CHAPTER 2

# Literature Review

Research on Computer Vision-based gait recognition started around 2005. However, most of the initial approaches rely on the extraction of hand-crafted features. These methods can be conveniently implemented on low-end systems with limited RAM and processing power. In the later years, several Deep Learning-based gait recognition approaches have been developed that are capable of handling challenging scenarios such as walking speed variation, viewpoint variation, etc. This chapter provides an overview of the trend of research in Computer Vision-based gait recognition, starting from the initial hand-crafted feature-based approaches to the modern Deep Learning techniques. The literature survey is presented by categorizing the existing gait recognition approaches into two broad categories, namely, (i) Traditional Approaches, i.e., methods that were developed prior to the introduction of Deep Learning methodologies which can be further sub-divided as either appearance-based or model-based, and (b) Modern Approaches, i.e., methods that use the latest discriminative or generative Deep Learning architectures.

## 2.1 Traditional Approaches

Traditional gait recognition techniques in the literature can be broadly categorized as either model-based or model-free (i.e., appearance-based). Among these, the model-based approaches [22, 23, 24] use a pre-defined human body structure such as a skeletal model and try to fit the movements of body parts into this model. The performance of these approaches is significantly dependent on the availability of high-resolution image/video, and these are generally invariant to viewpoint and scale variation [25]. A limitation of model-based approaches is their high computational cost while fitting the images to the models. Model-based gait features are of two types: (a) approaches working with kinetic parameters, and (b) those working with spatio-temporal parameters. Kinetic parameters require measurement of body structure variation over a gait cycle such as the angle variation among pairs of skeleton joints. On the other hand, spatio-temporal parameters include features such as walking speed, phase length, gait cycle period, etc. [26]. In appearance-based (or model-free) gait recognition approaches, the gait features are extracted directly from the silhouette sequence of a walking person. These approaches are robust to noise and are less computationally intensive, and hence achieved significant popularity. Most of these carry out silhouette extraction from video frames, silhouette normalization, and finally classification either in image space or latent space to identify an individual [18]. The first among these model-free methods is the one that uses the Gait Energy Image (GEI) feature [27] that is computed by averaging all the silhouettes corresponding to the frames present in a gait cycle. Later, several extensions to the GEI feature have been developed, namely, the Motion Silhouettes Image (MSI) [28], Gait History Image (GHI) [29], Enhanced GEI (EGEI) [30], Motion Energy Image (MEI) [31], Frame Difference Energy Image (FDEI) [32], Active Energy Image (AEI) [20], Gait Entropy Image (GEnI) [33], Gait Flow Image (GFI) [34], Pose Energy Image (PEI) [1] and Chrono-Gait Image (CGI) [35], etc. These approaches are further classified into two groups as explained in the following two sub-sections.

### 2.1.1 Using Complete Gait Cycle Aggregation

The gait recognition performance can be affected by varying types of co-variate factors such as clothing, carrying objects, elapsed time, and others. While matching a probe subject with the gallery set samples, often large intra-subject variations are caused due to the addition of co-variate objects that dominate over the subtle inter-subject variations, resulting in incorrect recognition. In [36], a method is discussed that handles the above problem using a joint intensity and spatial metric by suppressing the high intra-subject differences and simultaneously enhancing the inter-subject differences. The work in [37] presents another gait recognition technique based on two feature extraction methods, namely, the Multiscale Local Binary Pattern (MLBP) and Gabor filter bank considering co-variate conditions such as walking on different surfaces. The gait features of the same person are expected to change with time, and the work in [38] carries out a thorough study of the influence of this elapsed time on the accuracy of gait recognition. Later, an extension of this work is presented in [39] where experiments are conducted using a more extensive data set.

