# MULTIVIEW HUMAN GAIT ANALYSIS USING THE FIRST AND THIRD PERSON DATA

A DISSERTATION
SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF
MASTER OF TECHONOLOGY
IN

## SIGNAL PROCESSING AND DIGITAL DESIGN


Submitted by:
**NIKITA MALIK**
**2K19/SPD/13**
Under the supervision of
**Assistant Prof. SUDIPTA MAJUMDAR**



**DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING**
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042


JANUARY 2021

# <u>CANDIDATE'S DECLARATION</u>

I **Nikita Malik** student of MTECH (Signal Processing and Digital Design), hereby declare that the project Dissertation titled "**Multiview Human Gait Analysis using the First and Third Person Data**" which is submitted by me to the Department of Electronics and Communication Engineering, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Associateship, Fellowship or other similar title or recognition.

Place: Delhi

**Nikita Malik**

Date: 30/ 07 /2021

# DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

# CERTIFICATE

I hereby certify that the Project Dissertation titled "**Multiview Human Gait Analysis using the First and Third Person Data**" which is submitted by **NIKITA MALIK**, **2K19/SPD/13** of Electronics and Communication Department, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Master of Technology, is a record of the project work carried out by the student under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: Delhi **Asst Prof. SUDIPTA MUJAMDAR**

Date: 30/ 07 /2021 **SUPERVISOR**

To My Parents,

**Mrs. Savita Malik & Mr. Satish Kumar Malik**

And

All My Teachers

# ACKNOWLEDGEMENT

A successful project can never be prepared by the efforts of the person to whom the project is assigned, but it also demands the help and guardianship of people who helped in completion of the project.
With profound sense of gratitude, I thank **Asst. Prof. SUDIPTA MAJUMDAR**, my Research Supervisor, for her encouragement, support, patience, and her guidance in this research work. I could not have imagined having a better guide and advisor for my MTech study.

I am also grateful to **Prof. N.S. RAGHAVA**, HOD, Department of Electronics and Communication Engineering, DTU for providing the right academic resources & environment for this work to be carried out.

Last but not the least, my whole-heartedly thanks to all my family and friends who have patiently extended all sorts of help for accomplishing this undertaking.

**NIKITA MALIK**

# ABSTRACT

The gait of a person is often used as a biometric tool to identify or classify a person based on gender and age. Because of its ability to track a person from afar, gait recognition has found use in a variety of fields, including forensics, surveillance, and health monitoring departments. Biometric systems is a rapidly emerging area that necessitates the development of new methods to address problems that have plagued previous approaches. Human gait is a less-explored region in the field of biometrics. In this Project, two different types of gait datasets have been reported and presented. The FP (First Person) data carrying the camera motion collected from the movement of the volunteer's body and the TP (Third Person) data captured from a distant view were recorded at the same time. A total of 24 subjects (15 males and 9 females) are included in the dataset. The discussion is extended to include a comparison of the results obtained using TP and FP data.

This report also provides an extensive survey of the stages involved in the framework of gait recognition by analysing the different methods used in each stage along with the description of the feature extraction process and the state-of-the-art techniques used in appearance-based and human-pose-based methods. Moreover, a brief comparative description on the recent data reduction or feature selection methods has been provided. Furthermore, it will provide a first-hand knowledge about the public datasets that is motion capture databases and the datasets simply used for human gait recognition. In comparison to other biometric methods, gait recognition has a lot of potential for future work, Researchers working in the fields of biometrics, human pose estimation, monitoring, human gait recognition and analysis will benefit from the review given in the survey.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# CHAPTER 1

# INTRODUCTION

## 1.1 Human Gait

We have known biometric systems for a very long time, but it is still a growing field as some new and effective human identification methods are getting added to the list. Human gait recognition is one of those emerging methods in which lots of research has been done and still needs to be done. Human gait simply includes the study of human motion that uses brain and eye observations together with instruments that measures the body movements. The technology is useful for remote biometric identification without the involvement of observer. Because of its property of recognition from a distance, it is excelling in the field of biometrics. It is a technique that identifies a person by extracting features from the body movements.

A gait is a sequence of movements of the extremities produced during locomotion. The different ways in which a person can move, either naturally or because of advanced training, are human gaits. Human gait is characterized as bipedal, biphasic forward propulsion of the human body's centre of gravity, in which alternating sinuous motions with the least energy expenditure of different segments of the body occur.

In different forms, human gaits are graded. In general, each gait can be classified as either normal (one that humans use instinctively) or learned (a non-instinctive gait learned via training). Hand walking and advanced gaits used in martial arts provide examples of the latter.

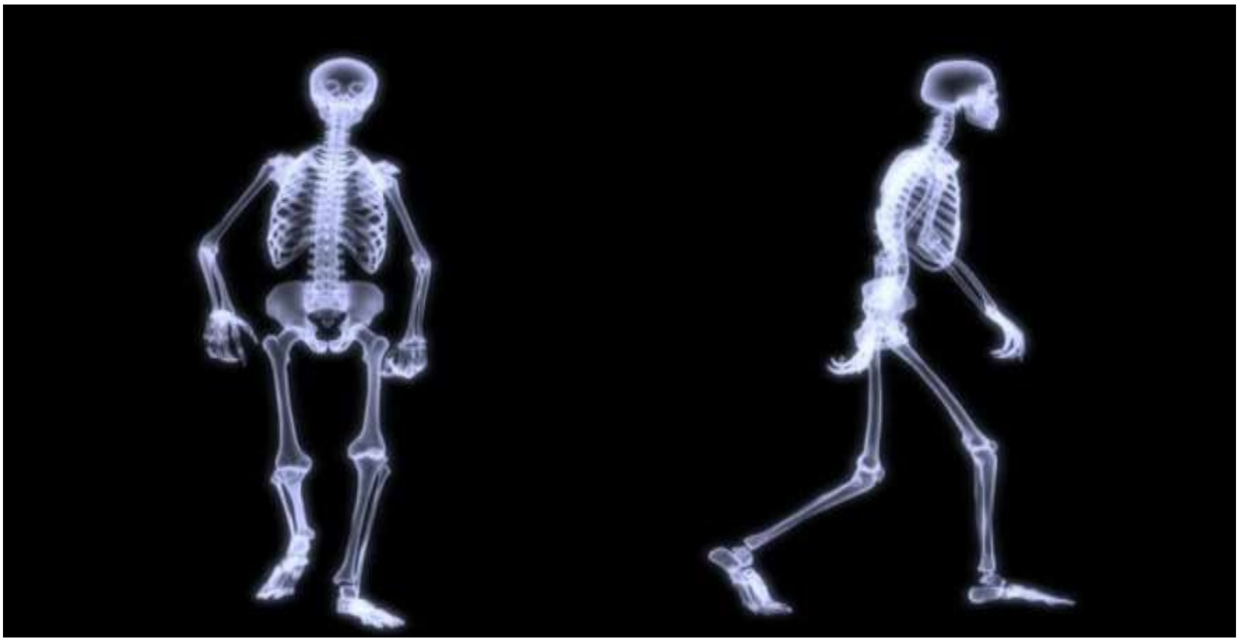


**Fig. 1 Human Skeleton Gait**

## 1.2 Applications of Human Gait Recognition

Gait recognition has many useful applications in departments like security, medical, forensics, sports, healthcare and so on, which made it a very popular choice among researchers. [1] Access control can be secured using gait recognition for particular protected facilities like on airports, it can be used for automatic security checkups. In border crossings, railway stations or any other travelling platform it can be used as an early authorization method and to organize the list of travelers into categories. In ATM security, the PIN system can be combined with gait characteristics of a person which will be a very helpful method to enhance the security [1]**.** Home security can be improved to very great extent using the gait recognition systems as a biometric method which may or may not allow certain people to enter into a building without a proper permit. With the advance in technologies gait identification can be a very useful solution to make peoples life easier. Gait is different for every other individual depending upon their age and gender, it shows variations according to the age of the person, some clothing variations and it even varies if the person is hurt or not able to walk with his/her natural gait. [2] Researchers are continuously evaluating the gait systems and its applications are few as of now, however there has been a significant increase in the number of uses of gait recognition systems in the surveillance departments which are mainly used for improving the security. Pattern recognition from the acquired CCTV footage is a growing choice for these departments. Such data is sufficient for identification of an individual and there has been extensive research done using the gait data as a forensic tool [3], [4], [5].

Gait information can be acquired with numerous gait acquisition devices such as wearable sensors data [6] (accelerometer, gyroscopes, pressure and force sensors), high-definition cameras or non-wearable sensors data. Gait Identification techniques are divided into two categories: 1) model and 2) appearance-based techniques. The parameters for a pre-defined model are incorporated in model-based approaches, while sculpted gait features are extracted from images or videos in appearance-based approaches. Gait Recognition is still a very challenging task. Many factors, such as wearing different clothes or shoes, having an injury, or carrying an item, can affect a person's natural walking style and appearance. Different walking rates, viewing angles, and environmental influences all contribute to the difficulty.

## 1.3 Objective

Two different types of gait datasets have been reported and presented. The first person (FP) data carrying the camera motion collected from the movement of the volunteer's body and the third person (TP) data captured from a distant view were recorded at the same time. A total of 24 subjects (15 males and 9 females) are included in the dataset. The discussion is extended to include a comparison of the results obtained using TP and FP data.

## 1.4 Organization of the Report

In this report, multiview gait analysis is done on a first person recorded data to add more to the research area. Chapter 1 introduces human gait and its applications; Chapter 2 describes some state-of-the-art surveys used in the literature review; Chapter 3 provides a detailed analysis of the gait recognition system framework; Chapter 4 covers the proposed work section, in which some new methods for Multiview human gait analysis are introduced; The dataset and description analysis are covered in Chapter 5, the results and simulation achieved using the first person dataset are covered in Chapter 6, and the dissertation is summarized in Chapter 7 by concluding and suggesting a future scope for the project.

# CHAPTER 2

# LITERATURE REVIEW

Some earlier state-of-the-art surveys are available for gait recognition which provides a detailed overview regarding the progress made so far in the area of human gait recognition systems. The evaluation part has been a history and the area of interest for many researchers is the application part in which not much evaluation has been done till now. [7] Some surveys contain a detailed overview about the current available vision-based methods available for gait recognition**.** In their work they performed a survey-based analysis on gait recognition in terms of the methods used for classification, feature extraction approaches and the public datasets available for identification. Some aspects were included in the survey which was particularly not available in other gait surveys like accelerometer-based identification, environmental issues affecting the gait, floor-sensors and so on. [8] Some influencers reviewed many of the revolutionary methods available for gait recognition in deep learning. They have categorized the gait data acquisition devices into following categories: video data acquired from camera, wearable sensor data, floor sensor data and also the public gait datasets available for gait analysis. They compared and evaluated different parameters acquired by different categorized devices using available neural network architectures. Also, in [9] they focused on reviewing the earlier methods available for human gait identification. They have done an extensive research on different datasets and performed a comparative analysis on different techniques and algorithms available to estimate parameters at each stage of the gait recognition framework.

