Using Deep Learning for Gait Abnormality Classification

João Firmino dos Santos Loureiro

Thesis to obtain the Master of Science Degree in

Electrical and Computer Engineering

Supervisor(s): Prof. Paulo Luís Serras Lobato Correia

Examination Committee

Chairperson: José Eduardo Charters Ribeiro da Cunha Sanguino
Supervisor: Prof. Paulo Luís Serras Lobato Correia
Member of the Committee: Prof. João Pedro Oliveira

November 2019

Declaration

I declare that this document is an original work of my own authorship and that it fulfills all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

Acknowledgments

I want to start by thanking everyone that guided me through the development of this thesis. Professor Paulo Correia, for giving me the opportunity of working in this subject with him and for his consistent support and time dedicated to my guidance. Tanmay, for all the previous work made available for me, that surely helped in the starting point of this work, and for the all the advices whenever i needed them. And Sérgio, with whom I shared feedback, ideas and company during our days in the lab. I would also like to thank Instituto de Telecomunicações, with whom i developed this work.

Next, I would like to thank my parents, my sister and the rest of my family, who made this five year journey possible and supported me through all my decisions and difficulties along the way.

Finally, I want to thank my group of friends João, Marco, Mariana, Pedro, Sara and Vitor, with whom I shared these amazing five years and am glad to have had by my side since the beginning.

Abstract

Gait has been the target of many recent research works. Many pathologies affect the normal gait patterns and the analysis of gait indicators can be a useful way of assessing different gait disorders and their evolution, while also providing valuable information for rehabilitation. Current gait assessment is mostly either based on subjective analysis, that lacks precision and quantitative measurements, or done in dedicated laboratories with complex and expensive equipment setups, which is impractical. However, recent studies have been made with the objective of obtaining precise gait analysis and classification using simpler systems, composed of a variety of sensors. This work first reviews the different techniques developed over the last decades to gain insight on the subject, and then presents a method for classification of abnormal gait types from videos of gait sequences. The proposed system is based on the use of gait representation images, like the Gait Energy Image (GEI) and a novel representation we call Skeleton Gait Energy Image (SGEI), that uses information of the skeleton, obtained with a pose machine. These representations are used as input to a VGG-19 convolutional neural network, that is used to perform classification or to extract a feature vector that is then classified using machine learning methods such as Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA). The system is trained and tested in a newly acquired gait dataset, the GAIT-IST dataset, that simulates four different types of abnormal gait and is also tested with another dataset, to assess its generalization ability.

Keywords: Gait analysis, Gait disorders, Gait recognition, Abnormal gait, Gait classification, Deep learning

Resumo

A marcha tem sido alvo de vários trabalhos de investigação. Muitas patologias afetam os padrões normais da marcha e a sua análise pode ser útil para avaliar os diferentes distúrbios, a sua evolução e fornecer informação relevante para reabilitação. Atualmente, a análise de marcha é baseada numa análise subjectiva, que tem falta de precisão e medidas quantitativas, ou é feita em laboratórios com equipamentos complexos, dispendiosos e impráticos. Contudo têm sido feitos estudos com o objetivo de realizar análise e classificação de marcha precisas, usando sistemas simples, compostos por uma variedade de sensores. Este trabalho começa por rever as técnicas desenvolvidas nas ultimas décadas e depois apresenta um método para realizar classificação de tipos de marcha patológica a partir de sequências de vídeo. O sistema proposto baseia-se no uso de imagens de representação da marcha como a *Gait Energy Image* (GEI) e uma nova representação chamada *Skeleton Gait Energy Image* (SGEI), que usa informação do esqueleto, obtida com uma *pose machine*. Estas imagens são usadas como entrada numa rede neuronal convolucional VGG-19, que é usada para realizar classificação ou extrair um vector de características que é depois classificado, usando métodos de aprendizagem automática como *Support Vector Machines* (SVM) e *Linear Discriminant Analysis* (LDA). O sistema é treinado e testado numa nova base de dados de marcha, que simula quatro tipos de marcha patológica e é também testado com outra base de dados, para avaliar a sua capacidade de generalização.

Palavras-chave: Análise de marcha, Indicadores de marcha, Distúrbios de marcha, Reconhecimento de marcha, Marcha patológica, Classificação de marcha, Aprendizagem profunda, Transferência de aprendizagem

Contents

Acknowledgments v
Abstract vii
Resumo ix
List of Tables xiii
List of Figures xv
List of Acronyms xix

1 Introduction **1**
1.1 Motivation 1
1.2 Problem definition 1
1.3 Objectives 2
1.4 Contributions 2
1.5 Outline 3

2 State of the Art **5**
2.1 Wearable Sensors 5
2.1.1 Electrogoniometer 5
2.1.2 Textile sensors 6
2.1.3 Ultrasonic Sensors 8
2.1.4 Electromagnetic Tracking System (ETS) 9
2.1.5 Inertial Sensors 10
2.1.6 Force and Pressure Sensors 13
2.1.7 Electromyography (EMG) 14
2.2 Non Wearable Sensors 15
2.2.1 Floor Sensors 15
2.2.2 Vision-based Methods 15
2.3 Commercial Systems and Gold Standards 22
2.4 Deep Learning Vision Methods 23
2.4.1 Image Pre-processing 23
2.4.2 Convolutional Neural Networks 26
2.4.3 Optical Flow in CNN 27

2.4.4 3D Convolutional Neural Networks . . . 28
2.4.5 Recurrent Neural Networks (RNN) . . . 30
2.4.6 Generative Adversarial Networks . . . 31
2.4.7 Classification Methods . . . 33
2.5 State of the Art Summary . . . 36

3 GAIT-IST Dataset Acquisition **39**
3.1 Existing Gait Datasets . . . 39
3.2 Description of Simulated Types of Gait . . . 39
3.3 Dataset Characterization . . . 41

4 Proposed Gait Analysis System **43**
4.1 Image Pre-processing . . . 43
4.1.1 Background Subtraction . . . 44
4.1.2 Gait Energy Image (GEI) Computation . . . 46
4.1.3 Pose Extraction . . . 48
4.1.4 Skeleton Gait Energy Image (SGEI) Computation . . . 49
4.2 CNN Feature Extraction . . . 50
4.3 Classification . . . 52

5 Results **53**
5.1 CNN Hyper-parameter Tuning . . . 53
5.2 CNN Fine-tuning and Classification . . . 56
5.3 CNN Feature Extraction and Classification . . . 57
5.4 Cross-database Scenario . . . 58
5.5 GEI and SGEI Result Fusion . . . 60
5.6 CNN Deep Feature Visualisation . . . 60
5.6.1 Saliency Maps . . . 61
5.6.2 Class Activation Maps . . . 61
5.7 Results Discussion . . . 64

6 Conclusions and Future Work **65**
6.1 Achievements . . . 65
6.2 Future Work . . . 65

Bibliography **67**

List of Tables

5.1 Confusion matrix with the mean validation accuracy of the CNN GEI classification across all cross-validation folds, for the configuration where the VGG-19 is trained starting at the third convolution block. The overall accuracy was **94.5%**. 57
5.2 Confusion matrix with the mean validation accuracy of the CNN SGEI classification across all cross-validation folds, for the configuration where the VGG-19 is trained starting at the second convolution block. The overall accuracy was **97.4%**. 57
5.3 Confusion matrix with the mean validation accuracy of the LDA classification using GEIs across all cross-validation folds, for the configuration where the VGG-19, used for feature extraction, is trained starting at the third convolution block. The overall accuracy was **89.6%**. 58
5.4 Confusion matrix with the mean validation accuracy of the LDA classification using SGEIs across all cross-validation folds, for the configuration where the VGG-19, used for feature extraction, is trained starting at the second convolution block. The overall accuracy was **96.4%**. 58
5.5 Confusion matrix depicting the mean validation accuracy of the CNN classification across all cross-validation folds, when testing with the DAI2 dataset. The overall accuracy was **43.3%**. 59
5.6 Confusion matrix depicting the mean validation accuracy of the LDA classification across all cross-validation folds, when testing with the DAI2 dataset. The overall accuracy was **76.7%**. 60
5.7 Confusion matrix depicting the mean validation accuracy of the CNN classification across all cross-validation folds, when fusing the results of the GEIs and SGEIs. The overall accuracy was **98.5%**. 60
5.8 Summary of the overall accuracy results using the GAIT-IST dataset with the methods performed considering various fine-tuning configurations of the VGG-19 and GEIs or SGEI as the system input and results of the cross-database experiments. 64

List of Figures

2.1 Knee and hip angle measuring [18]. 6
2.2 Sock Sensor voltage output and gait event detection [21]. 7
2.3 Horizontal foot displacement and autocorrelation plots of normal gait [26]. 8
2.4 Horizontal and Vertical autocorrelation plots of abnormal gait [26]. 9
2.5 Detection of gait events based on foot angular velocity [35]. 10
2.6 Foot angle of healthy and cerebral thrombosis patient [35]. 11
2.7 Shank angular velocity with identification of the gait events. IC: initial contact or HS; FO: foot off or TO. Peaks corresponde to MS [36]. 12
2.8 Operating principle of the optical sensor [59]. 14
2.9 External heel and toe estimation [84]. 16
2.10 Internal heel and toe estimation [84]. 17
2.11 Different binary silhouettes and corresponding Gait Energy Image (GEI) [84]. 17
2.12 Plot and silhouettes representing foot flat overlap ratio [86]. 18
2.13 Joint detection based on frontal and sagittal silhouettes and their vertical and horizontal segmented pixel projections [94]. 19
2.14 Comparison between the knee angle obtained with vision system and measurements from a goniometer [94]. 20
2.15 COG computation on head and torso [84]. 20
2.16 COG computation on thighs and foot location [84]. 21
2.17 Result of the approximated skeleton computation [84]. 21
2.18 Representation of a Qualisys gait lab setup composed of an array of cameras and floor force plates [22]. 22
2.19 Representation of the BTS GAITLAB that includes the use of different optic sensors, wearable inertial sensors, floor force sensors and wireless EMG [40]. 23
2.20 General block diagram of the proposed gait analysis system. 23
2.21 Typical challenges of Background Subtraction in outside scenarios [107]. 24
2.22 Different binary silhouettes and corresponding Gait Energy Image (GEI) [84]. 25
2.23 Optical flow. (a) and (b) represent a pair of consecutive frames. (c) is a close-up dense optical flow of the highlighted area. (d) and (e) are the horizontal and vertical displacement vector fields respectively. [109]. 25

2.24 Convergence of the estimation for the right elbow position in a convolutional pose machine [110]. . . . 26
2.25 Warping neighbouring heatmaps to correct pose estimate errors [116]. . . . 26
2.26 Feature extraction (VGG-19) architecture [119]. . . . 27
2.27 DeepGait architecture [121]. . . . 27
2.28 Two-stream architecture for video classification [109]. . . . 28
2.29 3D CNN architecture for human action recognition [124]. . . . 29
2.30 Multiresolution CNN architecture [125]. . . . 29
2.31 3D Convolutional network and LSTM architecture for human action recognition [126]. . . . 30
2.32 Dual channel LSTM architecture [131]. . . . 31
2.33 Pose heatmap LSTM architecture [134]. . . . 31
2.34 Pose-based gait recognition workflow [135]. . . . 31
2.35 Pixel level domain transfer architecture [136]. . . . 32
2.36 TS-GAN architecture [138]. . . . 33

3.1 Depiction of the four pathological gaits simulated in the GAIT-IST dataset. . . . 41

4.1 Detailed diagram of the image pre-processing implemented, corresponding to the pre-processing block in Figure 2.20. . . . 44
4.2 Background Subtraction results using the MOG2 algorithm. . . . 45
4.3 Application of median blur and OPEN and CLOSE morphological filters. . . . 45
4.4 Application of method to fill in foreground objects [150]. . . . 45
4.5 Application of method to eliminate noise blobs detected as foreground objects. . . . 46
4.6 Computation of the bounding box of the silhouette. . . . 46
4.7 Final resized silhouette image with feet bounding box (left) and the plot of its width along a full gait sequence (right). The inconsistency of the middle peak represents the change of walking direction in the gait sequence. . . . 47
4.8 Down-sampled set of silhouette frames corresponding to one complete gait cycle (left) and respective GEI (right). . . . 48
4.9 Representation of the pose data extracted with the OpenPose algorithm [114]. . . . 49
4.10 Example of the result of a skeleton extraction. Original frame with rendered pose on the left and the resulting skeleton on the right. . . . 49
4.11 Down-sampled set of skeleton frames corresponding to one complete gait cycle (left) and respective SGEI (right). . . . 50
4.12 Detailed diagram of the feature extraction step implemented, corresponding to the feature extraction block in Figure 2.20. . . . 51
4.13 Architecture of the VGG-19 network, with the third fully connected layer modified. The first FC-4096 corresponds to the layer whose output is used as a feature vector. . . . 51
4.14 Detailed diagram of the classification step implemented, corresponding to the classification block in Figure 2.20. . . . 52

5.1 Training (blue) and validation (orange) accuracies (left) and losses (right) during training of one fold of cross-validation for learning rate = 0.001. 54
5.2 Training (blue) and validation (orange) accuracies (left) and losses (right) during training of one fold of cross-validation for learning rate = 0.001 and using Nesterov Momentum. . 55
5.3 Training (blue) and validation (orange) accuracies (left) and losses (right) during training of one fold of cross-validation with learning rate = 0.0002 and using Nesterov Momentum. 56
5.4 Results of the application of the Keras Visualization Toolkit on GEI examples belonging to each class of the GAIT-IST dataset, resulting in Saliency Maps and Class Activation Maps. 62
5.5 Results of the application of the Keras Visualization Toolkit on SGEI examples belonging to each class of the GAIT-IST dataset, resulting in Saliency Maps and Class Activation Maps. 63

List of Acronyms

CNN Convolutional Neural Network.

COG Center Of Gravity.

EMG Electromyography.

ETS Electromagnetic Tracking System.

GEI Gait Energy Image.

GRF Ground Reaction Forces.

HS Heel Strike.

IMU Inertial Measurement Unit.

LDA Linear Discriminant Analysis.

LSTM Long Short-Term Memory.

MS Mid-Swing.

PCA Principal Component Analysis.

SGEI Skeleton Gait Energy Image.

SVM Support Vector Machine.

TO Toe Off.

Chapter 1

Introduction

1.1 Motivation

The human gait has been the topic of many research projects in the field of medicine. As gait results of the coordination between the nervous and the musculoskeletal systems, several medical doctors and researchers have recognized a relationship between the way humans walk and disease symptoms [1], such as dementia [2–4], Parkinson's disease [5–7], multiple sclerosis [8–10], strokes [11, 12] and aging related impairments [13–15]. As such, reliable and accurate ways of assessing gait and its indicators, and monitoring them over time can help specialists with early and precise diagnosis of the mentioned disorders.

1.2 Problem definition

Gait can be defined as a set of movements of the limbs that, in a cyclic manner, allows for the locomotion of humans [16]. As mentioned before, certain neurodegenerative disorders, injuries and aging can affect this set of movements leading to what can be called an impaired or abnormal gait and, although each individual has its own unique gait, some gait indicators, such as step or stride length and joint angles, can be analyzed in order to distinguish normal from abnormal gait and even discriminate which disorder is present.

The most common practice of gait assessment consists of a specialist simply observing the walking cycle of the patient and looking at indicators such as gait speed and step or stride length. This approach lacks the precision and information that can be obtained with instrumented methods that provide quantitative and objective measures and can have a negative impact on the diagnosis and posterior treatment of the conditions.

On the other side of the spectrum, there are solutions from specialized companies that use a wide variety of, mainly vision based, sensors in laboratories and sophisticated equipment that can acquire motion data with high accuracy. However, this uses high cost technology in specialized environments, which is not always available in clinical conditions and often require equipment setup and calibration

procedures that can be complicated and time-consuming.

To tackle the problems of these two methods, there has been a rise of new technologies and methods aimed at solving the problem of obtaining objective measures of a varied selection of gait indicators and classifying gait between normal and abnormal and even between the different possible pathologies. These methods use a variety of sensors, that can be wearable or not, to measure the different gait indicators. Wearable sensors are attached to parts of the body in order to measure gait kinetics and kinematics. These methods, while being precise and not restricting their usage to a controlled environment, can be considered intrusive. This means that they may cause discomfort to the patients or even affect their gait, producing inaccurate measurements. The non wearable sensors include floor sensors, whose main limitation is their length, which restricts how many steps can be taken and measured, and vision based sensors. The last ones are used in the specialized laboratories mentioned before, but, lately, there has also been an increased number of techniques using only a couple, or even just one camera. For instance, less costly devices, such as the Kinect and basic RGB cameras are capable of obtaining precise results on gait indicators and perform an accurate classification of abnormal gait. The main constraint on the non wearable sensors is the fact that they cannot be used permanently during the day-to-day activities of the patients and the analysis of gait must rely on a small sample to draw results.

In order to test all these methods it is necessary to have a representative database of walking cycles both of normal and abnormal gait. This can be a set back due to the difficulty of having samples of gait from real patients, that suffer from the pathologies to be detected. Because of this, most studies have voluntaries simulating the different types of gait. However, this solution can be more time consuming as it can take some attempts to reproduce the types of gait with a desirable degree of accuracy and, even then, the gait samples are not a perfect simulation of a real patient.

1.3 Objectives

This work will review the state of the art concerning the different techniques used to analyze human gait, retrieve gait indicators and classify between different disorders and normal gait. After assessing what are the best methods, considering both the accuracy and the simplicity of the approach, the objective is to identify and develop a simple and non-intrusive solution for classification between normal and different types of abnormal gait. More specifically, the method to be developed will be able to process inputs in the form of videos to obtain usable images in a deep learning architecture that will then extract features of the images, in order to perform classification over the feature vectors.

1.4 Contributions

Besides presenting a review of the state-of-the-art regarding the different methods used for gait analysis and classification and the implementation of a system that is able to perform classification between normal gait and four different types of abnormal gait using the widely used Gait Energy Image (GEI) representation, there are three main contributions of this work:

- The acquisition of a new gait database, where 10 subjects simulate the four types of gait with different types of severity, which is important given the lack of datasets regarding pathological gait.
- The usage of pose data obtained from a pose machine algorithm to compute a new gait representation we call the Skeleton Gait Energy Image (SGEI), which is not affected by the body shape or clothes being used, in contrast with silhouette based representations, and which outperforms the usage of GEIs on the same system.
- An analysis of the results, attempting to validate the features automatically extracted by the algorithm, also presenting and analysing results from a deep feature visualisation toolbox.

1.5 Outline

This report is organized as follows. In Chapter **2**, the different techniques for gait analysis and classification will be analyzed and discussed. Chapter **4** presents the proposed solution for a system that is able to classify between normal and different types of abnormal gait. In Chapter **5**, the results of the proposed system are presented and discussed. Final conclusions are drawn in Chapter **6** and future work possibilities are presented.

Chapter 2

State of the Art

In this chapter, the state-of-the-art concerning the techniques for gait assessment are is reviewed. The methods analyzed are divided in wearable sensors and non wearable sensors, where the vision-based methods are included in the latter. In the end of the chapter the commercialized systems developed by companies for high accuracy gait analysis are discussed.

2.1 Wearable Sensors

A great number of gait analysis methods based on wearable sensors have been studied and proposed as an inexpensive and less complex alternative to specialized laboratories, as their usage is not restricted to a controlled environment. These methods consist on the placement of different types of measuring devices on the feet, knees and waist to obtain indicators relating to three distinct categories: kinematics, kinetics and electromyography (EMG).

The kinematic gait analysis focuses on the movements of the lower limbs and joints and includes the usage of electrogoniometers, textile sensors, ultrasonic sensors, electromagnetic tracking systems (ETS) and inertial sensors, such as accelerometers, gyroscopes and magnetic sensors.

Kinetics study the moments and forces involved in producing movement in the human body. The main focus of this area are the forces related to the contact of the feet with the ground. Ground reaction forces (GRF) can be measured with wearable force and pressure sensors and, combined with kinematic measures of the leg segments, can be used to obtain kinetic information on the joints of the lower limbs.

An EMG shows the small electric activity that results from muscle contractions during locomotion. It can be obtained using surface electrodes or, more intrusively, with wire or needle electrodes.

In this section, the methods regarding the wearable sensors mentioned above will be reviewed.

2.1.1 Electrogoniometer

A goniometer is an instrument that can be used to measure angles between the lower limb segments, like the knees, ankles and hips. The most common type is the strain gauge-based flexible electrogoniometer. This sensor has a resistance that increases when it is flexed, proportionally to the angle applied.

Rowe et al. [17] demonstrated the validity of measuring knee kinematics with an electrogoniometer, by obtaining measurements of the knee angles at different walking speeds using electrogoniometers at the same time as using a Vicon VX motion analysis system. When comparing the mean values of the knee joint angles for the different walking sequences, the observed errors were of only 2^o (corresponding to 1,5% of the measuring range) when comparing the results of the two systems.

Maranesi et al. [18] presented a simple method for assessing gait parameters based on knee and hip joint angle measurements using electrogoniometers. Using two goniometers they measured the hip and knee angles in the sagittal plane, on the same lower limb as represented in Figure 2.1. The β angle is taken at its maximum, considered to be the moment of heel strike and the α angle is taken at the toe off event, which happens when the thigh is in the vertical position, i.e., when β has value zero. Knowing r_1 as the shank length (distance from the knee to the ankle), and r_2 as the leg length, the distances d_1 and d_2 can be calculated and the sum of the two is taken as the step length. The step duration is estimated as half the time between two consecutive heel strikes, when β is at a maximum value. Step speed is calculated as the ratio between step length and duration.

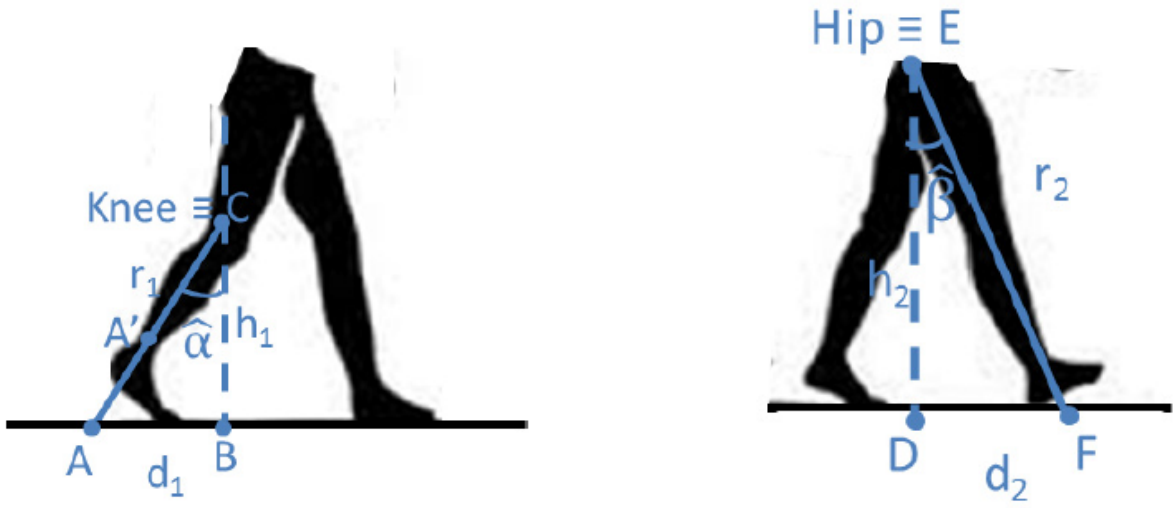


Figure 2.1: Knee and hip angle measuring [18].

There are other variations, such as the mechanical and inductive goniometers. New types of goniometers have also been developed, such as one based on encoders that Domínguez et al. [19] used to measure knee joint position, obtaining a correlation of 0,9943 compared with the measures from a commercial motion analysis system, and an optical fiber goniometer, Donno et al. [20], applied to human joint movement monitoring and obtaining an average RMS error in the range of 1% to 1,5%.

2.1.2 Textile sensors

This technology enables the incorporation of electronic and sensing capabilities into fabric, that can be molded into a flexible wearable piece, being less intrusive than other types of sensors attached to the body. This type of sensor can be obtained by coating regular fabrics with inherently conductive polymers or by weaving electrically conductive yarn into the fabric. The most common goal is to produce a fabric

with piezoresistive properties in order to be used, similarly to a goniometer, as a strain gauge sensor.

Preece et al. [21] developed a method to estimate gait events from the voltage data of a knitted resistive strain sensor in the form of a sock. The estimations were made based on the knowledge that, during gait, the ankle motion is made of periods of rapid movement followed by periods of slow movement. The three gait events detected from the sock voltage, represented in figure 2.2, were heel lift (HL), toe off (TO) and heel strike (HS), and their calculation was based on the estimation of five parameters (a_1 - a_5).

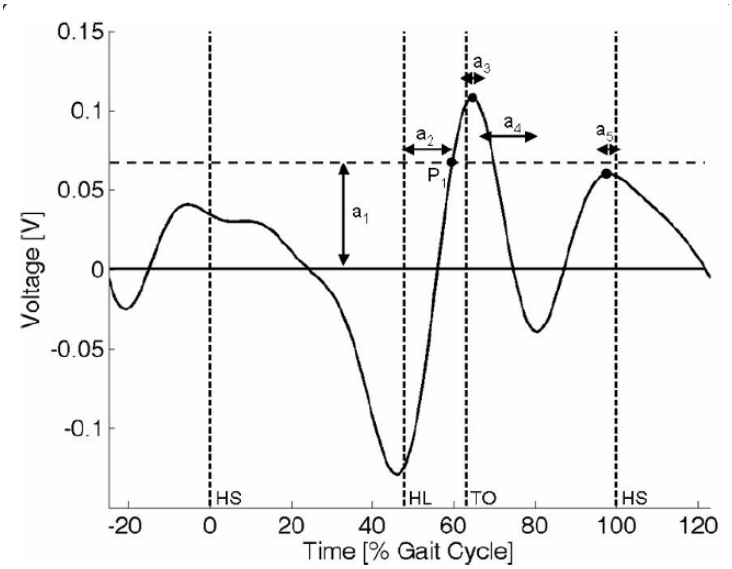


Figure 2.2: Sock Sensor voltage output and gait event detection [21].

These parameters are obtained from five of the ten walking sequences of each participant through a search algorithm. The first parameter (a_1) is a threshold and the other ones are distances in time between the point where this threshold is met (P_1) or maximum values of the signal and the gait event times, that are obtained in the remaining five walking trials using the estimated parameters. Using reflective markers and a ten-camera Qualisys Pro Reflex system [22], kinematic data was obtained and compared to the sock sensor output and, although the two signals didn't always match, the algorithm was able to predict the three gait events with mean errors of 1-1.6% of gait cycle for HL and TO and 2.6-3.3% for heel strike, showing significant accuracy.

