# Person Identity in Gait Patterns: A Spatial-Temporal Approach for Non-Invasive Gait Identification in Realistic Environments Using Deep Machine Learning and Computer Vision

Luke Kenneth Topham

A thesis submitted in partial fulfilment of the requirements of Liverpool John Moores University for the degree of Doctor of Philosophy

June 2024

*I dedicate this work to my Dad, who always encouraged my education.*

# Acknowledgements

I would first like to express my sincere gratitude to my advisors, Dr. Wasiq Khan, Professor Abir Hussain, and Professor Dhiya Al-Jumeily, for their unwavering support and encouragement throughout my PhD studies. This project would not be possible without their expert guidance. Special thanks go to the School of Computer Science and Mathematics, Liverpool John Moores University, for supporting this project. In particular, I would like to thank Tricia Waterson for her continuous help and support in navigating the many processes and systems encountered along the journey.

The final mention must go to my family. To my wife, Lisa, my daughter Isla, my son George (from whom we eagerly await an appearance), my sisters, and my parents. They have provided the love and support that ultimately made this research possible. I can now stop 'hiding' in the office.

# Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

# Abstract

Person identification is a substantial problem of vital importance with critical applications in public security and law enforcement, such as missing person identification and crime suspect identification. However, existing works have severe limitations, for example, many approaches rely upon face identification, which can easily be avoided by hiding or disguising the face. Furthermore, these approaches are limited in real-world environments due to the requirement of unobstructed high-quality images, which are not always available in uncontrolled environments.

An alternative method is gait identification, where people are identified by their manner of walking. This is a convenient approach as it may be performed without the knowledge or consent of the subject, it does not require high-quality images, and gait is a difficult biometric to disguise. However, many current approaches are limited and often affected by appearance, recording environment, viewing angle, and occlusion.

This thesis addresses the above issues of gait identification by developing a skeleton-based approach, where bodily keypoints, such as limb joints, are estimated and used to extract gait information. For the first time, a novel method of occluded gait identification is introduced, proposing a composite of state estimation and spatiotemporal deep learning to reconstruct missing keypoints. Next, a viewing angle classifier is implemented to identify the appropriate person identification model for each category of viewing angle. The reported results suggest that this approach provides state-of-the-art performance and is not affected by changes in appearance, viewing angle, or environments.

# Contents

**Supporting Publications** ..... **10**

**List of Acronyms** ..... **12**

**List of Figures** ..... **15**

**List of Tables** ..... **17**

**1. Introduction** ..... **20**

1.1. Motivation ..... 21

1.2. Research Challenges ..... 22

1.3. Aim and Objectives ..... 23

1.4. Major Contributions ..... 25

1.4.1. A Diverse and Multi-Modal Public Gait Dataset ..... 25

1.4.2. Sensor-based Gait Analysis and Person Identification 26

1.4.3. Non-Invasive Spatial-Temporal Gait-based Person Identification ..... 26

1.4.4. Uncertainty Handling in Dynamic Environments ..... 26

1.5. Organisation ..... 27

**2. Literature Review** ..... **28**

2.1. Person Identification ..... 28

2.1.1. Appearance-based Identification ..... 29

2.1.2. Appearance-free Identification ..... 30

2.2. Gait-based Person Identification ..... 30

2.2.1. Gait Datasets ..... 30

2.2.2. Silhouette-based Gait Identification ..... 33

2.2.3. Skeleton-based Gait Identification ..... 38

2.3. Human Pose Estimation ..... 41
2.3.1. Human Pose Estimation Datasets ..... 42
2.3.2. 2D Human Pose Estimation ..... 44
2.3.3. 3D Human Pose Estimation ..... 53
2.4. Handling Uncertainty ..... 59
2.4.1. Overcoming Varying Viewing Angles ..... 59
2.4.2. Overcoming Occlusion ..... 60
2.5. Discussion ..... 61
2.6. Summary ..... 64
**3. Primary Dataset Compilation ..... 66**
3.1. Dataset Collection Methodology ..... 67
3.1.1. Participants ..... 67
3.1.2. Experimental Tasks ..... 70
3.1.3. Video Capture ..... 73
3.1.4. Digital Goniometer Sensor Capture ..... 74
3.1.5. Data Collection ..... 77
3.1.6. Data Processing ..... 77
3.2. Data Records ..... 79
3.3. Technical Validation ..... 82
3.3.1. Sensor Placement and Function ..... 82
3.3.2. Camera Placement ..... 82
3.3.3. Quality Assurance ..... 83
3.4. Summary ..... 83
**4. Sensor-based Gait Identification ..... 85**
4.1. Methodology ..... 86
4.1.1. Data Windowing ..... 87

4.1.2. Feature Extraction ..... 89

4.1.3. Feature Selection and Dimensionality Reduction ..... 91

4.1.4. Machine Learning Classification Models for Gait Identification ..... 93

4.1.5. Two-Stage Output Filtering ..... 97

4.1.6. Experimental Design ..... 98

4.2. Results of the Sensor-based Gait Identification Method ..... 100

4.3. Discussion ..... 106

4.4. Summary ..... 110

**5. Non-Invasive Gait Identification Using Spatiotemporal Biometrics ..... 112**

5.1. Methodology ..... 113

5.1.1. Dataset ..... 113

5.1.2. Gait Identification Method Overview ..... 114

5.1.3. Spatiotemporal Feature Extraction ..... 115

5.1.4. Occlusion State Estimation ..... 116

5.1.5. Kinematics-Informed Deep Angle Detection ..... 118

5.1.6. Gait-based Person Identification Model ..... 120

5.1.7. Experimental Design ..... 121

5.2. Results of Non-Invasive Gait Identification ..... 124

5.2.1. Viewing Angle Detection ..... 124

5.2.2. Gait Identification Experiment ..... 125

5.2.3. Gait Identification with Occlusion Filtering ..... 126

5.2.4. Gait Identification with Alternative Appearance ..... 128

5.2.5. Gait Identification in Realistic Environments ..... 129

5.2.6. Cross-Dataset Validation ..... 131

5.2.7. Validation on the SOTON large Dataset ..... 131

5.2.8. Validation on the GRIDDS Dataset ..... 132

5.3. Discussion ..... 133

5.4. Summary ..... 136

**6. Conclusions and Future Work ..... 138**

6.1. Research Outcomes ..... 138

6.1.1. Diverse and Multi-Modal Gait Dataset ..... 139

6.1.2. Gait Identification using Limb Joint Movement ..... 139

6.1.3. Non-Invasive Gait Identification in Realistic Environments using Deep Learning and Spatiotemporal Biometrics ..... 140

6.2. Future Work Directions ..... 141

**References ..... 142**

# Supporting Publications

Some key aspects, ideas, results, and figures from this thesis have previously appeared in the following publications:

1. Topham, L.K., Khan, W., Al-Jumeily, D., Waraich, A. and Hussain, A.J., 2023. A diverse and multi-modal gait dataset of indoor and outdoor walks acquired using multiple cameras and sensors. Nature Scientific Data, 10(1), pp.320-31. DOI: 10.1038/s41597-023-02161-8.
2. Topham, L.K., Khan, W., Al-Jumeily, D., and Hussain, A.J., 2023. Enabling non-invasive gait identification in realistic environments using deep learning and spatiotemporal biometrics. IEEE Transactions on Information Forensics and Security. [**Under Review**].
3. Topham, L.K., Khan, W., Al-Jumeily, D. and Hussain, A., 2022. Human body pose estimation for gait identification: A comprehensive survey of datasets and models. ACM Computing Surveys, 55(6), pp.1-42. DOI: 10.1145/3533384.
4. Topham, L., Khan, W., Al-Jumeily, D., Waraich, A. and Hussain, A., 2022, August. Gait Identification Using Hip Joint Movement and Deep Machine Learning. In International Conference on Intelligent Computing (ICIC). Cham: Springer International Publishing, pp.220-233. DOI: 0.1007/978-3-031-13832-4_19.
5. Topham, L.K., Khan, W., Al-Jumeily, D., Waraich, A. and Hussain, A.J., 2022. Gait identification using limb joint movement and deep machine learning. IEEE Access, 10, pp.100113-100127. DOI: 10.1109/ACCESS.2022.3207836.

6. Topham, L. and Khan, W., 2022. 360 Degree Gait capture: A diverse and multi-modal gait dataset of indoor and outdoor walks acquired using multiple video cameras and sensors. DOI: 10.1038/s41597-023-02161-8.
7. Khan, W., Topham, L., Alsmadi, H., Al Kafri, A. and Kolivand, H., 2024. Deep face profiler (DeFaP): Towards explicit, non-restrained, non-invasive, facial and gaze comprehension. Expert Systems with Applications, 254, p.124425.
8. Khan, W., Khayam, U., Topham, L. and Hussain, A.J., 2023. Person de-Identification: A Comprehensive Review of Methods, Datasets, Applications, and Ethical Aspects Along-With New Dimensions. IEEE Transactions on Emerging Topics in Computational Intelligence. **[Under Review]**.
9. Badii, A., Khan, A. and Topham, L, 2023. Research Frontiers of Non-Intrusive Kinematics-informed Multi-Person Interpretive Occluded Pose Estimation and Context-aware Dynamic Postural Classification using multi-kinect 3D Skeleton-based Gait Analytics and Stream Rule Learning. Book Chapter. **[Under Review].**

# List of Acronyms

| | |
|---|---|
| **AFDEI** | Accumulated Frame Difference Energy Image |
| **ANN** | Artificial Neural Network |
| **AUC** | Area Under the Curve |
| **AP** | Average Precision |
| **CAGE** | Contrastive Attention-based Gait Encodings |
| **CCTV** | Closed-Circuit Television |
| **CNN** | Convolutional Neural Network |
| **COCO** | Microsoft Common Objects in Context |
| **CPM** | Convolutional Pose Machine |
| **CSI** | Channel State Information |
| **DRM** | Decoupled Regression model |
| **Depth-GEI** | Depth Gait Energy Image |
| **DGHEI** | Depth Gradient Histogram Energy Image |
| **DCNN** | Deep Convolutional Neural Network |
| **DNN** | Deep Neural Network |
| **EEG** | Electroencephalograms |
| **EMG** | Electromyography |
| **FR** | Facial Recognition |
| **FLIC** | Frames Labelled in Cinema |
| **FMT** | Face Matching Tool |
| **FoG** | Freezing of Gait |
| **FPN** | Feature Pyramid Network |
| **GAN** | Generative Adversarial Network |

| | |
|---|---|
| **GCN** | Graph Convolutional Network |
| **GEI** | Gait Energy Image |
| **GEV** | Gait Energy Volume |
| **GRIDDS** | Gait Recognition Image and Depth Dataset |
| **HMM** | Hidden Markov Model |
| **HMOR** | Hierarchical Multi-person Ordinal Relations |
| **HPE** | Human Pose Estimation |
| **IEF** | Iterative Error Feedback |
| **ILP** | Integer Linear Program |
| **IMU** | Inertial Measurement Unit |
| **IoU** | Intersection over Union |
| **KNN** | K-Nearest Neighbour |
| **LSP** | Leeds Sports Pose |
| **LSPE** | Leeds Sports Pose Extended |
| **LSTM** | Long Short-Term Memory |
| **MAD** | Mean Absolute Deviation |
| **mAP** | Mean Average Precision |
| **MLP** | Multilayer Perceptron |
| **MPII** | Max Planck Instituit Informatik |
| **MPJAE** | Mean Per Joint Angular Error |
| **MPJPE** | Mean Per Joint Position Error |
| **MPPE** | Multi-Person Pose Estimation |
| **N-MPJE** | Normalised Mean Per Joint Position Error |
| **NMS** | Non-Maximum Suppression |
| **NN** | Nearest Neighbour |
| **OKS** | Object Keypoint Similarity |
| **ORPM** | Occlusion-Robust Pose-Maps |
| **OVA** | One-Vs-All |
| **OVO** | One-Vs-One |
| **PC** | Principal Component |
| **PCA** | Principal Component Analysis |

| | |
|---|---|
| **PCK** | Percentage Correct Keypoints |
| **PCKh** | Head-Normalised Probability of the Correct Keypoint |
| **PCP** | Percentage Correct Parts |
| **PDJ** | Percentage of Detected Joints |
| **PGOFI** | Parted Gait Optical Flow Image |
| **P-MPJE** | Procrustes analysis Mean Per Joint Position Error |
| **PRM** | Pyramid Residual Module |
| **PTP** | Periodicity-inspired Temporal feature Pyramid aggregator |
| **RFE** | Recursive Feature Elimination |
| **RMB** | Rotating Multi-Beam |
| **RNN** | Recurrent Neural Network |
| **ROM** | Range of Motion |
| **RPN** | Region Proposal Network |
| **SEM** | Standard Error of Mean |
| **SMPL** | Skinned Multi-Person Linear |
| **SPPE** | Single-Person Pose Estimation |
| **SGEI** | Skeleton Gait Energy Image |
| **STD** | Standard Deviation |
| **SURREAL** | Synthetic hUmans foR REAL tasks |
| **SVM** | Support Vector Machine |
| **TGLSTM** | Time-based Graph Long Short-Term Memory |
| **TUM** | Technische Universität München |
| **TUM-GAID** | TUM Gait from Audio, Image, and Depth |
| **UREC** | University Research Ethics Committee |
| **VTM** | View Transformation Model |
| **3DPCK** | 3D Percentage of Correct Keypoints |

# List of Figures

**Figure 1.** Example of a binary silhouette of a human subject extracted from a single gait image. ..... 34
**Figure 2.** An example of a skeleton representation of an image from our dataset was extracted using the OpenPose [74] 2D HPE tool. ..... 38
**Figure 3.** A taxonomy of HPE approaches ..... 42
**Figure 4.** Box plots of dataset participants' age (years), height (cm), and mass (kg) distributions. ..... 70
**Figure 5.** Histograms of participants' age (years), height (cm), and mass (kg) distributions. ..... 70
**Figure 6.** Camera layout and walking paths for all tasks in experiments 1 and 2 (experiment 2 is the Reverse of this Image). ..... 71
**Figure 7.** Camera layout and walking paths for all tasks in experiment 3 ... 71
**Figure 8.** Recorded camera angles of the three experiments (E1, E2, and E3) with the subject at the centre (facing the 180° arrowhead) and the camera at the arrowhead. ..... 72
**Figure 9.** Still image samples of camera angles and backgrounds from the recorded data including (from the top left moving clockwise) indoor experiment 1 90°, indoor experiment 1 45°, indoor experiment 3 0°, and outdoor experiment 3 180°. Images shared with participants' consent. .... 74
**Figure 10.** Digital goniometer sensor placements (forehead, right arm, and right leg). Image shared with participant's consent. ..... 77
**Figure 11.** A comparison of the original RGB frame (left) and the extracted OpenPose [74] HPE skeleton (right). The image was cropped for comparative purposes. Image shared with participant's consent. ..... 79

**Figure 12.** The dataset folder structure per participant contains 3 experiment folders (e.g., Experiment 1), each containing 7 task subfolders (e.g., E1_T1 represents Experiment 1 Task 1). ..... 80
**Figure 13.** Overview of the sequential process of the sensor-based gait identification methodology..... 87
**Figure 14.** Box plots showing the sampled gait durations in seconds calculated from five samples per participant. ..... 88
**Figure 15.** Overview of the gait identification methodology, including input RGB images, skeleton keypoint extraction, occlusion filtering of any missing keypoints or frames, bodily joint angle extraction, sampling, angle group classification (sagittal, oblique, and frontal/rear), LSTM person identification and the resulting output in terms of participant ID number..... 115

# List of Tables

**Table 1.** An overview of existing image-based gait datasets. ..... 32
**Table 2.** A comparison of silhouette-based gait identification works, including the features used, implemented models, and reported performance. ..... 34
**Table 3.** A comparison of skeleton-based gait identification works, including the features Used, implemented models and reported performance. ..... 39
**Table 4.** An overview of existing human pose estimation datasets. ..... 42
**Table 5.** A comparison of 2D SPPE methods and models with their reported performance scores. ..... 48
**Table 6.** A comparison of 2D MPPE methods and models with their reported performance scores. ..... 50
**Table 7.** A comparison of 3D SPPE methods and models with their reported performance scores. ..... 55
**Table 8.** A comparison of 3D MPPE methods and models with their reported performance scores. ..... 57
**Table 9.** Anthropometry information of the dataset participants, including participant number (P_No), gender (M = Male, F = Female), age, height (cm), mass (kg), and ethnicity. ..... 68
**Table 10.** Description of digital goniometer data column and corresponding units. ..... 76
**Table 11.** A summary of the results of the proposed gait identification model with varying window and sample sizes. ..... 89
**Table 12.** Statistical features extracted from the overlapping windowed IMU input variables. ..... 90

**Table 13.** The thirty most important features identified during feature selection. ..... 92
**Table 14.** LSTM model configuration. ..... 94
**Table 15.** Feed Forward ANN Model Configuration. ..... 96
**Table 16.** A comparison of IMU-based gait identification methods. ..... 103
**Table 17.** Results of experiment 1-A, using leg data only (Exp.1-A). ..... 104
**Table 18.** Results of experiment 1-B, using leg data only (top 15 features only) (Exp.1-B). ..... 104
**Table 19.** Results of experiment 2-A, using arm data only (Exp.2-A). ..... 104
**Table 20.** Results of experiment 2-B, using arm data only (top 15 features only) (Exp.2-B). ..... 105
**Table 21.** Results of experiment 3-A, using synchronised leg and arm data (Exp.3-A). ..... 105
**Table 22.** Results of experiment 3-B, using synchronised leg and arm data (top 30 features only) (Exp.3-B). ..... 105
**Table 23.** Description of the four varieties of simulated occlusion implemented during the evaluation of the occlusion state estimation module. ..... 118
**Table 24.** LSTM model configuration for angle classification. ..... 120
**Table 25.** LSTM model configuration for gait-based person identification. ..... 121
**Table 26.** Statistical results of the proposed angle detection model were validated over a purely unseen partition of our dataset. ..... 125
**Table 27.** Results from the gait identification model for each angle group. ..... 126
**Table 28.** Results of gait identification with simulated occlusion, indicating the adverse effect suffered without state estimation. ..... 127
**Table 29.** Results of gait identification with simulated occlusion with state estimation filtering applied to alleviate the effects. ..... 128
**Table 30.** Results of gait identification with subjects wearing alternative clothing. ..... 129

**Table 31.** Results of gait identification using indoor and outdoor recordings. ..... 130
**Table 32.** Results of the STONI-GID gait identification method validated on the SOTON Large dataset. ..... 131
**Table 33.** Reported gait identification results in other works validated on the SOTON Large dataset. ..... 132
**Table 34.** Results of STONI-GID gait identification method validated on the GRIDDS dataset. ..... 133

# Chapter 1

# Introduction

Person identification is a significant problem in the security domain, with important applications such as identity verification, missing person identification and crime suspect identification. However, recent media reports have highlighted significant limitations with existing tools. For example, the media reports that security authorities (i.e., the police) have been using Face Matching Tools (FMT), which provide false positive rates of up to 98% [1]–[3]. Such issues may result in false alarms and the collection of images of innocent people. The performance of FMT tools is limited in dynamic real-world environments as they are affected by pose variation, lighting, facial expression, image resolution, makeup, masks and occlusion [4]. Therefore, a reliable alternative to FMT is required, which can operate in dynamic real-world environments.

Gait analysis measures and assesses the movement patterns that occur during walking and the resulting interpretation [5]. It has been shown to provide insights in healthcare [6], sports [7], and security domains [8], with diverse applications such as physiotherapy [9], diagnosis of neurological diseases (e.g., Parkinson's) [10], sports coaching [11], and criminal identification [12]. Furthermore, gait has been demonstrated to be a unique biometric identifier which is difficult to disguise, indicating its promise as a person identification method.

Considering the known issues regarding FMT, gait identification, whereby a person is identified by their manner of walking, has received significant attention. Gait identification is a convenient method as it may be performed covertly from a distance, without the knowledge or consent of

the subject, and it is a difficult biometric to disguise [13]. Moreover, gait identification may be performed at a lower image resolution than that required by FMT [14], potentially increasing the range of devices that may be used for surveillance; for example, domestic doorbell cameras.

## 1.1. Motivation

Person identification has vital lifesaving applications, such as identification of crime suspects and missing persons. Closed-circuit television (CCTV) is available in abundance to implement such applications. For instance, there are approximately 5.2 million CCTV cameras in the UK alone [15]. However, the vast quantity of footage produced daily makes it impossible to analyse all the footage manually. Therefore, accurate and robust autonomous analysis and detection could significantly increase the amount of footage analysed and help to find more missing people and crime suspects.

Currently, police and security services worldwide are trialling or implementing face recognition applications [16]. However, such applications have been reported to fail minimum ethical and legal standards, resulting in misidentification and false arrests [16]. This is likely due to the known limitations of appearance-based person identification methods. For example, they are affected by changes in appearance (e.g., wearing or removing accessories such as hats, glasses, masks, etc.), lighting conditions, occlusion, and more [17].

Alternatively, gait-based person identification may represent a more robust identification method than facial identification due to reduced dependence on appearance. Therefore, gait identification is more resilient to factors such as appearance change [18]. However, a significant limiting factor in gait-based person identification research is the lack of a comprehensive and diverse image-based dataset [19]. Without such a dataset, it is impossible to train and evaluate gait-based person

identification methods capable of overcoming the challenges of dynamic real-world environments. Therefore, despite the reduced dependence on appearance, current gait identification methods remain affected by common real-world events such as changed appearance, viewing angle, clothing, walking speed, and occlusion [20].

## 1.2. Research Challenges

Person identification is a challenging task due to variations in the appearance of a single person, which may be caused by clothing, accessories, viewing angle, lighting, environment, and background scenery [21]. Moreover, person identification has important applications, such as identity verification, and the identification of crime suspects and missing persons. Therefore, critical requirements for such applications include a non-invasive approach with high performance which is not severely affected by dynamic real-world environments.

However, without a dataset containing diverse subjects and recording conditions, it is impossible to train models to overcome these conditions. Therefore, it is necessary to obtain or collect a dataset containing diverse participants (i.e., age, gender, ethnicity, height, and weight) and recording conditions (i.e., viewing angles, backgrounds, and indoors and outdoors). Furthermore, examples of alternative appearances of subjects and environmental occlusion are necessary to address some of the more complex issues surrounding dynamic real-world environments. However, likely due to difficulties of recording and the resources required, in addition to the sensitive nature of the data, such datasets are complex to obtain and are often not made publicly available.

These key challenges motivate us to propose a skeleton-based method of gait identification in combination with viewing angle detection and state-estimation-based occlusion filtering to overcome the challenges

of dynamic real-world environments. To achieve this, we further propose a diverse and multi-modal gait dataset that contains a variety of conditions that are to be expected in real-world settings.

Below, several research questions are provided that guide the research presented in this thesis:

**Research Question 1 (RQ1).** What diverse aspects can be considered while compiling a primary gait dataset to advance the research frontiers of gait identification?

**Research Question 2 (RQ2).** Can persons be identified using human-understandable spatiotemporal gait features and deep machine learning?

**Research Question 3 (RQ3).** Can gait-based person identification be implemented so that it is not significantly affected by dynamic real-world conditions such as changes in viewing angle, subject appearance, or recording environment?

**Research Question 4 (RQ4).** Is it possible to reduce the effect of occlusion by estimating the position of body parts or joints that are missing?

## 1.3. Aim and Objectives

This thesis aims to address the issue of person identification in dynamic real-world environments. In particular, to increase the identification of subjects under currently unsolved conditions such as changes in appearance, varying viewing angles, differing real-world environments, and occlusion. Below, several objectives are considered to address this aim:

**Objective 1 (Obj1): Gait-based Person Identification.** To investigate and demonstrate the use of gait identification to perform person identification in real-world environments.

**Objective 2 (Obj2): Review of Datasets.** To gain an in-depth understanding of the relevant datasets for implementing gait-based person identification.

**Objective 3 (Obj3): Dataset Collection.** To collect and process a world-class dataset for the problem of gait identification. The dataset must contain sufficient diversity in subjects, subject appearance (e.g., alternative clothing), viewing angles, image backgrounds, and environments (e.g., indoors and outdoors).

**Objective 4 (Obj4): Dataset Processing.** To process the dataset in preparation for the training and evaluation of deep machine learning models, including video cropping and quality assurance.

**Objective 5 (Obj5): Kinematic Gait Signature and Feature Analysis.** To extract bodily keypoints to estimate each subject's skeletal position. Then, to calculate joint angles given the estimated skeleton. Also, analyse and select the features available from both sensor and video data.

**Objective 6 (Obj6): Sensor-based Gait Analysis and Person Identification.** To prototype the implementation of deep machine learning for gait-based person identification using time-series limb segment data obtained via motion sensors.

**Objective 7 (Obj7): Spatial-Temporal Person Identification.** To investigate the implementation of deep machine learning and computer vision methods for gait-based person identification from video data recorded in realistic environments.

**Objective 8 (Obj8): Uncertainty Handling and Gait Dynamics.** To investigate the application of the skeleton-based approach to gait-based person identification on dynamic issues such as changes in appearance, environment, or viewing angle. Also, to assess the application of filtering methods to mitigate the impact of occlusion.

## 1.4. Major Contributions

The research reported in this thesis has contributed to the literature on gait-based person identification. The major contributions are outlined below.

### 1.4.1. A Diverse and Multi-Modal Public Gait Dataset

A new dataset has been collected and made publicly available to overcome the limitations of existing datasets. It contains 65 subjects, recorded from 8 angles in indoor and outdoor real-world environments, and was collected using digital cameras and wearable motion measuring devices. Moreover, the dataset includes other diversities, such as alternative appearances for the subjects (e.g., changes in clothing) and skeleton keypoint labels extracted from the images, thus increasing the possible range of experiments. The dataset contains approximately 3,120 videos, approximately 748,800 labelled image frames, approximately 18,720,000 labels (i.e., 25 bodily keypoints per image frame), and approximately 1,026,480 motion data points. The dataset is publicly available at LJMU OpenData [22], and a dataset descriptor has been published in Nature Scientific Data [23].

### 1.4.2. Sensor-based Gait Analysis and Person Identification

A wearable motion sensor-based prototype has been developed to explore and demonstrate using skeleton-based movement data for gait-based person identification. The sensor data included angular movement from three joints: the hip, the shoulder, and the neck. The results collected suggest that angular movement data from an individual joint or a combination of joints can be used to identify subjects accurately. Furthermore, an extensive range of gait features is analysed, and the most critical features for gait identification are reported. The findings from this contribution were reported here for hip movement [24] and here for various limbs and joints [25].

### 1.4.3. Non-Invasive Spatial-Temporal Gait-based Person Identification

Although sensor data provides a convenient prototyping method, it is not appropriate for real-world use due to its invasive nature, which would require the knowledge and cooperation of subjects. Therefore, a non-invasive computer vision approach to gait-based person identification is presented. Given an input image, bodily keypoints are estimated from which joint angles can be mathematically calculated. Thus, the skeleton approach to gait identification can be performed non-invasively and from a distance. The suggested method demonstrated state-of-the-art performance during cross-dataset validation.

### 1.4.4. Uncertainty Handling in Dynamic Environments

The performance of many existing works is limited in real-world dynamic environments as they are affected by occlusion and changes in viewing

angles, subject appearance, and recording environments. We consider several aspects of handling uncertainty in dynamic environments. Firstly, a viewing angle estimation model classifies the gait sequence into one of three viewing angle classes which have been empirically identified. The viewing angle classifier enables identification models to be trained on specific subsets of viewing angle types, thus reducing the effect of viewing angles on identification performance. Secondly, the non-invasive spatial-temporal gait identification method results demonstrate that changes in the subject's appearance or viewing environment do not significantly affect it. Thirdly, a state estimation-based filtering approach is applied to estimate missing keypoint positions caused by occlusion. Experiments show that the filtering approach significantly restores performance when occlusion occurs.

## 1.5. Organisation

The remainder of this thesis is organised as follows. Chapter 2 provides an overview and review of the literature regarding person identification methods, models, and datasets. Chapter 3 describes the newly collected and publicly available multi-modal gait dataset compiled for this work. It describes the collection methodology, collected subjects, processing and feature extraction. Chapter 4 reports the implementation and experimental results of our newly proposed sensor-based gait identification method. Moreover, we report the analysis of the available gait features and present the most useful and human-understandable features for person identification. Chapter 5 presents a novel, non-invasive spatial-temporal approach to gait-based person identification. Furthermore, we report the implementation and evaluation of a state-estimation-based filtering method to reduce the effect of occlusion on person identification. Finally, Chapter 6 presents the conclusions of the thesis and suggests potential avenues for future work.

# Chapter 2

# Literature Review

This chapter provides an overview of the current state of the art in person identification and gait identification research. Several related topics are described, including human pose estimation, gait identification, and gait datasets. We outline the challenges of the problem of person identification and explore current methods of addressing them. Moreover, we identify several challenges which have not been satisfactorily solved in the literature. Furthermore, we focus on gait-based person identification methods and describe the advantages compared to traditional methods, such as using FMT tools. Also, we provide a significant review and taxonomy of the Human Pose Estimation (HPE) literature, which is necessary to understand the modern skeleton-based approach to gait identification.

## 2.1. Person Identification

This thesis aims to improve the performance of person identification, particularly in dynamic real-world environments, without depending on appearance-based features that often change. An understanding of the existing person identification literature is necessary to facilitate this aim. Therefore, this section describes the current state of the person identification literature. We focus on developing an understanding of the traditional appearance-based methods, such as Face Recognition (FR), and their related limitations. Moreover, we describe alternative person identification methods, namely, gait identification.

As the name suggests, person identification refers to methods of identifying or re-identifying an individual. Identification methods include appearance-based methods (e.g., FR) and appearance-free methods (e.g., fingerprint identification, iris recognition, and gait identification).

### 2.1.1. Appearance-based Identification

FR is the most common method by which humans recognise each other [26]. Furthermore, such appearance-based person identification methods are a common approach to computational solutions to the person identification problem. Moreover, FR is a convenient identification method as it can be performed from a distance without the cooperation or knowledge of the subject [27].

The FR literature may be broadly split into 2D and 3D approaches. 2D FR methods based on deep Convolutional Neural Networks (CNNs) have outperformed traditional methods, such as Support Vector Machines (SVMs), hidden Markov models, random forests, and K-Nearest-Neighbours (KNN) [28]. However, 2D FR methods have been shown to suffer decreased performance under conditions such as varied facial expressions, pose variance and occlusion [28]. Alternatively, 3D FR methods such as [11], [12] are proposed to address these issues. However, advances in 3D FR approaches are limited by a lack of available 3D datasets [29].

Recently, a margin-based loss function, where a margin is added to the SoftMax function as learned features are not sufficiently discriminative, has been applied in several works, such as [30], [31]. Moreover, in [32], an adaptive margin has been used to address the performance issues under varying image quality.

However, current FR methods perform poorly in non-ideal conditions such as variable illumination, posture, and expression [33]. Similarly, FR is limited by changes in appearance, such as ageing, cosmetics, and occlusion [28]. For example, surgical masks commonly worn since the COVID-19

pandemic have been shown to impair the person identification performance of both humans and computer vision applications [34].

### 2.1.2. Appearance-free Identification

Biometric identification methods, such as fingerprint, iris, retina, and voice recognition, have been successfully used for decades [35]. More recently, machine learning methods such as CNNs, SVMs, and Hidden Markov Models (HMMs) have been employed for biometric identification, for example, using finger veins [36], electroencephalograms (EEG) [37], and voice [38].

However, unlike appearance-based approaches, appearance-free identification cannot be performed at a distance without the subject's knowledge. Therefore, such methods are not ideal for non-invasive surveillance applications, which must be performed inconspicuously.

## 2.2. Gait-based Person Identification

Gait-based person identification may offer the advantages of appearance-free and appearance-based approaches. Like appearance-free methods, gait is a robust biometric identifier which cannot be easily disguised [39]. Furthermore, like appearance-based methods, it is a convenient identification method as it can be performed from a distance without the subject's knowledge [40].

### 2.2.1. Gait Datasets

A dataset of relevant gait examples is necessary to train and evaluate machine learning models. Such a dataset must include necessary diversity and realism to provide appropriate real-world conditions. Therefore, a gait dataset should contain sufficient diversity regarding the participants (e.g.,

gender, age, height, weight, and ethnicity) and recording environments (e.g., indoor and outdoor environments, different backgrounds, and various viewing angles). Moreover, it would be advantageous to include alternative appearances for subjects to ensure that any person identification models cannot be easily fooled simply by changing appearance.

Moreover, the literature describes many datasets which are appropriate for gait identification. Table 1 provides an overview of the relevant gait datasets. It describes the number of walking sequences (i.e., walks), the number of steps, the sensors used to acquire the dataset, the level of participant diversity present, whether alternative appearances are available for participants, whether the dataset provides diversity in terms of real-world environments (i.e., indoors and outdoors), and the type of data labelling that is required.

Despite the range of practical datasets, several limitations are apparent. For example, many datasets, such as [41]–[44], contain relatively low numbers of subjects, making it difficult to include sufficient levels of participant diversity (i.e., gender, age, height, weight, and ethnicity). Similarly, as described in Table 1, only [44], [45] and our dataset [23] provide alternative appearances for subjects. Moreover, as shown in Table 1, only [46] and our dataset [23] offer a range of real-world environments, including indoor and outdoor locations. Furthermore, although data labelling can help train machine learning models, only minimal labelling is available; for example, in [41], activities (i.e., standing, sitting, etc.) are labelled, while in [44], [46] silhouettes are extracted, and HPE keypoints are provided in our dataset [23].

*Table 1. An overview of existing image-based gait datasets.*

| Ref | Dataset Name | No. Subjects | No. Walks | No. Steps | Sensors | Participant Diversity | Alternative Appearance | Real-World Environment | No. Camera Angles | Labelling |
|---|---|---|---|---|---|---|---|---|---|---|
| [41] | Gait Silhouette Dataset | 30 | 300 | 100,000 | Camera | Limited | × | × | 1 | Activities |
| [42] | IST Database | 21 | 252 | - | Camera | Not described | × | × | 4 | None |
| [43] | Human Gait Phase Dataset | 21 | - | 25,306 | Motion capture | Limited | × | × | 1 | None |
| [44] | Gait Recognition Image and Depth Dataset (GRIDDS) | 35 | 350 | - | Depth camera, microphone | Limited | ✓ | × | 1 | Silhouette images |
| [46] | SOTON Large | 123 | 984 | ~1,200 | Camera | Not described | × | ✓ | 4 | Silhouette images |
| [45] | Tum Gait from Audio, Image and Depth (TUM-GAID) | 305 | 3,370 | - | Camera, microphone | Limited | ✓ | × | 2 | None |
| [23] | Ours | 64 | 3,120 | - | Camera, digital goniometer | Strong diversity | ✓ | ✓ | 8 | Human Pose Estimation |

In addition to the image-based gait datasets, many datasets are available, composed of data from alternative sensors. For example, the dataset described in [47] comprises data from several wearable Inertial Measurement Units (IMUs) and two Electromyography (EMG) sensors. Moreover, [48] was collected using accelerometers within mobile phones, [49] contains data from force plates, [50] contains only audio, and [51] was collected from sensors within the insole of shoes. However, despite the range of important gait data these sensors provide, they are not appropriate for domains such as security, where it is impossible to make subjects cooperate with such sensors.