### 2.1 Normalized cross-correlation

Stereo vision, motion tracking, image mosaicing, and other computer vision activities have all benefited from normalised cross-correlation. The simplest yet most powerful approach for determining similarity is normalised cross-correlation, which is insensitive to linear brightness and contrast variations. It is suitable for real-time applications due to its simple hardware implementation [10]**.** The normalised cross correlation (NCC) metric has long been used to assess the degree of similarity (or dissimilarity) between two images. The NCC has the advantage of being less susceptible to linear changes in the intensity of illumination in the two compared images than the cross correlation [11]**.** Furthermore, the NCC is limited to a range of 1 to 1. Cross correlation is much more difficult to set a detection threshold than it is to set a detection threshold. The reference and contrast signals are denoted by the letters f(i) and $f'$(i) respectively, where i is the sample index (1≤ i ≤ S, S = total number of samples). The reference and comparison windows' normalized cross-correlation, $R_{NCC}$, is defined as:

$$R_{NCC}(u, \mu) = \frac{\sum_{i=u}^{u+W-1} f(i)\mathrm{f}'(i+\Gamma)}{\sqrt{\sum_{i=u}^{u+W-1} f^2(i) \sum_{i=u}^{u+W-1} \mathrm{f}'^2(i+\Gamma)}} \text{ , } (\Gamma_1 \leq \Gamma \leq \Gamma_2),$$

where [u, u + W − 1] is the interval of location of reference window, u represents the origin, W= size of window, $\Gamma$ is depicting the shift between the windows, and [$\Gamma_1$, $\Gamma_2$] = range of search of physiologic displacements. The normalized cross-correlation estimation includes three terms, i.e., reference window's energy ($\sum_{i=u}^{u+W-1} f^2(i)$ )in the denominator, ($\sum_{i=u}^{u+W-1} \mathrm{f}'^2(i+\Gamma)$) =comparison window's energy in the denominator, and the standard (i.e., non-normalized) cross-correlation among the 2 windows ($\sum_{i=u}^{u+W-1} f(i)\mathrm{f}'(i+\Gamma)$) in the numerator.

Instead of conducting the normalized cross-correlation directly in this paper, the average horizontal and vertical projection vectors were determined first, and then the vectors were normalized to estimate the cross-correlation further. The evaluations, as well as the perfect shifts calculations and plot, are shown in the proposed work.

# CHAPTER 3

# FRAMEWORK OF GAIT RECOGNITION

The gait identification system comprises of 5 stages which are: Gait data acquisition, pre-processing of the acquired data, feature extraction, data reduction or feature selection, and classification and recognition as demonstrated in fig.2.

### 3.1 Gait Data Acquisition

Many data acquisition devices are used to collect gait data. It is the initial stage of a gait recognition system, the efficiency and accuracy of a gait recognition system depends upon this initial step of acquiring image frames data of the framework. Data acquisition can be done by using sensor-based, video-based or radar-based data acquisition techniques. These are categorized further into subcategories in fig.3.

### 3.2 Pre-processing of the acquired Gait Data

In this step noise removal and background modelling of the acquired data is done using filters (Gaussian filter, mean filters etc.) and by using some background/foreground subtraction methods (Frame difference, GMM etc.) respectively.

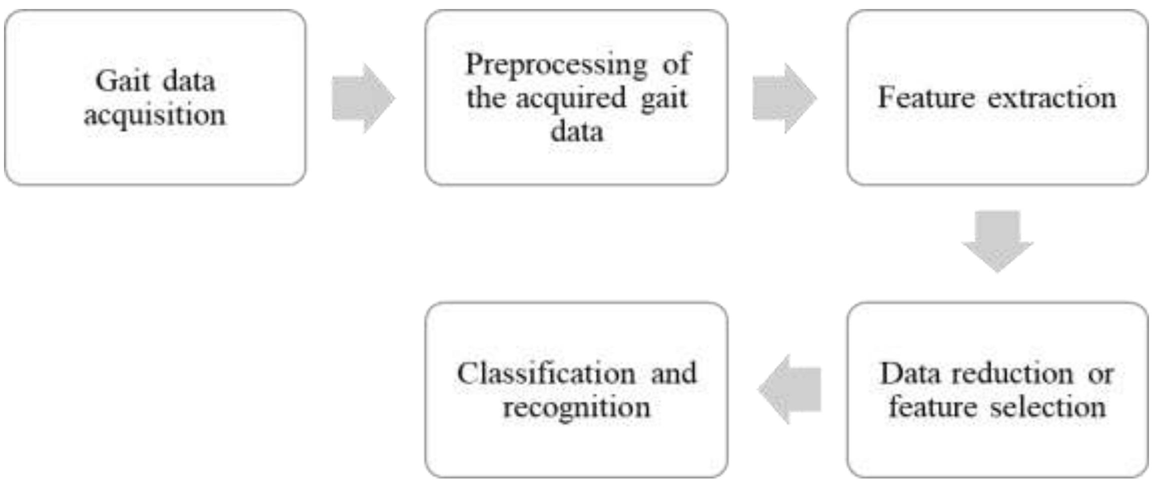



**Fig. 2 Framework of Gait Recognition**

### 3.3 Feature Extraction

After the pre-processing of acquired data, the next stage requires to extract some features from the acquired gait data which is segmented using the background subtraction methods and is noise-free. Features can be extracted with the help of various available methods. We have discussed here some appearance-based and human-pose based methods.

#### 3.3.1 Appearance-based approach

In past few years, traditional appearance-based gait recognition methods have been used extensively in many researches contributed in the area of computer-vision and pattern-recognition for the gait recognition or identification of a person. The method includes extracting 2D silhouettes from the key frames present across gait sequences and then matching them. There are many advantages of the appearance-based methods, some of them include ease in implementation, high threshold to noisy video and they seem to work well in a variety of walking gaits. The appearance-based approach considers several methods like spatiotemporal methods based on motion, physical parameter and statistical method, for extracting useful information for gait recognition. In video sequences, the spatiotemporal approach handles both space and time details. A low computational complexity is the main advantage achieved while using the spatiotemporal based methods. They are further divided into two subcategories that is, temporal template approach and spatiotemporal approach. In the template approach a temporal spatial comparison is done, on frames sequentially, this is achieved by directly performing a comparative analysis between probe input images and the gallery sequences.

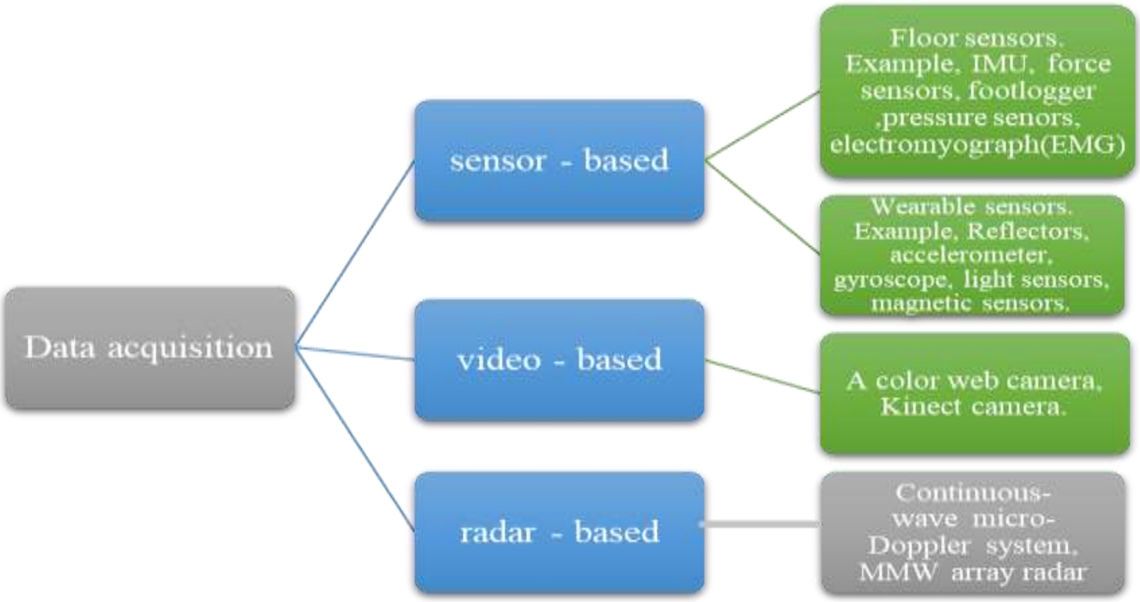



**Fig. 3 Data Acquisition Methods and their types**

In recent years many researchers have used this feature extraction method for recognition like Sarkar et al. [12] proposed an algorithm which successfully used background subtraction to extract silhouettes afterwards recognition is done by analyzing silhouettes through temporal correlation. R.T. Collins et al. [13] proposed another appearance-based template matching technique of body silhouette in which cyclic gait sequences were performed to extract key frames. Similarly, in [14] a temporal template matching hidden markov models (HMMs) is demonstrated for recognition of individuals from their gait. By following the same pattern, Kale et al. [15] proposed an approach which involves recognizing gait using HMM model. In the proposed method, hidden markov model estimates the likelihood corresponding to an individual by analyzing the temporal data.

Over the past years, studies have shown that the temporal template-based approaches came with some challenging issues as it involves analyzing images on a frame-by-frame basis. Whereas, the other spatiotemporal or patio-temporal motion summary techniques typically, superpose sequences of binary silhouettes according to certain rules. Some approaches use accumulated motion features which increase the effectiveness. In the spatiotemporal approach, the appearance and some dynamic characteristic features of gait are stored. These methods allow us a low computational complexity-based implementation and leaves us with a reduced feature vector. Several issues, like variations in camera orientation and changes in appearance were observed over the time while using these approaches as well due to walking speed, variations in clothing styles and some baggage carrying conditions. László Havasi et al. [16] focused on a new information symmetric extraction and temporal tracking algorithm used for classification of human movement patterns that work on cluttered video image sequences provided a spatio-temporal information as an input. Also, in [17] silhouettes of people transformed over time into low-dimensional feature vectors, which consists of average pixel distances from the silhouette's centre and is demonstrated using a novel feature extraction method that is evaluated on seven probe conditions.

