The Journal of Supercomputing
https://doi.org/10.1007/s11227-024-06172-z

A survey of appearance-based approaches for human gait recognition: techniques, challenges, and future directions

Pınar Güner Şahan[1] · Suhap Şahin[1] · Fidan Kaya Gülağız[1]

Accepted: 27 April 2024

Abstract
Gait recognition has become an important biometric feature for human identification, in addition to data such as face, iris, and fingerprint. The goal of human gait recognition is to identify people based on walking images. Artificial intelligence technologies have revolutionized the field of gait recognition by enabling computers to automatically learn and extract intricate patterns. These techniques examine video recordings to determine key features in an individual's gait, and these features are used to identify the person. This paper examines the existing appearance-based gait recognition methods that have been published in recent years. The primary objective of this paper is to provide an informative survey of the state-of-the-art in appearance-based gait recognition techniques, highlighting their applications, strengths, and limitations. Through our analysis, we aim to highlight the significant advance that has been made in this field, draw attention to the challenges that have been faced, and identify areas of prospective future research and advances in technology. Furthermore, we comprehensively examine common datasets used in gait recognition research. By analyzing the latest developments in appearance-based gait recognition, our study aims to be a helpful resource for researchers, providing an extensive overview of current methods and guiding future attempts in this dynamic field.

Keywords Gait recognition · Gait datasets · Artificial intelligence · Neural networks

1 Introduction

Gait recognition is a sort of biometric technology that identifies people based on their distinct walking patterns [1]. It evaluates how a person walks by capturing and quantifying numerous gait variables such as step width, stride length and foot angle (the angle between the foot and the horizontal) during heel strike and toe-off (pre-swing). These metrics are used to derive a gait signature for each person that

✉ Pınar Güner Şahan
pinar.guner@kocaeli.edu.tr

[1] Department of Computer Engineering, Kocaeli University, Kocaeli, Turkey

Published online: 15 May 2024

can be compared to a database of recognized signatures to help identify them [2]. The most beneficial advantage of gait as a biometric feature is that it can be used for identifying people at a distance. Furthermore, it does not necessitate the user's participation unlike other features [3]. These advantages make gait useful for video surveillance-based applications. Gait recognition has potential uses in security and surveillance, including the identification of people in crowded public places and the tracking of criminal suspects [4]. It could also have medical uses, such seeing variations in gait patterns that might point to illnesses or injuries [5]. Among the abovementioned advantages, gait recognition performance can be negatively affected by certain factors related to human pose analysis. Human pose analysis in computer vision faces several challenges, including occlusions, changing lighting conditions, and low image quality.

The following steps are often included in a gait recognition system [6]: (1) Data collection. To recognize an individual's gait, it is necessary to collect data about their gait patterns. Many techniques, including video recordings, pressure sensors, floor sensors and motion capture systems, can be used to obtain this data. (2) Feature Extraction. To identify an individual's gait, it is necessary to extract features that are unique to their walking pattern, such as stride length, walking speed and foot angle. (3) Dimension Reduction. In general, features extracted from gait data cannot be used for classification directly because in the feature representation step, the dimensionality of features (the number of features) collected from raw data is higher than the number of samples in the training data. Consequently, it is preferred to use a dimension reduction approach prior to classification. (4) Classification. To identify the individual based on their gait features extracted in the previous step, classification is performed using a machine learning or a deep learning algorithm.

Gait recognition problem approaches in computer vision are generally classified into two categories: model-based and appearance-based (model-free) [7]. Model-based gait recognition approaches utilize mathematical models to represent the walking motion of a person. In this approach, the kinematics of joint angles are modeled when people walk. Appearance-based gait recognition approaches extract features from the visual appearance of a person's walking pattern, such as body shape and limb movements. In this approach, silhouettes are analyzed from a gait sequence that embed both appearance and movement information, ensuring that the analysis encompasses the entire body structure, including key joints, without isolating them [8].

Appearance-based methods do not require extra sensors or subject consent because they depend on visual data obtained from security cameras. This makes them useful for real-world applications. Although model-based methods have benefits like providing detailed motion information and explicitly modeling skeletal systems, they also have disadvantages such as resource-intensive processing requirements or inaccurate key point estimation. Consequently, when compared to appearance-based approaches, they exhibit lower performance in recognition tasks [9–11]. Such reasons have led to the widespread research and establishment of appearance-based methods in the field. They have a solid foundation in the existing literature, with many methods and datasets available. Hence, the purpose of this paper is to survey appearance-based gait recognition methods that rely mostly on deep learning. Although there are many existing surveys [6, 8, 12–15] conducted on gait recognition, it is the first survey paper based only on recent appearance-based gait recognition studies as far as we know. By focusing entirely on appearance-based methods, the paper gives a full and extensive evaluation of many approaches used in gait recognition. This provides for a better understanding of the specific strategies that rely only on visual clues from gait patterns. Detailed information about the existing surveys and the number of references and citations from Web of Science are shown in Table 1.

The paper aims to provide an extensive overview of the appearance-based gait recognition methods. The paper summarizes the important methods and models used in this area, allowing readers to get a deep understanding of some of the most recent advances. The main contributions of this survey are as follows:

- The survey provides a comprehensive and systematic examination of appearance-based gait recognition methods. It analyzes the current literature and provides a comprehensive assessment of the state-of-the-art in this field of gait recognition.
- The survey evaluates the performance of gait recognition techniques. This evaluation provides useful insights for researchers in determining usability of appearance-based gait recognition methods.
- The survey provides a thorough examination of various publicly available datasets used in the literature.
- The survey highlights challenges in gait recognition. It suggests researchers in new and significant directions within this domain by suggesting prospective options for future study.

We employed a review methodology in parallel with these purposes. We first identified potential papers using search engines (e.g., Google Scholar [16]) and online archives (e.g., IEEE Xplore [17], ScienceDirect [18]). Our search string was a combination of different keywords such as “gait recognition”, “deep learning”, “human identification”, and “gait dataset”. We have included search results after 2018 because we want to focus on the studies of recent years. We then excluded the papers that use model-based gait recognition approaches, do not provide a unique solution, use private datasets for performance assessment, or do not evaluate their performance in comparison to the state-of-the-art. Finally, we identified a series of papers that have applied deep learning to gait recognition.

The reminder of this survey is organized as follows. Section 2 introduces the conceptual framework of gait recognition. Gait datasets and evaluation criteria are shown in Sect. 3. Section 4 reviews and compares appearance-based gait recognition approaches published in recent years. Section 5 discusses some challenges and future trends in gait recognition. Section 6 concludes and ends the paper.

Table 1 Recent surveys on gait recognition summarized by year

Reference & Year	Journal	Focused topics	Category	Number of references	Number of citations
[6], 2019	ACM computing surveys	Surveyed deep learning research in gait recognition based on video and sensors	Appearance-based and model-based	195	131
[8], 2020	IET biometrics	Surveyed machine learning and deep learning research in gait recognition	Appearance-based and model-based	103	4
[12], 2021	Archives of computational methods in engineering	Surveyed research in vision- based gait recognition	Appearance-based and model-based	230	23
[13], 2022	IEEE transactions on pattern analysis and machine intelligence	Surveyed recent developments in gait recognition with deep learning	Appearance-based and model-based	257	54
[14], 2023	Multimedia tools and applications	Discussed research in gait recognition with traditional and machine learning classification techniques	Appearance-based and model-based	127	3
[15], 2023	Engineering applications of artificial intelligence	Discussed model-based strategies for gait recognition using deep learning	Model-based	91	–

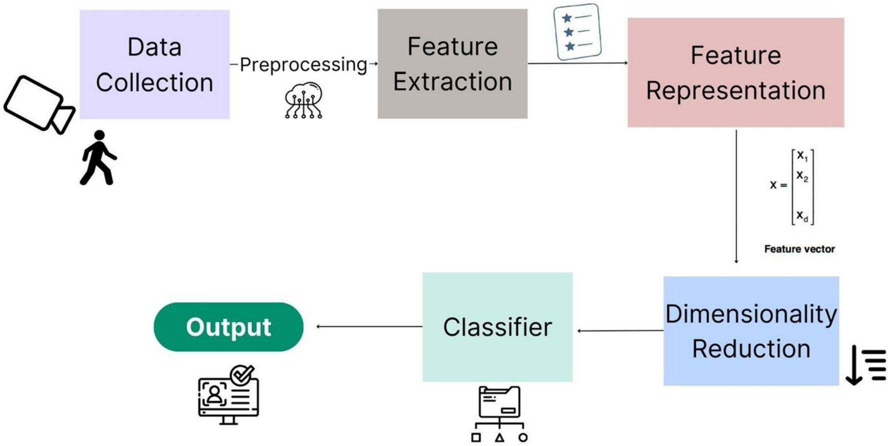

Fig. 1 Conceptual framework of traditional gait recognition

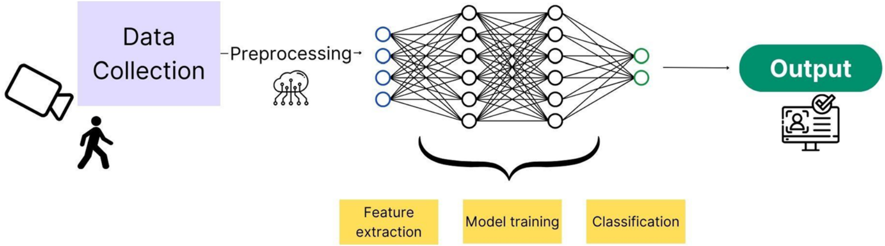

Fig. 2 Deep gait recognition pipeline

2 Gait recognition

In order to help demonstrate a general structure for understanding gait recognition approaches, which will be discussed in the following sections, we present the conceptual framework of gait recognition (Fig. 1). It includes obtaining different types of input data, feature extraction and representation, dimension reduction and classification. Deep pipelines for gait recognition require fewer steps than traditional pipelines, because the deep learning model can perform feature extraction and classification in a single step (Fig. 2). This can improve efficiency and reduce the likelihood of errors introduced by human-defined feature extraction and selection techniques. Deep pipelines, on the other hand, may need additional data and computational resources for training and evaluation, as well as deep learning skills. In this section, we initially described data collection processes conducted independently of the methodologies. Subsequently, we provided an overview of the general framework for gait recognition in both machine learning and deep learning, examining in detail the deep learning techniques employed in the methods examined within this article.

