International Journal of Intelligent Systems

WILEY

RESEARCH ARTICLE OPEN ACCESS

Enabling Passive Gait Identification in Realistic and Uncontrolled Environments Using Deep Learning and Spatiotemporal Biometrics

Luke K. Topham[1] | Wasiq Khan[1] | Dhiya Al-Jumeily[1] | Hoshang Kolivand[1] | Omar Aldhaibani[1] | Abir Hussain[2]

[1]School of Computer Science and Mathematics, Liverpool John Moores University L3 3AF, Liverpool, UK | [2]Department of Electrical Engineering, University of Sharjah, Sharjah, UAE

Correspondence: Luke K. Topham (l.k.topham@ljmu.ac.uk)

Received: 26 September 2025 | **Revised:** 16 December 2025 | **Accepted:** 12 March 2026

Academic Editor: Floriano Cuccureddu

ABSTRACT

Person identification is a pivotal challenge in the security domain, with important and impactful applications such as identifying crime suspects and locating missing persons. One convenient person identification method is gait identification, where individuals are identified by their unique walking style. However, traditional methods of gait identification are often affected by variations in appearance and occlusion. This work introduces a novel and robust spatiotemporal kinematics-informed non-invasive gait identification (STONI-GID) method that uses human pose estimation, occlusion state estimation and deep machine learning. Furthermore, unlike some existing methods, we demonstrate that our method remains unaffected by everyday appearance changes, environment, or viewing angle. Our approach achieved identification accuracy of up to 98.66% when evaluated using our primary dataset of 65 diverse participants in real-world environments. Moreover, the model outperformed existing methods during cross-dataset validation on the large Southampton dataset and the Gait Recognition Image and Depth Dataset (GRIDDS), achieving identification accuracies of 97.68% and 99.12%, respectively. Our findings will particularly advance the research frontiers of real-world gait identification and impact interdisciplinary areas of security and healthcare applications.

1 | Introduction

Person identification is a crucial problem in the security domain, with critical applications such as missing-person identification and crime-suspect identification. However, despite its importance, the media has reported several cases where law enforcement services have used person identification tools, resulting in high false-positive identification rates of up to 98% [1]. The reported tools utilise face matching tools (FMT), which are limited in real-time dynamic environments due to their sensitivity to pose variation, lighting, facial expression, image resolution, makeup, masks and occlusion [2]. Therefore, a reliable alternative to FMT is needed that incorporates machine automation.

Alternatively, gait identification, in which a person is identified by their manner of walking, may provide a robust person identification method, as gait cannot be easily disguised, and attempting to do so impedes natural movement [3]. Furthermore, gait-based person identification may be considered convenient, as it can be performed from a distance and without the knowledge or cooperation of the individual, making it an ideal approach for security applications such as identifying crime suspects.

In addition, existing computer vision methods leverage the benefits of gait-based person identification and can be performed using videos, such as those from security cameras. Traditional computer vision approaches to gait identification, as in Shiraga

et al. [4], are performed on gait energy images (GEIs), which are the average of the binary silhouettes over time. However, silhouette-based approaches are affected by appearance changes; for example, clothes or accessories can alter the shape of a silhouette [5]. Therefore, a cleaner representation of gait is desirable to separate gait features from the subject's appearance. One such approach is provided by skeleton-based gait identification methods, which use human pose estimation (HPE) to identify keypoints, such as bodily joints, to build a representation of the subject's skeleton to be used to analyse gait [5]. However, occlusion and viewing angles present a particular challenge to this approach.

This work proposes a new type of skeleton-based person identification in which joint angles are automatically calculated throughout the gait instance. Moreover, the approach aims to overcome several unresolved challenges commonly faced in existing solutions, namely, viewing angle, appearance and occlusion. Firstly, a deep learning model is presented to classify the viewing angle into one of three predefined groups (sagittal, oblique and frontal/rear). Secondly, a joint-based person identification approach is introduced to mitigate the effect of a modest appearance change. Lastly, a dynamic state estimation module is proposed, which substantially mitigates the adverse impact of gait occlusion on identification accuracy.

This work employs automatically derived joint angles, obtained using computer vision methods, for gait-based person identification under controlled benchmark evaluation settings. Our previous work Refs. [6, 7], which utilised a digital goniometer to measure joint angles, suggested the feasibility of the proposed method. Furthermore, this study investigates the effect of joint-angle–based representations on three challenges affecting traditional gait identification methods: viewing angle, appearance change and occlusion, within the constraints of the adopted evaluation datasets.

This study makes the following key contributions:

- The novel application of Kalman filtering for occlusion handling in unconstrained gait scenarios.
- A joint angle-based feature representation that enhances robustness across appearance changes and varying viewpoints.
- A multi-angle evaluation protocol using a self-collected dataset that includes clothing and viewpoint variation, complemented by cross-dataset evaluation on existing benchmarks where such variations are more limited.
- A practical system design that achieves high accuracy across multiple conditions without reliance on appearance-based features.

The remainder of this work is organised as follows. Related works are explored in Section 2. Our methodology is described in Section 3. Experimental results are presented in Section 4. A discussion of the results, limitations and future directions is provided in Section 5. Finally, conclusions are provided in Section 6.

2 | Related Work

As early as the 1970s, psychologists demonstrated that people can be recognised by their gait [8]. Early attempts at computationally identifying people, as in Popovic et al. [9], focussed on rule-based systems; however, significant progress was made after the introduction of machine learning. The literature considers three important aspects: existing gait datasets, existing gait-based person identification methods, and methods of overcoming occlusion in gait identification, as described in the following subsections.

2.1 | Existing Gait Datasets

Developing gait identification methods, particularly when using machine learning, requires gait datasets with many participants, each with many walking instances. However, many existing datasets are limited by the number and diversity (e.g., age, height, weight, gender, ethnicity, etc.) of participants, the diversity of real-world environments (e.g., environment), the number and consistency of viewing angles (e.g., comprising frontal perspective only), and accurate data labelling [10].

Table 1 presents existing video-based gait datasets comparing different aspects appropriate for gait-based person identification. Moreover, Table 1 suggests several common limitations exist in existing gait datasets; for example, Refs. [11, 12] provide only one instance per subject, making them unsuitable for gait identification.

Existing video-based gait datasets that compare different aspects relevant to gait-based person identification are described in our previous work [10]. Moreover, the dataset comparison in Topham et al. [10] suggests several common limitations exist in existing gait datasets; for example, Refs. [11, 12] provide only one instance per subject, making them unsuitable for gait identification. Similarly, only single instances are provided in Refs. [13, 14], as videos are captured in uncontrolled outdoor public environments where the recording angle and other important factors cannot be controlled; additionally, ethical issues such as consent may be a concern. Furthermore, availability is sometimes a significant issue, likely due to ethical issues surrounding personal data. For example, Refs. [15–17] are no longer publicly available, Refs. [18, 19] provide only silhouettes and not the original RGB video data, and Nunes et al. [20] provide only still images. Moreover, one of the most common limitations is a limited number of participants, as seen in Verlekar et al. [18], and a lack of diversity among participants [17, 21]. Appearance is known to influence person identification; however, only three datasets [15, 22, 23] provide subjects with changes in appearance. Similarly, only three datasets [16, 23, 24] provide gait records for both indoor and outdoor environments. CASIA-B [22] and OU-ISIR [25] are two of the most widely used gait identification datasets, owing to their large subject pools, diverse view angles and inclusion of various walking conditions, making them ideal for evaluating gait recognition algorithms under realistic and challenging scenarios. However, as described in Table 1, some limitations remain. In particular, both datasets are collected in controlled, lab-based environments, which may not fully reflect the complexities of real-world gait scenarios. In particular, OU-ISIR [25] is limited to treadmill walking.

Our recently published dataset, described in the Nature Scientific Data descriptor [23] and freely available at Topham and Khan [27], aims to overcome these limitations. Moreover, only our dataset [23] provides significant participant diversity, appearance variation commonly encountered in real-world scenarios,

TABLE 1 | A comparison of video-based gait datasets for person identification.

Ref.	No. subjects	No. instances	No. viewing angles	Indoor and outdoor	Alt. appearance	Labelling	Limitations
[18]	21	72	2	No	No	None	No RGB, silhouette only. Outdoor only.
[19]	294	2940	1	No	No	None	No RGB, silhouette only.
[20]	35	350	1	No	No	Depth sensor	Still images only.
[15]	305	3370	3	No	No	Depth sensor	Not available.
[16]	118	2280	4	Yes	Yes	Silhouettes	Not available.
[17]	60	1140	1	No	Yes	Emotions	Not available. Acted emotions, little participant diversity.
[21]	20	240	3	No	No	None	Little participant diversity.
[22]	124	13,640	11	No	No	None	Little participant diversity.
[11]	10,307	144,298	14	No	No	None	Only one instance per subject.
[26]	71	483	7	No	Yes	Depth sensor	Limited viewing angles.
[13]	1261	~20,000	6	No	No	None	Uncontrolled outdoor env.
[14]	18	216	12	No	No	None	Uncontrolled outdoor env.
[12]	983	1228	2	No	No	None	Mostly one instance per subject. Uncontrolled public env.
[24]	43	170	4	Yes	No	None	Uncontrolled outdoor env.
[25]	4007	4007	1	No	No	Silhouettes	Silhouette only, treadmill
Ours	64	3120	8	Yes	Yes	Keypoints, joints	No instances of multiple people.

multiple viewing angles and indoor and outdoor recordings. Our dataset [23] provides most of the advantages whilst minimising the limitations reported in Topham et al. [10], as further described in Section 3.1.

2.2 | Existing Gait Identification Methods

The two primary methodologies for gait identification are the traditional silhouette-based and skeleton-based approaches. In the silhouette approach, a binary silhouette image of the subject is extracted from an input image. The literature highlights several successful silhouette-based methods [28, 29]. Silhouette images are often further processed for gait identification; for example, GEI, gait energy image on depth data (depth-GEI), gait energy volume (GEV) and depth gradient histogram energy image (DGHEI) may be generated [15]. GEI, one of the most common silhouette methods, averages the silhouettes captured over a complete cycle and therefore assumes that all required gait information can be extracted from a single cycle [30]. However, silhouette-based approaches are affected by the subject's appearance, for example, clothing and accessories that alter the silhouette shape [5]. Furthermore, the literature fails to address the extraction of silhouettes from complex and dynamic real-world environments [5].

Alternatively, skeleton-based approaches leverage HPE for extracting body keypoints, such as joints, to construct a skeleton, which can then be used to analyse the subject's gait. Assuming the reliable skeleton extraction, this approach can effectively address the challenges arising from appearance or clothing. Moreover, the extracted skeletons may still be used to extract appearance-free silhouettes; for example, in the study by Yao et al. [31], skeleton gait energy images (SGEIs) are extracted, occlusion in gait identification.