Despite the availability of several appropriate image-based gait datasets, the current datasets have significant limitations. Consequently, a new dataset is required to overcome the limitations described. In particular, an image-based gait dataset should include a sufficient number of diverse (i.e., gender, age, height, weight, and ethnicity) participants and a range of real-world environments (i.e., indoor and outdoor environments, different backgrounds, and various viewing angles). Accordingly, we present a novel multi-modal gait dataset described in Chapter 3, Primary Dataset Compilation and previously reported in [23].

### 2.2.2. Silhouette-based Gait Identification

Earlier methods of gait identification used silhouette images of gait. Silhouette images are binary images extracted so that only the outline of the human body is visible in the image, and the background is removed. Many examples of silhouette extraction methods are available [52]–[54]. For example, [55] uses a Canny edge detector [56] to extract the silhouette and then fills the inner contour using closed morphology. An example of a silhouette image is presented in Figure 1.

*Figure 1. Example of a binary silhouette of a human subject extracted from a single gait image.*

Furthermore, silhouette images can be used to build a representation of a subject's gait. For example, a Gait Energy Image (GEI) is created by averaging silhouette images over one complete gait cycle [57]. Similarly, depth cameras have been used to develop Gait Energy Images on Depth Data (depth-GEI) [58]. Further, Depth Gradient Histogram Energy Images (DGHEI) are created by extracting depth gradients at each position and then aggregating them into direction histograms [58]. Alternatively, Gait Energy Volume (GEV) is presented as a 3D version of a GEI, where instead of averaging silhouettes, the averaging is performed on reconstructed voxel volumes [59]. Another example is Accumulated Frame Difference Energy Image (AFDEI), which begins with a silhouette averaging method similar to GEI. However, it uses a weighted average to reflect time characteristics [60].

*Table 2. A comparison of silhouette-based gait identification works, including the features used, implemented models, and reported performance.*

| Ref | Year | Features | Models / Classifiers | Validation Dataset | Identification Accuracy |
|---|---|---|---|---|---|
| [60] | 2015 | GEI, AFDEI | NN | CASIA-B | 88.70% |
| [61] | 2022 | GEI | GAN | CASIA-B | 51.80% |
| [62] | 2016 | GEI | CNN | OU-ISIR | 89.70% |
| [63] | 2015 | GEI | CNN | CASIA-B | 90.00% |
| [64] | 2020 | Depth-GEI | KNN | CASIA-B | 89.80% |
| [64] | 2020 | GEI | KNN | CASIA-B | 91.30% |
| [65] | 2015 | DGHEI | ANN | CMU MoBo | 78.00% |
| [66] | 2023 | PGOFI | CNN | CASIA-B | 90.94% |

Table 2 presents an overview of relevant silhouette-based gait identification methods. The results reported in Table 2 are collected from the cited literature. Therefore, some aspects of the frameworks used, such as the datasets used, may not be consistent between the works. According to Table 2, GEIs present a popular form of gait representation and are implemented in [60]–[64]. Alternatively, Depth-GEI is used in [64], and DGHEI is used in [65]. From these, various classification models are used to identify the subjects. For example, CNNs are used in [62], [63], KNNs are used in [64], Artificial Neural Network (ANN) is used in [65], Nearest Neighbour is used in [60], and Generative Adversarial Network (GAN) is used in [61].

Despite the prevalence and reported success of GEIs, they do not consider the time normalisation of each silhouette [60]. To address this, [60] introduces Accumulated Frame Difference Energy Image (AFDEI), which reflects time characteristics. Moreover, [60] extracts moment invariants from GEIs and AFDEIs as features, which are then classified using a Nearest Neighbour (NN) classifier. It is claimed that [60] can run in real-time. However, computational performance is not reported. Specifically, as this work uses both GEIs and AFDEIs, it is likely more computationally expensive than using GEIs alone. Furthermore, the fusion of GEIs and AFDEIs does not address the issues that affect silhouette-based features, such as appearance, viewing angle, and occlusion.

In contrast, [61] addresses the problem of viewing angle variance, which affects silhouette-based methods. The approach uses a GAN to convert gait images into a set of normal views (i.e., 54°, 90°, and 126°) and performs gait identification using a KNN on the extracted invariant features [61]. The results suggest that using multiple normal views is superior to a single view, as presented in earlier works such as [67]–[69], which are more sensitive to viewing angle and state changes. Despite this, [61] reported only 51.8% accuracy, as shown in Table 2. This is the lowest-performing work validated using the CASIA-B dataset described in Table 2.

Similarly, [62] reportedly addresses the issue of variance in viewing angles. The proposed method uses a CNN to perform gait identification on GEIs. Although [62] reports superior performance compared to alternative view-invariant methods such as [61], it does not report how using CNNs overcomes the view variance issue. Therefore, it is not possible to confirm the cause of the improved performance. Furthermore, the results reported in [62] suggest that the viewing angle may reduce gait identification performance by up to 14.8%. Moreover, only a limited range of angles is evaluated (55° to 75° in 5° increments), making it difficult to assess the success of the view invariance of gait identification.

Alternatively, [63] proposes a feature learning method aiming to prevent overfitting by focusing on the co-occurrences and frequencies of discriminative features. However, it ignores motion between adjacent frames [63]. Gait data is time-series by nature, and it is not clear whether important information is discarded when ignoring the motion between adjacent frames. Despite this, Table 2 indicates that [63] reports superior performance to most works identified when validated on the CASIA-B dataset. However, [63] reports significant variance in identification performance depending on the viewing angle.

Unlike [63], [66] uses optical flow information when dividing body segments for its novel Parted Gait Optical Flow Image (PGOFI), achieving 90.94% accuracy when validated on the CASIA-B dataset. However, in addition to its use of silhouette images, which are affected by appearance, [66] also extracts appearance features for person identification. Therefore, [66] cannot be considered a purely gait-based person identification method.

In contrast to most of the works in Table 2, [64] uses a long-range stereo depth camera. The reported results indicate superior performance when combining subject height with the standard GEI compared to depth GEI alone, as shown in Table 2. However, the reason for the increased performance is unclear as it is logical to assume that the depth-GEI should provide additional gait clues to aid identification. Moreover, [64] reports no experiments combining subject height with depth-GEI. Although the

calculated subject height may be significantly helpful to person identification, this is not purely a gait feature, and it is unclear how this would be extracted in uncontrolled environments and without depth cameras. Furthermore, although the camera used was demonstrated to work well in indoor and outdoor environments, unlike short-range sensors such as the Kinect used in [52], it remains a specialist camera, which is not widely available in outdoor environments.

Similarly, [65] also uses a specialist sensor, a Rotating Multi-Beam (RMB) lidar, which is not widely available, especially in public areas. Despite the use of the specialist sensor, Table 2 indicates that the proposed method underperforms compared to most of the GEI works. Despite this, [65] demonstrated that traditional silhouette-based techniques of gait identification can be efficiently adapted to measurement sequences from an RMB lidar, which may be beneficial in situations where efficiency is more vital than accuracy and where specialist sensors are available.

Overall, the best-performing work validated on the CASIA-B dataset was [64], which achieved an identification accuracy of 91.3% when using a GEI for gait representation and a KNN for classification. However, this method required stereo cameras [64], which are not widely available in public areas, thus limiting the range of potential applications and domains.

In summary, despite the success demonstrated, several limitations are associated with the silhouette-based approach to gait identification. For example, silhouette images can lose fine-grained spatial information, and most works in the literature do not explain how silhouette images are extracted from noisy and complex scenes [70]. Furthermore, they are affected by the clothing and accessories worn by the subject and may contain appearance-based visual clues [71]. Moreover, silhouettes are also influenced by viewing angle and occlusion [72]. Although works such as [61] address this by transforming the viewing angles of GEIs, such methods are not widely used in the silhouette-based literature, likely because the proposed method reported the weakest performance, as shown in Table 2.

### 2.2.3. Skeleton-based Gait Identification

The skeleton-based approach is a relatively recent alternative to the silhouette-based approach to gait identification. This approach is implemented using HPE to extract a skeleton representation. To extract the presentation, a set of bodily keypoints (i.e., joints) are identified in an image; the keypoints are then joined to form a skeleton [73], as shown in Figure 2. As this approach extracts a skeleton representation rather than a silhouette from the original image, unlike the silhouette-based approach, it is unaffected by appearance [74].

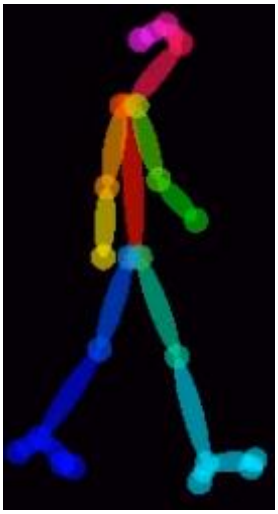

*Figure 2. An example of a skeleton representation of an image from our dataset, extracted using the OpenPose* [75] *2D HPE tool.*

Table 3 presents a comparison of skeleton-based gait identification works. The results reported are collected from the cited literature. Therefore, some aspects of the frameworks used, such as the datasets used, may not be consistent between works. Table 3 shows that various gait features are utilised in the identified works. For example, [70], [76]–[79] take 2D skeleton poses as input. Similarly, [80] takes a skeleton; however, they extract Contrastive Attention-based Gait Encodings (CAGE), which preserve inter-sequence locality. Furthermore, [81] introduces a Periodicity-inspired Temporal feature Pyramid aggregator (PTP) to extract temporal gait features from skeletons. Moreover, [70], [78], [81] then use a Graph Convolutional

Network (GCN) to extract gait features from graph-based models of skeletons; GCNs are based on CNNs but adapted to work with graph structures. Alternatively, [82], [83] use Skeleton Gait Energy Images (GEI), the skeleton equivalent of silhouette GEIs previously described in Section 2.2.2. Silhouette-based Gait Identification

Table 3 also shows that various machine-learning models have been used for feature extraction and classification. For example, Long Short-Term Memory (LSTM) models are used in [77], [80]. LSTM is a particular type of Recurrent Neural Network (RNN), a category of ANNs with a bi-directional flow that creates an internal state akin to memory. Moreover, [77] adapts LSTM, extends LSTM, and introduces Time-based Graph Long Short-Term Memory (TGLSTM). Furthermore, many works such as [70], [78], [79], [81] use graph structures in their representation of skeletons and therefore use GCNs. Alternatively, [76] uses decision trees, a hierarchical non-parametric tree-like model. Although [76] is not the highest-performing model, the decision tree-based method would likely be more interpretable (i.e., more understandable to humans) due to its lower complexity than the other methods [84].

*Table 3. A comparison of skeleton-based gait identification works, including the features Used, implemented models and reported performance.*

| Ref | Year | Features / Extraction | Models / Classifiers | Validation Dataset | Identification Accuracy |
|---|---|---|---|---|---|
| [81] | 2023 | PTP | GCN | CASIA-B | 89.50% |
| [85] | 2022 | 3D Skeleton | LSTM | UPCV1 | 96.30% |
| [79] | 2022 | Skeleton | GCN | CASIA-B | 91.60% |
| [70] | 2021 | Skeleton | GCN | CASIA-B | 87.70% |
| [80] | 2021 | 3D Skeleton | LSTM | CASIA-B | 92.00% |
| [78] | 2021 | Skeleton | GCN | EMOGAIT | 92.20% |
| [82] | 2019 | SGEI | CNN | CASIA-B | 82.30% |
| [76] | 2019 | 3D Skeleton | Decision Tree | PAVIS RGB-D | 92.30% |
| [83] | 2018 | SGEI | CNN | CASIA-B | 92.10% |
| [77] | 2018 | Skeleton | LSTM | TUM-GAID | 98.40% |

As shown in Table 3, [77]is a graph-based version of LSTM to exploit structured data and temporal information. Unlike most other works, instead of HPE, [77] generates an alternative skeleton representation using polygonal approximation on binary silhouettes. Despite the success reported, using a binary silhouette may still be affected by issues such as appearance, viewing angle, and occlusion, as reported in Section 2.2.2. Silhouette-based Gait Identification. Furthermore, the 98.4% reported accuracy was achieved on the TUMGAID [45] dataset, which contains only one viewpoint. Moreover, only 155 subjects were used during testing.

Similarly, [85], also used an LSTM model to classify subjects. However, unlike other works, such as [55], a 3D representation of the skeleton was used rather than a 2D representation. The 3D approach, also used in [80], allows for additional features to be calculated, such as 3D distances between various joints. However, the proposed model has 189,210 parameters [85], which is likely more than required for a 2D approach. Moreover, this may prevent real-time processing, which may be required for applications such as security and surveillance. However, [85] and [80] do not comment on real-time processing. Further experiments comparing the trade-off between the complexity and accuracy of 2D and 3D skeleton representations would be beneficial to guide future works.

In addition, [76] and [80] also use 3D skeleton representations. However, depth cameras are used in both cases to aid the skeleton extraction. Such cameras are not readily available in public areas, which may hinder its adoption. Furthermore, the proposed method also uses 3D representations of the subject's face, which means that the method depends on appearance-based features, which are problematic, as described previously in Section 2.1.1. Appearance-based Identification

Alternatively, [79] reduces the number of keypoint locations tracked (e.g., the head and one arm are not considered). This aims to reduce complexity and prevent overfitting. However, it is not clear if valuable clues are being removed. For example, previous works, such as [86], suggest that information regarding the head may be useful for gait identification. This

may explain why, despite strong performance, numerous works report better performance, as shown in Table 3.

Another approach is used in [82] and [83]; both works use Skeleton Gait Energy Images (SGEI) representation, a method of GEI that averages skeleton rather than binary silhouette images. The works report robustness in person identification regarding clothing variances [82], [83]. However, the reported results showed that for some viewing angles, the reported identification accuracy decreased, in the worst case, to 18.87% [82]. Therefore, although SGEI may mitigate the appearance variance problem when using GEIs, it does not solve the viewing angle problem.

In Summary, Table 3 suggests that skeleton-based methods utilising LSTM perform particularly well on their given validation dataset. In particular, Table 3 shows that [77] and [85]; both use LSTM to classify subjects and achieve identification accuracies above 96% on their individual validation datasets. However, as with the silhouette approach (Table 2), very few works use interpretable or understandable models, which may prevent them from being adopted in critical domains requiring greater transparency [87].

## 2.3. Human Pose Estimation

Before performing skeleton-based gait identification, it is necessary to extract the skeletal representation of the human body from the input image. The extraction of the skeleton is a crucial step as it enables the subsequent analysis of the individual's gait. As such, it is imperative to ensure that the extraction is carried out accurately, with minimum error, to ensure the reliability of the gait identification system.

The skeleton representation is extracted using HPE methods, which extract keypoints (i.e., bodily joints) that can be connected to obtain a skeleton. The traditional taxonomy of HPE, as described in the literature [72], [88]–[90], is presented in Figure 3. HPE can be broadly divided into two

categories: 2D and 3D. In 2D HPE, the 2D coordinates (i.e., x and y) of a keypoint are estimated, and in 3D HPE, the depth of the keypoint is also identified. 2D HPE can subsequently be divided into Single-Person Pose Estimation (SPPE) and Multi-Person Pose Estimation (MPPE). 2D SPPE approaches include regression and body part detection methods. 2D MPPE approaches include the top-down and bottom-up approaches. In comparison, 3D HPE may be divided into single-view and multi-view approaches. Single-view approaches may be further divided into 3D SPPE and 3D MPPE. 3D SPPE methods include the model-free and model-based approaches. Like the 2D equivalent, 3D MPPE methods include the top-down and bottom-up approaches. These approaches will be described in further detail in the following subsections.

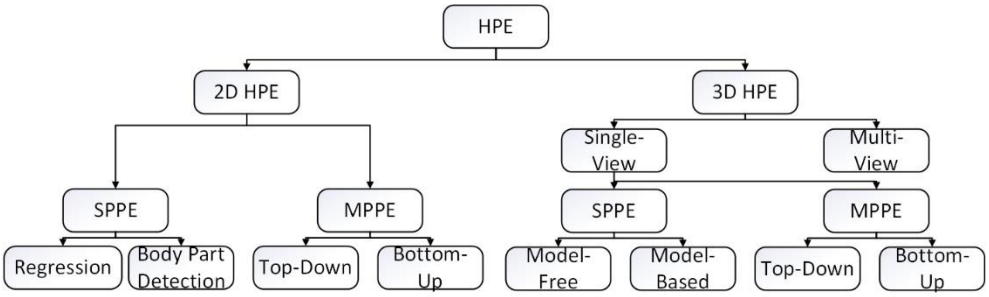


*Figure 3. A taxonomy of HPE approaches.*

## 2.3.1. Human Pose Estimation Datasets

To train HPE models, a dataset of labelled examples is required. For this purpose, a plethora of datasets are available. However, each dataset varies significantly regarding the number of subjects, instances, poses, etc. An overview of important HPE datasets is presented in Table 4.

*Table 4. An overview of existing human pose estimation datasets.*

| Ref | Dataset Name | No. Instances | 2D | 3D | Labelling |
|---|---|---|---|---|---|
| [91] | Common Objects in Context (COCO) | 200,000 | ✓ | × | Keypoints |
| [92] | Human3.6M | 3,600,000 | ✓ | ✓ | Motion Capture |
| [93] | PASCAL-Person-Part | 3,533 | ✓ | × | Keypoints and segmentation |

| Ref | Dataset Name | No. Instances | 2D | 3D | Labelling |
|---|---|---|---|---|---|
| [94] | Synthetic hUmans foR REAL tasks (SURREAL) | 67,582 | × | ✓ | 3D pose, depth maps, and segmentation |
| [95] | HumanEva | 40,000 | ✓ | ✓ | Keypoints (2D and 3D) |
| [96] | MuCo-3DHP | 400,000 | × | ✓ | Markerless motion capture and occlusion labels |
| [96] | MuPoTs-3D | 8,000 | × | ✓ | |
| [97] | PANOPTIC Studio | 297,000 | × | ✓ | Motion Capture |
| [98] | Leeds Sports Pose (LSP) | 2,000 | ✓ | × | Keypoints |
| [99] | Leeds Sports Pose Extended (LSPE) | 10,000 | ✓ | × | Keypoints |
| [100] | Frames Labelled in Cinema (FLIC) | 20,928 | ✓ | × | Keypoints |
| [101] | MPII Human Pose | 40,000 | ✓ | × | Keypoints, 3D Orientation (torso and head only), occlusion labels, and activity labels |
| [102] | Total Capture | 1,892,000 | ✓ | ✓ | Markerless motion capture |
| [103] | Channel State Information (CSI) Image | 5,400,000 | ✓ | ✓ | Keypoints (2D and 3D) |
| [104] | MARker-Less Motion Capture in Outdoor and Indoor Scenes (MPI-MARCOnI) | | ✓ | ✓ | Markerless motion capture |
| [105] | 3D Poses in the Wild (3DPW) | 51,000 | × | ✓ | 3D models |
| [106] | MPI-INF-3DHP | 1,300,000 | ✓ | ✓ | Keypoints (2D and 3D) |
| [107] | Joint Track Auto (JTA) | 500,000 | ✓ | ✓ | Keypoints |
| [108] | Unite the People (UP-3D) | 27,652 | × | ✓ | Keypoints and segmentation |

Data labelling is vital for HPE; the type of labelling provided by the datasets will determine the type of HPE method they can train. From Table

4, most of the datasets, such as [91]–[93], [95], [98], [101], provide 2D keypoint annotations. However, several datasets, such as [92], [95], also provide depth information, which enables the training of both 2D and 3D HPE methods. However, some datasets, such as [96], [97], provide only 3D labelling, usually because they have been captured in a motion-capture laboratory. Similarly, [94] provides 3D pose labelling but not 2D, because the data has been collected from a 3D simulation. Moreover, some datasets provide additional labelling, such as pixel-wise segmentation in [93], [94], depth maps in [94], occlusion labels in [96], [101], and activity labels (i.e., standing, walking, sitting, etc.) in [101]. Despite the additional information provided by the alternative labelling methods, the literature suggests that they have not been significantly adopted in the HPE and gait identification literature.

Despite the range of datasets available, several limitations exist. For example, the use of simulated humans in [94] has not been sufficiently validated against actual human data. Moreover, 3D datasets such as [97] are most often collected in motion capture laboratories due to the requirements of the motion capture equipment (i.e., cameras, sensors, stands/scaffolding, wiring, computers, etc.), which means that the collected images do not contain examples of dynamic real-world conditions and environments.

Furthermore, the existing datasets provide many common poses, such as standing and walking, but relatively few examples of unusual poses (i.e., upside down, yoga poses, etc.). Therefore, training a generalised model capable of performing HPE on various poses may be difficult [90]. While [98] includes some examples of unusual poses by including sports poses, further instances of unusual poses are required.

### 2.3.2. 2D Human Pose Estimation

2D HPE aims to extract 2D coordinates (i.e., x and y), representing the positions of bodily joints (e.g., hip, shoulder, etc.) from an input image or video frame. The coordinates may then be connected to form a skeleton

representation of the pose. For example, a skeleton is shown in Figure 2, extracted from our dataset [23] using the OpenPose 2D HPE tool [75]. In this image, each circle represents an identified keypoint, all keypoints are connected via edges, and each limb or body section is assigned a colour (e.g., dark blue for the left leg and light blue for the right leg).

#### *2.3.2.1. 2D HPE Evaluation Metrics*

Many metrics are available to evaluate the performance of 2D HPE, as previously described in [72]. For instance, Average Precision (AP) involves computing the average precision value for recall values over 0 to 1. In this context, precision pertains to the accuracy of the model's predictions, expressed as a percentage, as in (1. At the same time, recall evaluates the model's capability to identify all positive cases in a given dataset, as in (2. TP, FP, and FN correspond to true positives, false positives, and false negatives, respectively. In summary, AP finds the area under the precision-recall curve defined in (3 where p is precision and r is recall [109]. Precision and recall range from 0 to 1, making AP between 0 and 1. A related metric, Mean Average Precision (mAP), averages the AP computed for each class [109].

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$AP = \int_0^1 p\,(r)\,dr \quad (3)$$

The metric Percentage of Correct Parts (PCP) measures the detection rate of limbs. A limb is considered to be detected if the distance between the predicted and true locations is less than half that of the limb length [110],

[111]. A higher PCP score indicates superior model performance. Nevertheless, PCP is becoming less popular as it disproportionately penalises shorter limb lengths.

Alternatively, the Percentage of Detected Joints (PDJ) and Percentage of Correct Keypoints (PCK) do not penalise shorter limb lengths. In the case of PDJ, a joint is considered to be detected if the distance between the predicted location and the true location is within a specific factor of the torso diameter, defined as the distance between the right hip and the left shoulder [110]. For example, to achieve PDJ@0.1, the predicted and true locations must be within 0.1 * torso diameter. Similarly, PCK also measures the distance between estimated and true joint locations. However, PCK uses the maximum side length (i.e., threshold) of the external rectangle of ground-truth body joints instead of torso diameter [110], [112]. Therefore, a joint is correctly identified if the distance between the predicted and true locations is within a specified threshold factor. A higher PCK indicates superior model performance.

Moreover, a modified version of PCK is provided by the Head-Normalised Probability of the Correct Keypoint (PCKh). When using PCKh, a joint is correctly identified if the distance between the predicted and actual location is within a specified threshold [110], [113]. The threshold should be adapted for each subject based on their size and should be within $\alpha l$ pixels of the true position, where $\alpha$ is a constant and $l$ is the head size that corresponds to 60% of the diagonal length of the ground-truth head bounding box [110].

A further metric, Area Under the Curve (AUC), determines which model makes the best predictions. This is achieved by measuring the different ranges of PCK thresholds (e.g., $\alpha$ variance range of 0 to 0.5) [110] and is commonly approached by plotting true positives against false positives. The closer the AUC is to 1, the better the model performance becomes; a higher AUC is indicative of higher model performance.

Additionally, Intersection over Union (IoU) measures how much the predicted boundary overlaps with the actual boundary, as defined in (4

[109]. Sometimes, a predefined IoU threshold (e.g., 0.5) is used to classify whether a prediction is true or a false positive [109].

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \quad (4)$$

Alternatively, Object Keypoint Similarity (OKS) has a similar role to IoU in object detection but instead measures how close a predicted keypoint is to the actual location. It is calculated from the scale of the subject and the distance between the predicted keypoint location and the actual location [75]. A higher OKS score indicates higher performance and a higher overlap between the predicted and actual locations [110].

#### *2.3.2.2. 2D SPPE*

As mentioned previously, 2D HPE may be divided into SPPE and MPPE. An overview of a range of 2D SPPE methods and models is presented in Table 5, including the dataset used to train and evaluate the model, the SPPE approach, the machine learning model used to implement it, the metric used to assess the performance (as described in Section 2.3.2.1. 2D HPE Evaluation Metrics), and the performance score (in terms of the metric used).

In SPPE, the pose is extracted for only one person in an image. 2D SPPE may then be divided into regression and body part detection approaches. Regression-based approaches seek to employ an end-to-end framework to learn the mapping between input images and body joints automatically [111].

Like many other aspects of HPE, the progress of the regression approaches reported in the literature is accelerated by deep learning approaches, as evident in Table 5. For example, [111] and [112] are implemented using a Deep Neural Network (DNN), a more complex Artificial Neural Network (ANN) usually containing at least two layers. Similarly, [114]–[117] use Convolutional Neural Networks (CNNs), a deep learning

architecture inspired by natural visual perception mechanisms [124]. For instance, DeepPose [111] identifies keypoints in input images using a cascaded deep neural network regressor based on the architecture presented in [125]. A further example, Iterative Error Feedback (IEF) [114], is a self-correcting network based on GoogLeNet [126], which feeds the estimation error back into the input [90]. Alternatively, [116] uses soft-argmax to convert feature maps directly into body joint coordinates, resulting in a fully differentiable framework. Moreover, it indirectly learns heat map representations without generating an artificial ground truth.

*Table 5. A comparison of 2D SPPE methods and models with their reported performance scores.*

| Ref | Year | Dataset | Approach | Model | Metric | Score |
|---|---|---|---|---|---|---|
| [111] | 2016 | MPII, FLIC | Regression | DNN | PCP | 61.00% |
| [112] | 2018 | MPII, LSP | Regression | DNN | PCKh@0.5 | 91.20% |
| [114] | 2016 | MPII, LSP | Regression | CNN | PCP | 81.00% |
| [115] | 2018 | MPII, Human3.6M | Regression | CNN | PCKh@0.5 | 86.40% |
| [116] | 2019 | MPII, LSP | Regression | CNN | PCKh@0.5 | 89.10% |
| [117] | 2019 | MPII, LSP | Regression | CNN | PCKh@0.5 | 91.10% |
| | | | | | AUC | 65.90% |
| [118] | 2020 | MPII, LSP | Body Part Detection | CNN | PCP | 72.80% |
| | | | | | PCK@0.2 | 94.50% |
| | | | | | PCKh@0.5 | 92.70% |
| [119] | 2018 | MPII, LSP | Body Part Detection | CNN RNN | PCK@0.2 | 93.60% |
| [120] | 2018 | COCO | Body Part Detection | CNN | mAP | 76.70% |
| [121] | 2019 | FLIC | Body Part Detection | CNN | PCK@0.2 | 96.40% |
| [122] | 2017 | MPII, LSP | Body Part Detection | CNN, RNN | PCKh@0.5 | 88.10% |
| | | | | | AUC | 58.80% |
| [123] | 2017 | MPII, LSP | Body Part Detection | DCNN | PCKh@0.5 | 92.00% |

In contrast, the body part detection approach aims to train a body part detector to estimate the joint locations in an input image [90]. Typically, this requires two stages. First, all keypoints in an input image are identified. Second, the keypoints are assembled into body poses. Therefore, much of the recent literature, such as [75], [113], [127], [128], considers body part

detection as a heatmap detection problem. Heatmaps provide richer supervision information compared to using keypoint coordinates by preserving spatial location information [90]. Furthermore, heatmaps are ideal for training CNNs, which has led to a growing interest in CNNs for HPE, such as [119], [120], [122], [123], as shown in Table 5.

For instance, [119] uses a multilayer CNN and an LSTM model. This approach utilises temporal information present in a sequence of frames [119]. Similarly, in [122], a CNN-based model learns to represent keypoint appearance and configuration context (i.e., its place in the skeleton) by regressing a heatmap representation for each keypoint. Furthermore, [123] uses a Pyramid Residual Module (PRM) to enhance the scale invariance of Deep Convolutional Neural Networks (DCNNs).

Although Table 5 reports high-performance examples of both regression and body part detection approaches, it is apparent that body part detection methods outperform regression methods. This is likely because regression methods do not exploit the structural information contained in the body pose [115]. However, despite the superior performance, body part detection approaches are prone to estimating false positives [129].

In summary, various 2D SPPE works are compared in Table 5, including both body part detection-based and regression-based approaches. However, the performances reported in Table 5 are provided by the cited literature. Therefore, not all the works were validated using the same datasets, which prevents direct comparisons.

#### *2.3.2.3. 2D MPPE*

However, in MPPE, HPE may be performed for multiple people in an image. 2D MPPE may then be split into top-down and bottom-up approaches. Table 6 compares top-down and bottom-up approaches from the literature. It also includes the datasets used to train and evaluate the works, the model used to implement it, the metric used to assess it, and the performance score in terms of the evaluation metric.

*Table 6. A comparison of 2D MPPE methods and models with their reported performance scores.*

| Ref | Year | Dataset | Approach | Model | Metric | Score |
|---|---|---|---|---|---|---|
| [127] | 2020 | COCO, MPII | Top-Down | CNN | mAP | 79.20% |
| [130] | 2020 | COCO | Top-Down | Encoder-Decoder | mAP | 76.50% |
| [131] | 2020 | COCO, MPII | Top-Down | Encoder-Decoder | mAP | 76.20% |
| [113] | 2018 | COCO, MPII | Top-Down | CNN | mAP | 68.70% |
| [132] | 2017 | COCO, MPII | Top-Down | CNN | mAP | 76.70% |
| [75] | 2021 | COCO, MPII | Bottom-Up | CNN | mAP | 75.60% |
| [133] | 2020 | COCO | Bottom-Up | CNN | mAP | 71.40% |
| [128] | 2019 | COCO | Bottom-Up | CNN | mAP | 42.80% |
| [134] | 2019 | COCO, MPII | Bottom-Up | CNN | mAP<br>PCKh@0.5 | 77.00%<br>92.30% |
| [135] | 2019 | COCO | Bottom-Up | CNN | mAP | 66.70% |
| [136] | 2019 | COCO, MPII, PASCAL | Bottom-Up | CNN | mAP | 78.50% |
| [137] | 2018 | COCO | Bottom-Up | CNN | mAP | 68.70% |
| [138] | 2018 | MPII | Bottom-Up | CNN | mAP<br>PCKh@0.5 | 75.10%<br>85.60% |
| [139] | 2016 | MPII, LSP | Bottom-Up | CNN | PCP<br>PCK<br>mAP | 86.50%<br>63.60%<br>54.10% |
| [140] | 2016 | MPII, LSP | Bottom-Up | CNN | AP | 82.00% |

The top-down approach consists of two stages. First, each person in the image is separated and placed inside bounding boxes. Often, this stage is implemented using standard person detection methods such as Faster R-CNN [141]. Second, HPE is performed on each identified person, as in the 2D SPPE approach. Keypoints are estimated for each identified subject in the input image and then joined to form individual poses [90]. For instance, PoseTrack [113] assembles detected keypoints into a body pose by

surrounding detected people with bounding boxes and groups the maxima of the heatmaps. The computing time for the top-down approach is positively correlated with the number of individuals in the input image.

In contrast to the top-down approach, the bottom-up approach first identifies all keypoints in an image and then groups and connects them to form individual poses [90]. For instance, [139] uses an Integer Linear Program (ILP) to perform graph-based joint parsing. Alternatively, [75] reduces complexity by performing greedy parsing on a tree structure, rather than a graph. Moreover, both [142] and [140] adapt the Residual Network (ResNet) [143] to the sliding-window-based body part detection problem.

Like the SPPE methods described in Table 5, the top-down and bottom-up 2D MPPE methods rely heavily on CNN implementations. Table 6 shows that almost all the MPPE works identified use CNNs. However, [129] and [130] instead, use an encoder-decoder model. This model is made up of an encoder and a decoder. The encoder converts the input into a single-dimension vector, and then the decoder converts the hidden vector into an output [144]. Typically, encoder-decoder models provide faster computation as no features are dilated [145].

A problem related to 2D MPPE is the variability in the scale of individuals in an input image. Both [111] and [146] use multi-scale feature learning strategies to address this issue. For example, in [111], the initial-stage regressor roughly estimates the keypoint position in the input image. Additional regressors then take sub-images to learn the offset between the estimated and the actual location of the keypoints. Similarly, [146] follows a multi-stage process by concatenating heatmaps from previous stages with low-level features to prevent vanishing gradients during training. However, despite reporting a PCK of 90.5%, [146] reports difficulties separating individuals' poses when multiple subjects are close together.

A further issue is overfitting, which can often affect machine learning tasks. This problem occurs when the model fits closely to the training data but does not generalise to new instances. Several works, such as [147] and [148], have addressed this issue by augmenting the training data. For

example, in [148], a GAN is used to augment the training data by scaling, rotating, and occluding the images.

Additionally, MPPE methods are often affected by redundant person detections, which may lead to a false report of improved IoU scores of person detectors. Several works, such as [132], [149]–[151], have addressed this issue by removing redundant boundary boxes. For instance, [149] replaced IoU with OKS in Non-Maximum Suppression (NMS). Similarly, [150] used soft-NMS to remove redundant bounding boxes, which achieved low scores. Alternatively, using Parametric Pose NSM [103] removes invalid keypoints in the post-processing stage.

Several studies, such as [118]–[120], [134], [152], [153], have explored the use of temporal information to increase the performance of HPE. In particular, [152] highlights the use of motion features alone in outperforming some traditional methods of HPE. Similarly, [154] suggests that motion cues such as optical flow may provide valuable clues when extracting keypoints, as rigid points may indicate the location of joints. Moreover, [153] uses optical flow to predict better keypoints between frames where motion cues are available. However, despite the potential benefits of motion cues, two significant obstacles are preventing their adoption. Firstly, motion cues are often extracted as a pre-processing stage and are not performed in real-time. Similarly, using machine learning to extract motion cues, such as optical flow directly, is an open problem. Secondly, consecutive frames are required to extract motion cues. However, many datasets do not provide multiple frames (i.e., video sequences).

In summary, Table 6 reports various 2D MPPE methods, including both top-down and bottom-up approaches. As reported in Table 6, the best-performing top-down approach [127] outperformed the best-performing bottom-up approach [136] when evaluated using the MPII dataset, with reported mAP scores of 79.2% and 78.5%, respectively. However, despite the disadvantage in terms of accuracy, the bottom-up approach has an advantage in terms of processing speed as the input image passes through the model only once, unlike the top-down approach [89].

