extracted from the SOTON covariate and CASIA-B database and how the feature subsets are evaluated. In the following, we detail the processing steps, feature selection and classification methods that we considered for this comparative study. Specifically, our experimental work investigates the following points:

1. whether filter and embedded methods with their low computational cost advantages as compared to wrapper methods can effectively improve the performance in feature selection tasks;

2. whether search-based methods perform better than statistical approaches for feature selection, and the trade-offs in terms of computational cost;

3. to what extent the alleged improvement in classification accuracy that feature selection brings generalises to the use of different classifiers and on different datasets.

4.2.1 *Gait Pattern Extraction and Representation*

The original human silhouette images are obtained from the SOTON covariate database [97]. This database was used to evaluate the recognition rate of the walking subjects with different covariate factors. In most of human silhouette images, shadow is chronically found near the feet. This will hinder the gait feature extraction as it interferes with the body joint identification. Morphological opening is initially applied in order to reduce the background noise on the extracted human silhouette images. The result of the morphological operations is an enhanced silhouette that is then measured for its width and height. Next, each of the enhanced human silhouettes is segmented into eight body segments based on anatomical knowledge. The lower body joints that define the pivot points in human gait are automatically identified and the joint trajectories are computed. After that, step size and crotch height are measured. In total, five joint angular trajectories are extracted. There are various properties of gait that might serve as recognition features. For instance, the angular trajectories for a walking sequence from right to left are hip angular trajectory (θ_1), left knee angular trajectory (θ_2), right knee angular trajectory (θ_3), left ankle angular trajectory (θ_4) and right ankle angular trajectory (θ_5). To construct the feature vector, maximum hip trajectory (θ_1^{max}) is determined during a walking sequence from right to left. When θ_1^{max} is identified, the corresponding left knee angular trajectory (θ_2), right knee angular trajectory (θ_3), left ankle trajectory (θ_4), right ankle trajectory (θ_5), step size (S), width (W), height (H) and crotch height (CH) are also determined. To obtain optimal performance, nine additional features are used. These features are the average of the local maxima detected for hip trajectory (A^{θ_1}), left knee angular trajectory (A^{θ_2}), right knee angular trajectory (A^{θ_3}), left ankle angular trajectory (A^{θ_4}), right ankle angular trajectory (A^{θ_5}), step-size (A^S), width (A^W), height (A^H) and crotch-height (A^{CH}). Thus, eighteen gait features are used to construct the feature vector as it is shown below:

$$\begin{aligned}\vec{X} = \{&\theta_1^{max}, \theta_2, \theta_3, \theta_4, \theta_5, W, H, S, CH, A^{\theta_1}, \\ &A^{\theta_2}, A^{\theta_3}, A^{\theta_4}, A^{\theta_5}, A^W, A^H, A^S, A^{CH}\}.\end{aligned} \tag{5}$$

These specific features have been chosen based on a preliminary study; they all pertain to the lower body area that is expected to be more relevant to the gait pattern and more robust to covariate factors. Moreover, their extraction is not computationally expensive; for more details about the computation of these gait features, the interested reader may refer to our previous work [119, 122].

4.2.2 *Statistical Methods for Gait Feature Selection*

Feature selection aims to select a relevant subset of features, fulfilling certain objectives. In pattern recognition, a natural metric would be the classification accuracy or inversely the Bayes classification error. Since the Bayes error can not be minimised analytically, several alternative statistics that are easier to evaluate have been proposed [44]. These statistics are gauged in a preprocessing step, and, based on their value, it is possible to score and rank the features accordingly, in order to discard those that do not seem to carry relevant information for the classification task. Here we focus on typical approaches to gait feature selection using *Analysis of variance (ANOVA)* and *Mutual Information (MI)*, which are introduced as follows.

4.2.2.1 *Analysis of Variance (ANOVA-Fclass)*

The one-way ANOVA test is based on F statistic. The greater the F ratio is, the stronger the discriminative capability of the gait feature is. As a preliminary step, feature measurements are grouped by the target label; that is, individuals in the training set are treated as levels of a grouping factor, and by performing ANOVA we want to test if the average feature value differs across targets. The F ratio is then calculated for each of the gait feature x in this study as follows:

$$F = \frac{MS_{\text{between targets}}}{MS_{\text{within targets}}} = \frac{\sum_i n_i(\bar{x_i} - \bar{x}_G)^2/(k-1)}{\sum_{ij}(x_{ij} - \bar{x}_i)^2/(N-k)} \tag{6}$$

where N is the total number of samples, k is the number of targets or subjects, $\bar{x}_i$ is the average score of the considered gait feature for the i_{th} subject, $\bar{x}_G$ is the grand mean of feature x considering all subjects and all measurements, x_{ij} is the j_{th} measurement of feature x on the i_{th} subject, and n_i is the number of measurements for the i_{th} subject.

4.2.2.2 *Mutual Information (MI)*

Related work [44] has already considered feature selection for gait recognition from the point of view of information theory. In information theory, mutual information measures the statistical dependence between two variables [77]. Mutual information reveals the amount of information carried by one random variable, X, about another random variable, Y. It can be computed for each of the extracted gait features according to the following formula:

$$MI(X,Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \left(\frac{p(x,y)}{p(x), p(y)} \right) \tag{7}$$

where X represents the values of a gait feature, Y represents the subjects' labels, $p(x)$ and $p(y)$ are probability mass functions for the variables X and Y respectively, and $p(x,y)$ is the joint probability mass function.

In practice, probability densities, entropies and mutual information scores are estimated using non-parametric methods based on k-nearest neighbors distances [65].

4.2.3 *Search-Based Methods for Gait Feature Selection*

All aforementioned methods allows one to rank the features according to their relative importance or information content; it only remains to be decided how many of the top k features to retain in order to maximise the gait classification accuracy. In our implementation, instead of setting a pre-defined threshold, we select the best value of k by cross-validation. However, the subset selection problem can also be cast as a black-box optimisation problem and tackled by means of search-based heuristics. Since our focus is on feature selection rather than algorithm design, before devising yet another fancy metaheuristic [100] we prefer to experiment with the prototypical ones. In this section, we focus on a local search approach to gait feature selection using a simple Hill-climber.

4.2.3.1 *Hill-climbing Strategies (HC)*

A *hill-climbing algorithm* (or climber) is a basic local search strategy that guides through the search space by allowing only non-deteriorating moves [10]. Given an initial configuration called *starting point*, a traditional climber iteratively moves to better neighbors, until it reaches a local optimum. This search mechanism also known as the *iterative improvement*, encompasses several variants, depending on how the neighbourhood is explored and when a candidate neighbour is accepted to replace the current solution. Two strategies, *first* and *best* improvement, constitute the most widely used *pivoting rules* that are briefly discussed hereafter. These rules define how to select a better neighbor from a non-locally-optimal configuration [118]. More precisely, the *best-improvement* strategy (or *greedy hill climbing*) consists in selecting, at each iteration, the neighbor that achieves the highest fitness. This requires to generate and evaluate the whole neighborhood at each step of the search, which can imply high computational cost if an incremental evaluation of all neighbors cannot be performed. Conversely, the first-improvement strategy accepts the first evaluated neighbor that satisfies the moving condition. This avoids the systematic enumeration of the entire neighborhood and, possibly, it also allows for more robust results. In fact, given a fixed solution of departure, as long as the neighbourhood is visited in a randomised order, a first-improvement hill-climber is a stochastic local search algorithm. In contrast, a best-improvement hill-climber is a deterministic procedure that is guaranteed to remain in the basin of attraction of the local optimum to which the initial solution belongs.

For our tests we implement a randomised first-improvement (HC-First), outlined in Algorithm 1. Feature subsets are binary-encoded: each bit indicates whether or not the corresponding feature has been selected. The performance of each candidate subset is evaluated in terms of average classification accuracy from cross-validation. The climber departs from the full set of gait features as an initial solution, and it explores the neighborhood induced by the single bit-flip mutation; in other words, it tries to improve the current subset by iteratively adding or removing one feature at a time. The search stops when no further improvement is possible or, in the worst case, after 100 iterations of the pivot rule. We refer to [10] for a comparison of climbers search strategies.

Algorithm 1: First-Improvement Hill-Climber (HC-First)

Input: $x_0 = \{1\}^n$ vector of initial gait features (1=selected, 0=removed)
t_{max} max number of selections or removals (bitflips)
$f() : \{0,1\}^n \mapsto \mathcal{R}$ cross-validated CCR with selected features

Output: $x = \{0,1\}^n$ vector of selected gait features

$x \leftarrow x_0$ `// init current selection`
$t \leftarrow 0$
repeat
  $x' \leftarrow$ flip a bit of x at random w/o replacement `// mutate`
  $t \leftarrow t + 1$
  if $f(x') \geqslant f(x)$ **then**
    $x \leftarrow x'$ `// accept if not worse`
  end
until $f(x) > f(x')\ \forall x'$ *or* $t = t_{max}$

4.2.4 *Embedded Methods for Gait Feature Selection*

Embedded methods differ from other feature selection methods in the way feature selection and learning interact. In contrast to filter and wrapper approaches, in embedded methods the learning part and the feature selection part can not be separated - the structure of the class of functions under consideration plays a important role. In this section, we consider the use of variable importance from totally randomized trees (TRT) and recursive feature elimination from a linear support vector machine (RFECV) as embedded approaches to gait feature selection. We have also experimented with sparsity inducing techniques such as L1 regularisation, but with inconclusive results due to the non-sparse nature of our data.

4.2.4.1 *Variable Importance from Totally Randomized Trees (TRT)*

Decision trees recursively partition their input data by performing binary splits on the input features. Splits are chosen so that the resulting partitions become increasingly homogeneous, according to an impurity measure such as Entropy or Gini coefficient. Therefore, the Mean Decrease in Impurity (MDI) obtained when splitting on a given feature X_j can be used to measure its importance $\mathrm{Imp}(X_j)$, as originally proposed by Breiman [21]. This generalises to a forest of M random trees by summing up the impurity decreases for all the splits in a tree and for all the trees in the forest, according to the formula [22, 23]:

$$\mathrm{Imp}(X_j) = \frac{1}{M} \sum_{m=1}^{M} \sum_{t \in \varphi_m} \mathbb{1}(j_t = j) \left[p(t) \Delta i(s_t, t) \right] \tag{8}$$

where the importance of feature X_j in predicting Y is calculated by adding up, for all splits t where X_j is used $\mathbb{1}(j_t = j)$, the impurity decreases $\Delta i(s_t, t)$ weighted by the proportion of samples $p(t)$ in node t; the importances are then averaged over all the φ_m trees in the forest, $m \in \{1, \ldots, M\}$.

Louppe et al. [76, 75] have shown how, in the asymptotic conditions of an infinitely large training sample and infinitely large ensemble of fully-developed trees, the MDI feature importances expressed by equation 8 can be interpreted in terms of mutual information between the target labels Y and the set of predicting features X_j, $j \in \{1, \ldots, p\}$. In practice, to mitigate errors due to empirically estimating impurity on a finite training set, we bound the depth of the trees. Moreover, to mitigate the masking effect among correlated predictors, we employ totally randomised trees, i.e. trees in which each node t is partitioned using a feature j selected uniformly at random in $\{1, \ldots, p\}$.

4.2.4.2 *Recursive Feature Elimination (RFECV)*

The objective of recursive feature elimination (RFE) is to explore features by recursively considering smaller and smaller sets of features. To this end, we employ a SVM classifier with a linear kernel, where we take the absolute value of the linear coefficients as an indication of feature importance. Notice that input features undergo a standardisation procedure so that they all display zero mean and unitary standard deviation. Each time the classifier is trained on the input set of features, weights (that is, coefficients) are assigned. Then, the feature with the absolute smallest weight is pruned from the current set of features and the process is iteratively repeated on the pruned set until an optimal number of features to select is eventually achieved. This optimal number is chosen via cross-validation. We refer to [46] for details.

4.2.5 *Joint Mutual Information Maximisation (JMIM)*

Several state-of-the-art methods for feature selection make use of mutual information between the selected features and the target class labels as a surrogate for classification accuracy; based on mutual information estimation and selection heuristics, such methods can rank or pick the relevant features without the computational burden of training a classifier. For our comparison, we implement the recent Joint Mutual Information Maximisation (JMIM) proposed in [15]. The idea behind this method is to iteratively build a good feature subset by selecting features that are maximally informative w.r.t. the target class and complementary w.r.t. the features already selected. Such an iterative greedy procedure is sketched in Algorithm 2. As a stopping criterion, we observe that the joint mutual information of the subsequently added features monotonically decreases to zero, therefore we decide to stop when it becomes ill-defined. Again, mutual information is estimated using k-nearest neighbours.

Algorithm 2: Forward greedy search

Input: initial set of n features F, target class vector C
Output: subset with the selected features S

$S \leftarrow \emptyset$
foreach $f_i \in F$ **do** `// computation of MI with target class`
  compute $I(C; f_i)$
end
$f^0 \leftarrow \operatorname{argmax} f_i \in F I(C; f_i)$ `// first feature: max MI`
$F \leftarrow F \backslash \{f^0\}$
$S \leftarrow \{f^0\}$
repeat `// subsequent features`
  $f^* = \operatorname{argmax} f_i \in F \backslash S \min_{f_s \in S} I(f_i, f_s; C)$ `// max min joint MI`
  $F \leftarrow F \backslash \{f^*\}$
  $S \leftarrow S \cup \{f^*\}$
until $|S| = n$ *or* $\max_{f_i \in F \backslash S} \min_{f_s \in S} I(f_i, f_s; C) \leqslant 0$

4.3 EXPERIMENTS

In this section, we carry out several experiments to evaluate the effectiveness of the investigated approaches. We first give a brief description of datasets and experimental settings. Then, in the following section, we empirically compare the considered methods among themselves and with other state-of-the-art feature selection algorithms.

4.3.1 *Design of experiments*

The considered feature selection algorithms are paired with three well-known classifiers (KNN [31], SVM [30] and ExtraTrees [40]) and applied to two well-known datasets. The final classification performance after feature selection is evaluated in terms of Correct Classification Rate (CCR). All our experiments, including classifiers training, feature selection, and parameters tuning, are conducted using scikit-learn[1] [87]. Hereby, the SVM implementation is based on the LIBSVM package [25]. Regarding the classifiers parameters, for KNN we perform cross-validation on each dataset to select the number of neighbors, with $K \in [1, 3, 5]$. We train the SVM with a radial-basis-function kernel that we tune via cross-validation and grid search separately on each dataset; we pick the penalty constraint $C \in [50, 100, 150]$ and the kernel exponent $\gamma \in [0.2, 0.5, 1, 2]$. As for ExtraTrees, in our experiments we use forests of 1000 ExtraTrees in the final classification models and 300 trees in the evaluation step of the search-based feature selection methods. We only tune two tree hyper-parameters, namely the ratio of the random subset of features that are considered at each split in $[0.4, 0.5, 0.6, 0.7, 0.8, 0.9]$, and the minimal number of observations required in a leaf node in $[1, 3]$. The best configuration for each experiment is obtained by grid search with cross-validation.

4.3.2 *Database Description*

Towards robust gait recognition, the experimental analysis depends largely on a sufficient number of training and testing samples. Hence, large-scale gait databases are essential for our experiments. For this purpose, the gait database should contain a large number of subjects as well as a variety of covariate factors. There are some publicly available well-know gait databases, such as the SOTON covariate database [97], CASIA B database [126, 129] and USF gait database [94]. In order to evaluate the effectiveness of the considered feature selection approaches on overall gait recognition performance under different scenarios, we carried out all our experiments on these two benchmark databases, namely SOTON covariate and CASIA-B gait datasets. Moreover, from the research perspective, there are different advantages from these two databases: (1) the SOTON covariate gait database includes a various covariate factors; and (2) the CASIA-B gait database a large number of subjects. These scenarios should be fully investigated to design robust and reliable feature selection algorithms for model-based gait recognition.

- The SOTON covariate dataset originates from eleven subjects walking parallel to a static camera, with fifteen covariate fac-

1 http://scikit-learn.org/stable/

tors. Each subject was captured wearing a variety of clothes (rain coat, trench coat and normal), footwear (flip flops, bare feet, socks, boots, trainers and own shoes) and carrying various bags (barrel bag slung over the shoulder or carried by hand on shoulder, hand bag held in hand, and rucksack). They were also recorded walking at different speed (slow, fast, and normal speed). For each subject, there are approximately twenty sets of walking sequences from right to left and vice-versa. Overall, the dataset consists of 3178 walking sequences from 11 subjects spanning 15 covariates. These 3178 walking sequences are the data samples used for training and testing purposes. In order to estimate the classification accuracy on this dataset, we performed random sub-sampling stratified cross-validation, also known as Monte Carlo cross-validation, with a training/test ratio of 80/20.