Another unresolved challenge in gait identification is occlusion, which can manifest in various ways, such as self-occlusion, in which a part of the subject's body occludes another. Moreover, subjects may be wholly or partially occluded by other people or aspects of the environment, such as trees or vehicles. Several methods have been used to mitigate the effect of occlusion, often inferring missing joints in the case of partial occlusion. For example, in the study by Mehta et al. [32], occlusion-robust pose maps (ORPMs) are employed to produce full-body HPE despite substantial partial occlusion by utilising body part association to enable the inference of 3D poses. Similarly, in the study by Zhen et al. [33], a set of 2.5D representations of body parts is regressed, and 3D poses are reconstructed using a depth-aware part-association algorithm by reasoning about inter-person occlusion and bone-length constraints. Furthermore, in Ref. [34], interpolation is used to estimate gait information lost due to occlusion. However, the method requires depth sensors, which are not readily available in real-world dynamic environments. Alternatively, in Ref. [35], an occlusion-aware model is presented, which fits a skinned multiperson 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., 60% or more of the body is occluded).

Furthermore, recent works in gait recognition exhibit similar limitations that hinder their effectiveness and generalisability in real-world applications. For example, Zhang et al. [36] present an AutoEncoder-based method called GaitNet, which explicitly disentangles pose and appearance features from RGB images. A long short-term memory (LSTM) integrates pose features over

time to produce the gait representation. Their approach learns features directly from walking videos rather than from GEIs or skeleton data and demonstrates notable success in challenging frontal-view scenarios. However, while aiming to separate pose and appearance, the method does not provide strong guarantees against latent leakage, where appearance information may inadvertently contaminate the pose embedding—a well-documented issue in unsupervised disentanglement frameworks, as highlighted by Locatello et al. [37]. Alternatively, Ye et al. [38] propose BigGait, which leverages large vision models (LVMs) to learn robust gait representations directly from silhouette sequences, achieving improved recognition performance with minimal task-specific supervision. Nonetheless, silhouette-based models are prone to encoding appearance-based features such as carried objects, which complicates their use as purely gait-identification methods. In contrast, Ma et al. [39] introduce a framework employing a trainable part-based prompt pool and dynamic attention within a transformer architecture to enhance gait recognition under real-world challenges, including occlusions and clothing variations. Despite these claims, their work lacks a detailed breakdown of performance relative to occlusion severity or clothing changes, thereby limiting the ability to fully evaluate the method's robustness under these critical conditions. Although these studies report strong accuracy and computational efficiency, they fall short of fully addressing key challenges such as appearance variation, environmental diversity, participant heterogeneity and occlusion, thereby constraining their applicability in realistic, real-world environments.

Unlike the majority of studies that rely on computer vision, an increasing number of recent investigations have explored the use of reflected radio frequency (RF) signals for unobtrusive gait-based person identification. For instance, Jiang et al. [40] propose RF-Gait, which utilises commercial off-the-shelf RFID tags and readers to capture and decompose walking-induced signal fluctuations through multivariate variational mode decomposition, subsequently employing a support vector machine (SVM) classifier to achieve approximately 96% accuracy in identifying individuals within medium-sized groups in realistic indoor environments. Similarly, Yang et al. [41] introduce RF-UI, a low-cost RFID-based system for continuous user identification that robustly separates and analyses gait patterns by leveraging phase energy fluctuations and a Joint Similarity Matrix, demonstrating over 94% accuracy even under interference and multiuser conditions. Alternatively, Li et al. [42] present RDGait, a single-chip millimetre-wave (mmWave) radar system integrating a velocity-based 'ghost' detection algorithm, where 'ghosts' denote false or spurious reflections unrelated to actual human motion, and an attention-enhanced recurrent neural network (RNN) to reliably extract gait features, achieving over 95% recognition accuracy across 125 individuals in complex indoor environments and maintaining above 90% accuracy with minimal fine-tuning in novel settings. However, despite these promising advances, the availability of RF data remains limited in public spaces, restricting current applications predominantly to controlled environments such as airports; furthermore, the efficacy of RF-based gait identification in dynamic, outdoor public settings warrants further investigation, particularly given challenges related to signal interference from environmental factors [43], variability in user behaviour, privacy concerns and the need for scalable, real-time processing under diverse and unpredictable conditions.

Despite the advances, solutions for performing HPE and gait identification under temporary complete occlusion, for example, when the subject walks behind obstructions such as trees or vehicles, have not been described in the literature. Despite the plethora of available literature, several gait identification problems remain unsolved. For example, although methods such as SGEIs may potentially overcome the issue of appearance, this has not yet been satisfactorily proven [10]. Furthermore, the lack of a comprehensive multi-appearance dataset has likely limited the exploration of such research, a gap we endeavoured to address in our recently published dataset [23]. Additionally, the problem of both self-occlusion and environmental occlusion has not been sufficiently addressed in real-world dynamic environments, thereby limiting its applicability.

3 | Methodology

3.1 | Dataset

This work leverages our recently published multimodal (video, sensor) gait dataset, described in Ref. [10] and publicly available at Ref. [27]. The dataset contains 64 participants with diverse ages, genders, heights, weights, ethnicities and recording environments, as described in Table 2. Only subjects aged 18 years or older with a healthy gait are included. The dataset comprises approximately 3120 video sequences containing approximately 748,800 image frames, with 25 keypoints labelled per frame, yielding approximately 18,720,000 keypoints in total.

Further diversity is presented in terms of the recording environment, including indoor and outdoor recordings, various viewing angles and a variety of real-world backgrounds, a sample of which is shown in Figure 1. The dataset includes modest appearance changes, primarily involving the addition or removal of outerwear to reflect realistic yet limited variations in subject appearance. Each participant completed 24 walks, which were recorded using two digital cameras and the MOTI digital goniometer, a device commonly used by physiotherapists to measure joint range of motion. MOTI goniometer data were used in Refs. [6, 7] for prototyping joint-based gait identification. However, it is not used in this work, as such sensors cannot be used covertly, unlike video data.

The recordings were partitioned into three experiments, each with a different camera layout. In each recording, the subject performs their usual walk for 8 m. Two cameras are positioned on a line parallel to the walking line, 5.8 m from it. In Experiment 1 (E1), one camera is placed at the centre of the 8 m recording line at a 90° angle to the walking angle, and the second camera is placed at one end of the line at a 45° angle to the walking line. In Experiment 2 (E2), the walking and recording lines are swapped to generate a mirrored setup relative to E1. In Experiment 3 (E3), the cameras are placed at each end of the walking line to provide a frontal and rear view of the subject. Full descriptions and illustrations of the camera setup for each experiment and recording 'take' are supplied [23]. Furthermore, the variety of experiments and scenarios generates eight different viewing angles, providing a 360° view of each participant's gait in 45° increments, as shown in Figure 2.

For this study, the dataset is partitioned into training and testing subsets, which are processed independently. Specifically, 80% of

TABLE 2 | The diversity of participants present in our gait dataset.

Gender	Age	Height (cm)	Mass (kg)
Male (57.81%)	35.14 (±12.44)	170.83 (±10.86)	81.81 (±20.18)
Female (42.19%)	Max: 65 Min: 18	Max: 192 Min: 137.16	Max: 133.36 Min: 51

the gait videos for each subject and each recording-angle group (as defined in Section 3.2) are allocated to the training set, while the remaining 20% are allocated to the test set. Although both subsets undergo the same preprocessing steps outlined in Section 3.2, only the training data are used to train the proposed models, with the test set reserved exclusively for evaluation.

Further evaluation and cross-dataset validation of the proposed spatiotemporal kinematics-informed non-invasive gait identification (STONI-GID) method are performed using additional unseen partitions from Gait Recognition Image and Depth Dataset (GRIDDS) [20] and SOTON Large [16]. Both datasets were selected because they are publicly available, provide numerous instances per participant, and offer sufficient diversity among participants. Unlike many popular modern datasets, they provide real-world environments. In addition, the long-term availability of SOTON Large [17] has provided a plethora of benchmark results from several works that may be used to compare the performance of STONI-GID. For example, the GRIDDS dataset provides 10 walking sequences for each of its 35 participants in indoor environments [20]. In contrast, the SOTON Large dataset comprises eight walking sequences for 118 participants, recorded in both indoor and outdoor environments. Cross-dataset validation enables the evaluation of the proposed STONI-GID approach in terms of its reliability and generalisation 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 [16, 20].

3.2 | Gait Identification Method Overview

Figure 3 demonstrates an overview of the proposed STONI-GID methodology; each component is described in further detail in the following subsections. Firstly, each image frame is extracted from the input video recording. Secondly, as in related work such as Ref. [44], we used the OpenPose [45] HPE method, which extracts 25 bodily keypoints from each frame. Next, the position of each occluded keypoint is estimated using the Kalman state estimation filtering [46]. Then, 10 joint angles are calculated for the left and right sides, including the elbow, shoulder, hip, knee and ankle. Then, the data are 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.

3.3 | Spatiotemporal Feature Extraction

Algorithm 1 provides an overview of the spatiotemporal feature extraction method. Using the OpenPose [45] HPE method, 25 bodily keypoints are extracted from each frame. Next, the position of each occluded keypoint is estimated using the Kalman state estimation filtering [46]. Then, 10 joint angles, including elbow, shoulder, hip, knee and ankle for both the left and right sides, are calculated as per Equation 1, described in further detail in Neill [47]. Given an angle A $\longrightarrow$ B $\longrightarrow$ C, where A, B and C are keypoints, and B represents the desired joint to be measured.

FIGURE 1 | Examples of frames from the gait dataset include diverse participants and various environments and viewing angles. From the top left, moving clockwise, these examples show an indoor female frontal view (180°), an indoor female right sagittal view (90°), an outdoor male right oblique profile (315°), and an indoor male left sagittal profile (270°), respectively. Images used with participants' consent.

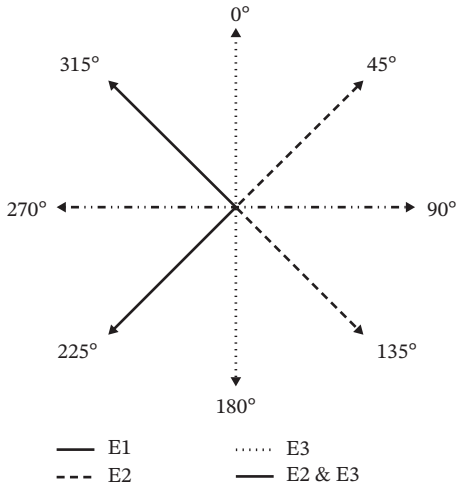

FIGURE 2 | Recorded camera angles with the subject at the centre and the camera at the arrowhead from the three experiments (E1, E2 and E3).

Multiple studies have shown these joints, such as Refs. [6, 7], possess relevant gait information for person identification.

$$\theta = \arccos\left(\frac{\overrightarrow{BA}\cdot\overrightarrow{BC}}{\|\overrightarrow{BA}\|\|\overrightarrow{BC}\|}\right). \quad (1)$$

The data are then transformed into three-dimensional samples (i.e., several instances are 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 is equivalent to the average gait cycle length, found to be 0.965 s [6]. This length ensures that each sample contains sufficient patterns regarding each stage of the gait cycle. Next, an LSTM [48] model classifies 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).

LSTM is a RNN that learns order dependence in sequential problems [48]. It contains four neural networks and various memory blocks known as cells. The cells are manipulated by three gates: the forget, input and output gates. See the original article for further details [48].

