Gait Recognition Based on Fusion of HOG and LBP

Chaoran Liu

2019

School of Engineering, Computer and Mathematical Science

Abstract

In this paper, a GEI recognition algorithm based on layered fusion of Local Binary Pattern (LBP) and Histogram of Oriented Gradient(HOG) is proposed. This paper also proposes and designs a feature learning network based on convolutionally constrained Boltzmann machine. The network uses an unsupervised convolutionally constrained Boltzmann machine to train the Gait Energy Image(GEI) to generate a network model with gait feature extraction and convolution with different sizes of convolution kernels to obtain different types of features and merge them. At the same time, in order to realize the gait-based identification function, this paper uses the supervised all-connected layer network to learn and identify the extracted gait features.

KEY WORDS: Gait Recognition, LBP, HOG, Convolutional Boltzmann Machine.

Contents

1. Introduction

1.1 Background and Motivation

Identity technology has become a research hot-spot with the development of computer vision and information security. Biometric technology, which is based on the unique physiological or behavioral characteristics of individuals, is developing rapidly. As biometric is not as easily transferred or stolen as physical documents such as ID cards, which is safer, more reliable and more convenient. The gait recognition is based on the walking posture of the human body. Compared with fingerprint recognition, speech recognition, face recognition and other technologies, gait recognition has the advantages of easy collection, long distance, non-contact and difficult to camouflage, which is a research hot-spot in the fields of biometric, computer vision and information security.

Gait analysis was first posed in psychology. Researchers found that everyone has their own unique walking posture and walking pace (Liu, 2009). According to medical research, gait is a unique human body characteristic that depends on hundreds of kinematic parameters of the human body. Different individual movements have different steps, gait cycles, etc. Therefore, considering all the information of gait, it is unique and can be used as a biometric strategy to identify the human body. At first, Johansson discovered that the walking pattern of the human body can be judged by observing the movement of the light tied on a human body (MLD) through experiments in 1973. Based on this thesis, Kozlowshi and Cutting found that the gender of the person or close friends can be identified through the MLD experiment.