- The CASIA-B is a large multiview gait dataset with 124 different subjects under the variations in viewing angle and walking status (normal, in a coat, or with a bag) [126]. Besides, there are 6 sequences from each normal walking subject under one of the 11 viewing angles. This is one of the most widely used gait database since it contains large view variations ranging from frontal view (0°) to back view (180°) at an 18°interval. The spatial resolution and frame rate of video files is 320×240 pixels and 25 frames per second, respectively. In our experiment, we only consider the use of videos with a lateral view angle (view angle of 90°). This database are recorded with ten sequences for each subject: six normal sequences (SetA: NM-01 to NM-06); two sequences with a coat (SetB: CL-01 to CL-02,); two sequences with a bag (SetC: BG-01, to BG-02). Hence, for each subject, we have at most 6 video sequences of the same condition. Therefore, in order to tune the classifiers' hyper-parameters, we perform 6-fold stratified cross validation. That is equivalent to leaving one sequence out for each subject. Then, for feature selection, in the normal case we train on five normal sequences and test against the held-out normal sequence. In the covariate cases, we train on three normal sequences and one sequence from each of the covariate factor (clothes and bag) and we test against the held-out covariate sequence.

4.4 RESULTS AND DISCUSSIONS

4.4.1 *Exploratory Analysis*

Figure 27 and 28 report via box-and-whiskers plots the distribution of classification accuracy across experiments, namely the median average CCR and its interquartile range. Six different types of feature

selection, plus no feature selection, are included in this comparative study. Each method is tested together with each of the three considered classifiers on each of the covariate factors. These classifier-covariate combinations, on the SOTON and on the CASIA-B datasets, account for the 18 and 9 experimental units displayed in Figure 27 and 28 respectively.

As shown in Figure 27, which illustrates the experiment with the SOTON dataset, we observe that the ExtraTrees and SVM classifiers achieve a better CCR than the KNN classifier. As for the covariate factors, carrying and clothing seem to be the more challenging ones, but we can also observe variations with the classifier in use. For example, the SVM accuracy appears slightly lower in the presence of different clothing conditions as compared to different carrying conditions, whereas KNN seems to be also affected by varying walking speed. ExtraTrees, on the contrary, appear to consistently yield a good performance. To further test the effectiveness of the methods, the larger gait database of CASIA-B is also used in the evaluation process, and the experimental results are shown in Figure 28. It can be observed that the ExtraTrees classifier achieves better results than the SVM and KNN classifiers across all covariate factor. Amongst the covariate factors, the carrying condition appears to be the most challenging one in the CASIA-B dataset.

Regarding the effect of feature selection, the accuracy improvements achieved by the investigated methods seem to vary across classifiers and covariates; moreover, visual inspection suggests comparatively close figures from different methods. Therefore, in order to better assess their relative impact, we move onto a statistical analysis of the experimental results.

4.4.2 *Statistical Analysis*

Question 1. *Is there a significant effect of feature selection on gait classification accuracy?*

4.4.2.1 *Three-way ANOVA*

Let CCR_{ijkr} denote the random variable giving the classification accuracy from replication r of random subsampling cross-validation when applying feature selection method i and classifier j to the walking sequences with varying covariate factor k. In our case, feature selection (*fs*), classifier (*clf*), and covariate (*cov*), are the experimental factors whose effect on classification accuracy we want to investigate. In SOTON dataset, we considered 7 feature selection methods, 3 classifiers, and 6 covariates. On the contrary, we considered 7 feature selection methods, 3 classifiers, and 3 covariates for CASIA-B dataset. Since we measure CCR at every combination of *fs*, *clf*, and *cov*, our factors are *crossed* and we have a balanced $7 \times 3 \times 6$ ANOVA design with

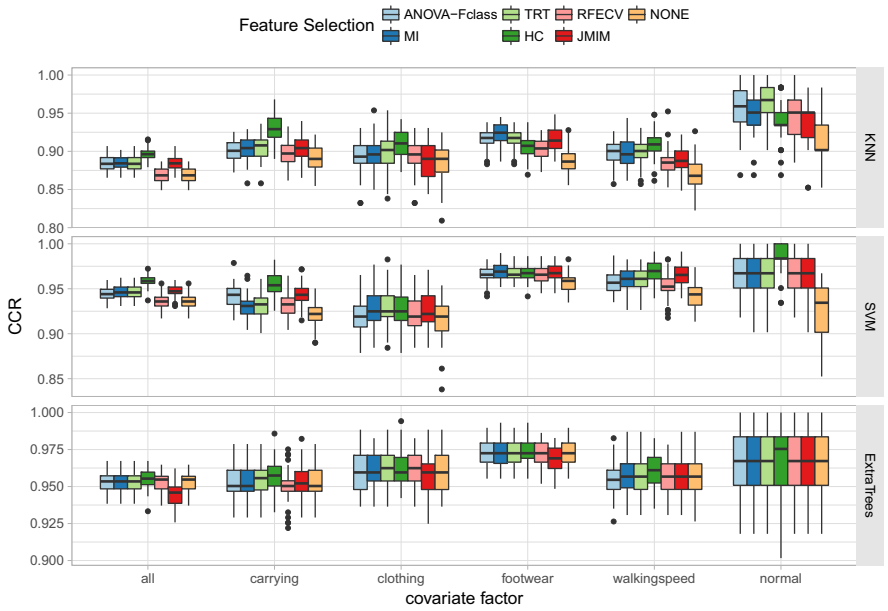


Figure 27: Box plots of test accuracy of the models on SOTON dataset. Note the different scales in the y-axis, that is, the range of ExtraTrees accuracies is shorter and closer to 1. Feature selection methods, from left to right: ANOVA, MI, TRT, HC, RFECV, JMIM, None.

50 observations per experimental unit for SOTON dataset. As for the CASIA-B dataset, we have a balanced $7 \times 3 \times 3$ ANOVA design with 80 observations per experimental unit. In the complete three-way model:

$$CCR_{ijkr} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkr} \tag{9}$$

where μ is the overall average CCR, α_i is the main effect of feature selection method i, β_j is the main effect of classifier j, γ_k is the main effect of covariate factor k, and $(\alpha\beta)_{ij}$ $(\alpha\gamma)_{ik}$ $(\beta\gamma)_{jk}$ and $(\alpha\beta\gamma)_{ijk}$ are the two-way and three-way interaction terms, respectively; ε_{ijkr} are the i.i.d. unexplained residuals.

In essence, this is a simple additive linear model. We use it to perform a classical analysis of variance in order to tests whether or not the average CCR differs significantly across the experimental units defined by the blocking factors (*fs*, *clf*, and *cov*) and their interactions. Table 10 and 11 report the ANOVA results for SOTON and CASIA-B datasets, respectively.

As it can be noticed by looking at the sums of squares, most of the observed variance in correct classification rate can be explained by considering the classifier in use, the covariate dataset it is applied to, and their combination. Feature selection has a significant effect ($P < 2.2e^{-16}$), meaning that at least one of the feature selection methods in this study produces a significant impact on the average classification accuracy when compared to the other methods. However,

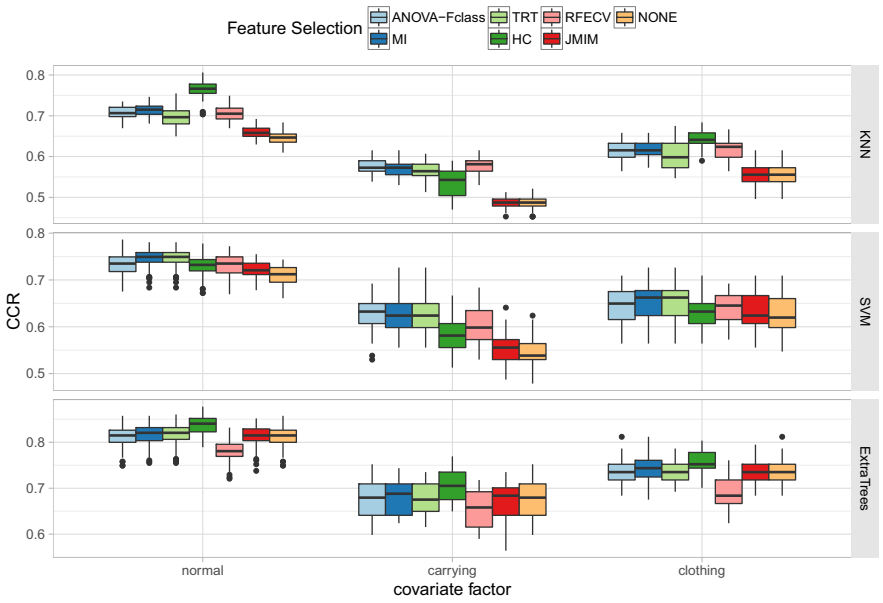


Figure 28: Box plots of test accuracy of the models on CASIA-B dataset. Notice again that, in contrast with the subsequent interaction plots, the y-axis have here different ranges. Feature selection methods, from left to right: ANOVA, MI, TRT, HC, RFECV, JMIM, None.

the presence of significant three-way interaction (*fs:clf:cov*) does not allow us to discuss the main effects or the two-way interaction either. In other words, the effect of feature selection on average model accuracy seems to depend on both the classifier and the dataset.

Figure 29 and 30 provides an illustration of this three-way interaction in both datasets. In the absence of it, the lines of average CCR by feature selection method would be parallel to each other across the blocks of classifier and covariate factors. This would allow us to aggregate results and directly examine the effect of feature selection. Instead, we can visually appreciate differences. Therefore, we need to follow-up the analysis by fixing one of the blocking factors and performing series of two-way ANOVAs.

Let us start with the SOTON dataset. If we subset our data by classifier (tables not reported), we still have a significant two-way interaction between feature selection method and covariate factors for both KNN and SVM classifiers. With ExtraTrees though, the interaction is not significant ($P = 0.915379$), but the main effects of feature selection ($P = 0.001076$) and covariate factors ($P < 2.2^{-16}$) are. In this case, we can perform a post-hoc Tuckey's Honestly Significant Difference (HSD) Test [54] to see which feature selection method has a significantly-different average CCR. We obtain that only HC brings an improvement over the baseline model with no feature selection; all other methods are indistinguishable. But all CCR differences are

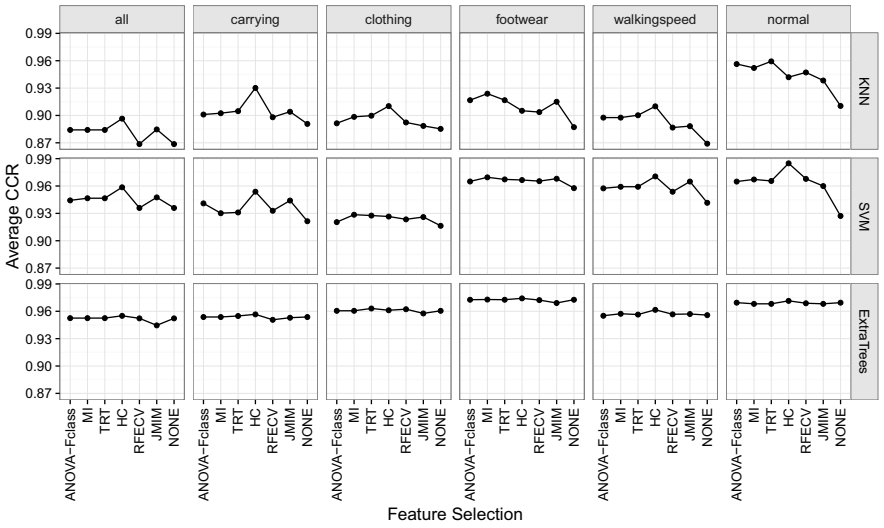


Figure 29: Interaction plot. Mean CCR by feature selection method for each level of the blocking factors "classifier" (by row) and "covariate" (by column).

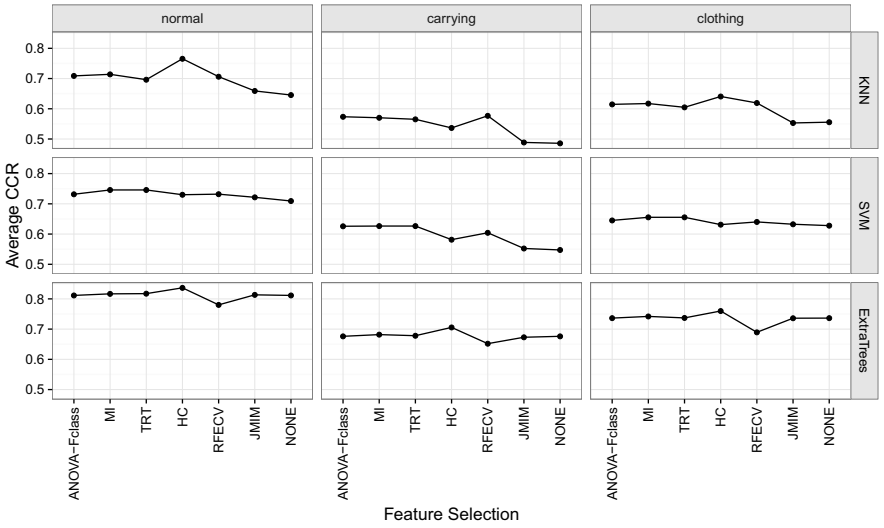


Figure 30: Interaction plot. Mean CCR by feature selection method for each level of the blocking factors "classifier" (by row) and "covariate" (by column).

Table 10: Analysis of variance - SOTON dataset.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
fs	6	0.201	0.0335	131.85	$< 2.2e^{-16}$
clf	2	3.677	1.8385	7241.78	$< 2.2e^{-16}$
cov	5	0.902	0.1804	710.55	$< 2.2e^{-16}$
fs:clf	12	0.099	0.0082	32.49	$< 2.2e^{-16}$
fs:cov	30	0.078	0.0026	10.28	$< 2.2e^{-16}$
clf:cov	10	0.497	0.0497	195.73	$< 2.2e^{-16}$
fs:clf:cov	60	0.080	0.0013	5.28	$< 2.2e^{-16}$
Residuals	6174	1.567	0.0003		

Abbreviation note: fs = feature selection, clf = classifier, cov = covariates.

Table 11: Analysis of variance - CASIA-B dataset.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
fs	6	1.408	0.2347	304.94	$< 2.2e^{-16}$
clf	2	14.124	7.0620	9175.40	$< 2.2e^{-16}$
cov	2	17.432	8.7161	11324.49	$< 2.2e^{-16}$
fs:clf	12	1.303	0.1086	141.12	$< 2.2e^{-16}$
fs:cov	12	0.274	0.0229	29.69	$< 2.2e^{-16}$
clf:cov	4	0.111	0.0277	35.98	$< 2.2e^{-16}$
fs:clf:cov	24	0.257	0.0107	13.89	$< 2.2e^{-16}$
Residuals	4977	3.831	0.0008		

Abbreviation note: fs = feature selection, clf = classifier, cov = covariates.

in the range $[-0.005, 0.005]$: the effect is statistically significant but practically irrelevant.

Indeed, if we follow up the two-way analysis by fixing also the covariate factor and hence performing series of one-way ANOVAs, we see that feature selection has a significant impact on the accuracy of ExtraTrees models only in the case of all conditions together ($P = 7.544^{-12}$). In contrast, KNN accuracy significantly differs across feature selection methods on all covariate subsets. With SVM, finally, we can reject the null hypothesis of all feature selection methods yielding the same average CCR in the cases of five covariate factors out of six. The only exception is clothing ($P = 0.02908$).

Similarly, on the CASIA-B dataset, we always have a significant two-way interaction between the feature selection method and covariate factors for both KNN and SVM classifiers. With ExtraTrees again, the interaction is not significant ($P = 0.02355$), but the main effects of

feature selection ($P < 2.2^{-16}$) and covariate factors ($P < 2.2^{-16}$) are. Only if we follow up the two-way analysis by fixing also the covariate factor and hence performing series of one-way ANOVAs, we then find that feature selection has a significant impact on the accuracy of all classifier models and in the case of all covariate conditions.

These results tell us that the gait recognition task can be more or less complicated by varying covariate factors, depending on the feature extraction process and on the classifier in use. Moreover, feature selection can have a significant impact on classification accuracy, but this in turn depends on both the classifier and the covariate subset. Because of such interaction, a simple ANOVA does not allow us to discuss and compare feature selection methods in isolation.

Question 2. *How can we generalise? can we find a common effect of feature selection methods across the experimental units defined by the combinations of classifier and covariate factors?*

4.4.2.2 *Mixed Model*

The linear model presented in equation (9) allows us to make inferences only considering the classifiers and covariate factors under study. These choices, naturally, do not cover all combinations of all possible classifiers and covariates. Moreover, we did not set to discuss the effectiveness of those specific classifiers or the hardness of those specific covariate factors. Nevertheless, as we have seen, classifier and covariate (and their interaction) have a strong impact on classification accuracy, which we need to model. We could regard our classifiers and covariates as drawn at random from a hypothetical population of all possible classifiers and covariates, and thus treat their combinations as random factors. This would allow us to estimate and compare the eventual impact of the chosen feature selection methods, while still taking into account the experimental conditions and the interdependencies that they generate among our observations. In statistical jargon, we will then refer to feature selection as the *fixed-effect* in which we are interested, and to the experimental units of classifier-covariate combination as the *random-effect* from which we would like to abstract.