3.4 | Occlusion State Estimation

To address the challenges arising under occlusion, an occlusion-filtering module based on a Kalman state-estimation filter [49] is employed for the first time in this study and is described in Algorithm 2. The procedure takes a 2D array as input, where each row represents an image frame and contains 50 columns for the 25 skeletal keypoints (each keypoint has x and y coordinates in separate columns). This procedure iterates over each row of the given array and, for each cell (keypoint coordinate) that is null (i.e., occluded), applies state estimation. The filtering process is skipped for a cell if there are no occlusions or missing values.

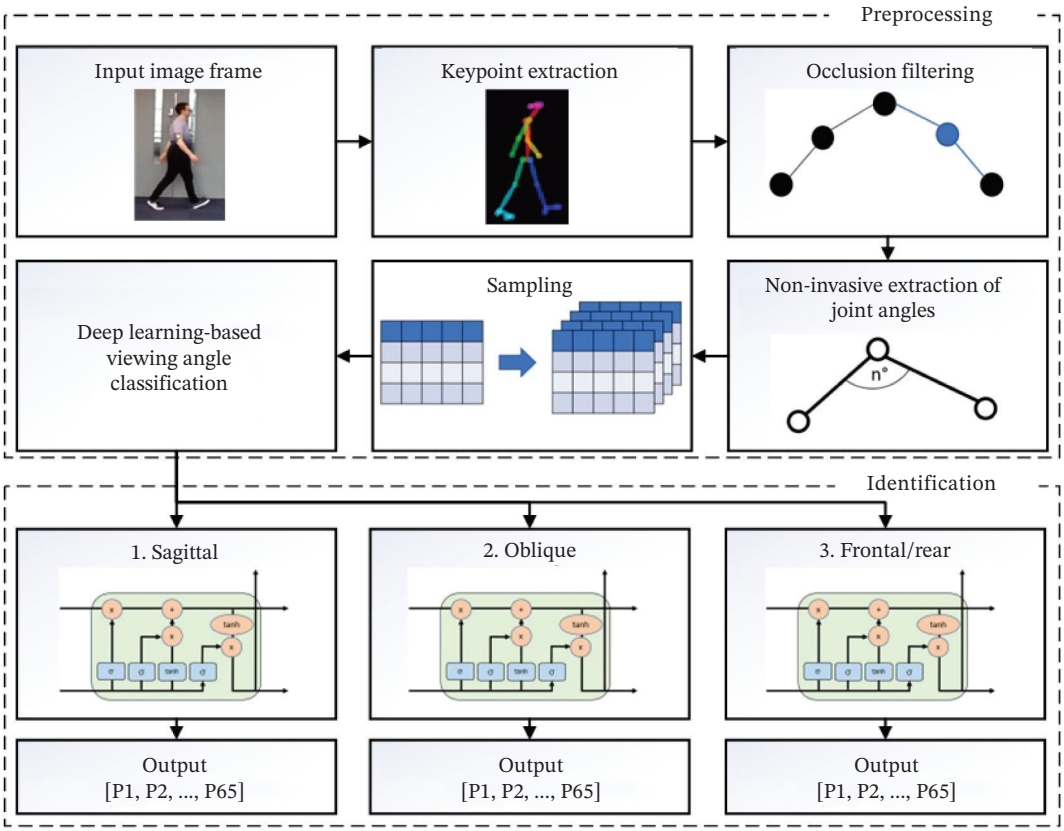

FIGURE 3 | Overview of the gait identification methodology, including input RGB images, skeleton keypoint extraction, occlusion filtering of any missing keypoints or frames, bodily joint angle extraction, sampling, angle group classification (sagittal, oblique and frontal/rear), LSTM person identification and the resulting output in terms of participant ID number.

ALGORITHM 1 | Feature Extraction.

1. Let $\boldsymbol{V}$ be a collection of videos, $\boldsymbol{v}$ be a video in $\boldsymbol{V}$, $\boldsymbol{F}$ be a video frame, $\boldsymbol{K}$ be an array of keypoints, $\boldsymbol{k}$ be a subarray of $\boldsymbol{K}$, $\boldsymbol{l}$ be the window length (number of frames in average gait cycle), $\boldsymbol{w}$ be an array containing a single window, $\boldsymbol{Fe}$ be an array of features, $\boldsymbol{s}$ be a range of statistical methods, $\boldsymbol{J}$ be a bodily joint containing three keypoints ($A \longrightarrow B \longrightarrow C$), and θ be a joint angle in degrees.

```
procedure EXTRACTKEYPOINTS(v)
    for F in V do
        i = 0
        row = []
        while i < 25 do
            row = row.append(OpenPose.ExtractKeypoint())
            i = i + 1
        end while
        Arr = Arr.append(row)
    end for
    return Arr
end procedure
procedure EXTRACTFEATURES(K)
:   i = 0
    Fe = []
    while i < K.length() do
        k = K[i: i + l]
        w = CalculateFeatures(k)
        Fe.append(w)
        i = i + (l)
    end while
    return Fe
end procedure
procedure CALCULATEFEATURES(k)
    C = []
    W = []
    for J in k do
        θ = arccos((BA⃗ · BC⃗)/(||BA⃗|| ||BC⃗||))
        C.append (θ)
    end for
end procedure
K = ExtractKeypoints(v)
Arr = ExtractFeatures(K)
```

After each iteration, the filtered row is appended to an array of filtered rows. The filtered rows can then be passed to the gait identification task.

As described in Section 2.1, no existing dataset contains labelled occlusion. Therefore, it is necessary to simulate the required occlusion, as both described and supported in Peng et al. [50]. However, it does not guarantee full-body occlusion. Furthermore, to the best of the authors' knowledge, no prior work has simulated occlusion with significant full-body occlusion; hence, there is a need to develop an occlusion simulation strategy. When occlusion affects HPE methods, the corresponding keypoints are

typically assigned null values, indicating a failure to localise joints. To simulate such conditions, we adopt a strategy of assigning null values to keypoints, effectively modelling occlusion. To evaluate the occlusion filtering model under severe conditions, we assume a full-body occlusion in which all keypoints are entirely undetectable. Accordingly, each entry in the array representing the body's keypoints is set to null, thereby simulating complete loss of pose information.

In this work, we simulate four distinct levels of occlusion. Table 3 describes the four varieties of simulated occlusion implemented to evaluate the occlusion filtering method. The array of keypoints for each variety is partitioned into blocks based on the corresponding block size. The first n rows are filled with null values, while the remaining rows remain unchanged. The 1-in-2 and 1-in-10 profiles simulate light occlusion or reduced frame rates, with occlusion effects that 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.

3.5 | Kinematics-Informed Deep Angle Detection

Gait identification is affected by the viewing angle [10]. Therefore, it is impossible to perform gait identification without either multiple identification models or preprocessing to remove the effect of the angle. For the first time, this study explores non-restrictive perspectives on gait identification, which is often required in realistic settings. Through systematic testing, we found that gait identification could be performed robustly when the gait data were divided into three groups based on viewing angle (i.e., camera perspective). Group 1 consisted of sagittal profile walks captured at 90° and 270° angles relative to the subject. Group 2 includes the oblique walks captured at 45°, 135°, 225° and 315°. Group 3 contains the frontal and rear views captured at 0° and 180°, respectively. Angle classification is performed for each gait cycle; thus, if an incorrect angle is predicted for a single cycle or the subject changes direction, the impact is mitigated by the correct classification of the other cycles within the sequence.

The LSTM model configuration for angle detection is provided in Table S1 of the supporting information. The model receives 60 input features, with the first 50 corresponding to the x and y coordinates for each of the 25 keypoints, and the remaining 10 corresponding to the joint angles listed previously (Section 3.3). The model consists of five layers: two LSTM layers, one dropout layer, and two dense layers, with the final layer producing the relevant viewing angle classification. The first LSTM layer has an output shape of 29,100, corresponding to the sequence length of 29, which matches the average gait cycle length when measured as the number of data points, as described previously. The model is trained with the training dataset described in Section 3. Training is performed for 50 epochs with a batch size of 5. This process is repeated 10 times for cross-validation. The model is then evaluated with the purely unseen reserved testing dataset described in Section 3. Detailed statistical results of the proposed LSTM-based angle detection are provided in Section 4.

TABLE 3 | Descriptions 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 second 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 sequential rows are nullified for every 10 rows.
10-in-20	20	10	10	Simulates environmental occlusion. Ten sequential rows are nullified for every 20 rows.

ALGORITHM 2 | State Estimation Filtering for Occluded Samples.

Let $\boldsymbol{V}$ be a collection of videos, $\boldsymbol{F}$ be a video frame, ***Arr*** be an array of keypoints where one row provides an entire skeleton of keypoints [25], $\boldsymbol{R}$ be a row in ***Arr*** and ***Col*** be a column in ***Arr***, ***Ce*** be a cell in a row of ***Arr***, where a cell contains a coordinate (x or y) of a keypoint, $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ be adaption matrices to convert input states to process states, $\boldsymbol{H}$ be the conversion matrix, $\boldsymbol{K}$ be the Kalman gain, $_K$ be the iteration number, $\boldsymbol{P}$ be the process covariance matrix, $\boldsymbol{Q}$ be the process noise covariance, $\boldsymbol{U}$ be the control variable, $\boldsymbol{W}$ be the predicted state noise, $\boldsymbol{X}$ be the skeletal keypoint state matrix, $\boldsymbol{Y}$ be the measurement of the keypoint state, and $\boldsymbol{Z}$ be the measurement of noise.

Inputs: Array (***Arr***) of keypoints and joint angles.

Outputs: Filtered array ($\boldsymbol{F}$) of keypoints and joint angles.

procedure UPDATE()

$\widehat{X}_K = AX_{K-1} + BU_K + W_k$

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

$Y_K = CX_K + Z_K$

return $\widehat{X}_K, \widehat{P}_K, Y_K$

end procedure

procedure PREDICT()

$K = (\widehat{K}_K H)/(H\widehat{P}_K H^T)$

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

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

return P_K

end procedure

for each Row in Arr **do**

  for each Col in Row **do**

    if is null(Ce) **then**

      Ce = Predict(X_K, P_K)

      X_K, P_K = Update()

    else

      Skip occlusion filtering for non-occluded cells

    end if

  end for

end for

3.6 | Gait-Based Person Identification Model

A description of the LSTM model configuration for gait-based person identification is provided in Table S2 of the supporting information. The configuration shown in Table S2 was empirically selected due to its success in our previous work, for which full details are available in Topham et al. [7]. As with the angle classification model described in Section 3.5, 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. Similar to the angle classification model, this model comprises five layers: two LSTM layers, one dropout layer and two dense layers, with the final layer producing the predicted viewing angle. Again, the first LSTM layer has an output shape of 29,100, corresponding to the sequence length of 29, which matches the number of data points equivalent to the average gait cycle length, as described previously.

As described in Section 3 and Figure 3, three LSTM models exist for person identification, one for each viewing-angle group. Each of the three models is trained with the training dataset described in Section 3. However, each model is trained only on samples relevant to the viewing-angle group, with the remainder excluded. Training is performed for 50 epochs with a batch size of 5. This process is repeated 10 times for evaluation. The models are then evaluated with the reserved unseen testing dataset described in Section 3. The results of this process are provided in Section 4.