## 2.3.3. 3D Human Pose Estimation

3D HPE is a more challenging problem than 2D HPE due to the requirement to estimate the depth of keypoints and their 2D location in an input image. Furthermore, an additional challenge relates to the availability and quality of existing 3D HPE datasets. For instance, Table 4 suggests that fewer 3D than 2D HPE datasets are available, likely due to the extra labelling and equipment requirements. Moreover, as 3D HPE datasets require specialist devices such as motion capture systems and depth cameras, capturing data in dynamic real-world environments is more difficult. Therefore, most datasets, such as [92] and [97], are restricted to laboratory environments and others, like [94], instead use synthetic data.

Moreover, as described previously, Figure 3 illustrates the taxonomy of HPE, with the right branch representing the 3D HPE taxonomy [72]. Therefore, 3D HPE can be broadly divided into single-view and multi-view approaches. Next, similar to 2D HPE, the single-view approach can be further divided into SPPE and MPPE. Furthermore, SPPE methods include model-free and model-based approaches, whereas, like 2D MPPE, 3D MPPE includes top-down and bottom-up methods. The aforementioned approaches are described in further detail in the following subsections, including relevant metrics for their evaluation.

#### *2.3.3.1. 3D HPE Evaluation Metrics*

Section 2.3.2.1. 2D HPE Evaluation Metrics introduced some of the metrics available to evaluate 2D HPE. With the exception of AUC, these methods cannot be used to assess 3D HPE due to their 2D measurements. Therefore, additional metrics are required. Nevertheless, some 2D metrics may be extended for 3D HPE; for example, the 3D Percentage of Correct Key Points (3DPCK) extends PCK [106]. Moreover, 3DPCK has been reported to be robust against inaccurate ground truths. However, it cannot indicate the precision of subtle pose details [155].

Alternatively, specific 3D HPE metrics are available, most of which aim to estimate the error between estimated and ground truth joints. For example, Mean Per Joint Angular Error (MPJAE) calculates the average geodesic distance (i.e., the shortest line between two points on a curved surface) between estimated and ground-truth joint orientation [105].

Similarly, Mean Per Joint Position Error (MPJPE) represents the mean of the Euclidean distances between an estimated and ground-truth joint. It is calculated after the estimated and ground-truth poses are aligned, often using the pelvis as the root [156], [157]. A lower MPJPE represents superior performance. Furthermore, two alternative methods of aligning the estimated and ground truth poses are available. Firstly, Normalised Mean Per Joint Position Error (N-MPJPE) aligns the estimated and ground-truth poses in the least-square sense in scale only before computing MPJPE [156]. Secondly, Procrustes analysis Mean Per Joint Position Error (P-MPJPE) aligns the ground truth in the least-square sense in scale, rotation, and translation before computing MPJPE [156].

#### *2.3.3.2. Single-View*

As the name suggests, single-view methods of 3D HPE require only a single image as input. As mentioned previously, similar to 2D HPE, the single-view approach to 3D HPE can be further divided into SPPE and MPPE approaches.

As with the 2D equivalent, 3D SPPE aims to perform HPE for a single subject in an input image. SPPE approaches may be further divided into model-based and model-free approaches. An overview of 3D SPPE methods and models with their reported performance scores for the associated validation dataset is presented in Table 7.

*Table 7. A comparison of 3D SPPE methods and models with their reported performance scores.*

| Ref | Year | Dataset | Approach | Model | Metric | Score |
|---|---|---|---|---|---|---|
| [103] | 2021 | CSI Image | Model-Based | DNN | P-MPJPE | 29.7mm |
| [164] | 2019 | Human3.6m, HumanEva | Model-Based | CNN, DNN | MPJPE<br>P-MPJPE | 42.9mm<br>32.8mm |
| [165] | 2019 | Human3.6M, HumanEva, 3DPW | Model-Based | ResNet | MPJPE | 77.8mm |
| [161] | 2018 | SURREAL, Unite the People | Model-Based | CNN | MPJPE | 49.0mm |
| [166] | 2018 | Up-3D | Model-Based | CNN | MPJPE | 59.9mm |
| [106] | 2018 | Human3.6m, MPI-INF-3DHP | Model-Based | CNN | MPJPE<br>3DPCK<br>AUC | 54.6mm<br>57.30%<br>28.00% |
| [156] | 2018 | Human3.6m | Model-Based | CNN | MPJPE<br>N-MPJPE<br>P-MPJPE | 131.7mm<br>122.6mm<br>98.2mm |
| [157] | 2017 | MPII, MPI-INF-3DHP | Model-Based | CNN | PCK<br>AUC | 69.20%<br>36.90% |
| [167] | 2017 | Human3.6M | Model-Based | CPM | MPJPE | 88.4mm |
| [168] | 2019 | Human3.6M, HumanEva, MPI-INF-3DHP | Model-Free | CNN, ResNet | MPJPE | 39.9mm |
| [169] | 2019 | Human3.6M, MPII, MPI-INF-3DHP | Model-Free | CNN | MPJPE | 52.7mm |
| [170] | 2017 | MPII, Human3.6M, HumanEva, KTH, LSP | Model-Free | CNN | MPJPE | 69.7mm |
| [159] | 2017 | Human3.6M, MPII | Model-Free | CPM | MPJPE | 82.7mm |
| [171] | 2017 | HumanEva, Human3.6M | Model-Free | CPM | MPJPE | 87.3mm |
| [160] | 2017 | HumanEva, Human3.6M, KTH | Model-Free | CNN | MPJPE | 71.9mm |

Sometimes referred to as the generative model approach, the model-based approach to 3D SPPE uses a parametric body model template to estimate a pose [72], [88], [158]. The model-based approach reduces the search space by considering the human body's appearance, structure and motions [129]. Furthermore, model-based works such as [96], [105], [159], [160] use kinematic (i.e., skeleton) models where joint positions and limb orientations represent the human body. In contrast, works such as [161]–[163] use volumetric models for pose reconstruction, and unlike kinematic models, they can represent texture and shape information. Despite this, kinematic models are more widely used as they can be easily represented using graph data structures [105].

Alternatively, model-free approaches automatically learn the mapping between appearance and body pose for faster, precise estimation of specific actions [129]. 3D HPE may be performed directly on 2D images (i.e., without first performing 2D HPE), as reported in [115], [160], [172]. Alternatively, 2D to 3D lifting is applied in [159], [171], where 2D HPE is performed as an intermediary step. 2D to 3D lifting has been reported to outperform the direct estimation approach [90]. For example, conventional background subtraction methods produce poor pose generalisation [129].

Most works described in Table 8 use CNNs or related methods to extract features. For example, in [174], a Feature Pyramid Network (FPN) [180] is used to compute pyramid feature maps at several resolutions and semantic levels. Similarly, [173], [178], [179] use a Region Proposal Network (RPN) [141] to extract full-image convolutional features. For example, [175] introduces RootNet to estimate the camera-centred coordinates of each detected person. Then, they propose PoseNet to estimate the root-relative 3D pose of each person [175]. Alternatively, [173] introduces Hierarchical Multi-person Ordinal Relations (HMOR) to maintain global consistency by encoding interaction information as the ordinal relations of depths and angles.

*Table 8. A comparison of 3D MPPE methods and models with their reported performance scores.*

| Ref | Year | Dataset | Approach | Model | Metric | Score |
|---|---|---|---|---|---|---|
| [173] | 2020 | MuPoTS-3D, CMU Panoptic | Top-Down | ResNet, RPN | 3DPCK<br>MPJPE | 43.80%<br>30.5mm |
| [174] | 2020 | JTA, MuPoTS-3D, CMU Panoptic | Top-Down | CNN | 3DPCK | 83.20% |
| [175] | 2019 | COCO, Human3.6m, MuPoTS-3D | Top-Down | ResNet | 3DPCK | 31.50% |
| [176] | 2018 | Human3.6M, CMU Panoptic | Top-Down | DNN | MPJPE | 48.0mm |
| [105] | 2018 | 3DPW, Total Capture | Top-Down | CNN | MPJPE<br>MPJAE | 26.0mm<br>12.10% |
| [177] | 2020 | MPI-MarCOnI, 3DPW MuCo-3DHP, MPI-INF-3DHP | Bottom-Up | ResNet | MPJPE<br>3DPCK<br>AUC | 98.4mm<br>82.80%<br>45.30% |
| [178] | 2020 | Human3.6M, MPII LSPE, MuPoTS-3D | Bottom-Up | CNN, RPN | 3DPCK | 74.00% |
| [96] | 2018 | MuCo-3DHP, Human3.6M, MuPoTs-3D | Bottom-Up | CNN | 3DPCK | 75.20% |
| [179] | 2017 | Human3.6M, MPII, LSPE | Bottom-UP | CNN, RPN | MPJPE | 72.7mm |

Alternatively, the bottom-up approach begins with estimating joint locations and depth maps. The identified joint locations are then associated with individual poses according to the root depth and relative depth [90]. One of the main limitations of the bottom-up approach is the difficulty of dealing with occlusion of the joint location; therefore, many works directly address this issue. For example, [96] presents Occlusion-Robust Pose-Maps (ORPM), which uses body-part association to infer 3D poses without explicit bounding boxes. Similarly, [181] describes the regression of 2.5D representation body-part representations (i.e., 2D keypoint locations and

root depth map) to reconstruct 3D poses using a depth-aware part association algorithm by reasoning about inter-person occlusion and limb-length constraints [72].

Unlike most works, which require stationary cameras, [105] can recover 3D poses from a moving camera. However, the proposed approach requires multiple IMUs and a mobile phone, which is inappropriate for domains such as security, which do not have the advantage of subject cooperation.

Alternatively, [182] recently proposed a single-stage alternative to the top-down and bottom-up approaches. The proposed method reduces computation costs by removing redundant pipelines using a Decoupled Regression Model (DRM) [182]. The proposed approach outperformed many existing works in terms of error. However, some works outperformed DRM regarding running time [182]. The computation expense may not have been reduced as much as desired, as the approach still requires two models, though they run concurrently.

#### *2.3.3.3. Multi-View*

In contrast to single-view methods, multi-view methods of 3D HPE require multiple input images from various viewing angles. As mentioned previously, 3D HPE is a more challenging problem than 2D HPE due to the need to estimate depth, which may be ambiguous in 2D images. Using multi-view images can reduce the ambiguity of depth due to the availability of multiple perspectives [183]. However, fusing information from multiple images or cameras is a challenging problem [183].

Several methods are available for fusing information from multi-view images [183]. Such as the fusion of multi-view heatmaps [184]–[186], the enforcement of multi-view consistency [156], [187], triangulation [188], [189], and the use of the Skinned Multi-Person Linear (SMPL) model [190].

Like single-view 3D HPE, dataset availability is a limiting factor of multi-view 3D HPE. Relatively few datasets provide synchronised multi-view images. Unlike many datasets, such as [49], Total Capture [102] provides

multi-view images without the use of markers, which is time-consuming and may restrict movement [97]. However, like other 3D datasets, the requirement for multiple cameras restricts the dataset to a controlled indoor laboratory environment.

## 2.4. Handling Uncertainty

As mentioned previously, person identification has many vital applications in the security domain, such as missing persons and crime-suspect identification. Such applications must perform well under dynamic real-world conditions such as varying viewing angles, appearance changes, and occlusion. The problem of appearance change, which most significantly affects silhouette-based approaches, has been addressed in Sections 2.2. Gait-based Person Identification and 2.3. Human Pose Estimation.

### 2.4.1. Overcoming Varying Viewing Angles

Many existing gait datasets provide a limited range of viewing angles, as reported in Table 1. This has previously limited the implementation and evaluation of gait identification methods. For example, many silhouette-based works construct GEIs using only images from the convenient side (90°) [191]. Although GEIs could be produced for any viewing angle, the resulting image provides significantly different information depending on the angle, as demonstrated in [191], making it a problematic approach for overcoming viewing angles. Despite this, GaitSet [192], [193] attempts to overcome the problem of varying viewing angles in silhouette-based gait identification by using gait as a deep set. In the deep set, several frames are integrated by a global-local fused network inspired by left-and right-hemisphere processing in the brain [192], [193]. However, the reported results indicate that although the method outperforms traditional GEI methods, it is still

significantly affected by viewing angle, which can reduce accuracy by up to 10%.

Similarly, varying viewing angles affect skeleton-based gait identification methods. However, several works have attempted to overcome the problem. For example, [194] uses a 3D skeleton to extract gait features without geometric assumptions of the sensor's position to the subject. However, a Kinect depth sensor extracts the skeleton, making the method impractical for real-world applications. Moreover, only a relatively small range of angles was evaluated. Another example [82] uses a View Transformation Model (VTM) to neutralise the effect of viewing angles. Despite this, the reported results indicate that the performance remains significantly affected by viewing angle, and using GEIs may introduce appearance-based bias. Alternatively, [61] uses a GAN to convert a source image into three normal states (54°, 90°, and 126° viewing angle). However, it underperformed in comparison to some VTM methods. Furthermore, another limiting factor for this research aspect is the lack of a comprehensive dataset that provides a significant range of viewing angles, as shown in Table 1.

Despite efforts to reduce it, the problem of viewing angle remains a limiting factor in gait identification. It is particularly problematic for silhouette-based gait identification methods, as silhouette images and related outputs such as GEIs are especially impacted by viewing angles.

### 2.4.2. Overcoming Occlusion

In dynamic real-world environments, occlusion is a common problem that must be overcome to achieve robust gait identification in uncontrolled environments. This may include self-occlusion, where the subject's body blocks other aspects of themselves based on the viewing angle (e.g., the torso blocking the view of one arm or one leg blocking the view of the other) [90]. It may also include environmental occlusion, where other people or objects in the environment wholly or partially occlude the subject.

A unique approach is proposed in [42]; shadows are used to obtain less occluded silhouettes. However, this method has significant issues; for example, it is severely affected by lighting conditions and slopes, and environments can distort or obscure shadows [42]. Alternatively, [195] uses a GAN to reconstruct occluded silhouettes. However, despite claiming state-of-the-art silhouette reconstruction, it is not clear if the reconstruction is accurate enough for gait identification; for example, it is not reported if it is sensitive enough to produce standard features such as GEIs. Similarly, [196] uses time index and temporal context to reconstruct gait frames by exploiting gait periodicity. However, this method relies on silhouettes, which are affected by appearance. In contrast, [197] uses various numerical interpolation techniques to reconstruct gait patterns. However, the method used depth sensors and IMUs, which are unavailable in uncontrolled real-world environments.

A further limiting factor in this research is the lack of availability of occlusion labelling in datasets, as previously reported in [72]. The datasets reported in Table 1 include very few examples of datasets with any occlusion labelling. For this reason, works such as [195] have had to resort to using simulated occlusion, and at present, there is no standardised method of occlusion simulation.

## 2.5. Discussion

Traditionally, the silhouette-based approach to gait identification has received significant attention in the literature. An overview of existing silhouette-based approaches is presented in Table 2. The highest-performing silhouette-based gait identification method was [64] with 91.3%. However, this method used GEI images, which, like other silhouette-based approaches, are sensitive to appearance changes, such as when subjects change their clothes and accessories or if they are carrying items [198].

In contrast, the skeleton-based approach is a more recent approach to gait identification, using skeleton representations extracted using HPE. An overview of skeleton-based gait identification methods is provided in Table 3. The highest-performing method was described in [77], which reported an accuracy of 98.4%.

In comparison, the skeleton-based approach to gait identification is superior for uncontrolled dynamic real-world environments. As described previously, the performance of the silhouette-based approach is affected by appearance and viewing angle. Therefore, the silhouette-based approach is unlikely to meet the requirements for gait identification in the security domain, where such conditions are expected. Furthermore, another limiting factor of gait identification is the availability of a dataset with significant diversity and realism in terms of participants (i.e., age, gender, ethnicity, height, weight, etc.) and environments (i.e., indoors, outdoors, viewing angles, backgrounds, dynamism, etc.).

HPE is a necessary step to implement skeleton-based gait identification and may broadly be divided into 2D and 3D approaches. The performance of 2D HPE has recently been accelerated by using deep learning techniques, as shown in Table 5 and Table 6. Moreover, several deep learning methods indicate high performance for 2D HPE for SPPE [111], [143] and MPPE [75], [132]. Furthermore, numerous studies [118]–[120], [134], [152], [153] have highlighted the use of temporal and spatial information in improving the performance and robustness of HPE. However, it is not always possible to extract temporal information as many datasets do not provide sequences of images.

Moreover, Table 5 and Table 6 reported a range of high-performing 2D HPE methods. Of the 2D SPPE methods, the highest-performing body part detection method validated on the MPII dataset was [118] with a PCKh@0.2 of 94.5%, and the highest-performing regression-based method validated on the MPII dataset was [112] with a PCKKh@0.5 91.2%. Furthermore, of the 2D MPPE methods, the highest-performing top-down method validated on the COCO dataset was [127], with a reported mAP of 79.2%, and the highest

bottom-up approach validated on the COCO dataset was [136], with a reported mAP of 78.5%. Despite the top-down approach having superior performance in terms of accuracy, the bottom-up approach has a processing speed advantage, as each image is processed only once, compared to the need to process each subject individually when using the top-down approach [89].

As described previously, 3D HPE is a more challenging problem than 2D HPE, as the depth for identified keypoints must be estimated in addition to their 2D coordinates. Furthermore, as evident in Table 4, the reliance on motion capture for 3D data labelling restricts most datasets to controlled laboratory environments, and examples of 3D labelling in real-world settings are rare. The datasets have been successfully used to train various 3D HPE methods, as reported in Table 7 and Table 8. The highest-performing model-free 3D SPPE method validated on the Human3.6m dataset was [168], with an MPJPE of 39.9mm. This outperformed the best performing model-based method validated on the Human3.6m dataset which reported an MPJPE of 42.9mm. In addition to the accuracy advantage of model-free methods, they also indicate faster performance [129]. Moreover, model-based HPE methods have been demonstrated to perform poorly on gait sequences [83], thus limiting their potential for gait identification applications. Furthermore, Table 8 indicated that the highest-performing 3D MPPE method validated on the Human3.6m dataset was the top-down method [176], with a reported MPJPE of 48.0mm. This method outperformed [179], the highest-performing bottom-up method validated on the Human3.6m dataset, which reported an MPJPE of 72.7mm. However, bottom-up approaches indicate faster performance as input images are processed by the model at once, whereas top-down approaches process each identified subject separately [72].

In summary, 3D HPE presents a more challenging problem than 2D HPE without offering significant advantages to gait identification. Although multi-view images may reduce the difficulty of depth estimation in 3D HPE, they present issues in fusing information from multiple cameras [183].

Furthermore, in uncontrolled dynamic real-world environments, multiple cameras may not be available, and if they are, they are not guaranteed to capture the same area simultaneously. Therefore, the multi-view approach is likely inappropriate for the security domain, which requires high performance with a single camera input. Recent works, such as [70] and [82], suggest that 2D HPE is sufficient to produce high-performance skeleton-based gait identification models. Moreover, to the best of the author's knowledge, the literature does not report any performance advantages gained by using 3D HPE for skeleton-based gait identification.

To the best of the author's knowledge, no existing gait identification method deals with all three of the most common problems of uncertainty: change of appearance, viewing angle, and occlusion. In particular, reducing the effects of occlusion is an open problem in gait identification. Future gait identification methods must successfully solve these aspects of uncertainty if gait identification is to be performed consistently and accurately in dynamic real-world environments. Furthermore, [197] proposes a promising approach to reconstructing gait information using numerical interpolation methods. However, the method requires specialist equipment such as depth sensors and IMUs. Despite this, the technique may apply to other types of gait data; for example, it may be used for skeletal reconstruction where HPE keypoints are occluded.

## 2.6. Summary

This chapter has summarised the existing literature regarding gait-based approaches to person identification. It has reviewed silhouette and skeleton-based approaches to gait identification. In addition, it has explored the 2D and 3D HPE methods available to extract the skeleton representations necessary for skeleton-based gait identification.

For the reasons stated in Section 2.5. Discussion, the research presented in the remainder of this thesis will follow a skeleton-based approach to gait identification. Moreover, skeleton representations are extracted using single-view 2D HPE. Specifically, we use OpenPose [75] because it is a high performing 2D MPPE method, with associated advantages in processing speed, the ability to label multiple subjects in an image, and can label in real-time. Similarly, LSTM models were found to be particularly powerful for skeleton-based gait identification [77], likely because temporal information has been shown to aid gait identification significantly [81]. Furthermore, as previously described, the existing datasets, as described in Table 1, are limited (Obj2). Therefore, the implementation reported in subsequent sections is trained using our novel diverse gait dataset, as described in [23].

# Chapter 3

# Primary Dataset Compilation

As previously described in Chapters 1 and 2, gait research is currently severely limited by the absence of a diverse, comprehensive, and openly available gait dataset. Such a dataset must include sufficient diversity, including various viewing angles, various real-world backgrounds, and indoor and outdoor environments. Similarly, there should be sufficient diversity in the dataset participants, regarding gender, ethnicity, age, height, mass, and alternative appearances (e.g., change of clothes and accessories). Moreover, comprehensive and consistent data labelling (e.g., HPE, occlusion, etc.) may aid future work.

Currently, the literature is limited in investigating the effect of variables such as age, gender, height, and mass on a person's gait [72]. This is due to the lack of a dataset providing relevant gait data with anthropometric labelling. Similarly, as reported in Table 1, many existing datasets are recorded in controlled laboratory settings that lack many dynamic factors in uncontrolled real-world environments.

We aimed to collect an improved world-class dataset to address the aforementioned limitations related to the existing gait datasets. The collected dataset contains a relatively large number of subjects with diverse characteristics and was recorded from multiple angles in real-world environments. For ethical and safety reasons, only participants over the age of 18 and without disabilities or injuries that would hinder walking were considered for the dataset recordings. The dataset, including RGB video, is publicly available at LJMU OpenData [22], and a dataset descriptor is published with Nature Scientific Data [23].

# 3.1. Dataset Collection Methodology

## 3.1.1. Participants

The dataset includes 65 participants of diverse age, ethnicity, height, and mass. Only participants over 18 years without disabilities or injuries that would affect their gait or pose safety issues were considered. Ethical approval was provided by the University Research Ethics Committee (UREC) at Liverpool John Moores University (LJMU) (Ref: 21/CMP/004). Furthermore, all participants read the participant information sheet explaining the recordings and the use of the acquired data, including a statement explaining that they would be identifiable in the footage. The participants then signed a consent form and completed a questionnaire before completing the recordings.

The participant recruitment process aimed to recruit diverse subjects in terms of age, ethnicity, gender, height, and weight. To achieve this, advertisements were sent internally at LJMU and externally using the author's personal and professional networks. Participant diversity is reported in Table 9, where the complete anthropometry information is contained, as provided by the participant questionnaires and manual height measurements. Missing data cells (denoted "-") correspond to participants who chose to withhold some or all information. Moreover, all participants were sequentially numbered (e.g., participant number 1) to minimise the amount of personally identifiable information recorded and to simplify experiments.

Furthermore, as described in Table 9, of the 65 participants in the dataset that reported their gender, 37 (57.81%) were male, and 27 (42.19%) were female. The age range of 18 to 65 covers most of the range of adult working ages in the UK. The participants' average height (170.83 cm ±10.86) and mass (81.51 kg±20.18) are close to those of the average European. The

standard deviations suggest ample dispersion of values in the dataset. Although most participants (62.5%) recorded their ethnicity as white European, this is lower than the national average of England and Wales (80.5% [199]), suggesting that the dataset is more diverse than the nation in which it was recorded. The remaining 37.5% of participants were Black European, Arab, Asian, or African.

*Table 9. Anthropometry information of the dataset participants, including participant number (P_No), gender (M = Male, F = Female), age, height (cm), mass (kg), and ethnicity.*

| P_No | Gender | Age | Height (cm) | Mass (kg) | Ethnicity |
|---|---|---|---|---|---|
| 1 | F | 27 | 153.0 | 60.0 | White British |
| 2 | F | 32 | 159.0 | 63.0 | White British |
| 3 | M | 30 | 175.0 | 65.0 | Asian |
| 4 | F | 57 | 153.0 | 95.0 | White British |
| 5 | M | 29 | 170.0 | 80.7 | White British |
| 6 | M | 27 | 177.0 | 108.0 | White British |
| 7 | F | 53 | 155.0 | 63.5 | White British |
| 8 | M | 27 | 169.0 | 80.0 | White British |
| 9 | F | 42 | 169.0 | 76.2 | White British |
| 10 | M | 40 | 178.0 | 77.0 | - |
| 11 | M | 37 | 180.0 | 70.0 | Asian |
| 12 | M | 37 | 171.0 | 67.0 | Mixed |
| 13 | F | 32 | 165.0 | 75.0 | British Mixed White -Asian |
| 14 | M | 21 | 185.0 | 95.0 | Asian |
| 15 | M | 19 | 181.0 | 90.0 | Asian |
| 16 | M | 32 | 180.0 | 86.0 | Asian |
| 17 | M | - | - | - | Asian |
| 18 | M | - | - | - | Asian |
| 19 | M | - | - | - | Asian |
| 20 | M | 28 | 183.0 | 70.0 | White British |
| 21 | M | 31 | 185.0 | 117.5 | White British |
| 22 | M | 25 | 183.0 | 79.5 | White British |
| 23 | M | 65 | 165.0 | 82.5 | White British |
| 24 | F | 42 | 152.0 | 133.0 | White British |
| 25 | M | 25 | 188.0 | 77.0 | White British |
| 26 | F | 24 | 162.5 | 57.0 | White British |
| 27 | M | 25 | 175.0 | 67.0 | African |
| 28 | M | 40 | 172.0 | 95.3 | Black British |
| 29 | M | 22 | 172.0 | 63.0 | White British |
| 30 | M | 20 | 173.0 | 74.0 | White British |
| 31 | F | 45 | 160.0 | 69.9 | White British |
| 32 | M | 44 | 167.0 | 79.0 | Black |

| P_No | Gender | Age | Height (cm) | Mass (kg) | Ethnicity |
|---|---|---|---|---|---|
| 33 | M | 34 | 177.0 | 107.0 | Greek |
| 34 | F | 62 | 173.0 | 62.0 | White British |
| 35 | F | 50 | 165.0 | - | White British |
| 36 | M | 23 | 182.0 | 85.7 | White British |
| 37 | F | 52 | 162.5 | 52.0 | White British |
| 38 | M | 20 | 180.0 | 70.0 | Arab |
| 39 | M | 52 | 180.0 | 113.0 | White British |
| 40 | F | 34 | 165.0 | 51.0 | Asian |
| 41 | M | 47 | 190.5 | 88.9 | White British |
| 42 | - | - | - | - | - |
| 43 | M | 21 | 182.0 | 116.0 | Arab |
| 44 | F | 39 | 167.6 | 60.8 | White British |
| 45 | F | 40 | 137.2 | 133.4 | White British |
| 46 | M | 62 | 177.8 | 100.7 | White British |
| 47 | F | 29 | 170.2 | 57.2 | White British |
| 48 | F | 21 | 173.0 | 61.0 | White Dutch |
| 49 | F | 27 | 160.0 | 67.1 | White British |
| 50 | F | 31 | 157.5 | 83.5 | White Irish |
| 51 | F | 39 | 160.0 | 53.0 | White Italian |
| 52 | F | 26 | 152.4 | 74.9 | White British |
| 53 | F | 27 | 157.5 | 99.8 | White British |
| 54 | M | 47 | 177.0 | 72.0 | White Mixed |
| 55 | F | 23 | 157.0 | 56.0 | British Bangladeshi |
| 56 | F | 51 | 165.0 | 92 | White British |
| 57 | M | 61 | 174.0 | 98.5 | White British |
| 58 | F | 18 | 162.5 | 54 | White British |
| 59 | M | 36 | 180.4 | 85 | White British |
| 60 | F | 25 | 162.6 | 98 | White British |
| 61 | F | 26 | 170.2 | 80 | White British |
| 62 | M | 30 | 165.0 | 66.0 | Arab |
| 63 | M | 40 | 180.0 | 81.0 | African |
| 64 | M | 29 | 180.0 | 90.0 | White Polish |
| 65 | M | 26 | 192.0 | 110.0 | Mixed Race |
| **Summary** | **M: 37 (57.81%) F: 27 (42.19%)** | **35.14 (±12.44) Max: 65 Min: 18** | **170.83 (±10.86) Max: 192 Min: 137.16** | **81.51 (±20.18) Max: 133.36 Min: 51** | **White European: 62.5% Other: 37.5%** |

Furthermore, Figure 4 and Figure 5 display the distribution of the participants' ages, heights, and masses in the dataset via box plots and histograms, all suggesting significant diversity. In addition to participant

diversity, various real-world backgrounds are provided in the dataset, including indoor and outdoor settings. Outdoor recordings are provided for participants 14 to 19 inclusively, accounting for approximately 8% of all recordings; the remainder were recorded in indoor environments.

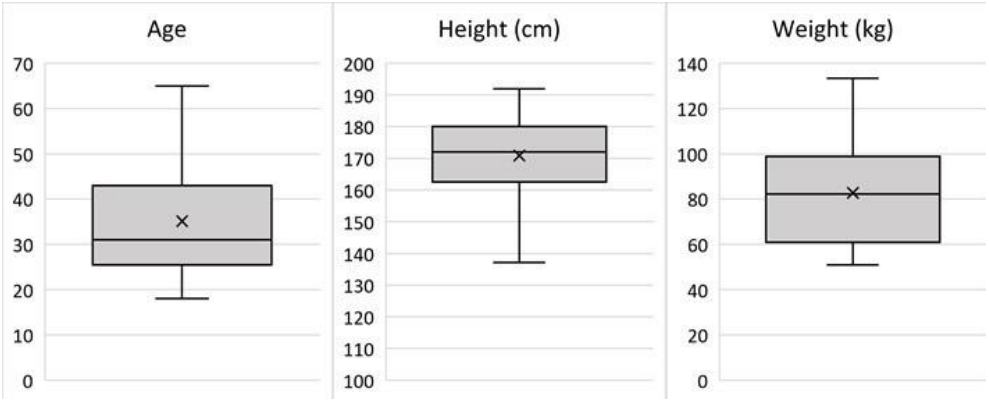


*Figure 4. Box plots of dataset participants' age (years), height (cm), and mass (kg) distributions.*

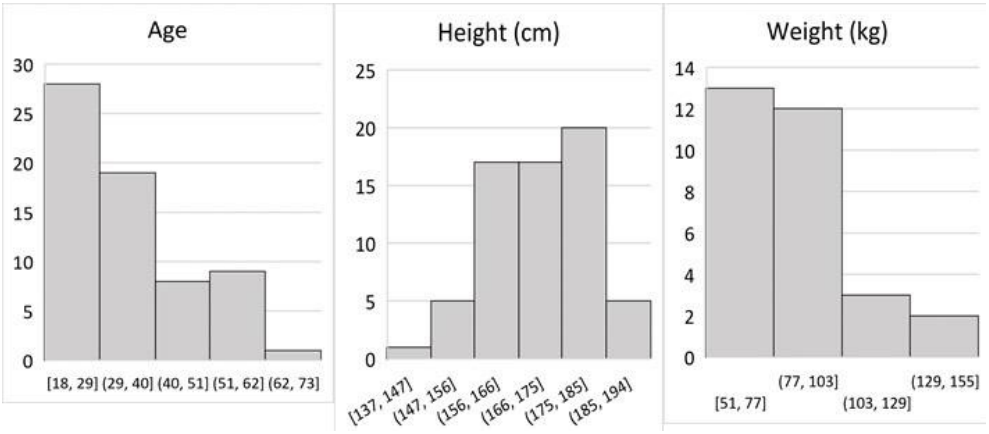


*Figure 5. Histograms of participants' age (years), height (cm), and mass (kg) distributions.*

## 3.1.2. Experimental Tasks

Three groups (henceforth known as experiments) of 6 walking tasks were required of each participant (extended to 8 tasks by repetition of tasks 1 and 2 where alternative clothing is available). Each experiment corresponded to a different camera layout, and each task required the participants to walk for 8 metres. The participants were instructed to walk casually in their usual manner in terms of walking style and speed. To aid the participants, they were asked to imagine taking an un-hurried walk to a shop.

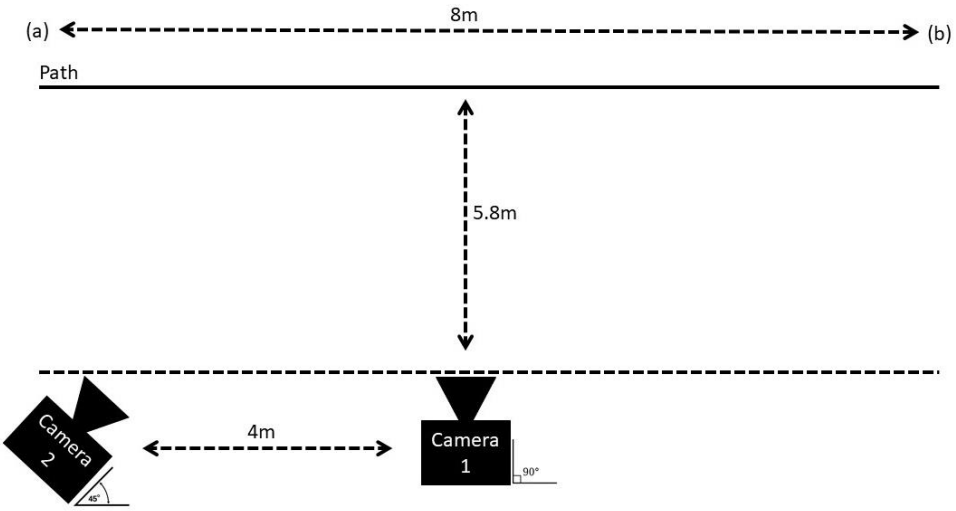


*Figure 6. Camera layout and walking paths for all tasks in experiments 1 and 2 (experiment 2 is the Reverse of this Image).*

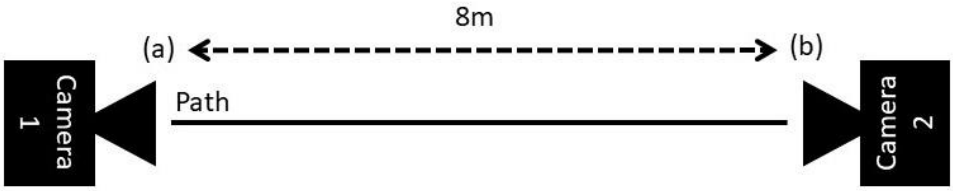


*Figure 7. Camera layout and walking paths for all tasks in experiment 3.*

The complete list of experiments and tasks is as follows:

- Experiment 1 (E1): Camera layout and walking path as in Figure 6.
  - Task 1: Sensor placement on the right leg, walk from a to b.
  - Task 2: Sensor placement on the right leg, walk from b to a.
  - Task 3: Sensor placement on the right arm, walk from a to b.
  - Task 4: Sensor placement on the right arm, walk from b to a.
  - Task 5: Sensor placement on the forehead, walk from a to b.
  - Task 6: Sensor placement on the forehead, walk from b to a.

- Experiment 2 (E2): Camera layout and walking path as in Figure 6 (mirrored).
  - Task 1: Sensor placement on the right leg, walk from a to b.
  - Task 2: Sensor placement on the right leg, walk from b to a.
  - Task 3: Sensor placement on the right arm, walk from a to b.
  - Task 4: Sensor placement on the right arm, walk from b to a.

- Task 5: Sensor placement on the forehead, walk from a to b.
- Task 6: Sensor placement on the forehead, walk from b to a.