Appearance-based gait examining methods require some pre-processing techniques which affects the performance of the whole recognition framework by increasing the time complexity of the system and also by causing some imperfect background-foreground subtraction algorithms. Some clothing and baggage carrying variations also severely affects the appearance-based approaches. To handle such kind of issues Kusakunniran et al. [18] in his work proposed an approach in which they directly extracted the gait features from raw videos. In the mentioned approach, concatenation of "histogram of oriented gradients" (HOG) and a "histogram of optical flow" (HOF) was used as STIP descriptor by detecting "space-time interest points" (STIP) in the spatiotemporal domain. R.T. Collins et al. in [13] in his survey mentioned another drawback of the appearance-based methods that it is view dependent. Since the approach is based on coordinating 2D silhouettes, test subjects viewed from substantially different angles than training subjects cannot be identified.

Some enhanced spatiotemporal approaches were discussed later on keeping in mind the raised issues in appearance-based methods. Enhanced spatiotemporal approaches contains several ground-breaking techniques in order to make an incomplete framework slightly more effective. The traditional spatiotemporal approaches, were dependent purely on the shape of silhouette or the motion of a person. The enhanced spatiotemporal approaches rely on the recorded gait images sequences. Usually, the enhanced spatiotemporal methods extract the individual silhouettes from the photographs or videos and push different details for gait recognition, but some issues were faced like consumption of massive storage space and a very huge amount of computational time taken for evaluation. To tackle such issues, J. Man et al. [19] proposed a spatiotemporal gait representation, gait-energy-image (GEI). A GEI it is single image obtained by collecting and merging the human silhouettes information frame by frame over a complete gait-cycle, it represents a motion information in the form of a template image useful for human identification. GEI represents a temporal average information over a complete gait cycle of the segmented silhouettes of a subject. While comparing GEI to the traditional gait representation by binary silhouette image sequences, GEI not only saves storage capacity and computation time, but it also makes individual frames less susceptible to silhouette noise [19]**.** GEI has been used in time domain exclusively in many methods as the baseline feature to represent the motion and appearance of a subject.

The gait energy image is represented as:

$$G'(p,q) = \frac{1}{N'} \sum_{t=1}^{N'} I(p,q,t)$$

Where $N'$ is the number of extracted frames of the silhouettes over a gait-cycle, t is the frame count or frame no. at a moment of time in a gait cycle and I (p, q) is the original extracted silhouette image with (p, q) values representing image coordinate in 2D.


**Fig. 4 Some examples of well-aligned silhouette image frames with their GEI (Gait Energy Image).**

The GEI templates have been very effective due to their less vulnerability to noise, needs less computational time and storage space. GEI+PCA is one of the popular and simple approach of GEI-based methods, it gives very good and accurate results when there are no absolute changes. But it is believed to fail in conditions with view angles, variations in clothes and other things. Few cases of well-aligned and normalized silhouettes frames in different human walking sequences with GEI are shown in the fig. 4. A gait recognition approach is shown in [21] which is based on GEI template representation by using a modified phase only correlation. Following a supervised feature extraction method that identifies most relevant discriminative features for human identification under various conditions like clothing, baggage handling or carrying and intraclass covariates, they employed an efficient approach to match images with low texture features which includes a spectral weighting function which is band-pass-type to improve the phase only correlation (POC) method that in turn improves the performance of recognition. In another work [20] an approach is proposed which reduces intra-class variations by selecting discriminative human body parts using the group lasso to segment GEI templates.

As GEI prove to be an efficient gait silhouette template for human recognition, on the other hand some unavoidable issues were also faced while using GEI in a gait sequence for recognition such as it loses some information which disrupts the performance due to changes caused by covariate factors such as clothes, baggage carrying situations and changes in view. There was a loss of a considerable amount of physical appearance attribute related details due to silhouettes extraction of a person from an image or video. However, a silhouette also provides valuable details regarding a human body's shape and posture. An important aspect for appearance-based gait identification is the static-shape information of human gait silhouette as it is effective than kinematics for most of the silhouette-based gait recognition approaches [22]. Khalid Bashir et al. [23] proposed a method to select features between each pair of probe and gallery sequences, the method was termed as gait entropy image (GenI). The method suggested that if the most relevant features of gait were selected which may be invariant to changes in gait covariate conditions then a gait recognition system can be developed which works without subject. In their work, they proposed to measure the Shannon entropy to compare the static and dynamic areas of GEI at each pixel location. GenI can be calculated as:

$$\mathrm{GenI} = \mathrm{H}(\mathrm{m}, \mathrm{n}) = -\sum_{t=1}^{N'} p_t(\mathrm{m}, \mathrm{n}) log_2 p_t(\mathrm{m}, \mathrm{n}),$$

where m, n are the pixel coordinates and $p_t(\mathrm{m}, \mathrm{n})$ is the probability that the pixel takes on the *t* value. If the silhouettes are binary images, then K=2, $p_1(m,n) = \frac{1}{N'} \sum_{t=1}^{N'} I(m, n, t')$ (i.e., the GEI) and $p_0(m,n) = 1 - p_1(m,n)$. Which depicts that there is a relation between GenI and a GEI. Moving on from GEI another method is proposed, to preserve the temporal information from loss, in [24] a method is demonstrated which is based on multi-channel temporal encoding termed as "chrono-gait image" (CGI). In this method, the authors have explained how a single CGI pattern is generated by encoding the gait frame's contour image with multi-channel mapping feature.

To tackle the problems raised due to view-condition variations, [25] the stacked progressive auto-encoders (SPAE) suggested that gait images be transformed from arbitrary angles to a particular view. The algorithms presented in his work can handle the conditions with view changes, variations in clothes and baggage carrying conditions to a certain extent. However, the effectiveness of these techniques is still questionable, since gait energy image would result in a loss of temporal information.

Some other appearance-based methods came into picture after some extensive years of research, like a method was introduced by Nikolaos V. et al. [26] in which they proposed a method based on radon transform of binary silhouettes for feature extraction in which, the templates were calculated by transforming the silhouettes, for each gait sequence. In [27] Shiqi Yu et al. found the key fourier descriptors in frequency domain by analysing the spatio-temporal human characteristic of moving silhouettes. Also, Hazem El-Alfy et al. [28] proposed a "geometric view transformation model" (GVTM) to improve the robustness of gait identification under cross-view conditions. GVTM does not corrupt the gait features unlike the other appearance-based methods, GVTM retains the spatial proximity of features while also transforming them more flexibly and effectively than simplistic weak perspective projection-based geometric approaches and 3D model-based approaches. In addition, Liang Wang et al. [29] presented a greatly improved background subtraction algorithm that used statistical shape analysis through each image frame sequence to extract moving silhouettes of a walking subject. Recently, some methods were developed which directly take the silhouettes of a person as input data instead of using the averaged information. The method presented by Rijun Liao et al. in [30] was the first method which includes a deep CNN model as a feature extraction algorithm from human silhouette sequence, and it outperforms the previous methods by an enormous amount. A similar conclusion can also be found in [31] showing how these methods can attain high accuracy in case where cross-view conditions are present. However, it is still a challenging task to handle the conditions such as cross-carrying and cross-clothing appropriately as the subject's appearance, posture and shapes may vary greatly under certain variations. But the recognition accuracy of an individual is highly based on the effective and precise segmentation of silhouette from the background information, which is still a challenging problem in literature. The construction of the gait descriptor may get damaged due to an inaccurate segmentation of silhouette which may also lead to failure in recognition or identification of a subject.

### 3.3.2 Human-Pose based approach

Another kind of feature extraction method is the human-pose-based estimation methods. Modeling of human body-based approach is dependent upon the information that is extracted prior. Modeling or monitoring of body components such as limbs, thighs, arms, and legs is achieved using this process to derive various gait signatures which are further used for identification or recognition of a subject. In this method, gait features are obtained by fitting models (such as 3D models, 2D models) on human body. These methods are used practically enormously as they have certain advantages over the other methods like they're quick to comprehend, invariant of view, scale-invariant, and are untouched by background clutter and noise. The main objective of human gait recognition is to analyze the style of walking and distinguish subjects on the basis of discriminative features. Earlier research approaches were based on 2D gait video sequences like the one described by Yoo *et al.* in their work

[32] described a set of sequential 2D-stick figures based on motion parameters to represent gait signature data for feature extraction. It consists of a back-propagation neural network for recognizing humans by their gait. Due to the fixed camera perspective, self-occlusion, and surface variations, several issues emerged that prevented correct and reliable results in 2D gait recognition. To get around the limitations of 2D video data, 3D gait recognition is recorded simultaneously by more than two cameras that have been calibrated in a static manner. 3D gait recognition obtains motion sequences as applicable in the practical applications by overcoming drawbacks raised in 2D images such as difficulties in fixed viewpoint, variations in surface and self-occlusion. In 3D human pose-based methods human gait recognition is achieved very efficiently, here a subject's gait which is used for reconstruction of 3D human structure is tracked with the help of 3D human models which in turn helps in the extraction of dynamic feature to do further recognition analysis. Keeping in mind the drawbacks of 2D human pose-based recognition methods, in [33] a model is designed for gait analysis which is a 3D temporal motion model robust to illumination changes, occlusions, independent to view and clothing changes. They have used data points for tracking and used an optical motion capture system to take the samples of 4 different subjects, where each volunteer walked with 9 different speed variations from 3-7 km per hr with an increment of 0.5 km per hr on a treadmill. To tackle variations in the angle of view and the issues raised in 2D video-based gait identification, [34] a 3D skeleton-based dataset is created that has 3D information of joints and also 2D silhouette image information, the dataset was recorded by using a second-generation Kinect V2 tool and is used to extract the static and dynamic features for human identification. Following the same techniques which are based on 3D gait recognition to effectively handle the effects of variations in surface and view, Zhao *et al*. [35] developed a method in which multiple cameras were used to record video sequences which are further used to create a 3D human model. Static and dynamic feature sets were recorded including main segment lengths and lower limb motion trajectories respectively for recognition. An accuracy of 70 percent is achieved by using the model developed in his work by inferring the static and dynamic sets from each other. In [36] a marker-less pose recovery method is proposed to capture the 3D human joints and to record the pose parameters sequentially from the data in volume.