Let CCR_{ijr} be a random variable representing the classification accuracy observed from replication r of Monte-Carlo cross-validation when testing a model that applies feature selection method i to the experimental conditions j, where j indicates one possible classifier-covariate combination. The full *mixed-effect* ANOVA model [27, 11] can be specified as:

$$CCR_{ijr} = \mu + \alpha_i + B_j + (\alpha B)_{ij} + \varepsilon_{ijr} \tag{10}$$

where μ is the overall average CCR, α_i is the parameter representing the effect of feature selection method i, B_j is a random variable

Table 12: Fixed effects estimates from the mixed model for SOTON dataset. Estimated gain in average CCR across experimental units with respect to the baseline of not using any feature selection method.

	Estimate	Std. Error	t value
(Intercept)	0.9265***	0.0081	113.88
fsANOVA-Fclass	0.0127***	0.0034	3.76
fsMI	0.0138***	0.0033	4.14
fsTRT	0.0141***	0.0034	4.14
fsHC	0.0199***	0.0039	5.06
fsRFECV	0.0090**	0.0029	3.09
fsJMIM	0.0113***	0.0029	3.84

*p<0.1; **p<0.05; ***p<0.01

representing the random effect of experimental block j, and $(\alpha B)_{ij}$ is a random variable representing the possible random deviation between the common effect α_i and the specific effect that feature selection method i displays in experimental condition j. It has to be noticed that, strictly speaking, B_j and $(\alpha B)_{ij}$ are not model parameters (they are obtained after the model fit, as model residuals ε_{ijr} are). Indeed, they are random variables with zero mean, whose variances and covariances are model parameters. Hereby, model residuals are now i.i.d. only within each experimental block. Notice also that we explicitly include the second-level random effect $(\alpha B)_{ij}$ since we are aware of significant interactions between effects, but also because using the maximal random-effect structure that is supported by the data, allows us to avoid anti-conservative results in the estimation of α_i, the fixed-effects of interest [5]. That is to say, we want to obtain the most generalisable inferences that the dataset under study could support.

Table 12 and 13 report the maximum-likelihood estimates of the coefficients μ, the intercept, and α_i, the fixed-effects parameters for SOTON and CASIA-B dataset, respectively. Each of the latter represents the difference in expected CCR between feature selection method i and the control group of models without any feature selection. Each estimate is tested against the null hypothesis $\alpha_i = 0$ and the respective p-values are reported, symbolically encoded. All feature selection methods seem to yield a significant gain in accuracy, which, after all, was the sole optimisation objective. However, in SOTON dataset, the first observation to make is that such an improvement is only marginal, ranging from 0% to 2.15% over the baseline model with no feature selection (Intercept). As for CASIA-B dataset, we observe a better improvement of CCR, ranging from 0% to 6.77% over the baseline (Intercept). In particular, the wrapper search-based method

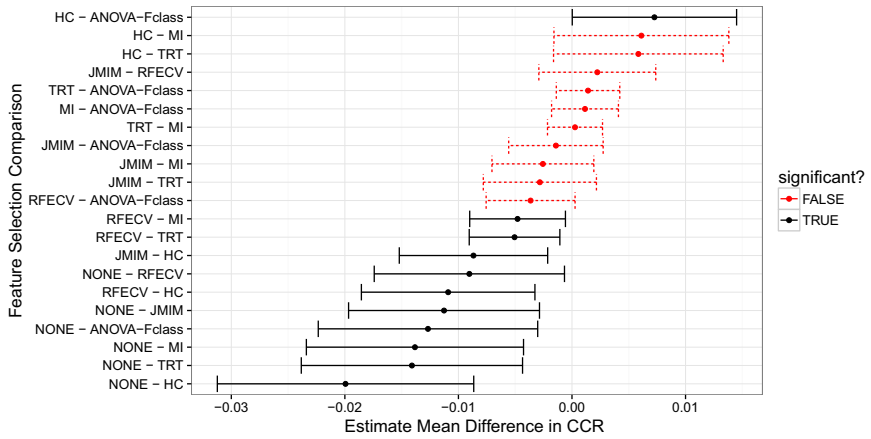


Figure 31: Simultaneous estimation of all pairwise comparisons for the fixed-effect of feature selection in the mixed model. 95% confidence intervals are adjusted following Tukey's method.

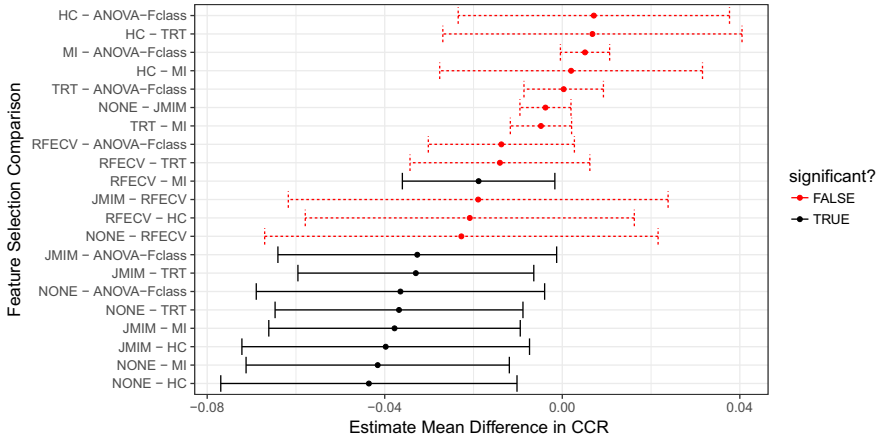


Figure 32: Simultaneous estimation of all pairwise comparisons for the fixed-effect of feature selection in the mixed model. 95% confidence intervals are adjusted following Tukey's method.

Table 13: Fixed effects estimates from the mixed model for CASIA-B dataset. Estimated gain in average CCR across experimental units with respect to the baseline of not using any feature selection method.

	Estimate	Std. Error	t value
(Intercept)	0.6440***	0.0342	18.82
fsANOVA-Fclass	0.0365**	0.0119	3.06
fsMI	0.0416***	0.0109	3.82
fsTRT	0.0368***	0.0102	3.59
fsHC	0.0436***	0.0122	3.56
fsRFECV	0.0227	0.0163	1.40
fsJMIM	0.0038	0.0021	1.80

*p<0.1; **p<0.05; ***p<0.01

seems to have a slight edge, but the relative rankings are different on the two datasets and, more importantly, care needs to be taken when performing multiple comparisons.

To address the problem of family-wise errors in multiple hypothesis testing, Figure 31 and 32 report the simultaneous estimates of fixed-effects' confidence intervals with Tukey's HSD adjustments for SOTON and CASIA-B dataset, respectively. There is enough empirical evidence to state that, on SOTON dataset, JMIM, RFE and first-improvement hill-climber HC approaches outperform conventional filter approaches in terms of final gait recognition accuracy. Nevertheless, we can observe only minor differences and, notably, no statistically-significant difference between MI-based rankings and rankings based on the analysis of variance. However, in CASIA-B dataset, the conventional method of MI-based rankings outperforms other statistical approaches in terms of final gait recognition accuracy. That makes it difficult to generalise results from one dataset to another. But we notice that, in the case of the CASIA-B dataset, the number of training samples per class is very small. Therefore, the k-nearest neighbors estimation of joint-mutual information within the JMIM method and the built-in estimation of feature importances of embedded methods, are most probably noisy and biased due to a lack of subject-specific data. Hence, the performances of those approaches are hard to disentangle and also rather close to what can be obtained by the simpler MI or ANOVA univariate rankings.

4.4.3 *Analysis of Computational Time and Best Feature Subset Length*

Figure 33 and 34 show, for all the combinations of feature selection method, classifier, and covariate factors, the average computational time required by feature selection and classification (in secs) and the

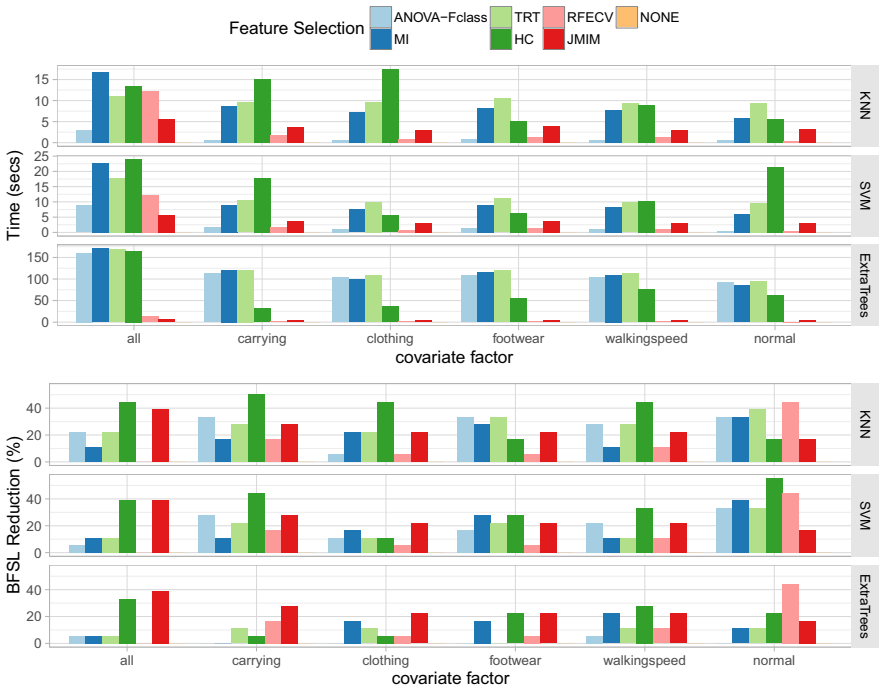


Figure 33: Top: Computational Time (in secs) on SOTON. Bottom: reduction (%) in Best Feature Subset Length on SOTON. Notice the different y-axis time scales. Feature selection methods, from left to right: ANOVA, MI, TRT, HC, RFECV, JMIM, None.

length of the best feature subset (in percentage of the original feature set cardinality) for SOTON and CASIA-B dataset, respectively. In both datasets, we notice that the KNN classifier requires a computation time significantly lower than the other two classifiers. However, its recognition accuracy is also lower than that of SVM and ExtraTree across the different covariate factors in this dataset. ExtraTrees are, comparatively, the slowest classifier in this study: their computational complexity is exposed by the Monte Carlo cross-validation, which at each iteration requires to fit and test a large ensemble of unpruned trees. SVMs, in contrast, produce sparse models that are not only faster to fit but also relatively faster to poll for predictions when testing.

As for the feature selection methods, we observe that conventional statistical approaches are much cheaper to evaluate than search-based wrapper approaches. In particular, ranking features according to univariate statistics is almost immediate, except when extracting variable importances from randomised trees. What adds to the computation time then is the linear search for the best number of features to retain, which is again based on cross-validation. Overall, the hill-climber (HC) method took longer than all other feature selection methods across all types of covariate factors. As can it can be seen in the bot-

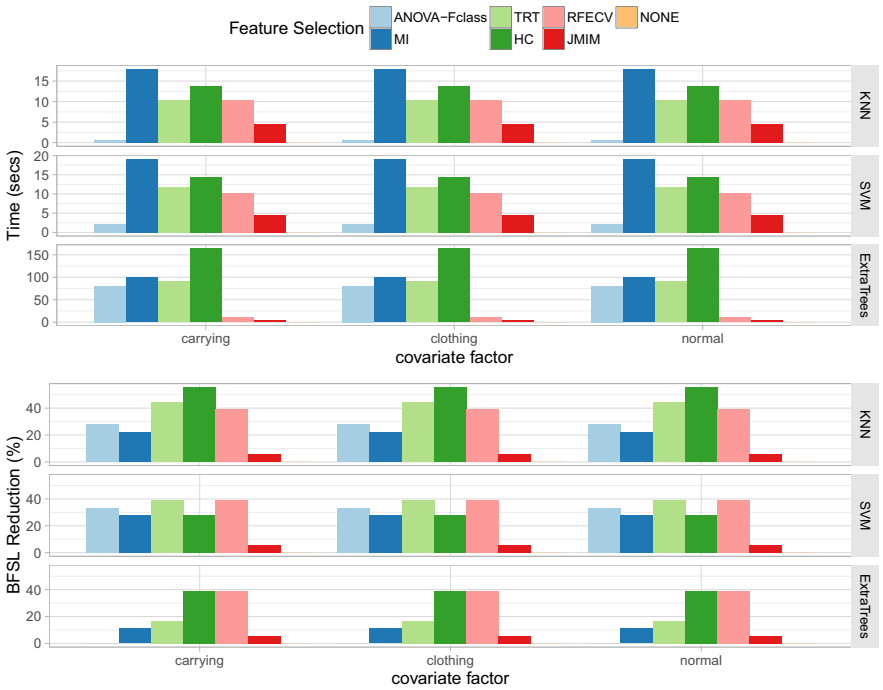


Figure 34: Top: Computational Time (in secs) on CASIA-B. Bottom: reduction (%) in Best Feature Subset Length on CASIA-B. Notice the different y-axis time scales. Feature selection methods, from left to right: ANOVA, MI, TRT, HC, RFECV, JMIM, None.

tom parts of Figures 33 and 34, HC also seems to select smaller sets of optimal gait features. On both datasets however, embedded methods took relatively less time to achieve comparable results, both in terms of accuracy and feature set reduction. This would suggest that, when the feature set is not overly complex, simple embedded methods should probably be the preferred solution for feature selection in gait classification.

Regarding the parsimony of the selected models, search-based methods discarded more aggressively the "irrelevant" features, resulting overall in smaller feature sets, despite the fact that the size of the selected subset was not part of the search objective. Nonetheless, such behaviour again depends on the classifier in use. Decision trees for instance, internally select the most discriminative features to split upon, hence the smaller reduction in feature subset length when optimising the final classification accuracy of ExtraTrees.

As for the selected features, on both datasets and for all classifiers, the features that are most often discarded by all selection procedures are related to the angular trajectories. In particular, the total angle of the trailing knee appears to be the less reliable one in the presence of occlusions.

Table 14: Correct classification rate (%) of different methods on the SOTON database [97].

Feature Extraction	#Covariates	#Subject	#Sequences	FS	Classifier	CCR (%)
Bouchrika et al. [20]	11	10	440	None	KNN	73.4
Our approach						74.5
Pratheepan et al. [90]	4	10	180	None	SVM-Ln	86.0
Our approach						93.0
Ng et al. [85]	13	11	2722	None	SVM-Ln	84.0
Our approach						93.0
Our approach	15	11	3178	None	SVM-RBF	93.6
				HC		96.0
				JMIM		95.0

4.4.4 *Further Comparisons with Conventional Methods*

Table 20 summarizes the comparison of results with respect to other conventional techniques for gait classification that use the SOTON covariate database. To have a fair comparison, we evaluate our approach under the same experimental settings and data subset. That is to say, we train the same classifier as in the approach we compare to, but using our own feature extraction technique and without performing any feature selection.

Overall, we seem to obtain higher classification accuracy than the results published in Bouchrika et al. [20], Pratheepan et al. [90] and Ng et al. [85], after taking into account the number of subjects, the number of covariate factors and the number of walking sequences that were used for training and testing. The feature extraction method proposed by Bouchrika et al. requires to manually label model template in order to describe the joints' motion, which could result in decreased classification accuracy. Moreover, in comparison with the approach by Pratheepan et al., we do not incorporate the removal of occluded sequences for estimating the gait cycle; instead, our approach uses the entire walking sequence for gait analysis. It can also be noticed from the results that our approach outperforms Ng et al.'s as their approach have difficulty in extracting gait features from the silhouette and self-occluding silhouette model, which can lead to an erroneous extraction of the gait features.

In conclusion, our technique for feature extraction shows an improvement in recognition accuracy as compared to previous results in the literature, without any feature selection and by using all the respective walking sequences in the SOTON dataset. This further helps putting into perspective the relative accuracy gain that feature selection methods can provide.

Table 15: Comparisons with other existing methods on the CASIA gait database B using lateral view (90°). The bold font highlights the highest accuracy for each test scenario in the experiment.

Method	Normal	Carrying	Clothing	Overall
Yu et al. [126]	97.60	32.70	52.00	60.77
Han et al. [81]	99.60	57.20	23.80	60.20
Bashir et al. [8]	**100.00**	78.30	44.00	74.10
Bashir et al. [9]	97.50	**83.60**	48.80	76.63
Bashir et al. [7]	99.40	79.90	31.30	70.20
Dupuis et al. [37]	98.80	73.80	63.70	78.77
Rida et al. [92]	95.97	63.39	72.77	77.38
Hu et al. [55]	94.00	45.20	42.90	60.70
Kusakunniran et al. [68]	95.40	60.90	52.00	69.43
Kusakunniran et al. [67]	94.50	60.90	58.50	71.30
Jeevan et al. [59]	93.36	56.12	22.44	57.31
Our method (None-KNN)	64.60	48.60	55.60	56.30
Our method (None-SVM)	70.90	54.70	62.80	62.80
Our method (None-ETC)	81.10	67.60	73.60	74.1
Our method (RFECV-KNN)	70.60	57.70	61.90	63.40
Our method (TRT-SVM)	74.60	62.60	65.50	67.57
Our method (HC-ETC)	83.70	70.60	**76.00**	76.77

Then, to investigate the performance of gait recognition on the CASIA-B database, an experiment was carried out to compute the CCR, and the experimental results are listed in Table 15. Results have been compared with those obtained by other approaches available in the literature. It can been seen that methods investigated in this work display good results, especially against changes in clothing types. This can be explained by the fact that our method eliminates noisy features which, in turn, improves the recognition performance in the clothing conditions when these features are considerably affected. However, the overall performance is comparable to that of other conventional methods.