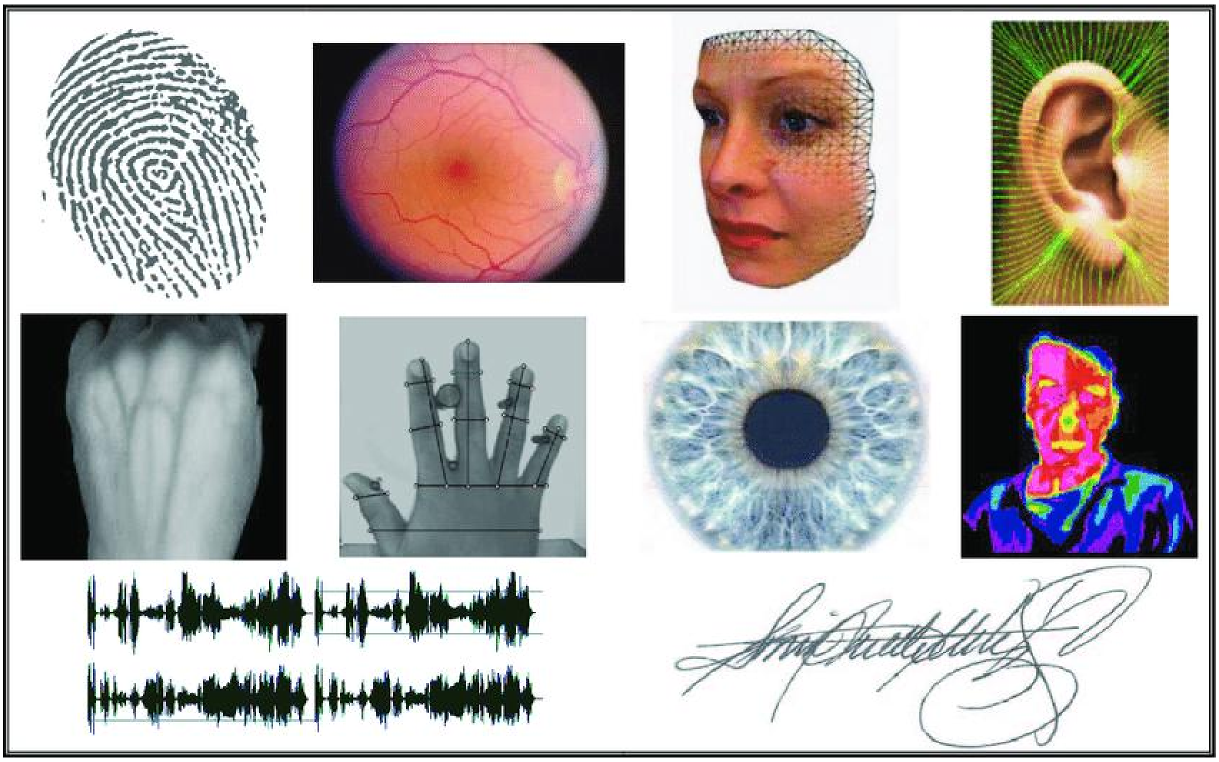


Figure 1. Examples of common biometrics

Actually, the theory of human recognition by gait had not been proposed until 1994 by Niyogi and AdelSon. In 1996, Murase and Sakai proposed a space-time correlation matching algorithm to distinguish different gait of different people. Another theory of human recognition by gait, which extracted frequency and phase features from the optical flow image, was proposed by Little and Boyd in 1998. This provides a good feature basis for further subsequent gait recognition researches. The famous Support Vector Machine (SVM) algorithm was applied to gait recognition in 2005. Later after 6 years, a new gait feature representation, Gait Flow Images (GFI), came to the world, which further improved the accuracy of gait recognition. Murodelaherran introduced the application of wearable technology and non-wearable technology in the field and expressed that the gait analysis of portable systems based on human sensors was a promising method at the end of article. Rida proposed a new approach combing Statistical Dependence (SD) feature selection with Global Locality Preserving Projections (GLPP) to mitigate the effects of intra-class changes and improve the negative impact of clothing changes and carrying conditions on recognition performance in 2016.

Some other theories and articles with breakthrough progress, for instance, a simple and effective gait recognition algorithm using spatial time profile analysis in low dimensional feature space applying supervised pattern classification technique, a step based on generalized tensor discriminant analysis and Gabor feature state recognition method, a more advanced strategy extracting gait features with Frame Difference Energy Image (FDEI) to suppress the effects of contour incompleteness, a new Patch Distribution Feature (PDF) for human gait recognition which achieved a better recognition rate, a sparse tensor discriminant position alignment algorithm for human gait feature representation and dimension reduction algorithm and a subspace set learning using the full correction (SEL-TCB) framework and its extension method based on tensor and partial patch are used for gait recognition algorithm.

In 2000, American Defense Advance Research Project Agency (DARPA) proposed a major project called Huid Identification at a Distance (HID) to study the human body's feature recognition under long-distance conditions, which greatly promoted the development of human gait recognition technology. At present, some large company like Jaguar Land Rover are applying a combination of face recognition and gait recognition to car alarm systems. Gait recognition applications are becoming more widespread. In recent years, scientific research institutions around the world have paid more and more attention to the study of gait recognition technology.

1.2 Research Question

This research aims at gait recognition and identification and implementation of a LBP and HOG based human gait recognition. The analysis of the methods and approaches helps with choosing the best strategies to implement this objective. Based on the reasons above, the research question should be:

Question:

How to create a working APP to recognize gait from the input video or instant surveillance camera for further research, application and analysis based on deep learning?

The core subjective of this research is gait recognition. Besides, the methods that had been adopted in the research or past work will be introduced and evaluated so that this research can follow the best strategies to achieve a superior level.

1.3 Influencing Factors and Difficulties

In order to fully explain the gait recognition, it is necessary to consider various factors that affect its recognition rate and the difficulties in achieving gait recognition.

1.3.1 Influencing Factors

In order to improve the recognition rate of gait recognition, it is necessary to first analyze various factors affecting it, and then take corresponding measures to overcome the above effects to achieve the desired performance. It should be pointed out that depending on the difference of the specific application environment and the difference in feature extraction, the factors affecting gait recognition are also different. According to Sarkar, Lee, Nixon and Boyd, the following factors are the main ones that influence recognition rate:

Time- Time changes generally cause changes in background, clothing, shoes and hats, etc. Especially when there is a large span of time, take six months for example, it will cause really big differences in many ways. There is no doubt that with the increase of age, people's gait will change. However, the current researches are mainly for adults, and the time span is generally within one year. In shape-based gait recognition, the influence of time on recognition performance is much greater than other factors; in model-based gait recognition, the change of contour caused by time variation also has a certain impact on feature extraction (Sarkar, 2005).

Apparel- Apparel changes have a relatively large impact on the extracted human silhouette.

Walking Pavement- The effects of walking the road surface mainly include the contour extraction below the calf (especially below the ankle) and the impact of the special road surface on the way people walk. For example, when walking on the grass, the grass will block part of the human foot, which will have a great impact on the contour extraction; while walking on the road, we can clearly see the various parts of the human body. According to Sarkar and Lee's researches in 2005 and 2004, the impact of walking pavement on recognition performance is second only to the effect of time variation.

Perspective- The change in perspective causes a large change in the extracted human contour, which has a large impact on the selection and extraction of features. According to the current research, it seems that the human body contour of the side view contains more valuable information. Most of the feature extraction is also based on the side silhouette. Of course, there are also perspective-independent feature extraction and recognition methods.

Other factors affecting gait recognition performance include occlusion, terrain, injury, fatigue, walking speed, specific training (soldier's gait), and psychological changes. Among many factors, current research focuses on time, clothing, and walking. Impact of pavement, viewing angle and shoes and hats. The extractions of gait features are mostly based on the extracted human silhouette images. The influence of various factors on the recognition performance is indirectly reflected by the influence of the human silhouette images. Therefore, in shape-based gait recognition, the change of human silhouette images have a great influence on the recognition performance; in model-based gait recognition, the change of human silhouette images will also have certain influence on feature extraction and recognition performance.