The method fuses the pose recovery and classification for action and human gait identification by Using a common viewpoint-free structure. Krezeszowski *et al*. in his work [37] described another marker-less tracking algorithm in which recognition is achieved on a view-independent gait-based data. The data is obtained by the marker-less 3D motion tracking algorithm. He used a particle swarm optimization algorithm for motion tracking. Also, Rijun Liao et al. in [38] proposed a PoseGait named gait recognition method, that takes 3D human body poses (body joints) which are very compact as features. Experiments were conducted on publicly available datasets like CASIA B and CASIA E. They improved the recognition rate by combining three different spatio-temporal features with body pose based on prior human knowledge. The model uses CNN (Convolutional Neural Network) to estimate human 3D pose from images which are used as input features for human gait identification. Typically, no issues or problems were faced in estimating the 3D poses or the coordinates of joints of the human body when changes were observed in external factors or views. Some recent works include the 3D pose estimation in videos using neural networks over 2D keypoints, as Dario Pavllo et al. demonstrated in his work [39] how 3D poses in a video can be efficiently calculated with the help of a fully convolutional model which is dependent upon dilated temporal convolutions over 2D keypoints. In their work they talked about back-projection, which is a semi-supervised method used for training that uses videos data which are unlabeled.

### 3.4 Data Reduction or Feature Seelection

The feature extraction process leads to useful information for gait analysis along with some inadequate or irrelevant features. The high dimensional data acquired after the feature extraction process contains some unimportant features which may disrupt the effectiveness of the whole process. So, to prevent the efficiency from disruption some data reduction or feature selection methods are used. They select some features from the output of the feature extraction stage which effectively describes that data by simultaneously decreasing the noise's effects or unimportant variables and give some commendably predictable or classified results. Some dimension reduction approaches include LDA (linear discriminant analysis), PCA (principal component analysis), DFT (discrete fourier transform), DCT (discrete cosine transform), outlier's removal and so on. Dimension reduction or feature selection is a necessary step before the classification as it increases the effectiveness of a gait analysis system.

A comparative analysis of some data reduction algorithms has been done in table 1.

**Table 1. Survey of some state-of-the-art feature selection algorithm**

| Feature sel. | Algorithm | Advantages |
|---|---|---|
| CSA+DATER [40] | CSA which is matrix-based is the pre-processing algorithm which removes noise and retains the best information whereas to increase the classification ability DATER is used. | CSA is better than PCA as it is based on matrix representation. |
| RF [41] | Random Forest algorithm ranks the features whereas a backward feature elimination search strategy searches throughout the subspaces. | Reduces the computational cost required to recognise a subject. |
| PCA [42] | PCA performs the feature sel. on the wavelet features. Up to the 6th principal component, dimensionality features are effectively separable for the classifiers to be used as features. | Classification performance is increased and computational complexity of the clustering process is decreased. |
| WPCA [43] | The square root of the eigenvalues was used to separate eigenvectors in WPCA. The positive impact of eigenvectors that provide knowledge about specifics grows, while the negative impact of eigenvectors with greater eigenvalues decreases. | Overcomes the issues faced in PCA such as holding low frequencies that show illumination and expression features while removing high frequencies that contain discriminating information. |
| BD [44] | BD applied on random features extracts some successful features using the properties of binomial distribution (mean and variance). Based on which a simple algo. is implemented which selects/extracts the best features. | It is effective in terms of recognition accuracy as well as execution time. |
| FEcS [45] | Fused feature vector (FV) calculates entropy and skewness vectors and selects best subsets of features using FEcS approach. | The proposed technique gave some promising results. |
| Firefly algorithm and Skewness based approach [46] | Firefly algo. (a meta-heuristic approach) implemented on fused vector. Skewness-based method selects best features using fused vectors. | Significantly accurate than other methods. |
| LDA [47] | LDA uses an orthogonal transformation tech. to get *m*-dimensional feature data which are unrelated and are used for classification, it minimizes information loss (m< n where n = dimension of feature data). The algo. selects the direction of large data variance. As the axis of the orthogonal transformations. | Data is easier to identify after feature sel., reduced correlation of different data and transformational result highlights the differences of feature data. |

### 3.5 Classification and Recognition

It is the last stage in the framework of gait recognition. The accuracy of this stage is highly dependent on the previous stages. Classification is done by using some machine learning algorithms which may be supervised, unsupervised or some reinforcement learning algorithms. Classifiers like nearest-neighbour (NN), KNN (k-nearest neighbour), multilayer perceptron, decision tree, SVM, naïve bayes, random forest, DCNN, regression techniques, random tree, bayesNet, QDA (quadratic discriminant analysis), neuro-fuzzy classifier, CART (classification and regression tree), PNN (probabilistic neural network), FIS (Fuzzy interference system)were used for human gait classification in the recent research work.

Rosa Altilio et al. [48] in their survey have performed a comparative analysis using various classifiers such as KNN, SVM. They explained that no specific changes are required to a dataset as most of the classifiers use the same dataset to match it to the classification model. Some distinguishing patterns were present in the extent of extracting information from the data by different algorithms such as KNN and SVM classifiers do not provide a mathematical model of classifier from the training data. Support vectors are produced as outputs using SVM whereas output labels are extracted from the nearest patterns which are classified using KNN. They also mentioned some other statistical and fuzzy logic-based classification algorithmic techniques which finds the parameters of the mathematical model by using training data, for the pattern under classification which might further used to find the probability or the fuzzy membership to a class, respectively. Also, another classifier named CART was proven to show an intermediate behaviour which was used to find the decision tree by training the data. Working on the same note, Aybuke Kececi et al. in their work [49] compared three algorithms named IB1, bayesian net and random forest and shown that they have a very high accuracy rate more than almost 99% for different kinds of activities, walking was the major one among them. The structure of their work includes the use of all the three different classification methods to make the absolute final classified decision based on the majority voting analysis to make a more suitable prediction. According to the data provided in [49] it has been shown that the random forest algorithm produces the best results, with a total accuracy of over 99 percent, amongst all the algorithms. The Bar graph plot representation of the data provided in [49] has been shown in fig. 5. Which shows a comparable bar-plot comparing the implemented results of all six algorithms seen in the work and clearly predicting the RF algorithm as the winner.

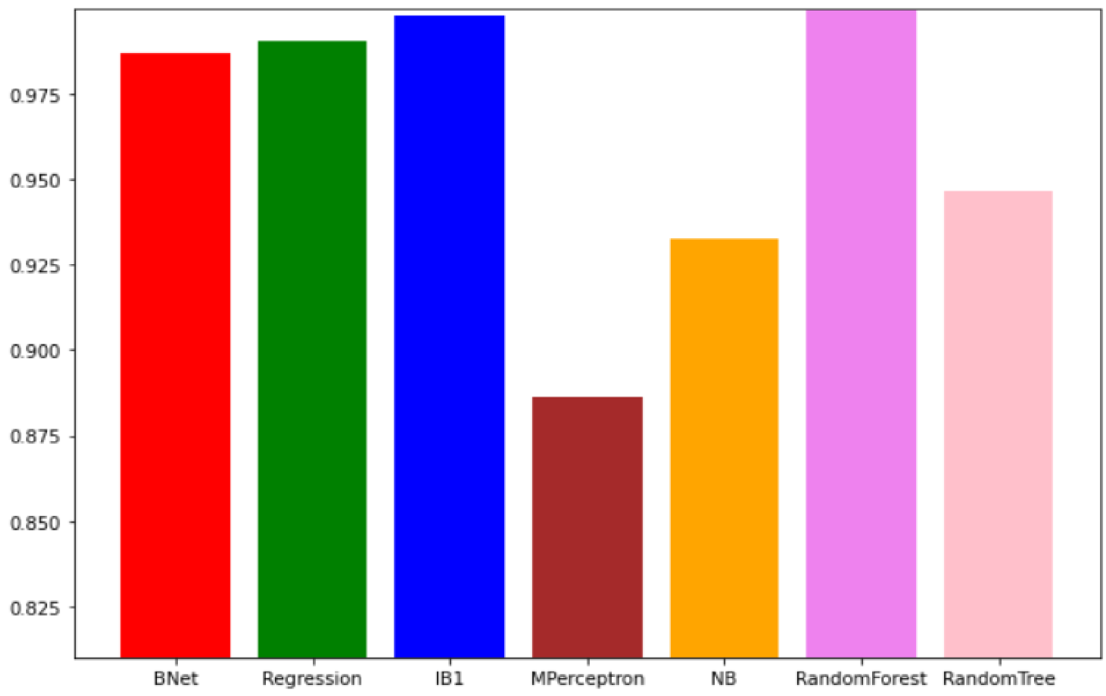



**Fig. 5 Implementations results of 6 algorithms discussed in [49] is presented in the form of a bar plot for walking activity.**

The classification of the sample is done by euclidean distance which is used for new sample classification each time to the nearest training cases. The naïve bayes algorithm is used when the dimensions of the inputs are high whereas the KNN model assigns the majority signs by considering the k nearest neighbor. Naïve bayes is a very simple and effective classifier which sometimes leaves behind other classification methods when their performances are compared. The existence of a specific feature belonging to one class is unrelated to the presence or absence of any other feature, as predicted or estimated by the naive bayes classifier. If the features are dependent upon each other, naïve bayes classifier contributes to the probability by considering all these specifications. Anup Nandy et al. in their work [51] compared the four classification algorithms that is KNN, naïve bayes, random forest and decision tree and analysed the results obtained on each segmented feature. The ROC analysis, the accuracy and precision analysis has been in their work for all the four classifiers on the right range of characteristics. They demonstrated how in this case, the KNN can be explicitly declared as the simple winner classifier. The KNN classifier demonstrated the strongest classification results above the line of prejudice. Therefore, it was said to be treated as optimal classifier. Anup Nandy et al. in [52] experimented with different classifiers such as k-nearest neighbor (k-NN), minimum distance classifier (MDC), and support vector machine (SVM) and checked the efficiency of the proposed method used for feature selection by considering the results obtained using these classifiers. The classifiers were compared and their performance were checked with respect to mean errors and standard deviation of error metrics. The performance checking criteria used was the N- fold cross validation method on misclassification rates to compare the classification results. The classifier is said to be an optimal classifier if the mean and standard deviation produced by it are comparatively smaller than produced by other classifiers. Faezeh Tafazzoli et al. in [50] experimented with two different classifiers: KNN and Naïve Bayes algorithm. He demonstrated how the NN classifier achieves a very high consistent performance comparatively than other classifiers which are used for classification without any prior assumptions regarding the distributions from which the training examples are extracted.  Both positive and negative training sets are present in this classifier.

# CHAPTER 4

# PROPOSED WORK

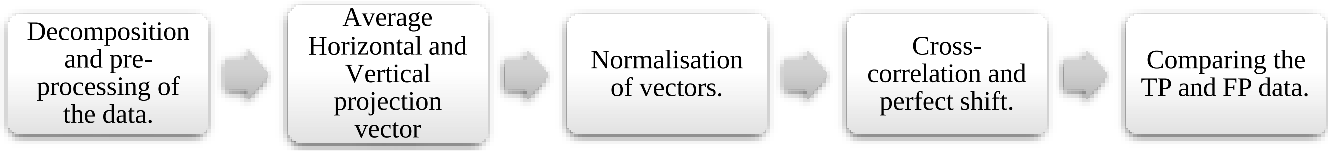



**Fig.6. Framework of the proposed approach**

### 4.1 Decompostion and pre-processing of the video data

The first-person (FP) video is decomposed into frames and some pre-processing steps were performed to remove the empty and turning parts of the video.