Vu and Kim [23] developed a method using muscle pants with similar properties in order to identify types of movement between walking, running, sprinting and jumping. This was done by extracting three parameters from the sensor output signal and processing them using machine learning algorithms in order to make the classification. The parameters used were the average amplitude, its standard deviation to the mean value and the average cycle duration. The method achieved a classification accuracy of 90% when using a random forest algorithm.

Other works using this type of technology include the ones by Tognetti et al. [24] and Lorussi et al. [25] that propose methods for identification and reconstruction of body segments position, focusing on the upper limb, using kinaesthetic garments with piezoresistive properties.

2.1.3 Ultrasonic Sensors

Ultrasonic sensors are devices that are able to send or detect ultrasound waves, which are sound waves at high frequencies, above the human audible spectrum. This type of sensor is most commonly used to detect distances between two points, either by having a transceiver that sends and receives the reflected signal from the object being detected or by having a transmitter sending a signal and a receiver detecting it, where one is typically static (anchor) and the other is mobile. Knowing the speed at which the signal travels and measuring the time it takes to be received, Time-of-Arrival (TOA), the distance can be calculated. If multiple anchors are present it is also possible to calculate the position of the mobile body based on the Time-Difference-of-Arrival (TDOA). In gait analysis this can be used to measure step or stride length, distance between feet and even the distance between the foot and the ground.

Qi et al. [26] proposed a method, based on this technology, to track the 3D displacement of walking subjects using a transmitter attached to the foot and four receivers. In order to assess gait regularity and symmetry, the autocorrelation coefficients of the horizontal (direction of progression) and vertical foot displacement was plotted as a function of the phase shift in the number of observations, which will show peak values when the phase shift used corresponds to the gait cycle duration. Then, it can be assumed that the gait symmetry is related to the regularity of the autocorrelation plot. Figures 2.3 and 2.4 show these plots for normal and abnormal gait.

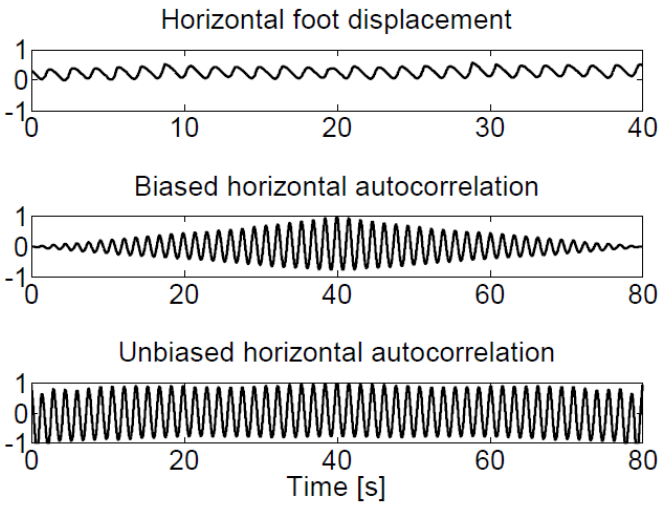


Figure 2.3: Horizontal foot displacement and autocorrelation plots of normal gait [26].

In this work gait indicators, such as stride length and duration, velocity and cadence were also estimated. Stride length was calculated as two times the difference between maximum and minimum horizontal foot displacement and stride duration as the time difference between those two instances. Velocity and cadence, which is the number of strides in a second, were derived from the previous parameters. These estimations were obtained with and average Root Mean Square Error (RMSE) of 2.7% when validated against the Motion Analysis Eagle System [27] using eight high speed cameras.

Wahab and Bakar [28, 29] considered the mid-swing (MS) phase of the foot, in which it is closer to the ground, also know as Minimum Foot Clearance (MFC), a very important moment when assessing fall

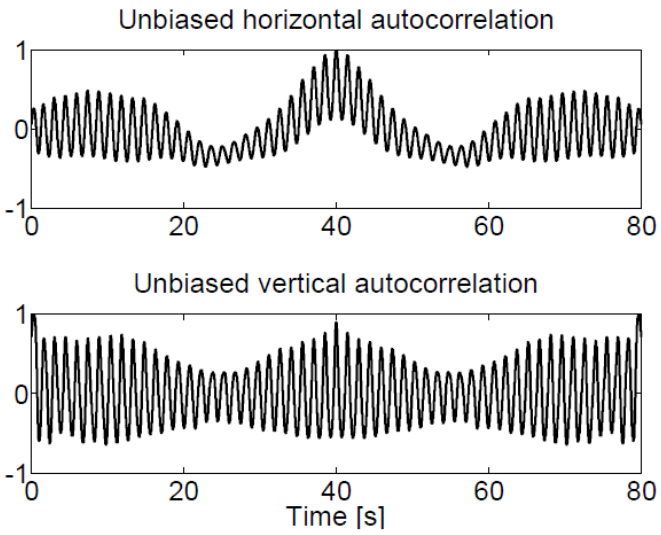


Figure 2.4: Horizontal and Vertical autocorrelation plots of abnormal gait [26].

.

or trip risk. In their works, they proposed a method to measure foot distance to the ground using an ultrasonic transceiver in order to measure MFC.

2.1.4 Electromagnetic Tracking System (ETS)

The ETS uses Faraday's law of magnetic induction to detect the 3D position and orientation of body segments [30]. It works by having a fixed device that generates a magnetic field and some sensors composed of three orthogonal coils attached to parts of the body that will have varying induced voltages depending on their position and orientation relative to the controlled magnetic field. The main restrictions of this method are the limited tracking space, defined by the range between the magnetic field generator and the sensor coils, and eventual interference from metallic objects near the tracking system [31].

Kobayashi et al. [32] used an ETS composed of a fixed transmitter on a wall and four sensors that measured the position and orientation of the pelvis, thigh, leg and foot. After compensating for environmental influences and residual errors, the joint angles for the hip, knee and ankle were calculated in the sagittal, frontal (coronal) and transverse planes from the absolute orientations measured. The results showed estimation errors of less than 1 degree. However their method was not compatible with the widely used joint coordinate system (JCS), proposed by Grood and Suntay [33], as reported by Mills et al. [34], who developed a ETS technique consistent with the JCS and assessed its repeatability of gait kinematic measurements during treadmill locomotion. For this they had participants undergo two gait assessment trials performed by the same two testers each day, for two days. The same angles as before were estimated, in the same planes and, overall, the repeatability of the kinematic measurements were equal or superior to results reported in previous literature, despite having a slightly lower score in the frontal and transverse plane repeatability when compared to the sagittal plane.

2.1.5 Inertial Sensors

Inertial sensors, namely accelerometers, gyroscopes and some magnetic sensors, are devices that can be used to measure the linear or angular velocity and acceleration and the orientation of objects and are one of the most used type of sensors for gait analysis.

An accelerometer's operation is based on the Newton's Laws of Motion, specifically on the knowledge that the sum of the forces applied to an object is proportional to its acceleration. So by having a reference mass connected to a suspension system, acceleration can be calculated by measuring the displacement of the mass when external acceleration is applied.

Gyroscopes are used to measure an objects orientation and angular velocity. Although there are many variations of gyroscopes, based on different principles, such as MEMS, fibre optic and quantum gyroscopes, the basic functioning principle of the gyroscopes is that a disc spinning around an axis will have a rotational inertia. This means that it will resist any change to its orientation. So when mounted in a frame that allows the disc's rotation axis to have any orientation, tilting or rotating the mounting will not affect the orientation of the spinning axis, which is used as a reference to determine changes in the mounting frame's orientation.

Magnetic sensors, or magnetometers, are sometimes also used in gait analysis mainly for detecting the orientation of body segments. These sensors can measure the strength, direction or change of a magnetic field around them and thus determine their orientation relative to the magnetic north.

One commonly used device that aggregates all these sensors is called Inertial Measurement Unit (IMU). This sensor is typically seen in air and spacecrafts, incorporated into navigation systems but has lately been used in motion capture technology and gait analysis. For example, Zhao et al. [35] developed a gait analysis system using a IMU in each foot in order to assess the rehabilitation of patients with gait disorders. The experiment was performed both on healthy subjects and patients with diagnosed cerebral thrombosis, which leads to weakness on only one side of the body, for result comparison. The method consisted, firstly, in estimating gait events based on the foot angular velocity, along the medio-lateral axis, perpendicular to the gait direction and parallel to the ground, of each foot, using peak detection, zero-crossing detection and flat-zone detection as depicted in Figure 2.5. The events detected were heel-strike (HS), foot-flat (FF), heel-off (HO) and toe-off (TO).

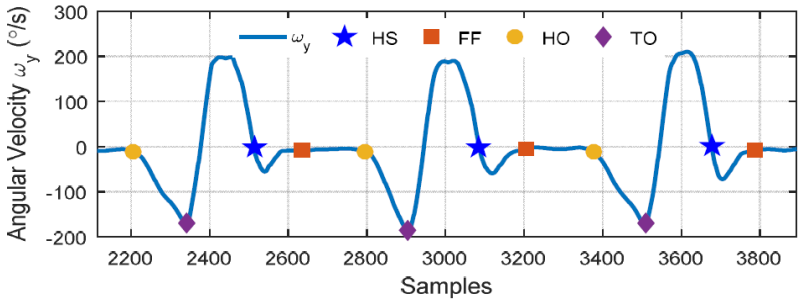


Figure 2.5: Detection of gait events based on foot angular velocity [35].

Based on these events, the duration of the phases between them, labeled heel-strike, stance (St), heel-off (HO) and swing (Sw), were calculated. These indicators were used to assess gait symmetry and

were compared between the healthy subjects and the patients. The results showed a greater asymmetry in the patients characterized by a lower duration of St and HS on the weak foot and a higher duration of Sw. These results were expected from the nature of the pathology, however further analysis was made to the reduced HS phase which was due to a weaker extension-flexion of the affected foot. This was corroborated by the measurement of the foot angle during gait as seen in Figure 2.6, that shows a large difference in both the plantar flexion and dorsiflexion, between the healthy subject and the cerebral thrombosis patient.

Furthermore, one patient was tested with the system once per week during a month in order to track his evolution. The results showed and improvement, overtime, of the gait symmetry, based on the duration of gait phases in each foot, and also and increase in the amplitude of the extension-flexion of the foot, showing a good recovering of the patient and demonstrated the capability of the system to assess rehabilitation of patients with this kind of gait disorder.

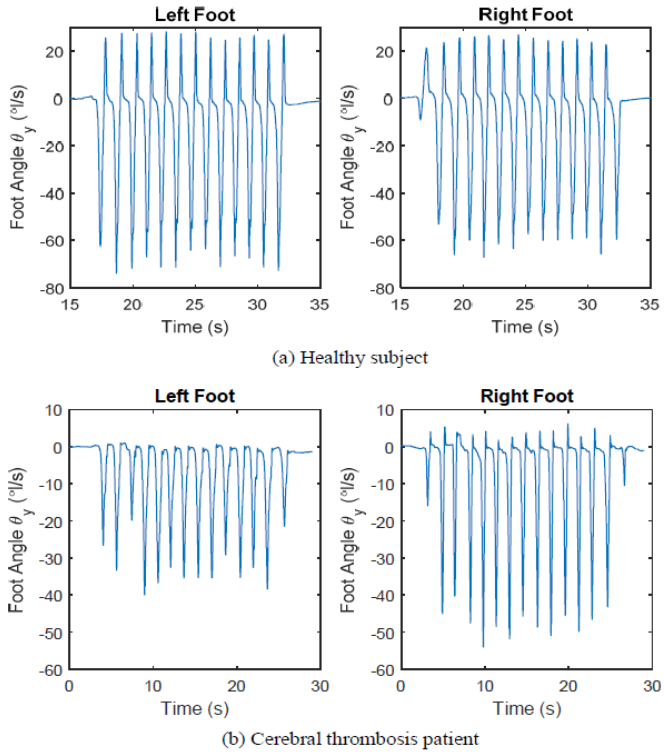


Figure 2.6: Foot angle of healthy and cerebral thrombosis patient [35].

Ferrari et al. [36] presented a method to detect gait events on subjects with movement disorders, using IMUs attached to the thighs and shanks of the subjects. This method was based on a study made by Lee and Park [37] who measured the medio-lateral axis angular velocity of the shank with gyroscopes and noticed that, during gait cycle, it showed a pattern of two negative peaks on each side of a positive peaks. These points were associated, in temporal order, to the events HS, MS and TO as ilustrated in Figure 2.7. Once these events were estimated, Ferrari et al. [36] used a double pendulum model (DMP) of the lower limbs to determine step lengths. Tests were made in healthy adults and children affected by Cerebral Palsy (GMFCS I) both with the proposed system and an optoeletronic system by NDI (Optotrak Certus) [38] for comparison, obtaining errors below 5% of the median step length estimations.

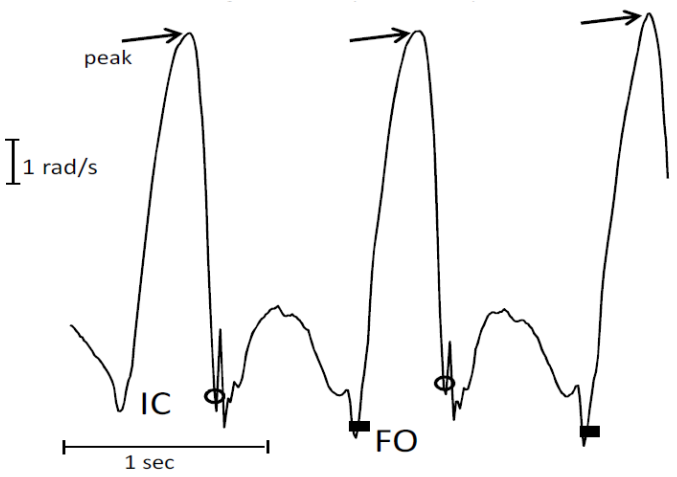


Figure 2.7: Shank angular velocity with identification of the gait events. IC: initial contact or HS; FO: foot off or TO. Peaks corresponde to MS [36].

A similar method had implemented by Salarian et al. [39] who estimated some more indicators on healthy subjects and Parkinson's Disease (PD) patients while a subthalamic nucleus deep brain stimulation (STN-DBS) was ON and OFF, verifying results closer to those of healthy subjects in the first case. The gait indicators estimated showed a high correlation with Unified Parkinson's Disease Rating Scale (UPDRS) subscores and an error of less that 8% when compared with a camera-based system by BTS [40]. Salarian et al. [41] also later studied an approach to predict the movements of thighs based on the movements of shanks, which reduces the number of wearable sensing units necessary for gait analysis.

Another identical method of detecting gait events, namely HS and TO, using shank-mounted gyroscopes was made by Catalfamo et al. [42] which also assessed incline walking and, compared with the pressure measurement insoles F-Scan by Tekscan [43], validated in [44], obtained errors below 25ms and 75ms for HS and TO respectively.

After suggesting a new method for measuring gait symmetry [45, 46], using symbol based symmetry indexes that could distinguish between healthy subjects and PD patients, Sant'Anna et al. [47] evaluated the usage of inertial sensors for gait symmetry and normality assessment. Gait normality assessment was done using the same type of information of the proposed gait symmetry method, comparing it to the average results of a set of healthy subjects, instead of comparing the information between right and left foot, as it's done in the symmetry assessment. The system was validated, with good results, in-lab against a 3D kinematic measures from the Qualisys MCU 240 motion capture system [22] and was also tested *in situ* on hip-replacement patients, correlating well with the clinical assessments performed.

Beravs et al. [48] developed and validated a system using IMUs for measuring hip, knee and ankle joint angles obtaining errors below 5 degrees when tested on a single healthy subject, validated against the Optotrak Certus (NDI) [38] system. More recently, together with Novak et al. [49], a system was developed, using inertial sensors and pressure sensor insoles, to detect gait initiation and termination in real time using a classification tree, a supervised machine learning method. Nine IMUs are used, in the foot, shank, thigh and arm of each side and on the back. The gait initiation detection consists, firstly, in the detection of movements that precede the lifting of the foot and the TO event itself. The set of

movements, named gait onset, is detected based on the hip and knee angles and angular velocity. The gait termination part segments the gait sequence into steps, using the pressure sensor insoles, and then analysis the first samples of each step, a buffer, in order to assess if it is the last one or not. It does this by analyzing the center of plantar pressure (COP) and ground reaction force (GRF), using the pressure sensors and the standard deviation of knee and hip angles from the inertial sensors. Tests made showed that gait initiation was predicted timely with a non-detection rate of less than 3% and gait termination was successfully detected, before the final foot contact with the ground, 80% of the time, when using a step analysis buffer shorter than the average step time.

Other studies using this type of sensor include the work by O'Donovan et al. [50], who used the same type of inertial sensors including magnetic sensors to measure joint angles, and Riva et al. [51], who developed a system to assess fall risk that does not require step detection. This was done by using an inertial sensor placed in the trunk of the subjects while walking in a treadmill, collecting kinematic trunk indicators and associating them to fall history in the subjects.

2.1.6 Force and Pressure Sensors

The study of gait kinetics, which focuses of the forces involved with the production of movement is most commonly done with pressure and force sensors. The main product produced by these type of sensors that is used in gait analysis are the so called ground reaction forces (GRF), that are the forces exerted by the ground on the feet. A distintion is made between pressure sensors, that measure only the vertical GRF and force sensors, that measure each compoment of the GRF on every axis. The most common use of these sensors is the integration in shoes [52] or insoles [53] of resistive [54, 55] or capacitive [56] sensors.

Some innovative studies involving these type of sensors include the work by Savelberg and de Lange [57], who managed to determine the horizontal, fore-aft component of GFR using an insole pressure sensor and artificial neural networks. Wang et al. [58] presented a way of measuring plantar pressure and shear stress based on the attenuation of intensity observed in an array of optical fibers due to its deformation. Lincoln et al. [59] developed an insole capable of measuring every GRF component based on the reflected light from a reflective moving part, that changes its position when pressure or shear stress is applied, detected by an optical sensor, as shown in Figure 2.8.

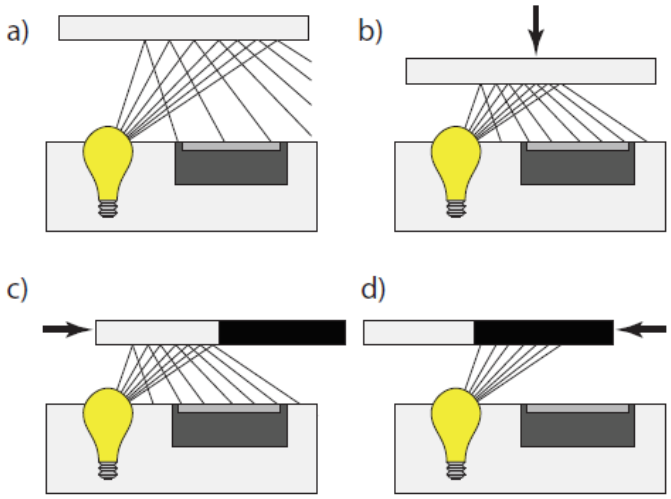


Figure 2.8: Operating principle of the optical sensor [59].

2.1.7 Electromyography (EMG)

An EMG is a medical technique for retrieving information on the activity of skeletal muscles. An electromyograph is able to detect the small electrical potential generated by muscle activity that can provide information on their contraction intensity and timing. Despite wire electrodes being able to produce information on the specific muscle they are inserted in [60], surface electrodes, that provide a more general insight in muscle activity, are enough for assessing the muscle activity in human gait [61], adding to the fact that they are less intrusive.

Bogey et al. [62] developed a technique to process EMG-to-Force and calculate ankle joint moments and muscle forces. The peak ankle tendon force estimated was closely matched with the values obtained with force transducers. Lloyd and Besier [63] presented an EMG-based musculoskeletal model of the human knee to preditc knee moments and muscle forces. Wentink et al. [64] used EMG to predict the onset, movements leading to gait initiation in transfemoral amputees. The study showes that EMG was able to detect initial movement up to 138 ms in advance when the prosthetic leg is leading.

Other works include Dietz et al. [65, 66] that studied leg muscle activation during gait in Parkinson's disease patients analysing adaptation and interlimb coordination during split-belt walking conditions and body unloading on a treadmill. A review of multichannel surface EMG in gait analysis is done by Frigo and Crenna [67].

2.2 Non Wearable Sensors

Non wearable sensors are mainly based on the use of floor force and pressure sensors, that get information of the feet pressure and force interactions with the ground, and the vision-based methods, that rely on information captured by an optic sensor to assess gait indicators from image processing techniques. These two methods are reviewed in the rest of this section.

2.2.1 Floor Sensors

Floor sensors are one of the most basic sensors for gait analysis including gait recognition. Like the wearable versions, they can be pressure sensors, that can measure the center of pressure and the pressure distribution under the foot, but can not measure every GRF component in 3D space, unlike the force sensors that are able to detect shear or horizontal forces. The main disadvantage of these sensors are their limited size, which restricts the number of steps they can measure at a time.

Leusmann et al. [68] developed a sensor floor composed of a dense network of piezoelectric sensors integrated in the house floor. The goal is to monitor the subject and detect abnormal gait patterns or falls that could indicate an emergency. Middleton et al. [69] developed a gait recognition system composed of an array of resistive sensors that extract gait indicators such as stride lenght and cadence and time on toe to time on heel ratio. The results showed 80% accuracy in gait recognition. Vera-Rodriguez et al. [70] analyzed a database of about 20,000 footsteps and more than 120 people and performed a comparative assessment and fusion of spatiotemporal informational for footstep recognition.

2.2.2 Vision-based Methods

A set of gait analysis and recognition approaches that have been largely used in the last years are the ones based on vision [71–76], in which gait data is captured with optic sensors and measurements of different gait indicators are obtained through image processing. These measurements can then be taken as features and used in machine learning techniques for classification or these methods can even be applied directly on the pre-processed images for automatic feature extraction.

Besides basic analog or digital cameras, other types of optic devices are used. For example, infrared sensors [77] provide a visual representation of the temperature of a body, based on its infrared thermal emissions. Time-of-Flight (ToF) cameras [78, 79] measure the distance of each captured pixel by the time it takes to receive a reflected light signal or by measuring the phase shift of a reflected modulated near infrared wave. ToF technology is used by the popular Microsoft Kinect, that is able to provide 3D information in the form of depth images. Another method used by the Kinect in previous versions was structured light [80], that is able to obtain 3D information of objects by capturing the projection of a calibrated light pattern on the target. The use of the Kinect for gait analysis was validated by Springer and Seligmann [81], who reviewed a number of papers describing methods for gait indicator measurements, such as gait events and kinetic and kinematic indicators, using this device. Another method for obtaining 3D information is through the use of multiple cameras, using stereoscopic vision [82, 83].

The main advantages of vision-based methods are the easy setup and data acquisition process and the absence of devices in contact with the subject, that could interfere with his normal gait cycle. Some setbacks are the occurrence of self-occlusions, where some body segments get in front of others, changes in the illumination and the background complexity, that can make it difficult to obtain some measurements.

Depending on whether the gait video features are computed with algorithms that make specified measurements on the image or are extracted automatically using more sophisticated deep learning techniques, we can divide video base methods into the ones that use handcrafted features and learned features, respectively. The methods that use handcrafted features can be further classified into model-based and shape-based. The rest of this section will focus on reviewing the state-of-the-art of the image processing approaches for gait indicator computing and classification tasks using handcrafted features. The methods involving deep learning vision-based methods will be reviewed in a separate section as those are the preferred approaches to be considered for the proposed solution of the Thesis.

Shape-based Methods

In the handcrafted features methods, shape-based approaches usually use image sequences obtained from a single camera and consist on the examination of the shape of the subject, most of the times in the form of a silhouette after background subtraction, and its changes over time to identify the gait indicators.

Nieto-Hidalgo et al. [84] proposed a method to classify normal and abnormal gait using a single RGB camera. In the first part of this work, after pre-processing the images to obtain a binary silhouette, shape-based techniques are applied to detect HS and TO events. The process starts by focusing on the lower 20% of the silhouette and defining a bounding box. Foot length is estimated as the width of the bounding box when its value is minimum during the gait sequence, corresponding to the moment when the feet are side by side. Then, external heel and toe of each frame are calculated as the minimum and maximum silhouette pixels in the horizontal axis (x) as shown in Figure 2.9. Next, using these points and the foot length, a bounding box for each foot is obtained and the internal heel and toe are estimated as the minimum and maximum values of x in each foot bounding box as seen in Figure 2.10.

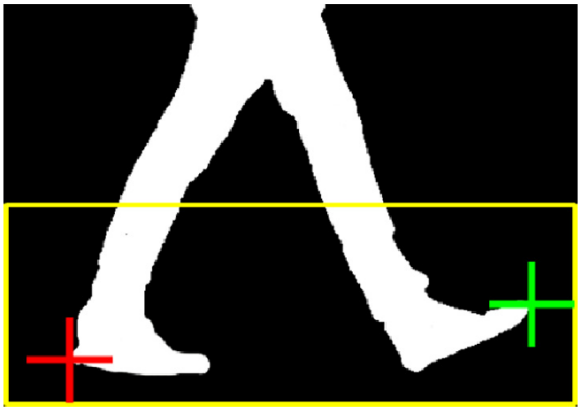


Figure 2.9: External heel and toe estimation [84].

Having the information on the horizontal position of foot heels and toes throughout the frames, a gradient analysis is done to determine HS and TO events, detecting moments where the gradient becomes

Figure 2.10: Internal heel and toe estimation [84].

or no longer is 0. After applying filters to handle false positives, the event detection gave results with RMSE lower than $100ms$ when compared with manual marking of the events.

Ortells et al. [85] propose a method to compute gait indicators from a single camera sequence of images, aimed at quantifying gait impairments, like gait asymmetry, based on the symmetry of extracted indicators, and estimating the falling risk. Three indicators are computed from the binary silhouettes, stance phase (St), swing phase (Sw) and step length (SL). These are obtained using a similar bounding box technique, where St occurs when the box stops growing, Sw when the box stars to decrease and SL corresponds to the width of the box at the start of St. For other indicators a representation known as the Gait Energy Image (GEI) is used. GEI is defined as the mean image of all the binary silhouette frames, during a complete gait cycle, aligned horizontally by their upper half centroid, as represented in Figure 2.11.

Figure 2.11: Different binary silhouettes and corresponding Gait Energy Image (GEI) [84].

From a GEI the indicators of Intensity, that represents the amount of movement and is based on the gray level of pixels and number of foreground pixels, and Amplitude, that represents the broadness of limb movement and is based on the ratio of foreground pixels. The five indicators obtained from silhouettes and GEI are then used to assess gait asymmetry.