1.3.2 Difficulties

The research difficulties of current gait recognition can be divided into the following aspects:

Motion segmentation- Motion segmentation has always been a problem in the field of computer vision. The main problem is that it is susceptible to illumination conditions, noise, shadows, etc. In addition, accurate motion segmentation becomes extremely challenging when the foreground target is similar to the background. For a given sequence of gait images, the primary task is to accurately segment the human target. Among the many motion segmentation methods, the background subtraction method is mainly used at present, but for a more complex background, an adaptive background model needs to be established to achieve accurate background update, thereby achieving effective motion segmentation. Therefore, one of the difficulties in current research is the adaptive background modeling problem in which the performance meets real-time requirements in complex backgrounds.

Occlusion Processing- The current research on gait recognition mainly focuses on the situation of single-person movement in the visual area, and this obviously cannot meet the requirements of practical applications. In the actual environment, when the occlusion occurs, since the human body is only partially visible or not visible, the human body region cannot be segmented from the image at this time. Therefore, the recognition rate will be reduced to some extent when occlusion occurs. In order to overcome the problems caused by occlusion, it is necessary to explore the motion characteristics of the human body and establish a good motion estimation model. At the same time, the use of multiple cameras will also help greatly.

Performance Evaluation- In view of the lack of a common gait database and its size limitations, the current research on the performance of recognition algorithms is only in its infancy. Throughout the current state of gait research, the largest database used is less than 200, and the background is relatively simple, basically does not involve the identification of individuals in complex backgrounds (such as in the crowd), which makes us gait as a creature The ability to recognize feature recognition technology cannot make an accurate assessment.

1.4 Objective and Structure of the Research

The main objective of this research is to design and implement a feasible approach for gait recognition for further study. In the first section, the background, research status and the influence factors for gait recognition would be introduced. The following section will describe a whole picture of gait recognition methods and it will summery some of the past method that had been adopted. Some brief evaluation would also be given in this section. In the third section, this paper would present the method that had been applied in this research and define the principle of these methods. Experiment result would be given in the fourth section. And then, the fifth section would analyse the result and compare with some other related works. In the final part, it would not describe only the conclusion but also some further work.

2. Literature Review

2.1 Introduction and Conception

Computer vision is one of the many research fields that have developed from gait recognition. Given video sequence containing one or more subject walking processes, the gait recognition process in a broad sense can be divided into four main phases: pedestrian detection, pedestrian segmentation, pedestrian tracking, and pedestrian recognition. The pedestrian detection phase locates the position of the pedestrian in the single frame image and determines the pedestrian size. The pedestrian segmentation stage performs pixel-level segmentation on the pedestrian detection result and removes the background information in the video. The pedestrian tracking phase determines the motion trajectory of the target and distinguishes different individuals in the video sequence. In the general sense of gait recognition, it refers to the pedestrian recognition stage, which uses the feature extracted from the pedestrian silhouette map to identify the person. In recent years, with the development of deep learning, the general detection segmentation framework, such as Mask Region-Based Convolutional Neural Network (Mask RCNN), has made it possible to apply gait recognition technology to practical complex scenes.

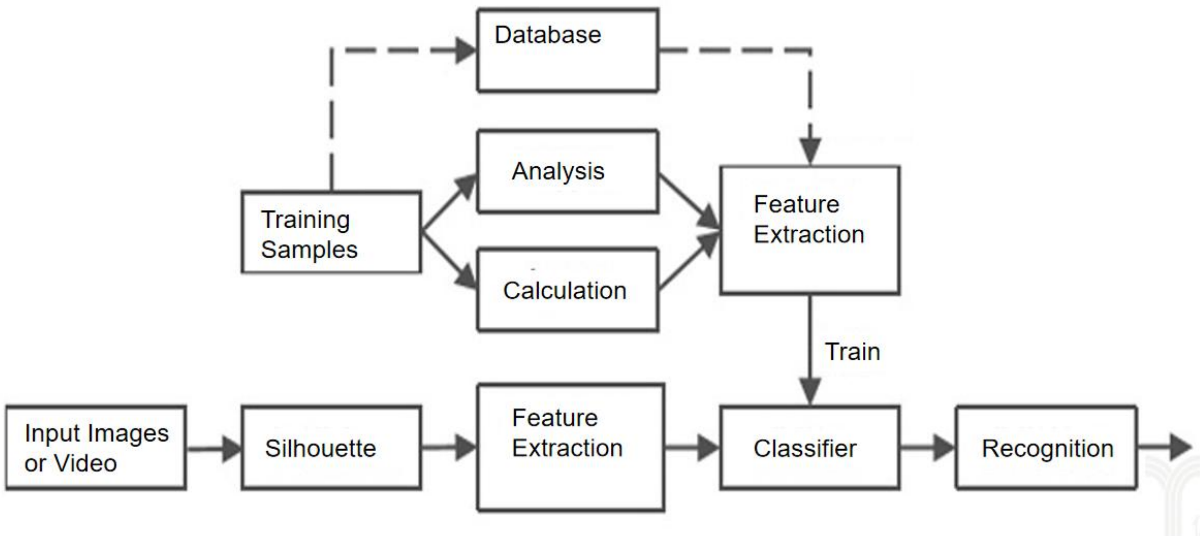


Figure 2. Flow Chart of General Steps of Gait Recognition

Gait recognition tasks can be divided into two categories according to task objectives.One type is the verification task, and the given sample and verification samples are judged according to a certain similarity index or a given threshold to identification task, that is, given the probe sample and the N samples in the gallery set, find the gallery sample with the same identity in the gallery set and the probe sample.