Gait recognition from a different viewpoint is addressed in [24, 40, 41], in which [24] utilized singular value decomposition (SVD) to transform the gallery and probe features into the same view. The work in [40] uses Support Vector Regression (SVR) model to reduce the effect of multiple views, where the SVR seeks to find a local region of interest (ROI) under a single view angle so that the related motion information can be predicted from a different view angle. In [41], multiple regression procedures are used to detect correlated walking motions that are encoded as gait features. Sparse regression based on the ElasticNet is adopted as the regression function that is free from the problem of overfitting to come up with more stable regression models for VTM construction. A Gaussian Process (GP) classification framework is used to estimate the view angle [42] and Canonical Correlation Analysis (CCA) is used to measure the correlation similarity. An arbitrary view transformation matrix has been constructed using a 3D gait model of multiple (non-target) subjects. The above methods are suitable for application if the input and target viewpoints are known. The method in [43] presents an arbitrary view transformation approach by constructing a transformation matrix associated with the two views using a 3D gait database consisting of the visual hulls from multiple non-target subjects. This matrix is used to transform the gallery gait features with varying views into features with the same view as the probe. Using this scheme, we can minimize the impact of the view difference. The gait energy image (GEI) based method [44] establishes a robust view transformation model via robust principal component analysis where partial least square is used as a feature selection. This method works for multi-view gait recognition under different wearing and carrying conditions. The work in [45] can handle arbitrary walking directions by cluster-based averaging of gait images. The subspace projection-based approaches project the gait features in a common subspace via asymmetric mapping. In [46], Coupled Bi-linear Transformation is used to project the GEI into a lower dimensional matrix subspace. The work in [47] modeled the gait data as a tensor and introduced three robust discriminative representations by tensor analysis along with criteria for tensorial coupled mapping. The work in [48] discusses about domain-dependent projection matrices for multi-view gait recognition to tackle single common discriminative space projection. These methods have been proven to be both accurate and efficient. But since these models transform view in pairs, there is a necessity of multiple models to carry out transformation between varying pairs of viewpoints.

### 2.1.2 Using Pose Features Approach

Pose-based gait recognition approaches [1, 49, 50], in general, provide higher accuracy than feature aggregation-based methods [31, 32, 51] due to their inherent capability of capturing the gait dynamics at a high resolution. In these methods, a set of human walking poses is first estimated either in the form of skeleton-representation or image-representation. Pose estimation techniques in the literature are grouped into either bottom-up (like DeepCut [52], OpenPose [53]) or top-down (like AlphaPose [54], and Mask-RCNN [55]). Pose estimation takes into consideration the dynamics of motion and hence the integration of a suitable pose estimation method before gait feature extraction usually results in a more accurate gait recognition performance. The method in [53] describes a simple 2D skeleton pose estimation technique from a particular view. This approach is only suitable to perform gait recognition with similar co-variate conditions and viewpoints. The most common image-based pose estimation techniques used in gait recognition, as described in [1, 50, 56], focus on deriving representative poses in a gait cycle by dividing the entire cycle into a certain number of smaller parts. A commonly used technique to segregate the gait cycle is the application of a suitable clustering algorithm to a set of image frames. The number of clusters in the clustering is determined by plotting a rate-distortion curve representing the variation of clustering error with increment in the number of clusters and then choosing the optimal point at which the curve attains saturation.

The work in [1] is the first attempt towards developing a key pose-based gait recognition approach in which a feature termed Pose Energy Image (PEI) is constructed. Here, a set of key poses in walking is determined first following which the frames of an input sequence are mapped to the appropriate key poses. The number of key poses used to compute the PEI feature is decided by plotting a rate distortion plot which shows the variation of clustering error with respect to the number of key poses. The mapping of frames to key poses is done by considering the following temporal constraints: (i) if there are $K$ number of key poses, and a frame is mapped to key pose $k$ (where $1 <= k <= K$), the subsequent frame must be mapped to either key pose $k$ or key pose $k+1$, (ii) if a frame is mapped to key pose $K$, i.e., the last key pose, the subsequent frame must be mapped to either key pose $K$, or key pose 1. Later key pose-based gait recognition approaches [57, 58] are also based on a similar concept. However, the procedure for mapping the frames into appropriate key poses, as discussed above, is likely to suffer if the video frame rate differs significantly during capturing of the training and test sets, or if an individual walks with different speeds during the training and test data capturing phases.