A last indicator is obtained called falling risk (FR) which relates the support area, measured as the distance from the toe of the front foot to the heel of the rear foot, with body tilt, obtained through the head horizontal position. Both measurements are observed when the feet are furthest apart from each other.

Another important product of this work is the INIT Gait Database, available to the research community, which consists of binary silhouette sequences of images taken from recordings of volunteers simulating different gait pathologies as well as their normal gait.

To tackle the self-occlusion problem accentuated in patients whose impairments reduce the step size significantly, Verlekar et al. [86] proposed a new way to determine step related indicators based on the detection of foot-flat (FF) events, where the foot is in full contact with the ground. Considering that

this event occurs between two HS, one of each foot, an average feet image (AFI) is obtained by averaging the lower 10% of each silhouette in this time interval. the HS events are obtained, like before, when the bounding box is maximum. By applying a threshold to the AFI [87], the FF position is obtained as the centroid of the resulting image. From this, step length is measured and other indicators are derived from it, such as step length symmetry, normalized step count and speed.

Using the FF image computed before, another indicator named foot flat ratio (FFR) is obtained by computing the overlap ratio between the FF and the silhouettes of the corresponding stance phase, as seen in Figure 2.12. This represents the amount of the stance phase where the foot is in complete contact with the ground, which can be a good indicator to distinguish between types of impairments.

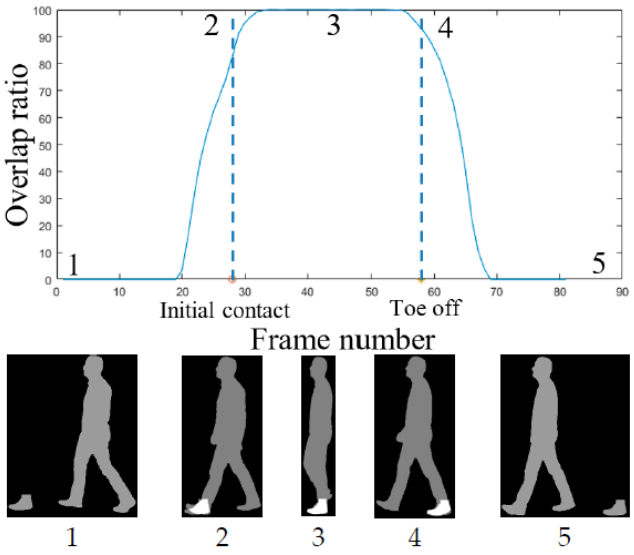


Figure 2.12: Plot and silhouettes representing foot flat overlap ratio [86].

Based on the GEI three more indicators are computed. Amount of movement (AOM) in every half gait cycle, that shows movement during body weight shifting from one side to another and can be a good indicator of one-sided impairments. Shift of center of gravity (COG) in relation with the center of support (COS), and torso orientation, that can identify changes in posture.

Using a classification method based on these indicators and using SVM, a classification accuracy of 98.8% was obtained when tested on the INIT gait database.

Other studies featuring shape-based methods for vision gait analysis include the work by Nguyen et al. [88], that measures motion variation using the motion history image (MHI) [89] and also builds a model of the walking cycle using hidden markov models (HMM). Bauckhage et al. [90] presented a homeomorphism between 2D lattices and binary shapes that provides a robust vector space embedding of segmented body silhouettes using it to track movements of limbs and detect abnormal gait. Stone and Skubic [91] used a system for 3D stride-to-stride measurements and compared the use of vision cameras and kinect sensors. Recently some interesting works aimed at gait recognition have been made using shadow biometrics [92, 93], which could also be adapted into the clinical gait assessment area.

Model-based Methods

When using model-based techniques, after acquiring images from one or more cameras, a model is matched with the sequence of images where gait is observed, usually in the form of a human skeleton model or by segmenting the different body parts into independent shapes. After having this information about segment and joint positions it is easy to obtain indicators such as joint angles, stride length etc.

Leu et al. [94] propose a method for gait analysis based on the extraction of joint positions and angles from 2D images captured from sagittal and frontal planes. After applying background subtraction and segmentation techniques to obtain the silhouette, the graphs for the number of segmented pixels in each row (vertical projection) and in each column (horizontal projection) are obtained, as seen in Figure 2.13 and used to determine the position of the neck joint and arm regions. The neck is considered to be at the y position of the global minimum at the top of the vertical projection graph and at the central x coordinate of the segmented image. Arm regions are at the x positions that correspond to the "hills" in each side of the horizontal projection graph.

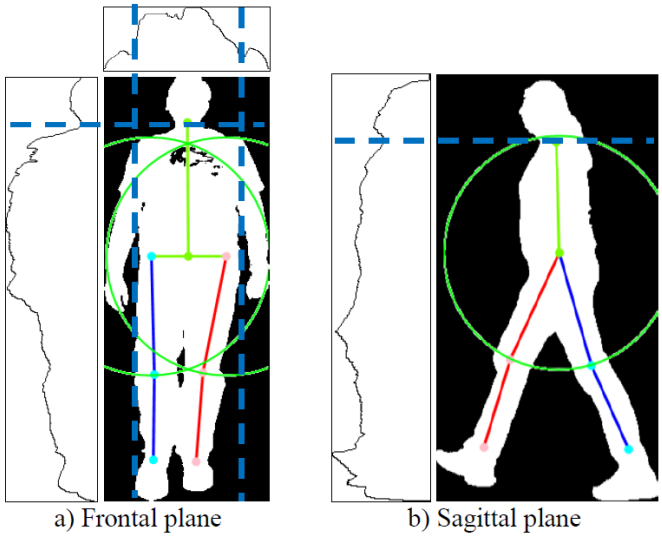


Figure 2.13: Joint detection based on frontal and sagittal silhouettes and their vertical and horizontal segmented pixel projections [94].

Other joints like the hip and segment lengths are obtained by combining the segmented image data, neck joint position, the height of the subject and statistical anatomical measures of the human body segments described in by Winter [95]. Considering the hip as the center of a circle with radius equal to the estimated length of the thigh, the knee joints are defined as the center point of the arcs of this circle that intersect the segmented parts of the image corresponding to the legs.

The method was validated against a goniometer by comparing the knee angle evolution during a gait cycle, shown in Figure 2.14 and by observing that the knee angle curves obtained from twenty healthy subjects matches the one found in literature [1, 16].

The second part of the study by Nieto-Hidalgo et al. [84] mentioned before, presents a similar algorithm to obtain an approximated full body 2D skeleton model from the silhouettes of the gait sequences captured with a single camera from the sagittal plane. They present however a different method of joint localization.

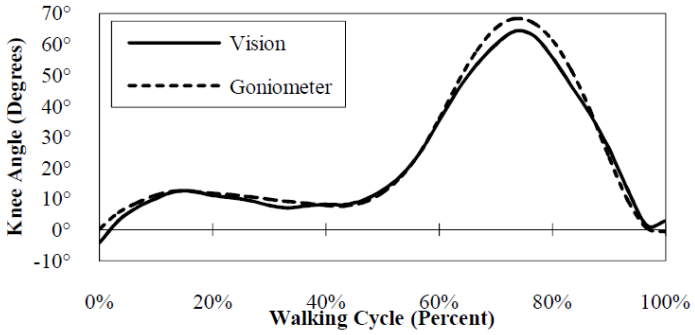


Figure 2.14: Comparison between the knee angle obtained with vision system and measurements from a goniometer [94].

The first step is dividing the silhouette image in fractions of the height corresponding to head, torso, thighs and feet, according to the model defined by Tafazzoli and Safabakhsh [96]. Then the head, torso and thighs fractions are split in half horizontally and in each half the COG is computed and projected up, for top halfs, and down, for bottom halfs, in the fraction frontiers. The thighs fraction is additionally splitted vertically and the COGs of each half are calculated and projected on the bottom frontier. The feet locations are obtained in the shape-based method described in the previous section. The computed points are represented in Figures 2.15 and 2.16

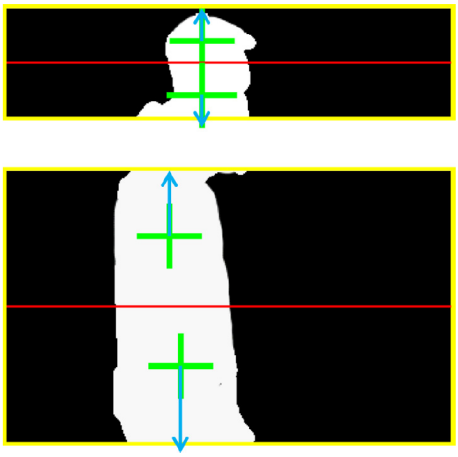


Figure 2.15: COG computation on head and torso [84].

Neck and hip joint position is then calculated as the medium point between the two projected points in the corresponding frontiers. Knee joint is defined as the intersection of a circle centered on the hip with radius equal to thigh length with circles centered on the feet and radius equal to shank length. The final joint positions are shown in Figure 2.17. Using the angles obtained through the estimated skeleton and indicators retrieved from the first part and different machine learning algorithms, classification of normal and abnormal gait was performed in a dataset of 30 healthy subjects performing normal and simulating abnormal gait. The classification accuracy ranged from 70% to 100% depending on the indicators and classifier used. The highest score was obtained using the leg angles and K-nearest neighbors with Dynamic Time Warping as distance function.

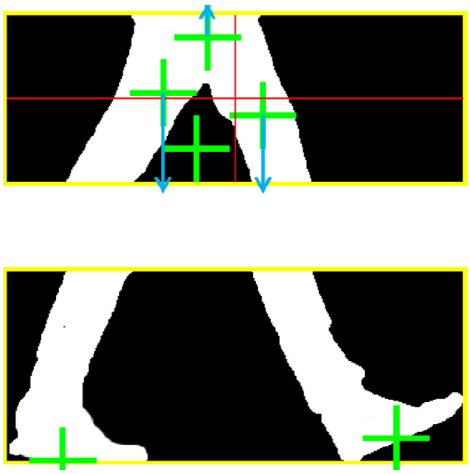


Figure 2.16: COG computation on thighs and foot location [84].

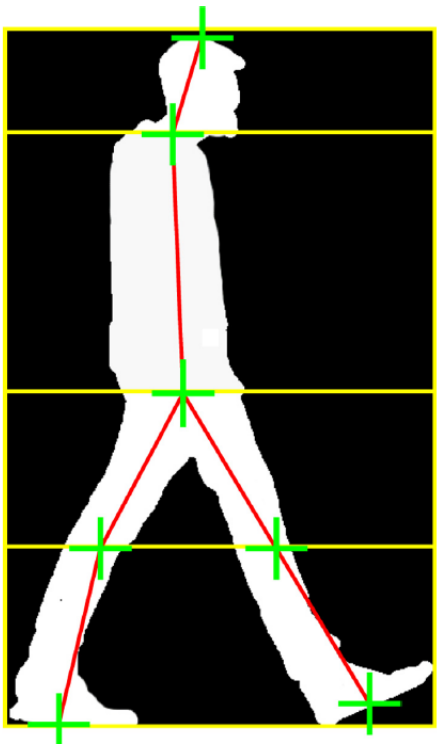


Figure 2.17: Result of the approximated skeleton computation [84].

Some other studies employing this type of method include Gabel et al. [97] who used the virtual skeleton provided by a kinect sensor. Chaaraoui et al. [98] presents a method for abnormal gait detection based on the skeleton pose representation obtained from depth data from RGB-D devices. A new feature called joint motion history (JMH) is introduced that encodes spatial and temporal information that can be continuously be obtained through time and compared with previous templates or learned with model-based machine learning algorithms. Hu et al. [99] used a structured light kinect and Hidden Markov Models for 3D pose tracking, and Jensen et al. [79] used a ToF camera to estimate pose based on an articulated model.

2.3 Commercial Systems and Gold Standards

There is a number of specialized companies that have developed various high precision gait analysis systems based on a variety of wearable and non wearable sensors, including the ones mentioned in this report, and dedicated laboratories for high accuracy motion capture and gait assessment. The most common are multiple camera-based systems used for very accurate motion capture, often using markers. These systems, however, require a large amount of equipment and calibration procedures that make them impractical and time-consuming, are expensive and are not available in most daily life situations. Nonetheless, they are an important part in the development of new gait assessment systems as they are considered as the gold standards for the clinical evaluations and the results of new systems and methods are usually validated against their measurements as ground truth.

Some examples of companies that develop highly accurate gait analysis systems in laboratory settings, using cameras and other types of sensors like IMU, force plates or EMGs, that have been used as gold standard in some of the studies mentioned above, include Vicon [100], Qualisys [22] (Figure 2.18) and BTS [40] (Figure 2.19).

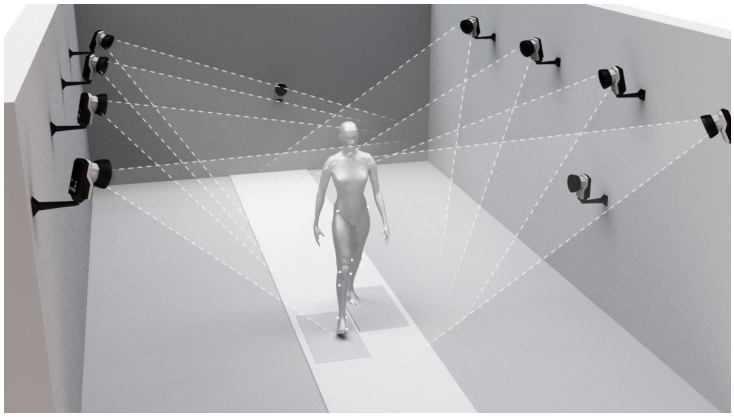


Figure 2.18: Representation of a Qualisys gait lab setup composed of an array of cameras and floor force plates [22].

There are also companies that produce smaller scale products like camera sensors or wearable inertial sensor-based systems. Vicon also has single high-speed cameras and IMUs. NDI [38] has optical measurement systems and Electromagnetic tracking systems (ETS). Xsens [101] develops inertial sensors and integrates them in full-body systems.

Figure 2.19: Representation of the BTS GAITLAB that includes the use of different optic sensors, wearable inertial sensors, floor force sensors and wireless EMG [40].

2.4 Deep Learning Vision Methods

Recently, many studies have used a variety of deep learning architectures in topics involving gait, like recognition, action recognition and medical analysis. These methods can take the raw images as input, or a pre-processed version, like silhouettes or GEIs, and then execute image feature extraction and classification, as illustrated in Figure 2.20. This section will review the methods that can be employed in each step of this process, starting with the pre-processing methods and then the types of deep-learning architectures and machine learning classification methods that are more commonly used.

Figure 2.20: General block diagram of the proposed gait analysis system.

2.4.1 Image Pre-processing

In the context of vision based methods for gait analysis, the raw data is composed of videos that capture gait sequences. However, these raw images are, as seen in the previous sections, rarely used directly as input. In fact, some pre-processing is required in order to simplify the input data while keeping the relevant information for the task in hand. Besides this, some methods use different types of input images, like GEIs, optical flow images or pose heatmaps, that need to be computed beforehand.

Background Subtraction

The first step of input pre-processing seen in almost every gait analysis method that was reviewed, consists in the application of background subtraction techniques, in order to obtain a sequence of binary images with silhouettes of the subjects. Background subtraction, associated with the foreground or movement

detection problems, is also done in many other applications, for example in video surveillance, and, as such, has been the target of a large amount of research, leading to the existence of a variety of different approaches to the problem. These methods have the common objective of building a model for the static background in a sequence of images and separating it from the moving objects to be detected and labeled as foreground.

Bouwmans [102] presented an overview on both the traditional and more modern methods for foreground detection. These methods range from basic background modeling, using the temporal mean, median [103] and mode [104] of the pixel values, to the usage of statistical models, for example using Gaussian distributions [105], neural networks and more advanced approaches. Bouwmans et al. [106] also made a review on more sophisticated methods, using deep neural networks, for background subtraction. Neves et al. [107] performed an in depth evaluation and comparison of different background algorithms, considering the *hardness* of real background scenarios and changes in the parameters configurations.

The advancements in background subtraction algorithms are motivated by a number of challenges present in the most complex scenarios, that exist mainly in the video surveillance research field. Some examples, depicted in Figure 2.21, are the existence of shadows and reflections, originated from both background and foreground elements, and also changes that may happen in the background or in the scene lighting.

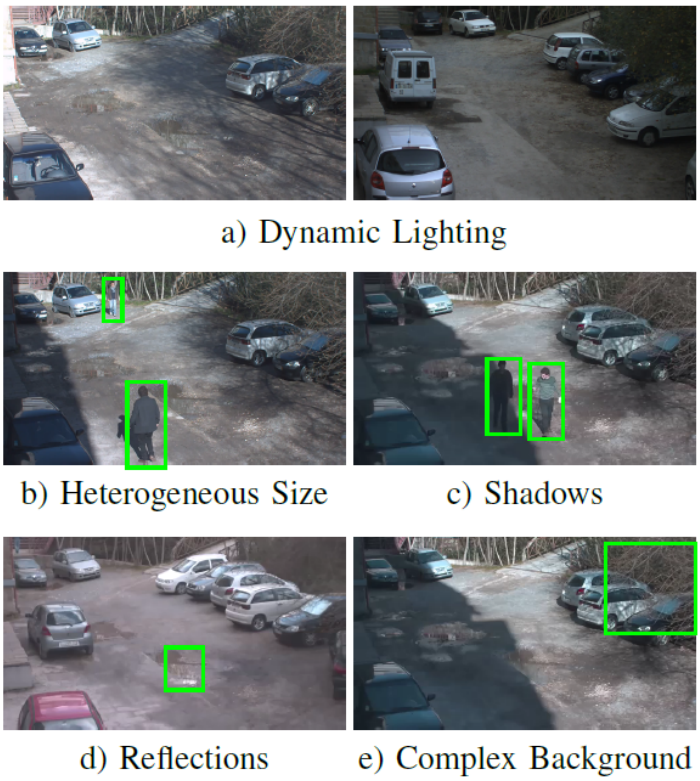


Figure 2.21: Typical challenges of Background Subtraction in outside scenarios [107].

Gait Energy Image (GEI) Computation

Another type of input that some gait analysis algorithms use is the GEI [108]. This gait representation image is useful in the way that it condenses the dynamic gait information of the silhouettes captured during a gait cycle into a single image. It is commonly computed by taking the mean image of the binary silhouette images that compose a complete gait cycle, aligned horizontally by the centroid of their upper halves. Besides visually representing relevant information, such as gait symmetry, amplitude of movement and body posture, it is also robust to imperfections that may exist in the binary silhouettes, as they are attenuated in the mean computation. It can be used directly in classification methods or be used to extract additional features. An example of the calculation of a GEI is shown in Figure 2.22.

Figure 2.22: Different binary silhouettes and corresponding Gait Energy Image (GEI) [84].

Optical Flow Computation

One of the problems, mostly in deep learning techniques, of using a single image as input is that the information available is mostly spatial, lacking dynamic and temporal information, which can be important in a task such as gait analysis, where the main features are present in the sequence of movements during gait.

An optical flow image, represented in Figure 2.23, created based on the stacking of displacement vector fields of consecutive frames, provides information on the movement along a sequence of frames in a single image. Simonyan and Zisserman [109] showed that it was advantageous to explicitly compute optical flow to encode motion information and using it as input to a CNN, rather than using single frames only, in the task of action recognition in videos. They also describe variations of this motion representation that can be used as different types of input in deep learning algorithms.

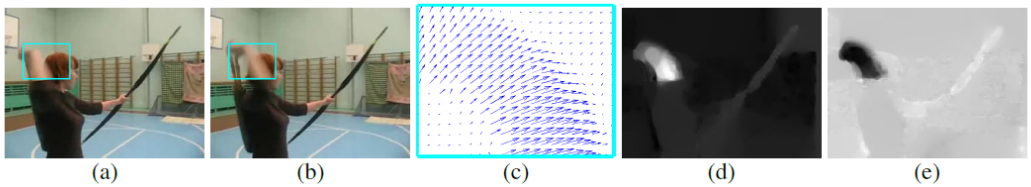


Figure 2.23: Optical flow. (a) and (b) represent a pair of consecutive frames. (c) is a close-up dense optical flow of the highlighted area. (d) and (e) are the horizontal and vertical displacement vector fields respectively. [109].

Pose Data Extraction

Another type of input that can be used to analyze the motion of body parts is pose data, which can be in the form of raw joint coordinates or images representing confidence heatmaps of the position of joints, as

seen in Figure 2.24. Pose data can also vary between 3D or 2D coordinates, of which the latter is easier to obtain from a single video sequence. This type of information eliminates some drawbacks of silhouette based methods, like the variance of the shape of each subject, due to different body shapes and clothes, that can help make datasets more uniform across subjects.

Some examples of 2D pose estimation methods include the works done by Wei et al. [110], Simon et al. [111], Cao et al. [112, 113] which resulted in the creation of the OpenPose system [114], which can detect human body, hand, facial and foot joint keypoints on single images and is available for free use online. Other pose extraction methods include the LSTM pose machine by Yue Luo et al. [115], which is based on the correspondent type of neural network 2.4.5. Pfister et al. [116] developed a similar system, also using a CNN, that uses optical flow to correct errors in the estimation of the joint positions by warping the heatmaps along the computed optical flow from neighbouring frames. This is represented in Figure 2.25.

Figure 2.24: Convergence of the estimation for the right elbow position in a convolutional pose machine [110].

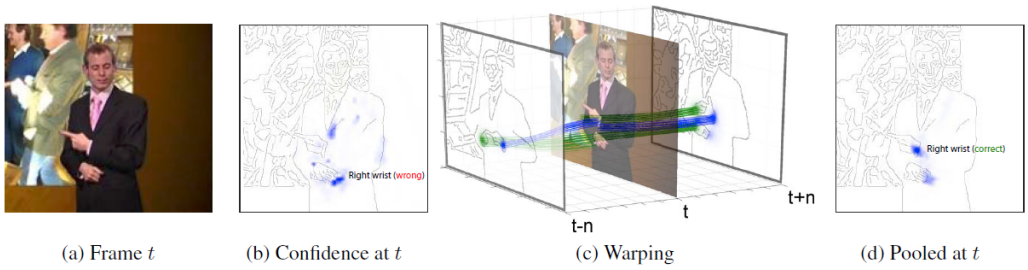


Figure 2.25: Warping neighbouring heatmaps to correct pose estimate errors [116].

2.4.2 Convolutional Neural Networks

One of the most used architectures is the Convolutional Neural Network (CNN), that has gained popularity as an advanced image recognition method among competitions such as the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [117]. The CNN is a deep neural network organized in parallel layers that use convolutional filters to detect features in images. While in the first layers basic shapes, like edges and curves are detected, in the subsequent layers, the features become more abstract, allowing for the detection of more complex structures in the image. When training data is insufficient for training

a CNN, which is a recurring problem in gait analysis, transfer learning [118] can be employed, using a pre-trained network and fine-tuning it with the available data, to adapt the network to the actual problem being explored. Moreover, instead of using the network as a classifier, the input to the fully connected layers can be used as feature vectors, using the CNN as an automatic feature extractor instead of using it directly for classification.

The use of the CNN for feature extraction in gait images has been explored by many authors, like Verlekar et al. [119], who fine-tuned a pre-trained VGG network to extract features from GEI images (Figure 2.26). The first fully connected layer is taken as a feature vector and Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are applied for classification between normal gait and four types of gait abnormality, achieving an accuracy of 95% using the DAI Gait Dataset 2 [120]. Li et al. [121] also used a VGG-like network (Figure 2.27) for feature extraction and, after applying PCA for dimension reduction, used Joint Bayesian for gait recognition, achieving an identification rate of about 89%, using the OU-ISIR dataset [122].

Another architecture is used by Shiraga et al. [123], who develop their own CNN called GEINet, that also takes GEI input images, and use it for cross-view gait recognition.

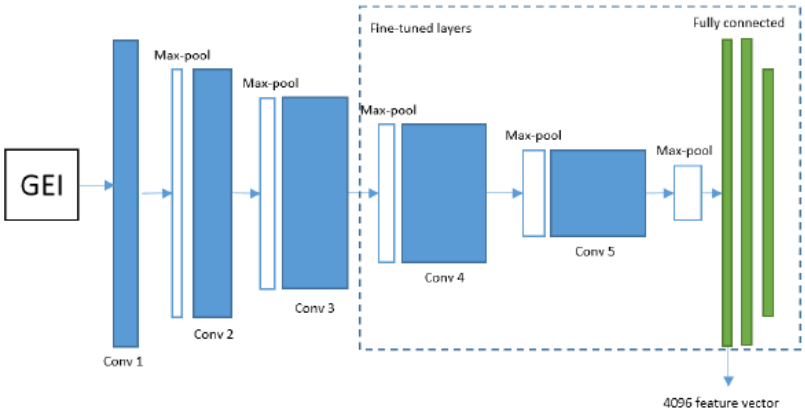


Figure 2.26: Feature extraction (VGG-19) architecture [119].

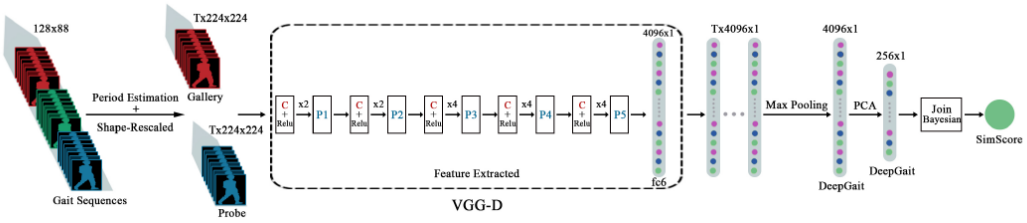


Figure 2.27: DeepGait architecture [121].

2.4.3 Optical Flow in CNN

Although the CNN is useful for image feature extraction, it only operates in single images, which provide spatial information only and ignores the temporal information that may exist because of the possible

correlation between different input instances such as the different frames in video sequences. In particular, when analyzing gait, movement information can be of interest in addition to the still image inputs that are normally used in CNCNNN architectures and although GEI images already provide some temporal information along a gait cycle, some techniques have been used to further capture motion information and use it as complement to the still image information.

One of the simplest ways of achieving this is the additional use of optical flow images as input in CNNs, as done by Simonyan and Zisserman [109], who developed a two-stream architecture that takes both a static image and a multi-frame dense optical flow image as input and performs action recognition in videos (Figure 2.28). The optical flow images are computed based on the displacement vector field detected between key points in consecutive frames and can be stacked across consecutive frames to create an image that represents motion in the selected sequence. Thus the network is composed of a spatial stream, that takes a still image frame of the video as input and a temporal stream, that takes the optical flow computed using a number of neighbor frames of the selected one. The softmax scores of the two streams are then fused using averaging or a linear SVM. The results showed not only that the temporal stream performs better than the spatial one, but, most importantly, that the fused network outperforms either one of them when used alone.