4.5 SUMMARY

In this chapter, we presented and compared feature selection techniques, designed to maximise the final gait classification accuracy. The overall approach constitutes a general framework for different machine learning algorithms, which we applied to the problem of feature selection under the effect of various covariate factors in a

model-based approach. The main goal of the present work was to find a best set of features in order to improve the overall classification accuracy.

The experiments were performed on a two well-known gait datasets, namely the SOTON covariate dataset and the CASIA-B dataset. The experimental results on both showed that search-based and embedded approaches can improve the Correct Classification Rate of the final model. Such accuracy gain is more pronounced when using a simple classifier as KNN, but not so evident in the case of decision-tree variants ExtraTrees. Nevertheless, we could confirm a common, significant improvement over not performing any feature selection and also over more conventional feature ranking techniques, across the considered combinations of classifiers and covariate factors. A mixed-effect statistical analysis of model accuracy gives empirical evidence for this conclusion. In our experiments, feature selection is able to select the discriminative input gait features and provides an improved overall classification performance, which is comparable to that of other state-of-the-art approaches. Hence, the proposed methodology provides a valid alternative for the selection of relevant features for model-based gait recognition.

Nevertheless, this improvement in classification accuracy remains marginal over the baseline accuracy of a given classifier, and over what could be gained switching feature extraction technique or classifier technology. It also comes at the cost of a manyfold increase in computation time. Therefore, if a wrapper approach is practically feasible for the dataset under study and if the feature space does not seem to require strong global-search capabilities, we would recommend a simple local search. In our experiment, a hill-climber showed to provide final classification rates comparable to those from a genetic algorithm but with a comparably smaller computational complexity.

Naturally, since all methods have problems for which they are especially well-suited and others for which they are not, a small change in the problem setup might change the performance rankings and the results of a comparative study. Hence, we invite the interested reader to be careful extrapolating our reasonable yet empirical conclusions to other problems or datasets [48]. Despite the fact that we did our best to avoid anti-conservative estimations, our findings are based on the analysed datasets. With this work, we hope to have provided not only the first comparative study of feature selection methods for gait classification, but also a modest example of analytical methodology.

This work can be extended in several directions in the future. For instance, the primary limitation of our approach is its dependency on the viewing angle and speed. The selected set of features would be uninformative and irrelevant if the viewpoint of the camera were to change, and also if the same subject were to walk at a significantly different speed. We intend to extend our approach to make it invari-

ant to speed and viewing angles changes by better understanding and correlating the set of features that have maximum discrimination power in order to overcome these limitations in the future. Moreover, it will be interesting to see how our approach can be modified to accommodate the case of multi-view gait identification under the effect of different covariate factors. The extension of this work will involve working with larger available gait databases and also on challenging gait covariate factors. This will not only further test the validity of our results but also reaffirm the reliability of our feature set. Therefore, our future goal will be to optimize the system for invariance in speed, viewing angle, and hence develop a reliable human gait recognition system. On a different note, the feature selection problem could also be addressed in a multi-objective optimisation context; in this way, we could analyse human gait patterns from other perspectives, and explicitly include models parsimony in the maximisation problem. In addition, we would like to explore the possibility of using deep neural networks in gait recognition tasks as means to automatic feature engineering. Deep learning approaches are very often effective, although at the expenses of a very high computational burden, yet they only require the use of minimal problem domain knowledge.

Part IV

WHEN GAIT RECOGNITION MEETS WITH DEEP LEARNING

In recent years, significant efforts have been put into the problem of human identification based on gait recognition. However, most of the conventional approaches have been used hand-crafted features for representing human gait, which might not generalize well to diverse datasets. In addition, in recent studies, deep learning approaches have shown promising results on various image classification tasks. In this part, we explore the use of deep learning-based approaches for learning and extracting gait features that suitable for gait recognition under challenging covariate factors such as changes in the clothing and variations in viewing angle.

5 DEEP LEARNING FOR GAIT RECOGNITION

5.1 OVERVIEW

Recently, deep learning has gained significant attention from the computer vision community. This is because deep learning models are capable of learning multiple layers of feature hierarchies by constructing high-level features from low-level features [3]. Hence, they are more generic since the feature construction process is fully automated. In addition, many recent studies have shown promising results for applying deep learning approaches to a variety of applications (e.g. image classification, text classification, natural language processing, scene labeling, etc.) [66, 128, 29, 39]. However, conventional state-of-the-art methods have mostly used hand-crafted features for representing the human gait. To the best of our knowledge, deep learning based approaches have not been well applied to address the problem of gait recognition yet. Therefore, in this work, we explore the use Convolutional Neural Networks (CNNs) with appropriate gait representation for gait recognition. Specifically, the variations in clothing types can cause the recognition rate decreased greatly. Therefore, we also employed a model based on stacked auto-encoders for clothing-invariant gait recognition. The model is trained to handle gait data with clothing variations. We also demonstrate the effectiveness of both proposed methods in gait recognition tasks using the publicly available gait database. We will describe the framework in the Section 5.3 and Section 5.4, respectively.

5.2 EXISTING WORK

Many promising works of human identification based on gait approaches have been introduced (for a recent review see [73, 79]). These works demonstrate the feasibility of using gait signature for human identification at a distance. However, gait recognition is still a complicated task because of the variations in view angles, clothing, carrying condition and other covariate factors. These factors can degrade the overall performance of gait recognition. This problem has recently gained considerable attention from the gait researchers.

Several gait approaches [80, 69] based on view transformation have been proposed which have the ability to deal with large viewing angle changes and do not rely on camera calibration. Recently, Makihara et al. [80] established a Singular Value Decomposition (SVD) based view transformation model to transform the gait features in probe

viewing angle to that in gallery viewing angles. In [69], Kusakunniran et al. established a View Transformation Model (VTM) from the different point of view by adopting Support Vector Regression (SVR) technique.

On the other hand, clothing type has been demonstrated to be the most challenging one [42]. It will drastically alter the individual's appearance with the variation of different clothing types, such as baggy pants, skirt, down jacket, and coats. The task of gait analysis becomes much more challenging [84]. In recent conventional methods, such as that in [42], proposed a technique by introducing a Random Subspace Method (RSM) framework for clothing-invariant gait recognition by setting up multiple inductive biases in a random manner. Islam et al. [57] conducted a study on human gait by dividing it into small window chunks and developed a Random Window Subspace Method (RWSM) for clothing invariant gait recognition.

However, previous conventional approaches have mostly used hand-crafted features for representing human gait. This hand-crafted features to try to capture the essence of different visual of gait patterns. Although most of the conventional approaches have satisfactory performance, but the main drawback of these approaches are highly problem-dependent. Therefore, this work studies an approach to gait based human identification via deep learning approaches.

5.3 CONVOLUTIONAL NEURAL NETWORK

The convolutional neural network (CNN or ConvNet) [41] model is a type of feed-forward artificial neural network which was inspired by biological processes [83]. It contains a variations of multilayer perceptrons which are designed to to use minimal amounts of preprocessing [71]. The connectivity pattern between its neurons within the network was inspired by the organization of the animal cortex [56], whose individual neurons are tiled in such a way that they respond to overlapping regions in the visual field. In general, a typical CNN architecture is formed by a sequence of layers that transform the input image volume into an output volume through a differentiable function. It is usually made up of a *convolution layer*, a *spatial pooling layer*, a *normalization layer* and followed by *fully-connected layer*.

A few distinct types of ConvNet layers are commonly used in deep learning networks and we explain them further below:

1. *Convolutional layer*: This layer's parameters consist of a set of learnable filters (or kernels), which have a small receptive field, but it extends through the full depth of the input data volume. Moreover, during the forward pass, each filter is convolved across the width and height of the input volume, and computing the dot product in between the entries of the filter and the input and creating a two dimensional activation map of that filter.

Hence, the network learns filters that activate when they see a particular type of visual feature at some spatial position in the input images. The full output volume of the convolution layer is formed by stacking the activation maps for all filters along the depth dimension. Therefore, every entry in the output volume can thus be considered as an output of a neuron that observes at a small region in the input and shares parameters with neurons in the same activation map.

2. *Pooling layer*: This layer divides the given input image into a set of non-overlapping rectangles and, for each such sub-region, it outputs the maximum. The exact location is not as important as its rough location corresponding to other features once a feature has been found. The aim of using pooling layer is to reduce the spatial dimensions of the feature. Eventually, it will minimize the amount of parameters and computation in the network, and hence to also control overfitting. It is a common practice in deep learning to regularly insert a pooling layer in between successive convolution layers in a CNN architecture. Moreover, the pooling operation provides a form of translation invariance.

3. *Fully-connected layer*: Finally, after several convolutional and max-pooling layers, the high-level reasoning in the neural network is done via fully connected layers. Hence, neurons in the fully-connected layer have full connections to all activations in the previous layer.

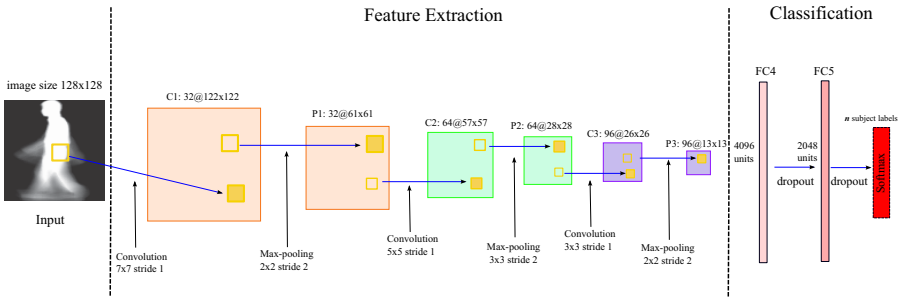
For a more detailed of the latest deep learning findings and the architecture of a CNN is designed, the interested readers may refer to the works in [35, 41].

5.3.1 *Methodology*

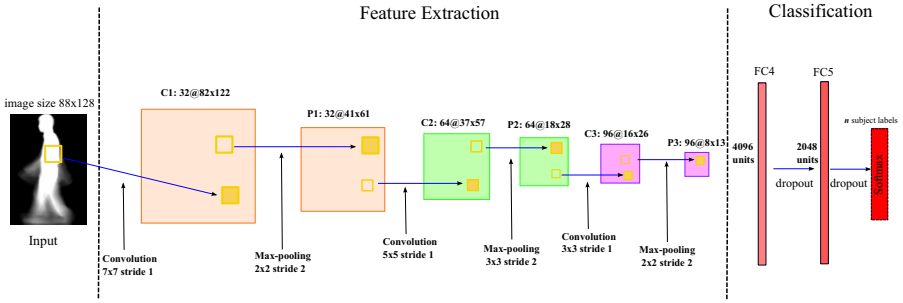
In this section, we describe our proposed method to address the problem of gait recognition using CNN for gait recognition tasks. The proposed framework for gait recognition based on CNN is illustrated in Figure 35a and Figure 35b for CASIA-B and OU-ISIR Treadmill B dataset, respectively.

5.3.1.1 *Input data*

In this work, we use a Gait Energy Image (GEI) [81] as the gait feature descriptor and input data to the CNN. Example GEIs belonging to one subject in the CASIA-B dataset and OU-ISIR Treadmill B dataset are shown in Figure 36a and Figure 36b, respectively. GEI is a spatiotemporal gait representation constructed using silhouettes. Given



(a) CASIA-B dataset



(b) OU-ISIR Treadmill B dataset

Figure 35: The proposed convolutional neural network (CNN) architecture for gait recognition. Top: CASIA-B dataset. Bottom: OU-ISIR Treadmill B dataset.

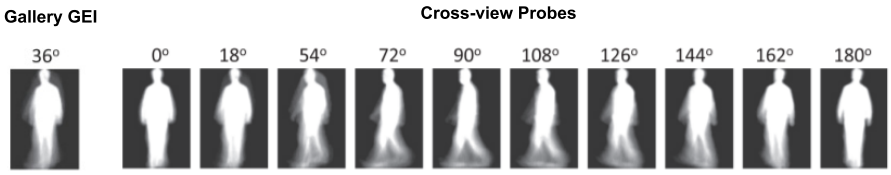
a size-normalized and horizontal aligned human walking binary silhouette sequenece $B(x, y, t)$, the grey-level GEI $G(x, y)$ is defined as follows

$$G(x, y) = \frac{1}{N} \sum_{t=1}^{N} B(x, y, t) \tag{11}$$

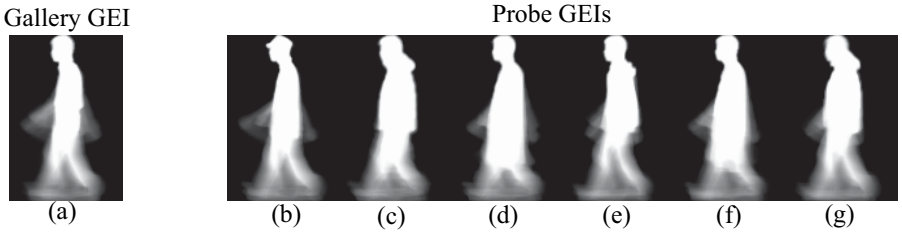
where N is the number of frames in complete cycles of the sequence, x and y are values in the 2D image coordinate.

5.3.1.2 *Network Structure*

The proposed method provides an end-to-end framework with convolutional neural network is shown in Figure 35a and Figure 35b. The network contains 6 layers, with 3 convolutional layers (Conv1, 2 and 3), and 2 fully connected layer (FC4-5) and last follow by softmax layer. In CASIA-B dataset, the first layer accepts GEI size of 128x128 which obtained from the sequence of GEI as input. On the hand, in OU-ISIR Treadmill B dataset, the first layer accepts GEI size of 88x128 pixels. The conv1, 2 and 3 yield 32, 64, 96 feature maps respectively. We set the number of neurons to 4096 and 2048 in FC4 and FC5, respectively. The input to the last layer, softmax layer, has n units, where n represents to the number of training subject samples taken from the dataset. Filter size for conv1 is set as 7x7 with strides of 1. While conv2 and conv3 use stride of 1, conv2 has filter size of 5x5,



(a) CASIA-B gait dataset: the leftmost GEI image is the gallery, GEI with view angle of 36°, while the righmost GEI image are the probe GEIs with view angle variations.



(b) OU-ISIR B dataset: the leftmost GEI image (a) is the gallery GEI in normal clothing condition, while the rightmost GEI image (b)-(g) are the probe GEIs with different clothing combinations.

Figure 36: Samples of GEIs of a similar person are computed by averaging silhouettes over the whole sequence.

and conv3 has filter size of 3x3. Local Response Normalization (LRN) is employed after conv1, conv2 and conv3 with similar settings in [66]. Due to large number of parameters in the architecture, overfitting is inevitable. Hence, it is important to use dropout as form of regularizer. We follow the implementation shown in [66] to use the dropout after the FC4 and FC5 as most of the parameters are concentrated in these layers. We employed Rectified Linear Unit (ReLU) used as a activation function for all convolution layers, except for softmax layer that uses softmax regression as activation and act as a multi-class classifier for gait classification. The activation function has been proved to yield a better performance and speed up training time as reported in [66].

5.3.2 *Experiments*

To evaluate the effectiveness of the performance of our proposed method for clothing-invariant gait recognition, we conducted all our experiments on the CASIA-B dataset [126] and OU-ISIR Treadmill dataset B [53] which contains the largest variations of view and clothing variations.