- Experiment 3 (E3): Camera layout and walking path as in Figure 7.
  - Task 1: Sensor placement on the right leg, walk from a to b.
  - Task 2: Sensor placement on the right leg, walk from b to a.
  - Task 3: Sensor placement on the right arm, walk from a to b.
  - Task 4: Sensor placement on the right arm, walk from b to a.
  - Task 5: Sensor placement on the forehead, walk from a to b.
  - Task 6: Sensor placement on the forehead, walk from b to a.

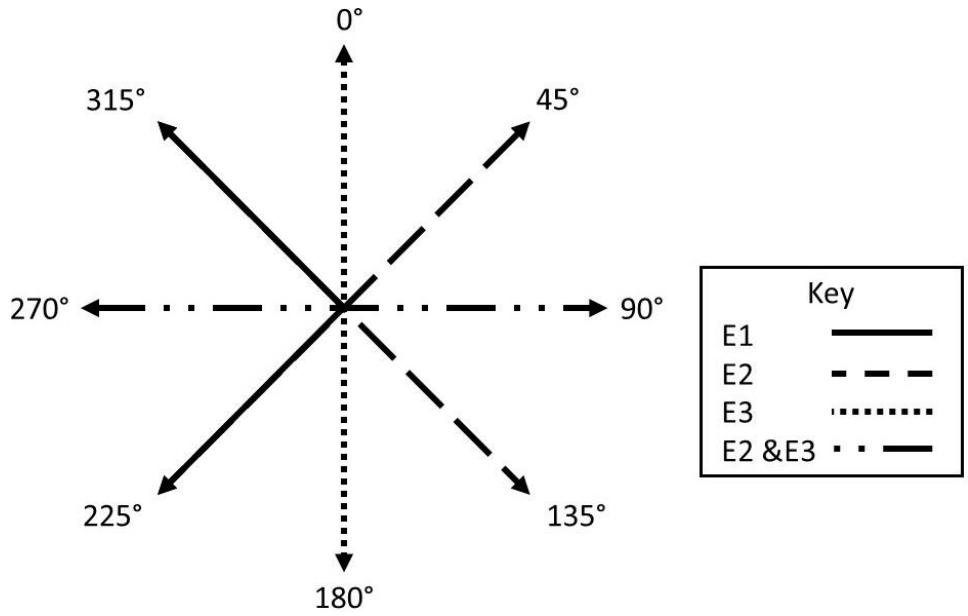


*Figure 8. Recorded camera angles of the three experiments (E1, E2, and E3) with the subject at the centre (facing the 180° arrowhead) and the camera at the arrowhead.*

Moreover, the range of experiments and tasks generated eight different viewing angles, providing a 360° view of each participant's gait in 45° increments, as shown in Figure 8. The figure also shows that each experiment provided two or three viewing angles and that all three experiments needed to be combined to recover a 360° perspective. Furthermore, Experiments 2 and 3 both provided 90° and 270° views (i.e., side profile), the most commonly used angles in the literature [72].

Similarly, the video sequences were divided into the following angle groups for subsequent image-based experiments. Group 1 contained sagittal profile walks, captured from 90° and 270° with respect to the subject. Group 2 included oblique walks captured at 45°, 135°, 225°, and 315°. Group 3 contained frontal and rear views, captured at 0° and 180°, respectively.

### 3.1.3. Video Capture

Two cameras were used to record each walking task, which for the three experiments provided 12 different recording angles for each participant, as the participants alternated the walking direction with each task (i.e., they walked in one direction and then walked from the opposite direction in the following task). The arrangement of the cameras and the variety of walking directions effectively provided a 360° view of each participant's walking in 45° increments.

All tasks for each experiment required the participants to walk along a measured 8-metre path, as displayed in Figure 6 and Figure 7. In experiment 1 (see Figure 6), one camera was placed perpendicular to the walking path at exactly halfway (4 metres) along the path, with a gap of 5.8 metres between the path and the camera. A second camera was placed opposite the starting position of the odd-numbered tasks. The starting position for the participants alternated between each end of the walking path with each task. The alternating starting point provided two benefits: firstly, both sides of the participant's body were recorded. Secondly, camera 2 captured the participant's walk from two different angles across each pair of recordings.

Similarly, experiment 2 may be considered a mirrored or reversed image of experiment 1 (see Figure 6). To achieve this, the camera positions and the walking path for experiment 1, as shown in Figure 6, were swapped. Again, this achieved additional recording angles of each participant's walk. with different backgrounds.

Furthermore, experiment 3 also included an 8-metre walk; however, the cameras were placed directly along the walking path, as shown in Figure 7. This provided a frontal and rear view of each participant's walking. Again, as with the previous experiments, the starting points of each task alternated; therefore, two different backgrounds were captured in this experiment.

*Figure 9. Still image samples of camera angles and backgrounds from the recorded data including (from the top left moving clockwise) indoor experiment 1 90°, indoor experiment 1 45°, indoor experiment 3 0°, and outdoor experiment 3 180°. Images shared with participants' consent.*

Still image examples of the various camera angles (4 samples from the available eight angles) and backgrounds acquired during the recording process are shown in Figure 9. Efforts were made to ensure that the locations and backgrounds used provided real-world environments for the data capture.

### 3.1.4. Digital Goniometer Sensor Capture

In each experiment, the participants also wore a digital goniometer sensor. Physiotherapists often use a goniometer to measure the Range of Motion (ROM) of bodily joints or limb segments [200]. In these experiments, a digital goniometer developed by MOTI [201], built using an Inertial Measurement Unit (IMU), was used. The IMU includes an accelerometer to measure acceleration, a gyroscope to measure angular velocity, and a magnetometer

to measure the Earth's magnetic field [202]. The tri-axial sensor data can then be harnessed to calculate the ROM of the sensor and, by extension, the limb segment responsible for the movement, similar to the methods described in [203], [204]. The sensor was attached to the participant by means of an adjustable strap, which enabled it to be quickly attached and detached. The position of the sensor was alternated with every second task in each experiment to record the thigh, upper arm, and head limb segments. These limb segments were chosen due to their relative height difference from the ground in relation to the participants. This will allow the authors and other researchers in the area of gait identification to prototype methods of gait identification which may be applied to the problem of occlusion (for example, computer vision-based approaches to gait identification where the subject is walking behind a wall or other obstacle).

In tasks 1 and 2, the sensor was attached to the participant's right leg, just above the knee. This position allowed the device to record the movement of the leg at the thigh. In tasks 3 and 4, the sensor was attached to the participant's right arm, just above the elbow, allowing the device to record the movement of the arm at the upper arm. Finally, in tasks 5 and 6, the device was attached to the participant's forehead using a double-sided sticker; this allowed the device to measure the movement of the head from the neck. Figure 10 displays the various sensor placements. All data was recorded in .csv format via an application on a mobile phone. These files were then copied to a desktop computer. A total of 1,560 motion data files were collected, providing approximately 1,026,480 time-stamped samples, with approximately 12,480 gait cycles present. The .csv file generated by the goniometer for each task contains various data from each IMU sensor, as described in Table 10.

*Table 10. Description of digital goniometer data column and corresponding units.*

| Column | Unit | Description |
|---|---|---|
| Timestamp | Seconds (s) | Experiment runtime. |
| q0 | Quaternion | Imaginary number satisfying Hamilton's equation [205]. |
| q1 | Quaternion | Imaginary number satisfying Hamilton's equation [205]. |
| q2 | Quaternion | Imaginary number satisfying Hamilton's equation [205]. |
| q3 | Quaternion | Imaginary number satisfying Hamilton's equation [205]. |
| ROM | Degrees (°) | Limb segment angular motion. |
| Roll | Degrees (°) | Rotation around the longitudinal axis. |
| Pitch | Degrees (°) | Rotation around the transverse axis. |
| Yaw | Degrees (°) | Rotation around the vertical axis. |
| accX | Metres Per Second Squared ($m/s^2$) | The accelerometer provides acceleration on the X-axis. |
| accY | Metres Per Second Squared ($m/s^2$) | The accelerometer provides acceleration in the Y axis. |
| accZ | Metres Per Second Squared ($m/s^2$) | The accelerometer provides acceleration in the Z axis. |
| gyrX | Degrees Per Second (°/s) | Rotation measurement around the X axis as a scalar value. |
| gyrY | Degrees Per Second (°/s) | Rotation measurement around the Y axis as a scalar value. |
| gyrZ | Degrees Per Second (°/s) | Rotation measurement around the Z axis as a scalar value. |
| magX | Millitesla (mT) | Measurement of the magnetic field for the X-axis. |
| magY | Millitesla (mT) | Measurement of the magnetic field for the Y axis. |
| magZ | Millitesla (mT) | Measurement of the magnetic field for the Z axis. |

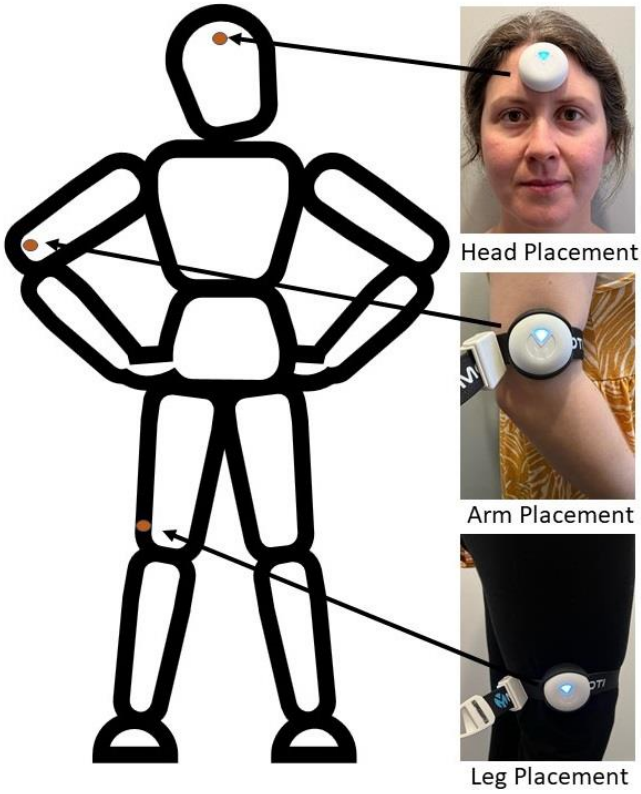


*Figure 10. Digital goniometer sensor placements (forehead, right arm, and right leg). Image shared with participant's consent.*

### 3.1.5. Data Collection

Approximately 415 minutes of video recordings, containing approximately 748,800 video frames (excluding the extracted skeleton videos), were collected across both cameras (i.e., ~207.5 minutes per camera). This corresponds to approximately 1,560 walking tasks, containing approximately 12,480 complete gait cycles. Moreover, while recording at the digital goniometer's default frequency of 84 hertz, approximately 1,026,480 time-stamped samples were collected.

### 3.1.6. Data Processing

All recorded data was transferred from the relevant devices to a desktop computer for post-processing. The goniometer sensor recording was

manually started and stopped at the beginning and end of each task and, therefore, does not require any post-recording processing.

The video was recorded continuously during each experiment, so data processing was required to separate each walking sequence. First, the video was cropped to separate each recorded walking task into individual video files. This was achieved using video editing software (Windows Photo application); each task was manually cropped so that each video starts at the first sign of movement after the audible start command and ends either when motion stops or where the participant disappears from the camera view, whichever occurs first. The video and sensor data were manually synchronised for each recording, with the sensor manually started and stopped when the command to the participants was verbally given. The videos were then carefully cropped to equal the duration of the sensor data. This was achieved by aligning the start and stop points using the motion clues available in the video (visible movement) and sensor data (detected movement). The videos provided in the dataset have all been cropped using this video cropping process.

Furthermore, additional labelling is provided in the form of Human Pose Estimation (HPE) to aid future work. Using the pre-trained OpenPose [75] 2D HPE system, 25 anatomical keypoints were labelled on the participant's body for each video frame. This processing provides estimated skeletal movement tracking, which may aid research areas, including gait analysis and identification. To ensure consistency and quality, a visual check of the labelling accuracy was performed before adding the data to the final version of the dataset. From the 1,560 collected walks, approximately 748,800 frames were labelled, which provides approximately 18,720,000 labelled keypoints. The labelling was provided in two ways. First, a folder was created for each video file to hold a JSON file for each frame; this file contains the location of the keypoints detected in the frame. Additionally, a video was generated from the JSON files so that each of the original videos also has a corresponding video (.avi format) containing only the extracted "skeleton" generated from the anatomical keypoints in the JSON files. An

example of an extracted and visualised skeleton is provided in Figure 11 alongside the original video frame (cropped for convenience here).

*Figure 11. A comparison of the original RGB frame (left) and the extracted OpenPose* [75] *HPE skeleton (right). The image was cropped for comparative purposes. Image shared with participant's consent.*

## 3.2. Data Records

The dataset described here is provided via the LJMU data repository (https://opendata.ljmu.ac.uk/id/eprint/133/) [22]. This repository contains a folder which includes a copy of the anthropometry information of the participants (Participant Information.xlsx) in addition to one folder per participant. Each participant's folder contains the entirety of the recorded data for that participant across all three experiments, as illustrated in Figure 12. There is one subfolder for each of the three experiments, and inside them are six folders for each of the walking tasks within the experiment, plus one folder for the experiments where the participant wore alternative clothing and repeated tasks 1 and 2. To make this clear to the user, the

repeated task files are appended with "_ALT" to specify that alternative clothing was worn for these repeated tasks. Finally, inside each of the task folders (e.g., E1_T1) the data for the task can be found, which includes two video files of the recorded walk in .mp4 format (i.e., two different angles of the same walking task), the corresponding extracted skeleton videos in .avi format, two folders containing the corresponding skeleton key points (one folder per video, one JSON file per frame), and a .csv file containing the digital goniometer data.

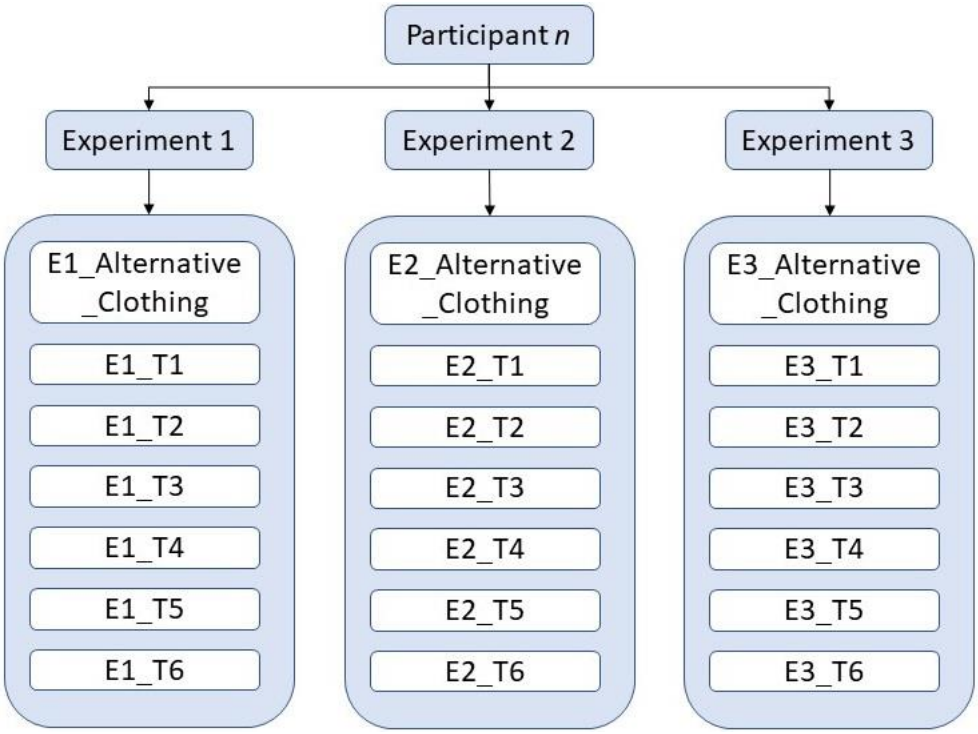


*Figure 12. The dataset folder structure per participant contains 3 experiment folders (e.g., Experiment 1), each containing 7 task subfolders (e.g., E1_T1 represents Experiment 1 Task 1).*

A simple file naming convention is used throughout the dataset to ensure each file is easily understood. A folder is assigned to each participant, labelled Participant_x, where x is the sequential number assigned to the participant. Inside each participant folder are three experiment folders, labelled Experiment_1, Experiment_2, and Experiment_3, where an experiment (e.g., Experiment 1) is a group of individual recorded walking tasks. Inside the individual experiment folders (e.g., Experiment 1), each of the seven task sub-folders is labelled to include the experiment number followed by the task numbers; for example, the task folder E1_T1

corresponds to Experiment 1, task 1. Moreover, inside the task folders, the files also contain the participant number, followed by the experiment number and task number. For example, the digital goniometer file P1_E1T1.csv corresponds to participant 1, experiment 1, task 1. Furthermore, the video files also include an appended camera angle. Here, _45 refers to a 45° recording angle, _90 refers to a 90° recording angle, _front represents a 0° recording angle (i.e., the participant is walking directly towards the camera), and _rear represents a 180° recording angle (i.e., the participant is walking away from the camera). For example, P1_E1T1_90 refers to the video recording of participant 1, experiment 1, task 1, recorded at a 90° angle (see Figure 6 and Figure 7 for an illustration of the camera layouts for each experiment). Finally, "_ALT" is appended to video files within the alternative clothing folder to signify that the participant has repeated the experiment with alternative clothing in these records, compared to those worn in the original task recordings.

Furthermore, the columns recorded by the digital goniometer device are described in Table 10. The columns include a timestamp and quaternions (q0, q1, q2, q3). The ROM column provides the calculated measurement of the limb segment ROM. Roll, Pitch, and Yaw provide Euler angles describing the movement of the sensor. The acceleration of the sensor for the x, y, and z-axis, as measured by the accelerometer, are provided in accX, accY, and accZ. The rotation of the sensor for the x, y, and z-axis, as measured by the gyroscope, is provided in gyrX, gyrY, and gyrZ. Finally, the magnetometer sensor readings for the x, y, and z axis are provided in magX, magY, and magZ.

## 3.3. Technical Validation

Significant effort and care were taken to ensure correct and consistent video and sensor data recording. This included carefully placing and testing equipment and reviewing the collected data for quality assurance.

### 3.3.1. Sensor Placement and Function

The sensors were placed on the participants according to the manufacturer's instructions [201], as shown in Figure 10. An adjustable strap was used to attach the sensor to the participants, ensuring that the sensor was firmly attached regardless of the type of clothing the participant wore. Each participant was requested to confirm that the placement and attachment of the sensor did not impede normal movement before each recording. Furthermore, before commencing each recording, the sensor signal was manually verified during a practice walk to ensure proper function. Manual inspections continued throughout the experiments to ensure proper function was maintained.

### 3.3.2. Camera Placement

Consistent camera placement (see Figure 6 and Figure 7) was ensured via manual measurement of the camera placement described in the methods section. The distance between cameras, the walking path, the position of the cameras, the height of the camera, and the angle of the cameras were all manually measured before each recording session.

### 3.3.3. Quality Assurance

A thorough review of the collected dataset was performed to ensure quality. The quality was assessed for sensor quality, image quality and video processing quality (e.g., video cropping, data synchronisation, keypoint annotation). Furthermore, checks were performed to ensure that only the experiment participants were captured in the video recording and that no bystanders were captured, which is an important ethical issue. Similarly, we confirmed that no unnecessary identifiable information was included in the videos, such as house numbers, street names, or car registrations.

## 3.4. Summary

We compiled a comprehensive multi-modal gait dataset (Obj3) containing 65 diverse subjects in various real-world environments, with alternative backgrounds and viewing angles. It was collected using two digital cameras and a wearable digital goniometer to capture visual and motion signal gait data, respectively. Traditional methods of gait identification are often affected by the viewing angle and appearance of the participant; therefore, this dataset pays special attention to diversity in various respects (e.g., participants' attributes, background variations, and view angles). The dataset is captured from 8 viewing angles in 45° increments, along with alternative appearances for each participant, for example, via a change of clothing. The dataset provides 3,120 videos, containing approximately 748,800 image frames with detailed annotations (Obj4), including approximately 18,720,000 bodily keypoint annotations, identifying 25 keypoints per video frame, and approximately 1,026,480 motion data points captured from a digital goniometer for three limb segments (thigh, upper arm, and head) (Obj5). Moreover, the dataset contains alternative

appearances for the participants to allow investigation into the effect of appearance change on gait identification methods.

# Chapter 4

# Sensor-based Gait Identification

This chapter proposes a new gait-based person identification approach using movement data regarding multiple body joints and machine/deep learning algorithms. Movement data is collected for the hip and shoulder joints using an Inertial Measurement Unit (IMU)--based digital goniometer device [201]. Physiotherapists commonly use goniometers to measure a body joint's range of motion (i.e., angle) [200]. Digital goniometers typically include IMU sensors containing accelerometers, gyroscopes, and magnetometers to measure movement [202]. This work may be considered a sensor-based prototype for the computer vision work described in Chapter 5. Non-Invasive Gait Identification Using Spatiotemporal Biometrics.

Existing works such as [206] use joint information during classification. However, [206] also relies on the Kinect camera to make physical measurements, such as limb length measurements, which are difficult to obtain without specialist sensors such as depth sensors. To the best of the author’s knowledge, this is the first gait identification method based only on the angle of a single joint. This approach may offer the opportunity to decrease the impact of problems such as occlusion in computer vision-based gait identification. The work was implemented and evaluated using the IMU data from our dataset described in Chapter 3, Primary Dataset Compilation. The dataset contains IMU movement data from both arm and leg joints (i.e., hip and shoulder joints) recorded during normal walking sequences via wearable IMU-based digital goniometers. Furthermore, we also present unique feature vectors generated by time-series analysis of the limb movement data from our acquired dataset. This

helps explain the distinct spatial and temporal gait information which can be used for robust person identification. Moreover, we describe a simple two-stage approach to filtering the gait identification predictions, which may be applied to alternative solutions.

## 4.1. Methodology

The proposed gait identification solution is a composite of several tasks mainly related to the data science cycle, as shown in Figure 13. In the first step, the IMU signal for the body joints is acquired from the real walking sequence. Then, the raw data is pre-processed (e.g., cleaned, standardised, etc.) using an overlapping (50%) fixed-width window (equivalent to the duration of the average gait cycle). A unique feature vector is then extracted through the overlapped windowing of the data, resulting in many statistical features (221 for leg movement and 221 for arm movement). A similar approach is taken in [207]–[209], where windowing data in this manner is expected to allow LSTM to learn temporal dependencies between windows, which provide complete information regarding an entire gait cycle at a given time.

The pre-processed data is then partitioned into two subsets; the training subset contains 80% of the data per participant and is used to train the required machine-learning models. The remaining 20% is reserved and not exposed to the machine learning models until the testing phase. As this is a person identification problem, it is necessary to provide training and testing data for each participant, as such models cannot identify specific individuals if they have not previously been trained using data associated with that individual.

From the 442 features available, the 30 most important features are identified using feature selection methods, including Principal Component Analysis (PCA), the Boruta algorithm, and Recursive Feature Elimination

(RFE). In the next step, an LSTM model is trained to identify the relevant participant using the extracted features. Finally, the model is evaluated with varying experimental configurations using the reserved and previously unseen test data (20% of the original data per participant is reserved for testing only). Here, the model predicts the participant associated with walking sequences to which it has not previously been exposed. At this stage, various alternative machine learning models are also evaluated to provide a comparison to the proposed LSTM model approach. Detailed descriptions of each component in the proposed method are presented in the following subsections.

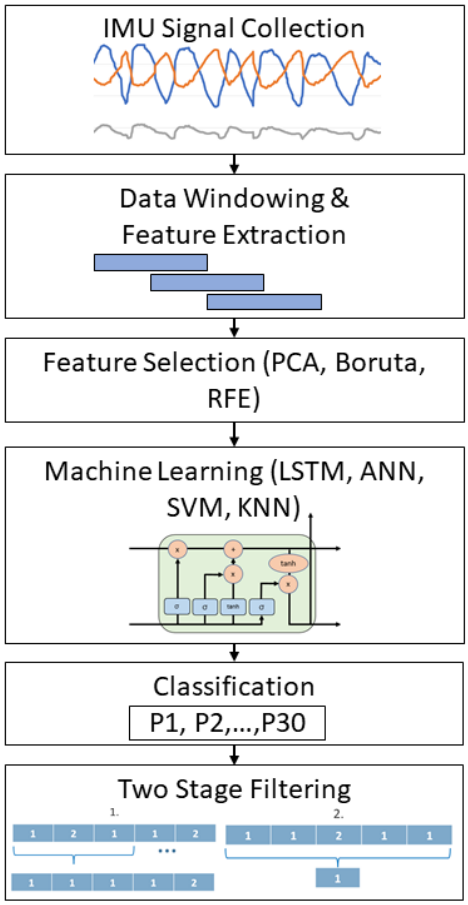


*Figure 13. Overview of the sequential process of the sensor-based gait identification methodology.*

## 4.1.1. Data Windowing

Similar works such as [210] have successfully used a fixed-length window with a 50% overlap of windows. This allows statistical features to be

calculated for each window, as described in Section 4.1.2. Feature Extraction. A fixed-length window equivalent to the size of a complete gait cycle was chosen as it would contain full gait information for a whole step. However, a single fixed length must be selected for all instances for all subjects; therefore, the fixed length would be equivalent to the average gait cycle duration with a 50% overlap. However, the literature does not agree upon an average duration. Thus, our dataset was sampled to calculate an average. The sample contained five random gait cycle instances from ten randomly selected participants. Figure 14 shows the median and quartiles of the sample gait cycles, with the overall average gait cycle found to have a 0.965-second duration. A second windowed dataset was created for comparison, with the same overlap but with a fixed-length window size equal to 50% of the average gait cycle.

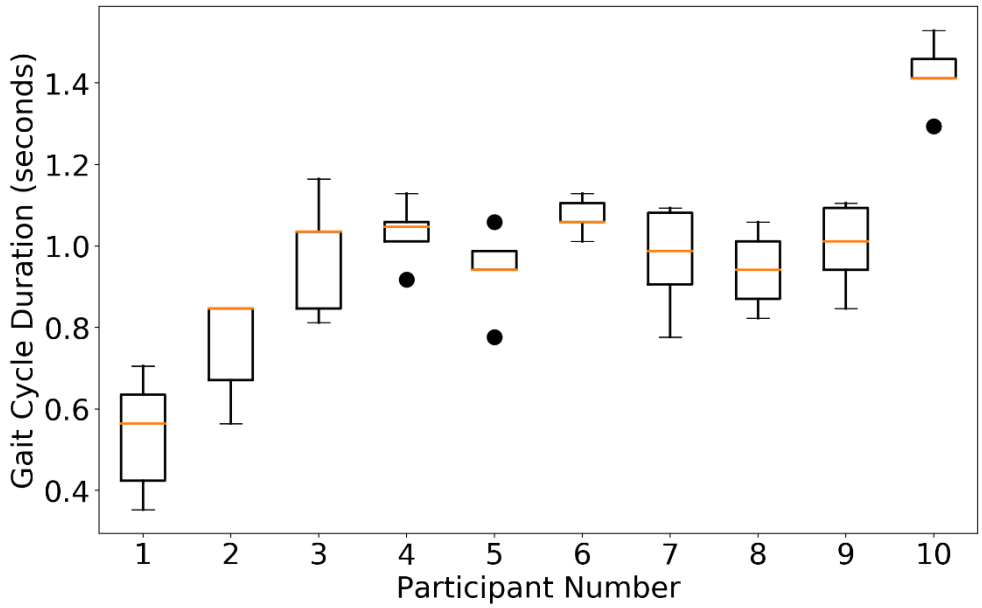


*Figure 14. Box plots showing the sampled gait durations in seconds calculated from five samples per participant.*

An average complete gait cycle is empirically selected as an optimal fixed-width window size, as shown in Table 11. Table 11 suggests that the optimal window size equals the average complete gait cycle (approximately 0.965s in this dataset), which comprises sufficient information concerning unique movement patterns. Furthermore, Table 11 also suggests that a sample size of 5 is optimal.

*Table 11. A summary of the results of the proposed gait identification model with varying window and sample sizes.*

| Sample Size | Window Size = 0.483 (s) | Window Size = 0.965(s) |
|---|---|---|
| 1 | 74.21% | 77.17% |
| 5 | 81.23% | 97.66% |
| 10 | 62.1% | 85.36% |

### 4.1.2. Feature Extraction

As described in Chapter 3 Primary Dataset Compilation, an IMU-based digital goniometer was used to obtain movement data for various limbs and joints. Table 10 describes the range of measurements and features provided by the sensor. The device provides 17 synchronised and timestamped variables (measured at a sampling frequency of 84 hertz). Feature extraction is then performed on the windowed data (as described in Section 4.1.1. Data Windowing) to obtain helpful gait clues to train machine-learning models.

A feature vector is produced by calculating each of the 13 statistical features described in Table 12, as illustrated in [208], [211], [212], for each window of the 17 data points defined in Table 10. As a result, a vector containing 221 features is generated (for leg movement) using overlapped window segmentation. Likewise, we extracted 221 features for arm movement, resulting in 442 features in total. Table 12 describes the statistical features calculated in each window for each data point described in Table 10, as inspired by [208], where similar features are used for an accelerometer only in detecting Freezing of Gait (FoG) as a symptom of Parkinson's disease. Further detail regarding the range of statistics used can be found in [212]. The statistical features selected in Table 12 provide a range of temporal and spatial features when combined with the data points in Table 10. For example, the maximum, minimum, and mean motion degrees provide information regarding the stride length, and kurtosis and skewness provide information regarding the spread of the data. Furthermore, Table 12 includes common time domain features, including

mean and standard deviation and frequency domain features such as entropy [212].

*Table 12. Statistical features extracted from the overlapping windowed IMU input variables.*

| Statistic | Description / Formula |
|---|---|
| Mean | $\frac{\Sigma \mathrm{X}}{N}$ |
| Standard Deviation (STD) | $\sqrt{\frac{\Sigma(x_i - \mu)^2}{N}}$ |
| Mean Absolute Deviation (MAD) | $\frac{1}{n}\sum_{i=1}^{n} \lvert x_i - m(X) \rvert$ |
| Maximum | $\max(x_j) - x_{j.n}$ |
| Minimum | $\min(x_j) - x_{j.1}$ |
| Variance (Var) | $\frac{\Sigma(x_i - \bar{x})^2}{N-1}$ |
| Quantile | $F_X(x) := \Pr(X \leq x) = p$ |
| Entropy | $\left(1 + \left(X_{1,i} + X_{2,i} + X_{3,i}\right)\right)^2 log\left(1 + \left(X_{1,i} + X_{2,i} + X_{3,i}\right)\right)^2$ |
| Max Index | Index of the maximum |
| Skewness | $\frac{1}{n}\sum_{i=1}^{n} \frac{(x_{j,i} - \bar{x})^3}{S_j^3}$ |
| Kurtosis | $\frac{1}{n}\sum_{i=1}^{n} \frac{(x_{j,i} - \bar{x})^4}{S_j^4}$ |
| Median | $med(x_j) = \begin{cases} x_{j,\frac{n}{2}}, if\ n\ is\ even \\ \frac{x_{j,\frac{n-1}{2}} + x_{j,\frac{n-1}{2}}}{2}, if\ n\ is\ odd \end{cases}$ |
| Standard Error of Mean (SEM) | $\frac{STD}{\sqrt{N}}$ |

### 4.1.3. Feature Selection and Dimensionality Reduction

A two-stage feature selection stage started with 442 features, as described in Section 4.1.1. First, highly correlated features were removed using Principal Component Analysis (PCA) to return the features with the highest variance. Second, the Boruta algorithm and RFE were performed on the reduced feature vector. Features selected by both Boruta and Recursive Feature Elimination (RFE) were returned to produce the final feature vector containing the most important features. Thirty features comprised 80% of the total variance among the available feature set. Therefore, these thirty features were selected as the most important; thirty features were also reported to be sufficient in [208].

#### *4.1.3.1. Principal Component Analysis*

PCA has been widely used as a transformation method [213] as well as the method of feature selection, as in [214], [215]. PCA can select the features with the highest importance to the classification task by selecting for high variance, thus ensuring variety in the feature vector [216]. The correlation coefficients between the statistical features extracted from gait data and the principal components (obtained through PCA) are represented by PCA component loadings. The maximised sum of variances of the squared loadings is then provided by the component rotations, where the absolute sum of component rotations produces importance values for the corresponding features. In the case of the proposed feature vector, the first 30 Principal Components (PCs) covered most of the variance (i.e., almost 80%) and were therefore used to calculate the features' importance. The mathematical formulation of PCA's attribute loadings and feature ranking can be found in [217].

#### *4.1.3.2. Boruta Algorithm*

As in [208], the Boruta algorithm was used for feature selection by identifying the most essential features for the classification task. The algorithm achieves this by implementing a random forest classifier and iteratively removing features that are deemed statistically less important [218]. The less important features are identified as statistically less relevant than random probes [218]. A detailed explanation of the Boruta algorithm can be found in [218].

#### *4.1.3.3. Recursive Feature Elimination (RFE)*

As introduced in [219], RFE recursively removes less important features until only the desired number of the most important features remains. At each stage, a machine learning model such as an SVM is trained, each feature is ranked by importance to the output, and finally, the least important features are removed [219].

*Table 13. The thirty most important features identified during feature selection.*

| **Leg Features** | **Arm Features** |
|---|---|
| MAD-MotionDeg | Min-Pitch |
| Max-Pitch | Max-gyrZ |
| Min-Pitch | Mean-accZ |
| STD-Yaw | Max-accX |
| MAD-accX | Max-Pitch |
| Max-accX | MAD-accY |
| Mean-accZ | Max-MotionDeg |
| Median-accZ | Max-accY |
| MAD-gyrX | Max-gyrX |
| Max-gyrX | Max-Pitch |
| Quantile-accZ | MAD-accZ |
| Var-gyrX | Min-accZ |
| Max-gyrY | Median-accZ |
| Min-gyrX | Min-gyrX |
| Skewness-gyrY | MAD-gyrZ |

Table 13 presents the thirty most important features (15 for the leg and 15 for the arm) identified using overlapped feature selection methods. Each feature is in the form ‘‘statistical feature–data point’’; the

corresponding sensor data points and statistical features can be found in Table 10 and Table 12, respectively.

As shown in Table 13, the most commonly appearing sensor data points in the list of most important features are accZ, gyrX, and pitch, suggesting that a range of IMU-based sensors (i.e., containing at least an accelerometer and gyroscope) may be of benefit to similar solutions. Conversely, none of the magnetometer features were selected in the list of the most important features; this suggests that such sensors may be less valuable to the gait identification task than the remaining sensors. Furthermore, the most commonly appearing statistical features in the list of essential features are Max, Min, and MAD.