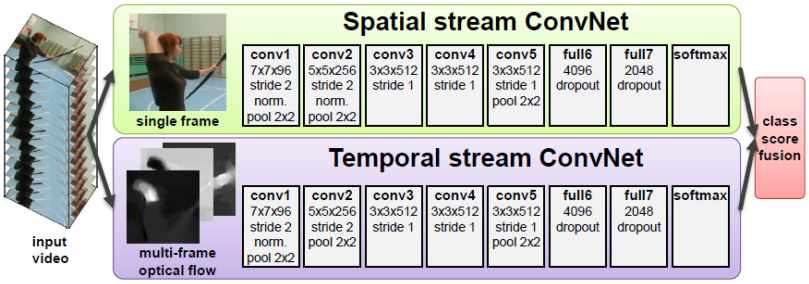


Figure 2.28: Two-stream architecture for video classification [109].

Many other types of architectures focusing on video classification, that already incorporate other ways of obtaining temporal features [124–126], actually include a hard-coded optical flow computation as additional input as it has shown to make small result improvements. This may be because, although we could expect a network that processes a sequence of frames, like the ones presented in the next sections, to automatically learn features related to optical flow, most of the time the training data may not be sufficient for that learning capacity, and therefore the addition of optical flow can improve the results of many types of architectures.

2.4.4 3D Convolutional Neural Networks

Another way of taking in consideration spatio-temporal information from videos with CNNs is the usage of 3D convolutions. This basically consists on extending the convolution filters to a third dimension, that corresponds to time, and performing convolutions in three dimensions, ending up with activation volumes as opposed to activation maps in 2D convolutions. This allows for the information of multiple frames, in a small time interval, to flow in the network and eventually fuse together to obtain spatio-temporal

features in the last layers.

One of the earliest implementations of this architecture was done by Ji et al. [124] in the context of action recognition (Figure 2.29). In their design the input is composed of seven frames, and so this is the size of the time dimension of the convolution filters. The original input is first pre-processed to generate multiple channels, including gray, gradient and optical flow versions of the frames, to which the convolutions are then applied separately. The average action recognition accuracy obtained was 90%.

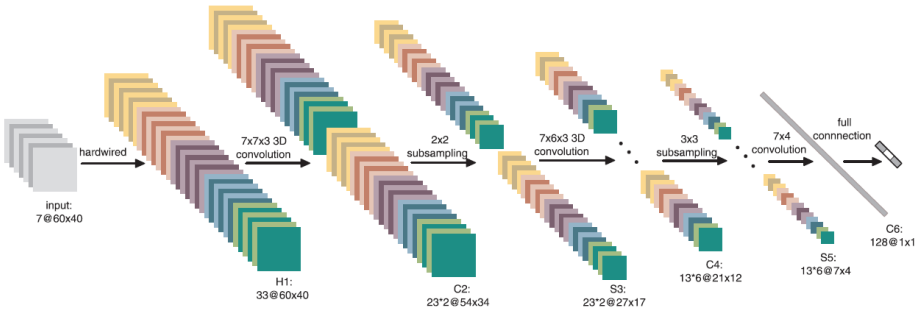


Figure 2.29: 3D CNN architecture for human action recognition [124].

The first large scale application of 3D convolutional networks was done by Karpathy et al. [125] using a large database of videos with the objective of classifying sports. They also explore different ways of fusing temporal information in videos using convolutions and design a network composed of a stream that takes low resolution images as input and another one that takes a higher resolution crop of the center of the same image (Figure 2.30). An interesting conclusion of this work is that, for their task, transitioning from a single image model to an image sequence model does not result in a very significant improvement, suggesting that local motion information may not be critically important for some, if not most, situations.

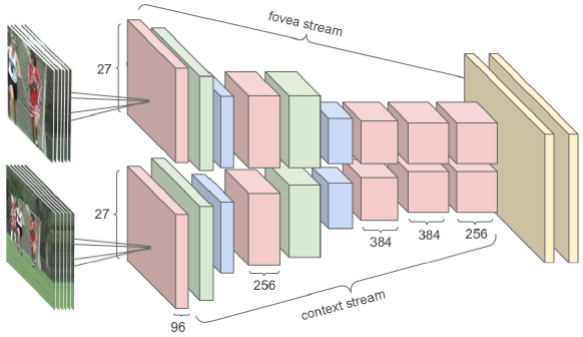


Figure 2.30: Multiresolution CNN architecture [125].

Other works include the ones by Tran et al. [127], who designed a simple and easy to train 3D CNN, based on small convolution filters, that can be compared to a 3D convolution version of the VGG network, and, in the context of gait analysis, Wolf et al. [128] developed a network that also takes gray scale and optical flow images as input, with a time dimension of 16 frames, and performs 3D convolutions in order to execute multi-view gait recognition. Results ranged from 96% to 99% using the Casia-B Dataset [129].

2.4.5 Recurrent Neural Networks (RNN)

Although the use of optical flow and 3D convolutions is effective in extracting local motion information from videos, these methods can not perceive possible dependencies between frames that are farther apart in time. Recurrent Neural Networks (RNN) however, are designed to extract this kind of information from data sequences, as they, theoretically, allow for infinite context of the previous time instances. Specifically the Long Short Termed Memory (LSTM) recurrent networks have recently been increasing in popularity due to their capability of modeling both short and long term dependencies between instances in data sequences.

One of the first applications of LSTM was in a work by Baccouche et al. [126] who used a combination of a 3D CNN, for feature extraction considering local motion, with an LSTM, that takes into account the global temporal evolution of the learned features (Figure 2.31).

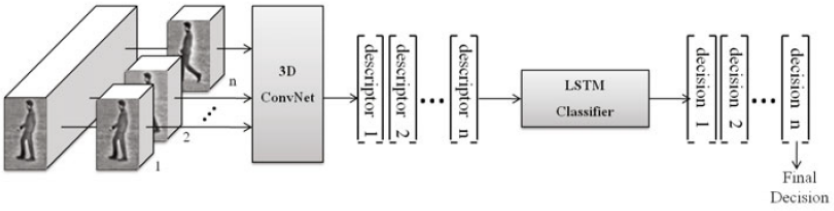


Figure 2.31: 3D Convolutional network and LSTM architecture for human action recognition [126].

More recently this architecture has been widely used in video classification and description as seen in the study by Donahue et al. [130] which also uses CNN for feature extraction in each image frame and then LSTM to classify each frame, taking into account temporal information.

In the context of gait analysis, Zhao et al. [131] designed a dual channel LSTM network (Figure 2.32) that takes as input gait data obtained with force sensors and classifies the gait between normal and three neurodegenerative pathologies, namely Parkinson's disease (PD), Huntingtons disease (HD) and Amyotrophic lateral sclerosis (ALS). The classification accuracy obtained, using data from a PhysioNet database [132], was of about 95%.

Khokhlova et al. [133] implemented an ensemble of bi-directional LSTM networks, that takes information from both the past and future at each of the time instances, using kinematic parameters obtained with a Kinect camera as input to classify between normal and two types of pathological gait. The algorithm was tested in their own dataset and obtained results of 91% accuracy.

Regarding the usage of pose data, Feng et al. [134] uses the estimation method developed by Pfister et al. [116], mentioned in Section 2.4.1, feeding the joint heatmaps to an LSTM to perform gait recognition (Figure 2.33). The LSTM is also trained to generate view invariant sequences similarly to a Generative Adversarial Network, that will be presented in the next chapter. The tests were made using Casia-B Dataset [129] and the results varied depending on the difference between the probe and gallery view angles, achieving, in the best cases, average recognition accuracy values of about 83%.

Liao et al. [135] proposed a Pose-Based Temporal-Spatial Network (PTSN), combining a CNN with a LSTM for spatial and temporal feature extraction from pose coordinate sequences for gait recognition

(Figure 2.34). Also using the Casia-B Dataset [129], a mean accuracy of 97% was obtained for the normal walking sequences, when using the same view angles for both gallery and probe sets.

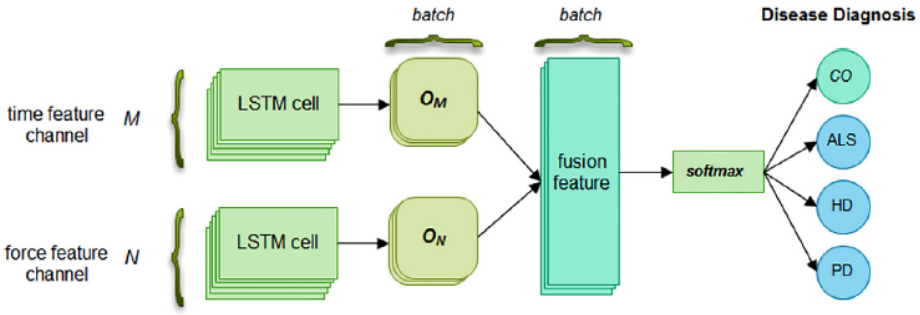


Figure 2.32: Dual channel LSTM architecture [131].

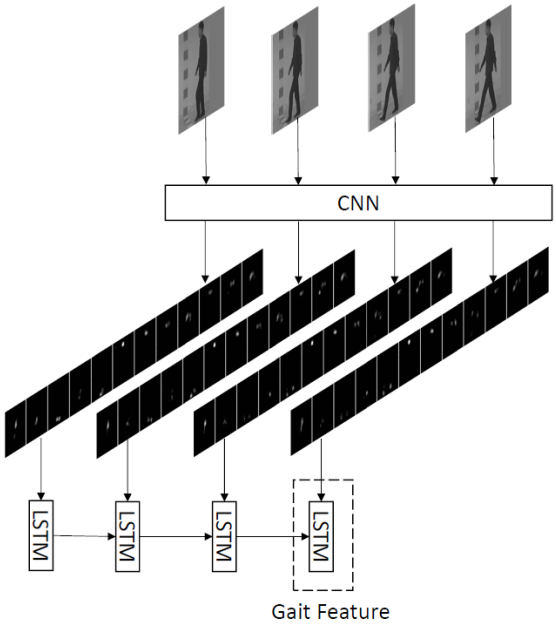


Figure 2.33: Pose heatmap LSTM architecture [134].

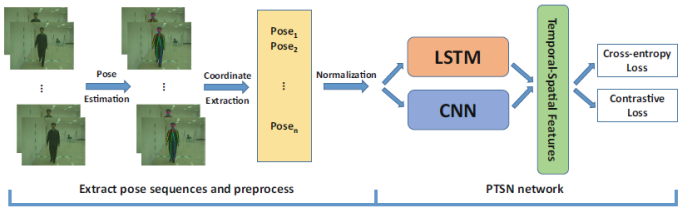


Figure 2.34: Pose-based gait recognition workflow [135].

2.4.6 Generative Adversarial Networks

Generative Adversarial Networks (GANs) are convolution based networks that rely on the generation of new data, similar to the training set, in order to make some kind of domain transformation and better

train its classification capabilities. In the case of gait analysis this domain transformation usually consists of changing the input view angle or eliminating the impact of clothes and objects being carried by the subject that may alter the shape of the silhouette used as input. A GAN is composed of a generator network and a discriminator network, each with its own purpose. The generative network is made of an encoder and decoder that encodes the input data and then generates new data based on the distribution of training data. The real/fake discriminator network takes as input either real data or generated data and tries to classify it as real or generated. In some cases there is also a domain discriminator, that takes as input real data and generated data and computes a probability of the two being related in the context of the problem. This architecture is based on the two person zero-sum game in Game Theory in which the generator and discriminator parts are competing. In this way, the generator is learning to generate data that would "fool" the discriminator, while the discriminator is learning to better distinguish between real and fake data created by the generator.

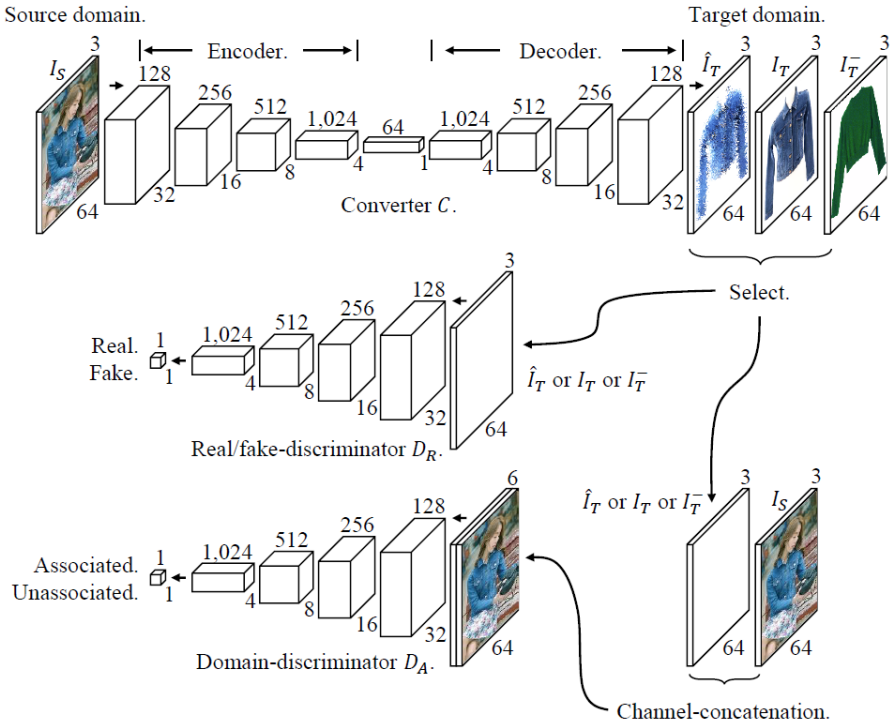


Figure 2.35: Pixel level domain transfer architecture [136].

Based on the design by Yoo et al. [136], that learns how to generate images of clothing pieces from pictures of models using them (Figure 2.35), Yu et al. [137] developed the GaitGan. This network takes GEI images of sequences filmed from different angles and where the subject carries different clothes and objects in order to generate a default GEI from a view angle of 90º (side view), preserving identification information of the gait sequence in order to perform gait recognition. The average recognition results, using the OU-ISIR dataset [122], were of about 94%.

Wang et al. [138] developed a similar architecture, named TS-GAN (Figure 2.36), with the novelty of adding a second stream to the network that performs the generation and discrimination separately in a GEI image divided into top, middle and bottom parts, in order to better learn local features. The

features from the global and part streams are then fused for gait recognition. Using the OU-ISIR dataset [122], cross-view recognition results ranged from 77% to 94%.

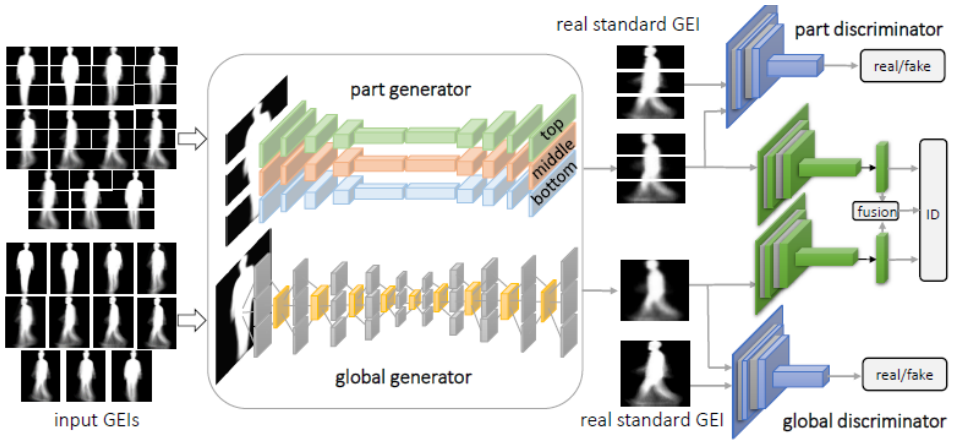


Figure 2.36: TS-GAN architecture [138].

2.4.7 Classification Methods

Sometimes, the deep learning architectures mentioned in the previous sections are not used directly for classification, but instead for extracting features of the input dataset. In this case, the feature vectors are classified using more conventional machine learning methods. This section will present a brief description of the dimensionality reduction and classification algorithms considered for this work.

Principal Component Analysis (PCA)

After obtaining the feature vectors to be used in the classification stage, a common first step is the application of PCA, a procedure created by Pearson [139] and later developed by Hotelling [140]. The main objective of PCA is to perform dimensionality reduction on the feature vectors in order to avoid overfitting. It basically takes the set of observations, that has correlated variables, and transforms it into a set of values of linearly uncorrelated variables, which are called principal components, removing redundant dimensions. The resulting principal components are responsible for explaining a percentage of the variance of the dataset, in descending order.

The covariance method for computing PCA is as follows. Given a dataset X, in which each row corresponds to a variable and each column is a dimension, the first step is to calculate the mean values of each dimension and subtract it from the original dataset, obtaining a mean-subtracted dataset

$$B = X - h\mu^T, \tag{2.1}$$

where μ is the dimension mean vector and h is a column vector of all ones, with the same size. Next the covariance matrix is calculated, given by

$$C = B^T B. \tag{2.2}$$

Finally, the matrix V of eigenvectors is computed, which satisfies

$$V^{-1}CV = D, \tag{2.3}$$

where D is the diagonal matrix of eigenvalues of C. By ordering these eigenvectors by the value of their corresponding eigenvalue in descending order and selecting a subset of the first ones we get the matrix W and the principal component decomposition of X is given as

$$T = WX. \tag{2.4}$$

For the classification task, only a subset of the first components is considered, such that it explains a chosen percentage of the original variance. This method also helps eliminating noisy components, as most of the noise will be concentrated in the later components, which are discarded.

Support Vector Machine (SVM)

The SVM [141] is a supervised machine learning method used for classification of data. Given a set of training data, it tries to calculate the hyperplanes that best separate the subsets of data belonging to each class. This creates a partition of the feature into the considered classes, that can be used to classify new inputs, based on its position on the feature space.

For the linear case, given a set of points $\vec{x}$ belonging to two different classes, we can define two parallel hyperplanes that divide the classes by the maximum margin possible as

$$\vec{w} \cdot \vec{x} - b = \pm 1, \tag{2.5}$$

where $\vec{w}$ and $\frac{b}{\|\vec{w}\|}$ are the normal vector and the offset from the origin of the hyperplane that lies halfway between the two margins, respectively. Finding this hyperplane corresponds to obtaining the $\vec{w}$ and b that minimize $\|\vec{w}\|$ subject to

$$y_i \left(\vec{w} \cdot \vec{x}_i - b\right) \geq 1., \tag{2.6}$$

where y_i is the true label of the data point.

When the data is not linearly separable, a new loss function is used, given by

$$\max\left(0, 1 - y_i \left(\vec{w} \cdot \vec{x}_i - b\right)\right) \tag{2.7}$$

and so the equation to be minimized becomes

$$\left[\frac{1}{n}\sum_{i=1}^{n} \max\left(0, 1 - y_i \left(\vec{w} \cdot \vec{x}_i - b\right)\right)\right] + \lambda \|\vec{w}\|^2, \tag{2.8}$$

where λ is a regularization term that defines the trade-off between the size of the margin and the number of data points that lie on the correct side.

In the case of nonlinear classification, a kernel method [142] is used, which basically maps the training points to a transformed feature space where a linear separation is now possible, making it easy to find

the separating hyperplane with the same method as before, that corresponds to a nonlinear classification in the original input space. There are many types of kernels that can be used that differ in the type of function used, for example, linear, quadratic etc.

When dealing with multiclass classification, there are different methods. A common solution is the creation of several binary classifiers and combining their results in order to classify each new input. This can be done with different configurations, such as one-versus-all, where each classifier distinguishes between one class and the rest, and one-versus-one, which is done by having a classifier for every pair or classes.

Linear Discriminant Analysis (LDA)

LDA is a method based on Fisher's linear discriminant [143], used for dimensionality reduction and classification. Similarly to PCA, it also tries to describe the data based on linear combinations of the variables. However LDA models the differences between classes in the data, unlike PCA, which does not take classes into account.

The goal of LDA is to project data onto a lower dimensional space with good class separability. To do this, first we need to calculate the within-class scatter matrix, which is an estimate of the correlation matrix, given by

$$\mathbf{S}_w = \sum_{i=1}^{n} \left(\mathbf{x}_i - \boldsymbol{\mu}_{y_i}\right) \left(\mathbf{x}_i - \boldsymbol{\mu}_{y_i}\right)^T, \tag{2.9}$$

where μ_k is the sample mean of the k-th class and x and y represent the vectors of samples and their labels, respectively, both of size n.

Next, assuming that each of the m classes has a mean μ_k and n_k number of samples and the overall sample mean is given by μ, the between-class scatter matrix is given by

$$\mathbf{S}_b = \sum_{k=1}^{m} n_k \left(\boldsymbol{\mu}_k - \boldsymbol{\mu}\right) \left(\boldsymbol{\mu}_k - \boldsymbol{\mu}\right)^T. \tag{2.10}$$

The last step is to calculate the lower dimensional space that maximizes the ratio of the between-class scattering to the within-class scattering, which is a set of linear combinations with coefficients w, also called Fisher's criterion and given by

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmax}} \frac{\mathbf{w}^T \mathbf{S}_b \mathbf{w}}{\mathbf{w}^T \mathbf{S}_w \mathbf{w}}. \tag{2.11}$$

Besides regular LDA, with a full covariance structure, there is also LDA with a diagonal covariance structure. This means that the elements of the covariance matrix that do not belong to the diagonal are assumed to be zero, which is equivalent to saying that the features are uncorrelated. This is preferred over the standard full covariance matrix if there are predictors with zero variance or if any of the covariance matrices of the predictors are singular. Another advantage is that it is faster, as the number of parameters to estimate is smaller.

2.5 State of the Art Summary

This section provides a review of the different types of sensors used in systems for clinical gait assessment. These systems are used to determine quantitative measurements of a variety of kinematic and kinetic gait indicators that can assist specialists in diagnosing gait impairing disorders and even help in predicting the next steps in patient rehabilitation. These indicators are also often used in combination with machine learning and classification algorithms to automatically classify between normal or abnormal gait or between a number of known pathologies and evaluate the severity of gait disorders. The systems were divided into two groups, based on their usage of wearable or non wearable sensors.

In the wearable sensors category (section 2.1), the most common type found in the literature were the ones using inertial sensors, such as accelerometers, gyroscopes and magnetic sensors, often seen in IMUs. These sensors provide kinematic gait information in the form of limb linear and angular acceleration and orientation. Other types of wearable sensors include goniometers, for joint angle measurement, textile sensors, that detect strain forces, ultrasonic sensors, to measure distances, ETS, for position detection, force and pressure sensors integrated in footwear, used to measure GRF and EMG for retrieving muscle contraction information. Most of these methods are used to detect gait events, like HS and TO and to calculate step and stride related indicators, like length, speed and cadence. The main advantage of these sensors is that they can be used without any restriction on the length of gait sequences. However they need to be attached to body parts, which which usually requires the help of a expert, and they can interfere with the normal gait of each subject, affecting the reliability of the results.

The non wearable systems (section 2.2) include the floor sensors used to detect center of pressure and GRF, similarly to the pressure and force wearable sensors. Floor sensors, however, restrict the number of steps that can be detected due to their limited size. The majority of non wearable sensors present in the literature belongs to the vision-based methods. These methods rely on the information provided by optical sensors such as infrared, ToF, structured light and, most commonly, normal video cameras. These methods compute gait indicators through image processing techniques applied on videos of gait sequences. Added to the estimation of indicators done by analyzing the shape of the captured subjects, some algorithms were developed in order to obtain skeleton models of the subjects obtaining their pose. This is done either with image processing or with the application of deep learning techniques, the so called pose machines. The main advantage of video-based methods is the fact that they are not intrusive and do not require a very complicated setup, in contrast with the laboratory systems developed by specialized companies that were also reviewed in section 2.3.

Besides the before mentioned advantages of a vision-based method, by working with images there are numerous pre-processing techniques that can be applied to obtain different types of data. Some examples are the GEIs, optical flow images and pose data, as discussed in section 2.4.1. This makes the vision-based methods the most versatile, in terms of the quantity of different types of information that can be derived from the originally captured images.

For these reasons the proposed analysis system, presented in the next chapter, will focus on images or videos as the base input data. After reviewing some possibilities for feature extraction and classification,

it was decided to focus on the use of CNNs to extract features and then use those features to perform classification using the same CNN or the machine learning algorithms described in 2.4.7.

Chapter 3

GAIT-IST Dataset Acquisition

3.1 Existing Gait Datasets

Regarding gait datasets, they can belong to two different categories, whether the objective is to perform gait recognition or pathological gait analysis. Regarding gait recognition databases, there is a significant amount of databases, as seen in the work by Verlekar [144], as their acquisition only requires subjects to walk normally. However, due to the difficulty of obtaining gait sequences from real patients, and also due to privacy and ethical issues, there are not many gait databases available dedicated to the study of pathological gait and the ones that exist are composed of simulated gait impairments and have a reduced number of samples, which is not optimal. The most notable pathological gait datasets are the following:

- The INIT Gait Dataset [85], with samples from 10 subjects, performing gait impairments in the arms and legs;
- The DAI Gait Dataset [145], with samples from 5 subjects that perform normal and random abnormal gait simulations;
- The DAI Gait Dataset 2 (DAI2) [120] that includes sequences of 5 subjects performing normal and four types of abnormal gait simulations.

These datasets also only provide the silhouettes and not the original videos or images and the silhouettes often present some segmentation errors. For the motives mentioned before, it was decided to make a new dataset acquisition, we called GAIT-IST, to be used in this work, in addition to the existing ones.

3.2 Description of Simulated Types of Gait

When deciding about the types of pathological gait to be analyzed and to be featured in the dataset, there is a decision to be made between considering the disorder that causes gait impairment or the type of abnormal gait. For this work, the latter was used and the types of gait chosen, besides normal gait, were the Diplegic, Hemiplegic, Neuropathic and Parkinsonian gaits [1, 16, 146]. This decision was made because each disorder that affects gait, such as Parkinson's, Multiple Sclerosis and Cerebral Palsy, is

often composed of a combination of different types of abnormal gait. In turn, each one of these types of gait are defined by a set of symptoms and movements that can be easily simulated. In addition, these four types of pathological gait are also the ones simulated in the DAI2 dataset, making it possible to perform some cross-database tests on the created model.