A 3D skeleton-based gait recognition method is presented in [59] in which the trajectories of the major joints are extracted to compute dynamic gait features, and the lengths of the segments between pairs of joints have been used as static parameters to aid in analysis, simulation and recognition. A 3D skeleton model is defined and a tree structure is applied to represent this skeleton, with each node of the tree corresponding to a particular joint. The main advantage of the method is that it can conveniently deal with viewpoint and surface variations that are quite difficult to solve using 2D analysis-based methods. Another view independent gait recognition approach has been developed by [60] where person identification is achieved using motion data obtained by markerless 3D motion tracking algorithm. The approach [61] is based on analyzing the leg and arm movements where the initial shape model is created based on anatomical [62] proportions which consists of three main regions. A posterior model is constructed upon the movements of the articulated parts of the body that are extracted via analyzing the variations of their angles with the vertical axis. The motion patterns of the moving parts are described by Fourier analysis. The main focus of this work is on increasing the discrimination capability of the model through extra features of the motion of the arms.

The introduction of the Kinect RGB-D camera has contributed to the growth of several gait recognition algorithms by utilizing the depth and skeletal streams captured by the device in real-time. The Kinect-based 3D gait recognition method in [63] utilizes the skeleton data to construct two features, namely the joint relative distance (JRD) and joint relative angle (JRA) computed among different skeletal joints. The method utilizes the spatio-temporal changes in the different JRD and JRA features over a complete gait cycle to represent the gait signatures mathematically. A genetic algorithm is used for joint-pair selection approach which evaluates every joint-pair based on its relevance and distinctiveness for JRD and JRA-based gait signatures, and a dynamic time warping (DTW)-based kernel is used for gait classification. Finally, a rank-level fusion of JRD and JRA features is done to further boost up the recognition performance [63]. The work in [64] also shows that gait and static body measurements are important biometric factors for passive human recognition.

Traditional view transformation methods use mapping techniques where gait image from one view is transformed to another view using a View Transformation Model (VTM). The work in [65] uses perspective projection of arbitrary view to side view and optical flow information for cross view to construct the VTM. The VTM described in [66] learns to transform images among various views in the frequency domain, whereas that in [67] use a quality dependent framework to compute the dissimilarities of the projection of subjects in a joint subspace. In another work, namely [68], a gait recognition method is described based on partitioning and Canonical Correlation Analysis (CCA). Here, a GEI image is separated into five non-overlapping parts, and for each part CCA is used to model the correlation. The work in [57, 58, 69, 70] also discuss few key pose-based gait recognition approaches using the RGB, skeleton, and depth streams from Kinect. These are based on the same idea as in [1] and hence are not discussed in depth here.

## 2.2 Modern Approaches

This section covers gait recognition approaches that make use of Deep Learning models. These approaches are further classified into two groups based on the characteristics of Deep Learning models, out of which the first group includes networks based on Convolutional Neural Networks (CNN) and the next group focuses on the use of Generative Adversarial Networks (GAN) models.

### 2.2.1 CNN-based Approach

Deep Learning-based approaches have also been developed for Computer Vision-based recognition, floor sensor-based recognition, and wearable sensor-based recognition. Out of these, a major focus has been given to the first category of approaches [71, 72, 73, 74, 75, 76]. The user authentication from wearables is a sensible methodology in constrained scenarios where physical participation is required from each and every subject [71]. For example, IDNet (IDentification Network) [71] is an architecture used for authenticating mobile users from smartphone acquired motion signals. The modern phones possess highly accurate inertial sensors [77, 78] that allow capturing of the motion data that can be conveniently used to perform gait recognition. Some other research work in this domain use the data acquired from a sensor in a known and fixed position [79, 80, 81, 82, 83, 84, 85, 86]. IDNet leverages Deep Convolutional Neural Networks [87] as universal feature extractors to discriminate between gait signatures from different subjects. Deep learning-based approaches are more useful for application when the gait recognition scenario becomes challenging, e.g., in situations where the silhouette shape changes due to the addition of co-variate objects, or situations where the training and test data are captured from different viewpoints. Convolution Neural Networks have been used as a classification model in [88] to come up with an approach that is robust to viewpoint variations, co-variate factors, and occlusion. Another view-invariant gait recognition technique using Deep AutoEncoder has been discussed in [89] in which view transformation model has been employed to convert silhouettes from one view into the target view. This model is based on a multilayered stacked Autoencoder that can effectively synthesize the gait features at different views in a progressive manner.