5.3.2.1 *Dataset Description*

OU-ISIR Treadmill B dataset The dataset [53] is an indoor gait dataset and consists of 68 subjects in total with 15 to 32 different clothing combinations as listed in Table 16, while the list of clothes available

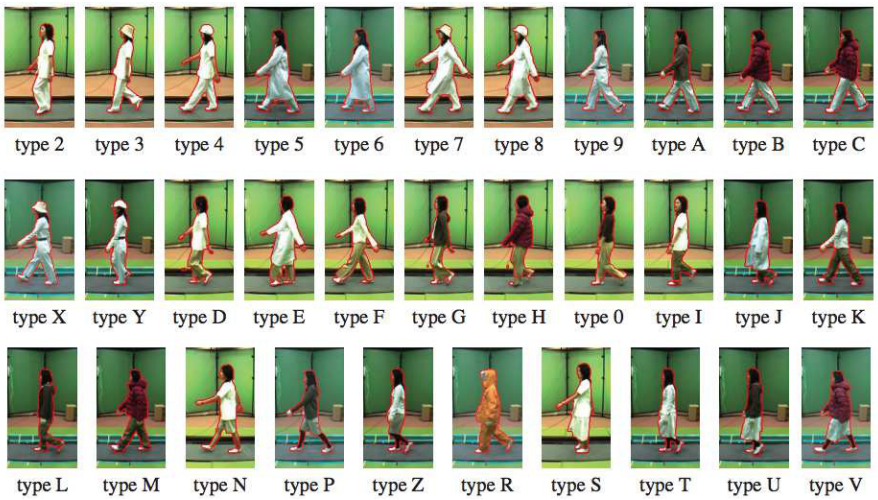


Figure 37: A sample of images for each clothing variations taken from the OU-ISIR-B dataset [53].

in the dataset shown in Table 17. The original dataset is divided into three subsets, i.e., training set, gallery set and probe set. For most of the conventional approaches, the experiments are evaluated using this test setup. They used a subset containing 446 sequences of 20 subjects with all possible clothing types for training purpose. For testing, they used a gallery set consisting gait sequences of the remaining of 48 subjects in standard clothing combination. The probe set comprising 856 gait sequences for these 48 subjects with all types of different clothing combinations excluding standard one in gallery set. This database setup does not suitable for most of the deep learning approaches as the training and test sets have been captured under different clothing conditions. The designed CNN network cannot learn the discriminative gait features that are invariant to various clothing conditions if the training data only consists samples of one normal clothing type. Hence, in this work, the whole dataset was split into into non-overlapping training and test set (using ratio of 80/20 respectively). For each subject in the test set, we used the gait image sequences which consists only standard clothing types as gallery set and whereas other clothes types to construct the probe set. Figure 37 shows sample images of different combinations of clothing variation in the dataset.

CASIA-B gait dataset The dataset [126] consists of the data from 124 subjects. The gait data was captured from 11 viewing angles, namely 0°, 18°, 36°, 54°, 72°, 90°, 108°, 126°, 144°, 162°, and 180°. There are 6 video sequences for each person under each different viewing angle. Therefore, we use a total of 8184 gait sequences. Figure 38 shows sample images of different combinations of clothing variation in the dataset.

Table 16: Different combinations of clothing variation in the OU-ISIR-B dataset[53].

Type	S_1	S_2	S_3	Type	S_1	S_2	Type	S_1	S_2
3	RP	HS	Ht	**0**	CP	CW	**F**	CP	FS
4	RP	HS	Cs	**2**	RP	HS	**G**	CP	Pk
6	RP	LC	Mf	**5**	RP	LC	**H**	CP	DJ
7	RP	LC	Ht	**9**	RP	FS	**I**	BP	HS
8	RP	LC	Cs	**A**	RP	Pk	**J**	BP	LC
C	RP	DJ	Mf	**B**	RP	DJ	**K**	BP	FS
X	RP	FS	Ht	**D**	CP	HS	**L**	BP	Pk
Y	RP	FS	Cs	**E**	CP	LC	**M**	BP	DJ
N	SP	HS	-	**P**	SP	Pk	**R**	RC	-
S	Sk	HS	-	**T**	Sk	FS	**U**	Sk	Pk
V	Sk	DJ	-	**Z**	SP	FS	-	-	-

Table 17: List of variations in clothing types used in the dataset (Abbreviation: Clothes type ID).

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

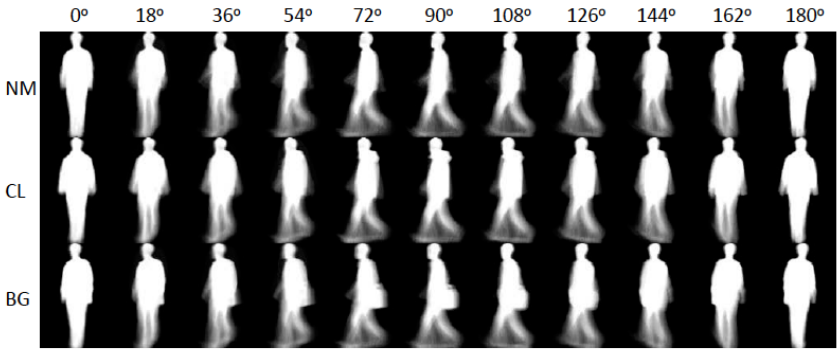


Figure 38: CASIA-B dataset: top row: normal walking (NM), mid row: with a coat (CL), bottom row: with a bag (BG).

5.3.2.2 *Implementation details*

We used stochastic gradient descent (SGD) to train our models with mini-batch of size 40 and a momentum value of 0.8. The weights are initialized by using Gaussian distribution with zero mean and a standard deviation of 0.01 for all trainable layers. All the bias terms are initialized with the constant zero. We set our initial learning rate to 0.01. When a network has a too small learning rate will eventually lead to slow convergence, while having a too big learning rate will make the weights and objective function diverge. We reduced the learning rate for all layers by factor of 10 every 1000 iterations prior to termination. We trained the model for 5000 iterations ($\sim 50 epochs$) for all experiments. We noticed that further training the model does not improve the results. The experiments was carried out using MatConvNet library [105] on NVIDIA GTX 970 4GB GPU. This library allows to prototype CNN architectures in an easy and fast manner using the Matlab environment. Besides, it takes advantage of CUDA and cuDNN [26] to improve the network's performance on classification tasks.

After we have acquired the gait features, the final stage (fully connected layer, FC5) consists of identifying those features to predict a subject identity. During this stage, we compute a common set of similarity feature to individual subjects using the Euclidean distance. During the matching between a corresponding pair of subjects (probe GEI P^i and gallery GEI G^i sequences), the similarity score between them can computed through a trained network as follows:

$$d(P^i, G^i) = \sum_{N}^{n=1} \left\| P^i(n) - G^i(n) \right\|_2 \tag{12}$$

where $d(P^i, G^i)$ is a distance between the gait signatures P^i and G^i. N is size of gait feature vector. The smaller value of d means the higher possibility that the gait signature in between of the given matching pair, P^i and G^i are belong to the same person. Although the top layer

Table 18: Comparison of results on CASIA-B dataset.

	Probe Angle			
Method	54°	90°	126°	Average
C3A [117]	56.64%	54.65%	58.38%	56.56%
SVR [69]	53.62%	45.90%	53.82%	51.11%
CNN+Softmax	57.50%	50.91%	58.39%	55.70%
CNN+SVM	60.19%	54.24%	61.39%	58.61%

of the CNN already comprises a softmax classifier, we can have the advantage of replacing the softmax layer with linear Support Vector Machine (SVM) classifier and using the extracted gait signatures as the input. Since our work is dealing with a multiclass problem, we use a binary SVM classifier with a linear kernel in a one-vs-all. In [103], they also indicated that this configuration of binary classifiers is suitable to obtain top-tier results in this problem. In our case, we L2-normalize the top fully connected layer before using it as a feature vector.

5.3.3 *Results and Discussion*

5.3.3.1 *Results on CASIA-B dataset*

The performance of our proposed approach on CASIA-B dataset can be observed from Figure 39. In overall, our method showed that these approaches can obtain good results when the data viewing angle is relatively small, i.e., 18°, however, the performance will eventually decreased due to view angle variations. In Table 18, the experimental results of C3A [117], SVR [69] and the proposed method are listed. From the experimental results, we can conclude the following key points. (1) CNN-SVM provides the highest accuracy than CNN-Softmax. (2) Being compared with conventional approach [69, 117], our proposed method performs better than the traditional approaches.

5.3.3.2 *Results on OU-ISIR Treadmill B dataset*

Performance analysis with clothing variations effect. We conducted all our experiments on OU-ISIR Treadmill B dataset to examine the performance of our proposed approach with the effect different clothing types. The performance for the proposed method can be observed from Figure 40. From the experiment results, we have the following observations:

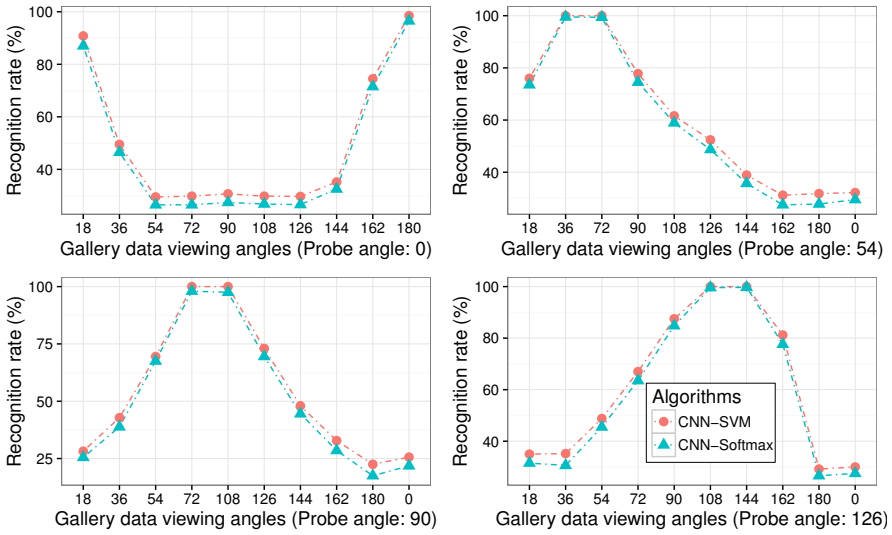


Figure 39: Performance evaluation on the CASIA-B dataset under the effects of view angle variation on gait identification.

In experiment 1 (Exp. 1), it consists of walking sequences that only involve basic clothing combinations (e.g., Y:regular pants+full shirt, I:baggy pants+half shirt, etc), the average CCR is above 94.0%. Besides from the basic clothing combinations in Exp. 1, we can see that in Experiment 2 (Exp. 2), the clothing combinations consists of walking sequences that include challenging clothing covariates (e.g., 7:regular pants+long coat, P:short pants+parker, etc). In this case, we are still able to achieve relatively high performance. The average CCR is above 88.0%. In Experiment 3 (Exp. 3), the recognition task becomes much more complicated when a complicated clothing type (e.g, down jacket or long coat) is combined with a basic clothing type (e.g., regular pants). However, based on results obtained by the proposed method, the performance still yields competitive results, with an average recognition rate above than 81.0%. The combinations of challenging clothing types shown in Experiment 4 (Exp. 4) lead to a decrease in gait recognition accuracy performance. We can observe the average CCR is further drop to 68.4% and 72.2% for CNN-Softmax and CNN-SVM, respectively. For instance, given a query gait pattern is under the effect of difficult combinations of clothing types (e.g. V:skirt+down jacket), the recognition rate is decreased to around 60.5% and 64.6% for CNN-Softmax and CNN-SVM, respectively. However, in the event of these circumstances, the recognition task would be difficult even for a human operator, as clothing occluded most of the discriminative features of the person. Interestingly, CNN is able to identify human gait under challenging combination of clothing types. In general, our experimental results using L2-SVMs show that by replacing softmax function with linear SVMs gives significant performance gains especially when the combination of cloth-

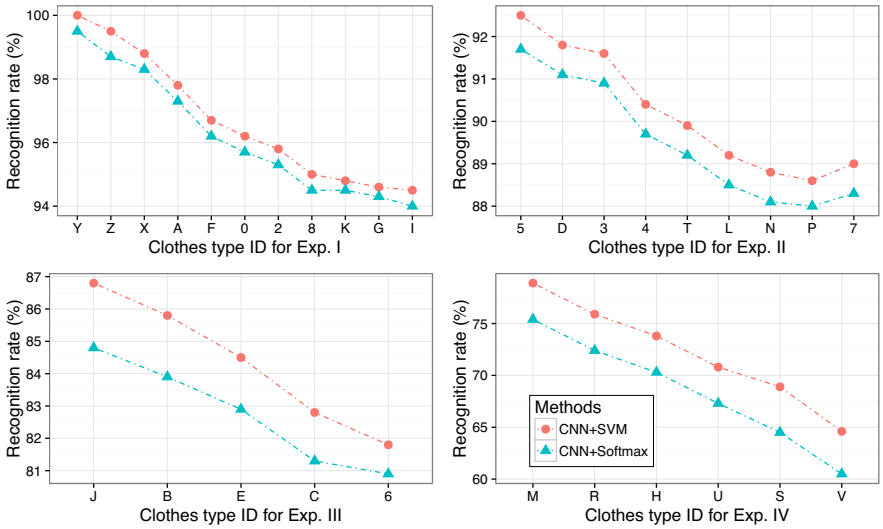


Figure 40: Performance evaluation on the OU-ISIR Treadmill B dataset under the effect of different clothing combinations.

Table 19: Comparison of results on OU-ISIR Treadmill B dataset.

Features	Algorithms	CCR (%)
Hand-crafted	Guan et al. (RSM) [42]	80.44
	Islam et al. (RWSM) [57]	78.54
Deep-learnt	Proposed method (CNN+SVM)	91.38
	Proposed method (CNN+Softmax)	87.80

ing types get more complicated. The performance of CNN in identifying people based on their gait signatures shows that CNN is able to learn and differentiate these non-representational patterns. In order to better understand why CNN works in identifying human based on their gait, further investigation is needed and we leave it for future work.

Comparison with conventional approaches. We apply five-fold cross-validation on this dataset. All the subjects in the dataset are randomly divided into five disjoint groups. In each run, keep one set for testing, and train a network with the remaining sets. Finally, the average correct classification accuracy are reported. Table 20 summarizes the comparison of results with respect to other conventional approaches (hand-crafted features) on the OU-ISIR Treadmill B dataset. Our approaches outperformed the methods proposed by Guan et al. [42] and Islam et al. [57]. Nevertheless, our CNN-based signature extractor has been trained in a fully automatic manner, in contrast to the hand-crafted steps need for computing GEI.

5.4 STACKED PROGRESSIVE AUTO-ENCODERS (SPAE)

In gait recognition, when the combinations between the clothing types is simple, it is easier to obtain an individual's gait patterns. However, if the type of clothing differs between the gallery and the probe, parts of the body seen in the silhouettes are likely to change and the ability to identify subjects decreases with respect to these body parts [53]. In this work, we proposed to extract clothing-invariant gait feature by learning the complex non-linear transform from the most challenging combinations of clothing types to normal condition. Specifically, our proposed method is inspired by the one [62] where a stacked progressive auto-encoders network is proposed to deal with face recognition across different poses. In this work, we propose a solution based on the extension of this principle to deal with the effect of challenging combinations of different clothing covariates. We summarize the contributions of this study as follows.

- We present a model Stacked Progressive Auto-Encoders for clothing-invariant gait recognition named as SPAEGait. It's designed to transform the input gait images from challenging combinations of clothing types to a normal one. The method tries to leverage gradually the information about the different clothing combinations in order to achieve clothing-invariant identification.
- We evaluate the performance of the proposed method on the challenging clothing-invariant of the OU-ISIR Treadmill-B dataset, achieving improved performance compared to other conventional approaches.

5.4.1 *Methodology*

In this section, we propose a solution for a clothing-invariant feature learning based gait recognition framework, which is especially effective for dealing with clothing variations. The proposed framework for gait recognition is represented in Figure 41. We will describe our proposed framework in the following sections.

5.4.1.1 *Input data (GEI)*

In this work, we employ a spatio-temporal gait representation called Gait Energy Image (GEI) [81] as the input raw data of our method. Example GEIs belonging to one subject are shown in Figure 42. GEI is constructed by averaging the silhouette in one complete gait cycle. Given a size-normalized and horizontal aligned human walking

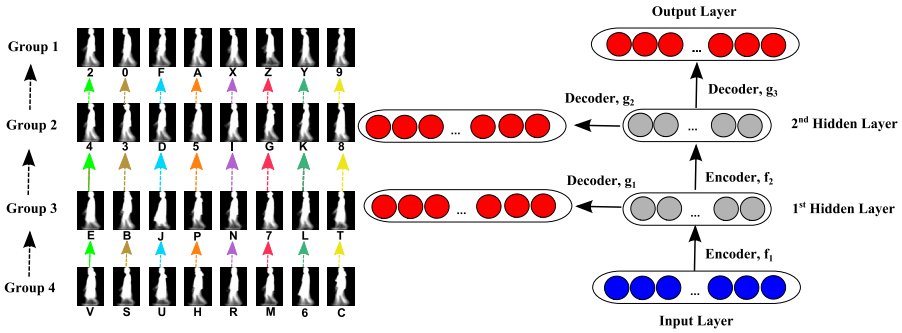


Figure 41: The schema of the proposed Stacked Progressive Auto-Encoders (SPAE) network for clothing-invariant gait recognition. We illustrate a model architecture of the stacked network with two hidden layers, which can deal with the variations in different combination of clothing types ranging from challenging ones to normal conditions. In training stage of our SPAE, each progressive auto-encoder aims at mapping the GEI images at challenging combinations of clothing types to normal condition, while keeping the GEI images with normal combination of clothing type unchanged. In the testing stage, given a GEI image, it is pass into the SPAE network, and the outputs of the topmost hidden layers with easy combination of simple clothing types are used as the clothing-invariant features for gait recognition.

binary silhouette sequenece $B(x, y, t)$, the grey-level GEI $G(x, y)$ is defined as follows

$$G(x, y) = \frac{1}{N} \sum_{t=1}^{N} B(x, y, t), \tag{13}$$

where N is the number of frames in complete cycles of the sequence, x and y are values in the 2D image coordinate.