### 4.2 Average Horizontal and Vertical projection vector

Each image frame is split into its R, G and B color components and for each components the horizontal and vertical projection vectors are calculated. The horizontal and vertical projection vectors are represented by $H_I$ and $V_I$ respectively.

$$H_I = \frac{1}{n}\sum_{i=1}^{m-1} M_i$$

$$V_I = \frac{1}{m}\sum_{i=1}^{n-1} N_i$$

Where m x n is the size of an image, m = no. of rows and n = no. of columns. $M_i$ and $N_i$ represent the elements of rows and columns.

After calculating three separate horizontal and vertical projection vectors for all the three components that is red, green and blue, the average projection vectors are calculated by simply performing the sample mean that is,

$$H_{Iav} = \frac{1}{3}(H_{IR} + H_{IG} + H_{IB})$$

$$V_{Iav} = \frac{1}{3}(V_{IR} + V_{IG} + V_{IB})$$

Where $H_{Iav}$ and $V_{Iav}$ represents the average horizontal and average vertical projection vectors respectively. $H_{IR}$, $H_{IG}$ and $H_{IB}$ represents the red, green and blue horizontal vector components. Similarly, $V_{IR}$, $V_{IG}$ and $V_{IB}$ represents the red, green and blue vertical vector components.

### 4.3 Normalisation of vectors

The normalized horizontal projection vector is represented as:

$$NH_{Iav} = \frac{(H_{Iav} - \mu_{H_{Iav}})}{(\sigma_{H_{Iav}} * N)}$$

Where $\mu_{H_{Iav}}$ is the average of $H_{Iav}$ , $\sigma_{H_{Iav}}$ is the standard deviation of $H_{Iav}$ and N represents the length of $H_{Iav}$. Similarly, the normalized vertical projection vector is represented as:

$$NV_{Iav} = \frac{(V_{Iav} - \mu_{V_{Iav}})}{(\sigma_{V_{Iav}} * M)}$$

Where $\mu_{V_{Iav}}$ is the average of $V_{Iav}$ , $\sigma_{H_{Iav}}$ is the standard deviation of $V_{Iav}$ and M represents the length of $V_{Iav}$.

### 4.4 Cross-correlation and perfect shift

$$r_H = \sum_{n=-\infty}^{\infty} NH_{Iav}(n) NH_{Iav}{}'(n - l_H)$$

for $l_H = 0, \pm1, \pm2...$

Where $r_H$ represents the cross-correlation value calculated for each consecutive normalised horizontal projection vector values that is $NH_{Iav}(n)$ and $NH_{Iav}{}'(n)$.

Similarly,

$$r_V = \sum_{n=-\infty}^{\infty} NV_{Iav}(n) NV_{Iav}{}'(n - l_V)$$

for $l_V = 0, \pm1, \pm2...$

Where $r_V$ represents the cross-correlation value calculated for each consecutive normalised vertical projection vector values that is $NV_{Iav}(n)$ and $NV_{Iav}{}'(n)$.

$$\max(r_H) = \sum_{n=-\infty}^{\infty} NH_{Iav}(n)NH_{Iav}{}'(n - l_H)$$

for $l_H = l'_H$

Max $(r_H)$ is the peak value of cross-correlation and $l'_H$ is the value of perfect shift or location at which the correlation peak occurs for horizontal projection vectors.

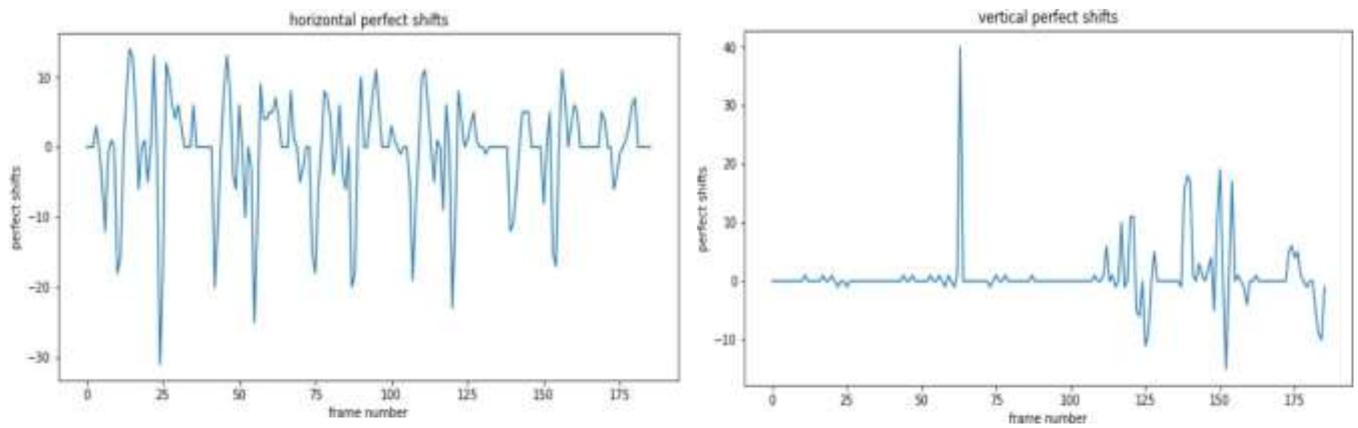
Whereas,

$$\max(r_V) = \sum_{n=-\infty}^{\infty} NV_{Iav}(n)NV_{Iav}{}'(n - l_V)$$

for $l_V = l'_V$

Max $(r_V)$ is the peak value of cross-correlation and $l'_V$ is the value of perfect shift or location at which the correlation peak occurs for vertical projection vectors. The values of all the horizontal perfect shifts and vertical perfect shifts are estimated between each consecutive frame of a video and are plotted on a plot against the frame number.

### 4.5 Comparing the TP and FP data




a) b)

**Fig. 7. Plots of the perfect shift measured between the pair of each consecutive frame of FP video data vs. the frame number for a) average horizontal projection data and b) average vertical projection data.**

The plots (Fig.7.) are drawn to estimate the number of positive peaks count for both the horizontal and vertical perfect shifts. The positive peaks are the absolute maxima peaks occurring in an oscillation. The peak counts are compared with the number of step counts estimated from the TP (third person) video data of the same volunteer for whom the perfect shift data is calculated.

# CHAPTER 5

## DATASET AND DESCRIPTION

The goal of recording this database, that contains two types of data is to serve research applications and practical demands.

### 5.1 Dataset Generation

A database of 24 subjects (15 males and 9 females) is collected, where two types of data is collected using two different phone cameras. One camera is fixed on a tripod stand which is used to collect the third person (TP) walking data, and the tripod stand is placed at a perpendicular bisector drawn at the line segment AB. The line segment AB is the line joining the point A - The point of start of walk and the point B - The point of return. The TP data is capturing the side view of the walking style of a person. The other camera is sticked with the volunteer’s body with the help of a waist belt, which is used to collect the first person (FP) data. The FP data is also captured with the help of a phone camera along with the timestamp information to sync it with the TP data. The FP data is used further to analyse the motion of the camera to estimate the body movements.

At a time, only one volunteer’s walking data is recorded with two cameras recording two different types of data. A person was told to simply walks his/her natural walk wearing the FP camera on his/her belt, simultaneously TP camera placed on the tripod stand is switched on approximately at the same time. Data of the volunteers of both the genders from following age groups was collected: below 5 years, 5-14, 15-24, 25-34, 35-49 and 50-65 years of age.

### 5.2 Precautions taken

- To infer the two types of data from each other, the number of steps in both the TP and FP camera recorded data should be the same.
- There should be no direct sunlight in the view of either of the cameras while recording the data. There should also be no sharp shadows cast by the subject or volunteer in the cameras' field of view. Recording data after sunset and before dark is needed, or recording should be done on cloudy days.

- A sufficient number of samples from various age groups should be available, with enough examples in each category to make the statistical significance suspicious. Any category can only be adequately analysed if it includes a large number of examples.

### 5.3 Analysis of Available Human GAIT Public Datasets

The development and implementation of a new model depends upon the implementation of it on a database. To get successful results for the model, testing on the database of suitable size and variant factors needs to be considered. In past few years a variety of datasets has been made available publicly for research purposes in the area of human gait recognition, estimating poses and motion tracking. Each and every other database is proved to be effective in terms of various factors such as baggage carrying conditions, clothing variations, covariates, change of location or view (that is change in the number of cameras leading to single or multiview), variation in footwear, indoor or outdoor surroundings and so on. A descriptive survey-based analysis of some human gait datasets available publicly has been provided in table 2. The datasets available in the table are categorized into various categories based on their method of acquisition that is video-based, sensor based, and radar-based datasets. In table 3, an analysis of available public motion capture databases has been done. Motion capture directly generates a data which is structured motion from the video data [53]**.** The MoCap data is acquired or collected using various devices and sensors such as Microsoft Kinect, Asus Vicon or Xtion, rgb-d sensors and so on. A human skeleton (graph depicting the joint angles) figure can be simply obtained by experimenting on the motion capture databases. The overall structure of the human body is analyzed by capturing the 3D anatomical positions of the persons movement or by the spatial co-ordinates of the body points. In the field of human gait recognition or pose estimation using MoCap data, obtaining the gait features which can be easily interpreted by humans is a common practice. The databases summarized in table 2 & 3 can be applicable to gait identification, pose estimation, movement tracking, surveillance purposes, clinical or healthcare analysis, sports, and many more.