2.1 Data collection

The first stage in the gait recognition framework involves collecting data to identify individual's gait patterns. Gait recognition can be performed using various input data such as RGB image, silhouette, GEI (Gait Energy Image), optical flow image, body skeleton and human mesh acquired by various sensors. In addition, movement data and pressure data from some wearable sensors can also be used for gait recognition. However, since the focus of this study is on vision-based gait recognition, the gait datasets mentioned in this study do not include them. Figure 3 contains examples of different input data types obtained from different gait dataset [19–22].

2.2 Machine learning techniques

2.2.1 Feature extraction and representation

This is the process of extracting features from the data that are most useful for identifying an individual's gait pattern. Feature extraction requires the ability to describe the distinctive characteristics of individuals and be robust to changing conditions. There are two main approaches for gait recognition as already mentioned: (1) model-based, (2) appearance-based. The key distinction between the two approaches

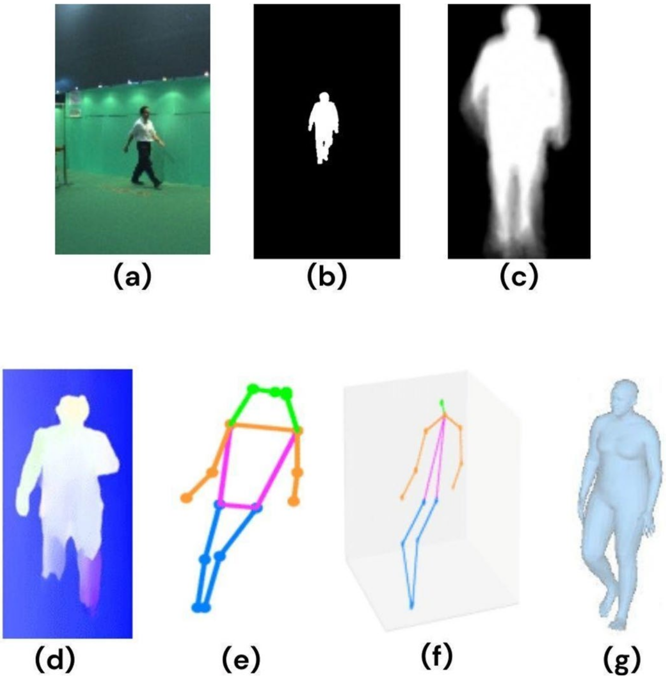

Fig. 3 Examples of input data types for gait recognition. **a** RGB image. **b** Silhouette. **c** GEI. **d** Optical flow. **e** 2D Skeleton. **f** 3D Skeleton. **g** 3D Mesh

is in how the features are extracted and the type of data used for recognition. Model-based gait recognition extracts features from a physical model of the human body that predicts joint angles and trajectories during walking. In appearance-based gait recognition, features are extracted by considering the entire movement pattern of the walking person's body. It handles occlusion better and contains more invariant features [14]. Feature representation for gait recognition transforming raw gait data into a set of features that can be utilized for classification. Appearance-based feature representation methods are statistical methods and spatiotemporal methods. Statistical features include shape (e.g., how high the leg is raised during the walking cycle), motion (e.g., speed) and texture (e.g., variations of clothing and carrying conditions). The spatiotemporal methods gather the motion characteristics and maintain both the spatial aspects (such as shape, distance, and direction) and the temporal aspects (like duration and occurrence time) of gait video sequences [12]. Human movement is usually represented through both spatial and temporal information.

2.2.2 Dimensionality reduction

The major goal of dimensionality reduction is to reduce the dimensionality of the feature vector that represents the gait patterns. Typically, the feature vector is high-dimensional and comprises a huge number of variables. This makes gait recognition methods computationally costly and time-consuming to compute. Dimensionality reduction aims to address this issue by reducing the dimensionality of the feature vector, while preserving essential information. There are different techniques for dimensionality reduction, such as principal component analysis (PCA) and linear discriminant analysis (LDA). These techniques attempt to transform the high-dimensional feature vector into a lower-dimensional space that still captures the important information. The classification algorithm is then fed the resulting lower-dimensional feature vector.

- PCA [23] transforms the feature vector into a set of orthogonal principal components, each of which is a linear combination of the original variables. Most of the information is included in the first few primary components, which are kept, while the other components are disposed.
- LDA [24] seeks to maximize the distance between the means of different classes, while minimizing the variance within each class. It aims to project the feature vector into a lower-dimensional space with the goal of maximizing the separation between the different classes.

2.2.3 Classification

In gait recognition, the classification step refers to the process of giving a label or class to a gait sequence. This stage is critical because it allows the system to recognize and distinguish between different individuals based on their gait patterns. The features selected in the previous steps are used to create a feature vector representing the gait sequence. In this stage, the feature vector is input to a classification algorithm that assigns the gait sequence to a specific class or label. During the classification process, it is learned to recognize patterns in feature vectors associated with certain individuals, and the gait sequence is assigned to the appropriate label.

In this section, it is useful to mention the two modes of biometrics, identification and verification. Person identification involves recognizing an individual from a group of known persons, which can be challenging due to the need to distinguish between highly similar gait patterns. Person verification compares a gait pattern to a single individual's known patterns to confirm or deny their identity. It is less useful in applications, where the identity of the individual is unknown or needs to be determined from a large number of possibilities.

In traditional machine learning approaches, the classification stage consists of applying an algorithm that can distinguish between the different classes (i.e., individuals) based on the features extracted from their gait. Because the process of extracting features is separated from the classification step. The similarity between features is measured by a vector similarity metric such as Euclidean distance, Cosine similarity, Manhattan distance or dynamic time warping (DTW). Euclidean distance measures the straight-line distance between two points in a multi-dimensional space. Instead of measuring distance, cosine similarity measures the cosine of the angle between two vectors. Manhattan distance sums the absolute differences of their cartesian coordinates. DTW defines an optimum path that can transform one signal into another [25, 26]. Siamese networks can also be used in gait recognition applications to learn how to differentiate between inputs, effectively learning a similarity metric. The siamese network can orient the similarity metric to be small for pairs of gait from the same individual and large for pairs from different people [27].

Finally, a label is assigned to each image by a classifier. The algorithm used depends on the type of feature set and the specific requirements of the recognition task (e.g. complexity of the data). Common algorithms used in this context are given below.

2.2.3.1 Support vector machine (SVM) Support vector machine (SVM) is a popular supervised machine learning algorithm used for the classification of gait patterns [28]. The basic idea behind SVM in gait recognition is to find a hyperplane that separates the data points representing the gait patterns of different individuals. The hyperplane is selected in such a way as to maximize the margin, which is the distance between the hyperplane and the closest data points from each class. After the SVM model has been trained, it can be used to classify new gait patterns using the features that were extracted. Depending on which side of the hyperplane the new data point falls, the SVM model will predict the class of the new gait pattern. In this point, it is important to note that SVM is basically a binary classification algorithm, aiming to distinguish between two classes by finding the optimal hyperplane that separates them in the feature space. However, gait recognition often involves identifying individuals from a set of multiple classes, requiring a multiclass classification technique. The one-vs-the-rest strategy is a popular way for adapting SVM for multiclass classification. This involves creating multiple, dedicated SVMs, each trained to distinguish between one of the classes and the sum of all other classes [29]. The authors cover the use of SVMs for automatic recognition of age-related gait changes in [30]. In [31], a gait recognition system is presented based on SVMs and acceleration data.

2.2.3.2 Hidden markov model (HMM)

Hidden markov model (HMM) can be used to represent the temporal properties of gait patterns in gait recognition [32]. The main concept is to describe the sequence of gait features as several states, each representing a different gait pattern. The transition probabilities between the states show the possibility of transitioning gait patterns. Given the current condition, the observation probabilities describe the likelihood of observing a specific gait feature. A training set of gait data is used to estimate the model parameters for an HMM. The model parameters include the transition and observation probability. The authors in [33] describe a potential approach for identifying people by their gait that involves modeling the dynamic silhouettes of a human body using a HMM. The research in [34] suggests utilizing a HMM to assess gait phases to examine a patient's gait for appropriate rehabilitation treatment.

2.3 Deep learning techniques

The key concept of the gait recognition using deep learning is automatically learning to identify individuals based on their unique gait patterns directly from the data. This ability brings the advantages of making them robust to variations in input data for the gait recognition task. The layered architecture of deep learning facilitates the incremental extraction of complex features from unprocessed data, eliminating the necessity for manually identifying important features, a process often demanding specialized expertise. This becomes especially significant in the context of analyzing gait patterns, where the automated identification of distinguishing features is crucial [35]. The automatic feature extraction concept in deep learning can include extracting and learning spatial features from individual frames and temporal features across sequences of frames.

When we look at the dimensionality reduction process in the context of deep learning, it is crucial to simplify models, increase their efficiency and reduce overfitting. Some prominent dimensionality reduction techniques used in deep learning are described below.

- Pooling is often applied to a set of values arranged in a grid-like structure, such as the feature maps produced by a convolutional neural network (CNN) in computer vision applications [36]. In order to produce a single output value, the pooling process divides the grid into non-overlapping or overlapping sub-regions and applies an aggregate function to the values within each subregion. The information stored within the subregion is then summarized using this output value. Maximum pooling and average pooling are the two most often used pooling functions. Max pooling includes taking the max value within each subregion, and average pooling involves taking the average value [36].
- Autoencoders (AEs) are neural networks designed to learn efficient representations (encodings) of the input data, typically for the purpose of dimensionality reduction. An autoencoder is composed of an encoder that reduces the input dimensions and a decoder that reconstructs the input data from the

reduced representation. The middle layer, also known as the code layer, has a lower dimensionality and acts as a reduced representation of the input data [37].

- Variational autoencoders (VAEs) are generative models that learn a latent variable model for the input data. They are similar to autoencoders but are intended to produce a probabilistic representation of the input data. Compared to the input space, the latent space learned by VAEs is generally significantly lower dimensionality [38].