2.2 Gait Silhouette Extraction

Gait silhouette extraction can be defined that the walking human body contour is separated from the background as the foreground, including the positioning of the walking human body and the humanoid segmentation. The existing gait silhouette extraction methods can be divided into the following two categories according to the detection method: image sequence based motion detection and static image based pedestrian detection.

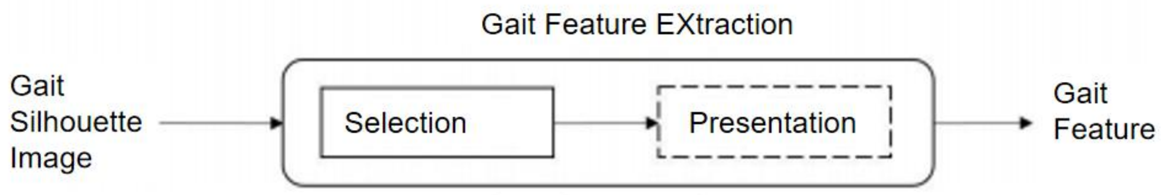


Figure 3. Flow Chart of Gait Feature Extraction

2.2.1 Image Sequence Based Motion Detection

Gait researches mainly focus on how the human body walks. The human contour target is moving, so it can be regarded as a moving target detection problem. Most of the existing gait studies use the mainstream gait database as the experimental object. However, all the background no matter indoors or outdoors in these databases are determined unchanged. In this case, there are three mainstream method to detect and locate the gait silhouette: frame difference method, optical flow method and background subtraction method.

The principle of the frame difference method is to differentiate and threshold the adjacent frames of consecutive image sequences. This method is simple to implement and efficient. It even can be applied in dynamic background environments and used to be one of the most commonly used methods of target detection and segmentation. But it generally extracts silhouette holes that are more serious and it is too sensitive to the environmental noise.

The optical flow method uses the characteristics of the optical flow which change with the motion of the object to detect the gait. Using the optical flow field, the complete motion information of the target can be obtained and the foreground target motion can be detected well from the background. However, due to factors such as illumination, the accuracy of the extracted moving targets is not always as good as expected.

Background subtraction method adopts the difference between the current image and the background image to detect the silhouette of the human body in the segmented motion. The most critical step of it is to create a background model. There are several technique for this, for example, time average method, time median image and mixed Gaussian model. Background subtraction method has poorer adaptability to dynamic background that it to static background. Therefore, it is more suitable for use on the current gait database.

2.2.2 Static Image Based Pedestrian Detection

The background in the actual application scenario is complex and variable, and the foreground may be walking by multiple people. It is difficult to accurately locate and extract the silhouette of the human body based on the motion detection method. Therefore, detecting and segmenting the silhouette of the human body for each frame of image becomes a countermeasure, which involves two steps: pedestrian detection and human silhouette segmentation.

The purpose of pedestrian detection is to detect and locate the location of the pedestrian from the image of humans. In the past ten years, it has been a hot issue of research, and there is a very rich pedestrian detection data set and perfect evaluation indicators. Pedestrian detection can also be seen as a classification problem. A series of windows of different sizes and different sizes are generated first, then the features are extracted, and finally the classifier determines whether the window contains pedestrians. Pedestrian detection methods are divided into two categories according to whether the feature extraction is a manually customized feature: pedestrian detection based on artificial features and pedestrian detection based on deep learning. The features that used for pedestrian detection are Haar wavelet, Histogram of Oriented Gradient (HOG), Local Binary Pattern (LBP), and Scale-Invariant Feature Transform (SIFT). these features are extracted directly from the original image, so they are really fast to generate and suitable for pedestrian detection in real-time scenes. These features combine Adaboost classification algorithm and SVM classifier separately or in combination, and achieved good detection results in the implementation environment. In recent years, with the rise of deep learning, the record of pedestrian detection test database has been continuously refreshed. Pedestrian detection based on deep learning has become the most attractive method in the present and future. There are two main categories for this method: One is based on regional nomination methods, the other one is an end-to-end target detection. The method based on regional nomination first needs to select some windows that may have targets as the input of deep neural network. End-to-end target detection starts directly from the original image and finally outputs the location of the pedestrian.