**Table 2. Available Public Datasets used for Human Gait Analysis. Acronym of Words Used in Table: Video-based(V), Radar-based(R), Sensor-based(S).**

| Data | Cate. | sub./seq. | location | cam/view/style |
|---|---|---|---|---|
| UMD | V | 55 sub. | In | 2 cam/ 4 view |
| CASIA A [54] | V | 20 sub/ 240 seq. | Out | 3 views |
| CASIA B [55] | V | 124 sub. | In | 11 views/ 4 style |
| CASIA C [56] | V | 153 sub./ 1530 seq. | Out | 1 cam |
| floor-sensor-based [57] | S | 15 sub. | in | Different footwears |
| Georgia-Tech [58] | V/S | 18 sub./ 20 seq.; 15 sub./ 268 seq. | out | multiple view |
| TUM-IITKGP [59] | V | 35 sub./ 840 seq. | In | 1 cam/ 6 style |
| HUMANID [60] | V | 74 sub./ 452 seq. | out | 2 cam/ 12 style |
| Motion-Recording-Sensor-Based Dataset [61] | S | 50 sub./ almost 100 seq. | - | 2 style |
| MIT [62] | V | 24 sub./ 194 seq. | in | 1 cam/ 1 view |
| Human Activities and Postural Transitions Dataset [63] | S | 30 sub. | - | style: 3 static postures/ 3 dynamic motions |
| HID UMD 1 and 2 [64] | V | 25 sub./ 100 seq.; 55 sub./ 222 seq. | out | 2 view |
| Ground Reaction forces [65] | S | 62 sub. | - | Data taken as Body mass |
| TUM-GAID [66] | V | 305 sub./ 3370 seq. | In +depth + a udio | Multimodal gait recognition, RGB-D sensor with a 4-channel microphone array |
| OU-ISIR A, B and D [67] | V | 34 sub./ 408 seq.; 68 sub./ 1350 seq.; 185 sub./ 370 seq. | in | 25 views/ 9 speed variations/ 32 style |
| UCSD [68] | V | 6 sub./ 42 seq. | out | 1 cam (stationary camera mounted at an elevation) |
| UWB Impulse Radar Prototype [69] | R | - | in | The individual marched/ walked with either one-arm or two-arm swings. |
| SOTON large dataset | V | 115 sub./ 2128 seq. | In/ out | multiple view |
| First wave-radar-based dataset [70] | R | 49 sub. | entrance to S-building on the MITRE | Signal reflected from the person's torso, legs, and arms. |

| | | | | |
|---|---|---|---|---|
| | | | Bedford campus. | |
| SOTON small dataset | V | 12 sub. | in | 3-4 view |
| SOTON temporal dataset [71] | V | 25 sub/ 2280 seq. | in | multi-view |

**Table 3. Some Public Motion Capture Datasets used for gait recognition and motion tracking**

| 3D motion cap. Data | application | Ac./ mo. | Sub/Seq./ cam view | author (year) |
|---|---|---|---|---|
| HumanEva-I | 2D/ 3D pose estimation/ tracking algo. | 6 ac. | 4 sub/ 6 cam | Rui Li and Alexandru Balan (2006-10) |
| HumanEva-II | 2D/ 3D pose estimation/ tracking algo. | 6 ac. | 2 sub/ 12 cam | Rui Li and Alexandru Balan (2006-10) |
| Carnegie Mellon MoCap Database | 3D data estimation. | 2030 mo. | 97 sub/ 12 infrared MX-40 cam | Carnegie Mellon University Motion Capture Database [72] |
| Leuven Action Database | 3D data estimation/ recognition | 22 ac. | 1 sub/ multi view | Karl Verfaillie and Jan Vanrie (2002) |
| Leuven Commu-nicative Interaction Database | Free motion capture data for research purposes | 20 mo. | 4 sub/ 4 view | Valeria Manera and Ben Schouten (2010) |
| MoCap Data-base HDM05 | Free motion capture data for research purposes | 1500 mo. | 5 sub/ 12 cam | Meinard M¨uller and Tido R¨oder (June 2007) [73] |
| Mocapdata | Free motion capture data for research | ≈18 ac. | Multiple actors/ 12 cam | JAPAN Co. Ltd. 2008-2009 |
| Human Identification at a Distance | Estimation/ recognition of 3D human gait data. | 6 ac. | 20 sub/ 3 view | Sponsored by Defence Advanced Research Projects Agency HID Program |

| | | | | |
|---|---|---|---|---|
| AMASS (archive of MoCap as surface shapes) | It comes in handy for generating deep learning training data, animation, and visualisation. | 11451 ac. | 346 sub | Naureen Mahmood and Nima Ghorbani (2019) [74] |
| Human 3.6M | Useful for 3D human sensing systems. | 17 ac./ 3.6 million poses | 11 sub/ 4 cam | Catalin Ionescu and Dragos Papava (2013) [75] |
| CMU MoBo Database | Human silhouette or gait recognition | 4 ac. | 25 sub/ 6 cam | Ralph Gross and Jianbo Shi (2001) [76] |

# CHAPTER 6

# RESULTS

The volunteers ranged from overweight to obese, and the camera motion was higher for obese people than for underweight people. Some discussions can be conducted using the tabular data provided in table 4., such as in the case of a 5-year-old healthy male, the number of positive peaks in an oscillation from the FP data plot was exactly double that of the number of steps obtained from the TP video data, the data obtained from the FP videos is almost three times that obtained from the TP videos in the next age group, which is 6-14 years old. There is a difference for non-healthy volunteers in terms of steps count and peak, and the FP data for underweight volunteers is less than three times that of the TP data, indicating less camera movement.

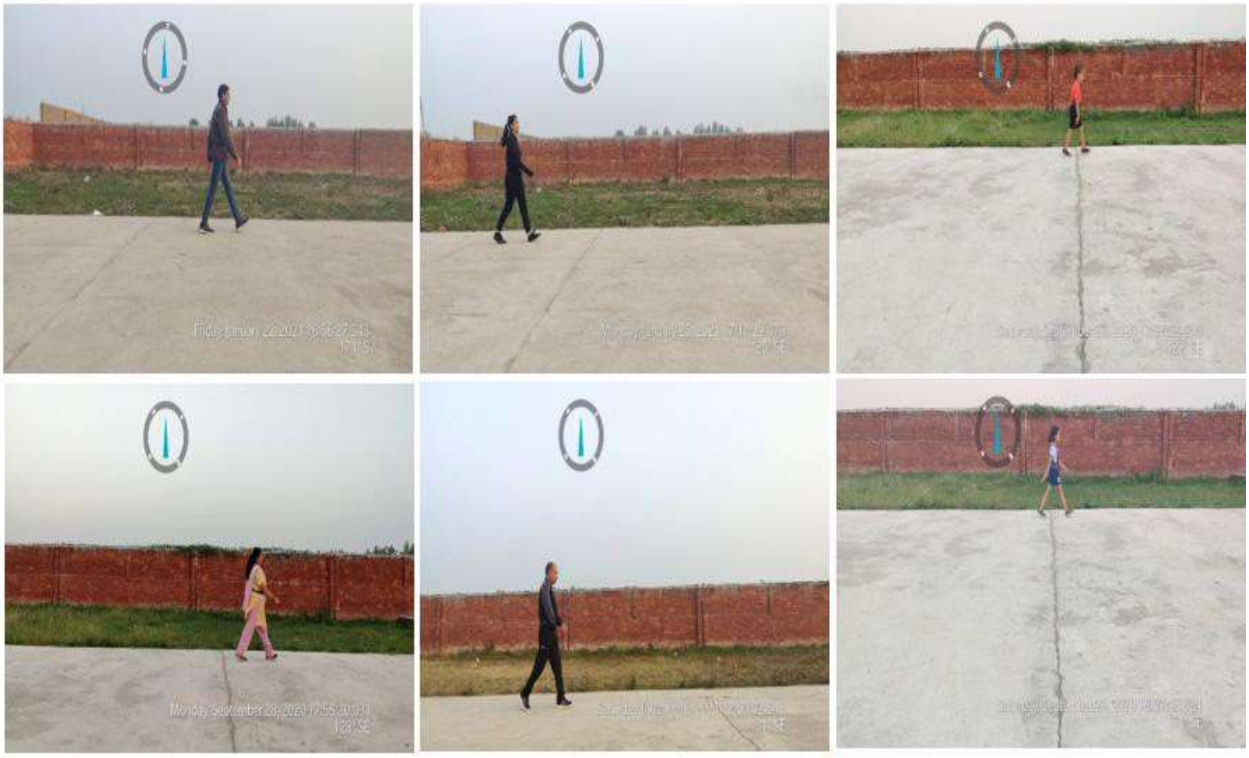


**Fig. 8 Decomposed frames from different volunteer's TP data belonging to different age groups.**

The data obtained from FP videos for the 15-30 range is double that obtained from FP videos, showing that in the case of obese and overweight volunteers, the camera movement was high in comparison to healthy volunteers. For volunteers aged 31 to 49, the peak counts from FP videos are nearly identical to the steps counts from TP videos depicting less camera movement while walking.

**Table 4. The steps count and positive peaks count from the obtained horizontal perfect shift plot for the volunteers classified as underweight, normal, overweight, and obese are used to compare the TP and FP video results.**

| **AGE GROUP (relation B/W TP and FP data)** | **GENDER – AGE** | **STEPS COUNT (TP)** | **PEAK COUNT (FP-HORIZO-NTAL)** | **BMI** | **Health** |
|---|---|---|---|---|---|
| **Below 6 (exactly double)** | M – 5 years | 38 | 76 | 16.5 | Normal |
| **6-14 (almost triple)** | F – 7 years | 31 | 70 | 13.2 | Underweight |
| | F – 10 years | 25 | 73 | 16.8 | Normal |
| | M – 12 years | 25 | 74 | 15.3 | Normal |
| | M – 12 years | 23 | 73 | 16.8 | Normal |
| | M – 13 years | 24 | 70 | 16.4 | Normal |
| | M – 13 years | 24 | 60 | 15.1 | Underweight |
| | M – 14 years | 22 | 54 | 15.4 | Underweight |
| **15-30 (almost double)** | M – 15 years | 22 | 45 | 19.2 | Normal |
| | F – 15 years | 27 | 56 | 21.3 | Normal |
| | F – 16 years | 27 | 55 | 20.6 | Normal |
| | M – 21 years | 26 | 53 | 27.1 | Overweight |
| | M – 22 years | 20 | 40 | 24.9 | Normal |
| | M – 22 years | 23 | 59 | 30.8 | Obese |

| | | | | | |
|---|---|---|---|---|---|
| | F – 25 years | 24 | 50 | 23.4 | Normal |
| | F – 25 years | 25 | 51 | 22.4 | Normal |
| **31-49 (almost same)** | F – 32 years | 24 | 25 | 23.9 | Normal |
| | M – 34 years | 24 | 27 | 27.5 | Overweight |
| | M – 42 years | 21 | 28 | 30 | Obese |
| | F – 44 years | 26 | 29 | 31.7 | Obese |
| | M – 47 years | 24 | 27 | 24.9 | Normal |
| | M – 48 years | 23 | 30 | 30.9 | Obese |
| **50-65 (almost double)** | F – 54 years | 29 | 58 | 27.6 | Overweight |
| | M – 55 years | 24 | 42 | 24 | Normal |

# CONCLUSION AND FUTURE DIRECTION

A small-scale database is suggested, consisting of data from 24 volunteers (15 males and 9 females). Two different types of collected data, FP (first person) and TP (third person), are used for research. FP data is used to appropriately perform camera motion analysis, while TP data is simply used to monitor gate cycles or number of steps count for various subjects. According to the findings, an underweight volunteer of any age group has fewer body movements and thus produces fewer camera movements than a healthy volunteer, while an obese and overweight volunteer of any age group has more body movements and thus produces more camera motion than a healthy participant. The camera moves differently depending on the age group and gender. Future analyses would be able to infer FP and TP data from each other, such as estimating joint angles and determining the cross-correlation between the two databases. Individual recognition can be achieved using their gait, which would include inferring the FP and TP data from each other.