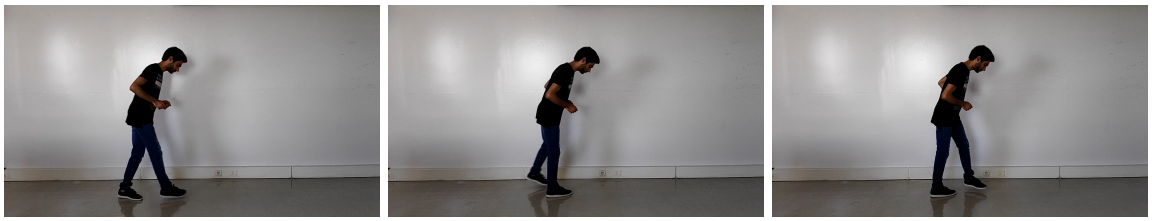
After a review of the main characteristics of the four types of pathological gait considered [1, 16, 146], subjects were instructed to replicate two versions of each type of gait, corresponding to different severity levels of the impairment in which the second level can have additional symptoms, besides a more intense version of the ones present in the first level. The different gait types, depicted in Figure 3.1, were then simulated based on the following symptoms:

- **Diplegic:** Posture characterized by a small flexion of the arms and body slightly leaning forward. There is an slight extensor spasticity causing the straightening of the legs. Tightness in hip abductors cause the legs to be close together, leading to a narrow base. Weakness on distal muscles causes foot drop, that causes a slight circumduction of the leg in each step, whit the tip of the foot pointing down and inwards. In the second severity level, hip abductors are tighter causing the crossing of the legs in what is known as scissors gait.

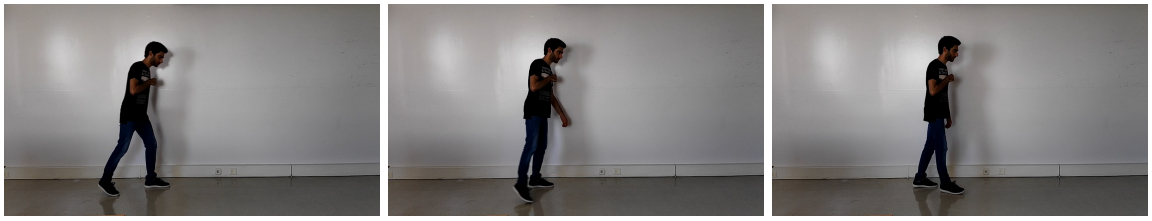
- **Hemiplegic:** Only one side of the body is affected. There is extensor spasticity and foot drop on the affected side, leading to the permanent extension and a circumduction on the steps of the affected leg, with the foot pointing downwards and making first contact with the ground using the toes. The arm on the affected side does not swing, resting still and extended in the first level while in the second level the arm is flexed and close to the chest. As the leg of the affected side is weaker there is a faster step of the unaffected side, in a limping manner, in order to minimize the duration of the stance phase of the affected leg.

- **Neuropathic:** The main symptom is the weakness of foot dorsiflexion causing foot drop. This causes the subject to lift the knee higher than usual in order to avoid dragging the tip of the foot on the ground. The first contact with the ground is done with the toe. The second level of severity consists only in a slight exageration of these symptoms.

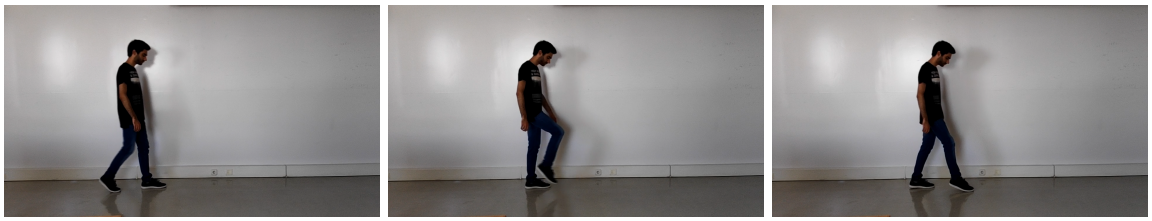
- **Parkinsonian:** Characterized by an universal body flexion, inability to perform fast movements and small shuffling and accelerating steps. In the second level the subject is asked to simulate a symptom called freezing of gait, which consists in the interruption of walking progression and posterior difficulty in resuming it accompanied by a set of small movements that display the intention. A slight shaking motion is also present in the second level, mostly in the arms and hands.



(a) Diplegic Gait



(b) Hemiplegic Gait



(c) Neuropathic Gait

(d) Parkinsonian Gait

Figure 3.1: Depiction of the four pathological gaits simulated in the GAIT-IST dataset.

3.3 Dataset Characterization

There were in total 10 subjects (8 males and 2 females) in the age range of 20-50 years old. The participants were voluntary and signed a consent form allowing the use of their images for research purposes. Each one performed 4 sequences for each of the 4 pathological gait types (2 per severity level) and 2 sequences for the normal gait, for a total of 180 sequences. In each one, the subject walks, parallel to the camera plane, one time from left to right and another from right to left, with a minimum of 2 complete gait cycles captured in each direction, which varied between gait types and subjects. For the situation of the freezing of gait, a gait cycle is considered to be the interval involving the step immediately before and after the episode.

The videos are in MP4 format were captured using a cellphone camera with a resolution of 720p, fixed on a tripod at a height of about 1,5 meters and at a distance of about 4 meters from the target. The illumination was fairly constant, mostly composed of the artificial light of the room. The background

was constant and uniform, consisting of a white wall and gray ground, with some shadows and reflections present. After pre-processing the images, the dataset is also composed of silhouette, GEI and SGEI images with size 224x224 pixels in PNG format.

Chapter 4

Proposed Gait Analysis System

The objective of the proposed gait analysis system consists in, starting with videos of subjects depicting normal gait and abnormal gait representing different pathologies, being able to classify the type of gait into one of those categories. A database of gait sequences will also be constructed to be used and help performing this task.

The development of an image classification system is often divided into different steps, as shown in the block diagram of Figure 2.20, which represents a general architecture of the system to be developed. The first step is the image pre-processing which firstly aims at obtaining a set of images that can be used as input in the next stages. In this work, two types of input are considered, and so this step will be different for each one. The resulting images are then used as input to a feature extraction module, although they can be used in the classification module directly. Finally, the resulting feature vector is used to perform classification. This chapter starts by describing the acquisition of the GAIT-IST dataset and the implementation of the different steps of the proposed system.

4.1 Image Pre-processing

After the video acquisition, the images have to go through a processing stage in order to simplify the data they represent, while maintaining the relevant information to be used in the feature extraction and classification steps, and even to create new forms of images that can be used as input in those steps. As discussed in section 2.4.1, there are many possibilities for image pre-processing techniques in order to obtain different types of input images. The ones used in this work were the GEI and a new, related representation we called Skeleton Gain Energy Image (SGEI), which is based on pose data in the form of a skeleton image. The process of obtaining these images is summarized in the diagram of Figure 4.1. This section will describe usage of these methods to classify the sequences available from the GAIT-IST dataset.

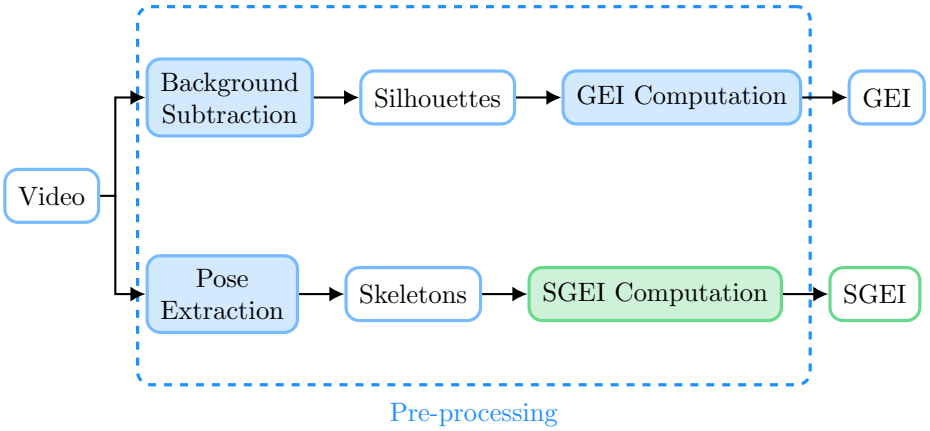


Figure 4.1: Detailed diagram of the image pre-processing implemented, corresponding to the pre-processing block in Figure 2.20.

4.1.1 Background Subtraction

In order to later obtain the GEI, the first step of image pre-processing is background subtraction, which results in a set of binary silhouette images. In the case of this work, similarly to other vision-based gait analysis works, the gait sequence videos were recorded in an indoors environment with fixed background and constant illumination settings throughout each acquisition. Furthermore, the background used is very simple, consisting mostly of a white wall. These factors make the background subtraction task less complicated, allowing for the adoption of a simple algorithm. In particular, the chosen algorithm is based on the use of Gaussian Mixture Models (GMM) for background and foreground segmentation [105]. Specifically, it was the MOG2, implemented in Python, present in the OpenCV library [147], which implements the descriptions made by Zivkovic and Van Der Heijden [148] and Zivkovic [149]. The basis of the algorithm is the creation of a number of Gaussian probability density functions for each pixel that model its probability of belonging to background or foreground.

Some parameters of the MOG2 algorithm, related to detection thresholds of objects and shadows and ratio of background on the image, were tuned by experimentation along the sequences of different subjects as there were differences in their acquisition, for example in lighting settings, as there was some natural light present and the acquisitions were taken at different times of the day, and color and types of clothing used. Furthermore, in the sequences that featured the freezing of gait, a parameter related to the amount of time that an object remains as foreground after becoming immobile also needed to be altered.

An illustration of the application of the GMM method is show in Figure 4.2.

As seen in Figure 4.2, after the background subtraction, the image is not perfect, as there is some noisy foreground detected and the silhouette has some holes and irregular edges. The first solution to deal with these segmentation errors is the application of some filters, namely a median filter with aperture size of 5 pixels for noise removal and a CLOSE followed by an OPEN morphological operations, both circle shaped and with sizes 5 and 3 respectively, for image smoothing. An example of the application of these filters is shown in Figure 4.3.

To deal with holes inside foreground objects, a simple method [150] was used which combines the

Figure 4.2: Background Subtraction results using the MOG2 algorithm.

Figure 4.3: Application of median blur and OPEN and CLOSE morphological filters.

original binary image with its inverse to fill every object, eliminating any closed regions inside them that were previously labeled as background. A result of applying this method can be seen in Figure 4.4

Figure 4.4: Application of method to fill in foreground objects [150].

The last step is the removal of noisy foreground objects resulting from noise or shadows in the image. The first measure taken towards this goal, as well as making sure that only images that contain the full body of the subject inside the frame were kept, was to delete any detected object that was in contact with the borders of the image. After this, a threshold was defined for the size of objects, to eliminate small blobs. At this stage, special attention was payed to cases in which, because of the clothing of the subject, the silhouette was not composed of a single connected object. In these cases the threshold had to be smaller, which allowed for a few larger blobs to not be deleted. However it was observed that, in almost every sequence, the position of these objects originated from noise was near the left boundary of the image. This is most likely due to it being the side where the natural light was coming from

during the video capturing. Taking this into consideration, the width of the bounding box containing every foreground object was computed and, when it exceeded a threshold defined based on the maximum width of a silhouette, the leftmost foreground object was deleted. In the end, a binary image only with the silhouette of the subject was obtained, as illustrated in Figure 4.5.

Figure 4.5: Application of method to eliminate noise blobs detected as foreground objects.

4.1.2 Gait Energy Image (GEI) Computation

In the first approach, which we will call CNN classification, a CNN was used for the classification task, namely the VGG-19, that uses 224×224 pixel images as input, so it was necessary to construct an image that condensed the information of one or more gait cycles into one single image. The format adopted was the widely used GEI for its before mentioned properties (see Section 2.4.1).

The GEI computation was done in Matlab and the first step is to obtain, a cropped image of each silhouette. This is done using a bounding box around the foreground, as seen in Figure 4.6.

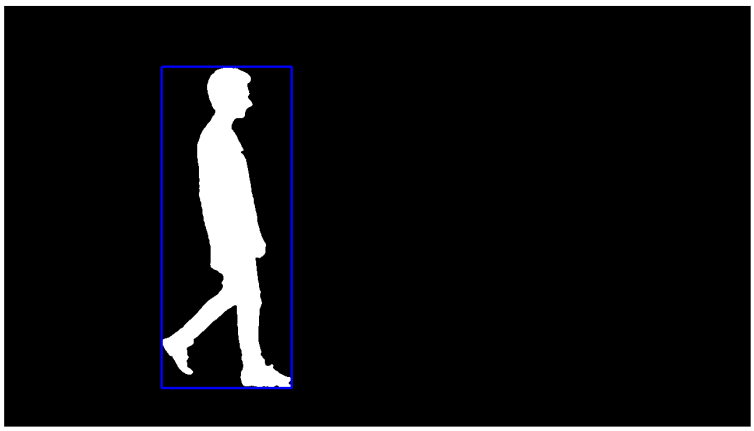


Figure 4.6: Computation of the bounding box of the silhouette.

The cropped image is then resized to the normalized height of 224 pixels, maintaining its aspect ratio. Next, the image is padded with zeros until its width is equal to 224 pixels as well. Finally, the image is shifted horizontally so that the centroid of the upper half of the silhouette is centered. An example of a final resized silhouette is presented in Figure 4.7 (left).

The next step is to identify each complete gait cycle from which a distinct GEI will be extracted. This

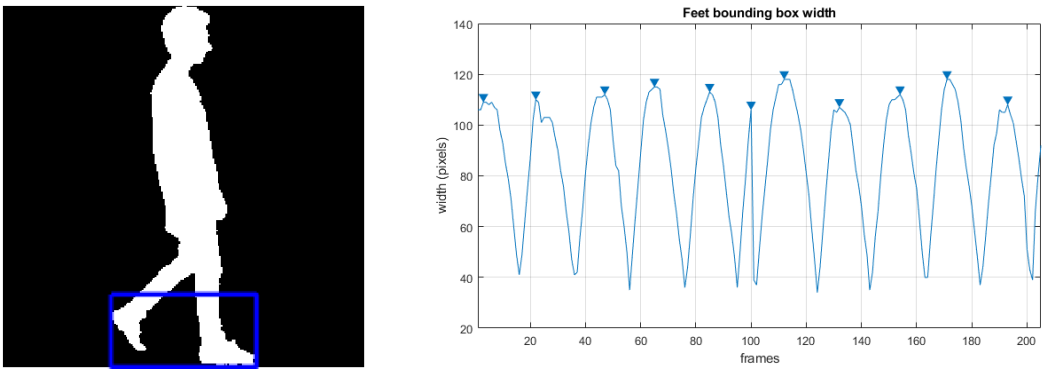


Figure 4.7: Final resized silhouette image with feet bounding box (left) and the plot of its width along a full gait sequence (right). The inconsistency of the middle peak represents the change of walking direction in the gait sequence.

is done by plotting and analyzing the width of the feet bounding box, as seen in Figure 4.7 (right), which corresponds to the distance between the heel and toe of the back and front foot, respectively. Assuming that the peaks of this signal correspond to the moment the feet are furthest apart after each step, a complete gait cycle can be said to take place once for every two occurrences of these peaks.

Knowing the numbers of the first frame a and the total number of frames N in each gait cycle, the GEI is defined as the average of the binary images I_t on this set of frames and computed as

$$GEI(x, y) = \frac{1}{a + N} \sum_{t=a}^{N} I_t(x, y), \tag{4.1}$$

as proposed by Han and Bhanu [108].

An example of some of the frames contained in one gait cycle and its correspondent GEI is presented in Figure 4.8. For every gait sequence, a GEI was computed for each individual complete gait cycle and then the frames of every GEI were used for computing a GEI of the whole sequence.

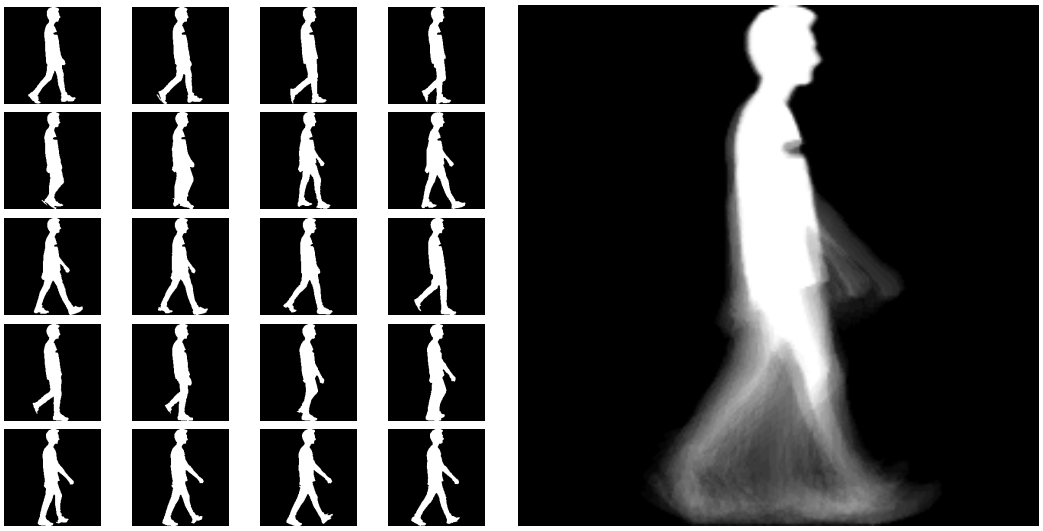


Figure 4.8: Down-sampled set of silhouette frames corresponding to one complete gait cycle (left) and respective GEI (right).

4.1.3 Pose Extraction

For the computation of the SGEI, first we need to compute the skeleton joints information from the gait videos. This is done using the OpenPose system [114], mentioned in Section 2.4.1. This algorithm can take the videos of walking people as input and returns, for each frame, a set of 2D coordinates, corresponding to different parts of the human body, as as illustrated in Figure 4.9, and labeled as follows:

0. Nose
1. Neck
2. Right Shoulder
3. Right Elbow
4. Right Wrist
5. Left Shoulder
6. Left Elbow
7. Left Wrist
8. Middle Hip
9. Right Hip
10. Right Knee
11. Right Ankle
12. Left Hip
13. Left Knee
14. Left Ankle
15. Right Eye
16. Left Eye
17. Right Ear
18. Left Ear
19. Left Big Toe
20. Left Small Toe
21. Left Heel
22. Right Big Toe
23. Right Small Toe
24. Right Heel

The OpenPose algorithm is able to detect multiple people simultaneously. Because of this, it was sometimes observed that, despite there being only one subject on each video, multiple people might be incorrectly detected. However, every time this type of error happens, the algorithm only detects one of each body part, dividing them between different detected people. Thus, in order to solve this problem, the coordinates from all the detected people are merged into a single person.

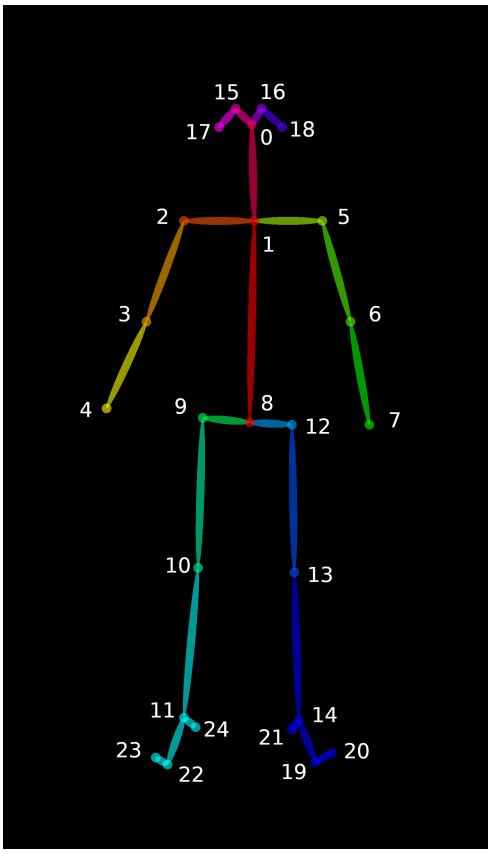


Figure 4.9: Representation of the pose data extracted with the OpenPose algorithm [114].

After having a set of body part coordinates for each frame of the videos, the OpenCV library [147] is used to draw an image of the skeleton, as illustrated on the right part of Figure 4.10, by drawing lines connecting the estimated coordinates according to the representation of Figure 4.9.

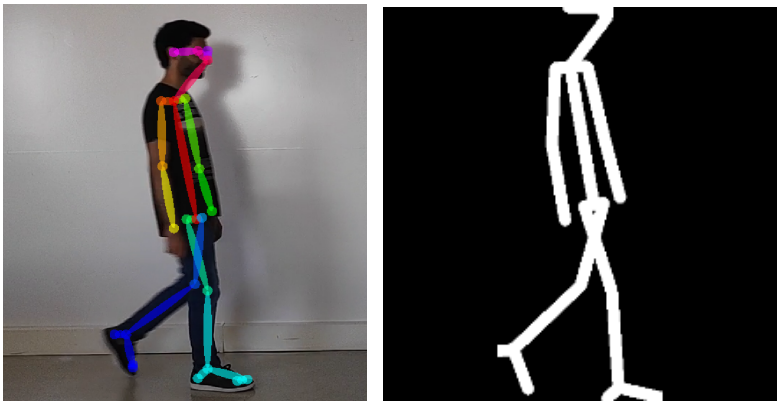


Figure 4.10: Example of the result of a skeleton extraction. Original frame with rendered pose on the left and the resulting skeleton on the right.

4.1.4 Skeleton Gait Energy Image (SGEI) Computation

After having a skeleton image for each frame of the video corresponding to one gait cycle, they can be combined together using a similar process as the one done with the silhouettes in Section 4.1.2, in order

to obtain the SGEI images. This step was done in Python, using the OpenCV library [147]. The skeleton images are cropped by their bounding box and then resized to a normalized height of 224 pixels and zero padded for same sized width. The resulting resized skeletons are aligned to place the horizontal average point between the neck and middle hip (points 1 and 8 in Figure 4.9) in the horizontal center of the image. Finally, the SGEI is computed in the same way as expressed in (4.1) using the exact same gait cycles calculated when computing the GEIs for each video sequence, in Section 4.1.2. An example of some of the skeletons contained in one gait cycle and the correspondent SGEI is shown in Figure 4.11. The SGEI, just like the GEI, allows us to have a compact representation of the movement along one gait cycle. However, due to the fact that it does not represent the physical constitution of the subjects, unlike the GEI representation, it is expected to focus more on the movement aspect of the sequences, allowing for a better description of the different types of abnormal gait.

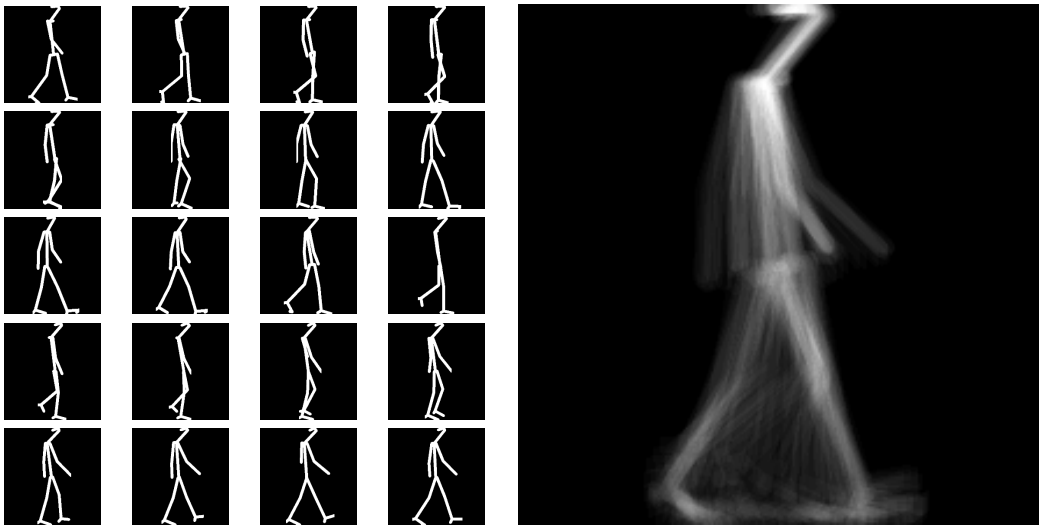


Figure 4.11: Down-sampled set of skeleton frames corresponding to one complete gait cycle (left) and respective SGEI (right).

4.2 CNN Feature Extraction

After performing image pre-processing and obtaining the desired input images, the next module is the feature extraction, performed by a CNN in order to obtain a feature vector for every image input, as shown in Figure 4.12, which will later be used for classification.

The chosen architecture for this task was the VGG-19 [151], implemented in Pyhon using the Keras deep learning library, which is a very deep CNN that has recently been successful in many image recognition and classification tasks, namely in the ILSVRC [117]. Other types of networks actually obtained better results in this classification competition, like the ResNet [152] and GoogLeNet [153]. However the VGG networks are easier to implement and train, as they converge more quickly, and are generally the best solution for feature extraction. The VGG-19 was chosen over the VGG-16 because it is a deeper network and will have a better performance overall [119]. Its input is a 224x224 image that passes through several blocks of convolutional layers composed of very small sized filters (3×3). The stride and padding

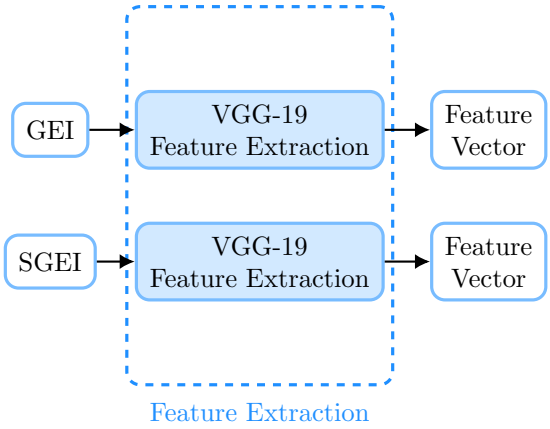


Figure 4.12: Detailed diagram of the feature extraction step implemented, corresponding to the feature extraction block in Figure 2.20.

are both fixed to 1 pixel. Spatial pooling after each convolutional block is done with max-pooling over a 2×2 pixel window and a stride of 2. The end of the network is composed of three fully connected layers, in which the first two have 4096 channels and the third one, that performs the classification, has originally 1000 channels, corresponding to the 1000-way classification of the ILSVRC, and is followed by a soft-max layer. The third fully connected layer is also changed to one with only 5 channels, corresponding to the the 5 classes of gait in our classification problem. The architecture of the VGG-19 used is presented in Figure 4.13.