## 4.1.4. Machine Learning Classification Models for Gait Identification

A series of relevant machine learning models were trained and evaluated using the extracted feature vector to compare their performance for gait identification. The various models were identified in the literature [72], with multiple studies reporting success with the specified models. The identified models were LSTM, SVM, Feed Forward ANN and KNN. All experiments and model training were performed on a computer with an AMD Ryzen Threadripper processor, four NVIDIA 2080 RTX graphics cards, each having 2944 CUDA cores for a total of 11776 CUDA cores, a base clock of 1.52GHz, total GPU combined memory of 32GB (8GB per GPU), and a 192-bit memory interface.

### *4.1.4.1. LSTM for Gait Identification*

First introduced in [220], LSTM [221] is a type of RNN capable of learning order dependence in sequential data, including long-term dependencies [222]. It was designed to improve performance for problems involving long sequences of data [223]. The main difference between LSTM and other RNNs is the construction of the LSTM, which contains three components. An LSTM

cell contains a forget gate that controls how much information is retained, an input gate that updates the values contained in the hidden states, and an output gate that updates the cell output value [223]. For the given task of person identification, the LSTM model is appropriate because of the time-series nature of the data involved and the ability of LSTM to learn temporal dependencies between data samples [119]. Moreover, LSTM has been successfully used for time-series prediction tasks in the literature [224].

Previously, Section 4.1.2. Feature Extraction described the feature extraction method. However, LSTM must be trained over temporal information; therefore, the data must be formatted into 3D batches of samples containing multiple sequential windows. In addition, the following training properties were used: 200 epochs, categorical cross-entropy loss function, and Adam optimiser.

For this work, Table 14 presents the LSTM model configuration implemented in this work, which was selected empirically by repeating the experiment with additional layers until the optimum configuration was found. The optimum configuration contains four layers: an LSTM layer, a dropout layer with a dropout rate of 50% to help prevent over-training, and two dense layers. Two dense layers are implemented to model more complex non-linear relationships in the data; a similar approach is described in [225]. The final layer has an output shape of 30 to allow for the classification of the 65 participants included in this work. The final output is provided as a single output feature ranging from 1 to 30.

*Table 14. LSTM model configuration.*

| Model Layer | Output Shape | No. of Parameters |
|---|---|---|
| LSTM | 100 | 115200 |
| Dropout | 100 | 0 |
| Dense | 100 | 10100 |
| Dense | 30 | 1010 |

#### *4.1.4.2. SVM for Gait Identification*

SVMs perform classification by selecting a hyperplane which appropriately separates observations into their relevant classes [226]. A predicted label is assigned to an observation instance based on which side of the hyperplane it falls into [227]. Further steps are required for multi-class classification; for example, One-Vs-One (OVO) or One-Vs-All (OVA) strategies. In OVO, a multi-class classification problem containing *m* classes is split into a series of binary classification problems where the number of such problems can be calculated using (5 [228]. Alternatively, OVA creates *m* binary problems for an *m* class problem in which there is a binary classifier for each class, responsible for distinguishing one class from all other classes [228]. As in [229], OVO was chosen as the most appropriate strategy for this gait identification application as it generally performs better. Additionally, the OVA approach is more likely to produce imbalanced data, since one class is compared against all remaining classes [228]. The SVM was configured with cost = 1, a decision function shape of OVO, and a linear kernel.

$$\frac{m(m-1)}{2} \qquad (5)$$

#### *4.1.4.3. Feed Forward ANN for Gait Identification*

ANNs are inspired by the human brain and contain a network of neurons which learn from example data using mathematical and optimisation algorithms [230]. They contain an input layer connected to an output layer via one or more layers of neurons, known as hidden layers. A more detailed description of ANNs is provided in [212]. The optimal ANN configuration was selected empirically for this work. The best-performing ANN contained seven hidden layers and 256 neurons in each hidden layer, providing a total of 400,242 parameters, as shown in Table 15. In addition, the following training properties were applied: 300 epochs, batch size of 16, categorical cross-entropy loss function, and Adam optimiser.

*Table 15. Feed Forward ANN Model Configuration.*

| Model Layer | Output Shape | No. of Parameters |
|---|---|---|
| Dense | 6 | 1128 |
| Dense | 256 | 1792 |
| Dense | 256 | 65792 |
| Dense | 256 | 65792 |
| Dense | 256 | 65792 |
| Dense | 256 | 65792 |
| Dense | 256 | 65792 |
| Dense | 256 | 65792 |
| Dense | 30 | 2570 |

#### *4.1.4.4. KNN for Gait Identification*

KNNs are an example of similarity searching. They are often used to find patterns within multi-dimensional data such as time-series data [224], [231], which is the case for gait identification. They are based on a distance function that measures the difference between two observations, instances *x* and *y*, often using the Euclidean distance function defined in (6 [232], where $x$ is described by an attribute vector $(a_1(x), a_2(x), \dots, a_n(x)$ [232]. The KNN for this study was configured using ten nearest neighbours.

$$d(x, y) \sqrt{\sum_{i=1}^{n} \big(a_i(x) - a_i(y)\big)^2} \quad (6)$$

As described in [232], the KNN algorithm assigns the most common class of $x$'s $k$ nearest neighbour as described in (7, where $c$ is the class variable, $\delta$indicates change in c, $y_1, y_2, \dots, y_k$are the $k$ nearest neighbours of $x$, given $k$ nearest neighbours. $\delta(c, c(y_i)$ = 1 if $c = c(y_i)$, else $\delta(c, c(y_i)$ = 0 [232]. A detailed description of the KNN algorithm can be found in [233]. Despite being a popular and effective method, the KNN algorithm has several disadvantages. For instance, it is sensitive to redundant features, can give poor run-time performance, especially with larger datasets, and has a relatively high computational cost [234].

$$c(x) = \underset{c \in C}{\arg\max} \sum_{i=1}^{k} \delta(c, c(y_i)) \quad (7)$$

### 4.1.5. Two-Stage Output Filtering

Due to the time-series nature of the dataset, filtering can be used to improve the overall accuracy of the results, as predictions can be made using a series of samples produced per participant. For example, given a series of data samples provided by one person, each sample can be classified concerning the target class (i.e., an individual). Where the predicted classes erroneously belong to more than one class (i.e., multiple persons are predicted where only one is present), filtering can be applied to correct noisy predictions or make predictions based on a relatively more significant number of samples (i.e., many consecutive samples from a single walking sequence).

In this work, we propose a simple two-stage filtering approach applied after a sequence of data samples has been classified, the algorithm for which is presented in Algorithm 1. Two-Stage Filtering. First, a moving average is applied to the samples (S) in the input feature vector (f), whereby every triplet of samples is evaluated. Where there is a mode (m) in the predicted classes (i.e., 2 out of 3 or 3 out of 3 matches), the minority class in the output (o) is overwritten with that of the majority (m). This approach is likely to be effective where accuracy is already high (i.e., >80%) as the chance of predicting the same incorrect class 2 out of 3 times is relatively unlikely. In the case of there being three different classes, the process moves to the next triplet without making changes. Second, as this work provides relatively large sequences of data, a final optional step may be taken where the mode is taken across all samples in a sequence (i.e., an entire 8-metre recording sequence for one individual), thereby making a single classification prediction for the whole sequence as opposed to making predictions per

sample. This is performed by replacing all sample predictions for that sequence with that of the mode. Both stages of the filtering algorithm are evaluated using the reserved test data (20% of the original dataset).

**Algorithm 1. Two-Stage Filtering**

- Let *C* be the classifier such that C ∈ {ANN, LSTM, SVM} used for the prediction for the current frame of feature vector (*f*)
- Let *O* be the list of output from *C* corresponding to each sample *S* in *f*
- Sequence (*Seq*) is the complete walking data of one experiment for one person
- Let *m* be the mode of a sublist
- Let *L* be the lower limit of the modal class
- Let *h* be the size of the modal class
- Let $f_m$ be the frequency of the modal class
- Let $f_1$ be the frequency of the class preceding the modal class
- Let $f_2$ be the frequency of the class succeeding the modal class

**Inputs:** Feature Array (*f*) from Table 13
**Output:** Predicted Person Identity (*p*)

**Procedure:**

**For** each windowed *Seq* per person:

**For** each sample (*s*) from *f*:
- Select the optimal *C (LSTM)*
- Predicted class (*p*)=predict(*s*) using *C*
- Store the outcomes *p* from *C* in list *O*

**End** Loop

$$m = L + h\frac{(f_m - f_1)}{(f_m - f_1) + (f_m - f_2)}$$

**If** isnull(m):
Skip
**Else:**
*p = m*
overwrite *p* in *o*
**End** If

**End** Loop

## 4.1.6. Experimental Design

This work contains three main experiments: the first uses the sensor data for the leg only, the second uses the data for the arm only, and the third uses synchronised data for both the leg and arm. In addition, three further experiments are described, using the most important selected features

rather than the large feature vector acquired before feature selection. To achieve this, three datasets were created using the aforementioned data, each containing a training set consisting of 80% of the total records for that experiment and a test set containing the remaining 20%. The test set was extracted carefully to ensure that consecutive sequences were extracted (i.e., not random) and remained unseen by the model during training.

These experiments aim to test which body joints provide sufficient information to classify individuals uniquely and whether combining joints improves identification accuracy. Moreover, the additional experiments will demonstrate whether the subset of the most important features can provide the same identification accuracy level as the larger original feature vector. Section 4.1.2. Feature Extraction describes the feature extraction approach, which is consistent for all three experiments. Note that for the third experiment, consisting of the combined data, the feature vector is double that of the feature vector provided by the individual joints (30, compared to 15 for the individual joints).

#### *4.165.1. Experiment 1-A (Exp.1-A) Leg Only*

In Exp.1-A, the motion data from only one of the two MOTI sensors is used (i.e., the one attached to the leg). By calculating the 13 statistical features in Table 12 for each of the 17 data points provided in Table 10, 221 potential features are generated for the hip motion. Such an experiment may be used to simulate gait identification under partial occlusion. For example, if the upper body was occluded (i.e., by omitting the arm data, we may simulate occlusion of the arm).

#### *4.1.6.2. Experiment 1-B (Exp.1-B) Leg Only – Selected Features*

As in Exp.1-B, only the leg data was used. However, instead of using the entire 221 feature vector, only the most important 15 features were used, as described in Section 4.1.3. Feature Selection and Dimensionality Reduction

#### *4.1.6.3. Experiment 2-A (Exp.2-A) Arm Only*

In Exp.2-A, the motion data from only one of the two MOTI sensors was used (i.e., the one attached to the arm). By calculating the 13 statistical features in Table 12 for each of the 17 data points provided in Table 10, 221 potential features were generated for the shoulder motion. Such an experiment may simulate gait identification under occlusion, such as if the lower body was occluded (i.e., by omitting the leg sensor data).

#### *4.1.6.4. Experiment 2-B (Exp.2-B) Arm Only – Selected Features*

As in Exp.2-B, only the arm data was used. However, instead of using the entire 221 feature vector, only the most important 15 features were used, as described in Section 4.1.3. Feature Selection and Dimensionality Reduction.

#### *4.1.6.5. Experiment 3-A (Exp.3-A) Leg and Arm*

In Exp.3-A, the motion data from both sensors generated 442 potential features (221 for the leg and 221 for the arm). This experiment evaluated combined joint movements (i.e., simulating no/less occlusion).

#### *4.1.6.6. Experiment 3-B (Exp.3-B) Leg and Arm – Selected Features*

As in Exp.3-A, the combined arm and leg data was used. However, instead of using the entire 442 feature vector, only the most important 30 features were used, as described in Section 4.1.3. Feature Selection and Dimensionality Reduction.

## 4.2. Results of the Sensor-based Gait Identification Method

Following the experimental design (Section 4.1.6. Experimental Design), statistical results were retrieved from multiple experiments and evaluated using the reserved, previously unseen test data described in Section 4.1. For

each experiment, results of the proposed LSTM model are provided in addition to the alternative models described in Section 4.1 to provide a comparison. The results of both stages of the two-stage filtering algorithm (i.e., the moving average and mode columns) are presented, as detailed in Algorithm 1. Both stages of the filtering algorithm are evaluated using the reserved test data. The first stage of the filtering algorithm, the moving average, is evaluated on a per-sample basis (i.e., many samples per walking sequence). Meanwhile, the mode metric is evaluated on a per-sequence basis due to the nature of the second stage (i.e., one prediction per walking sequence).

Table 17 presents the results of Exp.1-A, completed using leg data only, utilising all 221 available features. From Table 17, the highest accuracy was achieved using the proposed LSTM approach, with an accuracy of 97.3% when evaluated using purely unseen test data. However, the additional metrics, precision, recall, F1 score, and Cohen's Kappa are all highest for the SVM model, suggesting that this may provide a more stable and balanced classification in terms of a reduced rate of false positives.

Table 18 presents the results of Exp.1-B, which was also completed using the leg data only. However, only the 15 most important features were used in this experiment. As in the results from Exp.1-A, the proposed LSTM approach provides the best accuracy. Furthermore, when using only the top 15 as in Exp.1-B, all models report slightly increased results compared to all features in Exp.1-A (see Table 17 and Table 18), suggesting that an appropriate subset of the original feature vector has been correctly selected.

Table 19 presents the results of Exp.2-A, completed using the arm data only, using all 221 available features. As with Exp.1-A, the proposed LSTM architecture provides the highest accuracy with 98.9%, the highest reported across all experiments, which was achieved when evaluated by using the previously unseen test data. As with experiment 1, the SVM model provided higher scores for the remaining metrics.

Table 18, Table 20, and Table 22 provide the results for Exp.1-B, Exp.2-B, and Exp.3-B, respectively. Specifically, they provide the results

when performed using the most important features only, whose selection was described in Section 4.1.3. Feature Selection and Dimensionality Reduction. The accuracies reported by the LSTM models are similar for the three experiments when using the complete feature set and when using a reduced feature set, with no more than 1.2% difference. This suggests that the features identified as important in Section 4.1.3. Feature Selection and Dimensionality Reduction are sufficient for the gait identification task. The novel feature vector presented in Table 13 may be useful to future IMU-based gait identification works.

Table 20 presents the results of experiment Exp.2-B, which was also completed using the arm data only, using only the 15 most important features. As in Exp.2-A, described in Table 19, LSTM provides the best accuracy, as shown in Table 20. Although slightly lower accuracies were reported for all models when using only the top features for experiment 2, the difference reported by the LSTM model was only 0.5%.

Table 21 presents the results of Exp.3-A, completed using synchronised leg and arm movement data, using all 442 available features. Again, the proposed LSTM architecture provided the highest accuracy, with 97.5% achieved when evaluated on previously unseen test data. As with Exp.1-A and Exp.2-A, SVM provided heightened results for the additional metrics.

Table 22 presents the results of Exp.3-B, which was also completed using synchronised leg and arm data, using only the 30 most important features. As in the results of Exp.3-A shown in Table 21, LSTM provides the highest accuracy, as indicated in Table 22. Furthermore, when using the top features only, the LSTM model accuracy improves by 1.18%, as shown in Table 22 when compared to the results when using all features, as described in Table 21. However, all alternative models report lower accuracy than the complete feature set.

Like our proposed solution, an LSTM model was used in [207] to identify a smartphone user using accelerometer data only. Accuracies above 90% were achieved using combinations of the 21 total participants.

However, certain combinations of participants provided much lower identification rates; this is an unexplained issue the authors expect to be exacerbated by adding further participants [207].

*Table 16. A comparison of IMU-based gait identification methods.*

| Ref | No. Participants | Sensor Placement | Model | Accuracy |
|---|---|---|---|---|
| [207] | 21 | Trunk | LSTM | 90.00% |
| [235] | 24 | Ankle | CNN / SVM | 80.00% |
| [236] | 30 | Trunk | GMM-UBM | 80.30% |
| [237] | 10 | Body Network | CNN | 91.00% |
| [238] | 744 | Trunk | CNN | 97.16% |
| Ours (Exp.1-B) | 30 | Leg | LSTM | 97.30% |
| Ours (Exp.2-B) | 30 | Arm | LSTM | 98.90% |
| Ours (Exp.3-B) | 30 | Leg & Arm | LSTM | 98.68% |

Moreover, the results of Exp.1-A, Exp.2-A, and Exp.3-A, as presented in Table 16, Table 18, and Table 20, respectively, suggest an opportunity to explore joint-based gait identification with various body joints. Moreover, the LSTM-based solution outperformed the ANN, KNN, and SVM-based alternatives provided as benchmark comparisons for the newly proposed dataset. In each experiment, the LSTM-based solutions provided the highest accuracies, suggesting that LSTM is an appropriate approach for gait identification using time-series data from individuals or combinations of body joints. Exp.2-A, using arm data only, provided the highest accuracy. This is likely due to people's more varied arm movement, as discussed in Section 4.3. Discussion. This suggests an opportunity to explore arm and joint movement further to contrast the traditional approaches that focus on either the legs or the trunk of the body.

*Table 17. Results of experiment 1-A, using leg data only (Exp.1-A).*

| Model | Accuracy | Precision | Recall | F1 Score | Kappa | Moving Average | Mode |
|---|---|---|---|---|---|---|---|
| ANN | 86.00% | 59.70% | 59.00% | 58.50% | 55.40% | 86.67% | 100.00% |
| KNN | 72.90% | 75.30% | 73.80% | 72.10% | 71.90% | 74.21% | 83.33% |
| SVM | 94.40% | 94.90% | 94.30% | 94.40% | 94.30% | 94.46% | 100.00% |
| LSTM | 97.30% | 85.70% | 69.40% | 70.40% | 71.30% | 97.34% | 100.00% |

*Table 18. Results of experiment 1-B, using leg data only (top 15 features only) (Exp.1-B).*

| Model | Accuracy | Precision | Recall | F1 Score | Kappa | Moving Average | Mode |
|---|---|---|---|---|---|---|---|
| ANN | 85.36% | 54.31% | 52.72% | 52.44% | 51.80% | 87.30% | 93.3%% |
| KNN | 66.89% | 70.46% | 66.28% | 66.44% | 65.61% | 67.50% | 83.33% |
| SVM | 74.44% | 76.55% | 74.36% | 74.15% | 73.50% | 75.27% | 100.00% |
| LSTM | 98.23% | 86.94% | 71.46% | 71.31% | 71.91% | 98.64% | 100.00% |

*Table 19. Results of experiment 2-A, using arm data only (Exp.2-A).*

| Model | Accuracy | Precision | Recall | F1 Score | Kappa | Moving Average | Mode |
|---|---|---|---|---|---|---|---|
| ANN | 86.10% | 62.60% | 62.30% | 61.70% | 57.10% | 87.13% | 100.00% |
| KNN | 81.00% | 82.00% | 81.80% | 80.10% | 80.20% | 81.37% | 93.30% |
| SVM | 95.70% | 95.00% | 95.00% | 94.80% | 95.50% | 95.77% | 100.00% |
| LSTM | 98.90% | 93.40% | 84.30% | 78.80% | 81.20% | 98.90% | 100.00% |

*Table 20. Results of experiment 2-B, using arm data only (top 15 features only) (Exp.2-B).*

| Model | Accuracy | Precision | Recall | F1 Score | Kappa | Moving Average | Mode |
|---|---|---|---|---|---|---|---|
| ANN | 78.40% | 69.24% | 69.05% | 67.62 % | 67.29% | 78.76% | 93.30% |
| KNN | 77.61% | 80.10% | 77.97% | 77.12% | 76.77% | 78.30% | 93.30% |
| SVM | 79.39% | 81.62% | 80.75% | 80.21% | 78.59% | 80.16% | 100.00% |
| LSTM | 98.33% | 93.33% | 83.20% | 77.96% | 81.78% | 98.37% | 100.00% |

*Table 21. Results of experiment 3-A, using synchronised leg and arm data (Exp.3-A).*

| Model | Accuracy | Precision | Recall | F1 Score | Kappa | Moving Average | Mode |
|---|---|---|---|---|---|---|---|
| ANN | 86.10% | 61.90% | 62.30% | 61.90% | 57.80% | 87.21% | 100.00% |
| KNN | 88.40% | 89.90% | 88.30% | 88.20% | 88.00% | 90.02% | 100.00% |
| SVM | 96.20% | 98.40% | 98.20% | 98.30% | 98.20% | 97.10% | 100.00% |
| LSTM | 97.50% | 89.00% | 79.80% | 77.60% | 80.70% | 97.52% | 100.00% |

*Table 22. Results of experiment 3-B, using synchronised leg and arm data (top 30 features only) (Exp.3-B).*

| Model | Accuracy | Precision | Recall | F1 Score | Kappa | Moving Average | Mode |
|---|---|---|---|---|---|---|---|
| ANN | 77.80% | 79.09% | 79.93% | 78.47% | 76.96% | 79.10% | 100.00% |
| KNN | 89.78% | 90.36% | 89.10% | 88.86% | 89.41% | 89.81% | 100.00% |
| SVM | 92.44% | 93.06% | 92.70% | 92.57% | 92.16% | 92.65% | 100.00% |
| LSTM | 98.68% | 90.38% | 82.27% | 79.76% | 80.72% | 98.73% | 100.00% |

# 4.3. Discussion

As shown in Section 4.2. Results of the Sensor-based Gait Identification Method, when evaluated over unseen instances, each trained model indicated higher accuracy for the arm data (Exp.2-A) than the leg data (Exp.1-A). This is likely due to the more obvious differentiations in arm movements that can be visually observed. For example, some people are rigid and have minimal arm movement when they walk, some people allow their arms to swing naturally with the rhythm of their walking, and some people use more energy and swing their arms more flamboyantly as they walk [239].

Furthermore, it can be observed that overall, the results of Exp.2-A, arm data only, provided higher accuracy than those of Exp.3-A, the combination of leg and arm data. This was an unexpected result as it was anticipated that the unique traits of both limbs would provide more unique gait traits for each participant.

As presented in Section 4.2. Results of the Sensor-based Gait Identification Method, Table 16 compares the accuracy of the proposed solution to that of similar IMU-based gait identification works from the literature. In [235], movement data was gathered from 24 participants via an IMU device attached to the ankle, containing an accelerometer and gyroscope. However, unlike the proposed work, it did not include a magnetometer, which would provide additional features. A CNN was used in [235] to perform feature extraction, from which an SVM classified participants. An accuracy of 80% is reported when provided with data containing five complete gait cycles. In contrast, we use a window size equivalent to a single complete gait cycle and a sample size of 5 windows, achieving an accuracy of up to 98.9% using a similar amount of data, as shown in Table 19. Furthermore, using a CNN for feature extraction in [235],

compared to our hand-crafted features, will limit the explainability and interpretability of the model [240].

Similarly, in [236], IMU data was gathered from 30 participants via a smartphone attached to the trunk of the body. As in [235], magnetometer and gyroscope data were not included. However, as demonstrated in our proposed solution, both may provide useful features. As described in Section 4.2, attaching the sensor to the trunk means that only generic whole-body data is provided and does not allow for the evaluation of individual limbs and joints, unlike the proposed work. Furthermore, the approach described in [236] used the same number of participants as in the proposed solution but achieved only 80.3% accuracy, considerably lower than the 98.9% we report in Table 16.

Unlike the other works described in Table 16 and our newly proposed solution, [237] gathered movement data using a network of 5 IMU sensors. Despite only containing 10 participants, an accuracy of 91% is reported, substantially lower than the 98.9% achieved by the proposed work, which comprises 65 participants and requires fewer sensors.

In [238], motion data was collected using a smartphone and 3 IMU sensors attached to the waist for 744 participants [238]. Using this dataset [238] reports 97.16% accuracy, slightly lower than that reported in our work. However, similar to [235], this solution relies on a CNN for feature extraction. Again, unlike the statistical features in the proposed work, CNN-based feature extraction would likely make the model less explainable. Furthermore, only the pre-processed dataset is available, not the original raw data, thus limiting the ability to reuse or experiment with the original data. Likewise, similar to[236], the sensor used in [238] was attached to the trunk; again, this provides motion data for the body and does not allow for limb or joint-based approaches to be implemented or evaluated. A further limitation of the dataset is that only 2.5 seconds of data was recorded per participant, thus limiting the number of complete gait cycles [241]. In comparison, we provide approximately 60 seconds of data per participant in our primary dataset.

Overall, the proposed LSTM model outperformed the alternatives in all experiments, achieving classification accuracy of up to 98.9%. In addition, reducing the feature vector by more than 93% did not reduce the performance, and in 2 out of three experiments, the feature vector reduction led to improved classification accuracy. The results of the two-stage filtering approach described in Algorithm 1, as shown in Table 17, Table 18, Table 19, Table 20, Table 21, and Table 22, suggest that it leads to improved results in longer walking sequences (e.g., 8 metres in this case). The mode metric's success is partly due to the high accuracies achieved during the initial prediction. As up to 98.9% accuracy was achieved when evaluating the predicted samples, the mode (per sequence) must be correct as most of the samples within each sequence are correctly predicted, thus accounting for the 100% accuracy achieved in some experiments by the mode metric. However, further work and more data are required to validate the mode metric further, as each walking sequence provides only a single prediction to validate when using this approach.

The aforementioned statistical performance clearly validates the proposed proof-of-concept, which can be a baseline to explore further for computer vision-based implementation. This work demonstrates that when provided with accurate joint angles, high accuracy of gait-based identification is achievable. Therefore, it is likely that a high identification accuracy can be expected from future vision-based implementations if the joint angle can be accurately estimated using computer vision methods. Computer vision provides the potential for more practical implementations; for example, for security and policing, it can be performed covertly without the cooperation of the subjects. Such an implementation may also help to address the problem of occlusion that affects many of the current computer vision-based approaches to gait identification; this is possible as, except in cases of total occlusion, there is likely to be at least one or more body joints visually available for gait-based identification. As demonstrated in this work, a single joint may provide enough information to identify an individual accurately.

It is also important to note that the related problem of gait identification when unknown people have been detected (i.e., those not in the training set or those that we do not seek to identify) should be addressed to improve the real-world implication of the proposed solution. For example, when searching for missing people, the solution should exclude people not on the missing persons list (i.e., should not attempt to identify them). To the best of the author's knowledge, this is an important area of research that has gained little attention, and therefore, a robust solution is not currently available. One potential solution, OpenMax [242], is an alternative final layer for machine learning models which aims to estimate the probability of a sample belonging to an unknown class. It achieves this by adapting SoftMax, removing the requirement for possibilities over all classes to sum to 1, and including a category for unknown classes [242]. However, implementing and evaluating this would require collecting gait data from additional participants to act as the unknown classes. To the best of the author's knowledge, such a gait dataset is not currently available.

Furthermore, the authors are unaware of any existing works exploring how person identification can be performed using pathological gait (i.e., gait abnormalities caused by pain, reduced range-of-motion, or weakness, for example) or where the person has aged significantly since their gait sample was collected. As current works utilise datasets captured over a relatively brief period, the data provides information regarding the person's gait at the time of recording and, therefore, does not account for future ageing, injury, or illness. Moreover, it has been reported that the time between recording sessions has less impact on identification accuracy than other factors, such as clothing [243]. However, the dataset collected in [243] contains only a six-month gap between recordings, and the authors are unaware of alternative gait datasets with time gaps between recordings. An initial approach to further explore this problem would require the collection of a gait dataset recorded at significant intervals (i.e., years) containing instances of pathological gait. It may then be possible to identify unaffected gait features, which may aid in person identification from pathological gait

data. Solving this problem would potentially create an identification system capable of identifying individuals long after their gait samples have been collected without updating the gait data for each participant.

## 4.4. Summary

The newly proposed gait identification method reported in this section has demonstrated that both the hip and shoulder joint movements on their own and, in combination, possess sufficient information to provide accurate gait-based identification (Obj1, Obj6). To the best of the author's knowledge, a single joint has not been used for gait identification in previous works. These findings present the opportunity for future developments in joint-based gait identification, which could help overcome the occlusion problem in computer vision-based gait identification and, therefore, help reduce false positive identifications. Future work is required to implement this approach in computer vision and to develop the necessary algorithms to make predictions based on the subset of available joints in the event of occlusion.

Using the feature selection approach described in Section 4.1, the 30 most important features from the available 442 are presented in Table 13. Moreover, the most important features contain features from two IMU sensors (gyroscope and accelerometer). Therefore, including various sensors will likely benefit future works, as this is not always applied in the literature. For example, [235] does not use a gyroscope.

The simple filtering method based on the moving average and mode as described in Algorithm 1 (Section 4.1) has improved the accuracy in almost all cases, as shown in Table 17, Table 18, Table 19, Table 20, Table 21, and Table 22. Often, accuracy of 100% is reported by a variety of models when presented with a complete walking sequence (8 metres) per person. Furthermore, it should be noted that 8 metres is a relatively large walking distance to obtain, and it is therefore likely that this amount of data would

not be available in many circumstances, such as security camera footage in a small shop with limited walking space; therefore, the mode metric may be deemed an optional additional step.

# Chapter 5

# Non-Invasive Gait Identification Using Spatiotemporal Biometrics

Chapter 4 showed that spatiotemporal biometrics, specifically joint movement, can identify subjects accurately and consistently. However, the experiments required the use of wearable sensors. Despite being convenient methods of measuring ground truth, these require subject cooperation, which is impossible for applications such as crime suspect or missing person identification. Therefore, a non-invasive form of identification is required which can operate in realistic environments.

Despite the plethora of related works, as reviewed in Chapter 2, several open problems remain. For example, as described in Chapter 2, many existing works are affected by changes in appearance, such as clothing and accessories [70]. Another unresolved challenge in gait identification is occlusion, which may appear in multiple ways, such as self-occlusion, where part of the subject's body is occluding another part of their body. Moreover, subjects may be wholly or partially occluded by other people or aspects of the environment, such as trees or vehicles.

As described in Chapter 2, several works have reported some progress with this problem. For example, in [197], interpolation is used to estimate gait information lost due to occlusion. However, the method requires depth sensors, which are not readily available for real-world applications in dynamic environments. Alternatively, in [244], an occlusion-aware model is presented, which fits a skinned multi-person mesh model to subjects in an occluded input image. However, the reported failure cases

suggest that gait identification performance is limited under heavy occlusion (i.e., when 60% or more of the body is occluded).

Despite the advances described in Chapter 2, to the best of the author's knowledge, the challenges of viewing angle, appearance change, and occlusion remain unsolved for gait identification. In particular, methods for performing HPE and gait identification under temporary complete occlusion, such as when the subject walks behind obstructions such as trees or vehicles, have not been described in the literature. Therefore, this chapter presents a robust, spatiotemporal kinematics-informed, non-invasive gait identification method (STONI-GID). The technique is resilient to viewing angles, appearance change, and occlusion. Moreover, the relevant models were trained and evaluated using our novel dataset, as presented in Chapter 3 and [23].

# 5.1. Methodology

## 5.1.1. Dataset

This work is implemented and evaluated on our novel dataset, as previously described in Chapter 3 and [23]. Moreover, the work described in this chapter is based solely on the video data from each of the three recording experiments described in Chapter 3. No sensor data was used; instead, the skeletons extracted using HPE were used to calculate joint angles.

Moreover, the dataset was partitioned into training and testing subsets, which were processed separately for this work. The training dataset contained 80% of all gait videos per subject per recording angle group (explained in Chapter 3), and the testing dataset included the remaining 20%. The test data partition was processed the same way as the training data was processed, as explained in Chapter 3. However, only the training dataset

was used to train the proposed models, and the testing subset was reserved for evaluation purposes only.

Further evaluation and cross-validation of the proposed method were performed using additional unseen partitions from two datasets, i.e., GRIDDS and Soton Large [46]. The GRIDDS dataset only provides ten walking sequences for each of its 35 participants in indoor environments [44]. In contrast, the Soton Large dataset provides eight walking sequences for approximately 100 participants in indoor and outdoor environments. Cross-dataset validation enables the evaluation of the reliability and generalisability of the proposed approach for purely unseen instances captured in entirely different environments. Further details about the GRIDDS and Soton Large datasets can be found in the original studies [44], [46].

### 5.1.2. Gait Identification Method Overview

Figure 15 demonstrates an overview of the proposed STONI-GID methodology; each component is described in further detail in the following subsections. First, each image frame is extracted from the input video recording. Second, using the OpenPose [75] HPE method, 25 bodily keypoints are extracted from each frame. Next, the position of each occluded keypoint is estimated using Kalman state-estimation filtering [245], [246]. Then, ten joint angles are calculated for the left and right sides, including the elbow, shoulder, hip, knee, and ankle. The data is then sampled and formatted to meet the requirements of the person identification model. Using deep learning, the relevant angle group is detected for each sample. Finally, each angle group has an LSTM model trained to identify the query subject (i.e., person) as an output (e.g., P1, P2, …, P65), and each sample is input to the relevant model, determined by the detected angle. The following subsections describe the stages in further detail.

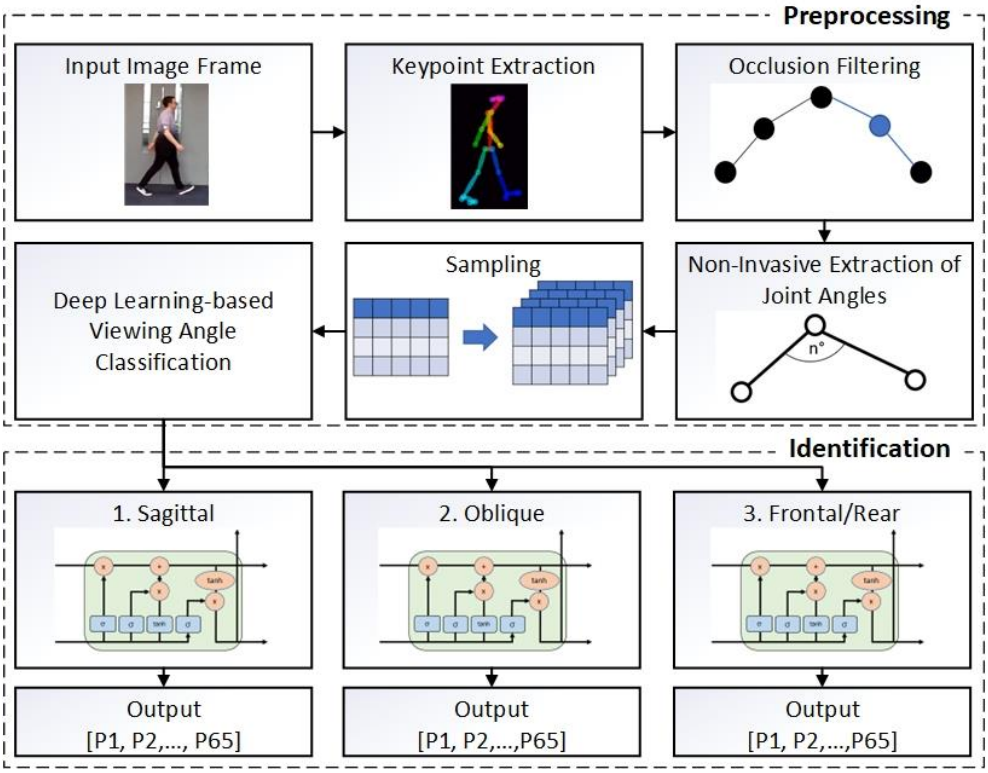


*Figure 15. Overview of the gait identification methodology, including input RGB images, skeleton keypoint extraction, occlusion filtering of any missing keypoints or frames, bodily joint angle extraction, sampling, angle group classification (sagittal, oblique, and frontal/rear), LSTM person identification and the resulting output in terms of participant ID number.*

## 5.1.3. Spatiotemporal Feature Extraction

Using the OpenPose [75] HPE method, 25 bodily keypoints were extracted from each frame. Next, the position of each occluded keypoint was estimated using Kalman state estimation filtering [245], [246]. Then, ten joint angles, including elbow, shoulder, hip, knee, and ankle for both the left and right sides, were calculated as per (8, described in further detail in [247]. Given an angle A → B → C, where A, B, and C are keypoints, and B represents the desired joint to be measured. Multiple works, such as [24], [25], have shown these joints to possess relevant gait information for person identification.

$$\theta = arccos\left(\frac{\overrightarrow{BA}\cdot\overrightarrow{BC}}{\left\|\overrightarrow{BA}\right\|\left\|\overrightarrow{BC}\right\|}\right) \quad (8)$$

The data was then transformed into three-dimensional samples (i.e., several instances were grouped into a sample, where an instance corresponds to all keypoints and features extracted from a single image frame) before being used by the identification model. The length of each sample was equivalent to the average gait cycle length, found to be 0.965 seconds in [24]. This length ensured that each sample contained sufficient patterns regarding each stage of the gait cycle. Next, an LSTM model classified the viewing angle into one of three relevant angle groups (i.e., sagittal, oblique, and frontal/rear). Finally, each angle group has an LSTM model trained to identify the query subject (i.e., person) as an output (e.g., P1, P2, …, P65).