5.4.1.2 *Auto-Encoder (AE)*

In recent years, auto-encoder [4, 14] played an important role in unsupervised learning and in deep architectures for transfer learning and other tasks. In general, a shallow auto-encoder neural network usually contains two parts: encoder and decoder. In addition, it is usually made up of a input layer, a hidden layer and followed by an output layer [106]. The encoder, denoted as $f(.)$ can transform the input data into a new representation in the hidden layer. It usually comprises of a linear transform and a nonlinear transformation as follows:

$$z = f(x) = s(Wx + b), \tag{14}$$

where W is the linear transform, b is the basis and $s(.)$ is the nonlinear transform, which is also called a element-wise "activation function", such as sigmoid function or tanh function:

$$s(x) = \frac{1}{1 + e^{-x}} \tag{15}$$

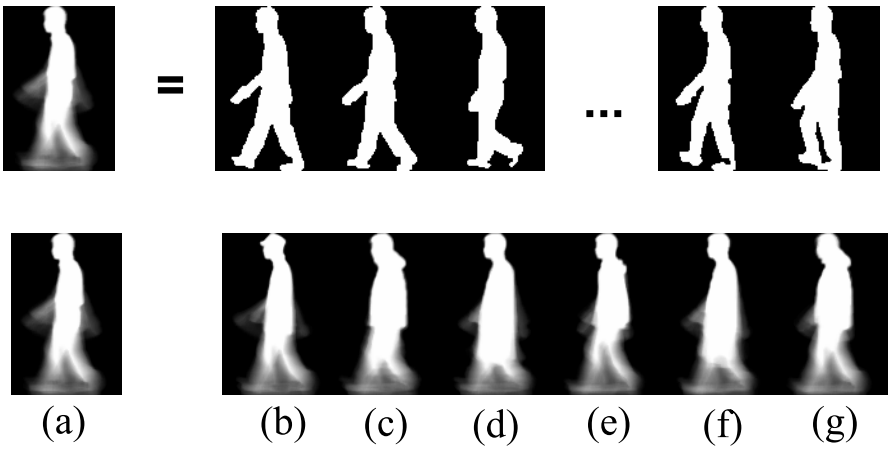


Figure 42: Samples of GEIs of a similar person are computed by averaging the silhouettes over one gait cycle taken from OU-ISIR Treadmill B dataset[53]. The leftmost GEI image (a) is the gallery GEI in normal clothing condition, while the rightmost GEI image (b)-(g) are probe GEIs with different clothing combinations.

or

$$s(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}. \tag{16}$$

The decoder, denoted as g, tries to transform the hidden representation z back to input data x, i.e.,

$$x^{'} = g(z) = s(W^{'}z + b^{'}), \tag{17}$$

where $g(.)$ denotes the decoder, $w^{'}$ and $b^{'}$ denote the linear transformation and basis in decoder and $x^{'}$ is the output data. In general, we usually employ least square error as the cost function to optimize the parameter of w, b, $W^{'}$ and $b^{'}$.

$$[\hat{W}, \hat{b}, \hat{W}', \hat{b}'] = \underset{W,b,W',b'}{\operatorname{argmin}} \sum_{i=1}^{N} \left\| x_i - x_i^{'} \right\|^2 \tag{18}$$

$$= \underset{W,b,W',b'}{\operatorname{argmin}} \sum_{i=1}^{N} \left\| x_i - g(f(x)) \right\|^2, \tag{19}$$

where x_i denotes the ith training sample of N samples and x_i' means the corresponding output of x_i. In our experiments, we train the autoencoder with Stochastic Gradient Descent (SGD). Our implementation for this network is based on the Caffe framework [60].

5.4.1.3 *SPAEGait for clothing variations*

In this paper, we propose a method by stacking multiple progressive auto-encoders together to deal with the effect caused by clothing variations. During the model training, the output is synthesized in a

progressive way. For gait recognition, a combination of simple clothing types contains more dynamic information about one person's gait pattern because the appearance changes are very minimal. Hence, we try to transform all the gait energy images to normal conditions. The first layer of auto-encoders is employed to handle the most challenging combination of clothing types. Then, after some layer of auto-encoders, all GEI images would gradually narrow down the clothing variation to normal condition as shown in Figure 41, improving the accuracy for gait recognition.

The combination of clothing types are {0,2,3,4,...,X,Y,Z}. In the first layer, the auto-encoder would transform the GEI images from challenging combinations of clothing types in Group 4 to Group 3. In other words, this progressive AE narrows down the clothing variations. Meanwhile it keeps the GEI images with normal condition unchanged. Similarly, the auto-encoder in the second layer is designed transform the GEI image which is slightly complicated in Group 3 to Group 2. Finally, the last progressive AE layer would transform all the GEI images to normal conditions but maintain the simple clothing types unchanged.

We train each progressive AE layer separately and the output of a hidden layer is the input of the next successive layer. The whole network is fine tuned by optimizing all layers together after training all the auto-encoders, as below

$$\left[\hat{W}_k|_{k=1}^{L}, \hat{b}_k|_{k=1}^{L}, \hat{W}'_L, \hat{b}'_L\right] = \underset{W_k|_{k=1}^{L}, b_k|_{k=1}^{L}, W'_L, b'_L}{\operatorname{argmin}} \sum_{i=1}^{N} \left\| x'_i - g_L(f_L(f_{L-1}(\ldots(f_1(x_i))))) \right\|^2 , \tag{20}$$

where k is the kth layer in all L layers.

5.4.1.4 *Clothing-invariant feature extraction and recognition*

As the GEIs with clothing variations are transformed to normal ones, the output representation of the topmost layer f_L should be almost reduced down to normal condition. However, the representation emdedded in the lower hidden layers do not contain clothing-invariant features, but only contain very small clothing variations. Therefore, we accumulate the output representation of multiple hidden layers as the clothing-invariant features as follows:

$$F = [f_{L-i}, f_{L-i+1}, \ldots, f_L] , \tag{21}$$

where $0 \leqslant i \leqslant L-1$. The resulting features, F learnt from SPAE-Gait model may not be class discriminative. We then employ Principal Component Analysis (PCA) for dimensionality reduction and the nearest neighbor for classification.

5.4.2 *Experiments*

In this section, we firstly describe the experimental settings for the evaluations including the datasets and implementation details; then investigate our proposed SPAEGait model on the effect of variation of different combination of clothing types; finally, compare it with the existing approaches on the OU-ISIR Treadmill dataset B [53].

5.4.2.1 *Dataset Description*

The OU-ISIR Treadmill dataset B [53] is one of the largest clothing variations gait database. It consists of 68 subjects in total with 15 to 32 different clothing combinations as listed in Table 16, while the list of clothes available in the dataset is shown in Table 17. The original dataset is divided into three subsets, i.e., training set, gallery set and probe set. During the training, same as in conventional approaches, we used a subset containing 446 sequences of 20 subjects with all possible clothing types for training purpose. For testing, we used a gallery set consisting gait sequences of the remaining 48 subjects in standard clothing combination. The probe set consists of 856 gait sequences for these 48 subjects with all types of different clothing combinations excluding the standard one in the gallery set. Figure 37 shows the sample images of different combinations of clothing variation in the dataset.

5.4.2.2 *Implementation Details*

In the experiments, the structure of our proposed autoencoder with 2 fully-connected hidden layers as shown in Figure 41 are used. During the training phase, we used the Caffe software [60], a very popular open source deep learning framework to train the model. Particularly, each progressive auto-encoder needs to be trained separately. The weights are initialized to random values using Gaussian distribution with zero mean and standard deviation of 0.01 for all trainable layers. All the bias terms are initialized with the constant zero. We set our base learning rate to 0.1. We reduced the learning rate for all layers by a factor of 10 every 1000 iterations prior to termination. We trained the model for 10,000 iterations and the activation function is sigmoid for all experiments. After this, the two hidden layers have been trained then combined in a stacked way, and fine-tuned as a whole in the model. In the fine-tuning, the base learning rate was set to 0.01 and the maximum number of iterations was 5000. One important parameter to the proposed model is the numbers of hidden layer neurons. Following the work in [13], our model is evaluated with different numbers of neurons from 500 to 3000. From the evaluation, we noticed that when the number is 2000 the model achieves the best recognition rate. So, we set this as the number of neurons for

each hidden layer in the model. Besides the feature from the topmost layer, we can also select features from lower layers as the invariant gait feature for gait recognition. We observed that the feature consisting of the multiple hidden layers did not help to improve recognition rates. However, in our case, the features from the topmost hidden layer achieved the best performance with a significant improvement. The final step of the invariant gait feature extraction was to reduce the dimension using PCA, same as in [62]. The feature dimension was reduced from 2000 to 100. The value 100 was chosen according the experiments. The features obtained from the topmost encoding layer are then compressed using PCA and used as input to a nearest-neighbour classifier (NN).

5.4.3 *Results and Discussions*

5.4.3.1 *On the effect of clothing variations*

We conducted all our experiments on the OU-ISIR Treadmill B dataset to examine the performance of our proposed approach with the effect of different clothing types. The performance for the proposed method can be observed in Figure 43. From the experiment results, we have the following observations:

In experiment 1 (Exp. 1), consisting of walking sequences that only involve normal clothing combinations (e.g., Y:regular pants+full shirt, 2:regular pants+half shirt, etc), we can see that the average CCR is above 98.5%. It is also clear from Figure 43 that the gait pattern is not affected by an appearance change caused by easy combinations of clothing types. Apart from the basic clothing combinations in Exp. 1, we can observe that in Experiment 2 (Exp. 2), the clothing combinations consists of walking sequences that involve some headwear covariates (e.g., 3:regular pants+half shirt+hat, P:regular pants+half shirt+casquette cap, etc). In this case, we are still able to achieve relatively high performance. The average CCR is above 94.2%. The headwear covariates does not affect the performance of gait recognition significantly.

In Experiment 3 (Exp. 3), we analyze the scenario of clothing combination type B and E when a complicated clothing type (e.g, down jacket or long coat) is combined with a basic clothing type (e.g., casual or regular pants), the recognition task becomes much more complicated. However, based on results obtained by the proposed method, the performance still yields competitive results, with an average recognition rate above than 85.5%. The combinations between the challenging clothing types shown in Experiment 4 (Exp. 4) lead to a decrease in gait recognition accuracy performance. We can observe the average CCR is reduced further to 77.3%. For instance, given a query gait pattern under the effect of difficult combinations of clothing types S, the recognition rate decreased to around 73.5%. Similarly, for clothing

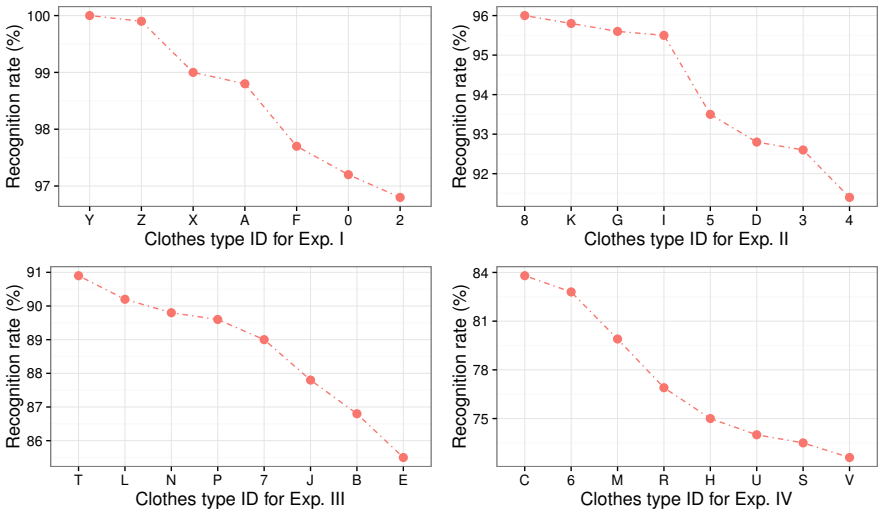


Figure 43: Performance evaluation on the OU-ISIR Treadmill B dataset under the effect of different clothing combinations.

type V, the recognition rate decreased to around 72.6%. Notice that for the same task a human would also have difficulties as clothes of this type occlude many discriminative features, nevertheless our proposed method is still able to recognize in most cases the individual's gait.

In general, these experiments revealed that the proposed approach is able to provide high classification accuracy even when the combination of clothing types get more complicated. The performance of our method in identifying subjects based on gait under the effect of clothing variations shows that the proposed stacked progressive AE is able to learn clothing-invariant features by converting challenging combinations clothing types to basic ones progressively which, in turn, means the clothing variations are narrowed down eventually.

5.4.3.2 *Comparison with conventional approaches*

The experimental results have been compared with the ones obtained by other approaches available in the literature. Table 20 summarizes the comparison of results with respect to other conventional approaches on the OU-ISIR Treadmill B dataset. It can been seen that our method displays good results, especially for the challenging clothing types. This can be explained by the fact that our method gradually narrows down the clothing variations which, in turn, improves the recognition performance in the challenging clothing combination conditions when these features are considerably affected. In addition, the overall performance outperforms all the state-of-the-art methods considered in our experiments [42, 57, 121]. This further demonstrates the specific effectiveness of using a stacked auto-encoder on gait recognition with clothing variations.

Table 20: Comparison of results on OU-ISIR Treadmill B dataset.

Algorithms	**CCR (%)**
Guan et al. (RSM) [42]	80.44
Islam et al. (RWSM) [57]	78.54
Yeoh et al. (CNN+SVM) [121]	91.38
Yeoh et al. (CNN+Softmax) [121]	87.80
Proposed method (SPAEGait)	**94.80**

5.5 SUMMARY

In this section, we employed convolutional neural network to build an end-to-end learning framework for clothing-invariant gait recognition. The experimental validation has been carried out on the challenging CASIA-B and OU-ISIR-B dataset, by using gait energy image (GEI) as a raw input to a CNN network. We designed convolutional neural network model so that it can extract a clothing-invariant and discriminative feature from the input of GEI images. The experimental results showed that the proposed method outperformed the state-of-the-art methods for both dataset.

We achieved better results than the previous conventional approaches even we trained a CNN from scratch. We also plan to investigate various fine tuning strategies to transfer learned recognition capabilities from general domains to the specific challenge of gait recognition task. In our future work, we are interested in visualizing the features in the CNN layers [127] for a better understanding of how CNN works in learning patterns in gait under various clothing types effect. In addition, we plan to extend our study by using our proposed approach to other large-scale datasets for gait recognition. Our research can also be extended to evaluate pre-trained CNNs such as AlexNet, GoogLeNet, ResNet and etc. It would be ideal to experiment whether these architectures can yield better classification accuracy as compared to the proposed approach. We also plan to investigate further the use deep network such as ResNet [49, 38] for our problem. This is because recently it has been reported that by training using deeper networks can obtain promising results.

We also have presented a stacked progresssive auto-encoders model to extract clothing-invariant gait feature for gait recognition under various clothing covariates. The proposed framework can tranform gradually GEI images from a challenging combination of clothing types to the normal ones by multiple shallow progressive auto-encoders. These features extracted from few topmost layers of stacked AE network only contain very small clothing variations, that are further

integrated with PCA for clothing-invariant gait recognition. The experiments were performed on a well-known and the largest clothing variation gait dataset, namely the OU-ISIR Treadmill B dataset. Experimental results show that our model can effectively improve recognition rate by reducing the clothing variations especially when there is a large clothing variation and achieves a far better performance over existing approaches.

As a future work, we intend to extend our approach to cope with more challenging covariate variations. Moreover, it will be interesting to see how our approach can be modified to accomodate the case of gait identification under the effect of various covariate factors. Besides, using another large gait database can also be investigated to evaluate the effectiveness of our method.

Part V

CONCLUSIONS

6

CONCLUSIONS AND FUTURE WORK

This chapter provides conclusions for each of the research objectives of this thesis and gives main findings and highlights from each individual chapter, then presents potential research areas for future work.

6.1 CONCLUSIONS

Gait is a relatively new and emergent biometric that pertains to the use of an individual's walking style to it. In recent years, human identification based on gait has received enormous interest due to their ability to apply it in automated visual security and surveillance systems. However, variations in covariate factors such as cloth, walking speed, object carrying, footwear and view angle can alter an individual's gait pattern. These variations make the task of gait analysis became much more complicated.

In this thesis, we have focused on covariate factor analysis by reducing their influence in gait recognition. We carefully analysed their effect and summarised the challenges to this problem. We proposed covariate-invariant methods to deal with the presence of various factors in order to mitigate their effect. Our proposed algorithms were evaluated on standard gait databases with various covariate factors (e.g., carrying condition, clothing, camera viewpoint, etc.) and experimental results confirmed their effectiveness. We also discussed several important and practical problems in gait recognition (e.g., feature representation, feature selection and feature learning) and provide solutions to them.

6.2 CONTRIBUTIONS AND DISCUSSIONS

We summarise our main contributions and the important research findings, which is helpful for the future research in gait recognition and related areas as follows.