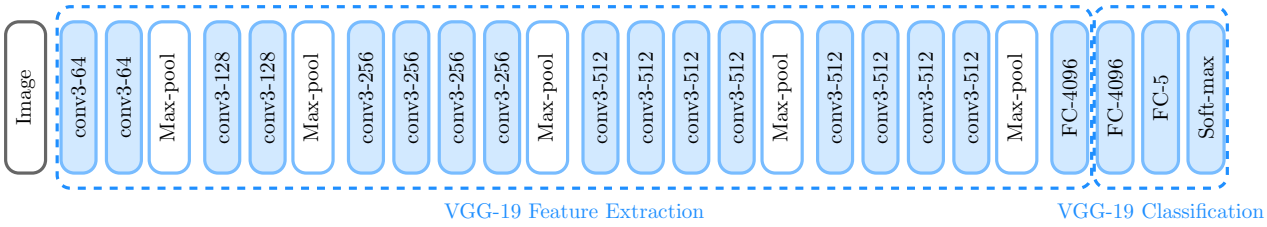


Figure 4.13: Architecture of the VGG-19 network, with the third fully connected layer modified. The first FC-4096 corresponds to the layer whose output is used as a feature vector.

As the available datasets for training are very small, it would be impossible to use them to completely retrain the network, as it would lead to problems such as overfitting. To overcome this issue, transfer learning is performed using a version of the VGG-19 pre-trained using the ImageNet database [117]. The practice of transfer learning, explored by Pratt et al. [118] is a widely used technique in machine learning, in which a model trained with a given dataset to perform some classification task is partially retrained with a second, usually smaller, dataset to perform a related task. In this case, as the first convolutional layers are responsible for detecting basic generic shapes in the images, they can be frozen during training, maintaining their weights. Succeeding layers are then fine-tuned using the GAIT-IST dataset in order for them to learn features that are more specific to our problem.

In a first stage the VGG-19 was fine-tuned and used to perform classification directly on the GEIs and SGEIs computed from the GAIT-IST dataset. However due to the low amount of data available and the high probability of overfitting when training the network, another classification method was considered.

Inspired by the work of Verlekar et al. [119], mentioned in section 2.4.2, the CNN was used as a feature extractor, taking the first fully connected layer as a feature vector for each GEI and SGEI. These vectors, which have a size of 4096, are then used for classification.

4.3 Classification

The classification step, represented in Figure 4.14 is done in two different ways. The first option is direct classification using the VGG-19 with the images obtained after pre-processing as input. The second option is, after obtaining feature vectors from the feature extraction module, performing dimensionality reduction using PCA and classification with either LDA or SVM.

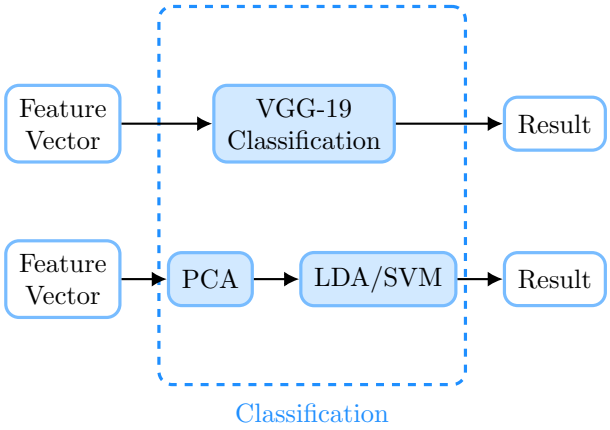


Figure 4.14: Detailed diagram of the classification step implemented, corresponding to the classification block in Figure 2.20.

The Matlab Classification Learner app [154] allows us to provide a set of data with labels and apply several types of supervised machine learning methods automatically, in order to evaluate what is the best one to be used. After using some of the extracted feature vectors as input for training all the methods available, the Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM) were observed to be the most promising ones, obtaining a higher training accuracy. The models of these classifiers were then saved and used to perform classification.

In both cases, PCA was applied to the feature vectors for dimensionality reduction, such that the resulting components explained 95% of the original variance.

Regarding the SVM, several models with different kernel functions were tested, namely Linear, Quadratic, Cubic and Gaussian kernels. It was later observed that the Linear kernel function produced the most consistent and generally best results.

In the case of LDA, both full and diagonal covariance structures were tested, of which the regular LDA with full covariance structure performed better.

Chapter 5

Results

In order to test the proposed system in the task of detecting and classifying different types of abnormal gait, results were first obtained using the GAIT-IST (see Section 3). Cross-validation tests were conducted, using the GAIT-IST dataset for training, first using the GEIs and then the SGEIs. After this, to assess the generalization capability of the system, the model trained with the GEIs was tested using the DAI2 dataset [120], which is composed of gait sequences of 5 individuals simulating the same types of gait as our dataset, allowing us to use one dataset for training and the other for testing. The computation of skeletons using the DAI2 dataset was not possible, as the original images were not available, only the silhouettes. This chapter presents the results obtained while fine-tuning the CNN, as well as the final results with the different methods proposed.

5.1 CNN Hyper-parameter Tuning

As mentioned in Section 4.2, the CNN network chosen for feature extraction and classification was the VGG-19 [151], with the architecture described in Section 4.2, pre-trained with the ImageNet database [117]. During the hyper-parameter tuning stage, only the fully connected layers were trained, while the convolutional layers kept their weights constant, as the effects of retraining more layers will be tested later. The batch size was fixed to 16, based on the memory available on the graphic card used. A higher value could help the model to converge more quickly, as more data would be used in each training epoch to update the network, however it would lead to problems when trying to allocate too much memory. The number of epochs was set to 35, as it was observed (see Figure 5.1) that the training accuracy converged close to 100% quickly and further training would lead to overfitting. Also an early stopping strategy was adopted, in which the training always runs for the set number of epochs, but the saved model is the first one to achieve the highest validation accuracy during training.

The chosen loss function was the categorical cross entropy, which is the default loss function to use for multi-class and single label classification, which means that each data instance belongs to only one of the available classes. Having M observations and N possible classes, the categorical cross entropy loss is given by

$$L(y, \hat{y}) = -\sum_{j=0}^{M}\sum_{i=0}^{N}\left(y_{ij} * \log\left(\hat{y}_{ij}\right)\right), \tag{5.1}$$

where, $\hat{y}$ is the predicted expected value and y is the observed value. For each observation, the ground truth is represented as a one-hot encoded vector, where the only non-zero value is at the position corresponding to the true class and has value 1. The output of the model is similar, but each entry of the vector is the predicted probability of the observation belonging to the correspondent class. This loss function basically compares both vectors, and the closer they are, the lower the loss.

After briefly testing some optimizers and observing the accuracies during training, the chosen one was the Stochastic Gradient Descent (SGD) [155], which is a version of Gradient Descent (GD) in which only part of the dataset is used to compute an estimate of the gradient, which achieves faster iterations despite converging more slowly. The GD optimization algorithm was first proposed by Cauchy [156] and it is used to calculate the minimum of a function by iteratively computing the gradient of the function ∇ and moving proportionally to its negative. The update equation can be written as

$$\theta_{t+1} = \theta_t - \eta \nabla l(\theta), \tag{5.2}$$

where η is the learning rate parameter and $l(\theta)$ is the loss function.

In a first experiment, with the purpose of adjusting the parameters of the optimizer, the learning rate was set to 0,001 and the network was fine-tuned with the GEIs of 5 subjects, performing 5-fold cross-validation using one subject as validation set. This means that results are obtained for five models, each using one of the subjects for validation and the remaining four for training. An example of the evolution of the training and validation accuracy and loss during one of the cross-validation folds is represented in Figure 5.1.

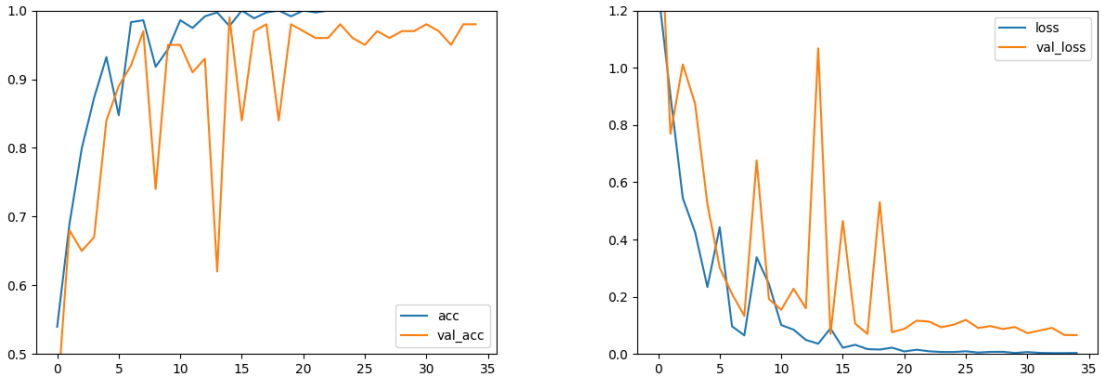


Figure 5.1: Training (blue) and validation (orange) accuracies (left) and losses (right) during training of one fold of cross-validation for learning rate = 0.001.

It can be observed that despite the accuracy and loss eventually reaching good values, there is a significant amount of oscillation that may lead to convergence problems. In order to improve this, momentum was introduced, namely a modification called Nesterov Momentum, developed by Nesterov

[157], in which the update rule is now

$$\begin{aligned} v_{t+1} &= \mu v_t - \eta \nabla l\left(\theta + \mu v_t\right) \\ \theta_{t+1} &= \theta_t + v_{t+1} \end{aligned}, \tag{5.3}$$

where μ is the momentum parameter, which defines how fast the velocity v can change. The use of momentum improves the convergence speed by giving more importance to the directions that decrease the loss function the most. After some experimentation the momentum parameter was set to 0.9 and after repeating the training of the model in the same conditions as before, an improvement in terms of the oscillations of the accuracy and loss was observed, as seen in Figure 5.2.

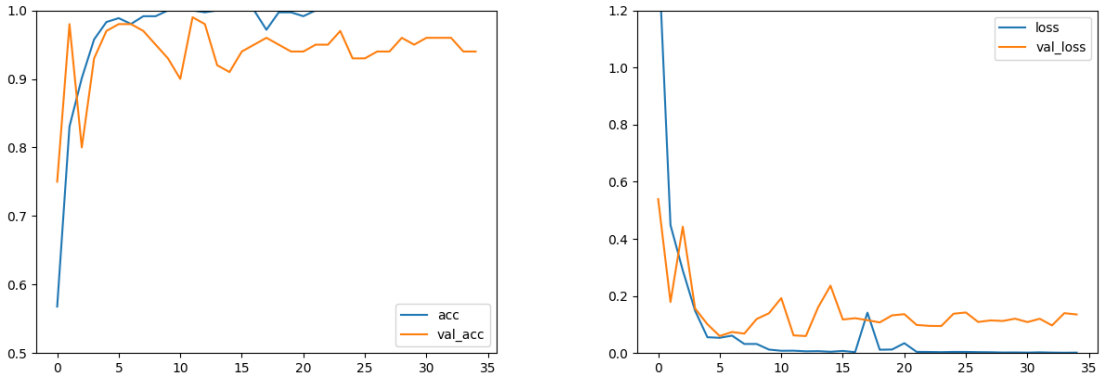


Figure 5.2: Training (blue) and validation (orange) accuracies (left) and losses (right) during training of one fold of cross-validation for learning rate = 0.001 and using Nesterov Momentum.

After this, training and validation was done using the full dataset, performing 10-fold cross-validation, separating the folds by subjects. The learning rate was adjusted, in order to further smoothen the accuracy curve. In the end, the value for the 10-fold cross-validation learning rate was set at 0.0002. The number of epochs was also increased to 50 to make sure there was no under-fitting, as the learning rate was decrease which causes a slower convergence speed. From the result of Figure 5.3 it is noted that the validation accuracy and loss stabilize early in training. Also, the training accuracy comes very close to 100% which may be a sign of overfitting which means that this number of epochs is sufficient.

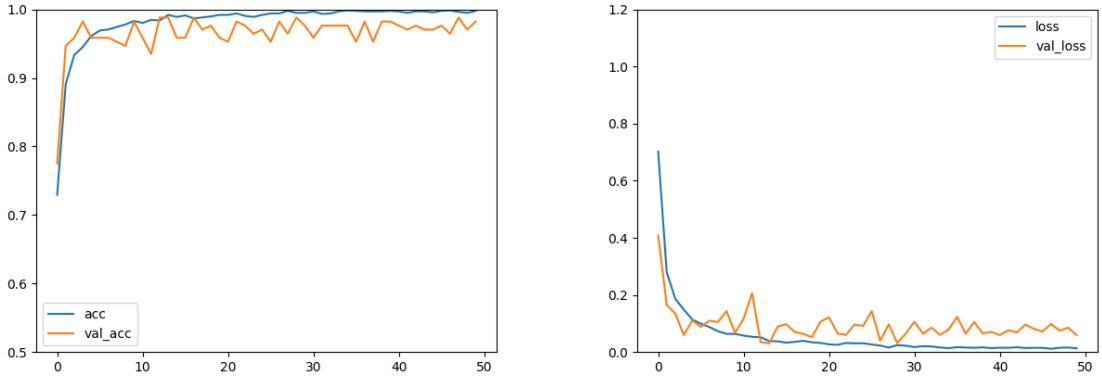


Figure 5.3: Training (blue) and validation (orange) accuracies (left) and losses (right) during training of one fold of cross-validation with learning rate = 0.0002 and using Nesterov Momentum.

5.2 CNN Fine-tuning and Classification

After the parameters of the CNN were selected as defined in the previous section, another parameter was tested while fine-tuning, which was what layers are actually being retrained with the GAIT-IST dataset. As the convolutional layers are responsible for learning shapes and patterns in the images, it can be advantageous to also fine-tune some of them using data related to the task being performed. Having this in mind, a set of tests were performed to evaluate which configuration of fine-tuning was best, in terms of blocks of layers to be retrained. In each of the tests, like before, 10-fold cross-validation was done, and for each fold a confusion matrix of the classification results for the validation set was obtained. Then a global confusion matrix is calculated, with the mean results from all the validation sets. This process was done first using the GEIs and then using the SGEIs as input.

If we consider that, in Figure 4.13 each set of convolutional layers, separated by a max-pool layer, corresponds to a convolution block, then in each experiment an additional convolution block was fine-tuned, starting from the end of the network with block 5 and ending at block 2. The global confusion matrices for the best experiments are presented in Tables 5.1 and 5.2, for the usage of GEIs and SGEIs respectively.

Regarding the confusion matrices, there are two interesting things to note. First, the Parkinsonian gait is the one most accurately classified, most noticeable when using GEIs. This is probably due to the fact that the GEIs of this class are the most distinguishable from the others due to the posture and very small step size, as described in section 3. It may also be because, as the steps are so small, more gait cycles were captured and thus more GEIs were able to be computed in each sequence, which means that there is more data belonging to this class in the dataset and the model is able to learn it better than the others. Also noticeable are the fact that the majority of the times that each gait type is misclassified, the wrongly predicted class is Hemiplegic gait. This occurrence is much less visible when using the SGEI, which suggests that this representation is better suited for this classification task. However, in this case, there is a consistent misclassification of Diplegic gait as Parkinsonian. This may be because of the similar

		Predicted Class				
		Diplegic	Hemiplegic	Neuropathic	Parkinsonian	Normal
True Class	Diplegic	**0.95**	0.02	0.00	0.03	0.00
	Hemiplegic	0.04	**0.90**	0.04	0.00	0.03
	Neuropathic	0.00	0.07	**0.93**	0.00	0.00
	Parkinsonian	0.00	0.00	0.00	**0.99**	0.00
	Normal	0.00	0.04	0.01	0.00	**0.95**

Table 5.1: Confusion matrix with the mean validation accuracy of the CNN GEI classification across all cross-validation folds, for the configuration where the VGG-19 is trained starting at the third convolution block. The overall accuracy was **94.5%**.

		Predicted Class				
		Diplegic	Hemiplegic	Neuropathic	Parkinsonian	Normal
True Class	Diplegic	**0.93**	0.00	0.00	0.07	0.00
	Hemiplegic	0.01	**0.98**	0.01	0.00	0.01
	Neuropathic	0.00	0.00	**0.98**	0.00	0.02
	Parkinsonian	0.01	0.00	0.00	**0.99**	0.00
	Normal	0.00	0.01	0.01	0.00	**0.98**

Table 5.2: Confusion matrix with the mean validation accuracy of the CNN SGEI classification across all cross-validation folds, for the configuration where the VGG-19 is trained starting at the second convolution block. The overall accuracy was **97.4%**.

posture that the two abnormal gaits present.

Regarding the global GEI classification accuracy, it can be noted that fine-tuning more convolution blocks generally leads to better results, as the best accuracies are obtained when fine-tuning from block 3 and 2 for GEI and SGEI respectively. Also noticeable is the general better results with the SGEIs experiments, which suggests that it is a more suited representation for this classification task.

5.3 CNN Feature Extraction and Classification

In order to avoid some overfitting due to working with a limited amount of data, we can use a classifier trained separately from the CNN, which is now only used for feature extraction.

Using the trained VGG-19 models obtained in the previous section, feature extraction was performed, on every GEI and SGEI, in the form of 4096-sized feature vectors obtained from the first fully connected layer. Like before, 10-fold cross-validation was performed for each fine-tuning experiment. PCA was applied to the feature vectors, for dimensionality reduction, keeping only the most representative components, and they were then used to train LDA and SVM models, to be used as classifiers. Each validation fold always corresponded to the GEIs and SGEIs used to train the CNN model from which those features were extracted, for performance comparison purposes.

Of the types of LDA and SVM tested, as mentioned in Section 4.3, the ones that achieved the best

results were the SVM with linear kernel the and standard LDA. Between these two, the latter achieved a slightly better overall accuracy and the confusion matrices are shown in Tables 5.3 and 5.4, for GEIs and SGEIs respectively. The results are very similar and it does not seem to exist any improvement when comparing to the ones using the CNN classification, obtained from the last fully connect layer of the network, as reported in Section 5.2. The same observations can be made about the high accuracy on the classification of the Parkinsonian gait, most of the misclassifications corresponding to a wrong prediction of Hemiplegic gait when using GEIs and the confusion between Diplegic and Parkinsonian gaits when using SGEIs. Also like before, the configurations that achieve the highest overall accuracy are the ones where fine-tuning is done starting at the third and second convolution block, for GEIs and SGEIs respectively. Finally, once more, SGEIs produce better results than GEIs.

		Predicted Class				
		Diplegic	Hemiplegic	Neuropathic	Parkinsonian	Normal
True Class	Diplegic	**0.79**	0.17	0.00	0.04	0.00
	Hemiplegic	0.0	**0.91**	0.05	0.00	0.04
	Neuropathic	0.00	0.11	**0.89**	0.00	0.00
	Parkinsonian	0.01	0.02	0.00	**0.97**	0.00
	Normal	0.00	0.05	0.05	0.00	**0.90**

Table 5.3: Confusion matrix with the mean validation accuracy of the LDA classification using GEIs across all cross-validation folds, for the configuration where the VGG-19, used for feature extraction, is trained starting at the third convolution block. The overall accuracy was **89.6%**.

		Predicted Class				
		Diplegic	Hemiplegic	Neuropathic	Parkinsonian	Normal
True Class	Diplegic	**0.91**	0.01	0.00	0.08	0.00
	Hemiplegic	0.0	**1.00**	0.00	0.00	0.00
	Neuropathic	0.00	0.02	**0.96**	0.00	0.02
	Parkinsonian	0.01	0.00	0.00	**0.99**	0.00
	Normal	0.00	0.02	0.01	0.00	**0.97**

Table 5.4: Confusion matrix with the mean validation accuracy of the LDA classification using SGEIs across all cross-validation folds, for the configuration where the VGG-19, used for feature extraction, is trained starting at the second convolution block. The overall accuracy was **96.4%**.

5.4 Cross-database Scenario

With the objective of testing the ability of generalization of the proposed system in the classification and feature extraction tasks, tests were made using a different gait database, namely the DAI Gait Dataset 2 (DAI2) [120]. This dataset is similar to the GAIT-IST, as it is composed of simulations of the same gait types, performed by 5 subjects. This makes it possible to use it for testing the models already trained

with the GAIT-IST dataset. As the original images from the DAI2 dataset were not available, it was not possible to make a cross-database test using the SGEI representation.

The first experiment consisted in doing CNN classification of the GEIs obtained from the DAI2 dataset using the VGG-19 fine-tuned with the GAIT-IST dataset. As previously noted, the configuration of fine-tuning the network starting at the third convolution block produced the best results for GEIs, so that was the model chosen for the realization of the cross-database testing. The DAI2 dataset was used to test each of the 10 models trained during the previous 10-fold cross-validation training and the mean confusion matrix and accuracy were computed, as seen in Table 5.5.

As expected, the results were significantly worse than when training and testing on samples of the same dataset, obtaining an overall accuracy of only 43.3%. In this case many factors can contribute to the decrease in accuracy. More specifically, the two datasets may have simulated each gait type emphasizing different characteristics of the pathologies and the GEIs may have been computed in slightly different ways. Besides, the silhouettes available from the DAI2 dataset have some segmentation errors, mainly in the form of holes in the silhouettes, which makes the classification task even harder. Still the same observations as before can be made. Again, the Parkinsonian gait is relatively well classified, second only to the Hemiplegic gait, which may be because of the, now much stronger, tendency of the network to classify every gait type with this category.

In order to improve this result, the alternative method of performing feature extraction using the CNN and then classification using PCA, LDA and SVM was tested. The feature vectors were extracted for all GEIs in the DAI2 dataset, using a model fine-tuned with the GAIT-IST dataset. Then the feature vectors were grouped by subject and 5-fold cross-validation was performed using the LDA and SVM models, in a way that each fold contains all the features that correspond to each subject. In this experiment the classification method with the best results was the LDA, with an overall accuracy of 76.7%. The confusion matrix is presented on Table 5.6. There is a significant improvement when comparing to the CNN classification, which shows that using the extracted features in a different independently trained classifier, improves the generalization ability of the system across different databases.

		Predicted Class				
		Diplegic	Hemiplegic	Neuropathic	Parkinsonian	Normal
True Class	Diplegic	**0.09**	0.50	0.18	0.23	0.00
	Hemiplegic	0.0	**0.97**	0.02	0.00	0.00
	Neuropathic	0.00	0.86	**0.09**	0.01	0.04
	Parkinsonian	0.17	0.19	0.04	**0.60**	0.00
	Normal	0.00	0.91	0.01	0.00	**0.08**

Table 5.5: Confusion matrix depicting the mean validation accuracy of the CNN classification across all cross-validation folds, when testing with the DAI2 dataset. The overall accuracy was **43.3%**.

True Class \ Predicted Class	Diplegic	Hemiplegic	Neuropathic	Parkinsonian	Normal
Diplegic	**0.53**	0.16	0.01	0.30	0.00
Hemiplegic	0.13	**0.72**	0.15	0.00	0.00
Neuropathic	0.09	0.22	**0.69**	0.00	0.00
Parkinsonian	0.08	0.00	0.00	**0.92**	0.00
Normal	0.00	0.00	0.00	0.00	**1.00**

Table 5.6: Confusion matrix depicting the mean validation accuracy of the LDA classification across all cross-validation folds, when testing with the DAI2 dataset. The overall accuracy was **76.7%**.

5.5 GEI and SGEI Result Fusion

After noticing the improvement on the results in every test, when using the SGEI representation as input, as opposed to the GEI, an experiment was done to try to merge the results of the two representations in order to further improve the results. To do this, for each gait sequence, the CNN was used to perform classification first using the GEI and the corresponding SGEI. Then, the scores for each category, obtained before the softmax layer, were observed. For each input, higher scores represent a higher probability that the input observed belongs to the corresponding category, according to the network. So for each pair of classification scores, corresponding to the usage of GEI or SGEI, the predicted category was chosen as the one with the higher score across the two sets. This can be interpreted as choosing the prediction of the model that performed the classification with higher certainty.

Using this method, an improvement on the overall accuracy was observed, when compared with the individual usage of GEI or SGEI, as observed in Figure 5.7.

True Class \ Predicted Class	Diplegic	Hemiplegic	Neuropathic	Parkinsonian	Normal
Diplegic	**0.96**	0.00	0.00	0.03	0.00
Hemiplegic	0.00	**0.98**	0.01	0.00	0.00
Neuropathic	0.00	0.01	**0.98**	0.00	0.01
Parkinsonian	0.00	0.00	0.00	**1.00**	0.00
Normal	0.00	0.01	0.00	0.00	**0.99**

Table 5.7: Confusion matrix depicting the mean validation accuracy of the CNN classification across all cross-validation folds, when fusing the results of the GEIs and SGEIs. The overall accuracy was **98.5%**.

5.6 CNN Deep Feature Visualisation

Neural networks are currently not very transparent and interpretable, as we do not fully know why and how models make decisions or learn features. As an application based on deep learning can raise questions about its reliability if there is no explanation given by the model, besides the final result, it is important to try to understand how these methods work. The Keras Visualisation Toolkit [158] includes several

techniques to give some insight about what is happening inside a CNN during classification. The two visualization techniques tested were the Saliency Maps [159] and the Class Activation Maps, or grad-CAM [160]. The results of the application of this toolkit is presented and discussed in the rest of this section.

5.6.1 Saliency Maps

The Saliency Map [159] is a way of visualising what part of the image the network is paying more attention to, in order to know if it is learning the correct features for each class or if it is looking at a part of the image that is not related with the class at all. The basis of this technique is the computation of the gradient of an output class with respect to a given input image. The end result is an image where the pixels that would produce an increase in the output value, if they were slightly changed, are highlighted.

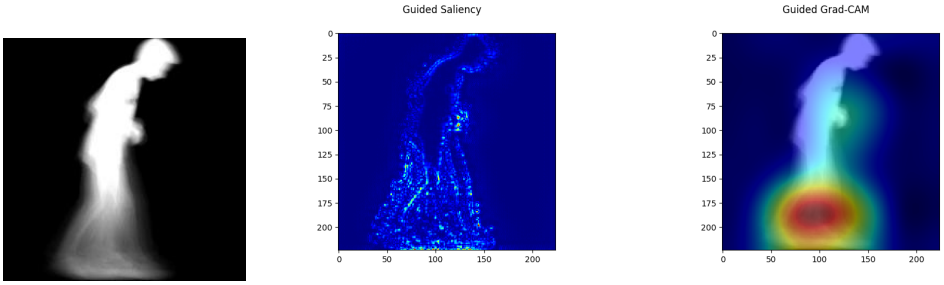
Five GEIs and SGEIs, one for each class, were chosen from the GAIT-IST dataset in order to serve as input in this technique, using the fine-tuned VGG-19 model that obtained the best results in the previous sections. The resulting Saliency maps are shown in Figures 5.4 and 5.5, in the middle image, for GEIs and SGEIs respectively.