Deep learning models offer an end-to-end learning approach, which means that the raw input is fed into the deep learning model, which then outputs the classification result directly. This smooth process optimizes the pipeline, while improving the model's ability to learn complex patterns. In the classification stage, the deep learning model uses the learned features to classify the gait data into predetermined classes, with each class representing an individual. This could be done through an activation function (e.g., softmax) in the output layer. The model is trained using a labeled dataset, where each gait sequence is associated with a specific individual. The training involves adjusting the model's weights via back propagation based on the difference between the predicted and actual labels, minimizing a loss function to improve classification accuracy over time.

2.3.1 Convolutional neural networks (CNN)

Convolutional neural network (CNN) [39] is a type of neural network that is commonly used in gait recognition. A CNN consists of many layers of interconnected nodes, such as convolutional layers, pooling layers, and fully connected layers. The convolutional layers are responsible for detecting and extracting features from the input data. The pooling layers then decimate the feature maps created by the convolutional layers, reducing the dimensionality of the data, while preserving the most critical information. Eventually, the fully connected layers classify the output from the previous layers into different gait patterns or individuals. A CNN can be trained to recognize the unique gait patterns of individuals using a huge dataset of labeled walking sequences. The network learns to extract relevant features from the input data and utilize them to make accurate predictions about the identity of the individual during training. Because CNN models are highly effective at learning spatial features, they are frequently trained using image data for gait recognition tasks. In these tasks, the CNN architecture allows the models to maintain the spatial or positional connections among the input data points. Besides that, CNN can be adapted to extract temporal features effectively by employing kernels that move in one direction across the temporal dimension of the data. This approach is typically realized through the use of one-dimensional (1D) convolutional neural networks (1D-CNNs), where the convolution operation is applied along the time axis of the input data [40].

Most of the studies analyzed in this survey (please check Table 3) used these properties of Convolutional Neural Networks (CNNs) for gait recognition.

2.3.2 Recurrent neural networks (RNN)

Recurrent neural networks (RNNs) perform well at processing sequential data, making them an ideal tool for gait recognition tasks that require evaluating the temporal dynamics of human walking patterns. RNNs are designed to recognize patterns in data sequences by storing previous inputs in their internal state (hidden layers), which is updated when new data points are processed. An RNN layer typically comprises multiple neurons that exhibit recurrent behavior, enabling the layer to accept a sequence of inputs and, in turn, output a sequence [41]. Their ability to learn from the sequence and duration of movement patterns allows for a detailed classification of distinct gait patterns.

However, traditional RNNs often struggle with the vanishing gradient problem when learning long sequences, making it hard to capture very long-term dependencies [42]. Solutions like long short-term memory (LSTM) [43] and gated recurrent units (GRU) [44] have been developed to address this issue. LSTM is a form of RNN designed to capture long-term dependencies in sequence data by using a set of gates to control the flow of information [43]. GRUs are a simplified version of LSTMs that try to capture dependencies in sequential data but use a more compact design that merges the forget and input gates into a single update gate, reducing complexity.

2.3.3 Generative adversarial networks (GAN)

Generative adversarial networks (GANs), offer novel approaches to gait recognition among other applications and can be used to generate synthetic gait data, improve feature extraction, and enhance the robustness of gait recognition approaches under various conditions. A GAN consists of two neural networks, the generator and the discriminator, which are trained simultaneously through adversarial processes [45]. GANs are especially useful in cross-view gait recognition, where the goal is to recognize individuals from different viewing angles. GANs can be used to produce gait data from unobserved angles, allowing the training of flexible gait recognition models that perform well from multiple perspectives. Applying GANs to gait recognition brings various challenges, including training stability and convergence concerns, which might result in low-quality or unrealistic synthetic data [46].

2.3.4 3D Convolutional neural networks (3D CNN)

3D convolutional neural networks (3D CNNs) enhance the capabilities of conventional CNNs by directly processing volumetric data, enabling them to collect both spatial and temporal information. This makes 3D CNNs ideal for video analysis applications such as gait recognition, which require an in-depth understanding of movement dynamics across time. 3D CNNs examine a sequence of frames as a single input, in contrast to 2D CNNs which process individual frames and may require additional mechanisms to integrate temporal information. This allows them to extract features that capture both the shape and the movement of the subject [47]. This means that 3D CNNs can recognize distinct patterns in the way a person walks by considering several frames together. Despite its benefits, 3D CNNs have several challenges, including the high computational cost of processing 3D data and the requirement for huge labeled datasets to adequately train the models [48].

2.3.5 Hybrid models

Hybrid models in gait recognition use the benefits of a number of neural networks to improve the accuracy and robustness of gait recognition systems. Compared to a single model employed on its own, these models are more suitable for capturing the complex spatial and temporal features of the human gait. Combining CNNs with RNNs or LSTM networks is a popular strategy. CNNs are used to extract spatial features such as the shape and posture of a walking person from individual frames, while RNNs or LSTMs are used to analyze temporal sequences by capturing gait dynamics of gait over time [49]. This hybrid strategy integrates the CNN's ability to recognize spatial patterns with the RNN/LSTM's ability to understand temporal associations, resulting in more accurate gait recognition.

3 Datasets and evaluation criteria

3.1 Datasets

Datasets are crucial for the gait recognition process because they are used to evaluate methods. Over the years, several gait recognition datasets have been developed to aid research and development in this field. Some publicly available gait datasets that are commonly used for gait recognition are shown in Table 2. This table provides an overview of the key features of these gait datasets. These features comprise the number of subjects (classes), the number of sequences, the number of cameras, resolution of the image, the frame rate captured per second, the number of training and testing subjects, the environment conditions, the type of data and variations in the appearance of the individual.

The CMU body movement (MoBo) database contains high-quality video recordings from multiple angles of subjects walking on a treadmill. This data collection, which includes different walking speeds and conditions of 25 subjects, provides a solid resource for analyzing and recognizing individual gait patterns [50]. The SOTON dataset [51] is a collection of gait videos acquired from a multi-camera system that captures people walking along a straight path. The dataset includes videos from 115 subjects and in indoor and outdoor environments. The CASIA-A [52] dataset is another dataset for gait recognition research, containing data from 20 subjects. The USF HumanID dataset [1] includes gait videos from 122 subjects, with variations in shoes, carrying briefcase, and with acquisition times. The videos were captured using two cameras. The CASIA-B dataset [53] is a large dataset containing gait cycles from 124 subjects, captured under various conditions such as normal walking (NM), different clothing (CL), and carrying a bag (BG). The CASIA-C dataset [54] includes gait videos 153 subjects walking in a cross-view scenario. The dataset also includes challenging variations such as three different walking

Table 2 Publicly available gait datasets that are commonly used in the literature

Dataset	Year	Number of subjects/ classes	Number of sequences	Number of cameras/camera angles	Image resolution	Frame rate captured per second (fps)	Number of training & testing subjects	Environment	Data type	Variations
CMU MoBo [50]	2001	25	600	6	640×480	30	not specified	Indoor	RGB, Silhouette	Slow walking, fast walking, incline walking and carrying condition
SOTON [51]	2002	115	2128	2	not specified	25	not specified	Indoor, outdoor	RGB, Silhouette	Normal walking
CASIA-A [52]	2003	20	240	3	352×240	25	not specified	Outdoor	RGB	Normal walking
USF HumanID [1]	2005	122	1870	2	720×480	30	not specified	Outdoor	RGB	Change in walking surface, carrying condition, time interval
CASIA-B [53]	2006	124	13,680	11	320×240	25	not specified	Indoor	RGB, Silhouette	Normal walking, clothing, and carrying condition
CASIA-C [54]	2006	153	1530	1	320×240	25	3 & 150	Outdoor	Infrared, Silhouette	Normal walking, slow walking, fast walking, and carrying condition

Table 2 (continued)

Dataset	Year	Number of subjects/classes	Number of sequences	Number of cameras/camera angles	Image resolution	Frame rate captured per second (fps)	Number of training & testing subjects	Environment	Data type	Variations
OU-ISIR Treadmill dataset [55] – Speed	2012	34	306	4	88 × 128	60	20 & 14	Indoor	Silhouette	Nine walking speeds
OU-ISIR Treadmill dataset [55]–Clothes	2012	68	2746	4	88 × 128	60	20 & 48	Indoor	Silhouette	32 different clothing combinations
OU-ISIR Large Population (OU-LP) [19]	2012	4,007 (V1), 4,016 (V2)	31,368	4	640 × 480	30	not specified	Outdoor	Silhouette	Normal walking
TUM GAID [56]	2012	305	3370	1	640 × 480	30	100 & 155 (50 subjects for validation)	Indoor	RGB, Depth, Audio	Time interval, carrying and clothing condition
OU-LP Bag [57]	2017	62,528	187,584	1	1280 × 980	25	29,097 & 29,102 for 58,199 subjects	Indoor	Silhouette, GEI	Carrying condition
OU-LP Age [58]	2017	63,846	63,846	1	640 × 480	30	31,923 & 31,923	Indoor	Silhouette, GEI	Different ages ranging from 2 to 90
OU-MVLP [20]	2018	10,307	259,013	14	1280 × 980	25	5,153 & 5,154	Indoor	Silhouette, GEI	Normal walking

Table 2 (continued)

Dataset	Year	Number of subjects/classes	Number of sequences	Number of cameras/camera angles	Image resolution	Frame rate captured per second (fps)	Number of training & testing subjects	Environment	Data type	Variations
CASIA-E [59]	2020	1,014	778,752	26	1920×1080	25	500 &400 (114 subjects for validation)	Multiple outdoor	Silhouette	Carrying and clothing condition, walking style, and soft biometric features
OU-MVLP Pose[60]	2020	10,307	259,013	14	1280×980	25	5153 & 5,154	Indoor	2D Skeleton	Normal walking
VersatileGait [61]	2021	11,000	1,000,000	33	280×200	not specified	not specified	Synthetic	Silhouette	Carrying and clothing condition
ReSGait [62]	2021	172	870	1	64×44	44 to 300	86 & 86	Indoor	Skeleton, Silhouette	Carrying and clothing condition, and phone usage
GREW [21]	2021	26,345	128,671	882	64×44	30	20,000 & 6000 (345 subjects for validation)	Wild	Silhouette, Optical flow, 2D and 3D Skeletons	Carrying and clothing condition, change in walking surface, different ages, and different walking speeds
OU-MVLP Mesh [63]	2022	10,307	not specified	14	1280×980	25	5153 & 5154	Indoor	3D Human Mesh	Normal walking