The work in [90] show the use of Deep-CNNs for cross-view recognition by learning similarities among various arbitrary views. The authors in [91] developed GEINet which use a CNN framework to project the GEI images into a latent space. This method has been seen to be effective for small view variations. The approach in [92] captures the static and dynamic information of a gait sequence through combined LSTM-CNN based networks. [93] used a CNN framework for pose-based gait recognition using optical flow corresponding to the various body parts across multiple frames and perform classification using VGG and ResNet models. The model has shown consistent and good performance for varying co-variate conditions but since it utilizes the silhouette shape directly, it is not expected to work well for dynamic pose differences. The GaitSet proposed in [94] views gait as a set of frames rather than a temporal sequence of frames. Here, gait features are derived by completely ignoring the temporal constraints of gait. Useful information from different sequences with varying co-variate conditions and different view angles are integrated to learn the intrinsic traits of a subject in a better way.

### 2.2.2 GAN-based Approach

Among the GAN-based approaches, the work in [95] describes an approach which is view invariant as well as co-variate condition invariant. Here, a model termed GaitGAN is introduced that is a form of Generative Network with two discriminators. The first discriminator preserves the identity while the second one indicates if an input image is real/fake. Next, an extended version of GaitGAN termed GaitGANv2 has been developed by the same authors in [96] by incorporating an additional contrastive loss. The contrastive loss is fused with softmax loss to preserve the identity of the subject in a better way. The work in [97] presents another Siamese-type Deep Neural Network termed JUCNet that can learn unique-gait and cross-gait representations. Here, a probe subject identity is labeled as either ”same” or ”different” with respect to a gallery subject by forwarding the gait features of the two subjects through the two parallel channels of the Siamese network. The network consists of three output branches, two of which learn unique-gait representation and the other one learn concatenated cross-gait representation. The quintuplet loss function used in this work helps the model to get trained with the inter-class differences and intra-class similarities effectively from the training sequences. A few work on handling co-variate conditions and viewpoint changes in gait recognition have also been done using Generative Adversarial Networks (GANs). Usually, GANs are trained to generate the desired images (i.e., either view-transformed or co-variate object removed) from random noise. However, this may also remove certain intrinsic identifiable features of the person which are essential for cross-view gait recognition. To overcome the shortcomings mentioned above, Multi-task Generative Adversarial Network (MGAN) [50] has been proposed to learn view-specific features from gait templates. Input images at the different poses are first transformed to images corresponding to the target view using the MGAN view transformation model. Training of the network is accomplished using cross-entropy loss and multi-task adversarial loss. The transformed multi-channel gait template is termed as the Period Energy Image and this feature has been seen to preserve a higher temporal and spatial information compared to many other appearance-based gait templates. During testing, a view-angle classifier is trained with cross-entropy loss to predict the view angle of the Period Energy Image before carrying out view transformation and recognition. The work in [98] presents a hybrid method that first removes the co-variate objects using a generation module (consisting of alpha blending GAN (Ab-GAN)), and next learns discriminant features from the generated templates using a discrimination module. There are two independent encoder-decoder networks: one that generates a gait template without the co-variate objects, and the other that estimates the alpha blending parameters (or, alpha matte). It avoids the unnecessary generation of those portions of the silhouette images which are unaffected by the co-variate objects to obtain the identity-preserved gait templates. In [99], first, the GEI gait template with co-variate objects is passed through an encoder network to obtain disentangled identity and co-variate features, following which a decoder is used to generate two images: (a) a reconstructed input GEI template, and (b) a GEI template without co-variate objects. Suitable reconstruction loss functions are employed to train the generator network effectively to predict the desired output gait templates without co-variate objects. A view normalizing conditional GAN is presented in [100] to learn the identity-related representations from multiple views. This model is inspired from Conditional GANs. In [101], a

StarGAN model has been used for multi-view transformation from a single view.

## 2.3 Limitations of Existing Approaches and Scopes for the Thesis

From the literature review, we point out a few limitations of existing gait recognition approaches as pointed out next.

1. Existing pose-based templates use binary silhouettes for feature extraction. The binary silhouettes preserve information about the entire shape of the person and do not put higher emphasis on the silhouette contours which preserve more kinematic or motion information. Since better gait signatures can be obtained from the kinematics of walking, it needs be studied if pose-based feature extraction using silhouette contours over a gait cycle can improve the accuracy of the existing methods.