The proposed LSTM-based identification model is trained in a supervised multiclass classification setting. Given an input gait sequence, the network outputs a probability distribution over all

identities using a softmax layer. Since class labels are encoded as integer values, the model is trained using sparse categorical cross-entropy loss, defined as

$$\mathcal{L} = -\log(\hat{y}_c), \tag{2}$$

where c denotes the ground-truth identity and $\hat{y}_c$ is the predicted probability for that identity.

Training is performed using the Adam optimiser with a fixed learning rate. The model is trained for a predefined number of epochs with mini-batch optimisation, and the final model is selected based on validation performance.

3.7 | Evaluation Metrics

We evaluate the performance of the proposed system using the following metrics. Let *TP*, *TN*, *FP* and *FN* denote true positives, true negatives, false positives and false negatives, respectively.

- **Accuracy:** the proportion of correctly classified instances:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{3}$$

- **Precision:** the proportion of correctly predicted positive instances among all predicted positive instances:

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{4}$$

- **Recall:** the proportion of correctly predicted positive instances among all actual positive instances:

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{5}$$

- **F1-score:** the harmonic mean of precision and recall:

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{6}$$

- **Cohen's Kappa for multiclass identification:** measures agreement between predicted and true labels, accounting for chance agreement across all classes:

$$\kappa = \frac{p_o - p_e}{1 - p_e}, p_o = \frac{\text{number of correctly classified instances}}{\text{total instances}}, p_e = \sum_{c=1}^{C} p_c^{\text{pred}} \cdot p_c^{\text{true}}, \tag{7}$$

where p_c^{pred} is the proportion of instances predicted as class c and p_c^{true} is the proportion of true instances in class c.

In this work, the task is one-to-many authentication (identification), where a probe instance is matched against a gallery of multiple subjects. This is distinct from one-to-one verification, where a probe is compared with a single claimed identity.

3.8 | Cross-Dataset Evaluation Protocol

In cross-dataset evaluation, the LSTM identification model is trained on 65 identities from our self-collected dataset and evaluated on external datasets (SOTON Large and GRIDDS), which contain different sets of identities and numbers of identities. For each external dataset, the model independently performs identity identification on the identities present in that dataset. All evaluation metrics (accuracy, precision, recall, F1-score and Cohen's kappa) are computed using the actual labels of the target dataset. This protocol ensures that the reported performance reflects the model's ability to generalise across datasets with different identities and that metric values are comparable within each evaluation set.

4 | Results

The following subsections provide the comprehensive results for the proposed gait identification method (STONI-GID) described in Section 3. This includes experimental outcomes of the proposed kinematics-informed angle detection model (Section 4.1), gait identification under normal circumstances (Section 4.2), and gait identification results with and without the gait occlusion state estimation module under simulated occlusion (Section 4.3). Furthermore, the results of experiments using alternative clothing (Section 4.4) and alternative environments (Section 4.5) are also described in this section. Finally, the results of cross-dataset validation performed on two additional datasets, SOTON Large [16] and GRIDDS [20], are provided (Section 4.6).

Algorithm 3 presents an overview of the experimental setup used to train and evaluate STONI-GID. It includes the setup for the angle detection module and various aspects of the person identification module, including experiments to test the effects of routine appearance changes, viewing angle changes, alternative participant appearances, changes in viewing environment, and cross-dataset validation.

4.1 | Angle Detection

The results of the proposed angle detection model (as described in Section 3.5) are shown in Figure 4 and Table S3 of the supporting information. The mean and standard deviation of 10 repetitions of training and evaluation are provided. A high accuracy of 98.87% is achieved when evaluating the viewing angle group classifier on the unseen partition of our primary dataset. Moreover, consistency is reported for the additional metrics F1, Kappa, precision and recall, which are reported as 98.82%, 98.29%, 98.92% and 98.75%, respectively. Thus, indicating a high level of performance.

Recent literature, as Wen et al. [51], highlights the need for an angle detection module. This literature indicates that viewing angles affect gait identification performance. Therefore, without a method to address the problems introduced by viewing angles,

ALGORITHM 3 | Experimental Steps for the Evaluation of the Method.

Let ***F*** be a feature vector containing all windowed features (from Algorithm 1), ***CA*** be the angle detection classifier (LSTM), ***CI*** be the person identification classifier (LSTM), ***OA*** {1,2, 3} be the output of the angle detection classifier (***CA***), ***OI*** {1, 2, ..., 64} be the output of the person for the identification classifier (***CI***), ***AP*** {1, 2, ..., 64} be the output for the alternative appearance test set, and ***E*** {1, 2, ..., 64} be the output for the alternative environment test set, ***ATS*** be the angle training set, ***AVS*** be the angle test set, ***ITS*** be the identification training set, ***IVS*** be the identification test set, ***APTS*** be the alternative appearance identification test set, and ***ETS*** be the alternative environment identification test set.

Inputs: Array (***Arr***) of windowed features (from Algorithm 1).

Outputs: ***OA***, ***OI***.

Prepare training data for CA:

1. ATS = random 80% of records per angle group from Arr
2. AVS = remaining 20% of records per angle group from Arr

Prepare training data for CI:

1. ITS = random 80% of records per participant from Arr, excluding alternative appearance and alternative environment
2. IVS = remaining 20% of records per participant from Arr, excluding alternative appearance and alternative environment
3. APTS = remaining alternative appearance records
4. ETS = remaining alternative environment records

Train classifiers until convergence:

1. Train CA on ATS
2. Train CI on ITS

Test classifiers (10 iterations): 1. OA = CA(AVS)

2. OI = CI(IVS)
3. OI_{AP} = CI(APTS)
4. OI_{E} = CI(ETS)

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 a gait identification model for each group of angles, we can overcome the issues caused by viewing angle, as reported in the following subsections.

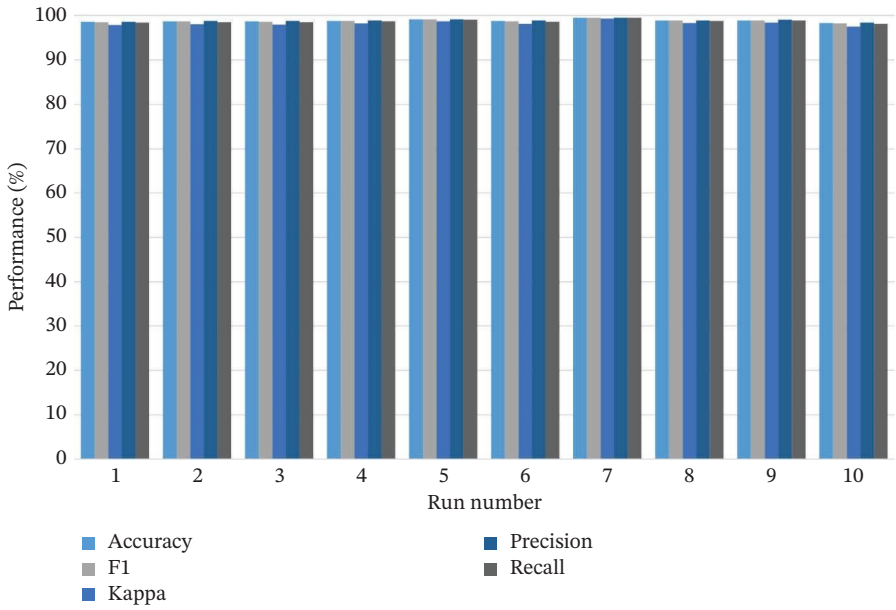

FIGURE 4 | Results of the proposed angle detection model (repeated with 10 models) were validated over a purely unseen partition of our test dataset.

TABLE 4 | Results from the gait identification models 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)

4.2 | Gait Identification Experimental Results

The statistical results of the three gait identification models (i.e., oblique view, sagittal view, frontal view and rear view) described in Section 3.5 with their corresponding angle groups, evaluated on our primary dataset, are shown in Table 4. Furthermore, the results, when tested with all gait samples for all angle groups combined, are shown in the final row of Table 4. It can be observed that all models exhibit robust performance and, importantly, similar outcomes across all three groups of angles. However, the group containing the front and rear-facing cameras (0° and 180°) performs slightly better. This suggests that this angle may be particularly useful for identifying subjects, perhaps because visible features (such as stride width) are more readily extracted than at other angles when using 2D HPE.

4.3 | Gait Identification With Occlusion Filtering Results

The statistical outcomes of gait identification under the simulated occlusion conditions, as described in Section 3.4, are shown in Table 5. It is important to note that the results shown in Table 5 are collected from our combined dataset (all angle groups) without applying the proposed occlusion state estimation model described in Section 3.4. As expected, Table 5 clearly shows that the simulated occlusion imposes significantly negative impacts on the gait identification performance, with varying levels of impact depending on the occlusion severity level. Moreover, the 1-in-2 occlusion method, in which every second frame is missing, has the greatest impact, likely due to the largest number of rows affected. In comparison, the 1-in-10 occlusion method has the least impact on the results, likely because it affects the fewest rows in the data. The 5-in-10 and 10-in-20 methods reduce person identification performance more than the 1-in-10 method, likely because more frames are affected. 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. Additionally, both profiles leave consecutive frames unaffected, a characteristic also present 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 by Fox [1] reports poor real-world performance of person identification by police forces. Therefore, this poses a significant challenge to achieving robust person identification, which will have clear benefits in the security domain, for example, for law enforcement agencies such as the police.

To address the challenge posed by occlusion, the Kalman filter-based approach to HPE recovery, as detailed in Section 3.5, is applied to data affected by the simulated occlusion. A qualitative comparison across five consecutive frames is presented in Figure 5. This figure contrasts the ground-truth poses ('Original'), derived from non-occluded sequences, with the estimated ('Recovered') poses obtained from data processed under our simulated occlusion method, described in Section 3.4. The results in Figure 5 indicate successful recovery with minimal visible errors. Additionally, Figure 6 provides a close-up view of frame 4 from this sequence, with both the original and recovered poses overlaid. Minor discrepancies are observable at the wrist, shoulder and right ankle; nonetheless, the overall recovery is accurate. Notably, the right arm, which was entirely self-occluded in the input, has been successfully estimated, demonstrating the effectiveness of the recovery approach under challenging conditions.

The results displayed in Table 6 demonstrate that both linear interpolation and the Kalman filter can recover occluded keypoints across a range of occlusion patterns. For short-term occlusions, such as 1-in-2, interpolation achieves a slightly lower RMSE and higher recovery rates, reflecting its ability to effectively fill small gaps. Recovery rate, defined as the proportion of occluded keypoints whose position is estimated within a 10-pixel error threshold, provides an interpretable measure of practical accuracy. As occlusion duration increases, for example, in the 10-in-20 pattern, the Kalman filter outperforms interpolation in both recovery rate and RMSE, demonstrating its strength in maintaining physically plausible trajectories over longer gaps.