2.2.3 Human Silhouette Segmentation

Human silhouette segmentation could be simply understood as a separate process of foreground and background.The traditional segmentation method takes into account the relationship between pixels and pixels, provides a threshold and divides the picture into two pictures.The most typical of these is the N-Cut graph cutting method. It describes the relationship information between pixels as a distance, and divides the image according to the distance difference, and generally adjacent similar color pixels are divided as a whole part from the others However, for pedestrians, they may wear different colors of clothing. This has a great impact on the segmentation, resulting in an unsatisfactory effect. Semantic segmentation based on deep learning, through the independent learning of the semantic information of human silhouette, the segmentation effect achieves satisfactory results.

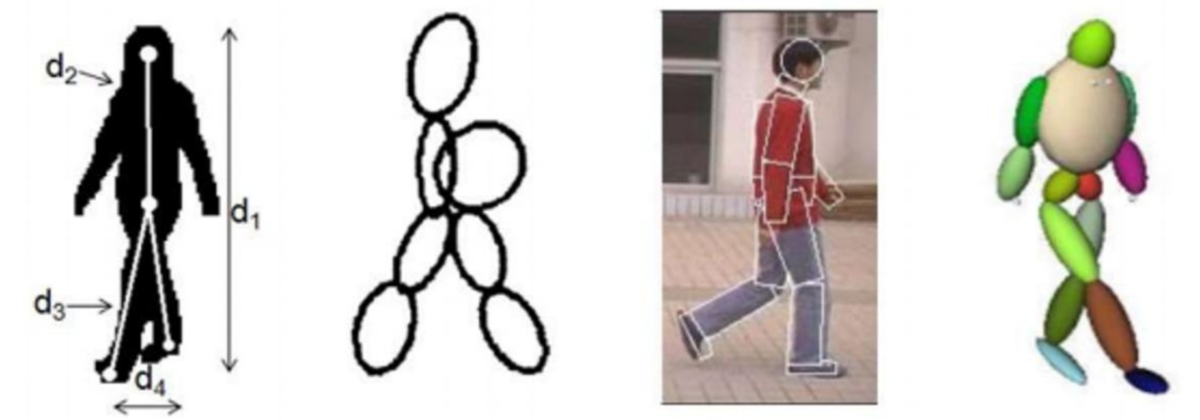
2.3 Gait Feature Extraction

Gait feature extraction is a key step in gait recognition. In the gait recognition process,the gait features extracted from the human silhouette image sequence directly determine the final recognition performance. This stage includes two partial: feature creation and feature learning. Feature creation is the use of empirical extract features from the original gait silhouette image sequence. Feature learning is to learn better gait features through statistical machine learning methods. The feature learning part is not essential. The new feature expression is learned on the basis of the created gait characteristics.

2.3.1 Model-based Gait Feature

The model-based approach tracks the parameters of the analytical model by establishing a static structure and motion model of the body part of the body. The gait model is usually established by using the trajectory of the joint position of the human body, the length of the limbs, and the angle formed by the legs. After modeling the human body structure or human motion based on the model-based method, the two-dimensional image sequence data is correlated with the model data to obtain the feature or model related parameters. The occlusion phenomenon may occur at any time in practical applications. People's walking images have the phenomenon of carrying a bag, an umbrella, a backpack, etc., which is enough to change the shape and cover part of the human body. In many cases, there is even a phenomenon of human body self-occlusion. For gait recognition, it is crucial to solve the occlusion problem. Model-based gait analysis has this advantage for occlusion problem because the models are built on the movement patterns of people in the sequence image, reflect current changes and can even estimate past and future changes.

Up to now, there are some mature methods for model-based which had been adopted in the past researches:



(a) Static Length Model (b) Elliptical Model (c) Geometry Model d) 3D Model

Figure 4. Common Model for Gait Feature

Elliptical model- Elliptical model which was founded by Lee used multiple ellipses to fit parts of the body and divided the human side image into seven parts. In this way, it represents the human body structure as seven ellipses that connected to each other. Each ellipse is uniquely determined by four parameters, together with the body's centred height, there are a total of 29 parameters that would be used for gait feature representation.

Pendulum model- The thigh is modeled as a linked pendulum and the gait feature is obtained from the frequency components of its tilt angle signal. Cunado first used the angle model based on hip and thigh to model the gait. The angle changed periodically with the human body walking. The frequency of the angle-changing signal was used as the feature to extract the motion frequency components of the hip joint. Therefore, they used Fourier descriptor represents the gait feature.

3D model- Some researchers believe that since the human body is a 3D figure, its motion description would be more accurate using a 3D model. Fua () built a 3D anatomical model even with skin. This more realistically reflects the structural characteristics of the human body, providing the possibility to use 3D models for gait recognition. Zhao used the video sequence acquired by the multi-camera to create a 3D skeleton model, which represented the human body with 15 joint points and 14 lines. It realized the motion tracking through the local optimization algorithm. The method extracts the length of some key areas as a static feature and uses the motion trajectory of the lower limbs as a dynamic feature. The experiment verified the validity of the 3D analysis and also proved the importance of lower limb joints and low trust of upper limb joints in gait recognition.