2. Another limitation of the existing pose-based gait templates is predetermining a fixed set of key poses before feature extraction. Such methods are not effective if the camera frame rates during the training and test video capturing times are different, and also if the walking speeds vary during the two times. It appears that this issue can be resolved if a dictionary of key poses is considered using several key pose sets, each with varying number of key poses.

3. The constraints used for mapping the frames of an input sequence to the key poses in any pose-based gait recognition technique are too strict in the sense that it only allows transition from a state to the same state or to the successive state, i.e., if a frame is mapped to key pose $k$, the subsequent frame can be mapped to either key pose $k$ or $k+1$. However, this assumption is too strict and is not effective if the number of frames in an input sequence is less than the number of key poses in a gait cycle. Relaxing the constraint by allowing transitions from a state to a few subsequent states may help in alleviating this problem to a certain extent, which needs further study.

4. Occlusion is an inevitable occurrence in any real-life surveillance scenario. However, existing GAN-based techniques in gait recognition focus mostly on view-invariant recognition. Focus has not been given to developing neural network-based approaches for occlusion reconstruction in gait recognition, which needs to be studied in the future.

In the thesis, we focus on providing effective solutions to the first three limitations to be described in detail in Chapters 3, 4, and 5 along with experimental results.

## 2.4 Gait Data Sets Used in The Thesis

Following the discussion in Section 1.1 of Chapter 1, a complete gait cycle is required to extract effective gait features. Further, any Computer Vision-based gait recognition method is suitable for application in places where the gait videos are captured from a distance and physical interaction with individual subjects is not possible. Examples of such application sites include airports, railway stations, shopping malls, etc. It may be noted that most subjects entering and leaving these zones have roughly symmetrical gait, i.e., both the two half gait cycles have almost similar appearance. Thus, instead of using a complete gait cycle to extract relevant gait features, a half cycle or multiples of a half gait cycle can also be used to recognize persons with symmetrical gait. However, such an approach will not be effective in identifying people with walking difficulties or lame people who have asymmetric gait. In such situations, information about a complete gait cycle must be utilized to compute the gait features. Since usually in a surveillance zone, majority of the persons will have symmetric gait, we rely on a half gait cycle only to derive their gait signatures. A few people with asymmetric gait, if any, can be manually tracked and identified easily.
In the thesis, we have focused on symmetrical gait only and used four datasets, namely, the CASIA B, CASIA C, TUM-GAID, and OU-ISIR TreadMill Data Set B, for the experiments. In all our experiments, we use the data captured at 90-degree view, i.e., that corresponding to the front-parallel or side-view of walking to prepare both the training and test sets. Also, it may be noted that we consider separate training and test scenarios for each experiment, as shown in Table 2.1.

The CASIA B data consists of 124 subjects, and gait sequences for each subject are captured under the following conditions: (a) six sequences with normal walk (*nm-01* to *nm-06*), (b) two sequences with carrying bag (*bg-01* and *bg-02*), (c) two sequences with wearing coat (*cl-01* and *cl-02*). Further, each gait sequence is captured from multiple viewpoints: 0 to 180 degrees with 18-degree intervals.

With reference to this table, scenario $C_1$ means that both the training and test sequences consist of normal walking videos of individuals without any co-variate objects. On the other hand, scenario $C_4$ implies that the training set is captured with normal walking while the test set is captured with persons carrying a bag. Similarly, for the scenario $C_5$, the training sequences are captured with normal walking while the test sequences are captured with people wearing coats, while each of $C_2$ and $C_3$ consists of the same conditional training and testing sequence, i.e., carrying bag and wearing coat, respectively.

The CASIA C [102] dataset consists of 153 subjects with four different walking conditions for each person captured by an infrared camera: (a) two sequences of slow walking (fs), (b) two sequences of fast walking (fq), (c) two sequences of normal walking with a bag (fb), and (d) four sequences of normal walking (fn). The four sequences with normal walking (fn) condition have been considered as training set, while others form as a testing set. We have formed three combinations of gallery-probe pairs denoted as $A_1$, $A_2$, $A_3$ in Table 2.1 and corresponding to each of these we have also shown the training and test sets.