One of the conclusions that can be made right away is that the most important pixels of the image are the ones on the border of the GEI, and on places that correspond to parts of the body that have a larger amplitude of movement. This was expected, as the inside of GEIs is mostly the same across all classes, specially in the upper body region, where there is almost no movement. This once again suggests that SGEI is a better representation as there is less information that is not used in the image. More evidence to this is the fact that the saliency maps on the SGEIs show a more distributed attention across the entire SGEI, meaning that more information is used by the CNN.

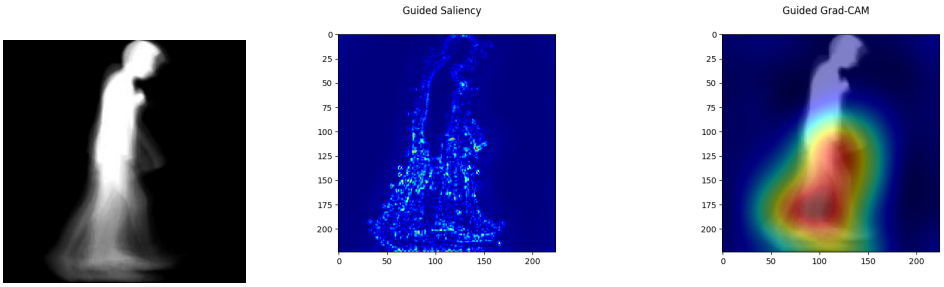
In every type of gait, except Parkinsonian, attention is given to the legs and arms. In the case of Parkinsonian gait this does not happen, probably due to the fact that there is not much movement on the legs and the features learned are more related to the posture and position of the hands.

5.6.2 Class Activation Maps

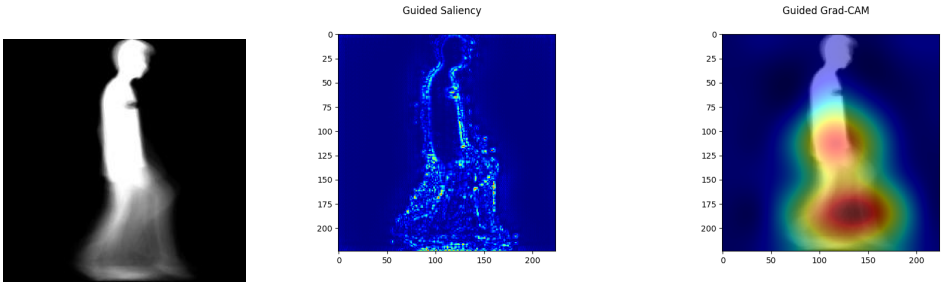
Another way of visualizing the attention is through the Class Activation Maps, or grad-CAM [160]. Instead of using the gradients with respect to the output, like Saliency, this technique obtains spatial information on the last convolution layer, which would then be lost in the dense layers that follow. Again, using the same 5 GEIs and SGEIs as before and the same VGG-19 model, the resulting images are the ones on the right of Figures 5.4 and 5.5. The same conclusions as before can be drawn, as the attention seems to be mostly focused on the leg area of the GEIs, except on the Parkinsonian gait, which gets more attention on the torso and hands. The Neuropathic gait on the SGEI also seems to get more attention on the torso, although no apparent explanation can be found for this.



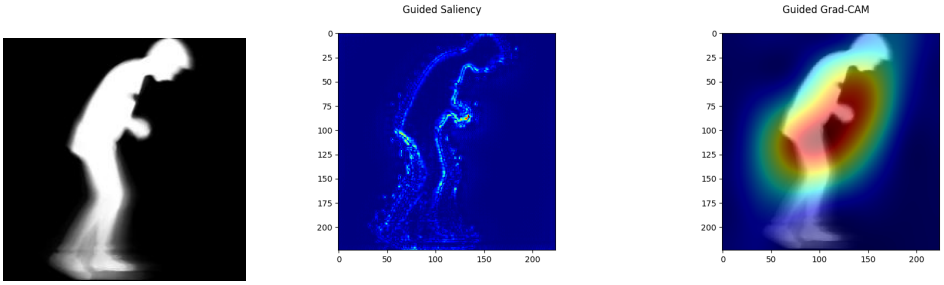
(a) Diplegic Gait



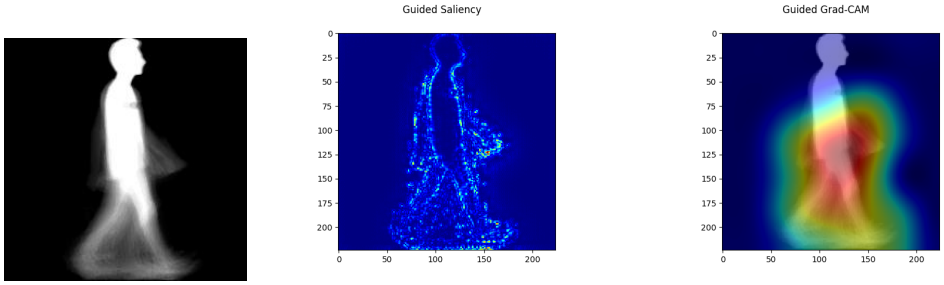
(b) Hemiplegic Gait



(c) Neuropathic Gait

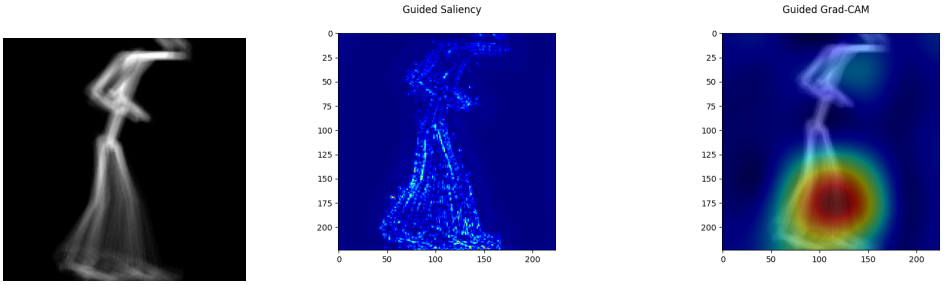


(d) Parkinsonian Gait

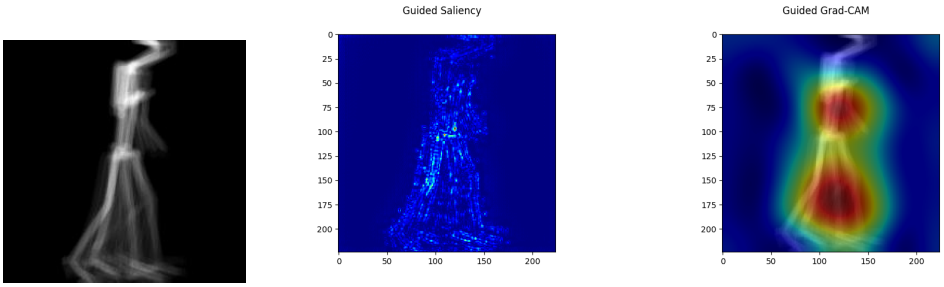


(e) Normal Gait

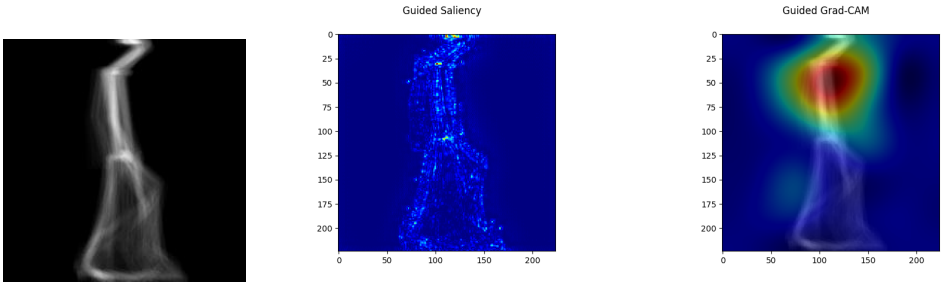
Figure 5.4: Results of the application of the Keras Visualization Toolkit on GEI examples belonging to each class of the GAIT-IST dataset, resulting in Saliency Maps and Class Activation Maps.



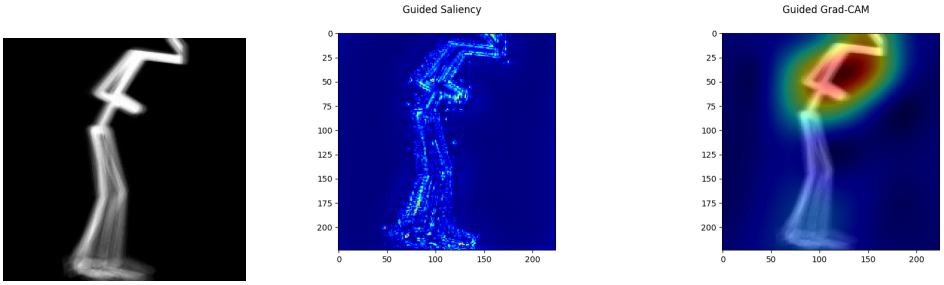
(a) Diplegic Gait



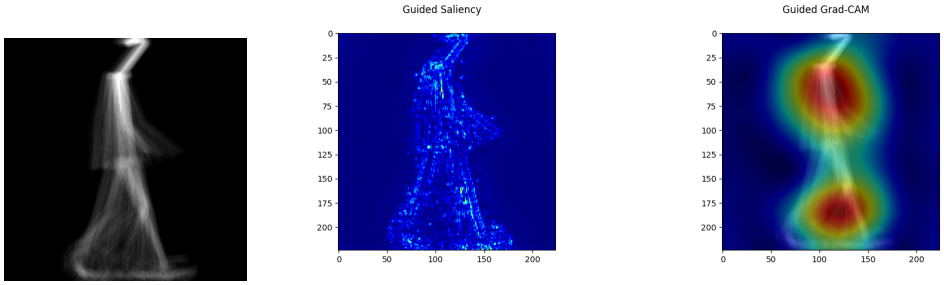
(b) Hemiplegic Gait



(c) Neuropathic Gait



(d) Parkinsonian Gait



(e) Normal Gait

Figure 5.5: Results of the application of the Keras Visualization Toolkit on SGEI examples belonging to each class of the GAIT-IST dataset, resulting in Saliency Maps and Class Activation Maps.

5.7 Results Discussion

Table 5.8 sums up the results obtained in all the experiments done. In summary, results were obtained for different sets of fine-tuned CNN layers, classification methods and sets of data.

In terms of the fine-tuning configuration of the CNN, it is concluded that, overall, the more layers are fine-tuned, the better the system performs, with observed peaks of accuracy when retraining all layers from the third convolutional block, when using GEIs, and from the second, in the case of SGEIs.

Regarding the classification methods, when using the GAIT-IST for training and cross-validation, there seems to be no improvement when using the CNN for feature extraction and then classifying the resulting feature vectors with LDA or SVM, in comparison to using the CNN directly for classification. However, in the cross-database scenario, when using the model trained with the GAIT-IST GEIs to classify GEIs from the DAI2 dataset, the improvement between the two classification methods is very significant, improving the accuracy from 43.3%, when using CNN classification, to 76.7%, with LDA. This result shows that using a classification method trained independently from the feature extraction module improves the generalization capability of the system significantly.

Finally, with respect to the datasets used, it can be observed that, when training and performing cross-validation with the GAIT-IST dataset, the system performs generally better when using the SGEI representation over the GEI. This is evidence that, as speculated before, the SGEI is a better gait representation than the GEI for this classification task, as there are less features in the image connected with the body shape and of each subject, emphasizing the movement component of the sequences, which better describes the different types of abnormal gait being detected. However, the further improved results obtained when performing the fusion of the two representations indicate that both representations have relevant and unique information and other fusion techniques may increase the performance of the system even more.

First fine-tuned VGG-19 block	**GAIT-IST GEI**			**GAIT-IST SGEI**			**GEI+SGEI**
	CNN	*LDA*	*SVM*	*CNN*	*LDA*	*SVM*	*CNN*
Fully connected	**89.4**	89.1	89.4	94.5	**95.6**	95.3	**96.2**
Conv. block 5	**93.3**	88.7	87.4	**96.7**	95.9	95.5	**97.6**
Conv. block 4	**92.7**	89.6	89.4	**97.3**	96.2	96.4	**98.5**
Conv. block 3	**94.5**	89.3	90.4	**97.1**	96.1	95.6	**97.8**
Conv. block 2	**93.7**	89.6	89.5	**97.4**	96.4	96.0	**98.0**

First fine-tuned VGG-19 block	**DAI2 GEI**		
	CNN	*LDA*	*SVM*
Conv. block 3	43.3	**76.7**	68.8

Table 5.8: Summary of the overall accuracy results using the GAIT-IST dataset with the methods performed considering various fine-tuning configurations of the VGG-19 and GEIs or SGEI as the system input and results of the cross-database experiments.

Chapter 6

Conclusions and Future Work

6.1 Achievements

A new gait dataset, GAIT-IST, is acquired, with sequences from 10 subjects simulating four different types of abnormal gait, to be used in gait analysis and classification systems.

A system is implemented with the objective of performing feature extraction and classification on datasets composed of different types of abnormal gait. The system uses the widely used GEI representation and the new proposed representation SGEI, which performs better in the proposed system. The system is also tested using a different dataset, the DAI2, in order to evaluate its generalization capability.

The results are analysed and discussed, in order to explain the classification output of the system. Furthermore, a deep feature visualisation toolkit is used to create images representing attention maps, indicating which parts of the input images are the focus of the CNN for feature extraction. This helps improving our understanding of the functioning of the network.

6.2 Future Work

There are three main propositions for future developments for this work.

Firstly, the importance of having large and representative datasets is of the utmost importance when dealing with classification tasks. The lack of data in the topic of pathological gait leads to problems when testing this kind of system, for instance, overfitting. Thus one important task is the acquisition of bigger datasets representing abnormal gait.

As the results show, the SGEI representation produces better classification accuracies than the GEI. However, preliminary results showed that the combination of the classification scores further improved the results. As such, a more in depth study of fusion techniques can be done, in order to further improve the performance of the system. This could be done at different levels. For instance a new image representation combining both of them or a combination of the features extracted at some point of the CNN.

The last suggestion is the use of a recurrent neural network such as the LSTM to perform the classification task. Although the GEI and SGEI representations already include some temporal information, the

use of an LSTM could further emphasize the time aspect of the gait sequences. Input to the LSTM could be in the form of skeleton coordinates data or feature vectors extracted from the silhouettes composing each gait cycle.

Bibliography

[1] J. Perry. *Gait Analysis: Normal and Pathological Function.* SLACK Incorporated, 1^{st} edition, 1992. ISBN:978-1-55642-192-1.

[2] L. M. Waite, D. A. Grayson, O. Piguet, H. Creasey, H. P. Bennett, and G. A. Broe. Gait slowing as a predictor of incident dementia: 6-year longitudinal data from the sydney older persons study. *Journal of the Neurological Sciences*, 229–230:89–93, Mar. 2005. doi:10.1016/j.jns.2004.11.009.

[3] J. Verghese, R. Lipton, C. Hall, G. Kuslansky, M. Katz, and H. Buschke. Abnormality of gait as a predictor of non-alzheimer's dementia. *New England Journal of Medicine*, 347(22):1761–1768, Nov. 2002. doi:10.1056/NEJMoa020441.

[4] J. Verghese, C. Derby, M. Katz, and R. Lipton. High risk neurological gait syndrome and vascular dementia. *Journal of Neural Transmission (Vienna)*, 114(10):1249–1252, June 2007. doi:10.1007/s00702-007-0762-0.

[5] M. Plotnik, N. Giladi, Y. Balash, C. Peretz, and J. M. Hausdorff. Is freezing of gait in parkinson's disease related to asymmetric motor function? *Annals of Neurology*, 57(5):656–663, May 2005. doi:10.1002/ana.20452.

[6] M. E. Morris, F. Huxham, J. McGinley, K. Dodd, and R. Iansek. The biomechanics and motor control of gait in parkinson disease. *Clinical biomechanics (Bristol, Avon)*, 16(6):459–470, July 2001. doi:10.1016/S0268-0033(01)00035-3.

[7] G. Yogev, M. Plotnik, C. Peretz, N. Giladi, and J. M. Hausdorff. Gait asymmetry in patients with parkinson's disease and elderly fallers: when does the bilateral coordination of gait require attention? *Clinical biomechanics (Bristol, Avon)*, 177(3):336–346, Mar. 2007. doi:10.1007/s00221-006-0676-3.

[8] G. Gehlsen, K. Beekman, N. Assmann, D. Winant, M. Seidle, and A. Carter. Gait characteristics in multiple sclerosis: progressive changes and effects of exercise on parameters. *Arch. Phys. Med. Rehabil.*, 67(8):536–539, Aug. 1986.

[9] C. L. Martin, B. A. Phillips, T. J. Kilpatrick, H. Butzkueven, N. Tubridy, E. McDonald, and M. P. Galea. Gait and balance impairment in early multiple sclerosis in the absence of clinical disability. *Multiple Sclerosis*, 12(5):620–628, Oct. 2006. doi:10.1177/1352458506070658.

[10] L. Comber, R. Galving, and S. Coote. Gait deficits in people with multiple sclerosis: A systematic review and meta-analysis. *Gait Posture*, 51:25–35, Jan. 2017. doi:10.1016/j.gaitpost.2016.09.026.

[11] J. C. Wall and G. I. Turnbull. Gait asymmetries in residual hemiplegia. *Arch. Phys. Med. Rehabil.*, 67(8):550–553, Aug. 1986.

[12] I. A. de Quervain, S. R. Simon, S. Leurgans, W. S. Pease, and D. McAllister. Gait pattern in the early recovery period after stroke. *J. Bone Joint Surg. Am.*, 78(10):1506–1514, Oct. 1996.

[13] R. Camicioli, D. Howieson, B. Oken, G. Sexton, and J. Kaye. Motor slowing precedes cognitive impairment in the oldest old. *Neurology*, 50(5):1496–1498, May 1998.

[14] K. Jahn, A. Zwergal, and R. Schniepp. Gait disturbances in old age. *Dtsch. Arztebl. Int.*, 107(17): 306–316, Apr. 2010. doi:10.3238/arztebl.2010.0306.

[15] J. Afilalo, M. J. Eisenberg, J. F. Morin, H. Bergman, J. Monette, N. Noiseux, L. P. Perrault, K. P. Alexander, Y. Langlois, N. Dendukuri, P. Chamoun, G. Kasparian, S. Robichaud, S. M. Gharacholou, and J. F. Boivin. Gait speed as an incremental predictor of mortality and major morbidity in elderly patients undergoing cardiac surgery. *J. Am. Coll. Cardiol.*, 56(20):1668–1676, Nov. 2010. doi:10.1016/j.jacc.2010.06.039.

[16] C. Kirtley. *Clinical Gait Analysis: Theory and Practice*. Churchill Livingstone, 1^{st} edition, 2006. ISBN:0-4431-0009-8.

[17] P. J. Rowe, C. M. Myles, S. J. Hillmann, and M. E. Hazlewood. Validation of flexible electrogoniometry as a measure of joint kinematics. *Physiotherapy*, 87(9):479–488, Sept. 2001. doi:10.1016/S0031-9406(05)60695-5.

[18] E. Maranesi, F. di Nardo, G. Ghetti, L. Burattini, and S. Fioretti. A goniometer-based method for the assessment of gait parameters. In *2014 IEEE/ASME 10th International Conference on Mechatronic and Embedded Systems and Applications (MESA)*, Senigallia, Italy, Oct. 2014. IEEE. doi:10.1109/MESA.2014.6935539.

[19] G. Domínguez, E. Cardiel, S. Arias, and P. Rogel. A digital goniometer based on encoders for measuring knee-joint position in an orthosis. In *2013 World Congress on Nature and Biologically Inspired Computing*, Fargo, ND, USA, Oct. 2013. IEEE. doi:10.1109/NaBIC.2013.6617835.

[20] M. Donno, E. Palange, F. D. Nicola, G. Bucci, and F. Ciancetta. A new flexible optical fiber goniometer for dynamic angular measurements: Application to human joint movement monitoring. *IEEE Transactions on Instrumentation and Measurement*, 57(8):1614–1620, June 2008. doi:10.1109/TIM.2008.925336.

[21] S. J. Preece, L. P. Keney, T. Dias, E. Lay, and B. T. Fernandes. Automatic identification of gait events using an instrumented sock. *J. Neuroeng. Rehabil.*, 8:32, May 2011. doi:10.1186/1743-0003-8-32.

[22] Qualisys — motion capture systems. *Qualisys.com*, Accessed in January 2019. https://www.qualisys.com/.

[23] C. C. Vu and J. Kim. Human motion recognition by textile sensors based on machine learning algorithms. *Sensors*, 18(9):3109, Sept. 2018. doi:10.3390/s18093109.

[24] A. Tognetti, R. Bartalesi, F. Lorussi, and D. de Rossi. Body segment position reconstruction and posture classification by smart textiles. *Transactions of the Institute of Measurement and Control*, 29(3–4):215–253, Aug. 2007. doi:10.1177/0142331207069487.

[25] F. Lorussi, W. Rocchia, E. Scilingo, A. Tognetti, and D. D. Rossi. Wearable, redundant fabric-based sensor arrays for reconstruction of body segment posture. *IEEE Sensors Journal*, 4(6):807–818, Dec. 2004. doi:10.1109/JSEN.2004.837498.

[26] Y. Qi, C. B. Soh, R. Gunawan, K. S. Low, and R. Thomas. Estimation of spatial-temporal gait parameters using a low-cost ultrasonic motion analysis system. *Sensors*, 14(8):15434–15457, Aug. 2014. doi:10.3390/s140815434.

[27] Motion analysis corporationg. *Motion Analysis*, Accessed in January 2019. https://motionanalysis.com/.

[28] Y. Wahab and N. A. Bakar. Microsystem based portable shoe integrated instrumentation using ultrasonic for gait analysis measurement. In *2011 4th International Conference on Mechatronics (ICOM)*, Kuala Lumpur, Malaysia, July 2011. IEEE. doi:10.1109/ICOM.2011.5937176.

[29] Y. Wahab and N. A. Bakar. Gait analysis measurement for sport application based on ultrasonic system. In *2011 IEEE 15th International Symposium on Consumer Electronics (ISCE)*, Singapore, Singapore, June 2011. IEEE. doi:10.1109/ISCE.2011.5973775.

[30] F. H. Raab, E. B. Blood, T. O. Steiner, and H. R. Jones. Magnetic position and orientation tracking system. *IEEE Transactions on Aerospace and Electronic Systems*, AES-15(5):709–718, Sept. 1979. doi:10.1109/TAES.1979.308860.

[31] A. D. Milne, D. G. Chess, J. A. Johnson, and G. J. King. Accuracy of an electromagnetic tracking device: A study of the optimal operating range and metal interference. *Journal of Biomechanics*, 29(6):791–793, June 1996. doi:10.1016/0021-9290(96)83335-5.

[32] K. Kobayashi, L. Gransberg, E. Knutsson, and P. Nolén. A new system for three-dimensional gait recording using electromagnetic tracking. *Gait & Posture*, 6(1):63–75, Aug. 1997. doi:10.1016/S0966-6362(96)01102-2.

[33] E. S. Grood and W. J. Suntay. A joint coordinate system for the clinical description of three-dimensional motions: Application to the knee. *Journal of Biomechanical Engineering*, 105(2): 136–144, May 1983. doi:10.1115/1.3138397.

[34] P. M. Mills, S. Morrison, D. G. Lloyd, and R. S. Barret. Repeatability of 3d gait kinematics obtained from an electromagnetic tracking system during treadmill locomotion. *Journal of Biomechanics*, 40(7):1504–1511, Aug. 2007. doi:10.1016/j.jbiomech.2006.06.017.

[35] H. Zhao, Z. Wang, S. Qiu, Y. Shen, and J. Wang. Imu-based gait analysis for rehabilitation assessment of patients with gait disorders. In *2017 4th International Conference on Systems and Informatics (ICSAI)*, Hangzhou, China, Nov. 2017. IEEE. 10.1109/ICSAI.2017.8248364.

[36] A. Ferrari, L. Rocchi, J. van den Noort, and J. Harlaar. Toward the use of wearable inertial sensors to train gait in subjects with movement disorders. In *Converging Clinical and Engineering Research on Neurorehabilitation*, volume 1, pages 937–940. Springer, Berlin, Heidelberg, 2013. ISBN:978-3-642-34545-6.

[37] J. K. Lee and E. J. Park. Quasi real-time gait event detection using shank-attached gyroscopes. *Medical & Biological Engineering & Computing*, 49(6):707–712, June 2011. doi:10.1007/s11517-011-0736-0.

[38] Measurement you can trust. *NDI*, Accessed in January 2019. https://www.ndigital.com/.

[39] A. Salarian, H. Russmann, F. J. G. Vingerhoets, C. Dehollain, Y. Blanc, P. R. Burkhard, and K. Aminian. Gait assessment in parkinson's disease: toward an ambulatory system for long-term monitoring. *IEEE Transactions on Biomedical Engineering*, 51(8):1434–1443, Aug. 2004. doi:0.1109/TBME.2004.827933.

[40] Home — bts bioengineering. *BTS Bioengineering*, Accessed in January 2019. https://www.btsbioengineering.com/.

[41] A. Salarian, P. R. Burkhard, F. J. G. Vingerhoets, B. M. Jolles, and K. Aminian. A novel approach to reducing number of sensing units for wearable gait analysis systems. *IEEE Transactions on Biomedical Engineering*, 60(1):72–77, Jan. 2013. doi:10.1109/TBME.2012.2223465.

[42] P. Catalfamo, S. Ghoussayni, and D. Ewins. Gait event detection on level ground and incline walking using a rate gyroscope. *Sensors*, 10(6):5683–5702, June 2010. doi:10.3390/s100605683.

[43] Pressure mapping, force measurement, & tactile sensors. *Tekscan*, Accessed in January 2019. https://www.tekscan.com/.

[44] P. Catalfamo, D. Moser, S. Ghoussayni, and D. Ewins. Detection of gait events using an f-scan in-shoe pressure measurement system. *Gait & Posture*, 28(3):420–426, Oct. 2010. doi:10.1016/j.gaitpost.2008.01.019.