Table 2 (continued)

Dataset	Year	Number of subjects/classes	Number of sequences	Number of cameras/camera angles	Image resolution	Frame rate captured per second (fps)	Number of training & testing subjects	Environment	Data type	Variations
Gait3D [22]	2022	4,000	25,309	39	1920×1080	25	3000 & 1000	Wild	RGB, Silhouette, 2D and 3D Skeletons, and 3D Meshes and SMPL	Different walking speeds, clothing condition

speeds (Normal walking—NM, slow walking—SW, fast walking—FW), and carrying a bag (BW). OU-ISIR Treadmill dataset [55] is a gait dataset that was collected at the University of Osaka in Japan. The speed dataset includes gait videos of 34 subjects walking on a treadmill at nine different speeds. The clothes dataset includes gait videos of 68 subjects with different clothes up to 32 options. OU-LP dataset [19] is a large-scale gait database that includes gait sequences of 4,007 subjects (in version 1). The gait sequences were collected using four camera angles. The OU-LP dataset includes a large number of participants with a wide range of gait patterns, all captured in a controlled environment to minimize external variables such as lighting and background variations, and subjects are typically dressed uniformly to reduce the impact of clothing variations on gait recognition. The TUM GAID dataset [56] incorporates audio, image (video), and depth data, providing a comprehensive set of modalities for gait analysis. It consists of 305 subjects and the 32 subjects in the subset enable study in clothing and time invariant gait recognition. The OU-LP Bag dataset [57] includes gait sequences of 62,528 subjects carrying an object, while walking. The dataset includes variations in types of carried objects. OU-LP Age dataset [58] includes gait sequences of 63,846 subjects at different ages. The OU-MVLP (Multi-View Large Population) dataset [20] is another large-scale gait database that includes gait sequences of 10,307 subjects captured from 14 different views ranging from 0 to 90, and 180 to 270. CASIA-E dataset [59] includes silhouettes from 1,014 subjects and variations in walking style, carrying objects, and wearing different clothing. The OU-MVLP Pose dataset [60] was created by taking the RGB images from the OU-MVLP and extracting pose skeleton sequences from them. VersatileGait [61] is a large-scale synthetic gait dataset produced using a gaming engine. The dataset includes nearly one million silhouette sequences of 11,000 participants, each with fine-grained features. This dataset intends to solve the shortcomings of existing real-world gait datasets, which frequently have small sample sizes and simple scenarios. The ReSGait dataset [62] consists of 172 subjects and 870 video clips that were collected over a period of 15 months. The dataset include gender, clothing and carrying conditions, and use of mobile phones. The GREW dataset [21] is known as the first extensive dataset for gait recognition in the wild. The dataset consists of gait sequences from 26,345 subjects collected from 882 cameras. Also, the dataset includes some information such as gender, age group, carrying and clothing condition. OU-MVLP Mesh [63] dataset was built upon OU-MVLP and it examines informative 3D human mesh model using parametric pose and shape features (i.e., SMPL). The Gait3D dataset [22] is a large-scale gait recognition dataset based on 3D representation. It contains 4,000 subjects taken from 39 cameras in the wild. The dataset also includes variations such as different walking speeds and clothing conditions.

3.2 Evaluation criteria

To evaluate gait recognition methods using different databases, there are two types of evaluation protocols that have been frequently used: subject-dependent and subject-independent [13]. Subject-dependent protocols involve training and testing the gait recognition method using the same set of subjects. In this scenario, the solution is trained on a subset of the gait data and then tested on the remaining data for each subject. The goal of this approach is to find out how well the method recognizes the gait patterns of an individual in the context of intraclass variations such as different walking speeds, clothing, and carrying conditions. Subject-independent protocols, on the other hand, involve training the gait recognition method on a set of subjects and testing it on a different set of subjects. This approach intended to evaluate how well the method generalizes to new individuals who were not included in the training data. The test data are subdivided into gallery and probe sets, and the learned model on the separate training subjects are utilized to extract features from these subsets. Overall, a classifier is used to compare the probe and gallery data to determine the most related gait patterns and categorize them as belonging to the same identity.

The gait recognition methods studied in this paper use the cumulative match characteristic (CMC) as an evaluation criterion. The CMC curve is a performance evaluation metric commonly used in biometrics and computer vision, particularly in tasks related to recognition systems such as face recognition, fingerprint identification, and gait recognition. It helps to assess the accuracy of identification systems. The CMC curve is essentially a rank-based metric and represents the probability that a query identity appears within the top K ranks of a sorted list of candidates generated by the system [64]. The CMC curve, despite being a widely used metric for measuring the precision of identification systems, has its limitations. It ignores the overall accuracy and confidence of matches as a result of focusing only on ranking performance. It provides limited insights on system performance across different circumstances, which might overlook the complex nature of real-world applications [65]. However, some studies reviewed in this survey focus solely on reporting rank-1 recognition accuracy that is the first point on a CMC curve. Consequently, in the subsequent sections, we will also use the rank 1 accuracy as our primary evaluation criteria.

4 Appearance-based gait recognition approaches

In the last decades, numerous approaches to gait recognition have been developed. We mentioned that these approaches are divided into model-based and appearance-based approaches. This section reviews the appearance-based gait recognition approaches published in recent years. Appearance-based techniques consider the complete human body structure or motion. This approach extracts gait features from human walking sequences, focusing on the silhouette shape and dynamic information needed for pattern matching.

Numerous methodologies exist in the literature, yet this section cannot cover all these methods. It does discuss the details of the state-of-the-art techniques. Table 3 summarizes the reviewed appearance-based gait recognition approaches arranged by the dates of publication. Section 4.1 contains a rigorous comparison of these approaches.

Table 3 Summary of recent gait recognition studies

Method	Reference	Gait feature	Dataset	Number of classes	Computation time	Resolution	Best accuracy reported (Rank-1) (%)	Methodology	Number of layers	Activation function	Code availability
ACLGait [66]	Zhang et al. (2019)	GEI, Silhouette	CASIA-B (local + temporal)	124 (training set: 74, testing set: 50)	1.5 min	not specified	96.0	CNN + LSTM	3 conv, 3 pooling layers	not specified	No
			OU-LP (local + temporal)	4007 (training set: 3075, testing set: 768)	not specified	not specified	99.3				
			OU-MVLP (GEI)	10,307 (training set: 5153, testing set: 5154)	45 min	not specified	89.0				
Gaitpart [67]	Fan et al. [67]	Silhouette	CASIA-B	124 (training set: 74, testing set: 50)	1.16 s per iteration	64 × 44	96.2	CNN + Attention + HP	6 conv, 2 pooling layers	Leaky ReLU	Yes
			OU-MVLP	10,307 (training set: 5153, testing set: 5154)	not specified	64 × 44	88.7				

Table 3 (continued)

Method	Reference	Gait feature	Dataset	Number of classes	Computation time	Resolution	Best accuracy reported (Rank-1) (%)	Methodology	Number of layers	Activation function	Code availability
GLN [68]	Hou et al. [68]	Silhouette	CASIA-B	124 (training set: 74, testing set: 50)	1.01 s per iteration	128×88	96.8	CNN + Lateral Feature Pyramid + HPM	3 conv, 2 pooling layers	ReLU	No
			OU-MVLP	10,307 (training set: 5153, testing set: 5154)	not specified	64×44	89.1				
SRN [11]	Hou et al. [11]	Silhouette	CASIA-B	124 (training set: 74, testing set: 50)	0.97 s per iteration	128×88	97.1	RNN + Dual Feature Pyramid + HPM	not specified	Leaky ReLU	No
			OU-MVLP	10,307 (training set: 5153, testing set: 5154)	not specified	64×44	89.1				
Multi-view gait recognition system [69]	Gul et al. [69]	GEI	CASIA-B	124 (divided 70: 30 for training and testing)	not specified	not specified	98.3	3D CNN	4 conv, 2 pooling layers	ReLU	No
			OU-LP	4016	not specified	128×88	93.1				

Table 3 (continued)

Method	Reference	Gait feature	Dataset	Number of classes	Computation time	Resolution	Best accuracy reported (Rank-1) (%)	Methodology	Number of layers	Activation function	Code availability
3DLocal [70]	Huang et al. [70]	Silhouette	CASIA-B	124 (training set: 74, testing set: 50)	not specified	64 × 44	97.5	3D local CNN + Set Pooling	6 conv, 2 pooling layers	ReLU	Yes
						128 × 88	98.3				
			OU-MVLP	10,307 (training set: 5153, testing set: 5154)	not specified	not specified	90.9				

Table 3 (continued)

Method	Reference	Gait feature	Dataset	Number of classes	Computation time	Resolution	Best accuracy reported (Rank-1) (%)	Methodology	Number of layers	Activation function	Code availability
CSTL [71]	Huang et al. [70]	Silhouette	CASIA-B	124 (training set: 74, testing set: 50)	not specified	64 × 44	98.5	CNN + MSTE + ATA	4 conv, 1 pooling layers	Leaky ReLU	Yes
						128 × 88	98.7				
			OU-MVLP	10,307 (training set: 5,153, testing set: 5,154)	not specified	not specified	91.0				
			GREW	26,345(training set: 20,000, validation set: 6000, testing set: 345)	not specified	not specified	50.6				

Table 3 (continued)

Method	Reference	Gait feature	Dataset	Number of classes	Computation time	Resolution	Best accuracy reported (Rank-1) (%)	Methodology	Number of layers	Activation function	Code availability
GQAN [72]	Hou et al. [72]	Silhouette	CASIA-B	124 (training set: 74, testing set: 50)	not specified	128×88	98.5	GQAN + FQBlock + PQBlock	not specified	ReLU + Sigmoid	No
			OU-MVLP	10,307 (training set: 5153, testing set: 5154)	not specified	64×44	89.7				
GaitSlice [10]	Li et al. [10]	Silhouette	CASIA-B	124 (training set: 74, testing set: 50)	SFE: 902.59 ms STSR: 20.94 ms FC: 223.4 ms (50 frame sequence)	64×44	96.7	CNN + SFE + SHP + STSR	2 conv, 4 focal conv, 2 pooling layers	Sigmoid	No
			OU-MVLP	10,307 (training set: 5153, testing set: 5154)	not specified	64×44	89.3				