The TUM-GAID gait database consists of 305 subjects, and for each subject there are six normal walking sequences: two sequences with carrying-bag conditions, and two sequences with people wearing shoes. These are labeled in the data set as $n1$, $n2$, $n3$, $n4$, $n5$, $n6$, $b01$, $b02$, $s01$, $s02$, respectively. The sequences present in the TUM-GAID data set are either from left to right walking direction or from right to left. Since human walking is symmetric in both the left and right limbs, flip, we horizontally flip all the images that correspond to left to right walking direction so that each sequence shows walking from right to left. To test the robustness of the proposed approach for varying training and test set combinations, we consider

Table 2.1: Training-test set combinations used for evaluating our approach

| Scenario | Training Set | Test Set |
|---|---|---|
| | CASIA B Data | |
| $C_1$ | *nm-03,nm-04,nm-05,nm-06* | *nm-01, nm-02* |
| $C_2$ | *bg-02* | *bg-01* |
| $C_3$ | *cl-02* | *cl-01* |
| $C_4$ | *nm-03,nm-04,nm-05,nm-06* | *bg-01, bg-02* |
| $C_5$ | *nm-03,nm-04,nm-05,nm-06* | *cl-01,cl-02* |
| | CASIA C dataset | |
| $A_1$ | *fn00,fn01,fn02,fn03* | *fb00, fb01* |
| $A_2$ | *fn00,fn01,fn02,fn03* | *fq00, fq01* |
| $A_3$ | *fn00,fn01,fn02,fn03* | *fs00, fs01* |
| | TUM-GAID Data | |
| $T_1$ | *n03,n04,n05,n06* | *n01, n02* |
| $T_2$ | *b02* | *b01* |
| $T_3$ | *s02* | *s01* |
| $T_4$ | *n03,n04,n05,n06* | *b01, b02* |
| $T_5$ | *n03,n04,n05,n06* | *s01,s02* |

five evaluation scenarios, namely, $T_1$, $T_2$, $T_3$, $T_4$, and $T_5$, as shown in the same Table 2.1.

The OU-ISIR TreadMill Data Set B [103] consists of gait silhouette sequences from the fronto-parallel view of 68 subjects with certain clothing variations. These variations include half shirts, medical doctor pants, down jackets, baggy pants, parka, half pants, etc., along with their combinations. The instruction files present in this data set specify the gallery and probe sets, as well as the description of the clothing varieties (i.e., co-variate conditions). One sequence corresponding to each subject indexed with ****90*, where *'9'* and *'0'* are the last two digits of the sequence number is considered to form the gallery set for gait recognition, and the rest of the other sequences are considered for evaluating the performance of the different gait recognition algorithms. The above-mentioned training and test conditions have been maintained throughout the thesis, and while discussing any experiment we have stated the scenario corresponding to the training and test sets considered in that experiment, as given in the first column of Table 2.1.

As already discussed in Section 2.3, the gait recognition methods proposed in the thesis focus on improving upon the existing pose-based approaches to gait recognition. Each of our methods have been developed considering the side-view of walking and normal walking pace of each person. However, our methods are robust enough to handle situations where the walking direction deviates slightly from side-view or the person walks with minor speed variations. In the future, our proposed pose-based gait recognition methods can be conveniently extended to handle even more challenging scenarios where the walking speed of the person differs drastically during the training and test data captures, or where there walking direction changes continuously. To train the gait recognition models effectively, we had to choose suitable datasets that bear the above characteristics, and also consist of a sufficient number of sequences that can be split into an adequate number of training and test samples.

Each of the CASIA B, CASIA C, TUM-GAID, and OU-ISIR TreadMill Data Set B has sequences with multiple gait cycles of normal walking from varying viewpoints under relatively uniform walking speed, out of which we specifically consider sequences from the 90 degree view, or the fronto-parallel view. From each of these datasets, a sufficient number of sequences can be extracted to form the gallery set for training gait recognition models and also for evaluation. We have mostly used traditional Machine Learning models which do not require extensive data to get trained and separate gait features computed from two-four gait cycles corresponding to each subject are sufficient to train such models. The approaches described in the thesis are specifically developed to advance state-of-the-art research in gait recognition from the front-parallel field of view. Basically, we have targeted to propose effective algorithms for pose-based gait recognition that can handle minor walking speed variations as well as different co-variate conditions in training and test sequences. The above-mentioned datasets are quite extensive and popularly used in the literature to evaluate algorithms developed for gait recognition. Further, these datasets also fully represent the variety of gait recognition scenarios considered in the thesis, as stated above.