1. In Chapter 2, we have provided a detailed literature review of gait recognition frameworks, feature representation categories, existing databases and the related works.

2. In Chapter 3, we have presented a robust approach for extracting human gait features from human silhouette images. The features are extracted by determining the joints from the body segments based on a priori knowledge of human body proportion. Once the body joints have been identified, the joint trajec-

tories can be computed using a straightforward approach. This approach has shown to be more effective as it is capable to to identify the joints from self-occluded human silhouettes. Then, the proposed feature post-process approach mitigates the effect of variation in covariate factors which complicate the individual's gait recognition process. Our experimental results demonstrated the proposed approach significantly outperforms the existing conventional techniques. In addition, the higher recognition rate also showed that proposed approach is robust and can perform well under different covariate factors.

3. In Chapter 4, we presented and compared feature selection techniques, designed to maximise the final classification accuracy. The overall approach constitutes a general framework for different machine learning algorithms, which we applied to the problem of feature selection under the effect of various covariate factors in a model-based approach. The experiments were performed on two-well known gait datasets, namely the SOTON covariate dataset and the CASIA-B dataset. The experimental results on both showed that the inevestigated approach is able to select the discriminative input gait features and achieve an improved classification accuracy on par with other state-of-the-art methods.

4. In Chapter 5, we employed convolutional neural networks to build an end-to-end learning framework for clothing and view-invariant gait recognition. The experimental validation has been carried out on the challenging dataset of OU-ISIR B treadmill and CASIA-B dataset, by using the gait energy image (GEI) as raw input to the CNN network. We designed a structure of CNN so that it can extract an invariant and discriminative features against clothing and view covariates. Experimental results showed that our method can achieves a far better performance over existing approaches in both these databases. In the later part of this chapter, we have presented a stacked progressive auto-encoders model to extract clothing-invariant gait features for gait recognition under various clothing covariates. The proposed framework can transform gradually GEI images from a combination of clothing types to more simpler ones by multiple shallow progressive auto-encoders. These features extracted from few topmost layers of the stacked AE network only contain very small clothing variations, that are further integrated with PCA for clothing-invariant gait recognition. The experiments were performed on a well-known and the largest clothing variation gait dataset, namely OU-ISIR Treadmill B dataset. Experiment results suggested that our model can effectively improve recognition rate by reducing the clothing variations especially

when there is a large variation and achieves a far better performance over existing approaches.

6.3 FUTURE RESEARCH DIRECTIONS

This thesis has contributed to several areas of a gait recognition system, however there are still many areas could address. Areas of future work, that could further improve techniques proposed in this thesis, as well potential new research are listed below:

6.3.1 *Investigations in the short-term*

1. **Experimental evaluation on large-scale gait databases**
   To ensure the gait recognition system's applicability on a real world environment it needs to be evaluated on large scale data. Thus, we intend to extend our work to large-scale gait dataset in order to examine how the results will generalize to larger data sets under various different covariate factors. So, it will be more meaningful to test the proposed method to show robustness with respect to different covariate factors potentially affecting performance.

2. **Transfer learning using convolutional neural network**
   Given insufficient training samples from small gait dataset, transferring feature representations learned from a larger gait dataset becomes necessary. Hence, as future work, we think fine-tuning our existing network with transfer learning could be more faster and easier than training a network with randomly initialized weights from scratch to any new gait dataset. Thus, we should go beyond existing gait datasets and consider much larger sources.

3. **Combining gait with other biometric traits**
   Face and iris are the commonly used biometrics in access control and identity management applications. A gait recognition system can be developed that combines these biometrics to provide a non-intrusive verification process and improve recognition performance. The fusion techniques within the different biometric modalities will be the important issues for future research.

6.3.2 *Investigations in the long-term*

1. **Gait recognition using shadow analysis**
   The identification of human beings based on their biometric body parts such as face, fingerprint, gait, iris, and voice, etc. Existing traditional biometrics rely on the direct observation (e.g. image of face/body), but it may be the case that a projection

may have more information than the direct signal. For example, the shadow of a person observed from a higher point, or from an overhead camera. To identify people based on their shadow, gait may be used to extract human characteristics. It is possible to exploit this biometric information in human shadow silhouettes for recognition of human identity.

2. **Towards efficient clinical gait analysis with wearable sensors** After decades of evolution, measuring instruments for quantitative gait analysis have become an important clinical tool for assessing pathologies manifested by gait abnormalities. However, such instruments tend to be expensive and require expert operation and maintenance besides their high cost, thus limiting them to only a small number of specialized centers. Consequently, gait analysis in most clinics today still rely on observation-based assessment. Recent advances in wearable sensors, especially inertial body sensors, have opened up a promising future for gait analysis. Not only can these sensors be more easily adopted in clinical diagnosis and treatment procedures than their current counterparts, but they can also monitor gait continuously outside clinics. Hence, we plan to extend our work to provide seamless patient analysis from clinics to free-living environments.

Part VI

APPENDICES

A

DATASET

The details of the datasets were presented in Section 2.3. This appendix shows additional samples of usage samples. Usable samples are defined as walking sequences in which the subject performs at least one full gait cycle.

A.1 SOTON COVARIATE DATABASE

Figure 44
shows the examples of silhouette data quality of the database.

A.2 OU-ISIR TREADMILL B DATASET

Figure 45
shows the examples of silhouette data quality of the database.

A.3 CASIA-B DATASET

Figure 46
shows the examples of silhouette data quality of the database.

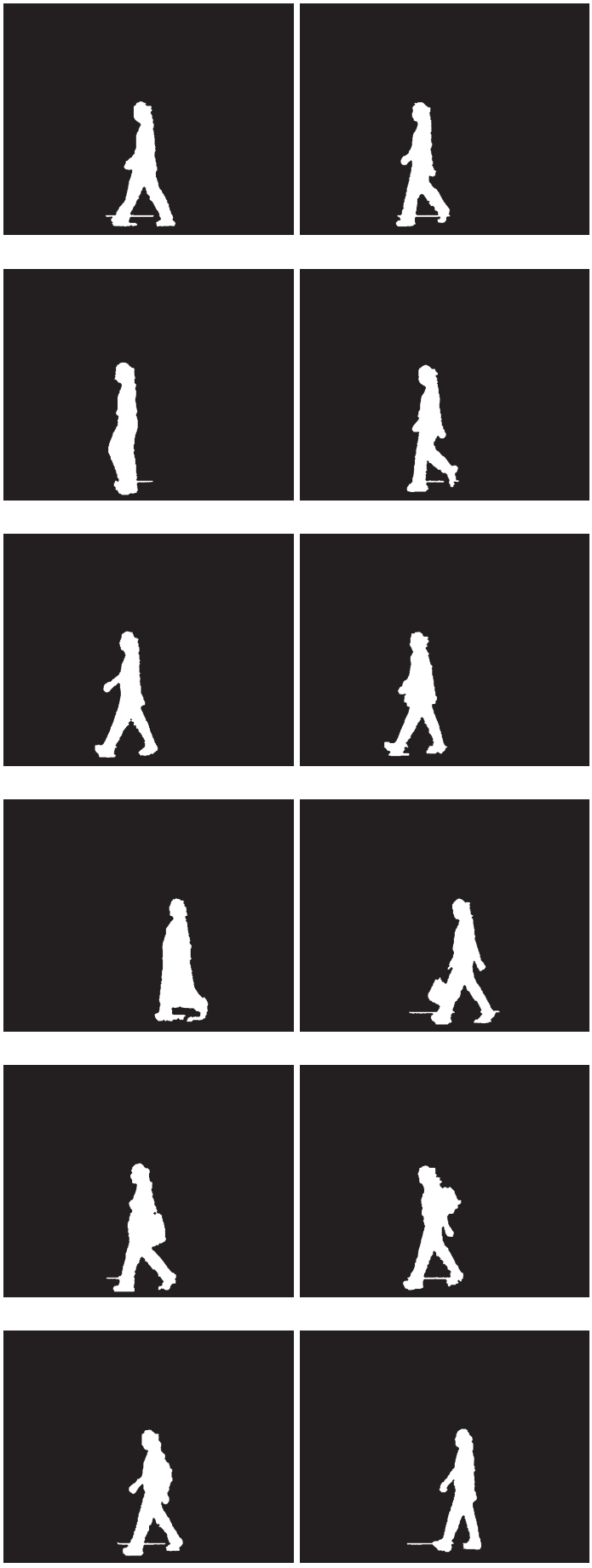


Figure 44: Pre-processed silhouettes with missing body parts, noise and shadows

Figure 45: Pre-processed silhouettes with missing body parts, noise and shadows

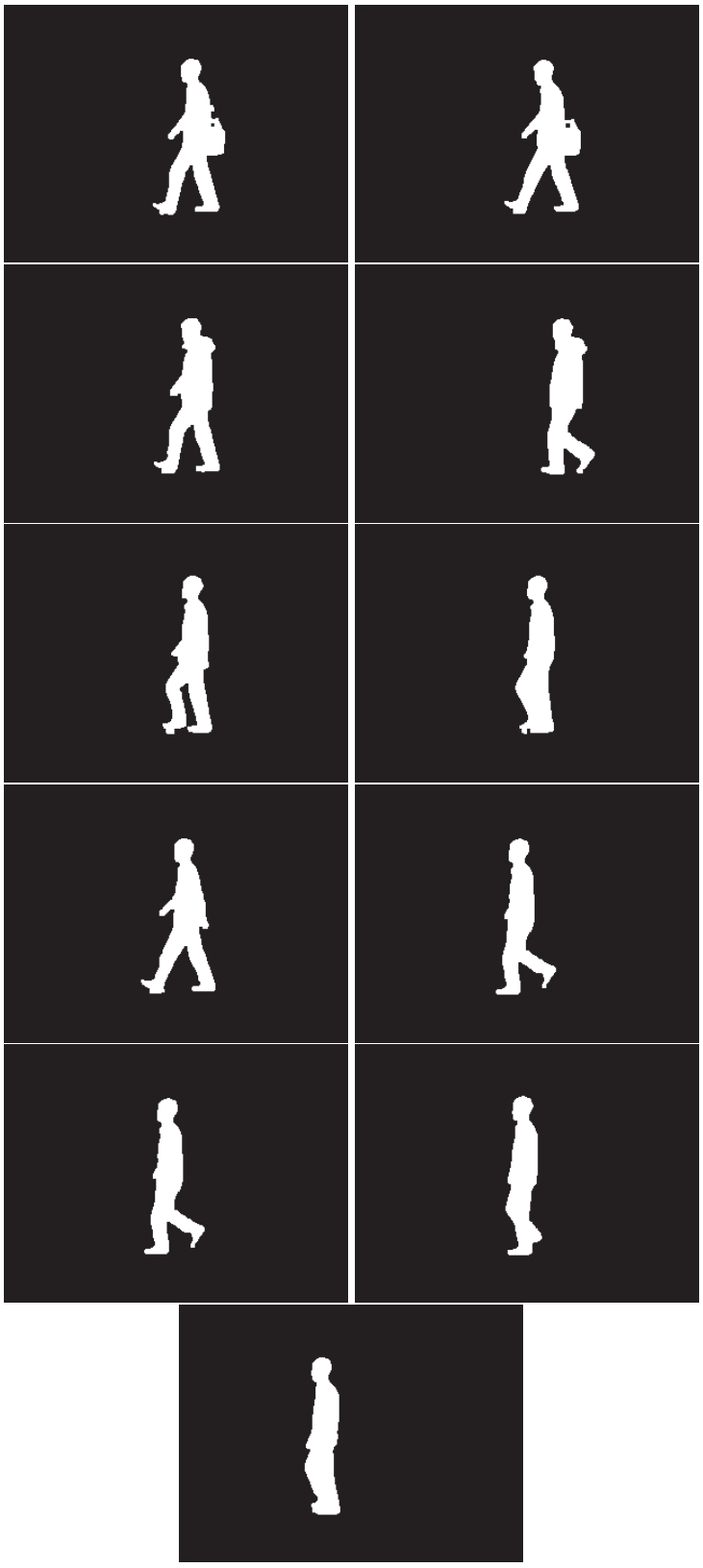


Figure 46: Pre-processed silhouettes with missing body parts, noise and shadows

BIBLIOGRAPHY

[1] Selim Aksoy and Robert M Haralick. Feature normalization and likelihood-based similarity measures for image retrieval. *Pattern recognition letters*, 22(5):563–582, 2001.

[2] Rosa Altilio, Luca Liparulo, Andrea Proietti, Marco Paoloni, and Massimo Panella. A genetic algorithm for feature selection in gait analysis. In *Evolutionary Computation (CEC), 2016 IEEE Congress on*, pages 4584–4591. IEEE, 2016.

[3] Moez Baccouche, Franck Mamalet, Christian Wolf, Christophe Garcia, and Atilla Baskurt. Sequential deep learning for human action recognition. In *International Workshop on Human Behavior Understanding*, pages 29–39. Springer, 2011.

[4] Pierre Baldi. Autoencoders, unsupervised learning, and deep architectures. In *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, pages 37–49, 2012.

[5] Dale J Barr, Roger Levy, Christoph Scheepers, and Harry J Tily. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3):255–278, 2013.

[6] David Barrett. One surveillance camera for every 11 people in britain, says cctv survey. *The Telegraph*, 10, 2013.

[7] Khalid Bashir, Tao Xiang, and Shaogang Gong. Feature selection on gait energy image for human identification. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, pages 985–988. IEEE, 2008.

[8] Khalid Bashir, Tao Xiang, and Shaogang Gong. Gait recognition without subject cooperation. *Pattern Recognition Letters*, 31(13):2052–2060, 2010.

[9] Khalid Bashir, Tao Xiang, Shaogang Gong, and Q Mary. Gait representation using flow fields. In *BMVC*, pages 1–11, 2009.

[10] Matthieu Basseur and Adrien Goëffon. Hill-climbing strategies on various landscapes: an empirical comparison. In *Proceedings of the 15th annual conference on Genetic and evolutionary computation*, pages 479–486. ACM, 2013.

[11] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*, 2014.

[12] Chiraz BenAbdelkader, Ross Cutler, Harsh Nanda, and Larry Davis. Eigengait: Motion-based recognition of people using image self-similarity. In *Audio-and Video-Based Biometric Person Authentication*, pages 284–294. Springer, 2001.

[13] Yoshua Bengio. Practical recommendations for gradient-based training of deep architectures. In *Neural networks: Tricks of the trade*, pages 437–478. Springer, 2012.

[14] Yoshua Bengio et al. Learning deep architectures for ai. *Foundations and trends® in Machine Learning*, 2(1):1–127, 2009.

[15] Mohamed Bennasar, Yulia Hicks, and Rossitza Setchi. Feature selection using joint mutual information maximisation. *Expert Systems with Applications*, 42(22):8520–8532, 2015.

[16] Aaron F Bobick and Amos Y Johnson. Gait recognition using static, activity-specific parameters. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–I. IEEE, 2001.

[17] Imed Bouchrika, John N Carter, and Mark S Nixon. Towards automated visual surveillance using gait for identity recognition and tracking across multiple non-intersecting cameras. *Multimedia Tools and Applications*, 75(2):1201–1221, 2016.

[18] Imed Bouchrika, Michaela Goffredo, John Carter, and Mark Nixon. On using gait in forensic biometrics. *Journal of forensic sciences*, 56(4):882–889, 2011.

[19] Imed Bouchrika and Mark S Nixon. Model-based feature extraction for gait analysis and recognition. In *International Conference on Computer Vision/Computer Graphics Collaboration Techniques and Applications*, pages 150–160. Springer, 2007.

[20] Imed Bouchrika and Mark S Nixon. Exploratory factor analysis of gait recognition. In *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*, pages 1–6. IEEE, 2008.

[21] L Breiman, J Friedman, R Olshen, and C Storne. Classification and regression trees. belmont, ca: Wadsworth international group; 1984. *Google Scholar*.

[22] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[23] Leo Breiman. Manual on setting up, using, and understanding random forests v3. 1. *Statistics Department University of California Berkeley, CA, USA*, 1, 2002.

[24] Christopher JC Burges. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167, 1998.

[25] Chih-Chung Chang and Chih-Jen Lin. Libsvm: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, 2(3):27, 2011.

[26] Sharan Chetlur, Cliff Woolley, Philippe Vandermersch, Jonathan Cohen, John Tran, Bryan Catanzaro, and Evan Shelhamer. cudnn: Efficient primitives for deep learning. *arXiv preprint arXiv:1410.0759*, 2014.

[27] Marco Chiarandini and Yuri Goegebeur. Mixed models for the analysis of optimization algorithms. *Experimental Methods for the Analysis of Optimization Algorithms*, 1:225, 2010.

[28] Robert T Collins, Ralph Gross, and Jianbo Shi. Silhouette-based human identification from body shape and gait. In *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pages 366–371. IEEE, 2002.

[29] Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, 12(Aug):2493–2537, 2011.

[30] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[31] Thomas Cover and Peter Hart. Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1):21–27, 1967.

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

[33] Brian DeCann and Arun Ross. Gait curves for human recognition, backpack detection, and silhouette correction in a nighttime environment. *Proc. SPIE, Biometric Technology for Human Identification VII*, 7667:76670Q–76670Q, 2010.