TABLE 5 | 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

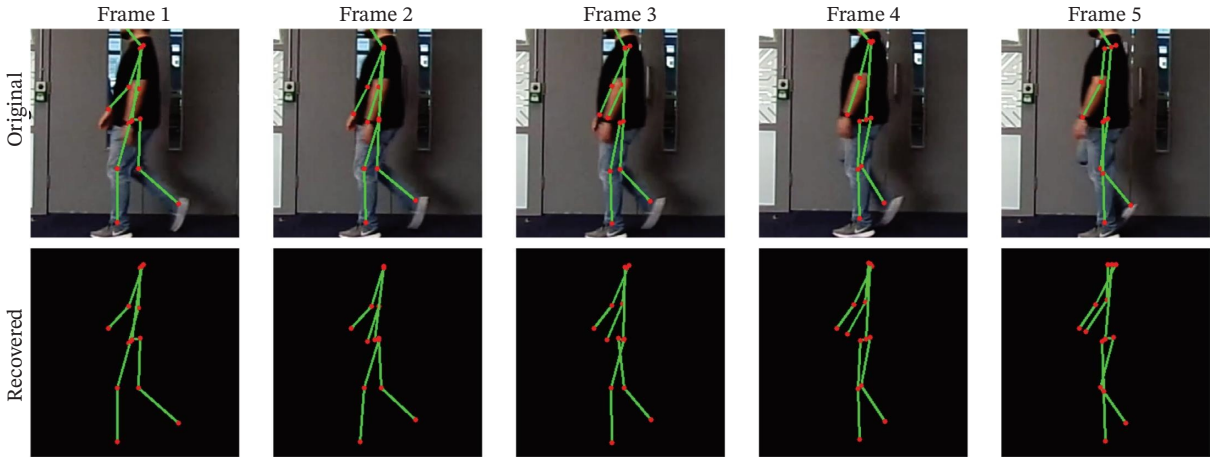

FIGURE 5 | Qualitative comparison of HPE across consecutive video frames. The top row ('Original') shows cropped RGB frames with the original detected poses. The bottom row ('Recovered') displays the corresponding Kalman-filtered poses rendered on black backgrounds. Each column represents a distinct image frame in the sequence, illustrating how poses may be recovered or estimated under temporary complete occlusion.

It is important to note that RMSE and recovery rate capture different aspects of performance. While the recovery rate focuses on the fraction of keypoints estimated very accurately, RMSE incorporates all keypoints, including those with large deviations, and is therefore sensitive to occasional extreme errors. This explains why RMSE values remain high even when a substantial fraction of keypoints are recovered within the 10-pixel threshold. Overall, these results highlight the predictive capability of the Kalman filter: by leveraging motion dynamics and backward smoothing, it maintains temporal continuity and robustness, which are critical for reliable person identification under partial or intermittent occlusion, particularly when occlusions are prolonged.

To validate the effectiveness of the proposed STONI-GID method under occlusion, the statistical outcomes of gait identification under the simulated occlusion conditions, with the occlusion state estimation filtering performed, are shown in Table 7. The results for all metrics significantly improve over those shown in Table 5 (without the proposed 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, the gait identification accuracy substantially dropped from 98.29% (as shown in Table 4) to 42.71% (as shown in Table 5). Moreover, F1, kappa, precision and recall are similarly affected. However, Table 7 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 5) and recovers to 95.17% with state estimation filtering (Table 7). Correspondingly, the 5-in-10 occlusion method causes identification accuracy to drop to 61.04% (Table 5), and state estimation filtering recovers the performance to 82.89% accuracy (Table 7). Finally, the 10-in-20 occlusion method causes the identification accuracy to drop to 52.27% (Table 5), and state estimation filtering recovers the performance to 69.98% accuracy (Table 7).

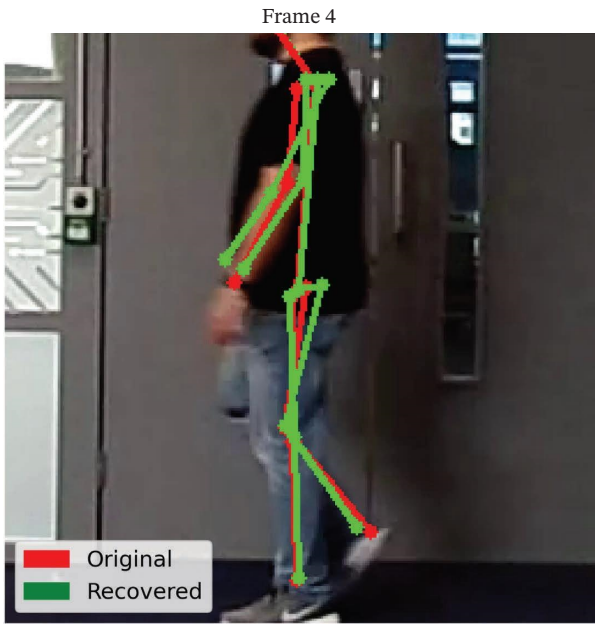

FIGURE 6 | Overlay of original (red) and recovered (green) poses on a single RGB frame.

4.4 | Gait Identification With Alternative Appearance

As discussed previously, appearance has been shown to affect the accuracy of person identification methods [31, 51], including some gait identification methods. 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 accurately identified regardless of appearance, then a subject's appearance should not impede their identification when using skeleton-based gait identification methods [52]. Existing available datasets did not provide alternative appearances to investigate this hypothesis sufficiently; however (as described in Section 3.1), our dataset [23] provides additional recordings where participants change their appearance, for example, 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 appearances. We utilise the previously unseen

TABLE 6 | Keypoint recovery performance under different occlusion scenarios using interpolation and the Kalman filter-based approach.

Occlusion pattern	Interpolation		Kalman filter	
	RMSE	Recovery rate	RMSE	Recovery rate
1-in-10	275.936	0.826	279.460	0.811
1-in-2	264.831	0.852	265.629	0.826
5-in-10	277.163	0.754	279.045	0.758
10-in-20	292.000	0.617	288.176	0.639

Note: RMSE is reported in pixels (lower is better), and recovery rate denotes the proportion of occluded keypoints recovered within a 10-pixel error threshold (higher is better).

TABLE 7 | 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

partition of our dataset [23], further dividing appearance into standard and alternative categories for comparison. Table 8 presents the average gait identification results. The evaluation set is partitioned into standard and alternative appearances and evaluated using the 10 iterations of the identification models described previously (Section 3.5). It can be noticed in Table 8 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 yields a p-value of 0.069 under an alpha of 0.05, suggesting no statistical significance between the means of the two experiments. Therefore, it is concluded that the proposed kinematics-informed deep identification model is not significantly affected by common appearance changes of the subject.

Moreover, the finding that STONI-GID can resolve appearance changes is significant, as previous work has identified appearance changes as a challenge to gait identification [52]. 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.

4.5 | Gait Identification in Realistic Environments

One advantage of our primary dataset [23] is that it includes examples of indoor and outdoor real-world environments. Previous work has shown that performance may decrease outdoors [53], likely due to the uncontrolled nature of the environments, and sensors such as depth cameras may be challenging to access outdoors [54]. Therefore, this dataset provides an opportunity to investigate the proposed STONI-GID's performance using indoor and outdoor data. Table 9 displays the average results of the gait identification, where the evaluation set is partitioned into indoor and outdoor recordings and evaluated using the 10 iterations of the identification models, as described previously. Overall, a 98.98% accuracy is achieved indoors, whereas 97.95% accuracy (with no significant difference) is achieved outdoors. Similar, consistent F1, kappa, precision and recall results are observed. Furthermore, a two-sample t-test yields a p-value of 0.9, assuming an alpha of 0.05; this suggests no statistical significance between the means of the two experiments. Therefore, it may be suggested that the recording environment does not significantly affect the proposed STONI-GID identification model.

To reiterate, gait identification in outdoor environments has been described as a particular challenge to real-world person identification [51]. 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 authors' 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 environmental factors is another crucial step towards a reliable, non-restraining and non-invasive solution for person identification.

4.6 | Cross-Dataset Validation

In addition to test data (Section 3.1), we employed cross-dataset validation to evaluate the proposed STONI-GID's performance and its 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 [16] and the GRIDDS dataset [20] while ensuring the same training and evaluation methods described in Section 3.6. Tables S4 and S5 in the supporting information present the statistical results obtained using the SOTON Large [16] and the GRIDDS [20] datasets, respectively. Similarly, the results are visualised in Figures 7 and 8.

4.6.1 | Validation on the SOTON Large Dataset

The SOTON Large dataset [16] is a video sequence gait dataset comprising 115 participants, recorded from two angles (including a side profile and an oblique view). Table S4 (supporting information) presents the statistical results of the proposed STONI-GID trained on entirely different data (our primary

TABLE 8 | 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)

TABLE 9 | Results of gait identification using indoor and outdoor recordings.

Environment	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)

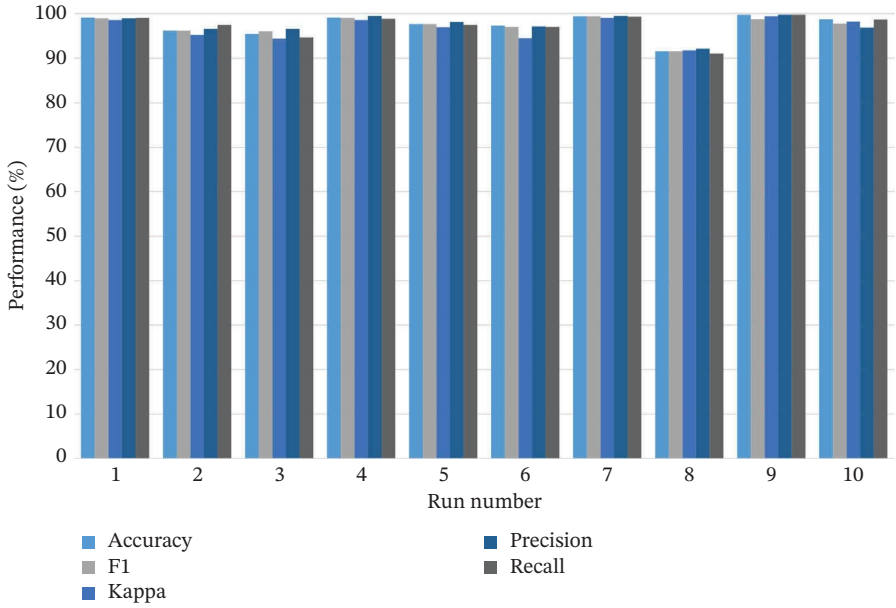

FIGURE 7 | Performance of the 10 models validated on the SOTON Large dataset [16].

dataset [23]) using the methodology described in Section 3. The results in Table S4 (supporting information) and Figure 7 include relevant metrics for each of the 10 models, in addition to the mean and standard deviation of each metric for each model trained for validation purposes (see Section 3), with the models achieving 97.68%, 97.29%, 96.71%, 97.56% and 97.38% in terms of average accuracy, F1, kappa, precision and recall, respectively.