2.3.2 Appearance-based Gait Feature

Different from the method of using the model, the appearance-based method does not need to locate the key parts of the human body, and only needs to perform statistical analysis on the spatio-temporal pattern generated by the moving human body in the gait image sequence to obtain the corresponding gait features.The appearance-based method has a relatively low computational complexity and can handle low-resolution gait images, which is the method used by most gait researchers.The research object based on the appearance method contains the outline or shape of the human gait. The gait shape is the outermost boundary of the gait image and is hollow; the gait silhouette is a binary image of the entire body and solid.

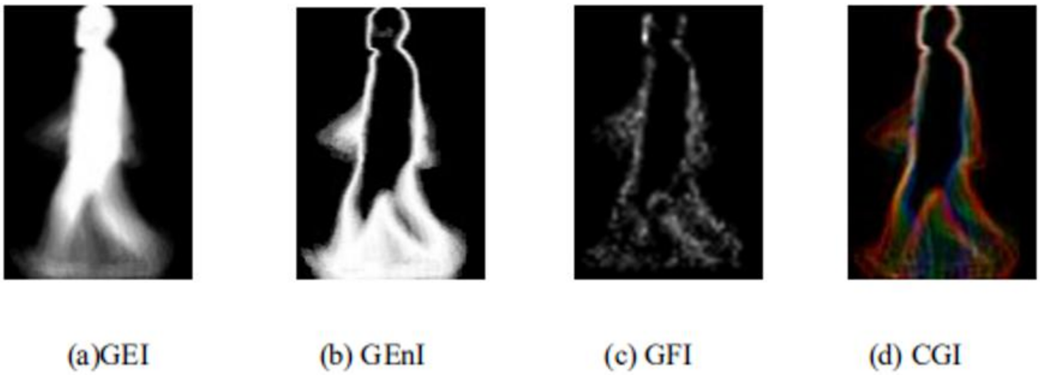


Figure 5. Time normalized gait image

Image sequence based method- This type of method directly uses image sequences to represent gait features. Since the length of the gait image sequence is not fixed, other techniques such as correlation coefficients and hidden Markov models (HMMs) are needed to calculate the similarity between the two gait sequences.

Space coding based method- Gait video is 3D data (2D space and one-dimensional time), which is computationally intensive, with many noises and low recognition efficiency. Some methods attempt to encode a two-dimensional gait image into a one-dimensional vector, ultimately obtaining a two-dimensional gait feature. Generally, the shape of the human body contour is regarded as a closed curve, and it is much easier to encode the curve with a vector.

Spatio-temporal description based method- Although the coding of the body contour can greatly reduce the complexity, since the gait is a periodic signal with an unfixed length, it is troublesome to match the gait characteristics in the recognition phase. For this, some methods directly extract spatio-temporal features, and do not need time series matching methods to help identify.

Time accumulation based method- Although the extracted gait features can be directly used for classification based on the spatio-temporal description method. The spatio-temporal signals extracted by these methods lose the spatial positional relationship information of the original body parts,which the information is proved to be quite significant for gait recognition. For the feature representation of the gait image sequence, a better solution is to use the time-cumulative statistics of the gait image at spatial locations. All contours of a gait cycle are accumulated according to certain rules to form a time normalized graph. For instance, Procrustes mean shape(PMS), Motion History Image(MHI), Gait Energy Image(GEI), Gait Entropy Image(GEnI), Gait Flow Image(GFI) and ChronoGait Image(CGI).

2.4 Classification and Recognition

The classification and recognition phase assigns the gait samples of the test to the corresponding category labels, which is, identifies the people. The process is to calculate the similarity between the test sample and the registered sample and complete the classification according to certain judgment rules. This stage is often combined with feature extraction to identify features that are suitable for classification. This type of method can be divided into two categories according to whether the gait feature is a time series feature.

Sequentially matching- The gait silhouette image sequence is a time series signal with a unfixed length of time. The sequentially matching techniques for gait mainly include: Dynamic Time Warping(DTW) and Hidden Markov Models(HMMs). DTW has the advantages of simple concept and robust algorithm. It was widely used in speech recognition in the early days and was later used to match human motion patterns. For DTW, even if the time scale of the test sequence pattern and the reference sequence pattern are not completely consistent, as long as the time order constraint exists, it can better complete the pattern matching between the test sequence and the reference sequence. HMMs are more mature techniques for matching time-varying data, and they are random state machines. Adopting HMMs involves two stages: training and classification. The training phase includes specifying a hidden state number of a hidden Markov model and optimizing the corresponding state transitions and output probabilities so that the resulting output symbols match the image features observed within the particular motion category. The classification phase involves the calculation of the probability that a particular HMMs may produce a sequence of test symbols corresponding to the observed image features. HMMs have better advantages than DTW in learning ability and processing undivided continuous data streams.