### 5.1.4. Occlusion State Estimation

To overcome the challenging issues that occur under occlusion, an occlusion filtering module based on a Kalman state estimation filter [248] was, for the first time, employed in this study and described in Algorithm 2. State Estimation Filtering for Occluded Samples. This procedure receives a 2D array as its input, where each row represents one image frame and contains 50 columns for the 25 skeletal keypoints (each keypoint has x and y coordinates in separate cells). This procedure iterates through each row of the given array, and for each cell (keypoint coordinate) in the array, state estimation is applied if the cell is null (i.e., occluded). A cell's filtering process is skipped if no occlusion or missing values exist. After each iteration, the filtered row is appended to an array of filtered rows. The filtered rows can then be passed for processing for the gait identification task.

**Algorithm 2. State Estimation Filtering for Occluded Samples**

-Let ***Arr*** be an array of keypoints where one row provides an entire skeleton of keypoints (25).
-Let ***Row*** be a row in ***Arr*** and ***Col*** be a column in ***Arr***
-Let ***Ce*** be a cell in a row of ***Arr***, where a cell contains a coordinate (x or y) of a keypoint
Let ***A***, ***B***, and ***C*** be adaption matrices to convert input states to process states, ***H*** be the conversion matrix, ***K*** be the Kalman gain, $\kappa$ be the iteration number, ***P*** be the process covariance matrix, ***Q*** be the process noise covariance, ***U*** be the control variable, ***W*** be the predicted state noise, ***X*** be the skeletal keypoint state matrix, ***Y*** be the measurement of the keypoint state, ***Z*** be the measurement of noise.
**Inputs**: Array (***Arr***) of keypoints and joint angles
**Outputs**: Filtered array (***F***) of keypoints and joint angles
**Procedure Update()**:

$$\hat{X}_k = AX_{K-1} + BU_K + Wk$$

$$\hat{P}_K = AP_{K-1}A^T + Q_K$$

$$Y_K = CX_K + Z_K$$

**End Procedure**
**Procedure Predict()**:

$$K = \frac{\hat{K}_K H}{H\hat{P}_K H^T}$$

$$X_K = \hat{X}_K + K(Y_K - H\hat{X}_K)$$

$$P_K = (I - KH)\hat{P}_K$$

**End Procedure**
**Procedure**:
  **For** each ***Row*** in ***Arr***:
    **For** each ***Col*** in ***Row***:
      **If** is null (***Ce***):
        $Ce = Predict(X_K, P_K)$
        $X_K, P_K = Update()$
      **Else**:
        Skip occlusion filtering for non-occluded cells
      **End If**
      ***F***.append(***C***)
    **End Loop**
  **End Loop**
**End Procedure**

As discussed in Chapter 2, no existing dataset contains a consistently labelled occlusion. Therefore, it is necessary to simulate the required occlusion in this study, a procedure supported by [196]. Although is in [196]

describes simulated occlusion, it does not guarantee simulation of full-body occlusion. Furthermore, to the best of the authors' knowledge, no existing work has simulated occlusion with guaranteed applicability to full-body occlusion; hence the need to develop an occlusion simulation strategy. When occlusion affects HPE methods, keypoints are usually assigned a null value, so this is one potential method of simulating occlusion. Full-body occlusion has been assumed to evaluate the occlusion filtering model. Thus, each cell in a row of the array containing keypoints for the entire body is assigned a null value. Four different levels of occlusion have been simulated in this work.

*Table 23. Description of the four varieties of simulated occlusion implemented during the evaluation of the occlusion state estimation module.*

| Occlusion Profile | Block Size | No. Occluded Rows | No. Standard Rows | Description |
|---|---|---|---|---|
| 1 in 2 | 2 | 1 | 1 | Simulates reduced frame rate. Every 2nd row is nullified. |
| 1 in 10 | 10 | 1 | 9 | Simulates reduced frame rate. Every 10th row is nullified. |
| 5 in 10 | 10 | 5 | 5 | Simulates environmental occlusion. Five rows are nullified for every ten rows (half of all rows). |
| 10 in 20 | 20 | 10 | 10 | Simulates environmental occlusion. Ten rows are nullified for every 20 rows (half of all rows). |

Table 23 describes the four varieties of simulated occlusion implemented to evaluate the occlusion filtering method. The array of keypoints for each type is partitioned into blocks given by the corresponding block size. The first n rows are filled with null values, with the remaining rows unchanged. The 1 in 2 and 1 in 10 profiles simulate light occlusion or reduced frame rates, where the occlusion effects are brief but regular. Meanwhile, the 5 in 10 and 10 in 20 simulate heavier occlusion, where the entire body is

regularly lost from the image for a substantial number of frames. To demonstrate the recovery of missing samples, occlusion is simulated before the joint angles are calculated and appended to the rows. This means that joint angles are not calculated, just as they would not be calculable if the keypoints were undetectable during true occlusion.

### 5.1.5. Kinematics-Informed Deep Angle Detection

Gait identification is affected by the viewing angle [72]. Therefore, it is impossible to perform gait identification without either multiple identification models or pre-processing to remove the effect of the angle. For the first time, this study takes an initiative towards non-restraining perspectives of gait identification, which is required in most realistic cases. Through systematic testing, we found that gait identification could be robustly performed when the gait data was divided into three groups based on the viewing angles (i.e., camera perspectives).

To perform the testing, an LSTM model was trained to classify the data, using the configuration shown in Table 25. First, the model was trained using the data from all viewing angles. However, the model was unable to identify participants. Next, the data was split into two groups, one containing the frontal and rear angles (at 0° and 180°, respectively) and the other containing the remaining angles. Although the model trained on the frontal/rear group performed well, the model trained on the second group performed poorly. Finally, the larger group was split into two groups. One group contained the sagittal viewing angles (90° and 270°) and the other the oblique viewing angles (45°, 135°, 225°, and 315°).

Therefore, three groups were finalised for testing. Group 1 contained the sagittal profile walks, captured from 90° and 270° in relation to the subject. Group 2 included the oblique walks captured at 45°, 135°, 225°, and 315°. Group 3 contained the frontal and rear views captured at 0° and 180°, respectively.

A description of the LSTM model configuration for angle detection is provided in Table 24. The model receives 60 input features, the first 50 corresponding to x and y coordinates for each of the 25 keypoints and the remaining 10 corresponding to the joint angles listed previously (Section 5.1.2. Gait Identification Method Overview). The model contains five layers, including 2 LSTM layers, one dropout layer, and two dense layers, with the final layer outputting the relevant viewing angle classification. The first LSTM layer has an output shape of 29,100, corresponding with the sequence length of 29, matching the average gait cycle length when measured as the number of data points, as described previously. The model was trained with the training dataset described in Chapter 3 Primary Dataset Compilation. For training, 50 epochs were performed, with a batch size of 5. This process was repeated ten times for cross-validation. The model was then evaluated with the purely unseen reserved testing dataset described in Section 4.1. Detailed statistical results of the proposed LSTM-based angle detection are provided in Section 5.2. Results of Non-Invasive Gait Identification.

*Table 24. LSTM model configuration for angle classification.*

| Model Layer | Output Shape | No. of Parameters |
|---|---|---|
| LSTM | 29,100 | 64400 |
| LSTM | 100 | 80400 |
| Dropout | 100 | 0 |
| Dense | 100 | 10100 |
| Dense | 3 | 100 |

## 5.1.6. Gait-based Person Identification Model

A description of the LSTM model configuration for gait-based person identification is provided in Table 25. The configuration shown in Table 25 was empirically selected due to its success in our previous work, the full details of which are available in [25]. As with the angle classification model described in Section 5.1.5, the gait identification model receives 60 input features, the first 50 corresponding to x and y coordinates for each of the 25

keypoints and the remaining 10 corresponding to the joint angles listed previously. As with the angle classification model, this model also contains five layers, comprising 2 LSTM layers, one dropout layer, and two dense layers, with the final layer outputting the relevant viewing angle classification. Again, the first LSTM layer has an output shape of 29,100, corresponding with the sequence length of 29, matching the number of data points equivalent to the average gait cycle length, as described previously.

*Table 25. LSTM model configuration for gait-based person identification.*

| Model Layer | Output Shape | No. of Parameters |
|---|---|---|
| LSTM | 29,100 | 64400 |
| LSTM | 100 | 80400 |
| Dropout | 100 | 0 |
| Dense | 100 | 10100 |
| Dense | 65 | 1010 |

As described in Section 5.1.2 and Figure 15, three LSTM models exist for person identification, one for each viewing angle group. Each of the three models was trained with the training dataset described in Section 4.1. However, each model was trained only with samples relevant to the viewing angle group, and the remainder were excluded. For training, 50 epochs were performed, with a batch size of 5. This process was repeated ten times for evaluation. The models were then evaluated with the reserved unseen testing dataset described in Section 5.1.1. The results of this process are provided in 5.2. Results of Non-Invasive Gait Identification.

### 5.1.7. Experimental Design

This chapter describes four experiments. The first aims to identify the appropriate angle group for a video sequence, as per the angle groups described in Section 3.1.2. After detecting the viewing angle, the second experiment aims to identify subjects in a video sequence. The third experiment investigates the effect of occlusion on the newly proposed

STONI-GID. The fourth experiment investigates the impact of appearance change on STONI-GID.

#### *5.1.7.1. Experiment 1 (Exp.1) Viewing Angle Classification*

In Exp.1, the viewing angle of an input video sequence was categorised as sagittal, oblique, or frontal/rear, as per Section 3.1.2. The classification was performed using an LSTM model, as described in Table 24.

#### *5.1.7.2. Experiment 2-A (Exp.2-A) Sagittal Angle Group*

Exp 2-A was a gait identification experiment implemented by training an LSTM model, as described in Table 25, on video sequences identified as belonging to the sagittal angle group.

#### *5.1.7.3. Experiment 2-B (Exp.2-B) Oblique Angle Group*

Exp 2-B was a gait identification experiment implemented by training an LSTM model, as described in Table 25, on video sequences identified as belonging to the oblique angle group.

#### *5.1.7.4. Experiment 2-C (Exp.2-C) Frontal/Rear Angle Group*

Exp 2-C is a gait identification experiment implemented by training an LSTM model, as described in Table 25, on video sequences identified as belonging to the Frontal/Rear angle group.

#### *5.1.7.5. Experiment 3-A (Exp.3-A) Without Occlusion Filtering*

Exp.3-A demonstrated the effect of simulated occlusion on the gait identification models developed during Exp.2-A, Exp.2-B, and Exp.2-C. The methodology followed that described in Figure 15. However, the occlusion filtering model was not included to demonstrate the negative effect of occlusion on gait identification performance.

#### *512.7.6. Experiment 3-B (Exp.3-B) With Occlusion Filtering*

Exp.3-B demonstrated the ability of the occlusion filter to recover the performance lost in Exp.3-A. The methodology followed that described in Figure 15, including the occlusion filtering model.

#### *5.1.7.7. Experiment 4 (Exp.4) Alternative Appearance*

The models trained during Exp.2-A, Exp.2-B, and Exp.2-C were tested using instances of the subjects with alternative appearances (i.e., alternative clothing). The performance was compared to that achieved during Exp.2-A, Exp.2-B, and Exp.2-C to ascertain whether the proposed STONI-GID method was significantly affected by a change in appearance.

#### *5.1.7.8. Experiment 5 (Exp.5) Recording Environments*

The models trained during Exp.2-A, Exp.2-B, and Exp.2-C were tested separately indoors and outdoors. The performance was compared to ascertain whether the proposed STONI-GID method was significantly affected by swapping between indoor and outdoor environments.

#### *5.1.7.9. Experiment 4 (Exp.6) Cross-Dataset Validation*

The STONI-GID models were trained and validated using two additional datasets: GRIDDS [44] and Soton Large [46], as per Exp.2-A, Exp.2-B, and Exp.2-C. This test aimed to demonstrate the proposed STONI-GID method on additional gait datasets.

# 5.2. Results of Non-Invasive Gait Identification

The following subsections provide comprehensive results of the proposed gait identification method (STONO-GID) described in Section 5.1.2. Gait Identification Method Overview. This includes experimental outcomes of the proposed kinematics-informed angle detection model (Sections 5.1.7.1. Experiment 1 (Exp.1) Viewing Angle Classification to 5.1.7.4. Experiment 2-C (Exp.2-C) Frontal/Rear Angle Group), gait identification under normal circumstances (Section 5.1.7.5. Experiment 3-A (Exp.3-A) Without Occlusion Filtering), and gait identification results with and without the gait occlusion state estimation module under simulated occlusion (Section 512.7.6. Experiment 3-B (Exp.3-B) With Occlusion Filtering). Furthermore, the results of experiments using alternative clothing (Section 5.1.7.7. Experiment 4 (Exp.4) Alternative Appearance) and alternative environments (Section 5.1.7.8. Experiment 5 (Exp.5) Recording Environments) are also described in this section. Finally, the results of cross-validation performed on two additional datasets, SOTON Large [46] and GRIDDS [44], are provided (Section 5.2.6.1 Validation on the SOTON large Dataset and

5.2.6.2. Validation on the GRIDDS Dataset).

## 5.2.1. Viewing Angle Detection

The results of the proposed angle detection model (as described in Section 5.1. Methodology and 5.1.7.1. Experiment 1 (Exp.1) Viewing Angle Classification) are shown in Table 26. The mean and standard deviation of 10 repetitions of training and evaluation are provided. A high level of accuracy, 98.87%, was achieved when evaluating the viewing angle group classifier on the unseen partition of our primary Dataset.

The need for an angle detection module is highlighted by recent literature, such as [61], which indicates that viewing angles affect the

performance of gait identification. Therefore, without a method of solving the problems introduced by viewing angle, gait identification performance in real-world environments will likely be limited to more controlled environments such as airports. However, by training an angle detection model and gait identification models for each group of angles, we can overcome the viewing angle issues, as reported in the following subsections.

*Table 26. Statistical results of the proposed angle detection model were validated over a purely unseen partition of our dataset.*

| Accuracy | F1 | Kappa | Precision | Recall |
|---|---|---|---|---|
| 98.61% | 98.54% | 97.89% | 98.64% | 98.47% |
| 98.75% | 98.68% | 98.10% | 98.82% | 98.56% |
| 98.70% | 98.66% | 98.03% | 98.78% | 98.56% |
| 98.85% | 98.83% | 98.25% | 98.90% | 98.76% |
| 99.18% | 99.15% | 98.76% | 99.21% | 99.10% |
| 98.80% | 98.74% | 98.18% | 98.86% | 98.64% |
| 99.57% | 99.56% | 99.34% | 99.58% | 99.55% |
| 98.89% | 98.87% | 98.32% | 98.92% | 98.84% |
| 98.99% | 98.94% | 98.47% | 99.04% | 98.86% |
| 98.37% | 98.25% | 97.52% | 98.42% | 98.13% |
| **98.87% (± 0.33)** | **98.82% (± 0.35)** | **98.29% (± 0.50)** | **98.92% (± 0.32)** | **98.75% (± 0.38)** |

## 5.2.2. Gait Identification Experiment

The statistical results of the three gait identification models (i.e., Exp.2-A oblique view, Exp.2-B sagittal view, and Exp.2-C frontal/rear view) described in Sections 5.1.7.2. Experiment 2-A (Exp.2-A) Sagittal Angle Group to 5.1.7.4. Experiment 2-C (Exp.2-C) Frontal/Rear Angle Group with their corresponding angle groups, evaluated on our primary dataset, are shown in Table 27, where the results, when tested with all gait samples for all angle groups combined, are shown in the final row. It can be seen that all models indicate robust performance and, importantly, similar outcomes for all three groups of angles. However, the group containing the front and rear-facing cameras (0° and 180°) performs slightly better. This indicates that this angle may be beneficial for identifying subjects, perhaps due to visible features (such as

stride width) that are not easily extracted when using 2D HPE at other angles.

*Table 27. Results from the gait identification model for each angle group.*

| Angle Group | Accuracy | F1 | Kappa | Precision | Recall |
|---|---|---|---|---|---|
| 1. Sagittal | 98.29% (± 0.22) | 98.26% (± 0.24) | 98.26% (± 0.22) | 98.34% (± 0.23) | 98.25% (± 0.24) |
| 2. Oblique | 98.24% (± 0.22) | 98.39% (± 0.24) | 98.21% (± 0.22) | 98.60% (± 0.23) | 98.40% (± 0.28) |
| 3. Frontal/ Rear | 99.88% (± 0.04) | 99.88% (± 0.04) | 99.87% (± 0.04) | 99.89% (± 0.04) | 99.88% (± 0.05) |
| **Combined** | **98.66% (± 0.18)** | **98.73% (± 0.19)** | **98.64% (± 0.18)** | **98.83% (± 0.18)** | **98.73% (± 0.21)** |

## 5.2.3. Gait Identification with Occlusion Filtering

The statistical outcomes of gait identification under the simulated occlusion conditions, as described in Section 5.1.4. Occlusion State Estimation, are shown in Table 28. It is important to note that the results shown in Table 28 are collected from our combined dataset (all angle groups) without applying the proposed occlusion state estimation model described in Section 5.1 and Exp.3-A. As expected, Table 28 clearly shows that the simulated occlusion imposes significantly negative impacts on the gait identification performance, with varying levels of impact depending on the severity of the occlusion level. Moreover, the 1 in 2 occlusion method, where every second frame is missing, has the most significant effect, likely due to it having the greatest number of rows affected. In comparison, the 1 in 10 occlusion method has the most negligible impact on results, likely because it affects the fewest rows in the data. The 5 in 10 and 10 in 20 methods reduce the person identification performance more than the 1 in 10 method, likely because more frames are impacted. Similarly, both the 5 in 10 and 10 in 20 methods affect performance less than the 1 in 2 method, likely because fewer frames are impacted, and both profiles leave consecutive frames unaffected, which also occurs in the 1 in 10 method. This variety of occlusion

methods enables the demonstration of the performance effects for differing levels (i.e., heavy or light) and styles of occlusion (i.e., single frames or blocks of frames), which may be encountered in real-world applications.

Such findings are significant as they demonstrate that real-world challenges such as occlusion can drastically reduce the performance of person identification and other similar tasks. Moreover, the previously described articles [1]–[3] report poor real-world performance of person identification by police forces. Therefore, this is a significant challenge to overcome to achieve robust person identification and will have clear benefits in the security domain, such as for the police.

*Table 28. Results of gait identification with simulated occlusion, indicating the adverse effect suffered without state estimation.*

| Occlusion Method | Accuracy | F1 | Kappa | Precision | Recall |
|---|---|---|---|---|---|
| 1 in 2 | 42.71% | 43.09% | 43.92% | 34.04% | 41.65% |
| 1 in 10 | 79.77% | 80.08% | 80.93% | 72.90% | 78.37% |
| 5 in 10 | 61.04% | 60.89% | 62.00% | 54.18% | 59.86% |
| 10 in 20 | 52.27% | 53.79% | 53.70% | 52.68% | 52.07% |

To validate the effectiveness of the newly proposed STONI-GID in occluded gait, the statistical outcomes of gait identification under the simulated occlusion conditions, with occlusion-state-estimation filtering implemented, are shown in Table 29 and described in Exp.3-B. The results significantly improve on those shown in Table 28 (without the proposed filtering solution), suggesting that our occlusion-state-estimation model successfully improves gait identification under occluded conditions. For instance, without the state estimation model and under the 1 in 2 occlusion simulation, gait identification accuracy dropped substantially from 98.29% (as shown in Table 27) to 42.71% (as shown in Table 28). However, Table 29 shows that with the inclusion of the state-estimation model, the accuracy recovers to 69.26%. Similarly, for the 1 in 10 occlusion method, accuracy drops to 79.77% (Table 28) and recovers to 95.17% with state-estimation filtering (Table 29). Correspondingly, the 5 in 10 occlusion method causes identification accuracy to drop to 61.04% (Table 28), and state-estimation

filtering restores the performance to 82.98% accuracy (Table 29). Finally, the 10 in 20 occlusion method causes identification accuracy to drop to 52.27% (Table 28), and state estimation filtering restores the performance to 69.98% accuracy (Table 29).

*Table 29. Results of gait identification with simulated occlusion with state estimation filtering applied to alleviate the effects.*

| Occlusion Method | Accuracy | F1 | Kappa | Precision | Recall |
|---|---|---|---|---|---|
| 1 in 2 | 69.26% | 69.25% | 70.72% | 69.14% | 69.15% |
| 1 in 10 | 95.17% | 96.88% | 96.61% | 96.67% | 94.89% |
| 5 in 10 | 82.89% | 83.76% | 84.09% | 81.79% | 82.14% |
| 10 in 20 | 69.98% | 71.50% | 71.41% | 70.39% | 69.78% |

### 5.2.4. Gait Identification with Alternative Appearance

As discussed previously, appearance has been shown to affect the accuracy of person identification methods [61], [83], including some forms of gait identification. However, unlike methods such as GEI, skeleton-based approaches to gait identification focus on the location of bodily keypoints, rather than appearance. Therefore, if keypoints are identified accurately regardless of appearance, then the appearance of a subject should not impede their identification when using skeleton-based gait identification methods [76]. Existing available datasets did not provide alternative appearances to investigate this hypothesis sufficiently; however, (as described in Section 3), our dataset [23] provides additional recordings where participants change their appearance, such as by wearing a coat or jacket on top of their standard clothing.

For the first time, this study investigates the effectiveness of a newly proposed gait identification approach for standard and alternative appearance, as in Exp.4. We used the previously unseen partition of our dataset [23], further portioning appearance into standard and alternative for comparison. Table 30 presents the average gait identification results. The evaluation set was partitioned into standard and alternative appearances and evaluated using the ten iterations of the identification models described

previously (Section 5.1). Table 30 shows that the average identification accuracy for the standard appearance is 98.34%, and for the alternative appearance, it is 98.02%, almost similar. Furthermore, a two-sample t-test produces a p-value of 0.069, assuming an alpha value of 0.1; this suggests no statistical significance between the means of both experiments. Therefore, it concludes that the proposed approach of the kinematics-informed deep identification model is not significantly affected by the subject's appearance changes.

*Table 30. Results of gait identification with subjects wearing alternative clothing.*

| Appearance | Accuracy | F1 | Kappa | Precision | Recall |
|---|---|---|---|---|---|
| Standard | 98.34% (± 0.20) | 98.57% (± 0.21) | 98.60% (± 0.20) | 98.71% (± 0.21) | 98.52% (± 0.21) |
| Alternative | 98.02% (± 0.24) | 97.93% (± 0.28) | 97.98% (± 0.26) | 98.10% (± 0.28) | 97.89% (± 0.27) |

Moreover, the finding that STONI-GID can resolve appearance change is significant as previous works have identified appearance change as a challenge to gait identification [76]. The challenge significantly impacts real-world applications; for example, for crime suspect identification, subjects may change their clothing to disguise themselves. Furthermore, to the best of the authors' knowledge, this is the first work to demonstrate successful gait identification with no significant statistical difference between appearance changes. This is a substantial initiative towards a reliable, non-restraining, non-invasive solution to person identification.

### 5.2.5. Gait Identification in Realistic Environments

One of the advantages of our primary dataset [23] is that it includes examples of indoor and outdoor real-world environments. Previous works have shown that performance may decrease outdoors [249], [250], likely due to the uncontrolled nature such environments, and sensors such as depth cameras may be challenging to access outdoors [251]. Therefore, this

dataset provides an opportunity to investigate the proposed STONI-GID's performance using indoor and outdoor data, as in Exp.5. Table 31 displays the average results of the gait identification, where the evaluation set was partitioned into indoor and outdoor recordings and evaluated using ten iterations of the identification models, as described previously. Overall, a 98.98% accuracy was achieved indoors, whereas 97.95% accuracy (with no significant difference) was achieved outdoors. Furthermore, a two-sample t-test produced a p-value of 0.09, assuming an alpha value of 0.1; suggesting that there is no statistical significance between the means of both experiments. Therefore, it may be suggested that the recording environment does not significantly affect STONI-GID.

*Table 31. Results of gait identification using indoor and outdoor recordings.*

| Appearance | Accuracy | F1 | Kappa | Precision | Recall |
|---|---|---|---|---|---|
| Indoors | 98.98% (± 0.20) | 98.45% (± 0.22) | 98.85% (± 0.20) | 98.64% (± 0.20) | 98.55% (± 0.21) |
| Outdoors | 97.95% (± 0.28) | 98.30% (± 0.30) | 98.32% (± 0.29) | 98.37% (± 0.30) | 98.30% (± 0.30) |

To reiterate, gait identification in outdoor environments has been described as a particular challenge to real-world person identification [61]. The challenge significantly impacts real-world applications; for example, subjects may move from indoor to outdoor environments in the case of crime suspect identification. Moreover, to the best of the author's knowledge, this is the first work to have statistically compared the performance of gait identification in indoor and outdoor environments. The finding that STONI-GID is not statistically affected by environments is another crucial step towards a reliable, non-restraining, non-invasive solution to person identification.

## 5.2.6. Cross-Dataset Validation

In addition to test data (5.1.1. Dataset), we employed cross-dataset validation to evaluate the proposed STONI-GID's performance and ability to generalise across multiple datasets, each representing different scenarios, environments, and data sources. For this purpose, we validated STONI-GID on the SOTON Large dataset [46] and the GRIDDS dataset [44] while assuring the same training and evaluation method described in Section 5.1 (5.1.7.9. Experiment 4 (Exp.6) Cross-Dataset Validation). Table 32 (Section 5.2.6.1) and Table 34 (Section 5.2.6.2) show the statistical results achieved using the SOTON Large [46] and the GRIDDS datasets [44], respectively.

### *5.2.6.1 Validation on the SOTON large Dataset*

The SOTON large dataset [46] is a gait dataset comprising 115 participants recorded from 2 angles (including side and oblique views). Table 32 presents the statistical results of the proposed STONI-GID trained over totally different data (our primary dataset [23]) using the methodology described in Section 5.1. Methodology. The results in Table 32 include relevant metrics for each of the ten models and the mean and standard deviation of each metric for each model trained for validation purposes (see Section 5.1), with the models achieving 97.68% accuracy on average.

*Table 32. Results of the STONI-GID gait identification method validated on the SOTON Large dataset.*

| Run Number | Accuracy | F1 | Kappa | Precision | Recall |
|---|---|---|---|---|---|
| 1 | 99.13% | 99.02% | 98.65% | 99.03% | 99.08% |
| 2 | 96.27% | 96.25% | 95.30% | 96.63% | 97.49% |
| 3 | 95.55% | 96.08% | 94.45% | 96.58% | 94.75% |
| 4 | 99.13% | 99.12% | 98.65% | 99.57% | 98.86% |
| 5 | 97.70% | 97.73% | 96.97% | 98.21% | 97.53% |
| 6 | 97.34% | 97.06% | 94.55% | 97.14% | 97.05% |
| 7 | 99.49% | 99.45% | 99.07% | 99.55% | 99.40% |
| 8 | 91.61% | 91.62% | 91.81% | 92.15% | 91.07% |
| 9 | 99.85% | 98.81% | 99.49% | 99.82% | 99.80% |
| 10 | 98.77% | 97.77% | 98.23% | 96.90% | 98.74% |
| **Average** | **97.68% (± 0.02)** | **97.29% (± 2.32)** | **96.71% (± 2.56)** | **97.56% (± 2.30)** | **97.38% (± 2.66)** |

Table 33 compares our best-performing STONO-GID model (out of the ten samples trained on our dataset) with relevant results reported in the literature. Table 33 suggests that KNNs are a prevalent approach to gait identification on the Soton Large dataset [46]. However, performance varies significantly between works, despite them using the same data. However, the relatively low performance reported by [252] is due to this early work only utilising height and footprint features. The statistical comparison clearly shows that the proposed STONI-GID approach outperforms all the other relevant models identified which were validated on the SOTON Large dataset. Specifically, STONI-GID achieved 97.68% accuracy on the SOTON Large dataset [46], outperforming the next best-performing work [253], which achieved 96.7% accuracy on the same dataset.

*Table 33. Reported gait identification results in other works validated on the SOTON Large dataset.*

| Ref | Year | Model / Algorithm | Accuracy |
|---|---|---|---|
| [252] | 2012 | KNN | 58.20% |
| [253] | 2011 | KNN | 96.70% |
| [254] | 2017 | SVM | 93.00% |
| [255] | 2016 | KNN | 95.70% |
| [256] | 2010 | KNN | 95.80% |
| **STONI-GID** | **2023** | **LSTM** | **97.68%** |

#### *5.2.6.2. Validation on the GRIDDS Dataset*

The GRIDDS dataset [44] contains 35 subjects recorded from a single angle (sagittal). The dataset provides 350 instances and contains depth information and RGB video. Table 34 presents the statistical results achieved using our STONI-GID approach, which was trained using the methodology described in Section 5.1. The results presented in Table 34 include the results of all ten models and the mean and standard deviation of each metric from the models trained for validation purposes, with the models achieving 99.12% accuracy on average. The improvement of the accuracy when using GRIDDS as compared to using our dataset (in Table 26) or SOTON large (in Table 27) may be at least partly due to the lower number of subjects or simplicity of experimental data (captured from a single perspective).

However, to the best of the author's knowledge, no existing gait identification method has been validated on the GRIDDS dataset.

The results presented in Table 32, Table 33, and Table 34 are important, as previous works such as [257] have highlighted model generalisation as a challenge to robust person identification in real-world environments, likely resulting in false-positive identifications. The results reported in Table 32 suggest that such issues do not limit the STONI-GID and can be generalised across additional datasets.

Despite the advances, one limitation of this work is a lack of a gait dataset containing labelled examples of real-world occlusion. This work has endeavoured to simulate such occlusion. However, future work with real-world examples would be necessary to validate STONI-GID and other related methods further.

*Table 34. Results of STONI-GID gait identification method validated on the GRIDDS dataset.*

| Run Number | Accuracy | F1 | Kappa | Precision | Recall |
|---|---|---|---|---|---|
| 1 | 98.63% | 98.64% | 98.18% | 98.66% | 98.66% |
| 2 | 98.63% | 98.64% | 98.18% | 98.66% | 98.66% |
| 3 | 99.32% | 99.33% | 99.09% | 99.36% | 99.30% |
| 4 | 97.95% | 97.98% | 97.27% | 97.97% | 98.01% |
| 5 | 99.36% | 99.49% | 99.00% | 99.30% | 99.33% |
| 6 | 99.32% | 99.31% | 99.09% | 99.30% | 99.33% |
| 7 | 99.66% | 99.65% | 99.54% | 99.67% | 99.64% |
| 8 | 99.66% | 99.66% | 99.54% | 99.70% | 99.63% |
| 9 | 99.32% | 99.35% | 99.09% | 99.40% | 99.30% |
| 10 | 99.32% | 99.32% | 99.09% | 99.30% | 99.33% |
| **Average** | **99.12% (± 0.54)** | **99.14% (± 0.54)** | **98.81% (± 0.71)** | **99.13% (± 0.54)** | **99.12% (± 0.51)** |

# 5.3. Discussion

To the best of the author's knowledge, this is the first work that addresses multiple unresolved challenges of gait identification, particularly occluded gait and realistic environments. We introduce a novel methodology as an initiative towards non-invasive and non-restraining gait identification, which

is highly demanding in the real world. For this purpose, our method extracts body joint angles using purely automated computer vision methods, fed to a composite of deep learning (LSTM) and state estimation (Kalman filter) to handle the dynamics of real-time gait identification. We report high identification accuracy in evaluating the newly proposed STONI-GID over our dataset [23], as well as for cross-dataset validation on the SOTON Large [46] and GRIDDS [44] datasets, with 98.66%, 95.48%, and 99.12% accuracy, respectively.

Moreover, as reported in Table 27, gait identification performance was slightly higher for the frontal/rear angle group than for the sagittal and oblique groups. This is likely due to the availability of gait features such as stride width, which are not easily extracted from other viewing angles when using 2D HPE. Therefore, this result indicated that the frontal/rear angle may be beneficial for gait identification methods in the future in applications where the viewing angle can be controlled; for example, in more controlled environments such as airports. Further research is required to identify the cause of this increased performance; however, to the best of the authors' knowledge, only our dataset [23] provides the relevant viewing angles. Furthermore, future research may attempt to obtain such benefits, for example, the extraction of stride width, without the need to adjust the viewing angle; 3D HPE may present an opportunity to achieve this.

Additionally, we show that, unlike some existing works, our approach is not significantly affected by the subject's appearance or the recording environment. Furthermore, we present a method of minimising the impact of occlusion using state estimation to estimate missing keypoints, which was shown to significantly reduce the loss of accuracy under occlusion. However, future work is required to minimise this lost accuracy further.

Although our gait dataset improves on numerous aspects of existing datasets, it could be further enhanced by recording and adding additional participants. Moreover, additional environments would be a further benefit, particularly including examples of real-world occlusion, which would remove the need to simulate occlusion as performed in this work. To the best of the

author's knowledge, no existing dataset provides real-world examples of occluded gait. Furthermore, a larger dataset could give further insights into gait; for example, clustering may provide some insights into types or categories of gaits, which could help to improve gait identification and may also have applications in gait analysis. Similarly, further research should identify whether the three angle groups described in this work (sagittal, oblique, and frontal/rear) are sufficient to resolve all viewing angles; such research would require additional data recorded from different viewing angles. Furthermore, if it is found that STONI-GID does not resolve all viewing angles, then the technical challenge of neutralising the viewing angle should be addressed. 3D, rather than 2D, HPE may provide a potential route to removing the effect of viewing angle.

Furthermore, we show that our approach was able to restore much of the performance lost due to occlusion. Related works such as [20] and [196] have also addressed occlusion. In [20], no attempt is made to recover information lost during occlusion. Instead, an occlusion awareness module is implemented to insert occlusion information into the gait identification backbone, alongside the available gait features. However, unlike [196] and our work, [20]does not demonstrate the performance effects of their simulated occlusion method.