Table 10 compares our best-performing STONI-GID model (out of the 10 samples trained on our dataset) with relevant results reported in the literature. Table 10 suggests that KNNs are a prevalent approach to gait identification on the SOTON Large dataset [16]. However, performance varies significantly across studies even when they use the same data. However, the relatively low performance reported by [57] may be due to the fact that this early work utilised only height and footprint to achieve this result. 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 [16], outperforming the next best-performing work [58], which achieved 96.70% accuracy on the same dataset.

4.6.2 | Validation on the GRIDDS Dataset

The GRIDDS dataset [20, 60] contains 35 subjects recorded from a single angle (sagittal). The dataset provides 350 instances and contains depth information and RGB video. Table S5 (of the supporting information) presents the statistical results achieved using our STONI-GID approach, which is trained using the methodology described in Section 3. The results presented in Table S5 (of the supporting information) and Figure 8 include the results of all 10 models and the mean and standard deviation of each metric from the models trained for validation purposes, with the models achieving 99.12%, 99.14%, 98.81%, 99.13% and 99.12% regarding average accuracy, F1, kappa, precision and recall, respectively. The improvement of the accuracy when using GRIDDS as compared to using our dataset (in Table S3 of the supporting information) or SOTON Large (in Table S4 of the supporting information) may be at least partly due to the lower number of subjects or the simplicity of experimental data

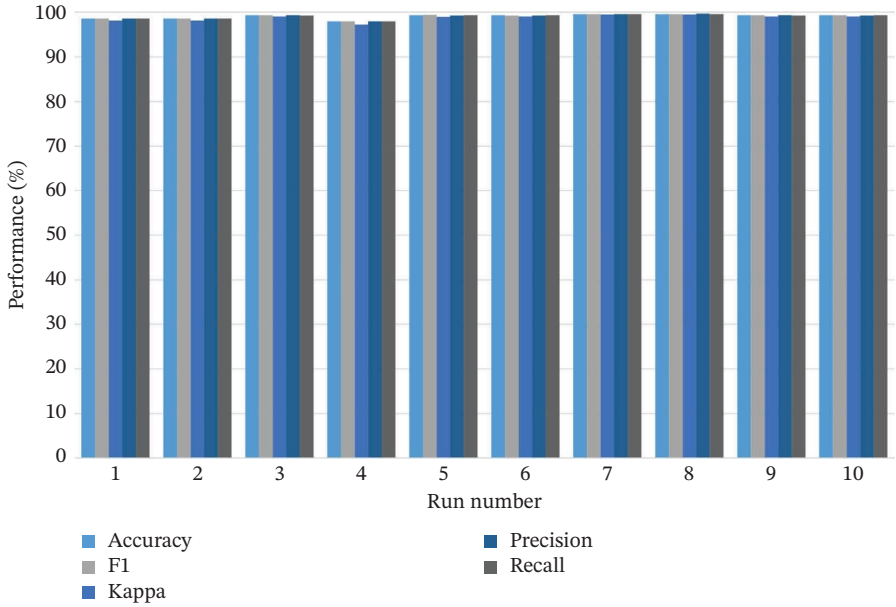

FIGURE 8 | Performance of the 10 models validated on the GRIDDS dataset [20].

TABLE 10 | Reported gait identification results validated on the SOTON Large dataset [16].

Ref.	Year	Model/algorithm	Accuracy (%)
[55]	2017	SVM	93.00
[56]	2016	KNN	95.70
[57]	2012	KNN	58.20
[58]	2011	KNN	96.70
[59]	2009	KNN	95.80
STONI-GID	**2024**	**LSTM**	**97.68**

Note: The results from the proposed method are shown in bold.

(captured from a single perspective). However, to the best of the authors' knowledge, no existing gait identification method has been validated on the GRIDDS dataset.

The results presented in Tables S4, 10, S5 are significant. Previous work, such as [59], has highlighted model generalisation as a challenge for robust person identification in real-world environments, leading to false-positive identifications. The results reported in Table S4 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 real-world occlusion. This work has endeavoured to simulate such occlusion. However, further validation of STONI-GID and related methods is necessary using real-world examples.

5 | Discussion

As reported in Table 4, 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 features such as stride width, which is not easily extracted from other viewing angles when using 2D HPE. Therefore, frontal/rear angles may be beneficial for gait identification in applications where the viewing angle can be controlled, such as in more controlled environments, like airports. Further research is required to identify the cause of the increased performance; however, to the best of the authors' knowledge, only our dataset [23] provides the relevant viewing angles. Furthermore, future research may aim to obtain such benefits, such as extracting stride width, without adjusting the viewing angle; 3D HPE may offer an opportunity to achieve this.

Recent works, such as Refs. [35, 61, 62], have made significant progress in mitigating the impact of occlusion on gait identification. However, these works address only partial occlusion, not full-body occlusion, which is arguably a more challenging problem. Unlike these works, STONI-GID has minimised the effect of full-body occlusion. Moreover, STONI-GID would be equally applicable to partial occlusion; future work will compare its performance under partial occlusion with that of existing works. Moreover, Ref. [61] only compensates for upper-body occlusion. However, the lower body provides significant gait information, which alone is likely sufficient for identification. Unlike Ref. [62], STONI-GID does not require additional machine learning models to address occlusion, thereby potentially reducing computational expense.

Similarly, many recent works, such as [50, 61, 62], continue to employ silhouette-based approaches for gait identification primarily. However, such approaches cannot be considered methods for pure gait identification, as they incorporate many appearance-based cues [10]. In contrast, STONI-GID performs gait identification without relying on appearance-based clues, and the statistical analysis suggests that it is not affected by changes in appearance.

Despite recent advances, fully automated gait-based person identification remains challenging for direct practical deployment, particularly in forensic retrieval scenarios where high accuracy is required. In operational settings, person re-identification methods

are more commonly adopted for large-scale retrieval due to their superior performance under appearance-rich conditions [63, 64]. In contrast, gait analysis in forensic science is typically employed in an observation-based or silhouette-based comparison framework, where gait serves as supportive or corroborative evidence rather than a standalone identification modality [65–68]. Within this context, the proposed joint angle-based approach is better positioned as a complementary tool for controlled comparison or verification, rather than as a fully automated identification system. Bridging the gap between research prototypes and forensic practice remains an open challenge, requiring improved robustness, transparency and validation protocols.

Although our gait dataset improves on numerous aspects of existing datasets, it could be further enhanced by adding additional participants. Moreover, including additional environments would be beneficial, particularly by including examples of real-world occlusion, thereby eliminating the need to simulate occlusion as in this work. To the best of the authors' knowledge, no existing dataset provides labelled real-world examples of occluded gait. Furthermore, a larger dataset could provide additional insights; for example, clustering may reveal gait types or categories, which could improve gait identification and 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.

5.1 | Limitations

Currently, the proposed STONI-GID approach does not account for the possibility that the detected gait patterns may belong to someone not included in the training dataset. Therefore, the model would attempt to assign a participant identifier when none is appropriate. In public applications, this would lead to false positive identifications. Thus, future work is required to prevent STONI-GID from predicting the identity of unknown people.

Moreover, the methodology described in Figure 3 assumes that during training, each participant has been recorded from each of the viewing angles depicted in Figure 2. However, this assumption may not hold for applications such as identifying missing persons. Therefore, future work is needed to address the assumption of multiple viewing angles for model training. Similarly, as with the alternative methods described in the literature, the camera height is not considered. This is an additional limitation of the existing datasets. Therefore, future data should be collected to test the effect of camera height, for example, by simulating CCTV cameras mounted on a tall building.

A related limitation is that currently, STONI-GID does not support dynamic switching between the sagittal, oblique and frontal/rear identification models within a single video sequence. In practical surveillance scenarios, a subject's walking direction relative to the camera may change; the method assumes that each sequence contains continuous walking in a single direction, and sequences are analysed separately according to their viewing angle. Angle classification is performed for each gait cycle, partially mitigating the effect of small deviations in walking direction within a sequence.

In addition, while the dataset offers valuable diversity in participants and conditions, it has several limitations. First, the appearance changes are relatively modest, typically limited to the addition or removal of outerwear, which may not fully reflect more substantial or deliberate attempts at disguise. Second, the recording environments are static, with minimal background motion or environmental complexity, which limits their applicability to dynamic real-world scenarios. Finally, although occlusion is simulated by nulling keypoints, the dataset does not include real instances of occlusion caused by objects or people due to ethical and practical constraints in uncontrolled settings.

Another limitation of the proposed STONI-GID framework is the relatively coarse discretisation of viewing angles into three groups. Recent studies have shown that even subtle differences in viewing direction can lead to noticeable drops in gait verification performance, particularly in cross-view scenarios [69, 70]. While the adopted angle grouping is sufficient to demonstrate the feasibility of the proposed joint-angle–based identification framework, it does not capture the fine-grained viewpoint variations present in more recent large-scale gait datasets. Extending the framework to support denser, more continuous representations of viewing angles is therefore an important direction for future work.

Finally, the STONI-GID framework does not currently make claims regarding real-time performance. The method has been developed and validated primarily for offline analysis of gait sequences, and no specific measures have been implemented to ensure real-time operation. While this does not affect the validity of the results presented here, it may limit immediate applicability in live surveillance or other scenarios requiring low-latency processing. Future work could explore algorithmic optimisations or hardware acceleration to enable real-time deployment of the framework.

5.2 | Future Directions

Several aspects of this research suggest potential avenues for future work. First, to address the aforementioned issue of handling samples from individuals not present in the training set, a threshold may be applied. For example, if the probability output is below a certain threshold, the output may be labelled as unknown. Alternatively, STONI-GID could be adapted to quantify uncertainty in its predictions. Again, a threshold could be applied, returning 'unknown' when the prediction's uncertainty exceeds a defined threshold. Moreover, uncertainty quantification presents additional opportunities in person identification. For example, when searching for a missing person or a crime suspect, high-confidence identifications may trigger an automated response, whereas medium levels of uncertainty may be forwarded to human professionals for confirmation.

Similarly, deep learning is commonly used in person identification due to its superior performance. However, deep learning methods may be considered 'black boxes' because it is difficult to understand how such models arrive at a decision. Therefore, future work should explore explainable AI (XAI) approaches which can provide human-understandable justifications for their decisions. Such approaches may promote the trust and ethical use of person identification, which is particularly important given the sensitive nature of crime suspect identification.

Another potential future work is the normalisation of viewing angles. For example, generative AI can be used to map an image's viewing angle to a predefined angle. Such an addition would eliminate the need for the angle-detection model and the multiple detection models in STONI-GID, simplifying the solution. Moreover, this would address the aforementioned issue of requiring training data to include multiple angles for each participant.

Similarly, identifying which specific joints contribute most to gait identification accuracy is an important avenue for future work. Such an analysis could guide the state estimation algorithm to prioritise the protection or more sophisticated modelling of these keypoints. A similar study in our previous work by Topham et al. [7] demonstrates how the most informative features can be determined, providing a foundation for this approach in the STONI-GID framework.

To address the limitations of the present study, future work could focus on expanding the dataset. While the dataset used includes real-world environments, diverse participants, varied appearances and multiple viewing angles [23], certain constraints remain. Notably, the recording environments are relatively static, with limited background dynamics such as moving crowds. Furthermore, the dataset contains minimal instances of real occlusion. Capturing such data in uncontrolled environments poses ethical and practical challenges, particularly regarding recording individuals without consent and accurately annotating occlusions in real-world settings.