Table 3 (continued)

Method	Reference	Gait feature	Dataset	Number of classes	Computation time	Resolution	Best accuracy reported (Rank-1) (%)	Methodology	Number of layers	Activation function	Code availability
GaitSet [73]	Chao et al. (2022)	Silhouette	CASIA-B	124 (training set: 74, testing set: 50)	0.96 s per iteration	64×44	96.1	CNN + Set Pooling + HPM	4 conv, 2 pooling layers	ReLU	Yes
			OU-MVLP	10,307 (training set: 5153, testing set: 5154)	not specified	64×44	87.9				
Light-weight-Deep [74]	Khan et al. [74]	Video Frame	CASIA-B	124 (divided 50: 50 for training and testing)	104.68 s (average classification time)	not specified	96.8	CNN + TL + DCA + ELM	16 conv, 5 pooling layers	ReLU 6	No
			TUM GAID	305 (divided 50: 50 for training and testing)	86.43 s (average classification time)	not specified	98.6				

Table 3 (continued)

Method	Reference	Gait feature	Dataset	Number of classes	Computation time	Resolution	Best accuracy reported (Rank-1) (%)	Methodology	Number of layers	Activation function	Code availability
STAR [9]	Huang et al. [9]	Silhouette	CASIA-B	124 (training set: 74, testing set: 50)	4.17 ms (30 frame sequence)	64×44	97.3	CNN + MDFG + STAI	4 conv, 1 pooling layers	Sigmoid	No
			OU-MVLP	10,307 (training set: 5153, testing set: 5154)	not specified	not specified	89.7				
GaitAMR [75]	Chen et al. [75]	Silhouette	CASIA-B	124 (training set: 74, testing set: 50)	not specified	64×44 128×88	98.1 98.6	CNN + MSFE	4 conv, 1 pooling layers	Sigmoid	No
			OU-MVLP	10,307 (training set: 5153, testing set: 5154)	not specified	not specified	88.3				

In [66] a new loss function for cross-view gait recognition called angle center loss (ACL) and a method for learning spatial-temporal features that combines learned horizontal partition and an LSTM attention model are proposed. Gait silhouettes are divided into four horizontal parts and each part is fed into a separate CNN. Attention weights for each part are used to average frame-level features. During training, various weighted features are fed into various loss functions, but during testing, the weighted features for each part are concatenated to form a feature vector. For both verification and identification tasks, cosine similarities are determined between these feature vectors. For each local part, several independent CNNs are used to learn the local gait features, and a simplified spatial transformer network is used to localize the informative parts. An LSTM-based temporal attention model is used to capture the temporal features. The proposed method is evaluated using silhouettes on three gait recognition datasets (CASIA-B, OULP, and OUMVLP with the accuracy of 96.0%, 99.3%, 89.0% respectively).

Fan et al. [67] introduces a deep learning-based solution for gait recognition (Gait Part) which recognizes people based on their walking patterns. The method uses a temporal part-based architecture consisting of Frame-level part feature extractor (FPFE) and micro-motion capture module (MCM), two separate components. FPFE aims to improve fine-grained learning of part-level features and while attention based MCM aims to derive local short-range spatiotemporal expressions. Experiments are performed on the CASIA-B and OUMVLP datasets and the averaged rank-1 accuracies of the method are 96.2% and 88.7% respectively.

The research in [68] proposes the Gait Lateral Network (GLN), a new network for learning discriminative and compact representations from silhouettes of gait sequences. For correct recognition, GLN takes advantage of the intrinsic feature pyramid in deep CNNs to extract discriminative features and lateral connections to integrate silhouette-level and set-level features. Furthermore, GLN has a Compact Block to considerably reduce the dimension of the gait representations, while maintaining accuracy. The experiments are conducted on CASIA-B and OUMVLP datasets with the accuracy of 96.8% and 89.1% respectively.

In [11] a Set Residual Network (SRN) is presented for silhouette-based gait recognition. It has a fundamental block named Set Residual Block (SRBlock) which builds the framework for feature learning from silhouettes. The SR Block is divided into two parallel branches: the silhouette-branch (learn features from each silhouette individually) and the set-branch (learn features from all silhouettes collectively). The features retrieved from the two branches are concatenated using a residual connection and Leaky ReLU. The paper also presents a Dual Feature Pyramid approach for learning more robust part representations for gait recognition using shallow layer features. The proposed SRN is tested on the CASIA-B (accuracy is 97.1%) and OUMVLP (accuracy is 89.1%) datasets.

In [69], a novel approach is proposed that uses 3D convolutional deep neural network (3D CNN) to extract the spatiotemporal features of a gait sequence, while adopting a holistic approach using GEIs. This network is made up of two sets of convolutional layers, each of which is succeeded by a pooling layer, followed by batch normalization, and two fully connected layers. The proposed model is evaluated on the CASIA-B and OULP datasets and to enhance performance, optimization techniques are applied. The best accuracy reported for CASIA-B dataset is 98.3% and OULP dataset 93.1%.

The research in [70], a unique approach for gait recognition, introduces the use of 3D local convolutional neural networks (CNNs) as building blocks. This block enables the retrieval of local 3D volumes sequentially with adaptable spatial and temporal scales, locations, and lengths. Location, sampling, feature extraction, and fusion modules make up the network. Additionally, a framework for interacting with and enhancing global and local 3D volume information in any layer of 3D CNNs is presented in the paper. The proposed approach evaluated on CASIA-B (accuracies are 97.5% and 98.3% for the resolution of 64×44 and 128×88 respectively) and OUMVLP datasets (accuracy is 90.9%).

The authors in [71] propose a method for gait recognition using a context-sensitive temporal feature learning (CSTL) network and salient spatial feature learning (SSFL) module. The authors highlight that by focusing on various temporal sequences with varying time scales, humans may distinguish between different gaits. The CSTL network uses relation modeling to evaluate the importance of multi-scale features, increasing the more significant scale and suppressing the less important one. The SSFL module solves the misalignment problem induced by temporal operations by selecting discriminative spatial hints throughout the sequence. The suggested method combines adaptive temporal learning with salient spatial mining. The experiments are conducted on three datasets: CASIA-B (accuracies are 98.5% and 98.7% for the resolution of 64×44 and 128×88 respectively), OUMVLP (accuracy is 91.0%) and GREW (accuracy is 50.6%). Although CSTL achieves Rank-1 scores more than 90% on both CASIA-B and OU-MVLP datasets it achieves a 50.6% success rate in recognizing sequences on the GREW dataset. GREW is an unconstrained benchmark for gait recognition, aiming to better simulate real-world conditions than its predecessors, such as CASIA-B and OU-MVLP. The significant variation in performance due to the GREW dataset's inherently challenging conditions. Unlike CASIA-B and OU-MVLP, which are partially controlled environments with limited variations, GREW considers a wider range of factors, such as different views, significant differences in clothing, and the presence of objects held by participants [21]. These factors provide an amount of complexity and unpredictability that better captures real-world circumstances, but they also present more difficulties for gait recognition systems.

The authors in [72] describe a method for gait recognition named gait quality aware network (GQAN). It directly evaluates the quality of each silhouette and each part, and it is made up of two blocks: the frame quality block (FQBlock) and the part quality block (PQBlock). FQBlock adjusts the features of every silhouette separately and combines the scores of all channels to generate a frame quality measure. Meanwhile, PQBlock calculates the weighted distance between the probe and gallery by estimating a score for each part. GQAN can be trained using only identity annotations at the sequence level by using a loss function called Part Quality Loss (PQLoss). CASIA-B and OUMVLP datasets are used to evaluate the proposed network model and the best accuracy reported is 98.5% and 89.7%, respectively.

GaitSlice proposed in [10] is a unique gait recognition model and it enhances recognition accuracy by refining spatial and temporal details of each portion of the human body. The model has slice extraction device (SED) and residual frame attention mechanism (RFAM) modules. SED divides the body into parts and connects features of neighboring body parts from head to toe, and for each body component, RFAM collects and emphasizes the significant frames of sequences. The GaitSlice model combines RFAMs that run in parallel with interrelated slice features in order to allow for flexible selection of the key frames of each body part. The model is tested on two gait recognition datasets: CASIA-B (accuracy is 96.2%) and OUMVLP (accuracy is 89.3%).

GaitSet proposed in [73] considers gait as a set of gait silhouettes and uses a deep learning model to recognize gaits. The paper emphasizes that the sequence of poses during a walking period is not the most important information for differentiating individuals, since the pattern of the sequence is universal. The GaitSet model extracts frame-level information from each silhouette using a CNN, then combines these features into a single set-level feature using Set Pooling. Using Horizontal pyramid mapping, the set-level feature is transformed into a space with more differentiation ability. The experiments are conducted on CASIA-B and OUMVLP datasets with the accuracy of 96.1% and 87.9% respectively.

The research in [74] proposes a sequential lightweight deep learning framework for gait recognition. The researchers modify two pre-existing deep learning models (VGG-19 and MobileNet-V2) and train them using transfer learning. Then, feature engineering is conducted on the VGG-19 and MobileNet-V2. Finally, using discriminant correlation analysis (DCA), the resulting features were merged. In order to select optimum features, a modified moth-flame optimization algorithm is proposed. The chosen features are then categorized using an extreme learning machine (ELM). The proposed method evaluated on CASIA-B (accuracy is 91.2%) and TUM-GAID datasets (accuracy is 98.6%).

STAR (Spatio-Temporal Augmented Relation Network) introduced in [9] is a novel approach for gait recognition. Multi-branch diverse-region feature generator (MDFG) and spatiotemporal augmented interactor (STAI) are the two modules that make up the STAR. The MDFG has the capability to identify body features within separate regions that do not overlap, while the STAI, uses the connections of these regions within a frame and across various frames to create intra- and inter-relation models. The introduced approach evaluated on CASIA-B and OUMVLP datasets and the best accuracy reported is 97.3% and 89.7%, respectively.