[45] A. Sant'Anna and N. Wickström. A symbol-based approach to gait analysis from acceleration signals: Identification and detection of gait events and a new measure of gait symmetry. *IEEE Transactions on Information Technology in Biomedicine*, 14(5):1180–1187, Sept. 2010. doi:10.1109/TITB.2010.2047402.

[46] A. Sant'Anna, A. Salarian, and N. Wickstrom. A new measure of movement symmetry in early parkinson's disease patients using symbolic processing of inertial sensor data. *IEEE Transactions on Biomedical Engineering*, 58(7):2127–2135, July 2011. doi:10.1109/TBME.2011.2149521.

[47] A. Sant'Anna, N. Wickström, H. Eklund, R. Zügner, and R. Tranberg. Assessment of gait symmetry and gait normality using inertial sensors: In-lab and in-situ evaluation. In J. Gabriel et al., editor, *Biomedical Engineering Systems and Technologies: 5th International Joint Conference, BIOSTEC 2012, Vilamoura, Portugal, February 1-4, 2012, Revised Selected Papers*, pages 239–254. Springer, Berlin, Heidelberg, 2013. doi:10.1007/978-3-642-38256-7_16.

[48] T. Beravs, P. Reberšek, D. Novak, J. Podobnik, and M. Munih. Development and validation of a wearable inertial measurement system for use with lower limb exoskeletons. In *2011 11th IEEE-RAS International Conference on Humanoid Robots*, Bled, Slovenia, Oct. 2011. IEEE. doi:10.1109/Humanoids.2011.6100914.

[49] D. Novak, P. Reberšek, S. M. M. D. Rossi, M. Donati, J. Podobnik, T. Beravs, T. Lenzi, N. Vitiello, M. C. Carrozza, and M. Munih. Automated detection of gait initiation and termination using wearable sensors. *Medical Engineering & Physics*, 35(12):1713–1720, Dec. 2013. doi:10.1016/j.medengphy.2013.07.003.

[50] K. J. O'Donovan, R. Kamnik, D. T. O'Keeffe, and G. M. Lyons. An inertial and magnetic sensor based technique for joint angle measurement. *Journal of Biomechanics*, 40(12):2604–2611, Mar. 2007. doi:10.1016/j.jbiomech.2006.12.010.

[51] F. Riva, M. J. P. Toebes, M. Pijnappels, R. Stagni, and J. H. van Dieën. Estimating fall risk with inertial sensors using gait stability measures that do not require step detection. *Gait & Posture*, 38 (2):170–174, June 2013. doi:10.1016/j.gaitpost.2013.05.002.

[52] J. Bae and M. Tomizuka. A tele-monitoring system for gait rehabilitation with an inertial measurement unit and a shoe-type ground reaction force sensor. *Mechatronics*, 23(6):646–651, Sept. 2013. doi:10.1682/JRRD.2004.07.0076.

[53] A. F. Cordero, H. J. F. M. Koopman, and F. C. T. van der Helm. Use of pressure insoles to calculate the complete ground reaction forces. *Journal of Biomechanics*, 37(9):1427–1432, Sept. 2004. doi:10.1016/j.jbiomech.2003.12.016.

[54] J. Pawin, T. Khaorapapong, and S. Chawalit. Neural-based human's abnormal gait detection using force sensitive resistors. In *The Fourth International Workshop on Advanced Computational Intelligence*, Wuhan, China, Oct. 2011. IEEE. doi:10.1109/IWACI.2011.6160007.

[55] A. M. Howell, T. Kobayashi, H. A. Hayes, K. B. Foreman, and S. J. M. Bamberg. Kinetic gait analysis using a low-cost insole. *IEEE Transactions on Biomedical Engineering*, 60(12):3284–3290, Dec. 2013. doi:10.1109/TBME.2013.2250972.

[56] I. Lackovic, V. Bilas, A. Santic, and V. Nikolic. Measurement of gait parameters from free moving subjects. *Measurement*, 27(2):121–131, Mar. 2000. doi:10.1016/S0263-2241(99)00063-9.

[57] H. H. C. M. Savelberg and A. L. H. de Lange. Assessment of the horizontal, fore-aft component of the ground reaction force from insole pressure patterns by using artificial neural networks. *Clinical Biomechanics*, 14(8):585–592, Oct. 1999. doi:10.1016/S0268-0033(99)00036-4.

[58] W.-C. Wang, W. R. Ledoux, B. J. Sangeorzan, and P. G. Reinhall. A shear and plantar pressure sensor based on fiber-optic bend loss. *The Journal of Rehabilitation Research and Development*, 42 (3):315–325, May 2005. doi:10.1682/JRRD.2004.07.0076.

[59] L. S. Lincoln, S. J. M. Bamberg, E. Parsons, C. Salisbury, and J. Wheeler. An elastomeric insole for 3-axis ground reaction force measurement. In *2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, Rome, Italy, June 2012. IEEE. doi:10.1109/BioRob.2012.6290838.

[60] R. Davis. Clinical gait analysis. *IEEE Engineering in Medicine and Biology Magazine*, 7(3):35–40, Sept. 1988. doi:10.1109/51.7933.

[61] A. Rainoldi, G. Melchiorri, and I. Caruso. Journal of neuroscience methods. *Journal of Neuroscience Methods*, 134(1):37–43, Mar. 2004. doi:10.1016/j.jneumeth.2003.10.014.

[62] R. Bogey, J. Perry, and A. Gitter. An emg-to-force processing approach for determining ankle muscle forces during normal human gait. *IEEE Transactions on Neural Systems and Rehabilitation Engineeringe*, 13(3):302–310, Sept. 2005. doi:10.1109/TNSRE.2005.851768.

[63] D. G. Lloyd and T. F. Besier. An emg-driven musculoskeletal model to estimate muscle forces and knee joint moments in vivo. *Journal of Biomechanics*, 36(6):765–776, June 2003. doi:10.1016/S0021-9290(03)00010-1.

[64] E. C. Wentink, V. G. H. Schut, E. C. Prinsen, J. S. Rietman, and P. H. Veltink. Detection of the onset of gait initiation using kinematic sensors and emg in transfemoral amputees. *Gait & Posture*, 39(1):391–396, Jan. 2014. doi:10.1016/j.gaitpost.2013.08.008.

[65] V. Dietz, W. Zijlstra, T. Prokop, and W. Berger. Leg muscle activation during gait in parkinson's disease: adaptation and interlimb coordination. *Electroencephalography and Clinical Neurophysiology/Electromyography and Motor Control*, 97(6):408–415, Dec. 1995. doi:10.1016/0924-980X(95)00109-X.

[66] V. Dietz, K. L. Leenders, and G. Colombo. Leg muscle activation during gait in parkinson's disease: influence of body unloading. *Electroencephalography and Clinical Neurophysiology/Electromyography and Motor Control*, 105(5):400–405, Oct. 1997. doi:10.1016/S0924-980X(97)00042-8.

[67] C. Frigo and P. Crenna. Multichannel semg in clinical gait analysis: A review and state-of-the-art. *Clinical Biomechanics*, 24(3):236–245, Mar. 2009. doi:10.1016/j.clinbiomech.2008.07.012.

[68] P. Leusmann, C. Mollering, L. Klack, K. Kasugai, M. Ziefle, and B. Rumpe. Your floor knows where you are: Sensing and acquisition of movement data. In *2011 IEEE 12th International Conference on Mobile Data Management*, Lulea, Sweden, June 2011. IEEE. doi:10.1109/MDM.2011.29.

[69] L. Middleton, A. Buss, A. Bazin, and M. Nixon. A floor sensor system for gait recognition. In *Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'05)*, Buffalo, NY, USA, Dec. 2005. IEEE. doi:10.1109/AUTOID.2005.2.

[70] R. Vera-Rodriguez, J. S. Mason, J. Fierrez, and J. Ortega-Garcia. Comparative analysis and fusion of spatiotemporal information for footstep recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(4):823–834, Apr. 2013. doi:10.1109/TPAMI.2012.164.

[71] D. Muramatsu, A. Shiraishi, Y. Makihara, and Y. Yagi. Arbitrary view transformation model for gait person authentication. In *2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, Arlington, VA, USA, Dec. 2012. IEEE. doi:10.1109/BTAS.2012.6374561.

[72] Y. Iwashita, R. Kurazume, and K. Ogawara. Expanding gait identification methods from straight to curved trajectories. In *2013 IEEE Workshop on Applications of Computer Vision (WACV)*, Tampa, FL, USA, Mar. 2013. IEEE. doi:10.1109/WACV.2013.6475018.

[73] T. T. Verlekar, P. L. Correia, and L. D. Soares. View-invariant gait recognition system using a gait energy image decomposition method. *IET Biometrics*, 6(4):299–306, June 2017. doi:10.1049/iet-bmt.2016.0118.

[74] T. Verlekar, P. Correia, and L. Soares. View-invariant gait recognition exploiting spatio-temporal information and a dissimilarity metric. In *2016 International Conference of the Biometrics Special Interest Group (BIOSIG)*, Darmstadt, Germany, Nov. 2016. IEEE. doi:10.1109/BIOSIG.2016.7736937.

[75] Y. Pratheepan, J. V. Condell, and G. Prasad. The use of dynamic and static characteristics of gait for individual identification. In *2009 13th International Machine Vision and Image Processing Conference*, Dublin, Ireland, Nov. 2009. IEEE. doi:10.1109/IMVIP.2009.27.

[76] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li. Support vector regression for multi-view gait recognition based on local motion feature selection. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, USA, June 2010. IEEE. doi:10.1109/CVPR.2010.5540113.

[77] Z. Xue, D. ming, W. Song, B. Wan, and S. Jin. Infrared gait recognition based on wavelet transform and support vector machine. *Pattern Recognition*, 43(8):2904–2910, Aug. 2010. doi:10.1016/j.patcog.2010.03.011.

[78] M. O. Derawi, H. Ali, and F. A. Cheikh. Gait recognition using time-of-flight sensor. In *BIOSIG 2011 - Proceedings of the Special Interest Group on Biometrics and Electronic Signatures*, Darmstadt, Germany, Sept. 2011. BIOSIG.

[79] R. R. Jensen, R. R. Paulsen, and R. Larsen. Analyzing gait using a time-of-flight camera. In *Scandinavian Conference on Image Analysis 2009*, Oslo, Norway, June 2009. Springer, Berlin, Heidelberg. doi:10.1007/978-3-642-02230-2.

[80] J. Geng. Structured-light 3d surface imaging: a tutorial. *Advances in Optics and Photonics*, 3(2): 128–160, Mar. 2011. doi:10.1364/AOP.3.000128.

[81] S. Springer and G. Y. Seligmann. Validity of the kinect for gait assessment: A focused review. *Sensors*, 16(2):194, Feb. 2016. doi:10.3390/s16020194.

[82] A. N. M. Gomatam and S. Sasi. Multimodal gait recognition based on stereo vision and 3d template matching. In *Proceedings of the International Conference on Imaging Science, Systems and Technology*, Las Vegas, Nevada, USA, June 2004. CISST.

[83] H. Liu, Y. Cao, and Z. Wang. Automatic gait recognition from a distance. In *2010 Chinese Control and Decision Conference*, Xuzhou, China, May 2010. IEEE. doi:10.1109/CCDC.2010.5498729.

[84] M. Nieto-Hidalgo, F. J. Ferrández-Pastor, R. J. Valdivieso-Sarabia, J. Mora-Pascual, and J. M. García-Chamizo. A vision based proposal for classification of normal and abnormal gait using rgb camera. *Journal of Biomedical Informatics*, 63:82–89, Oct. 2016. doi:10.1016/j.jbi.2016.08.003.

[85] J. Ortells, M. T. Herrero-Ezquerro, and R. A. Mollineda. Vision-based gait impairment analysis for aided diagnosis. *Medical & Biological Engineering & Computing*, 56(9):1553–1564, Sept. 2018. doi:10.1007/s11517-018-1795-2.

[86] T. Y. Verlekar, L. D. Soares, and P. L. Correia. Automatic classification of gait impairments using a markerless 2d video-based system. *Sensors*, 18(9):2743, Aug. 2018. doi:10.3390/s18092743.

[87] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, Jan. 1979. doi:10.1109/TSMC.1979.4310076.

[88] T.-N. Nguyen, H.-H. Huynh, and J. Meunier. Extracting silhouette-based characteristics for human gait analysis using one camera. In *SoICT '14 Proceedings of the Fifth Symposium on Information and Communication Technology*, Hanoi, Vietnam, Dec. 2014. doi:10.1145/2676585.2676612.

[89] M. A. R. Ahad. *Motion History Images for Action Recognition and Understanding*. Springer, 2013. ISBN:978-1-4471-4730-5.

[90] C. Bauckhage, J. K. Tsotsos, and F. E. Bunn. Automatic detection of abnormal gait. *Image and Vision Computing*, 27(1–2):108–115, Jan. 2009. doi:10.1016/j.imavis.2006.10.004.

[91] E. E. Stone and M. Skubic. Passive in-home measurement of stride-to-stride gait variability comparing vision and kinect sensing. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Boston, MA, USA, Dec. 2011. IEEE. doi:10.1109/IEMBS.2011.6091602.

[92] T. T. Verlekar, L. D. Soares, and P. L. Correia. Gait recognition in the wild using shadow silhouettes. *Image and Vision Computing*, 76:1–13, Aug. 2018. doi:10.1016/j.imavis.2018.05.002.

[93] Y. Iwashita, A. Stoica, and R. Kurazume. Gait identification using shadow biometrics. *Pattern Recognition Lettersg*, 33(16):2148–2155, Dec. 2012. doi:10.1016/j.patrec.2012.07.022.

[94] A. Leu, D. Ristić-Durrant, and A. Gräser. A robust markerless vision-based human gait analysis system. In *2011 6th IEEE International Symposium on Applied Computational Intelligence and Informatics (SACI)*, Timisoara, Romania, May 2011. IEEE. doi:10.1109/SACI.2011.5873039.

[95] D. A. Winter. *Biomechanics and Motor Control of Human Movement*. John Wiley & Sons, 3^{rd} edition, 2005. ISBN:978-0-470-39818-0.

[96] F. Tafazzoli and R. Safabakhsh. Model-based human gait recognition using leg and arm movements. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(8):1237–1246, Dec. 2010. doi:10.1016/j.engappai.2010.07.004.

[97] M. Gabel, R. Gilad-Bachrach, E. Renshaw, and A. Schuster. Full body gait analysis with kinect. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, San Diego, CA, USA, Aug. 2012. IEEE. doi:10.1109/EMBC.2012.6346340.

[98] A. A. Chaaraoui, J. R. Padilla-López, and F. Flórez-Revuelta. Abnormal gait detection with rgb-d devices using joint motion history features. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, Ljubljana, Slovenia, May 2015. IEEE. doi:10.1109/FG.2015.7284881.

[99] R. Z.-L. Hu, A. Hartfiel, J. Tung, A. Fakih, J. Hoey, and P. Poupart. 3d pose tracking of walker users' lower limb with a structured-light camera on a moving platform. In *CVPR 2011 WORKSHOPS*, Colorado Springs, CO, USA, June 2011. IEEE. doi:10.1109/CVPRW.2011.5981732.

[100] Motion capture systems. *VICON*, Accessed in January 2019. https://www.vicon.com/.

[101] Home - xsens 3d motion tracking. *Xsens 3D motion tracking*, Accessed in January 2019. https://www.xsens.com/.

[102] T. Bouwmans. Traditional and recent approaches in background modeling for foreground detection: An overview. *Computer Science Review*, 11-12:31–66, May 2014. doi: 10.1016/j.cosrev.2014.04.001.

[103] N. J. B. McFarlane and C. P. Schofield. Segmentation and tracking of piglets in images. *Machine Vision and Applications*, 8(3):187–193, 1995. doi: 10.1007/BF01215814.

[104] J. Zheng, Y. Wang, N. Nihan, and M. Hallenbeck. Extracting roadway background image: Mode-based approach. *Transportation Research Record: Journal of the Transportation Research Board*, 1944:82–88, 2007. doi: 10.3141/1944-11.

[105] P. W. Power, J. a. Schoonees, P. W. Power, and J. a. Schoonees. Understanding background mixture models for foreground segmentation understanding background mixture models for foreground segmentation. *Image and Vision Computing*, (November):267–271, 2002. doi: 10.1.1.9.5222.

[106] T. Bouwmans, S. Javed, M. Sultana, and S. K. Jung. Deep neural network concepts for background subtraction: A systematic review and comparative evaluation. *CoRR*, (November), 2018. URL `http://arxiv.org/abs/1811.05255`.

[107] J. C. Neves, K. Wysoczanska, and H. Proença. Evaluation of background subtraction algorithms for human visual surveillance. In *2015 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, Kuala Lumpur, Malaysia, Oct. 2015. IEEE.

[108] J. Han and B. Bhanu. Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(2):316–322, 2006. doi: 10.1109/TPAMI.2006.38.

[109] K. Simonyan and A. Zisserman. Two-stream convolutional networks for action recognition in videos. *Advances in neural information processing systems*, 1, June 2014.

[110] S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh. Convolutional pose machines. In *CVPR*, 2016.

[111] T. Simon, H. Joo, I. Matthews, and Y. Sheikh. Hand keypoint detection in single images using multiview bootstrapping. In *CVPR*, 2017.

[112] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *CVPR*, 2017.

[113] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. In *arXiv preprint arXiv:1812.08008*, 2018.

[114] Openpose. *GitHub*, Accessed in October 2019. https://github.com/CMU-Perceptual-Computing-Lab/openpose.

[115] J. R. Yue Luo, Z. Wang, W. Sun, J. Pan, J. Liu, J. Pang, and L. Lin. Lstm pose machines. In *CVPR*, 2018.

[116] T. Pfister, J. Charles, and A. Zisserman. Flowing convnets for human pose estimation in videos. In *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, Dec. 2015. IEEE. doi:10.1109/ICCV.2015.222.

[117] Imagenet large scale visual recognition competition (ilsvrc). *ImageNet*, Accessed in March 2019. http://image-net.org/challenges/LSVRC/.

[118] L. Pratt, J. Mostow, and C. Kamm. Direct Transfer of Learned Information Among Neural Networks. *AAAI-91 Proceedings*, pages 584–589, 1991. URL `www.aaai.org`.

[119] T. T. Verlekar, P. Lobat, and L. D. Soares. Using transfer learning for classification of gait pathologies. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Madrid, Spain, Dec. 2018. IEEE. doi:10.1109/BIBM.2018.8621302.

[120] M. Nieto-Hidalgo and J. M. García-Chamizo. Classification of pathologies using a vision based feature extraction. In *International Conference on Ubiquitous Computing and Ambient Intelligence*, Philadelphia, USA, Nov. 2017. Springer, Cham. doi:10.1007/978-3-319-67585-5_28.

[121] C. Li, X. Min, S. Sun, W. Lin, and Z. Tang. Deepgait: A learning deep convolutional representation for view-invariant gait recognition using joint bayesian. *Appl. Sci.*, 7(3):210, Feb. 2017. doi:10.3390/app7030210.

[122] H. Iwama, M. Okumura, Y. Makihara, and Y. Yagi. The ou-isir gait database comprising the large population dataset and performance evaluation of gait recognition. *IEEE Transactions on Information Forensics and Security*, 7(5):1511–1521, 2012. doi: 10.1109/TIFS.2012.2204253.

[123] K. Shiraga, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi. Geinet: View-invariant gait recognition using a convolutional neural network. In *2016 International Conference on Biometrics (ICB)*, Halmstad, Sweden, June 2016. IEEE. doi:10.1109/ICB.2016.7550060.

[124] S. Ji, W. Xu, M. Yang, and K. Yu. 3d convolutional neural networks for human action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):221–231, Jan. 2013. doi:10.1109/TPAMI.2012.59.

[125] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei. Large-scale video classification with convolutional neural networks. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, Sept. 2014. IEEE. doi:10.1109/CVPR.2014.223.

[126] M. Baccouche, F. Mamalet, C. Wolf, C. Garcia, and A. Baskurt. Sequential deep learning for human action recognition. In *HBU 2011: Human Behavior Understanding*, Amsterdam, The Netherlands, Nov. 2011. Springer, Berlin, Heidelberg. doi:10.1007/978-3-642-25446-8_4.

[127] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri. Learning spatiotemporal features with 3d convolutional networks. In *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, Feb. 2015. IEEE. doi:10.1109/ICCV.2015.510.

[128] T. Wolf, M. Babaee, and G. Rigoll. Multi-view gait recognition using 3d convolutional neural networks. In *2011 IEEE 12th International Conference on Mobile Data Management*, Phoenix, AZ, USA, Sept. 2016. IEEE. doi:10.1109/ICIP.2016.7533144.

[129] S. Yu, D. Tan, and T. Tan. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In *18th International Conference on Pattern Recognition (ICPR'06)*, Hong Kong, China, Aug. 2006. IEEE. doi:10.1109/ICPR.2006.67.

[130] J. Donahue, L. A. Hendricks, M. Rohrbach, S. Venugopalan, S. Guadarrama, K. Saenko, and T. Darrell. Long-term recurrent convolutional networks for visual recognition and description.

IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(4):677–691, Sept. 2017. doi:10.1109/TPAMI.2016.2599174.

[131] A. Zhao, L. Qi, J. Dong, and H. Yu. Dual channel lstm based multi-feature extraction in gait for diagnosis of neurodegenerative diseases. *Knowledge-Based Systems*, 145:91–97, Apr. 2018. doi:10.1016/j.knosys.2018.01.004.

[132] Gait dynamics in neuro-degenerative disease data base. *PhysioNet*, Accessed in May 2019. https://physionet.org/physiobank/database/gaitndd/.

[133] M. Khokhlova, C. Migniot, A. Morozov, O. Sushkova, and A. Dipanda. Normal and pathological gait classification lstm model. *Artificial Intelligence in Medicine*, 94:54–66, Mar. 2019. doi:10.1016/j.artmed.2018.12.007.

[134] Y. Feng, Y. Li, and J. Luo. Learning effective gait features using lstm. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, Cancun, Mexico, Dec. 2016. IEEE. doi:10.1109/ICPR.2016.7899654.

[135] R. Liao, C. Cao, E. B. Garcia, S. Yu, and Y. Huang. Pose-based temporal-spatial network (ptsn) for gait recognition with carrying and clothing variations. In *CCBR 2017: Biometric Recognition*, Beijing, China, Oct. 2017. Springer, Cham. doi:10.1007/978-3-319-69923-3_51.

[136] D. Yoo, N. Kim, S. Park, A. S. Paek, and I. S. Kweon. Pixel-level domain transfer. In *European Conference on Computer Vision*, Amsterdam, The Netherlands, Oct. 2016. Springer, Cham. doi:10.1007/978-3-319-46484-8_31.

[137] S. Yu, E. Liao, W. An, H. Chen, E. B. García, Y. Huang, and N. Poh. Gaitganv2: Invariant gait feature extraction using generative adversarial networks. *Pattern Recognition*, 87:179–189, Mar. 2019. doi:10.1016/j.patcog.2018.10.019.

[138] Y. Wang, C. Song, Y. Huang, Z. Wang, and L. Wang. Learning view invariant gait features with two-stream gan. *Neurocomputing*, 339:245–254, Apr. 2019. doi:10.1016/j.neucom.2019.02.025.

[139] K. Pearson. LIII. On lines and planes of closest fit to systems of points in space . *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901. doi: 10.1080/14786440109462720.

[140] H. Hotelling. Analysis of a complex of statistical variables into Principal Components. Jour. Educ. Psych., 24, 417-441, 498-520. *The Journal of Educational Psychology*, 24:417–441, 1933.

[141] K. Pearson. LIII. On lines and planes of closest fit to systems of points in space . *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901. doi: 10.1080/14786440109462720.

[142] E. Boser, N. Vapnik, I. M. Guyon, and T. B. Laboratories. Training Algorithm Margin for Optimal Classifiers. *Perception*, pages 144–152, 1992.

[143] S. Fisher, Ronald Aylmer. 138: The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, 7:179–188, 1936. URL `https://digital.library.adelaide.edu.au/dspace/handle/2440/15227`.

[144] T. T. Verlekar. Gait Analysis in Unconstrained Environments. 2019.

[145] M. Nieto-Hidalgo, F. J. Ferrández-Pastor, R. J. Valdivieso-Sarabia, J. Mora-Pascual, and J. M. García-Chamizo. Vision based extraction of dynamic gait features focused on feet movement using rgb camera. In *AmIHEALTH 2015: Ambient Intelligence for Health*, pages 155–166, Puerto Varas, Chile, Dec. 2015. Springer.

[146] Gait abnormalities. *Stanford Medicine 25*, Accessed in August 2019. https://stanfordmedicine25.stanford.edu/the25/gait.html.

[147] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[148] Z. Zivkovic and F. Van Der Heijden. Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern Recognition Letters*, 27(7):773–780, 2006. doi: 10.1016/j.patrec.2005.11.005.

[149] Z. Zivkovic. Improved adaptive gaussian mixture model for background subtraction. In *17th International Conference on Pattern Recognition, ICPR 2004, Cambridge, UK, August 23-26, 2004.*, pages 28–31, 2004. doi: 10.1109/ICPR.2004.1333992.

[150] Filling holes in an image using opencv (python / c++). *Learn OpenCV*, Accessed in August 2019. https://www.learnopencv.com/filling-holes-in-an-image-using-opencv-python-c/.

[151] K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. pages 1–14, 2014. URL `http://arxiv.org/abs/1409.1556`.

[152] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-December: 770–778, 2016. doi: 10.1109/CVPR.2016.90.

[153] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Computer Vision and Pattern Recognition (CVPR)*, 2015. URL `http://arxiv.org/abs/1409.4842`.

[154] Train models to classify data using supervised machine learning - matlab. *Mathworks.com*, Accessed in September 2019. https://www.mathworks.com/help/stats/classificationlearner-app.html.

[155] L. Bottou and O. Bousquet. The tradeoffs of large scale learning. *Advances in Neural Information Processing Systems 20 - Proceedings of the 2007 Conference*, 2009.

[156] A. Cauchy. Méthode générale pour la résolution des systemes d'équations simultanées. *Comp. Rend. Sci. Paris 25.1847*, pages 536–538, 1847.

[157] Y. E. Nesterov. A method for solving the convex programming problem with convergence rate $O(1/k^2)$. *Dokl. akad. nauk Sssr*, 269:543–547, 1983.

[158] R. Kotikalapudi and contributors. keras-vis. `https://github.com/raghakot/keras-vis`, 2017.

[159] K. Simonyan, A. Vedaldi, and A. Zisserman. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. pages 1–8, 2013. URL `http://arxiv.org/abs/1312.6034`.

[160] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Proceedings of the IEEE International Conference on Computer Vision*, 2017-October:618–626, 2017. doi: 10.1109/ICCV.2017.74.