In contrast, [196] attempts to estimate the occluded information by performing pixel-wise recovery of the gait silhouette from temporal information in the previous non-occluded frames. However, the method is computationally expensive and, therefore, not appropriate for real-time applications. Furthermore, the proposed method only recovers approximately 20% of the lost performance, even under only partial occlusion. Whereas our proposed method was able to recover 27% under the more challenging event of full-image occlusion.

Moreover, unlike [20] and [196], we demonstrated our approach using the more challenging problem of total occlusion, but it can also be applied to partial occlusion. Therefore, our proposed approach has the advantage of also being applicable to related problems such as heavily

reduced framerates. To the best of the authors' knowledge, this is the first work to address temporary total image occlusion.

This work has the potential to improve non-invasive person identification in unconstrained environments. Furthermore, it can potentially impact domains such as security, where existing applications have been shown to provide unacceptably high levels of false positive identifications [1]–[3]. Despite the advances reported in the literature, this work, to the best of the authors' knowledge, is the first to overcome the real-world problems of viewing angle, appearance change, and occlusion in a single method.

## 5.4. Summary

This chapter has introduced STONI-GID, a novel gait-based person-identification method. The method includes a deep-learning-based method of viewing angle detection, as described in Section 5.1. Our empirical experiments demonstrated that some viewing angles could be grouped, thus simplifying the model. Table 26 reports a 98.87% average accuracy for the angle detection model.

Furthermore, expanding on the method described in Chapter 4, STONI-GID extracts spatiotemporal biometric features from video sequences, thus providing non-invasive identification (Obj7). The method was 98.66% accurate across all viewing angles, as reported in Table 27, and was demonstrated not to be significantly affected by viewing angles, as described in Section 5.2. Moreover, the dataset was further validated on the SOTON Large [46] and GRIDDS [44] datasets, as reported in Table 32 and Table 34, respectively, indicating state-of-the-art performance. Unlike many existing works, STONI-GID was shown not to be significantly affected by appearance change or the recording environment, as reported in Table 30 and Table 31, respectively (Obj8).

Moreover, STONI-GID makes significant advances towards solving the important and difficult challenge of overcoming occlusion in gait identification. As reported in Section 5.1, a filtering module that recovers skeletal measurements (i.e., keypoints and joint angles) missing due to occlusion has been presented. As mentioned in Chapter 2, no existing dataset contains labelled examples of gait occlusion. Therefore, simulated occlusion was used as described in Section 5.1. The results reported in Table 28 indicate the efficacy of the simulated occlusion and its effect on gait identification performance when efforts are not made to recover lost data. Upon activating the occlusion filtering module of STONI-GID (Section 5.1), Table 29 indicates a significant performance recovery.

Despite the advances reported in this chapter, it would benefit the research community to collect further data containing labelled examples of dynamic real-world occlusion. Such a dataset could further validate the ability of STONI-GID to overcome occlusion. Similarly, Table 29 indicates that although STONI-GID can significantly recover performance lost due to occlusion, this recovery is not complete. Future work may investigate alternative filtering methods or enhancing partial information in the case of partial occlusion.

# Chapter 6

# Conclusions and Future Work

Skeleton-based gait identification is a promising method of person identification. It may be conveniently performed non-invasively from a distance, making it suitable for essential security applications such as identification of crime suspects and missing persons. However, the evidence from the literature suggests limitations in terms of existing datasets and identification methods, which fail to meet the challenges of viewing angles, appearance change, and occlusion.

In this thesis, we introduced STONI-GID, a robust skeleton-based gait identification method. Moreover, we demonstrated that it is resilient to dynamic real-world factors, such as viewing angles, realistic settings, appearance change, and occlusion. Additionally, we introduced a novel diverse gait dataset recorded using multiple cameras and sensors.

## 6.1. Research Outcomes

We commenced this research by reviewing the gait identification literature and the HPE theory, which underpins the skeleton-based approach to gait identification. Based on the findings of this research, we designed a skeleton-based gait identification method using deep learning and spatiotemporal biometrics.

### 6.1.1. Diverse and Multi-Modal Gait Dataset

The literature indicated that many of the existing gait datasets are limited by a lack of diversity in terms of the participants, appearance, viewing angle, environments, annotations, and availability (RQ1, Obj2). Therefore, we presented a primary gait dataset comprising 1,560 annotated casual walks from 65 participants in indoor and outdoor real-world environments (Obj3), as presented in [23]. We used two digital cameras and a wearable digital goniometer to capture visual and motion signal gait data, respectively. Traditional methods of gait identification are often affected by the viewing angle and appearance of the participant; therefore, this dataset purposely considers diversity in various respects (e.g., participants' attributes, variations in background, and view angles). The dataset was captured from 8 viewing angles in 45° increments, along with alternative appearances for each participant, such as by a change of clothing. The dataset provides 3,120 videos containing approximately 748,800 image frames with detailed annotations, including approximately 18,720,000 bodily keypoint annotations, identifying 75 keypoints per video frame, and approximately 1,026,480 motion data points captured from a digital goniometer for three limb segments (thigh, upper arm, and head) (Obj4, Obj5).

### 6.1.2. Gait Identification using Limb Joint Movement

Before exploring non-invasive methods of gait identification, a sensor-based implementation was developed to explore several aspects of the research (Obj6), as described in [24], [25]. In this work, we proposed a gait identification approach which uses limb joint motion data and deep machine learning models to identify the individuals. Distinct statistical features were identified and extracted from limb movement using a fixed-width sliding window to train a Long Short-Term Memory model. Moreover, we identified the 30 most important gait features from the 442 generated and further

showed that a single complete gait cycle was sufficient for identification (Obj5). Furthermore, the proposed solution outperforms the existing sensor-based methods, producing 98.87% accuracy when evaluated over unseen samples (RQ1). In addition, we proposed a simple two-stage filtering approach to increase the prediction accuracy up to 100% when larger data sequences (i.e., multiple complete gait cycles) are available for a subject.

### 6.1.3. Non-Invasive Gait Identification in Realistic Environments using Deep Learning and Spatiotemporal Biometrics

We introduced a robust spatiotemporal kinematics-informed non-invasive gait identification method (STONI-GID) using computer vision, human pose estimation, occlusion state estimation, and deep machine learning (Obj7). Furthermore, unlike some existing methods, we show that our method remains unaffected by appearance change, environment, or viewing angle (RQ3, Obj8). Our approach achieved identification accuracy of up to 98.66% when evaluated using our primary dataset in real-world environments (RQ2). Moreover, the model achieved state-of-the-art performance for cross-dataset validation over the SOTON Large and GRIDDS datasets with identification accuracies of 97.68 and 99.12%, respectively.

Moreover, we introduced a state-estimation occlusion filter for the first time. The results demonstrated that the filter was able to significantly recover performance lost during occluded sequences (RQ4, Obj8). As no existing dataset contains labelled examples of occlusion, we provide several strategies for simulating occlusion in gait sequences.

## 6.2. Future Work Directions

There are several potential directions for future research. For example, despite developing a world-class gait dataset, further dataset contributions could be made. For instance, no existing dataset contains labelled examples of genuine occlusion (i.e., self and environmental). Such data could further validate STONI-GID and other related methods under realistic occluded conditions. Similarly, the effect of time between data recording (i.e., for model training) and identification is relatively unstudied. To investigate this, a dataset would have to include examples of videos recorded at significant time intervals. Such a dataset does not currently exist.

Furthermore, despite demonstrating its ability to significantly restore performance lost due to occlusion, STONI-GID does not fully recover lost performance. Future research may explore alternative methods of data filtering or interpolation. Alternatively, methods of harnessing incomplete data may be explored instead of interpolating missing data. For example, Chapter 4 demonstrated that high performance can be achieved even when data is only available for a single joint .

Moreover, STONI-GID indicated high performance when evaluated on recorded data. However, to further enhance its practicality in real-world applications, the method should be expanded to integrate it with live video feeds, such as CCTV. To achieve this, a pipeline should be developed to process and format available data streams.

# References


[1] C. Fox, “Face recognition police tools ‘staggeringly inaccurate,’” *BBC News*, 2018. https://www.bbc.co.uk/news/technology-44089161 (accessed Mar. 24, 2022).

[2] J. Sharman, “Metropolitan Police’s facial recognition technology 98% inaccurate, figures show,” *Independent*, 2018. https://www.independent.co.uk/news/uk/home-news/met-police-facial-recognition-success-south-wales-trial-home-office-false-positive-a8345036.html (accessed Mar. 24, 2022).

[3] M. Burgess, “Facial recognition tech used by UK police is making a ton of mistakes,” *WIRED*, 2018. https://www.wired.co.uk/article/face-recognition-police-uk-south-wales-met-notting-hill-carnival (accessed Mar. 24, 2022).

[4] G. Guo and N. Zhang, “A survey on deep learning based face recognition,” *Comput. Vis. Image Underst.*, vol. 189, no. July, p. 102805, 2019, doi: 10.1016/j.cviu.2019.102805.

[5] R. Baker, “Gait analysis methods in rehabilitation,” *J. Neuroeng. Rehabil.*, vol. 3, pp. 1–10, 2006, doi: 10.1186/1743-0003-3-4.

[6] S. Majumder, T. Mondal, and M. J. Deen, “A Simple, Low-Cost and Efficient Gait Analyzer for Wearable Healthcare Applications,” *IEEE Sens. J.*, vol. 19, no. 6, pp. 2320–2329, 2019, doi: 10.1109/JSEN.2018.2885207.

[7] D. Xu *et al.*, “A new method proposed for realizing human gait pattern recognition: Inspirations for the application of sports and clinical gait analysis,” *Gait Posture*, vol. 107, no. 818, pp. 293–305, 2023, doi: 10.1016/j.gaitpost.2023.10.019.

[8] W. Xu *et al.*, “KEH-Gait: Using Kinetic Energy Harvesting for Gait-based User Authentication Systems,” *IEEE Trans. Mob. Comput.*, vol. 18, no. 1, pp. 139–152, 2019, doi: 10.1109/TMC.2018.2828816.

[9] C. Raccagni *et al.*, “Physiotherapy improves motor function in patients with the Parkinson variant of multiple system atrophy: A prospective trial,” *Park. Relat. Disord.*, vol. 67, no. July, pp. 60–65, 2019, doi: 10.1016/j.parkreldis.2019.09.026.

[10] S. Pardoel, J. Kofman, J. Nantel, and E. D. Lemaire, “Wearable-sensor-

based detection and prediction of freezing of gait in parkinson's disease: A review," *Sensors*, vol. 19, no. 23, pp. 1–37, 2019, doi: 10.3390/s19235141.

[11] D. Greenwood, K. Davids, and I. Renshaw, "The role of a vertical reference point in changing gait regulation in cricket run-ups," *Eur. J. Sport Sci.*, vol. 16, no. 7, pp. 794–800, 2016, doi: 10.1080/17461391.2016.1151943.

[12] D. Seckiner, X. Mallett, P. Maynard, D. Meuwly, and C. Roux, "Forensic gait analysis — Morphometric assessment from surveillance footage," *Forensic Sci. Int.*, vol. 296, pp. 57–66, 2019, doi: 10.1016/j.forsciint.2019.01.007.

[13] M. S. Nixon and J. N. Carter, "Advances in automatic gait recognition," in *Proceedings - Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, 2004, pp. 139–144, doi: 10.1109/AFGR.2004.1301521.

[14] F. M. Castro, M. J. Marín-Jiménez, N. Guil, and N. P. de la Blanca, "Automatic learning of gait signatures for people identification," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017, vol. 10306 LNCS, pp. 257–270, doi: 10.1007/978-3-319-59147-6_23.

[15] HR News, "Number of CCTV Cameras in the UK reaches 5.2 million," *HR News*, 2020. https://hrnews.co.uk/number-of-cctv-cameras-in-the-uk-reaches-5-2-million/ (accessed Dec. 12, 2023).

[16] M. Burgess, "Police Use of Facial Recognition Is Sweeping the UK," *WIRED*, 2023. https://www.wired.co.uk/article/uk-police-face-recognition-expansion (accessed Dec. 12, 2023).

[17] S. Komkov and A. Petiushko, "AdvHat: Real-world adversarial attack on arcFace face ID system," in *Proceedings - International Conference on Pattern Recognition*, 2020, pp. 819–826, doi: 10.1109/ICPR48806.2021.9412236.

[18] X. Lu, X. Li, W. Sheng, and S. S. Ge, "Long-Term Person Re-Identification Based on Appearance and Gait Feature Fusion under Covariate Changes," *Processes*, vol. 10, no. 4, pp. 1–19, 2022, doi: 10.3390/pr10040770.

[19] N. Li and X. Zhao, "A multi-modal dataset for gait recognition under occlusion," *Appl. Intell.*, vol. 53, no. 2, pp. 1517–1534, 2023, doi: 10.1007/s10489-022-03474-8.

[20] C. Xu, S. Tsuji, Y. Makihara, X. Li, and Y. Yagi, "Occluded Gait Recognition via Silhouette Registration Guided by Automated Occlusion Degree Estimation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 3199–3209.

[21] Q. Leng, M. Ye, and Q. Tian, “A Survey of Open-World Person Re-Identification,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 4, pp. 1092–1108, 2020, doi: 10.1109/TCSVT.2019.2898940.

[22] L. Topham and W. Khan, “360 Degree Gait capture: A diverse and multi-modal gait dataset of indoor and outdoor walks acquired using multiple video cameras and sensors.” LJMU OpenData, 2022, doi: https://doi.org/10.24377/LJMU.d.00000133.

[23] L. Topham, W. Khan, D. Al-Jumeily, A. Waraich, and A. J. Hussain, “A diverse and multi-modal gait dataset of indoor and outdoor walks acquired using multiple cameras and sensors,” *Sci. Data*, vol. 10, no. 320, 2023, doi: https://doi.org/10.1038/s41597-023-02161-8.

[24] L. Topham, W. Khan, D. Al-Jumeily, A. Waraich, and A. J. Hussain, “Gait Identification using Hip Joint Movement and Deep Machine Learning,” in *International Conference on Intelligent Computing*, 2022, pp. 220–233, doi: https://doi.org/10.1007/978-3-031-13832-4_19.

[25] L. Topham, W. Khan, D. Al-Jumeily, A. Waraich, and A. Hussain, “Gait Identification Using Limb Joint Movement and Deep Machine Learning,” *IEEE Access*, vol. 10, pp. 100113–100127, 2022, doi: 10.1109/ACCESS.2022.3207836.

[26] D. White and M. A. Burton, “Individual differences and the multidimensional nature of face perception,” *Nat. Rev. Psychol.*, vol. 1, pp. 287–300, 2022.

[27] D. A. Maharani, C. Machbub, P. H. Rusmin, and L. Yulianti, “Improving the Capability of Real-Time Face Masked Recognition using Cosine Distance,” in *6th International Conference on Interactive Digital Media (ICIDM)*, 2021, pp. 6–11, doi: 10.1109/ICIDM51048.2020.9339677.

[28] I. Adjabi, A. Ouahabi, A. Benzaoui, and A. Taleb-Ahmed, “Past, Present, and Future of Face Recognition: A Review,” *Electronics*, vol. 9, pp. 1–52, 2020.

[29] Y. Guo, H. Wang, L. Wang, Y. Lei, L. Liu, and M. Bennamoun, “3D Face Recognition: Two Decades of Progress and Prospects,” *ACM Comput. Surv.*, vol. 56, no. 3, 2023, doi: 10.1145/3615863.

[30] J. Deng, J. Guo, J. Yang, N. Xue, I. Kotsia, and S. Zafeiriou, “ArcFace: Additive Angular Margin Loss for Deep Face Recognition,” *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, vol. 44, no. 10, pp. 5962–5979, 2019, doi: 10.1109/TPAMI.2021.3087709.

[31] Y. Huang *et al.*, “Curricularface: Adaptive curriculum learning loss for deep face recognition,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2020, pp. 5900–5909, doi: 10.1109/CVPR42600.2020.00594.

[32] M. Kim, A. K. Jain, and X. Liu, “AdaFace : Quality Adaptive Margin for Face Recognition,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 18750–18759.

[33] L. Li, X. Mu, S. Li, and H. Peng, “A Review of Face Recognition Technology,” *IEEE Access*, vol. 8, pp. 139110–139120, 2020, doi: 10.1109/ACCESS.2020.3011028.

[34] D. J. Carragher and P. J. B. Hancock, “Surgical face masks impair human face matching performance for familiar and unfamiliar faces,” *Cogn. Res. Princ. Implic.*, 2020, doi: 10.1186/s41235-020-00258-x.

[35] R. De Luis-García, C. Alberola-López, O. Aghzout, and J. Ruiz-Alzola, “Biometric identification systems,” *Signal Processing*, vol. 83, no. 12, pp. 2539–2557, 2003, doi: 10.1016/j.sigpro.2003.08.001.

[36] K. Syazana-Itqan, A. R. Syafeeza, N. M. Saad, N. A. Hamid, and W. H. Bin Mohd Saad, “A Review of Finger-Vein Biometrics Identification Approaches,” *Indian J. Sci. Technol.*, vol. 9, no. 32, 2016, doi: 10.17485/ijst/2016/v9i32/99276.

[37] Z. Mao, W. X. Yao, and Y. Huang, “EEG-based biometric identification with deep learning,” in *International IEEE/EMBS Conference on Neural Engineering, NER*, 2017, pp. 609–612, doi: 10.1109/NER.2017.8008425.

[38] M. Aliaskar, T. Mazakov, A. Mazakova, S. Jomartova, and T. Shormanov, “Human voice identification based on the detection of fundamental harmonics,” in *ENERGYCON 2022 - 2022 IEEE 7th International Energy Conference, Proceedings*, 2022, no. Energycon, pp. 1–4, doi: 10.1109/ENERGYCON53164.2022.9830471.

[39] W. Li *et al.*, “An In-Depth Exploration of Person Re-Identification and Gait Recognition in Cloth-Changing Conditions,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 13824–13833, doi: 10.1109/cvpr52729.2023.01328.

[40] V. Narayan, S. Awasthi, N. Fatima, M. Faiz, and S. Srivastava, “Deep Learning Approaches for Human Gait Recognition: A Review,” in *2023 International Conference on Artificial Intelligence and Smart Communication, AISC 2023*, 2023, pp. 763–768, doi: 10.1109/AISC56616.2023.10085665.

[41] A. Chauhan, “Gait Silhouette Dataset,” *Kaggle*, 2020. https://www.kaggle.com/watermasterz/gaitsilhouettedataset?select=GaitDatasetC-silh (accessed Mar. 30, 2021).

[42] T. T. Verlekar, L. D. Soares, and P. L. Correia, “Gait recognition in the wild using shadow silhouettes,” *Image Vis. Comput.*, vol. 76, pp. 1–13, 2018, doi: 10.1016/j.imavis.2018.05.002.

[43] F. Hebenstreit, A. Leibold, S. Krinner, G. Welsch, M. Lochmann, and B.

M. Eskofier, “Effect of walking speed on gait sub phase durations,” *Hum. Mov. Sci.*, vol. 43, pp. 118–124, 2015, doi: 10.1016/j.humov.2015.07.009.

[44] J. F. Nunes, P. M. Moreira, and J. M. R. S. Tavares, “GRIDDS - A Gait Recognition Image and Depth Dataset,” in *In ECCOMAS Thematic Conference on Computational Vision and Medical Image Processing*, vol. 34, no. December, 2019, pp. 343–352.

[45] M. Hofmann, J. Geiger, S. Bachmann, B. Schuller, and G. Rigoll, “The TUM Gait from Audio, Image and Depth (GAID) database: Multimodal recognition of subjects and traits,” *J. Vis. Commun. Image Represent.*, vol. 25, no. 1, pp. 195–206, 2014, doi: 10.1016/j.jvcir.2013.02.006.

[46] J. Shutler, M. Nixon, and J. N. Carter, “On a Large Sequence-Based Human Gait Database,” in *4th International Conference on Recent Advances in Soft Computing*, 2002, pp. 66–71, doi: 10.1007/978-3-540-45240-9.

[47] R. Chereshnev and A. Kertész-Farkas, “HuGaDB: Human gait database for activity recognition from wearable inertial sensor networks,” in *International Conference on Analysis of Images, Social Networks and Texts*, 2017, vol. 10716 LNCS, no. March, pp. 131–141, doi: 10.1007/978-3-319-73013-4_12.

[48] A. Vajdi *et al.*, “Human Gait Database for Normal Walk Collected by Smart Phone Accelerometer,” 2019.

[49] C. Schreiber and F. Moissenet, “A multimodal dataset of human gait at different walking speeds established on injury-free adult participants,” *Nat. Sci. Data*, vol. 6, no. 1, pp. 1–7, 2019, doi: 10.1038/s41597-019-0124-4.

[50] E. Aydemir, T. Tuncer, S. Dogan, and M. Unsal, “A novel biometric recognition method based on multi kernelled bijection octal pattern using gait sound,” *Appl. Acoust.*, vol. 173, p. 107701, 2021, doi: 10.1016/j.apacoust.2020.107701.

[51] I. P. I. Pappas, T. Keller, S. Mangold, M. R. Popovic, V. Dietz, and M. Morari, “A Reliable Gyroscope-Based Gait-Phase Detection Sensor Embedded in a Shoe Insole,” *IEEE Sens. J.*, vol. 4, no. 2, pp. 268–274, 2004, doi: 10.1109/JSEN.2004.823671.

[52] W. Krajnik, Ł. Markiewicz, and R. Sitnik, “sSfS: Segmented Shape from Silhouette Reconstruction of the Human Body,” *Sensors*, vol. 22, no. 3, 2022, doi: 10.3390/s22030925.

[53] H. Luo, S. Li, and Q. Zhao, “Towards Silhouette-Aware Human Detection in Depth Images,” in *Proceedings of the International Joint Conference on Neural Networks*, 2021, vol. 2021-July, pp. 1–8, doi: 10.1109/IJCNN52387.2021.9534347.

[54] S. Sivolobov, “Human gait feature extraction method,” *Procedia Comput. Sci.*, vol. 193, pp. 220–227, 2021, doi: 10.1016/j.procs.2021.10.022.

[55] S. Ouellet and F. Michaud, “Enhanced automated body feature extraction from a 2D image using anthropomorphic measures for silhouette analysis,” *Expert Syst. Appl.*, vol. 91, pp. 270–276, 2018, doi: 10.1016/j.eswa.2017.09.006.

[56] J. Canny, “A Computational Approach to Edge Detection,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, 1986, doi: 10.1109/TPAMI.1986.4767851.

[57] M. Babaee, Y. Zhu, O. Kopuklu, S. Hormann, and G. Rigoll, “Gait energy image restoration using generative adversarial networks,” 2019, pp. 2596–2600.

[58] M. Hofmann, S. Bachmann, and G. Rigoll, “2 . 5D Gait Biometrics using the Depth Gradient Histogram Energy Image,” in *2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2012, pp. 399–403.

[59] S. Sivapalan, D. Chen, S. Denman, S. Sridharan, and C. Fookes, “Gait energy volumes and frontal gait recognition using depth images,” in *2011 International Joint Conference on Biometrics, IJCB 2011*, 2011, doi: 10.1109/IJCB.2011.6117504.

[60] J. Luo, C. Zi, J. Zhang, and Y. Liu, “Gait recognition using GEI and AFDEI,” *Int. J. Opt.*, vol. 44, no. 4, pp. 400–404, 2015, doi: 10.3969/j.issn.1003-501X.2017.04.003.

[61] J. Wen, Y. Shen, and J. Yang, “Multi-View Gait Recognition Based on Generative Adversarial Network,” *Neural Process. Lett.*, 2022, doi: 10.1007/s11063-021-10709-1.

[62] 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 2016*, 2016, doi: 10.1109/ICB.2016.7550060.

[63] Z. Wu, Y. Huang, L. Wang, and S. Member, “Learning Representative Deep Features for Image Set Analysis,” *IEEE Trans. Multimed.*, vol. 17, no. 11, pp. 1960–1968, 2015.

[64] A. Ramakić, D. Sušanj, K. Lenac, and Z. Bundalo, “Depth-Based Real-Time Gait Recognition,” *J. Circuits, Syst. Comput.*, vol. 29, no. 16, 2020, doi: 10.1142/S0218126620502667.

[65] B. Gálai and C. Benedek, “Feature selection for Lidar-based gait recognition,” in *2015 International Workshop on Computational Intelligence for Multimedia Understanding, IWCIM 2015*, 2015, pp. 12–16, doi: 10.1109/IWCIM.2015.7347076.

[66] J. Xu, H. Li, and S. Hou, “Attention-based gait recognition network with novel partial representation PGOFI based on prior motion information,” *Digit. Signal Process.*, vol. 133, p. 103845, 2023, doi: 10.1016/j.dsp.2022.103845.

[67] S. Zheng, J. Zhang, K. Huang, R. He, and T. Tan, “Robust view transformation model for gait recognition,” in *Proceedings - International Conference on Image Processing, ICIP*, 2011, pp. 2073–2076, doi: 10.1109/ICIP.2011.6115889.

[68] W. Kusakunniran, Q. Wu, H. Li, and J. Zhang, “Multiple views gait recognition using view transformation model based on optimized gait energy image,” in *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops 2009*, 2009, pp. 1058–1064, doi: 10.1109/ICCVW.2009.5457587.

[69] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi, “Gait recognition using a view transformation model in the frequency domain,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2006, vol. 3953 LNCS, pp. 151–163, doi: 10.1007/11744078_12.

[70] T. Teepe, A. Khan, J. Gilg, F. Herzog, S. Hörmann, and G. Rigoll, “Gaitgraph: Graph Convolutional Network for Skeleton-Based Gait Recognition,” in *Proceedings - International Conference on Image Processing, ICIP*, 2021, vol. 2021-Septe, pp. 2314–2318, doi: 10.1109/ICIP42928.2021.9506717.

[71] S. Gul, M. I. Malik, G. M. Khan, and F. Shafait, “Multi-view gait recognition system using spatio-temporal features and deep learning,” *Expert Syst. Appl.*, vol. 179, no. March, p. 115057, 2021, doi: 10.1016/j.eswa.2021.115057.

[72] L. Topham, W. Khan, D. Al-Jumeily, and A. J. Hussain, “Human Body Pose Estimation for Gait Identification: A Comprehensive Survey of Datasets and Models,” *ACM Comput. Surv.*, 2022.

[73] T. N. Nguyen, H. H. Huynh, and J. Meunier, “Skeleton-Based Abnormal Gait Detection,” *Sensors*, vol. 16, no. 11, pp. 1–13, 2016, doi: 10.3390/s16111792.

[74] T. Teepe, J. Gilg, F. Herzog, S. Hormann, and G. Rigoll, “Towards a Deeper Understanding of Skeleton-based Gait Recognition,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2022, vol. 2022-June, pp. 1568–1576, doi: 10.1109/CVPRW56347.2022.00163.

[75] Z. Cao, G. Hidalgo, T. Simon, S. E. Wei, and Y. Sheikh, “OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 172–186,

2021, doi: 10.1109/TPAMI.2019.2929257.

[76] P. Pala, L. Seidenari, S. Berretti, and A. Del Bimbo, “Enhanced skeleton and face 3D data for person re-identification from depth cameras,” *Comput. Graph.*, vol. 79, pp. 69–80, 2019, doi: 10.1016/j.cag.2019.01.003.

[77] F. Battistone and A. Petrosino, “TGLSTM: A time based graph deep learning approach to gait recognition,” *Pattern Recognit. Lett.*, vol. 126, pp. 132–138, 2019, doi: 10.1016/j.patrec.2018.05.004.

[78] W. Sheng and X. Li, “Multi-task learning for gait-based identity recognition and emotion recognition using attention enhanced temporal graph convolutional network,” *Pattern Recognit.*, vol. 114, 2021, doi: 10.1016/j.patcog.2021.107868.

[79] S. Gao, J. Yun, Y. Zhao, and L. Liu, “Gait-D: Skeleton-based gait feature decomposition for gait recognition,” *IET Comput. Vis.*, vol. 16, no. 2, pp. 111–125, 2022, doi: 10.1049/cvi2.12070.

[80] H. Rao *et al.*, “A Self-Supervised Gait Encoding Approach with Locality-Awareness for 3D Skeleton Based Person Re-Identification,” *IEEE Trans. Pattern Anal. Mach. Intell. (Early Access)*, 2021, doi: 10.1109/tpami.2021.3092833.

[81] N. Li and X. Zhao, “A Strong and Robust Skeleton-Based Gait Recognition Method with Gait Periodicity Priors,” *IEEE Trans. Multimed.*, vol. 25, pp. 3046–3058, 2023, doi: 10.1109/TMM.2022.3154609.

[82] L. Yao, W. Kusakunniran, Q. Wu, J. Zhang, Z. Tang, and W. Yang, “Robust gait recognition using hybrid descriptors based on Skeleton Gait Energy Image,” *Pattern Recognit. Lett.*, 2019, doi: 10.1016/j.patrec.2019.05.012.

[83] L. Yao, W. Kusakunniran, Q. Wu, J. Zhang, and Z. Tang, “Robust CNN-based Gait Verification and Identification using Skeleton Gait Energy Image,” in *2018 International Conference on Digital Image Computing: Techniques and Applications, DICTA 2018*, 2018, doi: 10.1109/DICTA.2018.8615802.

[84] M. Graziani *et al.*, *A global taxonomy of interpretable AI: unifying the terminology for the technical and social sciences*, vol. 56, no. 4. Springer Netherlands, 2023.

[85] M. Rashmi and R. M. R. Guddeti, “Human identification system using 3D skeleton-based gait features and LSTM model,” *J. Vis. Commun. Image Represent.*, vol. 82, p. 103416, 2022, doi: 10.1016/j.jvcir.2021.103416.

[86] M. W. Rahman and M. L. Gavrilova, “Kinect gait skeletal joint feature-based person identification,” in *Proceedings of 2017 IEEE 16th*

*International Conference on Cognitive Informatics and Cognitive Computing, ICCI*CC 2017*, 2017, pp. 423–430, doi: 10.1109/ICCI-CC.2017.8109783.

[87] H. Hagras, "Toward Human-Understandable, Explainable AI," *Computer (Long. Beach. Calif).*, vol. 51, no. 9, pp. 28–36, 2018, doi: 10.1109/MC.2018.3620965.

[88] Y. Chen, Y. Tian, and M. He, "Monocular human pose estimation: A survey of deep learning-based methods," *Comput. Vis. Image Underst.*, vol. 192, no. January, p. 102897, 2020, doi: 10.1016/j.cviu.2019.102897.

[89] Q. Dang, J. Yin, B. Wang, and W. Zheng, "Deep Learning-Based 2D Human Pose Estimation: A survey," *Tsinghua Sci. Technol.*, vol. 24, no. 6, pp. 663–676, 2019, doi: 10.26599/TST.2018.9010100.

[90] C. Zheng *et al.*, "Deep Learning-Based Human Pose Estimation: A Survey," *Tsinghua Sci. Technol.*, vol. 24, no. 6, pp. 663–676, 2020, doi: 10.26599/TST.2018.9010100.

[91] T. Y. Lin *et al.*, "Microsoft COCO: Common objects in context," in *European Conference on Computer Vision (ECCV)*, 2014, vol. 8693 LNCS, no. PART 5, pp. 740–755, doi: 10.1007/978-3-319-10602-1_48.

[92] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, "Human3.6M: Large scale datasets and predictive methods for 3D human sensing in natural environments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 7, pp. 1325–1339, 2014, doi: 10.1109/TPAMI.2013.248.

[93] F. Xia, P. Wang, X. Chen, and A. Yuille, "Joint multi-person pose estimation and semantic part segmentation," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, vol. 2017-Janua, pp. 6080–6089, doi: 10.1109/CVPR.2017.644.

[94] G. Varol *et al.*, "Learning from synthetic humans," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, vol. 2017-Janua, pp. 4627–4635, doi: 10.1109/CVPR.2017.492.

[95] L. Sigal, A. O. Balan, and M. J. Black, "HumanEva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion," *Int. J. Comput. Vis.*, vol. 87, no. 1–2, pp. 4–27, 2010, doi: 10.1007/s11263-009-0273-6.

[96] D. Mehta *et al.*, "Single-shot multi-person 3D pose estimation from monocular RGB," in *Proceedings - 2018 International Conference on 3D Vision, 3DV 2018*, 2018, pp. 120–130, doi: 10.1109/3DV.2018.00024.

[97] H. Joo *et al.*, "Panoptic Studio: A Massively Multiview System for

Social Interaction Capture," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 1, pp. 190–204, 2019, doi: 10.1109/TPAMI.2017.2782743.

[98] S. Johnson and M. Everingham, "Clustered pose and nonlinear appearance models for human pose estimation," in *British Machine Vision Conference, BMVC 2010 - Proceedings*, 2010, no. ii, pp. 1–11, doi: 10.5244/C.24.12.

[99] S. Johnson and M. Everingham, "Learning effective human pose estimation from inaccurate annotation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2011, pp. 1465–1472, doi: 10.1109/CVPR.2011.5995318.

[100] B. Sapp and B. Taskar, "MODEC: Multimodal decomposable models for human pose estimation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3674–3681, doi: 10.1109/CVPR.2013.471.

[101] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, "2D human pose estimation: New benchmark and state of the art analysis," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3686–3693, doi: 10.1109/CVPR.2014.471.

[102] M. Trumble, A. Gilbert, C. Malleson, A. Hilton, and J. Collomosse, "Total capture: 3D human pose estimation fusing video and inertial sensors," in *British Machine Vision Conference 2017, BMVC 2017*, 2017, pp. 1–13, doi: 10.5244/c.31.14.

[103] Y. Wang, L. Guo, Z. Lu, X. Wen, S. Zhou, and W. Meng, "From Point to Space: 3D Moving Human Pose Estimation Using Commodity WiFi," *IEEE Commun. Lett.*, vol. 25, no. 7, pp. 2235–2239, 2021, doi: 10.1109/LCOMM.2021.3073271.

[104] A. Elhayek *et al.*, "MARCOnI - ConvNet-Based MARker-less motion capture in outdoor and indoor scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 3, pp. 501–514, 2017, doi: 10.1109/TPAMI.2016.2557779.

[105] T. von Marcard, R. Henschel, M. J. Black, B. Rosenhahn, and G. Pons-Moll, "Recovering accurate 3D human pose in the wild using IMUs and a moving camera," *Eur. Conf. Comput. Vis.*, vol. 11214 LNCS, pp. 614–631, 2018, doi: 10.1007/978-3-030-01249-6_37.

[106] D. Mehta *et al.*, "Monocular 3D human pose estimation in the wild using improved CNN supervision," in *Proceedings - 2017 International Conference on 3D Vision, 3DV 2017*, 2018, pp. 506–516, doi: 10.1109/3DV.2017.00064.

[107] M. Fabbri, F. Lanzi, S. Calderara, A. Palazzi, R. Vezzani, and R. Cucchiara, "Learning to Detect and Track Visible and Occluded Body Joints in a Virtual World," in *European conference on computer vision*

*(ECCV)*, 2018, vol. 11208 LNCS, pp. 450–466, doi: 10.1007/978-3-030-01225-0_27.

[108] C. Lassner, J. Romero, M. Kiefel, F. Bogo, M. J. Black, and P. V. Gehler, "Unite the people: Closing the loop between 3D and 2D human representations," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, vol. 2017-Janua, pp. 4704–4713, doi: 10.1109/CVPR.2017.500.