In [75], GaitAMR is offered as a method for extracting discriminative subject features for gait recognition. GaitAMR uses a holistic and partial temporal aggregation technique that collects global and local body movement parameters. It is composed of four primary parts: a baseline, spatial extraction, temporal extraction, and view assessment. The baseline part uses silhouette information to convert gait samples into features. Then, a multi-scale feature extractor processes the features to provide richer motion data. The remaining sections analyze the features further to extract relevant information, solving appearance occlusion and silhouette misalignment challenges. After all the features from different domains have been combined, they are sent to the classification layer for recognition. The proposed method evaluated on CASIA-B (accuracies are 98.1% and 98.6% for the resolution of 64×44 and 128×88 , respectively) and OUMVLP datasets (accuracy is 88.3%).

4.1 Comparison of different approaches

Considering the methods reviewed in this survey, it has been observed that the CASIA-B and OUMVLP datasets are the preferred primary datasets for evaluating appearance-based gait recognition applications.

In this section, detailed explanations are provided on how each method differs from previous ones and how these differences have led to success compared to earlier methods. Since, the CASIA-B dataset is commonly used across all examined studies, the papers are organized in ascending order based on the Rank-1 accuracy rates achieved on this dataset. The accuracy rates for the CASIA-B dataset in Table 3, correspond to the accuracy under normal walking conditions.

The study conducted in [66], a loss function is proposed that enhances robustness, especially when different spatial-temporal features are used. Loss functions in deep learning have the advantage of learning discriminative features or metrics. Prior to this study, gait recognition methods typically employed classical loss functions like softmax. The loss function proposed in this paper has been shown to improve performance when compared to previous works. Additionally, the study combines different parts of silhouettes with certain weight values. It is stated that the features obtained in this way have increased the accuracy of the model, but this process has brought along computational cost and feature dimension problems. Finally, the LSTM attention model used to extract temporal features is mentioned to be insufficient in terms of efficiency due to the length of the testing sequence and low parallel computing capacity.

In [73], a new method named GaitSet is proposed to obtain spatial and temporal information, differing from existing methods that view walking as a template or sequence. The study demonstrates that using additional feature extraction methods alongside deep networks yields more successful results than those found in the literature.

GaitPart [67] performs individual gait recognition by considering both static appearance features and dynamic temporal information. Previous studies have been conducted without detailed acquisition of temporal features. GaitPart stands out with its detailed modeling of temporal features.

In [10], GaitSlice is proposed to refine gait recognition features in both spatial and temporal dimensions, based on the logic that the less information included in gait silhouettes, the more significant the role of key frames of body parts. The proposed model has particularly improved gait recognition accuracy under cross-view conditions and complex walking conditions.

In [74], the VGG-19 and MobileNet-V2 models were trained using deep transfer learning. Subsequently, a new moth-flame optimization algorithm was developed to select the best features. It has been stated that combining lightweight model features with the developed algorithms is time-consuming, but accuracy has been increased in this way. Additionally, it has been determined that the optimization algorithm reduces computation time and increases accuracy.

In GLN [68], features at the silhouette-level and set-level were extracted at different stages within the deep network backbone and were combined from top to bottom via lateral connections. This approach aggregated more visual details, thereby enhancing the accuracy of gait recognition. Additionally, the size of the gait representations was reduced using a compact block. The proposed method has outperformed previous studies in the literature in terms of both accuracy and size.

SRN [11] differs from previous studies mainly by its method of coordinating silhouette-level and set-level information for set-based feature learning from silhouettes. Additionally, SRN proposes a method to leverage shallow layer features to better learn part representations. In particular, compared to GLN, which uses silhouette-level and set-level information, it has been stated that upsampling or lateral connections are unnecessary. Therefore, SRN suggests a method that utilizes only marginal memory cost and takes advantage of shallow layer features to learn more robust part representations. The proposed approach is superior to its counterparts in terms of accuracy especially under challenging conditions.

The study conducted in [9] introduces a new spatiotemporal augmented relation network (STAR). It facilitates the generation of visual clues in various regions for fine-grained feature learning through its contained modules and adaptively locates non-overlapped various regions that have significant identity information. With these aspects it offers, it enables better extraction of distinct information among frames and has improved accuracy compared to studies in the literature.

The method proposed in [70] extracts temporal features using its simple but effective three-dimensional CNN model. This method performs better than the other studies through this feature extraction technique.

In the study conducted in [72], unlike other methods, a module named FQBlock is proposed to measure the quality of each frame. FQBlock works on the number of feature channels, evaluating the features of each frame separately. Moreover, the attention values of each frame are based solely on its own features and do not change with permutation according to the silhouette pattern. FQBlock shares weights across different silhouettes, thus ensuring the comparability of attention values of frames in different sequences. These features have enabled the GQAN method to achieve more successful results than previous ones.

GaitAMR [75] is superior over other methods in both feature representation and temporal representation dimensions, due to its attention to potential silhouette error issues, the impact of local body features on final recognition, spatial occlusion errors, and appearance variation. It also performs better than other methods in terms of recognition performance within a smaller iteration period.

In [69], an effort was made to capture spatial features such as body shape and the temporal characteristics of walking patterns, specifically to address the challenges of person recognition encountered by gait recognition algorithms in open environments. GEI (gait energy images) and 3D CNN model was employed for both feature extraction and gait recognition. Additionally, the network's parameters were optimized using Bayesian optimization. Thanks to the proposed three-dimensional model and the conducted hyperparameter optimization, this method ranks among the successful studies in the literature.

In [71], a temporal modeling network is proposed to combine multi-scale temporal features. Additionally, a spatial feature learning module is also suggested to fix feature corruption problems resulting from temporal processes. Studies conducted on datasets have demonstrated the superiority of the model compared to current methods.

5 Challenges and future perspectives

Despite substantial improvement in recent years, there are still several challenges in human gait recognition. These possible challenges include variability in walking patterns, occluded views, environmental factors, lack of gait datasets, ethical and privacy concerns and learning challenges. The subsequent part of this section provides a detailed description of them. The accuracy, reliability, and usefulness of gait recognition systems can be improved by researchers by focusing on these challenges.

5.1 Variability of walking patterns

People walk in different ways, and the same person may exhibit many walking styles depending on circumstances such as walking speed, carrying and clothing, surface type, and aging. When people walk at different speeds, or on different surfaces, they naturally adjust gait parameters such as stride length and step width to maintain balance. Carrying a bag may cause the upper body to lean forward, resulting in a longer stride length. Tight or restricting clothes can limit hip and leg range of motion and high-heeled shoes can tilt the ankles forward, resulting in a shorter stride length. These adjustments can cause changes in the way a person walks and the features of their gait pattern. Most of the previous studies [9–11, 66–74] achieve promising results even with datasets with some of these conditions. Aging also can lead to changes in the walking pattern. Changes in joint flexibility and mobility can lead to a reduction in the stride length. Some data sets [21] contain gait data of the same people at different times. Even so, the longest time interval is 15 months. There is a need for further research over a much longer time frame.

5.2 Occluded views

In real-world scenarios, gait recognition systems can be blocked by obstacles such as bags, cars that occlude the view of a part of a person's body. This could be challenging to capture enough information about the gait pattern to identify an individual. The researchers who will create the new dataset can use multiple cameras or sensors to collect data from different angles and viewpoints to solve the problem of occluded views. Another way to solve this problem can be through human body alignment, where the system aligns various parts of the body, including the head, torso, and limbs. In this way, gait recognition algorithms can better detect and track the person's gait patterns, even in situations with partial obstacles. There are several studies [69, 71, 75] specified they improve gait recognition in occlusion conditions, and GaitPart [67] extracts gait features from different parts of the body and can partially solve the problem of occluded views.

5.3 Environmental factors

In real-world scenarios there are many uncontrolled factors in the environment such as lighting conditions, shadows, and camera angles that can affect gait recognition accuracy. Researchers can conduct experiments in a number of real-world environments to investigate the impact of these factors on gait recognition accuracy. They can identify which factors have the greatest impact on accuracy by collecting data in varying lighting conditions, for example, and develop algorithms that are more resilient to these variations. Varying lighting conditions affect the consistency and reliability of the captured gait data collected. Gait recognition can be highly sensitive to changes in lighting, which can alter the appearance of the subject's silhouette and overall visibility. Poor lighting can lead to incomplete or inaccurate silhouettes, making it difficult to extract reliable gait features [76]. Fluctuating lighting can introduce variability in important features, reducing the model's ability to recognize and classify gait patterns accurately. Strong lighting can create shadows that may be misinterpreted as part of the gait, leading to incorrect feature extraction and analysis [77]. By employing a combination of robust feature selection, preprocessing techniques, depth sensing, and adaptive machine learning approaches, it's possible to mitigate the impact of lighting variability and enhance the performance of gait recognition systems in diverse environments. Most of the current publicly available gait datasets were obtained under controlled conditions and are comparatively simple to recognize. ResGait [62] dataset is based on real scenarios and GREW [21] dataset is optimized for real-world applications.

5.4 Lack of gait datasets

Gait recognition systems rely on large amounts of data to accurately identify individuals. However, obtaining and labeling such data can be time-consuming and expensive. A possible solution for researchers to access more data is to generate synthetic gait data from virtual 3D human models. As synthetic data can be produced with remarkable control and accuracy, it can also be used to create data that captures specific variations in gait patterns that may be difficult to capture in real-world data. VersatileGait [61] is the only synthetic data in gait recognition as far as we know, and it contains gait data of 11,000 subjects. The use of unlabeled data, which is easily accessible via the videos on the Internet, can also help overcome the lack of gait data. But, since labeling these data one by one will be tedious and time-consuming, self-supervised learning can help researchers at this point. Self-supervised learning has shown potential to train such unlabeled data, as it can learn useful representations of the data without requiring human labeling [78].

5.5 Ethical and privacy concerns

Gait recognition is a form of biometric identification, and there are worries over data privacy and its exploitation. People may be worried about having their gait patterns captured and retained, particularly if they are unfamiliar with the technology or how their data will be used. Gait recognition could become a powerful tool for mass surveillance, as gait can be captured remotely without the person's knowledge or consent. Gait data could be repurposed for uses not originally intended, or it might fall into the hands of unauthorized individuals who could exploit it for illegal activities [79]. Addressing these concerns requires comprehensive regulatory frameworks. There should be strict guidelines on data collection, usage, and storage, ensuring individuals' rights are protected. The gait data should be maintained securely and protected from unauthorized access through secure storage and protective measures such as encryption and access control. Moreover, the development of gait recognition technologies must include ethical considerations from the outset, with ongoing assessments of their impact on society.