Moreover, although the dataset is among the few that incorporate variations in participant appearance, these changes are relatively minor, for example, the addition or removal of a coat. Future efforts should aim to capture more substantial appearance changes, including full outfit alterations. It would also be beneficial to include more challenging clothing types, such as long coats, robes, dresses and accessories, which may significantly impact pose estimation performance.

A related avenue for improvement involves expanding the participant pool, as in Zhao et al. [71], to better match the scale of large, widely used gait datasets, such as CASIA-B [22] and OU-ISIR [72]. While our dataset exhibits valuable diversity across gender, age, ethnicity, height and weight, scaling up the number of subjects would enhance generalisability and facilitate more comprehensive model evaluation. Importantly, any expansion should preserve this demographic diversity to ensure continued relevance to real-world applications. Additionally, future experiments should evaluate the effectiveness of the proposed STONI-GID methodology on benchmark datasets such as CASIA-B and OU-ISIR to validate its performance across larger and more varied subject populations.

Another important direction for future work is to evaluate the computational efficiency and runtime performance of the proposed STONI-GID method. While this study focussed on robustness to appearance changes, environmental variation and viewpoint differences, real-time performance is critical for deployment in practical surveillance settings. Future experiments should therefore measure inference speed, model size and resource requirements across various hardware platforms, including edge devices commonly used in surveillance scenarios. Such analysis would provide a more complete understanding of the trade-offs between accuracy and efficiency in real-world applications.

6 | Conclusion

To the best of the authors' knowledge, this is the first work to simultaneously address multiple unresolved challenges in gait identification, particularly occluded gait, realistic environments and appearance changes. We introduce a novel methodology as an initial step towards non-invasive, non-restraining gait identification for real-world applications. Unlike many recent works, such as Refs. [73, 74], which still rely on the silhouette-based approach, our method extracts body joint angles using purely automated computer vision methods, which are fed to a composite of deep learning (LSTM) and state estimation (Kalman filter) to handle the dynamics of real-time gait identification. Specifically, we perform state-estimation filtering to restore performance typically lost due to occlusion. While Kalman filters have previously been applied to estimate gait-related features in both humans and robots [75, 76], to the best of our knowledge, this work is the first to employ them for person identification under occluded gait conditions. In particular, unlike other works which focus on partial occlusion, this is the first to apply state estimation to the problem of temporary full-body occlusion.

Moreover, we report high identification accuracy for our dataset [23, 27], as well as for the SOTON Large [16] and GRIDDS [20] datasets, with 98.66%, 95.48% and 99.12%, respectively. Moreover, the proposed method offers comparable performance to the only known example of facial identification validation results reported for these datasets, which is the 96.7% reported in Reid and Nixon [77]. Additionally, we show that, unlike some existing works, our approach is not significantly affected by the appearance of subjects, the recording environment, or the viewing angle. These findings were enabled by using our novel dataset [23], which provides a unique combination of real-world factors.

This work has the potential to improve non-invasive person identification in unconstrained real-world environments. Furthermore, it can affect domains such as security, where existing applications have been shown to produce higher-than-acceptable false-positive rates [1]. Despite advances reported in the literature, to the best of the authors' knowledge, this work is the first to address the real-world problems of viewing angle, appearance changes and occlusion within a single method.

Funding

No funding was received to assist with the preparation of this manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are openly available in LJMU OpenData at https://opendata.ljmu.ac.uk/id/eprint/133/.

References

1. C. Fox, "BBC News Face Recognition Police Tools 'staggeringly Inaccurate," (2018), https://www.bbc.co.uk/news/technology-44089161.

2. G. Guo and N. Zhang, "A Survey on Deep Learning Based Face Recognition," *Computer Vision and Image Understanding* 189 (2019): 102805, https://doi.org/10.1016/j.cviu.2019.102805.

3. J. P. Singh, S. Jain, S. Arora, and U. P. Singh, "Vision-Based Gait Recognition: A Survey," *IEEE Access* 6 (2018): 70497–70527, https://doi.org/10.1109/access.2018.2879896.

4. K. Shiraga, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi, "Geinet: View-Invariant Gait Recognition Using a Convolutional Neural Network," in *2016 International Conference on Biometrics (ICB)* (Halmstad, Sweden: IEEE, 2016), 1–8.

5. T. Teepe, A. Khan, J. Gilg, F. Herzog, S. Hörmann, and G. Rigoll, "Gaitgraph: Graph Convolutional Network for Skeleton-Based Gait Recognition," in *2021 IEEE International Conference on Image Processing (ICIP)* (Anchorage: IEEE, 2021), 2314–2318.

6. L. Topham, W. Khan, D. Al-Jumeily, A. Waraich, and A. Hussain, "Gait Identification Using Hip Joint Movement and Deep Machine Learning," in *International Conference on Intelligent Computing* (Xi'an, China: Springer, 2022), 220–233.

7. L. K. Topham, W. Khan, D. Al-Jumeily, A. Waraich, and A. J. Hussain, "Gait Identification Using Limb Joint Movement and Deep Machine Learning," *IEEE Access* 10 (2022): 100113–100127, https://doi.org/10.1109/access.2022.3207836.

8. J. E. Cutting and L. T. Kozlowski, "Recognizing Friends by Their Walk: Gait Perception Without Familiarity Cues," *Bulletin of the Psychonomic Society* 9, no. 5 (1977): 353–356, https://doi.org/10.3758/bf03337021.

9. M. R. Popovic, T. Keller, S. Ibrahim, G. v. Bueren, and M. Morari, "Gait Identification and Recognition Sensor," in *Proceedings of 6th Vienna International Workshop on Functional Electrostiumlation* (Vienna, Austria, 1998).

10. L. K. Topham, W. Khan, D. Al-Jumeily, and A. Hussain, "Human Body Pose Estimation for Gait Identification: AComprehensive Survey of Datasets and Models," *ACM Computing Surveys* 55, no. 6 (2022): 1–42, https://doi.org/10.1145/3533384.

11. N. Takemura, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi, "Multi-View Large Population Gait Dataset and Its Performance Evaluation for Cross-View Gait Recognition," *IPSJ transactions on Computer Vision and Applications* 10 (2018): 1–14, https://doi.org/10.1186/s41074-018-0039-6.

12. M. Hirzer, C. Beleznai, P. M. Roth, and H. Bischof, "Person Re-Identification by Descriptive and Discriminative Classification," in *Image Analysis: 17Th Scandinavian Conference, SCIA 2011, Ystad, Sweden, May 2011. Proceedings 17* (Berlin, Germany: Springer, 2011), 91–102.

13. L. Zheng, Z. Bie, Y. Sun, et al., "Mars: A Video Benchmark for Large-Scale Person Re-Identification," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, the Netherlands, October 11–14, 2016, Proceedings, Part VI 14* (Berlin, Germany: Springer, 2016), 868–884.

14. L. Wei, Y. Tian, Y. Wang, and T. Huang, "Swiss-System Based Cascade Ranking for Gait-Based Person Re-Identification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 29, no. 1 (Austin, 2015), https://doi.org/10.1609/aaai.v29i1.9454.

15. 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," *Journal of Visual Communication and Image Representation* 25, no. 1 (2014): 195–206, https://doi.org/10.1016/j.jvcir.2013.02.006.

16. J. D. Shutler, M. G. Grant, M. S. Nixon, and J. N. Carter, "On a Large Sequence-Based Human Gait Database," in *Applications and Science in Soft Computing* (Berlin, Germany: Springer, 2004), 339–346.

17. W. Sheng and X. Li, "Multi-Task Learning for Gait-Based Identity Recognition and Emotion Recognition Using Attention Enhanced Temporal Graph Convolutional Network," *Pattern Recognition* 114 (2021): 107868, https://doi.org/10.1016/j.patcog.2021.107868.

18. T. T. Verlekar, L. D. Soares, and P. L. Correia, "Gait Recognition in the Wild Using Shadow Silhouettes," *Image and Vision Computing* 76 (2018): 1–13, https://doi.org/10.1016/j.imavis.2018.05.002.

19. A. Chauhan, *Kaggle Gait Silhouette Dataset* (2020).

20. J. F. Nunes, P. M. Moreira, and J. M. R. Tavares, "Gridds-A Gait Recognition Image and Depth Dataset," in *ECCOMAS Thematic Conference on Computational Vision and Medical Image Processing* (Berlin, Germany: Springer, 2019), 343–352.

21. L. Wang, T. Tan, H. Ning, and W. Hu, "Silhouette Analysis-Based Gait Recognition for Human Identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25, no. 12 (2003): 1505–1518, https://doi.org/10.1109/tpami.2003.1251144.

22. S. Yu, D. Tan, and T. Tan, "A Framework for Evaluating the Effect of View Angle, Clothing and Carrying Condition on Gait Recognition," *18th international conference on pattern recognition (ICPR'06)* 4 (2006): 441–444.

23. L. K. 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," *Scientific Data* 10, no. 1 (2023): 320, https://doi.org/10.1038/s41597-023-02161-8.

24. A. Das, A. Chakraborty, and A. K. Roy-Chowdhury, "Consistent Re-Identification in a Camera Network," in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part II 13* (Berlin, Germany: Springer, 2014), 330–345.

25. H. Iwama, M. Okumura, Y. Makihara, and Y. Yagi, "The Ou-Isir Gait Database Comprising the Large Population Dataset and Performance Evaluation of Gait Recognition," *IEEE Transactions on Information Forensics and Security* 7, no. 5 (2012): 1511–1521, https://doi.org/10.1109/tifs.2012.2204253.

26. F. Pala, R. Satta, G. Fumera, and F. Roli, "Multimodal Person Reidentification Using RGB-D Cameras," *IEEE Transactions on Circuits and Systems for Video Technology* 26, no. 4 (2015): 788–799, https://doi.org/10.1109/tcsvt.2015.2424056.

27. 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," (2022), https://opendata.ljmu.ac.uk/id/eprint/133/.

28. M. Gruosso, N. Capece, and U. Erra, "Human Segmentation in Surveillance Video With Deep Learning," *Multimedia Tools and Applications* 80, no. 1 (2021): 1175–1199, https://doi.org/10.1007/s11042-020-09425-0.

29. P. Nithyakani, A. Shanthini, and G. Ponsam, "Human Gait Recognition Using Deep Convolutional Neural Network," in *2019 3rd International Conference on Computing and Communications Technologies (ICCCT)* (Chennai, India: IEEE, 2019), 208–211.

30. J. Han and B. Bhanu, "Individual Recognition Using Gait Energy Image," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28, no. 2 (2005): 316–322, https://doi.org/10.1109/tpami.2006.38.

31. 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 Digital Image Computing: Techniques and Applications (DICTA)* (New York: IEEE, 2018), 1–7.

32. D. Mehta, O. Sotnychenko, F. Mueller, et al., "Single-Shot Multi-Person 3d Pose Estimation From Monocular RGB," in *2018 International Conference on 3D Vision (3DV)* (Verona, Italy: IEEE, 2018), 120–130.