[34] Wilfrid Taylor Dempster and George RL Gaughran. Properties of body segments based on size and weight. *Developmental Dynamics*, 120(1):33–54, 1967.

[35] Li Deng. A tutorial survey of architectures, algorithms, and applications for deep learning. *APSIPA Transactions on Signal and Information Processing*, 3, 2014.

[36] Jeffrey Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, and Trevor Darrell. Long-term recurrent convolutional networks for visual recognition and description. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2625–2634, 2015.

[37] Yohan Dupuis, Xavier Savatier, and Pascal Vasseur. Feature subset selection applied to model-free gait recognition. *Image and vision computing*, 31(8):580–591, 2013.

[38] Mohammad Sadegh Ebrahimi and Hossein Karkeh Abadi. Study of residual networks for image recognition.

[39] Clement Farabet, Camille Couprie, Laurent Najman, and Yann LeCun. Learning hierarchical features for scene labeling. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1915–1929, 2013.

[40] Pierre Geurts, Damien Ernst, and Louis Wehenkel. Extremely randomized trees. *Machine learning*, 63(1):3–42, 2006.

[41] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep learning. book in preparation for mit press. *URL< http://www. deeplearningbook. org*, 2016.

[42] Yu Guan, Chang-Tsun Li, and Yongjian Hu. Robust clothing-invariant gait recognition. In *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2012 Eighth International Conference on*, pages 321–324. IEEE, 2012.

[43] Yu Guan, Chang-Tsun Li, and Fabio Roli. On reducing the effect of covariate factors in gait recognition: a classifier ensemble method. *IEEE transactions on pattern analysis and machine intelligence*, 37(7):1521–1528, 2015.

[44] Baofeng Guo and Mark S Nixon. Gait feature subset selection by mutual information. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 39(1):36–46, 2009.

[45] Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar):1157–1182, 2003.

[46] Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1):389–422, 2002.

[47] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.

[48] David J Hand et al. Classifier technology and the illusion of progress. *Statistical science*, 21(1):1–14, 2006.

[49] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[50] Martin Hofmann, Jürgen Geiger, Sebastian Bachmann, Björn Schuller, and Gerhard Rigoll. The tum gait from audio, image and depth (gaid) database: Multimodal recognition of subjects and traits. *Journal of Visual Communication and Image Representation*, 25(1):195–206, 2014.

[51] Martin Hofmann, Stephan M Schmidt, A N Rajagopalan, and Gerhard Rigoll. Combined face and gait recognition using alpha matte preprocessing. In *Biometrics (ICB), 2012 5th IAPR International Conference on*, pages 390–395. IEEE, 2012.

[52] Emdad Hossain and Girija Chetty. Multimodal feature learning for gait biometric based human identity recognition. In *International Conference on Neural Information Processing*, pages 721–728. Springer, 2013.

[53] Md Altab Hossain, Yasushi Makihara, Junqiu Wang, and Yasushi Yagi. Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control. *Pattern Recognition*, 43(6):2281–2291, 2010.

[54] Torsten Hothorn, Frank Bretz, and Peter Westfall. Simultaneous inference in general parametric models. *Biometrical journal*, 50(3):346–363, 2008.

[55] Maodi Hu, Yunhong Wang, Zhaoxiang Zhang, De Zhang, and James J Little. Incremental learning for video-based gait recognition with lbp flow. *IEEE transactions on cybernetics*, 43(1):77–89, 2013.

[56] David H Hubel and Torsten N Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, 160(1):106–154, 1962.

[57] Md Shariful Islam, Md Rabiul Islam, Most Sheuli Akter, MA Hossain, and MKI Molla. Window based clothing invariant gait recognition. In *Advances in Electrical Engineering (ICAEE), 2013 International Conference on*, pages 411–414. IEEE, 2013.

[58] Anil K Jain, Arun Ross, and Salil Prabhakar. An introduction to biometric recognition. *IEEE Transactions on circuits and systems for video technology*, 14(1):4–20, 2004.

[59] Mahadevu Jeevan, Neha Jain, Madasu Hanmandlu, and Girija Chetty. Gait recognition based on gait pal and pal entropy image. In *Image Processing (ICIP), 2013 20th IEEE International Conference on*, pages 4195–4199. IEEE, 2013.

[60] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 675–678. ACM, 2014.

[61] George H John, Ron Kohavi, Karl Pfleger, et al. Irrelevant features and the subset selection problem. In *Machine learning: proceedings of the eleventh international conference*, pages 121–129, 1994.

[62] Meina Kan, Shiguang Shan, Hong Chang, and Xilin Chen. Stacked progressive auto-encoders (spae) for face recognition across poses. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1883–1890, 2014.

[63] Stefan Knerr, Léon Personnaz, and Gérard Dreyfus. Single-layer learning revisited: a stepwise procedure for building and training a neural network. *Neurocomputing: algorithms, architectures and applications*, 68(41-50):71, 1990.

[64] Ron Kohavi and George H John. Wrappers for feature subset selection. *Artificial intelligence*, 97(1-2):273–324, 1997.

[65] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. *Physical review E*, 69(6):066138, 2004.

[66] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[67] Worapan Kusakunniran. Attribute-based learning for gait recognition using spatio-temporal interest points. *Image and Vision Computing*, 32(12):1117–1126, 2014.

[68] Worapan Kusakunniran. Recognizing gaits on spatio-temporal feature domain. *IEEE Transactions on Information Forensics and Security*, 9(9):1416–1423, 2014.

[69] Worapan Kusakunniran, Qiang Wu, Jian Zhang, and Hongdong Li. Support vector regression for multi-view gait recognition based on local motion feature selection. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 974–981. IEEE, 2010.

[70] Peter K Larsen, Erik B Simonsen, and Niels Lynnerup. Gait analysis in forensic medicine. *Journal of forensic sciences*, 53(5):1149–1153, 2008.

[71] Yann LeCun et al. Lenet-5, convolutional neural networks. *URL: http://yann. lecun. com/exdb/lenet*, 2015.

[72] Lily Lee and W Eric L Grimson. Gait analysis for recognition and classification. In *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pages 155–162. IEEE, 2002.

[73] Tracey KM Lee, Mohammed Belkhatir, and Saeid Sanei. A comprehensive review of past and present vision-based techniques for gait recognition. *Multimedia tools and applications*, 72(3):2833–2869, 2014.

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

[75] Gilles Louppe. Understanding random forests: From theory to practice. *arXiv preprint arXiv:1407.7502*, 2014.

[76] Gilles Louppe, Louis Wehenkel, Antonio Sutera, and Pierre Geurts. Understanding variable importances in forests of randomized trees. In *Advances in neural information processing systems*, pages 431–439, 2013.

[77] Frederik Maes, Andre Collignon, Dirk Vandermeulen, Guy Marchal, and Paul Suetens. Multimodality image registration by maximization of mutual information. *IEEE transactions on medical imaging*, 16(2):187–198, 1997.

[78] Yasushi Makihara, Hidetoshi Mannami, Akira Tsuji, Md Altab Hossain, Kazushige Sugiura, Atsushi Mori, and Yasushi Yagi. The ou-isir gait database comprising the treadmill dataset. *IPSJ Transactions on Computer Vision and Applications*, 4:53–62, 2012.

[79] Yasushi Makihara, Darko S Matovski, Mark S Nixon, John N Carter, and Yasushi Yagi. Gait recognition: Databases, representations, and applications. *Wiley Encyclopedia of Electrical and Electronics Engineering*, 2015.

[80] Yasushi Makihara, Ryusuke Sagawa, Yasuhiro Mukaigawa, Tomio Echigo, and Yasushi Yagi. Gait recognition using a view transformation model in the frequency domain. *Computer Vision–ECCV 2006*, pages 151–163, 2006.

[81] Ju Man and Bir Bhanu. Individual recognition using gait energy image. *IEEE transactions on pattern analysis and machine intelligence*, 28(2):316–322, 2006.

[82] Raúl Martín-Félez and Tao Xiang. Uncooperative gait recognition by learning to rank. *Pattern Recognition*, 47(12):3793–3806, 2014.

[83] Masakazu Matsugu, Katsuhiko Mori, Yusuke Mitari, and Yuji Kaneda. Subject independent facial expression recognition with robust face detection using a convolutional neural network. *Neural Networks*, 16(5):555–559, 2003.

[84] Anup Nandy, Rupak Chakraborty, and Pavan Chakraborty. Cloth invariant gait recognition using pooled segmented statistical features. *Neurocomputing*, 191:117–140, 2016.

[85] Hu Ng, Hau-Lee Ton, Wooi-Haw Tan, Timothy Tzen-Vun Yap, Pei-Fen Chong, and Junaidi Abdullah. Human identification based on extracted gait features. *International Journal of New Computer Architectures and their Applications (IJNCAA)*, 1(2):358–370, 2011.

[86] Mark S Nixon, Tieniu Tan, and Rama Chellappa. *Human identification based on gait*, volume 4. Springer Science & Business Media, 2010.

[87] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(Oct):2825–2830, 2011.

[88] Florent Perronnin and Diane Larlus. Fisher vectors meet neural networks: A hybrid classification architecture. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3743–3752, 2015.

[89] P Jonathon Phillips, Sudeep Sarkar, Isidro Robledo, Patrick Grother, and Kevin Bowyer. The gait identification challenge problem: Data sets and baseline algorithm. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, volume 1, pages 385–388. IEEE, 2002.

[90] Yogarajah Pratheepan, Joan V Condell, and Girijesh Prasad. Individual identification using gait sequences under different covariate factors. In *International Conference on Computer Vision Systems*, pages 84–93. Springer, 2009.

[91] Imad Rida, Somaya Almaadeed, and Ahmed Bouridane. Gait recognition based on modified phase-only correlation. *Signal, Image and Video Processing*, 10(3):463–470, 2016.

[92] Imad Rida, Ahmed Bouridane, Gian Luca Marcialis, and Pierluigi Tuveri. Improved human gait recognition. In *International Conference on Image Analysis and Processing*, pages 119–129. Springer, 2015.

[93] Imad Rida, Xudong Jiang, and Gian Luca Marcialis. Human body part selection by group lasso of motion for model-free gait recognition. *IEEE Signal Processing Letters*, 23(1):154–158, 2016.

[94] Sudeep Sarkar, P Jonathon Phillips, Zongyi Liu, Isidro Robledo Vega, Patrick Grother, and Kevin W Bowyer. The humanid gait challenge problem: Data sets, performance, and analysis. *IEEE transactions on pattern analysis and machine intelligence*, 27(2):162–177, 2005.

[95] Bernhard Schölkopf, Christopher JC Burges, and Alexander J Smola. *Advances in kernel methods: support vector learning*. MIT press, 1999.

[96] Vijay Bhaskar Semwal, Joyeeta Singha, Pinki Kumari Sharma, Arun Chauhan, and Basudeba Behera. An optimized feature selection technique based on incremental feature analysis for bio-metric gait data classification. *Multimedia Tools and Applications*, pages 1–19, 2016.

[97] Jamie D Shutler, Michael G Grant, Mark S Nixon, and John N Carter. On a large sequence-based human gait database. In *Applications and Science in Soft Computing*, pages 339–346. Springer, 2004.

[98] Karen Simonyan and Andrew Zisserman. Two-stream convolutional networks for action recognition in videos. In *Advances in neural information processing systems*, pages 568–576, 2014.

[99] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[100] Kenneth Sörensen. Metaheuristics—the metaphor exposed. *International Transactions in Operational Research*, 22(1):3–18, 2015.

[101] Zehang Sun, George Bebis, and Ronald Miller. Object detection using feature subset selection. *Pattern recognition*, 37(11):2165–2176, 2004.

[102] Rawesak Tanawongsuwan and Aaron Bobick. Gait recognition from time-normalized joint-angle trajectories in the walking plane. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 2, pages II–II. IEEE, 2001.

[103] Yichuan Tang. Deep learning using linear support vector machines. *arXiv preprint arXiv:1306.0239*, 2013.

[104] Dacheng Tao, Xuelong Li, Stephen J Maybank, and Xindong Wu. Human carrying status in visual surveillance. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 1670–1677. IEEE, 2006.

[105] Andrea Vedaldi and Karel Lenc. Matconvnet: Convolutional neural networks for matlab. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 689–692. ACM, 2015.

[106] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, 11(Dec):3371–3408, 2010.

[107] David Kenneth Wagg and Mark S Nixon. On automated model-based extraction and analysis of gait. In *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, pages 11–16. IEEE, 2004.

[108] Ai-Hua Wang and Ji-Wei Liu. A gait recognition method based on positioning human body joints. In *Wavelet Analysis and Pattern Recognition, 2007. ICWAPR'07. International Conference on*, volume 3, pages 1067–1071. IEEE, 2007.

[109] Chen Wang, Junping Zhang, Liang Wang, Jian Pu, and Xiaoru Yuan. Human identification using temporal information preserving gait template. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2164–2176, 2012.

[110] Jin Wang, Mary She, Saeid Nahavandi, and Abbas Kouzani. A review of vision-based gait recognition methods for human identification. In *Digital Image Computing: Techniques and Applications (DICTA), 2010 International Conference on*, pages 320–327. IEEE, 2010.

[111] Junqiu Wang, Yasushi Makihara, and Yasushi Yagi. People tracking and segmentation using spatiotemporal shape constraints. In *Proceedings of the 1st ACM workshop on Vision networks for behavior analysis*, pages 31–38. ACM, 2008.

[112] Limin Wang, Yu Qiao, and Xiaoou Tang. Action recognition with trajectory-pooled deep-convolutional descriptors. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4305–4314, 2015.

[113] Tenika Whytock, Alexander Belyaev, and Neil M Robertson. On covariate factor detection and removal for robust gait recognition. *Machine Vision and Applications*, 26(5):661–674, 2015.

[114] Tenika P Whytock, Alexander Belyaev, and Neil M Robertson. Towards robust gait recognition. In *International Symposium on Visual Computing*, pages 523–531. Springer, 2013.

[115] TP Whytock, A Belyaev, and NM Robertson. Robust gait recognition via covariate factor mitigation. 2013.

[116] Zifeng Wu, Yongzhen Huang, and Liang Wang. Learning representative deep features for image set analysis. *IEEE Transactions on Multimedia*, 17(11):1960–1968, 2015.

[117] Xianglei Xing, Kejun Wang, Tao Yan, and Zhuowen Lv. Complete canonical correlation analysis with application to multi-view gait recognition. *Pattern Recognition*, 50:107–117, 2016.

[118] Mihalis Yannakakis. The analysis of local search problems and their heuristics. *STACS 90*, pages 298–311, 1990.

[119] Tze-Wei Yeoh, Wooi-Haw Tan, Hu Ng, Hau-Lee Tong, and Chee-Pun Ooi. Improved gait recognition with automatic body joint identification. In *International Visual Informatics Conference*, pages 245–256. Springer, 2011.

[120] Tze Wei Yeoh, Saúl Zapotecas-Martínez, Youhei Akimoto, Hernán E Aguirre, and Kiyoshi Tanaka. Feature selection in gait classification using geometric pso assisted by svm. In *International Conference on Computer Analysis of Images and Patterns*, pages 566–578. Springer, 2015.

[121] TzeWei Yeoh, Hernán E Aguirre, and Kiyoshi Tanaka. Clothing-invariant gait recognition using convolutional neural network. In *Intelligent Signal Processing and Communication Systems (ISPACS), 2016 International Symposium on*, pages 1–5. IEEE, 2016.

[122] TzeWei YEOH, Youhei AKIMOTO, and Hernan AGUIRRE. A gait-based human identification method under various covariate factors (special issue on computer vision and applications). *IIEEJ transactions on image electronics and visual computing*, 3(2):193–205, 2015.

[123] TzeWei Yeoh, Saúl Zapotecas-Martínez, Youhei Akimoto, Hernan Aguirre, and Kiyoshi Tanaka. Genetic algorithm assisted by a svm for feature selection in gait classification. In *Intelligent Signal Processing and Communication Systems (ISPACS), 2014 International Symposium on*, pages 191–195. IEEE, 2014.

[124] Jang-Hee Yoo and Mark S Nixon. Automated markerless analysis of human gait motion for recognition and classification. *Etri Journal*, 33(2):259–266, 2011.

[125] Jang-Hee Yoo, Mark S Nixon, and Chris J Harris. Extracting human gait signatures by body segment properties. In *Image Analysis and Interpretation, 2002. Proceedings. Fifth IEEE Southwest Symposium on*, pages 35–39. IEEE, 2002.

[126] Shiqi Yu, Daoliang Tan, and Tieniu Tan. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 4, pages 441–444. IEEE, 2006.

[127] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer, 2014.

[128] Xiang Zhang, Junbo Zhao, and Yann LeCun. Character-level convolutional networks for text classification. In *Advances in neural information processing systems*, pages 649–657, 2015.

[129] Shuai Zheng, Junge Zhang, Kaiqi Huang, Ran He, and Tieniu Tan. Robust view transformation model for gait recognition. In *Image Processing (ICIP), 2011 18th IEEE International Conference on*, pages 2073–2076. IEEE, 2011.