Also an overview about the recent developments made in the field of gait recognition has been provided here. It covers the recent techniques used to make each and every stage of the framework of gait recognition better. It addresses the difficulties that appearance-based methods face, as well as how numerous studies were performed to resolve them, such as the need for various pre-processing techniques, which directly affect the efficiency of the entire recognition mechanism by raising the system's time complexity, and how certain clothing and baggage carrying variations also have a major effect on appearance-based methods. Also, issues such as large storage space usage, a significant amount of computational time needed for assessment, loss of a significant amount of physical appearance attribute related details due to silhouettes extraction of a person from a picture or video, loss of temporal information, and damage to gait descriptor construction due to inaccurate silhouette segmentation. To fix all of these problems, authors have suggested a number of alternative solutions, including extracting gait features directly from raw images, using algorithms like GVTM to enhance gait recognition robustness under cross-view conditions, and using multiple deep learning models, such as a deep CNN as a feature extraction algorithm that outperforms the previous methods by an enormous amount. However, properly managing situations such as cross-carrying and cross-clothing remains a difficult task since the subject's appearance, stance, and shapes can differ greatly under certain variations. Moreover, an individual's recognition accuracy is highly dependent on the efficient and accurate segmentation of silhouette from background information, which is still a difficult problem to solve in literature. Similarly, many problems in human-pose-based approaches, such as fixed camera perspective, self-occlusion, and surface variations, prevented accurate and reliable results in 2D gait recognition, and 3D gait recognition became popular among researchers as a way to address these issues. The public gait recognition and the motion capture database provided an enormous amount of gait data for the benefit of research. It is still an open challenge for the researchers to accurately identify the age and gender of a person using their gaits.

Some issues are hard to resolve, and there are no alternatives available, such as when a person's gait is altered as a result of an injury or accident, or when a person changes his or her gait on intention. Such difficulties need to be addressed properly for a reliable gait recognition system. With the advent of technology, gait will undoubtedly become a specialised method for recognising an individual in the future. The advances in human gait recognition systems would support biometric systems, security, as well as forensic and criminal investigation units.

# REFERENCES


[1] J.E. Mason, I. Traoré and I. Woungang, "Applications of Gait Biometrics," in *Mach. Learning Tech. for Gait Biometric Recogn. Springer, Cham*, pp. 203-208, Feb 2016.

[2] Y. Makihara, D.S. Matovski, M.S. Nixon, J.N. Carter, and Y. Yagi, "Gait recognition: Databases, representations, and applications," *John Wiley & Sons, Inc*., pp 1-15, 2015.

[3] I. Bouchrika, J.N. Carter, J.N. and M.S. Nixon, "Recognizing people in non-intersecting camera views", in *ICDP*, pp. 1-6, 2009.

[4] Y. Makihara, A. Tsuji and Y. Yagi, "Speed-Invariant Gait Recognition," in *Sig. and Image Proces. for Biometrics. Lecture Notes in Elec. Eng.*, vol 292., pp. 209-229, 2014.

[5] N. Lynnerup and P.K. Larsen, "Gait as evidence," *IET Biometrics*, vol 3 no. 2, pp. 47-54, 2014.

[6] A Halim, A. Abdellatif, MI Awad and MRA Atia, "Prediction of human gait activities using wearable sensors," *Proc. of the Inst. of Mech. Engineers, Part H: Journal of Engineering in Medicine*, March 2021.

[7] P. Singh, S. Jain, S. Arora, and U. P. Singh, "Vision-based gait recognition: A survey," *IEEE Access*, vol 6, pp. 70497–70527, Nov 2018.

[8] A. S. Alharthi, S. U. Yunas and K. B. Ozanyan, "Deep Learning for Monitoring of Human Gait: A Review," *IEEE Sensors Journal*, vol 19, no. 21, pp. 9575-9591, 2019.

[9] I. Rida, N. Almaadeed and S. Almaadeed, "Robust gait recognition: A comprehensive survey," *IET Biometrics*, vol. 8, pp. 14-28, Jan. 2019.

[10] Feng Zhao, Qingming Huang and Wen Gao, "Image Matching by Normalized Cross-Correlation," *IEEE Int. Conf. on Acou. Speech and Sig. Proces. Proc., Toulouse, France,* vol. 2, pp. II-II, July 2006.

[11] Du-Ming Tsai, Chien-Ta Lin, “Fast normalized cross correlation for defect detection,” *Pattern Recognition Letters*, vol. 24, Issue 15, pp. 2625-2631, Nov. 2003.

[12] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother and K. W. Bowyer, "The humanID gait challenge problem: data sets, performance, and analysis," in *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 2, pp. 162-177, Feb. 2005.

[13] R. T. Collins, R. Gross and Jianbo Shi, "Silhouette-based human identification from body shape and gait," *Proc. of 5th IEEE Int. Conf. on Automatic Face Gesture Recognition*, *Washington, DC, USA*, pp. 366- 371, 2002.

[14] A. Sundaresan, Amit K. Roy-Chowdhury and R. Chellappa, “A Hidden Markov Model Based Framework for Recognition of Humans from Gait Sequences,” *IEEE Int. Conf. on Image Processing*, 2003.

[15] A. Kale, A. N. Rajagopalan, N. Cuntoor and V. Kruger, "Gait-based recognition of humans using continuous HMMs," *Proc. of 5th IEEE Int. Conf. on Automatic Face Gesture Recognition, Washington, DC, USA*, pp.336-341, 2002.

[16] László Havasi, Zoltán Szlávik and Tamás Szirányi, “Higher order symmetry for non-linear classification of human walk detection,” *Pattern Recognition Letters*, vol. 27, Issue 7, Pages 822-829, May 2006.

[17] Nikolaos V. Boulgouris, Konstantinos N. Plataniotis and Dimitrios Hatzinakos, “Gait recognition using linear time normalization,” *Pattern Recognition*, vol. 39, Issue 5, pp. 969-979, May 2006.

[18] W. Kusakunniran, "Recognizing Gaits on Spatio-Temporal Feature Domain," in *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 9, pp. 1416-1423, Sept. 2014,

[19] J. Han and B. Bhanu, "Individual recognition using gait energy image," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 2, pp. 316-322, Feb. 2006.

[20] I. Rida, S. Almaadeed and A. Bouridane, "Gait recognition based on modified phase-only correlation," in *SIViP* 10, pp. 463–470, Mar. 2016.

[21] I. Rida, X. Jiang and G. L. Marcialis, "Human Body Part Selection by Group Lasso of Motion for Model-Free Gait Recognition," in *IEEE Signal Processing Letters*, vol. 23, no. 1, pp. 154-158, Jan. 2016.

[22] A. Veeraraghavan, Amit Roy-Chowdhury, and R. Chellappa, Rama, "Role of Shape and Kinematics in Human Movement Analysis," *in Proc. IEEE Comput. Soc. Conf. on Comput. Vis. Pattern Recognit. (CVPR*), vol. 1, pp. 730-737, 2004

[23] Khalid Bashir, Tao Xiang and Shaogang Gong, "Gait recognition without subject cooperation," *Pattern Recognit. Let*., vol. 31, 13, pp. 2052-2060, 2010

[24] Chen Wang, Junping Zhang, Liang Wang, Jian Pu, and Xiaoru Yuan, "Human Identification Using Temporal Information Preserving Gait Template," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 34, 11, pp. 2164–2176, Nov. 2012.

[25] Shiqi Yu, Haifeng Chen, Qing Wang, Linlin Shen, and Yongzhen Huang, "Invariant feature extraction for gait recognition using only one uniform model," *Neurocomput*., vol. 239, pp. 81-93, 2017.

[26] N. V. Boulgouris and Z. X. Chi, "Gait Recognition Using Radon Transform and Linear Discriminant Analysis," in *IEEE Trans. on Image Processing*, vol. 16, no. 3, pp. 731-740, March 2007.

[27] S. Yu, L. Wang, W. Hu, and T. Tan, "Gait analysis for human identification in frequency domain," in *Proc. of the 3rd Int. Conf. on Image and Graphics*, pp. 282–285, Hong Kong, China, Dec 2004.

[28] H. El-Alfy, C. Xu, Y. Makihara, D. Muramatsu, and Y. Yagi, "A geometric view transformation model using free-form deformation for cross-view gait recognition," in *Proc. of the 4th Asian Conf. on Pattern Recognition (ACPR 2017 IEEE*, Nov 2017, p. (To appear).

[29] Liang Wang, Tieniu Tan, Weiming Hu and Huazhong Ning, "Automatic gait recognition based on statistical shape analysis," in *IEEE Transactions on Image Processing*, vol. 12, no. 9, pp. 1120-1131, Sept. 2003.

[30] Z. Wu, Y. Huang, L. Wang, X. Wang and Tieniu Tan, "A Comprehensive Study on Cross-View Gait Based Human Identification with Deep CNNs," *IEEE TransPattern Anal Mach Intell.*, vol. 39 no. 2, pp. 209-226, Feb. 2017.

[31] Yuqi Zhang, Yongzhen Huang, Liang Wang and Shiqi Yu, "A comprehensive study on gait biometrics using a joint CNN-based method", *Pattern Recognit.*, vol. 93, pp. 228-236, Sept. 2019.

[32] J. Yoo, D. Hwang, K. Moon and M. S. Nixon, "Automated Human Recognition by Gait using Neural Network," *First Workshops on Image Processing Theory, Tools and Applications, Sousse, Tunisia*, pp. 1-6, Nov. 2008.

[33] Urtasun R and Fua P, "3D tracking for gait characterization and recognition," *6th IEEE international conference on automatic face and gesture recognition. Seoul,*pp. 17–22, 2004.

[34] Y. Wang, J. Sun, J. Li and D. Zhao, "Gait recognition based on 3D skeleton joints captured by kinect," *IEEE International Conference on Image Processing (ICIP),Phoenix, AZ, USA*, pp. 3151-3155, 2016.

[35] Guoying Zhao, Guoyi Liu, Hua Li and M. Pietikainen, "3D gait recognition using multiple cameras," *7th International Conference on Automatic Face and GestureRecognition (FGR06), Southampton, UK,* pp. 529-534, 2006.