Statistical classification method- After feature extraction, the gait feature is generally a fixed length vector, picture or tensor, which can all be seen as a fixed length vector. There are many statistical classification methods corresponding to it, such as Nearest Neighbor classifier(NNC), Support Vector Machine(SVM), Bayesian classification, neural network and so on. NN directly uses the distance formula to calculate the similarity between the probe sample and the gallery sample and classifies the probe sample into the class of the most similar gallery sample. Some of the metric distances with similar similarities are: Euclidean distance, Manhattan distance and Mahalanobis distance. Based on Vapnik statistical learning theory, VC dimension theory and structural risk minimization principle, SVM seeks an optimal interface for different types of sample data. It itself is a linear classification method. But it can be extended to a nonlinear classification by a kernel function. The data set is mapped to a high-dimensional space and the hyperplane with the largest spacing is found in this space for classification. Bayesian classifiers are classification methods based on decision theory. It focuses on the specific probability distribution of the sample. There are two basic requirements when using this method for classification. The first is that the overall probability distribution of each category is known. Then the number of categories to be classified is certain. It utilizes the decision criterion of the maximum a posteriori probability as the minimum error probability criterion. If the cost caused by the actual decision error is considered, the criterion can become the minimum average conditional risk criterion. The neural network is essentially an adaptive nonlinear learning system, which refers to the theory of human neuron activity. It shows characteristics of adaptability, parallelism, robustness, fault tolerance and learning. In recent years, with the breakthrough of deep neural networks, neural network-based classification methods are widely used in image recognition, target tracking and other fields. Neural networks, especially deep neural networks, has become the hottest research topics.

2.5 Evaluation Method

The gait recognition method is similar to the evaluation of other biometric methods. The gait data set is divided into a training set and a test set.The training set is used to train to generate a classification model and the test set is used to evaluate the generalization performance of the classification model. The training set and the test set maintain a mutually exclusive relationship, that is, the test samples should not appear in the training set. Train the model with the training set and evaluate the pros and cons of the gait recognition method on the test set. From a practical point of view, the recognition rate is the primary performance indicator for measuring the gait recognition algorithm. Generally, the higher the recognition rate, the higher the performance of the algorithm. In addition, cumulative recognition rates and ROC curves are often used to evaluate performance. Moreover, real-time or not is also an important indicator for measuring gait recognition algorithms.

Correct Recognition Rate(CRR)- Correct recognition rate is also called correct classification rate(CCR). It is the most commonly used evaluation index for gait recognition, that is, the ratio of the correct number of samples to the total number of samples.

Cumulative Match Score(CMS)- The cumulative match score is a sort of matching results, in which the first k ranks contain samples of the real category, which is considered as successful identification. Generally speaking, gait is used as an auxiliary biometric in actual use. As long as the real category is included in the category of the top k recommended matches, it can be regarded as a successful recognition. The cumulative matching identification curve, that is, the CMS curve, refers to the relationship between the cumulative recognition rate and the number of ranks. CRR is a special case where the cumulative recognition rate is at rank 1.

Verification performance evaluation- In addition to identifying performance, verification performance is also a criterion for evaluating an algorithm's quality that with the ability to evaluate impersonation and malicious intrusion. Commonly used model evaluation indicators have accuracy, recall rate, F1, AUC and ROC curves. These indicators are related to each other and have different emphasis. In the biometric identification problem, the ROC curve is the most widely used evaluation method for verification performance. There are two main indicators in the ROC curve, which are True Positive Rate(TPR) and False Positive Rate(FPR). With FPR as the horizontal axis and TPR as the vertical axis, the threshold is changed to obtain a curve called Receiver Operating Characteristic(ROC). The area under the curve reflects the verification performance of the algorithm. The larger the area, the better the performance.

Time Cost- Since gait recognition technology is ultimately applied to practical applications, the time cost is an important indicator of gait recognition technology. The calculation time of the gait recognition system mainly has two aspects: one is for the training phase. The time taken by the gait recognition system to learn the process. The other is for the test phase that the time required for gait feature extraction and the matching recognition time. Usually the training of the gait recognition system is offline training. In an identification system, the test time cost is usually more significant than the training time cost.

3. Methodology

3.1 Gait Energy Image

Different from directly using the gait silhouette as an input to the deep network, Shiraga applied Gait Energy Images(GEI) as the input feature. GEI is a gait model of static and dynamic information in a sequence of mixed gait silhouette. The energy of each pixel in the model is obtained by calculating the average intensity of the silhouette pixels in a gait cycle.

$$GEI = \frac{1}{N}\sum_{t=1}^{N} B_t ,$$

In this formula, N is the number of frames in a single gait cycle, B_t is the silhouette of frame No.t in the gait cycle. In order to solve the cross-view problem in gait recognition, a network structure with two layers of convolutional layer is proposed. The feature representation of the view is invariant through a fully connected layer. In the training process, the recognition problem is regarded as the classification problem on the training data set. In the last layer of the network, the soft max function is used to calculate the cross entropy loss of the multi-classes. In the test phase, the nearest neighbor classifier is used to identify the features with the same view angle.