Among the other popular existing gait datasets, the OU-ISIR TreadMillA, GREW, TUM-IITKGP datasets have not been considered in the thesis since these datasets consist of scenarios such as occlusion, multiple views, varying walking speeds, etc., which must be addressed in addition to developing a gait recognition approach.

While the OU-ISIR TreadMillA data has been specifically made to evaluate approaches focusing on cross-speed recognition, the GREW dataset is prepared by capturing natural walking sequences with rich attributes in an unconstrained environment. On the other hand, the TUMIITKGP dataset is aimed at testing occlusion handling algorithms in gait recognition. Existing view-transformation models, cross-speed data tackling modules, or occlusion handling approaches can be suitably fused with our developed methods to tackle the different challenges in gait recognition considered in the above datasets.

## 2.5 Tools and Frameworks

We have mostly used the Python programming language with its functional tools and frameworks like Keras, TensorFlow and Pytorch for implementation of the existing methods as well as our proposed algorithms. We have also used MatLab to carry out some image preprocessing operations and visualization of results.

### 2.5.1 MatLab

MatLab is a high-performance programming language for technical computing that is specially designed for engineers and scientists to analyze and design different kinds of systems [104]. The programming language used in MatLab is a matrix-based language that allows the most natural expressions of computational mathematics [104]. It integrates computation, visualization and programming in an easy-to-use environment where problems and solutions are expressed in familiar mathematical notation. Its specific uses are Mathematics and computation, Algorithm development, Modeling and simulation, Prototyping, Data analysis, Exploration, Visualization, Scientific and Engineering Graphics, Application development, Graphical User Interface building [104].

### 2.5.2 Tensorflow

Tensorflow [105, 106] is a free and open-source software library developed by the Google Brain team in November 2015 for Machine Learning applications. TensorFlow is an end-to-end platform that makes it easy to build and deploy Machine/Deep Learning models [106]. It is an extensive library with several useful functions and can be used for a variety of Machine/Deep Learning related tasks. This library focuses particularly on providing an interface to distribute Deep Learning models. It is a symbolic math library based on Data Flow and Differentiable Programming. Tensorflow is now being used all over the world by various companies and research communities to implement Deep Learning models.

### 2.5.3 Keras

Keras [107] is open-source software released in March 2015 for implementing Deep Learning models in Python which runs on top of the machine learning platform TensorFlow. It was developed with a focus on enabling rapid experimentation with Deep Neural Networks and is user-friendly, extensible, and modular. It is first written and maintained by Google engineer François Chollet who is also the author of the XCeption Deep Neural Network model [108]. Apart from standard Deep Learning flexibility, it also provides a platform for users to distribute their Deep Learning models on smartphones and the web.

### 2.5.4 Pytorch

Pytorch [109] is a free-to-use open Machine Learning library released in September 2016 and developed by Facebook's AI Research lab (FAIR). It is based on Torch which is a scientific computing framework and scripting language based on Lua programming language. Pytorch provides two important features: Tensor computing (like NumPy) with strong acceleration via graphics processing units (GPU) and Deep Neural Networks built on a type-based automatic differentiation system. Deep learning frameworks have often focused on usability or speed, but not both. PyTorch [109, 110] is a machine learning library that shows that these two goals are indeed compatible: it was designed from first principles to support an imperative and Pythonic programming style that supports code as a model, debugging and is in line with other popular scientific computing libraries, while remaining efficient and supporting hardware accelerators such as GPUs.

## 2.6 Summary

In this chapter, we overview the related literature on gait recognition from the traditional hard feature computing algorithms to the modern Deep Learning-based algorithms. Both traditional and modern approaches can be classified as model-based and model-free. The traditional approaches are divided into two groups: the first group includes approaches where features are created by the aggregation of complete gait cycles while the next includes features that are created by dividing the gait cycle into smaller fragments. The modern approaches are also subdivided into two groups: the first group includes approaches where deep-learning based Convolution Neural Networks (CNNs) models are used for classification and/or encode the features while the next includes Generative Adversarial Networks (GANs) that transform the feature from one to another desired form. We also point out the limitations of the existing methods and provide a brief note on the frameworks used for implementation of the different algorithms in the thesis.