[109] J. Hui, "mAP(mean Average Precision) for Object Detection," *Medium2*, 2018. https://jonathan-hui.medium.com/map-mean-average-precision-for-object-detection-45c121a31173 (accessed Jul. 26, 2021).

[110] T. L. Munea, Y. Z. Jembre, H. T. Weldegebriel, L. Chen, C. Huang, and C. Yang, "The Progress of Human Pose Estimation: A Survey and Taxonomy of Models Applied in 2D Human Pose Estimation," *IEEE Access*, vol. 8, pp. 133330–133348, 2020, doi: 10.1109/ACCESS.2020.3010248.

[111] A. Toshev and C. Szegedy, "DeepPose: Human pose estimation via deep neural networks," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1653–1660, doi: 10.1109/CVPR.2014.214.

[112] G. Ning, Z. Zhang, and Z. He, "Knowledge-Guided Deep Fractal Neural Networks for Human Pose Estimation," *IEEE Trans. Multimed.*, vol. 20, no. 5, pp. 1246–1259, 2018, doi: 10.1109/TMM.2017.2762010.

[113] M. Andriluka *et al.*, "PoseTrack: A Benchmark for Human Pose Estimation and Tracking," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5167–5176, doi: 10.1109/CVPR.2018.00542.

[114] J. Carreira, P. Agrawal, K. Fragkiadaki, and J. Malik, "Human pose estimation with iterative error feedback," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, vol. 2016-Decem, pp. 4733–4742, doi: 10.1109/CVPR.2016.512.

[115] S. Liang, X. Sun, and Y. Wei, "Compositional Human Pose Regression," *Comput. Vis. Image Underst.*, vol. 176–177, pp. 1–8, 2018, doi: 10.1016/j.cviu.2018.10.006.

[116] D. C. Luvizon, H. Tabia, and D. Picard, "Human pose regression by combining indirect part detection and contextual information," *Comput. Graph.*, vol. 85, pp. 15–22, 2019, doi: 10.1016/j.cag.2019.09.002.

[117] F. Zhang, X. Zhu, and M. Ye, "Fast human pose estimation," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2019-June, pp. 3512–3521, 2019, doi: 10.1109/CVPR.2019.00363.

[118] B. Artacho and A. Savakis, “UniPose: Unified human pose estimation in single images and videos,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2020, pp. 7033–7042, doi: 10.1109/CVPR42600.2020.00706.

[119] Y. Luo *et al.*, “LSTM Pose Machines,” in *European Conference on Computer Vision (ECCV)*, 2018, pp. 5207–5215.

[120] B. Xiao, H. Wu, and Y. Wei, “Simple baselines for human pose estimation and tracking,” in *European conference on computer vision (ECCV)*, 2018, vol. 11210 LNCS, pp. 472–487, doi: 10.1007/978-3-030-01231-1_29.

[121] B. Debnath, M. O’brien, M. Yamaguchi, and A. Behera, “Adapting MobileNets for mobile based upper body pose estimation,” in *Proceedings of AVSS 2018 - 2018 15th IEEE International Conference on Advanced Video and Signal-Based Surveillance*, 2019, pp. 1–6, doi: 10.1109/AVSS.2018.8639378.

[122] V. Belagiannis and A. Zisserman, “Recurrent Human Pose Estimation,” *Proc. - 12th IEEE Int. Conf. Autom. Face Gesture Recognition, FG 2017 - 1st Int. Work. Adapt. Shot Learn. Gesture Underst. Prod. ASL4GUP 2017, Biometrics Wild, Bwild 2017, Heteroge*, pp. 468–475, 2017, doi: 10.1109/FG.2017.64.

[123] W. Yang, S. Li, W. Ouyang, H. Li, and X. Wang, “Learning feature pyramids for human pose estimation,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 1281–1290, doi: 10.4028/www.scientific.net/AMR.988.290.

[124] J. Gu *et al.*, “Recent advances in convolutional neural networks,” *Pattern Recognit.*, vol. 77, pp. 354–377, 2018, doi: 10.1016/j.patcog.2017.10.013.

[125] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *NeurIPS*, 2012, doi: 10.1201/9781420010749.

[126] C. Szegedy *et al.*, “Going deeper with convolutions,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9, doi: 10.1002/jctb.4820.

[127] Y. Cai *et al.*, “Learning Delicate Local Representations for Multi-person Pose Estimation,” in *European Conference on Computer Vision (ECCV)*, 2020, vol. 12348 LNCS, pp. 455–472, doi: 10.1007/978-3-030-58580-8_27.

[128] D. Osokin, “Real-time 2D multi-person pose estimation on CPU: Lightweight OpenPose,” in *ICPRAM 2019 - Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods*, 2019, pp. 744–748, doi: 10.5220/0007555407440748.

[129] X. Perez-Sala, S. Escalera, C. Angulo, and J. Gonzàlez, “A survey on model based approaches for 2D and 3D visual human pose recovery,” *Sensors (Switzerland)*, vol. 14, no. 3, pp. 4189–4210, 2014, doi: 10.3390/s140304189.

[130] J. Huang, Z. Zhu, F. Guo, and G. Huang, “The devil is in the details: Delving into unbiased data processing for human pose estimation,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2020, pp. 5699–5708, doi: 10.1109/CVPR42600.2020.00574.

[131] F. Zhang, X. Zhu, H. Dai, M. Ye, and C. Zhu, “Distribution-Aware Coordinate Representation for Human Pose Estimation,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2020, pp. 7091–7100, doi: 10.1109/CVPR42600.2020.00712.

[132] H. S. Fang, S. Xie, Y. W. Tai, and C. Lu, “RMPE: Regional Multi-person Pose Estimation,” *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2017-Octob, pp. 2353–2362, 2017, doi: 10.1109/ICCV.2017.256.

[133] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.

[134] K. Sun, B. Xiao, D. Liu, and J. Wang, “Deep high-resolution representation learning for human pose estimation,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019, vol. 2019-June, pp. 5686–5696, doi: 10.1109/CVPR.2019.00584.

[135] S. Kreiss, L. Bertoni, and A. Alahi, “PifPaf: Composite fields for human pose estimation,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019, vol. 2019-June, pp. 11969–11978, doi: 10.1109/CVPR.2019.01225.

[136] X. Nie, J. Feng, J. Zhang, and S. Yan, “Single-stage multi-person pose machines,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, vol. 2019-Octob, pp. 6950–6959, doi: 10.1109/ICCV.2019.00705.

[137] G. Papandreou, T. Zhu, L. C. Chen, S. Gidaris, J. Tompson, and K. Murphy, “Personlab: Person pose estimation and instance segmentation with a bottom-up, part-based, geometric embedding model,” *Eur. Conf. Comput. Vis.*, vol. 11218 LNCS, pp. 282–299, 2018, doi: 10.1007/978-3-030-01264-9_17.

[138] M. Fieraru, A. Khoreva, L. Pishchulin, and B. Schiele, “Learning to refine human pose estimation,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2018, vol. 2018-June, pp. 318–327, doi: 10.1109/CVPRW.2018.00058.

[139] L. Pishchulin *et al.*, “DeepCut: Joint subset partition and labeling for multi person pose estimation,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, vol. 2016-Decem, pp. 4929–4937, doi: 10.1109/CVPR.2016.533.

[140] E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, and B. Schiele, “Deepercut: A deeper, stronger, and faster multi-person pose estimation model,” in *European Conference on Computer Vision (ECCV)*, 2016, vol. 9910 LNCS, pp. 34–50, doi: 10.1007/978-3-319-46466-4_3.

[141] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, 2017, doi: 10.1109/TPAMI.2016.2577031.

[142] E. Insafutdinov *et al.*, “ArtTrack: Articulated multi-person tracking in the wild,” in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, vol. 2017-Janua, pp. 1293–1301, doi: 10.1109/CVPR.2017.142.

[143] A. Newell, K. Yang, and J. Deng, “Stacked hourglass networks for human pose estimation,” in *European Conference on Computer Vision (ECCV)*, 2016, vol. 1, pp. 483–499, doi: 10.1007/978-3-319-46484-8.

[144] S. Toshniwal, A. Kannan, C. Chiu, Y. Wu, T. N. Sainath, and K. Livescu, “A comparison of techniques for language model integration in encoder-decoder speech recognition,” in *2018 IEEE Spoken Language Technology Workshop (SLT)*, 2018, no. d, pp. 369–375.

[145] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, “Encoder-Decoder with Atrous Seperable Convolution for Semantic Image Segmentation,” in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.

[146] S.-E. Wei, V. Ramakrishna, T. Kanada, and Y. Sheikh, “Convolutional pose machines,” *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 4724–4732, 2016.

[147] K. Toyoda, M. Kono, and J. Rekimoto, “Post-data augmentation to improve deep pose estimation of extreme and wild motions,” in *26th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2019 - Proceedings*, 2019, pp. 1570–1574, doi: 10.1109/VR.2019.8798221.

[148] X. Peng, Z. Tang, F. Yang, R. S. Feris, and D. Metaxas, “Jointly Optimize Data Augmentation and Network Training: Adversarial Data Augmentation in Human Pose Estimation,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2226–2234, doi: 10.1109/CVPR.2018.00237.

[149] G. Papandreou *et al.*, “Towards accurate multi-person pose estimation in the wild,” in *Proceedings - 30th IEEE Conference on*

*Computer Vision and Pattern Recognition, CVPR 2017*, 2017, vol. 2017-Janua, pp. 3711–3719, doi: 10.1109/CVPR.2017.395.

[150] J. Zhang, Z. Chen, and D. Tao, "Towards High Performance Human Keypoint Detection," *Int. J. Comput. Vis.*, pp. 1–24, 2021, [Online]. Available: http://arxiv.org/abs/2002.00537.

[151] R. Zhang *et al.*, "Exploiting Offset-guided Network for Pose Estimation and Tracking," in *CVPR Workshops*, 2019, pp. 20–28, [Online]. Available: http://arxiv.org/abs/1906.01344.

[152] A. Jain, J. Tompson, Y. LeCun, and C. Bregler, "MoDeep: A deep learning framework using motion features for human pose estimation," in *European Conference on Computer Vision (ECCV)*, 2015, vol. 9004, pp. 302–315, doi: 10.1007/978-3-319-16808-1_21.

[153] P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid, "DeepFlow: Large displacement optical flow with deep matching," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, no. Section 2, pp. 1385–1392, doi: 10.1109/ICCV.2013.175.

[154] W. Gong *et al.*, "Human Pose Estimation from Monocular Images: A Comprehensive Survey," *Sensors (Switzerland)*, vol. 16, no. 12, pp. 1–39, 2016, doi: 10.3390/s16121966.

[155] X. Ji, Q. Fang, J. Dong, Q. Shuai, W. Jiang, and X. Zhou, "A survey on monocular 3D human pose estimation," *Virtual Real. Intell. Hardw.*, vol. 2, no. 6, pp. 471–500, 2020, doi: 10.1016/j.vrih.2020.04.005.

[156] H. Rhodin, M. Salzmann, and P. Fua, "Unsupervised geometry-aware representation for 3D human pose estimation," in *European Conference on Computer Vision (ECCV)*, 2018, vol. 11214 LNCS, pp. 765–782, doi: 10.1007/978-3-030-01249-6_46.

[157] X. Zhou, Q. Huang, X. Sun, X. Xue, and Y. Wei, "Weakly supervised transfer for 3d human pose estimation in the wild," in *IEEE International Conference on Computer Vision*, 2017, vol. 206, p. 3.

[158] N. Sarafianos, B. Boteanu, B. Ionescu, and I. A. Kakadiaris, "3D Human pose estimation: A review of the literature and analysis of covariates," *Comput. Vis. Image Underst.*, vol. 152, pp. 1–20, 2016, doi: 10.1016/j.cviu.2016.09.002.

[159] C. H. Chen and D. Ramanan, "3D human pose estimation = 2D pose estimation + matching," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, vol. 2017-Janua, pp. 5759–5767, doi: 10.1109/CVPR.2017.610.

[160] G. Pavlakos, X. Zhou, K. G. Derpanis, and K. Daniilidis, "Coarse-to-fine volumetric prediction for single-image 3D human pose," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, vol. 2017-Janua, pp. 1263–1272, doi:

10.1109/CVPR.2017.139.

[161] G. Varol *et al.*, “BodyNet: Volumetric inference of 3D human body shapes,” in *European Conference on Computer Vision (ECCV)*, 2018, vol. 11211 LNCS, pp. 20–38, doi: 10.1007/978-3-030-01234-2_2.

[162] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black, “SMPL: A skinned multi-person linear model,” *ACM Trans. Graph.*, vol. 34, no. 6, pp. 1–16, 2015, doi: 10.1145/2816795.2818013.

[163] Zhou, Xingyi, X. Sun, W. Zhang, S. Liang, and Y. Wei, “Deep kinematic pose regression,” in *European Conference on Computer Vision*, 2016, pp. 186–201, doi: 10.1007/978-3-319-49409-8.

[164] Y. Cheng, B. Yang, B. Wang, Y. Wending, and R. Tan, “Occlusion-aware networks for 3D human pose estimation in video,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, vol. 2019-Octob, no. Iccv, pp. 723–732, doi: 10.1109/ICCV.2019.00081.

[165] A. Arnab, C. Doersch, and A. Zisserman, “Exploiting temporal context for 3D human pose estimation in the wild,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019, vol. 2019-June, pp. 3390–3399, doi: 10.1109/CVPR.2019.00351.

[166] M. Omran, C. Lassner, G. Pons-Moll, P. Gehler, and B. Schiele, “Neural body fitting: Unifying deep learning and model based human pose and shape estimation,” in *Proceedings - 2018 International Conference on 3D Vision, 3DV 2018*, 2018, pp. 484–494, doi: 10.1109/3DV.2018.00062.

[167] D. Tome, C. Russell, and L. Agapito, “Lifting from the deep: Convolutional 3D pose estimation from a single image,” in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, vol. 2017-Janua, pp. 5689–5698, doi: 10.1109/CVPR.2017.603.

[168] K. Zhou, X. Han, N. Jiang, K. Jia, and J. Lu, “HEMlets pose: Learning part-centric heatmap triplets for accurate 3D human pose estimation,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, vol. 2019-Octob, pp. 2344–2353, doi: 10.1109/ICCV.2019.00243.

[169] C. Li and G. H. Lee, “Generating multiple hypotheses for 3D human pose estimation with mixture density network,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019, vol. 2019-June, pp. 9879–9887, doi: 10.1109/CVPR.2019.01012.

[170] B. Tekin Pablo Márquez-Neila Mathieu Salzmann Pascal Fua EPFL, “Learning to Fusing 2D Uncertainty and 3D Cues for Monocular Body Pose Estimation,” in *In Proceedings of the IEEE International*

*Conference on Computer Vision (ICCV)*, 2017.

[171] F. Moreno-Noguer, "3d human pose estimation from a single image via distance matrix regression," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2823–2832, [Online]. Available: http://www.sciencedirect.com/science/article/pii/B008044854202229X.

[172] S. Li, W. Zhang, and A. B. Chan, "Maximum-margin structured learning with deep networks for 3D human pose estimation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, vol. 2015 Inter, pp. 2848–2856, doi: 10.1109/ICCV.2015.326.

[173] C. Wang, J. Li, W. Liu, C. Qian, and C. Lu, "HMOR: Hierarchical Multi-person Ordinal Relations for Monocular Multi-person 3D Pose Estimation," in *European Conference on Computer Vision (ECCV)*, 2020, vol. 12348 LNCS, pp. 242–259, doi: 10.1007/978-3-030-58580-8_15.

[174] A. Benzine, F. Chabot, B. Luvison, Q. C. Pham, and C. Achard, "Pandanet: Anchor-based single-shot multi-person 3D pose estimation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2020, pp. 6855–6864, doi: 10.1109/CVPR42600.2020.00689.

[175] G. Moon, J. Y. Chang, and K. M. Lee, "Camera distance-aware top-down approach for 3D multi-person pose estimation from a single RGB image," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, vol. 2019-Octob, pp. 10132–10141, doi: 10.1109/ICCV.2019.01023.

[176] A. Zanfir, E. Marinoiu, and C. Sminchisescu, "Monocular 3D Pose and Shape Estimation of Multiple People in Natural Scenes: The Importance of Multiple Scene Constraints," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2148–2157, doi: 10.1109/CVPR.2018.00229.

[177] D. Mehta *et al.*, "XNect: Real-time Multi-Person 3D Motion Capture with a Single RGB Camera," *ACM Trans. Graph.*, vol. 39, no. 4, 2020, doi: 10.1145/3386569.3392410.

[178] G. Rogez, P. Weinzaepfel, and C. Schmid, "LCR-Net++: Multi-Person 2D and 3D Pose Detection in Natural Images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 5, pp. 1146–1161, 2020, doi: 10.1109/TPAMI.2019.2892985.

[179] G. Rogez, P. Weinzaepfel, and C. Schmid, "LCR-Net: Localization-classification-regression for human pose," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, vol. 2017-Janua, pp. 1216–1224, doi: 10.1109/CVPR.2017.134.

[180] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature Pyramid Networks for Object Detection,” *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, pp. 428–431, 2017, doi: 10.1109/ICVRIS.2019.00110.

[181] J. Zhen *et al.*, “SMAP: Single-Shot Multi-person Absolute 3D Pose Estimation,” in *European Conference on Computer Vision (ECCV)*, 2020, vol. 12360 LNCS, pp. 550–566, doi: 10.1007/978-3-030-58555-6_33.

[182] L. Jin *et al.*, “Single-Stage is Enough : Multi-Person Absolute 3D Pose Estimation Institute of North Electronic Equipment,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 13086–13095.

[183] J. Wang *et al.*, “Deep 3D human pose estimation: A review,” *Comput. Vis. Image Underst.*, vol. 210, no. August 2020, p. 103225, 2021, doi: 10.1016/j.cviu.2021.103225.

[184] H. Qiu, C. Wang, J. Wang, N. Wang, and W. Zeng, “Cross view fusion for 3D human pose estimation,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, vol. 2019-Octob, pp. 4341–4350, doi: 10.1109/ICCV.2019.00444.

[185] D. Tome, M. Toso, L. Agapito, and C. Russell, “Rethinking pose in 3D: Multi-stage refinement and recovery for markerless motion capture,” in *Proceedings - 2018 International Conference on 3D Vision, 3DV 2018*, 2018, pp. 474–483, doi: 10.1109/3DV.2018.00061.

[186] G. Pavlakos, X. Zhou, K. G. Derpanis, and K. Daniilidis, “Harvesting multiple views for marker-less 3D human pose annotations,” in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, vol. 2017-Janua, pp. 1253–1262, doi: 10.1109/CVPR.2017.138.

[187] H. Rhodin *et al.*, “Learning Monocular 3D Human Pose Estimation from Multi-view Images,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8437–8446, doi: 10.1109/CVPR.2018.00880.

[188] M. Kocabas, S. Karagoz, and E. Akbas, “Self-supervised learning of 3D human pose using multi-view geometry,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019, vol. 2019-June, pp. 1077–1086, doi: 10.1109/CVPR.2019.00117.

[189] K. Iskakov, E. Burkov, V. Lempitsky, and Y. Malkov, “Learnable triangulation of human pose,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, vol. 2019-Octob, pp. 7717–7726, doi: 10.1109/ICCV.2019.00781.

[190] J. Liang and M. Lin, “Shape-aware human pose and shape

reconstruction using multi-view images," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, vol. 2019-Octob, pp. 4351–4361, doi: 10.1109/ICCV.2019.00445.

[191] P. C. Chang, M. C. Tien, J. L. Wu, and C. S. Hu, "Real-time gender classification from human gait for arbitrary view angles," in *ISM 2009 - 11th IEEE International Symposium on Multimedia*, 2009, pp. 88–95, doi: 10.1109/ISM.2009.81.

[192] H. Chao, Y. He, J. Zhang, and J. Feng, "GaitSet: Regarding gait as a set for cross-view gait recognition," in *33rd AAAI Conference on Artificial Intelligence, AAAI 2019, 31st Innovative Applications of Artificial Intelligence Conference, IAAI 2019 and the 9th AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019*, 2019, no. 16, pp. 8126–8133, doi: 10.1609/aaai.v33i01.33018126.

[193] H. Chao, K. Wang, Y. He, J. Zhang, and J. Feng, "GaitSet: Cross-View Gait Recognition Through Utilizing Gait As a Deep Set," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3467–3478, 2022, doi: 10.1109/TPAMI.2021.3057879.

[194] E. Gianaria and M. Grangetto, "Robust gait identification using Kinect dynamic skeleton data," *Multimed. Tools Appl.*, vol. 78, no. 10, pp. 13925–13948, 2019, doi: 10.1007/s11042-018-6865-9.

[195] M. Z. Uddin, D. Muramatsu, N. Takemura, M. A. R. Ahad, and Y. Yagi, "Spatio-temporal silhouette sequence reconstruction for gait recognition against occlusion," *IPSJ Trans. Comput. Vis. Appl.*, vol. 11, no. 1, 2019, doi: 10.1186/s41074-019-0061-3.

[196] Y. Peng, C. Cao, and Z. He, "Occluded Gait Recognition," in *Proceedings of the International Joint Conference on Neural Networks*, 2023, vol. 2023-June, pp. 1–8, doi: 10.1109/IJCNN54540.2023.10191651.

[197] A. Gupta and V. B. Semwal, "Occluded Gait reconstruction in multi person Gait environment using different numerical methods," *Multimed. Tools Appl.*, vol. 81, no. 16, pp. 23421–23448, 2022, doi: 10.1007/s11042-022-12218-2.

[198] Y. Iwashita, K. Uchino, and R. Kurazume, "Gait-based person identification robust to changes in appearance," *Sensors*, vol. 13, no. 6, pp. 7884–7901, 2013, doi: 10.3390/s130607884.

[199] Office for National Statistics, "Ethnicity and National Identity in England and Wales: 2011," 2011. https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/articles/ethnicityandnationalidentityinenglandandwales/2012-12-11 (accessed Oct. 03, 2022).

[200] M. De Marsico and A. Mecca, "A survey on gait recognition via wearable sensors," *ACM Comput. Surv.*, vol. 52, no. 4, 2019, doi:

10.1145/3340293.

[201] MOTI, “MOTI,” *MOTI*, 2021. moti.dk (accessed Apr. 12, 2022).

[202] N. Ahmad, R. A. R. Ghazilla, N. M. Khairi, and V. Kasi, “Reviews on Various Inertial Measurement Unit (IMU) Sensor Applications,” *Int. J. Signal Process. Syst.*, vol. 1, no. 2, pp. 256–262, 2013, doi: 10.12720/ijsps.1.2.256-262.

[203] T. S. Palsson, S. W. Christensen, M. H. Thomsen, and R. P. Hirata, “Assessment of range and quality of neck movement using a smartphone-based application,” *Musculoskelet. Sci. Pract.*, vol. 41, no. January, pp. 64–69, 2019, doi: 10.1016/j.msksp.2019.01.002.

[204] M. Rigoni *et al.*, “Assessment of shoulder range of motion using a wireless inertial motion capture device—A validation study,” *Sensors (Switzerland)*, vol. 19, no. 8, pp. 1–8, 2019, doi: 10.3390/s19081781.

[205] J. Voight, *Quaternion algebras*. Springer Nature, 2021.

[206] S. Rouzbeh and M. Babaei, “Human gait recognition using body measures and joint angles,” *Int. J. Sci. Knowl.*, vol. 6, no. 4, pp. 10–16, 2015.

[207] Y. Watanabe and M. Kimura, “Gait identification and authentication using LSTM based on 3-axis accelerations of smartphone,” *Procedia Comput. Sci.*, vol. 176, pp. 3873–3880, 2020, doi: 10.1016/j.procs.2020.09.001.

[208] N. Kleanthous, A. J. Hussain, W. Khan, and P. Liatsis, “A new machine learning based approach to predict Freezing of Gait,” *Pattern Recognit. Lett.*, vol. 140, pp. 119–126, 2020, doi: 10.1016/j.patrec.2020.09.011.

[209] S. Sony, S. Gamage, A. Sadhu, and J. Samarabandu, “Vibration-based multiclass damage detection and localization using long short-term memory networks,” *Structures*, vol. 35, no. November 2021, pp. 436–451, 2022, doi: 10.1016/j.istruc.2021.10.088.

[210] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, “Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2012, vol. 7657 LNCS, pp. 216–223, doi: 10.1007/978-3-642-35395-6_30.

[211] R. S. Witte and J. S. Witte, *Statistics*, 11th ed. Wiley, 2021.

[212] N. Kleanthous, A. J. Hussain, W. Khan, J. Sneddon, A. Al-Shamma’a, and P. Liatsis, “A survey of machine learning approaches in animal behaviour,” *Neurocomputing*, vol. 491, pp. 442–463, 2022, doi: 10.1016/j.neucom.2021.10.126.

[213] F. Song, Z. Guo, and D. Mei, “Feature selection using principal component analysis,” in *International Conference on System Science, Engineering Design and Manufacturing Informatization*, 2010, vol. 1, pp. 27–30, doi: 10.1109/ICSEM.2010.14.

[214] A. Larumbe, M. Ariz, J. J. Bengoechea, R. Segura, R. Cabeza, and A. Villanueva, “Improved Strategies for HPE Employing Learning-by-Synthesis Approaches,” in *IEEE International Conference on Computer Vision Workshops, ICCVW*, 2017, pp. 1545–1554, doi: 10.1109/ICCVW.2017.182.

[215] W. Khan, K. Crockett, J. O’Shea, A. Hussain, and B. M. Khan, “Deception in the eyes of deceiver: A computer vision and machine learning based automated deception detection,” *Expert Syst. Appl.*, vol. 169, no. February 2020, p. 114341, 2021, doi: 10.1016/j.eswa.2020.114341.

[216] A. K. Gárate-Escamila, A. Hajjam El Hassani, and E. Andrès, “Classification models for heart disease prediction using feature selection and PCA,” *Informatics Med. Unlocked*, vol. 19, 2020, doi: 10.1016/j.imu.2020.100330.

[217] H. Abdi and L. J. Williams, “Principal component analysis,” *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 2, no. 4, pp. 433–459, 2010, doi: 10.1002/wics.101.

[218] M. B. Kursa and W. R. Rudnicki, “Feature selection with the boruta package,” *J. Stat. Softw.*, vol. 36, no. 11, pp. 1–13, 2010, doi: 10.18637/jss.v036.i11.

[219] I. Guyon, J. Weston, and S. Barnhill, “Gene selection for cancer classification using Support Vector Machines,” *Mach. Learn.*, vol. 46, pp. 389–422, 2002, doi: 10.1007/978-3-540-88192-6_8.

[220] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[221] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[222] Y. Yu, X. Si, C. Hu, and J. Zhang, “A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures,” *Neural Comput.*, p. 1270, 2019.

[223] C. Filipi Gonçalves dos Santos *et al.*, “Gait Recognition Based on Deep Learning: A Survey,” *ACM Comput. Surv.*, vol. 55, no. 2, pp. 1–34, 2022, doi: 10.1145/3490235.

[224] G. Van Houdt, C. Mosquera, and G. Nápoles, “A review on the long short-term memory model,” *Artif. Intell. Rev.*, vol. 53, no. 8, pp. 5929–5955, 2020, doi: 10.1007/s10462-020-09838-1.

[225] A. T. Prihatno, I. B. K. Y. Utama, and Y. M. Jang, “oneM2M-Enabled

Prediction of High Particulate Matter Data Based on Multi-Dense Layer BiLSTM Model," *Appl. Sci.*, vol. 12, no. 4, 2022, doi: 10.3390/app12042260.

[226] D. A. Pisner and D. M. Schnyer, "Support vector machine," in *Machine Learning: Methods and Applications to Brain Disorders*, Elsevier Inc., 2019, pp. 101–121.

[227] W. S. Noble, "What is a support vector machine?," *Nature Biotechnology*, vol. 24, no. 12, pp. 1565–1567, 2006.

[228] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, and F. Herrera, "An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes," *Pattern Recognit.*, vol. 44, no. 8, pp. 1761–1776, 2011, doi: 10.1016/j.patcog.2011.01.017.

[229] M. Koklu and I. A. Ozkan, "Multiclass classification of dry beans using computer vision and machine learning techniques," *Computers and Electronics in Agriculture*, vol. 174, no. June 2019, Elsevier, p. 105507, 2020.

[230] A. Krogh, "What are artificial neural networks?," *Nat. Biotechnol.*, vol. 26, no. 2, pp. 195–197, 2008, doi: 10.1038/nbt1386.

[231] H. Samet, "K-nearest neighbor finding using MaxNearestDist," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 243–252, 2008, doi: 10.1109/TPAMI.2007.1182.

[232] L. Jiang, Z. Cai, D. Wang, and S. Jiang, "Survey of improving K-nearest-neighbor for classification," in *Proceedings - Fourth International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2007*, 2007, vol. 1, pp. 679–683, doi: 10.1109/FSKD.2007.552.

[233] D. W. Aha, D. Kibler, and M. K. Albert, "Instance-Based Learning Algorithms," *Mach. Learn.*, vol. 6, no. Springer, pp. 37--66, 1991.

[234] S. B. Imandoust and M. Bolandraftar, "Application of K-Nearest Neighbor ( KNN ) Approach for Predicting Economic Events : Theoretical Background," *Int. J. Eng. Res. Appl.*, vol. 3, no. 5, pp. 605–610, 2013.

[235] M. Gadaleta, L. Merelli, and M. Rossi, "Human authentication from ankle motion data using convolutional neural networks," in *IEEE Statistical Signal Processing Workshop*, 2016, pp. 1–5, doi: 10.1109/SSP.2016.7551815.

[236] R. San-Segundo, J. D. Echeverry-Correa, C. Salamea-Palacios, S. Lebai Lutfi, and J. M. Pardo, "I-vector analysis for Gait-based Person Identification using smartphone inertial signals," *Pervasive Mob. Comput.*, vol. 38, pp. 140–153, 2017, doi: 10.1016/j.pmcj.2016.09.007.

[237] O. Dehzangi, M. Taherisadr, and R. ChangalVala, “IMU-based gait recognition using convolutional neural networks and multi-sensor fusion,” *Sensors (Switzerland)*, vol. 17, no. 12, 2017, doi: 10.3390/s17122735.

[238] H. Huang, P. Zhou, Y. Li, and F. Sun, “A lightweight attention-based cnn model for efficient gait recognition with wearable imu sensors,” *Sensors*, vol. 21, no. 8, pp. 1–13, 2021, doi: 10.3390/s21082866.

[239] J. E. Cutting and L. T. Kozlowski, “Recognizing friends by their walk: Gait perception without familiarity cues,” *Bull. Psychon. Soc.*, vol. 9, no. 5, pp. 353–356, 1977, doi: 10.3758/BF03337021.

[240] A. S. Wicaksana and C. C. S. Liem, “Human-Explainable Features for Job Candidate Screening Prediction,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2017, vol. 2017-July, pp. 1664–1669, doi: 10.1109/CVPRW.2017.212.

[241] H. Iwama, M. Okumura, Y. Makihara, and Y. Yagi, “The OU-ISIR Gait Database Comprising the Large Population Dataset and Performance Evaluation of Gait Recognition,” *IEEE Trans. Inf. FORENSICS Secur.*, vol. 7, no. 1, pp. 1511–1521, 2012, doi: 10.1186/s41074-017-0035-2.

[242] A. Bendale and T. E. Boult, “Towards open set deep networks,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, vol. 2016-Decem, pp. 1563–1572, doi: 10.1109/CVPR.2016.173.

[243] D. S. Matovski, M. S. Nixon, S. Mahmoodi, and J. N. Carter, “The effect of time on gait recognition performance,” *IEEE Trans. Inf. Forensics Secur.*, vol. 7, no. 2, pp. 543–552, 2012, doi: 10.1109/TIFS.2011.2176118.

[244] C. Xu, Y. Makihara, X. Li, and Y. Yagi, “Occlusion-Aware Human Mesh Model-Based Gait Recognition,” *IEEE Trans. Inf. Forensics Secur.*, vol. 18, pp. 1309–1321, 2023, doi: 10.1109/TIFS.2023.3236181.

[245] K. Chen *et al.*, “Patient-specific pose estimation in clinical environments,” *IEEE J. Transl. Eng. Heal. Med.*, vol. 6, no. May, pp. 1–11, 2018, doi: 10.1109/JTEHM.2018.2875464.

[246] R. E. Kalman, “A New Approach to Linear Filtering and Prediction Problems,” *Trans. ASME-Journal Basic Eng.*, vol. 82, pp. 35–45, 1960.

[247] H. Neill, *Trigonometry: A complete introduction*. Mobius, 2018.

[248] J. Durbin and S. J. Koopman, *Time series analysis by state space methods*. OUP Oxford, 2012.

[249] D. Gafurov, “A survey of biometric gait recognition: Approaches, security and challenges,” *Annu. Nor. Comput. Sci. Conf.*, no. June, pp. 19–21, 2007.

[250] R. Chellappa, A. K. Roy-Chowdhury, and A. Kale, “Human identification using gait and face,” *Circuits, Signals, Speech Image Process.*, no. May 2014, 2007, doi: 10.1109/CVPR.2007.383523.

[251] J. Ahn, K. Nakashima, K. Yoshino, Y. Iwashita, and R. Kurazume, “2V-Gait : Gait Recognition using 3D LiDAR Robust to Changes in Walking Direction and Measurement Distance,” in *IEEE/SICE International Symposium on System Integration (SII).*, 2022, pp. 1–6.

[252] G. Ariyanto and M. S. Nixon, “Marionette mass-spring model for 3D gait biometrics,” in *5th IAPR International Conference on Biometrics, ICB*, 2012, pp. 354–359, doi: 10.1109/ICB.2012.6199832.

[253] J. H. Yoo and M. S. Nixon, “Automated markerless analysis of human gait motion for recognition and classification,” *ETRI J.*, vol. 33, no. 2, pp. 259–266, 2011, doi: 10.4218/etrij.11.1510.0068.

[254] T. W. Yeoh, F. Daolio, H. E. Aguirre, and K. Tanaka, “On the effectiveness of feature selection methods for gait classification under different covariate factors,” *Appl. Soft Comput.*, vol. 61, pp. 42–57, 2017, doi: 10.1016/j.asoc.2017.07.041.

[255] I. Bouchrika, J. N. Carter, and M. S. Nixon, “Towards automated visual surveillance using gait for identity recognition and tracking across multiple non-intersecting cameras,” *Multimed. Tools Appl.*, vol. 75, no. 2, pp. 1201–1221, 2016, doi: 10.1007/s11042-014-2364-9.

[256] M. Goffredo, I. Bouchrika, J. N. Carter, and M. S. Nixon, “Self-calibrating view-invariant gait biometrics,” *IEEE Trans. Syst. Man, Cybern. Part B Cybern.*, vol. 40, no. 4, pp. 997–1008, 2010, doi: 10.1109/TSMCB.2009.2031091.

[257] A. Khatun, S. Denman, S. Sridharan, and C. Fookes, “End-To-End Domain Adaptive Attention Network for Cross-Domain Person Re-Identification,” *IEEE Trans. Inf. Forensics Secur.*, vol. 16, pp. 3803–3813, 2021, doi: 10.1109/TIFS.2021.3088012.