5.6 Learning challenges

The use of deep learning and machine learning techniques within gait recognition areas brings several crucial challenges that must be resolved in order to get reliable findings. An examination of a few key issues is provided below.

When a model learns the training set too well, overfitting occurs, and this leads to poor generalization to new unknown data [80]. This could mean that the model performs well when applied to known subjects but misrecognizes the gait patterns of unknown subjects. Overfitting can be addressed by regularization strategies, data augmentation approaches, and dropout layers (in deep learning models). Furthermore, using cross-validation to check model performance on unseen data during training might lead to the early termination of training to avoid overfitting [81].

Deep learning models, in particular, are considered black boxes due to their complex architectures and the high dimensionality of their learned feature spaces [80]. This lack of interpretability may cause issues in sensitive gait recognition systems, where it is important to understand the reasoning behind decisions. Model behavior may be partially understood by visualizing the parts of the input data that have the most impact on the model's decisions through the use of techniques such as layer-wise relevance propagation (LRP) [82].

Machine learning and deep learning models, especially the deep learning models, require large amounts of labeled data for training. It takes a lot of time and resources to gather and analyze a large number of gait patterns. Using synthetic data or unlabeled data might be an option as shown in 5.4.

Two major issues that can frequently arise in gait recognition with deep learning are catastrophic forgetting and low inter-class variance. Catastrophic forgetting happens when a neural network loses information from past tasks after training on a new task. Elastic weight consolidation (EWC), proposed to solve this problem, allows the network to learn new tasks while helping to preserve weights that are important for previous tasks [83]. Low inter-class variance indicates a situation in which distinct classes (i.e., gait patterns of different individuals) have highly similar features, making it challenging for the model to distinguish between them successfully. This can result in higher misclassification rates, since the model fails to identify unique identifying features that distinguish one person's gait from another's. Feature aggregation is an effective method for addressing the challenge of low inter-class variance in gait recognition and other tasks that require distinguishing between extremely similar classes [84]. To address the challenge of low inter-class variance in gait recognition, in [25] a novel approach is introduced through the development of a generalized inter-class loss. This strategy tackles the problem by focusing on both the sample-level and class-level feature distributions.

One of the most typical challenges in machine learning is class imbalance referring to a situation, where the number of instances of one class significantly outnumbers the instances of one or more other classes in a dataset. This imbalance can lead to biased models that tend to predict the majority class better than the minority classes. In gait recognition tasks, each class typically represents an individual. If the dataset contains approximately an equal number of walking examples for all individuals, significant class imbalance does not occur. However, there may be situations where some individuals have more examples than others. This could lead to the model recognizing some individuals better than others, which can cause problems, especially in sensitive applications. There are some strategies to solve the problem of class imbalance such as oversampling (increasing the number of minority class instances), undersampling (reducing the number of majority class instances) and using ensemble methods [85].

6 Conclusion

This survey provides an extensive examination of appearance-based methods for human gait recognition, covering the significant developments made in this field. The paper highlights the enormous advances that have been achieved in this field as well as the many strategies used for gait recognition using visual information. We have demonstrated the effectiveness of appearance-based methods in successfully recognizing individuals based on their unique gait patterns through careful examination. The paper also reviews publicly available gait datasets that are commonly used for gait recognition, and it underlines the significance of dataset size, quality, and diversity in the development of accurate and robust gait recognition algorithms. Furthermore, challenges such as variability in walking patterns, occluded views, environmental factors, lack of gait datasets, ethical and privacy concerns, and learning challenges are addressed, along with potential solutions proposed in recent research. In conclusion, while appearance-based human gait recognition demonstrates considerable promise and has achieved significant progress, there is still need for further exploration and improvement. Future research should focus on addressing the identified challenges, exploring the integration of different types of data, and improving the interpretability and generalizability of gait recognition models. Overall, appearance-based human gait recognition algorithms have a lot of potential for applications in surveillance, health and biometrics, and future progress in this subject will help to improve security, personal identity, and health tracking systems.

Author contributions Study conception done by PGŞ, SŞ, FKG; Study design done by PGŞ, SŞ, FKG; Supervision done by PGŞ, SŞ, FKG; Figure and Table preparation done by PGŞ; Materials done by PGŞ, SŞ, FKG; Data collection and/or processing done by PGŞ; Literature review done by PGŞ, SŞ, FKG; Manuscript preparation done by PGŞ, SŞ, FKG; and Critical review done by PGŞ, SŞ, FKG.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK).

Data availability Not applicable as no datasets were generated during the current study.

Declarations

Conflict of interests The authors declare that they have no competing interests.

Ethical approval Not applicable as the nature of the article being a survey.

References

1. Sarkar S, Phillips PJ, Liu Z et al (2005) The humanID gait challenge problem: data sets, performance, and analysis. IEEE Trans Pattern Anal Mach Intell 27:162–177. https://doi.org/10.1109/TPAMI.2005.39
2. Nixon MS, Carter JN, Cunado D et al (1999) Automatic gait recognition. In: Jain AK, Bolle R, Pankanti S (eds) Biometrics. Springer, Boston, pp 231–249
3. Wang L, Ning H, Tan T, Hu W (2004) Fusion of static and dynamic body biometrics for gait recognition. IEEE Trans Circuits Syst Video Technol 14:149–158. https://doi.org/10.1109/TCSVT.2003.821972
4. Wu Z, Huang Y, Wang L et al (2017) A comprehensive study on cross-view gait based human identification with deep CNNs. IEEE Trans Pattern Anal Mach Intell 39:209–226. https://doi.org/10.1109/TPAMI.2016.2545669
5. Chen J (2014) Gait correlation analysis based human identification. Sci World J 2014:1–8. https://doi.org/10.1155/2014/168275
6. Wan C, Wang L, Phoha VV (2019) A survey on gait recognition. ACM Comput Surv 51:1–35. https://doi.org/10.1145/3230633
7. Kale A, Sundaresan A, Rajagopalan AN et al (2004) Identification of humans using gait. IEEE Trans on Image Process 13:1163–1173. https://doi.org/10.1109/TIP.2004.832865
8. Kusakunniran W (2020) Review of gait recognition approaches and their challenges on view changes. IET Biom 9:238–250. https://doi.org/10.1049/iet-bmt.2020.0103
9. Huang X, Wang X, He B et al (2023) STAR: spatio-temporal augmented relation network for gait recognition. IEEE Trans Biom Behav Identity Sci 5:115–125. https://doi.org/10.1109/TBIOM.2022.3211843
10. Li H, Qiu Y, Zhao H et al (2022) GaitSlice: a gait recognition model based on spatio-temporal slice features. Pattern Recogn 124:108453. https://doi.org/10.1016/j.patcog.2021.108453
11. Hou S, Liu X, Cao C, Huang Y (2021) Set residual network for silhouette-based gait recognition. IEEE Trans Biom Behav Identity Sci 3:384–393. https://doi.org/10.1109/TBIOM.2021.3074963

12. Singh JP, Jain S, Arora S, Singh UP (2021) A survey of behavioral biometric gait recognition: current success and future perspectives. Arch Comput Methods Eng 28:107–148. https://doi.org/10.1007/s11831-019-09375-3
13. Sepas-Moghaddam A, Etemad A (2023) Deep gait recognition: a survey. IEEE Trans Pattern Anal Mach Intell 45:264–284. https://doi.org/10.1109/TPAMI.2022.3151865
14. Rani V, Kumar M (2023) Human gait recognition: a systematic review. Multimed Tools Appl. https://doi.org/10.1007/s11042-023-15079-5
15. Parashar A, Parashar A, Shabaz M et al (2024) Advancements in artificial intelligence for biometrics: a deep dive into model-based gait recognition techniques. Eng Appl Artif Intell 130:107712
16. Google Scholar. https://scholar.google.com/?hl=en&as_sdt=0,5. Accessed 4 Jul 2023
17. IEEE Xplore. https://ieeexplore.ieee.org/Xplore/home.jsp. Accessed 4 Jul 2023
18. ScienceDirect.global | Science, health and medical journals, full text articles and books. https://sciencedirect.global/. Accessed 4 Jul 2023
19. Iwama H, Okumura M, Makihara Y, Yagi Y (2012) The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition. IEEE Trans Inform Forensic Secur 7:1511–1521. https://doi.org/10.1109/TIFS.2012.2204253
20. Takemura N, Makihara Y, Muramatsu D et al (2018) Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition. IPSJ Trans Comput Vis Appl 10:4. https://doi.org/10.1186/s41074-018-0039-6
21. Zhu Z, Guo X, Yang T, et al. (2022). Gait Recognition in the Wild: A Benchmark. IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 2021, pp. 14769–14779.
22. Zheng J, Liu X, Liu W, et al. (2022). Gait Recognition in the Wild with Dense 3D Representations and A Benchmark. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 20228–20237).
23. Pearson K (1901) LIII. On lines and planes of closest fit to systems of points in space. Lond Edinburgh Dublin Philosophical Mag J Sci 2:559–572. https://doi.org/10.1080/14786440109462720
24. Fisher RA (1936) The use of multiple measurements in taxonomic problems. Ann Eugen 7:179–188. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x
25. Yu W, Yu H, Huang Y, Wang L. (2022). Generalized inter-class loss for gait recognition. In: Proceedings of the 30th ACM International Conference on Multimedia (pp. 141–150).
26. Crouse MB, Chen K, Kung HT. (2014). Gait Recognition using Encodings with Flexible Similarity Metrics. In: 11th International Conference on Autonomic Computing (ICAC 14) (pp. 169–175).
27. Zhang C, Liu, W, Ma H, Fu H. (2016). Siamese neural network based gait recognition for human identification. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 2832–2836). IEEE.
28. Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20:273–297. https://doi.org/10.1007/BF00994018
29. Bishop CM, Nasrabadi NM (2006) Pattern recognition and machine learning, vol 4. Springer, New York, p 738
30. Begg RK, Palaniswami M, Owen B (2005) Support vector machines for automated gait classification. IEEE Trans Biomed Eng 52:828–838. https://doi.org/10.1109/TBME.2005.845241
31. Gou H, Yan L, Xiao J (2015) A gait recognition system based on SVM and accelerations. MATEC Web Conf 30:06001. https://doi.org/10.1051/matecconf/20153006001
32. Rabiner LR (1989) A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE 77:257–286. https://doi.org/10.1109/5.18626
33. Suk H-I, Sin B-K (2006) HMM-based gait recognition with human profiles. In: Yeung D-Y, Kwok JT, Fred A et al (eds) Structural, syntactic, and statistical pattern recognition. Springer, Berlin, pp 596–603
34. Bae J, Tomizuka M (2010) Gait phase analysis based on a hidden markov model. IFAC Proc Vol 43:746–751. https://doi.org/10.3182/20100913-3-US-2015.00014
35. Dargan S, Kumar M, Ayyagari MR, Kumar G (2019) A survey of deep learning and its applications: a new paradigm to machine learning. Arch Comput Methods Eng 27:1–22
36. Zafar A, Aamir M, Mohd Nawi N et al (2022) A comparison of pooling methods for convolutional neural networks. Appl Sci 12:8643. https://doi.org/10.3390/app12178643
37. Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. Science 313(5786):504–507
38. Welling M, Kingma DP (2019) An introduction to variational autoencoders. Found Trends Mach Learn 12(4):307–392