[36] J. Gu, X. Ding, S. Wang, and Y. Wu, "Action and gait recognition from recovered 3-D human joints," *IEEE Trans. Syst. Man. Cybern. B. Cybern*. vol. 40 no. 4, pp. 1021-33, Aug. 2010.

[37] T. Krzeszowski, A. Michalczuk, B. Kwolek, A. Switonski and H. Josinski, "Gait recognition based on marker-less 3D motion capture," *10th IEEE International Conference on Advanced Video and Signal Based Surveillance, Krakow, Poland,* pp. 232-237, 2013.

[38] Rijun Liao, Shiqi Yu, Weizhi An and Yongzhen Huang, "A model-based gait recognition method with body pose and human prior knowledge," *Pattern Recognition*, vol. 98, Feb. 2020.

[39] D. Pavllo, C. Feichtenhofer, D. Grangier, M. Auli, “3D human pose estimation in video with temporal convolutions and semi-supervised training,” in *Proc. IEEE Conf. on Comput. Vis. Pattern Recognit.*, pp. 7753–7762, 2019.

[40] Dong Xu, Shuicheng Yan, Dacheng Tao, Lei Zhang, Xuelong Li and Hong-Jiang Zhang, "Human Gait Recognition with Matrix Representation," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 7, pp. 896-903, July 2006.

[41] Y. Dupuis, X. Savatier and P. Vasseur, “Feature subset selection applied to model-free gait recognition,” *Image and Vision Computing*, vol. 31, no. 8, pp. 580-591, Aug. 2013.

[42] Mitchell Yuwono, Steven W. Su, Ying Guo, Bruce D. Moulton and Hung T. Nguyen, “Unsupervised nonparametric method for gait analysis using a waist-worn inertial sensor,” *Applied Soft Computing*, vol. 14, Part A, pp. 72-80, Jan. 2014.

[43] Meryem Uzun-Per and Muhittin Gökmen, “Face recognition with Patch-based Local Walsh Transform,” *Signal Processing: Image Communication*, vol. 61, 2018, pp. 85-96, Feb. 2018.

[44] H. Arshad, M. A. Khan, M. Sharif, M. Yasmin, and M. Y. Javed, “Multi-level features fusion and selection for human gait recognition: an optimized framework of Bayesian model and binomial distribution,” *Int. J. Mach. Learn. & Cyber.*, vol.10, pp. 3601–3618, Mar. 2019.

[45] H. Arshad, MA Khan, MI Sharif, M.Yasmin, JM R. S. Tavares, YD Zhang and SC Satapathy, “A multilevel paradigm for deep convolutional neural network features selection with an application to human gait recognition,” *Expert Systems*, pp. e12541, Mar. 2020.

[46] A. Mehmood, MA. Khan, M. Sharif, Muhammad, S. Khan, M. Shaheen, T. Saba, Riaz, N. Riaz and I. Ashraf, “Prosperous Human Gait Recognition: An End-to-End System based on Pre-trained CNN Features Selection,” *Multimedia Tools and Applications*, Apr. 2020.

[47] H. Guo, B. Li, Y. Zhang, Y. Zhang, W. Li, F. Qiao, X. Rong and S. Zhou, "Gait Recognition Based on the Feature Extraction of Gabor Filter and Linear Discriminant Analysis and Improved Local Coupled Extreme Learning

Machine", *Mathematical Problems in Engineering*, vol. 2020, pp. 245-254, Apr. 2020.

[48] R. Altilio, A. Rossetti, Q. Fang, X. Gu and M. panella, “A comparison of machine learning classifiers for smartphone-based gait analysis,” *Med Biol Eng Comput*, vol. 59, pp. 535–546, Feb 2021.

[49] A. Kececi, A. Yildirak, K. Ozyazici, G. Ayluctarhan, O. Agbulut and I. Zincir, “Implementation of machine learning algorithms for gait recognition,” *Engineering Science and Technology, an International Journal*, vol. 23, no. 4, pp.931-937, Aug. 2020.

[50] Tafazzoli, Faezeh & Bebis, George & Louis, Sushil & Hussain, Muhammad, “Genetic feature selection for gait recognition,” *Journal of Electronic Imaging*, vol 24, pp. 31-36, 2015.

[51] Anup Nandy, Rupak Chakraborty and Pavan Chakraborty, “Cloth invariant gait recognition using pooled segmented statistical features,” *Neurocomputing*, vol. 191, pp. 117-140, May 2016.

[52] A. Nandy, A. Pathak and P. Chakraborty, “A study on gait entropy image analysis for clothing invariant human identification,” *Multimed Tools Appl*, vol. 76, pp. 9133–9167, 2017.

[53] M. Balazia and P. Sojka, "You are how you walk: Uncooperative MoCap gait identification for video surveillance with incomplete and noisy data," *IEEE International Joint Conference on Biometrics (IJCB), Denver, CO, USA*, pp. 208-215, 2017.

[54] S. Yu, D. Tan and T. Tan, "A Framework for Evaluating the Effect of View Angle, Clothing and Carrying Condition on Gait Recognition," *18th InternationalConference on Pattern Recognition (ICPR'06), Hong Kong, China*, pp. 441-444, Jan. 2006.

[55] Liang Wang, Tieniu Tan, Huazhong Ning and Weiming Hu, "Silhouette analysis-based gait recognition for human identification," in *IEEE Transactions on PatternAnalysis and Machine Intelligence*, vol. 25, no. 12, pp. 1505-1518, Dec. 2003.

[56] Daoliang Tan, Kaiqi Huang, Shiqi Yu and Tieniu Tan, "Efficient Night Gait Recognition Based on Template Matching," *18th International Conference on*

*Pattern Recognition (ICPR'06)*, Hong Kong, China, pp. 1000-1003, 2006.

[57] Robert J. Orr and Gregory D. Abowd., “The smart floor: a mechanism for natural user identification and tracking,” In *CHI EA '00: CHI '00 Extended Abstracts on Human Factors in Computing Systems, NY, USA*, pp. 275–276, Apr. 2000.

[58] AY. Johnson and AF. Bobick, “A Multi-view Method for Gait Recognition Using Static Body Parameters” in *AVBPA, Audio- and Video-Based Biometric Person Authentication*, pp. 301-311, Aug. 2001.

[59] Martin Hofmann, Shamik Sural, and Gerhard Rigoll, “Gait Recognition in the Presence of Occlusion: A New Dataset and Baseline Algorithms”, in *Proceedings of the 19th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, pp. 99–104, Jan. 2011.

[60] P. J. Phillips, P. Grother, S. Sarkar, I. S. Vega, and K. Bowyer, “Baseline Results for the Challenge Problem of Human ID Using Gait Analysis,” *5th IEEE International Conference on Automatic Face and Gesture Recognition (FGR), with CD-ROM, Washington, D.C., USA*, pp. 137-142, Jan. 2002.

[61] D. Gafurov, E. Snekkenes, and P. Bours, “Gait Authentication and Identification Using Wearable Accelerometer Sensor,” *IEEE Workshop on Automatic Identification Advanced Technologies*, pp. 220 – 225, Jul. 2007.

[62] R. T. Collins, R. Gross and Jianbo Shi, "Silhouette-based human identification from body shape and gait," *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition, Washington, DC, USA*, pp. 366-371, 2002.

[63] Jorge-L. Reyes-Ortiz, Luca Oneto, Albert Samà, Xavier Parra and Davide Anguita, “Transition-Aware Human Activity Recognition Using Smartphones,” *Neurocomputing*, vol. 171, pp. 754-767, 2016.

[64] A. Kale, N. Cuntoor, and R. Chellappa, “A framework for activity-specific human identification,” in *IEEE Proc. International Conference on Acoustics, Speech, andSignal Processing (ICASSP)*, vol. 4, pp. 3660–3663, 2002.

[65] J. Jenkins and C. Ellis, “Using Ground Reaction Forces from Gait Analysis: Body Mass as a Weak Biometric,” in *A. LaMarca, M. Langheinrich, and K.N. Truong (eds), Pervasive Comput., Lecture Notes in Computer Science*, vol 4480, pp 251-267, 2007.

[66] Martin Hofmann, Jürgen Geiger, Sebastian Bachmann, Björn Schuller and Gerhard Rigoll, “The TUM Gait from Audio, Image and Depth (GAID) database: Multimodal recognition of subjects and traits,” *Journal of Visual Communication and Image Representation*, vol. 25, no. 1, pp. 195-206, Jan. 2014.

[67] Y. Makihara, H. Mannami, A. Tsuji, Md. A. Hossain, K. Sugiura, A. Mori and Y. Yagi, “The OU-ISIR Gait Database Comprising the Treadmill Dataset,” *IPSJ Transactions on Computer Vision and Applications*, vol. 4, pp. 53-62, Apr. 2012.

[68] J.J. Little, J.J. and Jeffrey Boyd, “Recognizing People by Their Gait: The Shape of Motion.,” *J. Comput. Vis*. vol. 1, pp. 1–33, 1998.

[69] Y. Wang and A. E. Fathy, "Micro-Doppler signatures for intelligent human gait recognition using a UWB impulse radar," *IEEE International Symposium on Antennas and Propagation (APSURSI), Spokane, WA, USA*, pp. 2103-2106, 2011.

[70] Michael. Otero, “Application of a continuous wave radar for human gait recognition,” *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 5809, pp. 538-548, May 2005.

[71] D. S. Matovski, M. S. Nixon, S. Mahmoodi and J. N. Carter, "The Effect of Time on Gait Recognition Performance," *in IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 543-552, April 2012.

[72] M. Balazia and P. Sojka, “Gait recognition from motion capture data,” *ACM Trans Multimedia Comput Commun Appl*, vol. 14, no. 1s, pp. 22:1--22:18, Feb. 2018.

[73] M. Muller, T. Roder, M. Clausen, B. Eberhardt, B. Kruger, and A. Weber, Documentation mocap database hdm05. Technical Report CG-2007-2, *Universitat Bonn*, 2007.

[74] N. Mahmood and N. Ghorbani, N. F. Troje, G. P. Moll, M. J. Black, “AMASS: Archive of Motion Capture as Surface Shapes,” *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 5441--5450, Oct. 2019.

[75] C. Ionescu, D. Papava, V. Olaru, C. Sminchisescu, “Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments,” *IEEE Trans Pattern Anal Mach Intell.*, vol. 36, no. 7, pp.1325-39, Jul. 2014.

[76] R. Gross and J. Shi, “The CMU motion of body (MoBo) database," *Robot. Inst., Pittsburgh, PA, USA, Tech. Rep. CMU-RI-TR-01-18,* Jun. 2001.