33. J. Zhen, Q. Fang, J. Sun, et al., "Smap: Single-Shot Multi-Person Absolute 3d Pose Estimation," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XV 16* (Berlin, Germany: Springer, 2020), 550–566.

34. A. Gupta and V. B. Semwal, "Occluded Gait Reconstruction in Multi Person Gait Environment Using Different Numerical Methods," *Multimedia Tools and Applications* 81, no. 16 (2022): 23421–23448, https://doi.org/10.1007/s11042-022-12218-2.

35. C. Xu, Y. Makihara, X. Li, and Y. Yagi, "Occlusion-Aware Human Mesh Model-Based Gait Recognition," *IEEE Transactions on Information Forensics and Security* 18 (2023): 1309–1321, https://doi.org/10.1109/tifs.2023.3236181.

36. Z. Zhang, L. Tran, X. Yin, et al., "Gait Recognition via Disentangled Representation Learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Long Beach, 2019), 4710–4719.

37. F. Locatello, S. Bauer, M. Lucic, et al., "Challenging Common Assumptions in the Unsupervised Learning of Disentangled Representations," in *International Conference on Machine Learning* (Long Beach: PMLR, 2019), 4114–4124.

38. D. Ye, C. Fan, J. Ma, X. Liu, and S. Yu, "Biggait: Learning Gait Representation You Want by Large Vision Models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Seattle, 2024), 200–210.

39. K. Ma, Y. Fu, C. Cao, S. Hou, Y. Huang, and D. Zheng, "Learning Visual Prompt for Gait Recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Seattle, 2024), 593–603.

40. S. Jiang, J. Jiang, S. Wang, et al., "RF-Gait: Gait-Based Person Identification With Cots RFID," *Wireless Communications and Mobile Computing* 2022, no. 1 (2022): 3638436, https://doi.org/10.1155/2022/3638436.

41. Z. Yang, Z. Zhen, H. Xu, Y. Zhang, and X. Feng, "RF-UI: Continuous User Identification Through Gaits Using Rfid," *IEEE Transactions on Cognitive Communications and Networking* (2024).

42. J. Li, B. Li, L. Wang, and W. Liu, "Passive Multiuser Gait Identification Through Micro-Doppler Calibration Using Mmwave Radar," *IEEE Internet of Things Journal* 11, no. 4 (2023): 6868–6877, https://doi.org/10.1109/jiot.2023.3312668.

43. J.-I. Cairó, J. Bonache, F. Paredes, and F. Martín, "Interference Sources in Congested Environments and Its Effects in UHF-RFID Systems: A Review," *IEEE Journal of Radio Frequency Identification* 2, no. 1 (2018): 1–8, https://doi.org/10.1109/jrfid.2018.2806738.

44. W. An, S. Yu, Y. Makihara, et al., "Performance Evaluation of Model-Based Gait on Multi-View Very Large Population Database With Pose Sequences," *IEEE transactions on biometrics, behavior, and identity science* 2, no. 4 (2020): 421–430, https://doi.org/10.1109/tbiom.2020.3008862.

45. Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime Multi-Person 2d Pose Estimation Using Part Affinity Fields," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Honolulu, 2017), 7291–7299.

46. R. E. Kalman, *A New Approach to Linear Filtering and Prediction Problems* (IEEE, 1960).

47. H. Neill, *Trigonometry: A Complete Introduction* (Teach Yourself, 2013).

48. S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation* 9, no. 8 (1997): 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735.

49. J. Durbin and S. J. Koopman, *Time Series Analysis by State Space Methods* (OUP Oxford, 2012).

50. Y. Peng, C. Cao, and Z. He, "Occluded Gait Recognition," in *2023 International Joint Conference on Neural Networks (IJCNN)* (Gold Coast, Australia: IEEE, 2023), 1–8.

51. J. Wen, Y. Shen, and J. Yang, "Multi-View Gait Recognition Based on Generative Adversarial Network," *Neural Processing Letters* 54, no. 3 (2022): 1855–1877, https://doi.org/10.1007/s11063-021-10709-1.

52. P. Pala, L. Seidenari, S. Berretti, and A. Del Bimbo, "Enhanced Skeleton and Face 3d Data for Person Re-Identification From Depth Cameras," *Computers & Graphics* 79 (2019): 69–80, https://doi.org/10.1016/j.cag.2019.01.003.

53. R. Chellappa, A. K. Roy-Chowdhury, and S. K. Zhou, "Human Identification Using Gait and Face," in *Circuits, Signals, and Speech and Image Processing* (Boca Raton: CRC Press, 2018), 27–1.

54. 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 *2022 IEEE/SICE International Symposium on System Integration (SII)* (New York: IEEE, 2022), 602–607.

55. 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," *Applied Soft Computing* 61 (2017): 42–57, https://doi.org/10.1016/j.asoc.2017.07.041.

56. 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," *Multimedia Tools and Applications* 75, no. 2 (2016): 1201–1221, https://doi.org/10.1007/s11042-014-2364-9.

57. G. Ariyanto and M. S. Nixon, "Marionette Mass-Spring Model for 3d Gait Biometrics," in *2012 5th IAPR International Conference on Biometrics (ICB)* (New Delhi, India: IEEE, 2012), 354–359.

58. J.-H. Yoo and M. S. Nixon, "Automated Markerless Analysis of Human Gait Motion for Recognition and Classification," *ETRI Journal* 33, no. 2 (2011): 259–266, https://doi.org/10.4218/etrij.11.1510.0068.

59. M. Goffredo, I. Bouchrika, J. N. Carter, and M. S. Nixon, "Self-Calibrating View-Invariant Gait Biometrics," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 40, no. 4 (2009): 997–1008, https://doi.org/10.1109/tsmcb.2009.2031091.

60. J. F. Nunes, P. M. Moreira, and J. M. R. Tavares, "Benchmark RGB-D Gait Datasets: A Systematic Review," in *VipIMAGE 2019: Proceedings of the VII ECCOMAS Thematic Conference on Computational Vision and Medical Image Processing, October 16–18, 2019* (Porto, Portugal: Springer, 2019), 366–372.

61. 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* (Paris, France, 2023), 3199–3209.

62. A. Gupta and R. Chellappa, "You can Run but Not Hide: Improving Gait Recognition With Intrinsic Occlusion Type Awareness," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (Waikoloa, 2024), 5893–5902.

63. Y. Zhai, X. Guo, Y. Lu, and H. Li, "In Defense of the Classification Loss for Person Re-Identification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (Long Beach, 2019), 0.

64. Y. Yuan, W. Chen, Y. Yang, and Z. Wang, "In Defense of the Triplet Loss Again: Learning Robust Person Re-Identification With Fast Approximated Triplet Loss and Label Distillation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (Seattle, 2020), 354–355.

65. I. Bouchrika, M. Goffredo, J. Carter, and M. Nixon, "On Using Gait in Forensic Biometrics," *Journal of Forensic Sciences* 56, no. 4 (2011): 882–889, https://doi.org/10.1111/j.1556-4029.2011.01793.x.

66. I. Birch, M. Birch, L. Rutler, et al., "The Repeatability and Reproducibility of the Sheffield Features of Gait Tool," *Science & Justice* 59, no. 5 (2019): 544–551, https://doi.org/10.1016/j.scijus.2019.04.001.

67. I. Birch, L. Raymond, A. Christou, M. A. Fernando, N. Harrison, and F. Paul, "The Identification of Individuals by Observational Gait Analysis Using Closed Circuit Television Footage," *Science & Justice* 53, no. 3 (2013): 339–342, https://doi.org/10.1016/j.scijus.2013.04.005.

68. D. Imoto, M. Hirabayashi, M. Honma, and K. Kurosawa, "Pre-Set Estimation-Based in-Silico Silhouette-Based Methodology for Improving the Robustness to Viewing Direction Difference for Assisting Forensic Gait Analysis," *Journal of Forensic Sciences* 68, no. 2 (2023): 470–487, https://doi.org/10.1111/1556-4029.15214.

69. D. Muramatsu, A. Shiraishi, Y. Makihara, M. Z. Uddin, and Y. Yagi, "Gait-Based Person Recognition Using Arbitrary View Transformation Model," *IEEE Transactions on Image Processing* 24, no. 1 (2014): 140–154, https://doi.org/10.1109/tip.2014.2371335.

70. D. Imoto, M. Hirabayashi, M. Honma, and K. Kurosawa, "Enhancing the Robustness of Forensic Gait Analysis Against Near-Distance Viewing

Direction Differences," *Multimedia Tools and Applications* 81, no. 18 (2022): 26199–26221, https://doi.org/10.1007/s11042-022-12751-0.

71. Y. Zhao, R. Liu, W. Xue, et al., "Effective Fusion Method on Silhouette and Pose for Gait Recognition," *IEEE Access* 11 (2023): 102623–102634, https://doi.org/10.1109/access.2023.3317437.

72. Y. Makihara, H. Mannami, A. Tsuji, et al., "The OU-ISIR Gait Database Comprising the Treadmill Dataset," *IPSJ Transactions on Computer Vision and Applications* 4 (2012): 53–62, https://doi.org/10.2197/ipsjtcva.4.53.

73. R. Wang, Y. Shi, H. Ling, et al., "Gait Recognition via Gait Period Set, IEEE Transactions on Biometrics, Behavior, and," *Identity Science* 5, no. 2 (2023): 183–195, https://doi.org/10.1109/tbiom.2023.3244206.

74. J. Chen, Z. Wang, P. Yi, K. Zeng, Z. He, and Q. Zou, "Gait Pyramid Attention Network: Toward Silhouette Semantic Relation Learning for Gait Recognition," *IEEE Transactions on Biometrics, Behavior, and Identity Science* 4, no. 4 (2022): 582–595, https://doi.org/10.1109/tbiom.2022.3213545.

75. S. L. Nogueira, S. Lambrecht, R. S. Inoue, et al., "Global Kalman Filter Approaches to Estimate Absolute Angles of Lower Limb Segments," *BioMedical Engineering Online* 16, no. 1 (2017): 58, https://doi.org/10.1186/s12938-017-0346-7.

76. X. Chen, C. Chen, Y. Wang, et al., "A Piecewise Monotonic Gait Phase Estimation Model for Controlling a Powered Transfemoral Prosthesis in Various Locomotion Modes," *IEEE Robotics and Automation Letters* 7, no. 4 (2022): 9549–9556, https://doi.org/10.1109/lra.2022.3191945.

77. D. A. Reid and M. S. Nixon, "Human Identification Using Facial Comparative Descriptions," in *2013 International Conference on Biometrics (ICB)* (Madrid, Spain: IEEE, 2013), 1–7.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. *(Supporting Information)*

The following supporting tables are available in the Supporting information document.

Section 1: LSTM model configurations.

Table S1. LSTM model configuration for angle classification.

Table S2. LSTM model configuration for gait-based person identification.

Table S3. Statistical results of the proposed angle detection model validated over a purely unseen partition of our test dataset.

Section 2: STONI-GID validation results.

Table S4. Results of the STONI-GID gait identification method validated on the SOTON Large dataset.

Table S5. Results of the STONI-GID gait identification method validated on the GRIDDS dataset.