39. Lecun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86:2278–2324. https://doi.org/10.1109/5.726791
40. Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT press, Cambridge
41. Medsker LR, Jain L (2001) Recurrent neural networks. Design Appl 5(64–67):2
42. Bengio Y, Simard P, Frasconi P (1994) Learning long-term dependencies with gradient descent is difficult. IEEE Trans Neural Netw 5(2):157–166
43. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780
44. Cho K, Van Merriënboer B, Gulcehre C, et al. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. preprint arXiv:1406.1078v3.
45. Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. (2014). Generative adversarial networks. arXiv:1406.2661.
46. Kodali N, Abernethy J, Hays J, Kira Z. (2017). On convergence and stability of gans. arXiv preprint arXiv:1705.07215.
47. Tran D, Bourdev L, Fergus R, et al. (2015). Learning spatiotemporal features with 3d convolutional networks. In: Proceedings of the IEEE International Conference on Computer Vision.
48. Nonis F, Dagnes N, Marcolin F, Vezzetti E (2019) 3D approaches and challenges in facial expression recognition algorithms—a literature review. Appl Sci 9(18):3904. https://doi.org/10.3390/app9183904
49. Shi X, Chen Z, Wang H, et al. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. Advances in Neural Information Processing Systems, 28.
50. Gross R, Shi J. (2001). The CMU motion of body (MoBo) database. Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep. CMU-RI-TR-01–18.
51. Shutler JD, Grant MG, Nixon MS, Carter JN (2004) On a large sequence-based human gait database. In: Lotfi A, Garibaldi JM (eds) Applications and science in soft computing. Springer, Berlin, pp 339–346
52. Wang L, Tan T, Ning H, Hu W (2003) Silhouette analysis-based gait recognition for human identification. IEEE Trans Pattern Anal Mach Intell 25:1505–1518. https://doi.org/10.1109/TPAMI.2003.1251144
53. Yu S, Tan D, Tan T. (2006). A Framework for Evaluating the Effect of View Angle, Clothing and Carrying Condition on Gait Recognition. In: 18th International Conference on Pattern Recognition (ICPR’06). IEEE, Hong Kong, China, pp 441–444
54. Tan D, Huang K, Yu S, Tan T. (2006). Efficient Night Gait Recognition Based on Template Matching. In: 18th International Conference on Pattern Recognition (ICPR’06). IEEE, Hong Kong, China, pp 1000–1003
55. Makihara Y, Mannami H, Tsuji A et al (2012) The OU-ISIR gait database comprising the treadmill dataset. IPSJ Trans Comput Vis Appl 4:53–62. https://doi.org/10.2197/ipsjtcva.4.53
56. Hofmann M, Geiger J, Bachmann S et al (2014) The TUM gait from audio, image and depth (GAID) database: multimodal recognition of subjects and traits. J Vis Commun Image Represent 25(1):195–206
57. Uddin MdZ, Ngo TT, Makihara Y et al (2018) The OU-ISIR large population gait database with real-life carried object and its performance evaluation. IPSJ T Comput Vis Appl 10:5. https://doi.org/10.1186/s41074-018-0041-z
58. Xu C, Makihara Y, Ogi G et al (2017) The OU-ISIR gait database comprising the large population dataset with age and performance evaluation of age estimation. IPSJ T Comput Vis Appl 9:24. https://doi.org/10.1186/s41074-017-0035-2
59. Song C, Huang Y, Wang W, Wang L (2022) CASIA-E: a large comprehensive dataset for gait recognition. IEEE Trans Pattern Anal Mach Intell. https://doi.org/10.1109/TPAMI.2022.3183288
60. An W, Yu S, Makihara Y et al (2020) Performance evaluation of model-based gait on multi-view very large population database with pose sequences. IEEE Trans Biom Behav Identity Sci 2:421–430. https://doi.org/10.1109/TBIOM.2020.3008862
61. Dou H, Zhang W, Zhang P, et al. (2021). VersatileGait: A Large-Scale Synthetic Gait Dataset with Fine-GrainedAttributes and Complicated Scenarios. ArXiv, abs/2101.01394.
62. Mu Z, Castro FM, Marin-Jimenez MJ, et al. (2021). ReSGait: The Real-Scene Gait Dataset. In: 2021 IEEE International Joint Conference on Biometrics (IJCB). IEEE, Shenzhen, China, pp 1–8.
63. Li X, Makihara Y, Xu C, Yagi Y (2022) Multi-view large population gait database with human meshes and its performance evaluation. IEEE Trans Biom Behav Identity Sci 4:234–248. https://doi.org/10.1109/TBIOM.2022.3174559

64. Phillips P, Grother R, Michaels D. (2003). FRVT 2002: Facial Recognition Vendor Test. Technical report, DoD.
65. Ye M, Shen J, Lin G (2021) Deep learning for person re-identification: a survey and outlook. IEEE Trans Pattern Anal Mach Intell 44(6):2872–2893
66. Zhang Y, Huang Y, Yu S, Wang L (2020) Cross-view gait recognition by discriminative feature learning. IEEE Trans on Image Process 29:1001–1015. https://doi.org/10.1109/TIP.2019.2926208
67. Fan C, Peng Y, Cao C, et al. (2020). GaitPart: Temporal Part-Based Model for Gait Recognition. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Seattle, WA, USA, pp 14213–14221
68. Hou S, Cao C, Liu X, Huang Y (2020) Gait lateral network: learning discriminative and compact representations for gait recognition. In: Vedaldi A, Bischof H, Brox T, Frahm J-M (eds) Computer vision— ECCV 2020. Springer International Publishing, Cham, pp 382–398
69. Gul S, Malik MI, Khan GM, Shafait F (2021) Multi-view gait recognition system using spatio-temporal features and deep learning. Expert Syst Appl 179:115057. https://doi.org/10.1016/j.eswa.2021.115057
70. Huang Z, Xue D, Shen X, et al. (2021). 3D Local Convolutional Neural Networks for Gait Recognition. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, Montreal, QC, Canada, pp 14900–14909
71. Huang X, Zhu D, Wang X, et al. (2022). Context-Sensitive Temporal Feature Learning for Gait Recognition. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 12909–12918).
72. Hou S, Liu X, Cao C, Huang Y (2022) Gait quality aware network: toward the interpretability of silhouette-based gait recognition. IEEE Trans Neural Netw Learn Syst. https://doi.org/10.1109/TNNLS.2022.3154723
73. Chao H, Wang K, He Y et al (2021) GaitSet: cross-view gait recognition through utilizing gait as a deep set. IEEE Trans Pattern Anal Mach Intell 44(7):3467–3478
74. Khan MA, Arshad H, Damaševičius R et al (2022) Human gait analysis: a sequential framework of lightweight deep learning and improved moth-flame optimization algorithm. Comput Intell Neurosci 2022:1–13. https://doi.org/10.1155/2022/8238375
75. Chen J, Wang Z, Zheng C et al (2023) GaitAMR: cross-view gait recognition via aggregated multi-feature representation. Inf Sci 636:118920. https://doi.org/10.1016/j.ins.2023.03.145
76. Lee TK, Belkhatir M, Sanei S (2014) A comprehensive review of past and present vision-based techniques for gait recognition. Multimed Tools Appl 72:2833–2869
77. Verlekar TT, Soares LD, Correia PL (2018) Gait recognition in the wild using shadow silhouettes. Image Vis Comput 76:1–13
78. Ohri K, Kumar M (2021) Review on self-supervised image recognition using deep neural networks. Knowl-Based Syst 224:107090. https://doi.org/10.1016/j.knosys.2021.107090
79. Boulgouris NV, Hatzinakos D, Plataniotis KN (2005) Gait recognition: a challenging signal processing technology for biometric identification. IEEE Signal Process Mag 22(6):78–90
80. Talaei Khoei T, Ould Slimane H, Kaabouch N (2023) Deep learning: systematic review, models, challenges, and research directions. Neural Comput Appl 35:23103–23124. https://doi.org/10.1007/s00521-023-08957-4
81. Jabbar H, Khan RZ (2015) Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study). Comput Sci Commun Instrum Devices 70(10.3850):978–981
82. Montavon G, Samek W, Muller K (2018) Methods for interpreting and understanding deep neural networks. Dig Signal Process 73:1–15
83. Kirkpatrick J, Pascanu R, Rabinowitz N et al (2017) Overcoming catastrophic forgetting in neural networks. Proc Natl Acad Sci 114(13):3521–3526
84. Zhang Z, Luo C, Wu H et al (2022) From individual to whole: reducing intra-class variance by feature aggregation. Int J Comput Vis 130(3):800–819
85. Al Musalhi N, Çelebi E. (2023). Age estimation in human gait extraction using a combination of multi-energy image with invariant moment. Preprints, 2023060186. https://doi.org/10.20944/preprints202306.0186.v1

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.