3.2 Calculating the Gait Cycle

There are many approaches for gait cycle calculation. Since the width and height of the silhouette of the human body change periodically as the person walks, it is defined as a gait cycle when the ratio of the width and height of the human silhouette appears twice in succession. Some researchers think that the time required to take the maximum distance from one right leg in a frame of a human target image in the gait sequence to the next right leg to take the maximum distance is a gait cycle. In this paper, the optical flow method is used to calculate the gait cycle.

The optical flow method describes the shape of the instantaneous motion, including the shape (space) of the moving object and the shape (time) of the motion. Therefore, different gaits can be distinguished from the periodic differences in the motion shape.The method can be calculated only by relying on the relative motion of adjacent frames without creating any model and it does not acquire the background of the image in advance. It carries not only the motion information of the moving object, but also rich information about the 3D structure of the scene.

It is reasonable to calculate the gait cycle by the periodic variation of the offset *v* in the optical flow method over the entire gait sequence. Therefore, the gait circle can be calculated by algebra of the *v* component in the gait flow field *t* and T(*t*).

$$T(t)=\sum_{x=1}^{W}\sum_{y=1}^{H} f_v(x,y,t)$$

In this formula, *x* and *y* indicate the sequence number of the row and column, W and H represent the number of rows and columns, $f_v(x, y, t)$ represents the optical component *v* of two neighbour frame at the coordinates(*x,y*). The graph below shows the periodic variation of T(t). Frames between every two adjacent peaks or valleys belong to the same period. The other one shows the process of GEI generation.

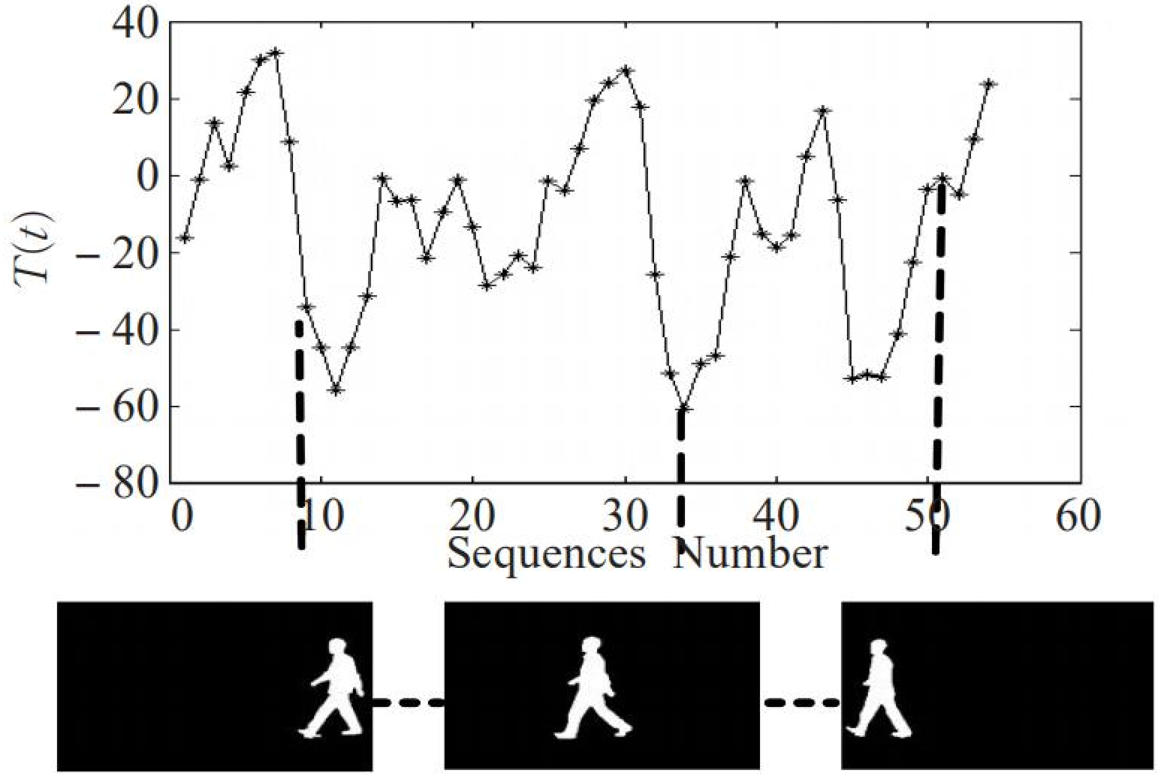


Figure 6. Calculation Processes of Gait Circle

3.3 Gait feature Generation Method

3.3.1 Local Binary Pattern(LBP)

LBP is an image description operator, which can effectively express the local texture features of the image. Compared with other image description operators, it has significant advantages such as strong classification ability, high computational efficiency, rotation invariance and gray invariance. It was originally proposed by Ojala et al. in 1996 to study the extraction of texture features. The earliest LBP operator is defined as being in a 3×3 grid. The gray value of the center point of the grid is used as the threshold. The gray values of the remaining 8 pixels in the grid are compared with each other in turn. Like it show in the graph below ,if the threshold is less than one of the adjacent 8 pixels in the grid, the position of the pixel is marked as 1, and if the threshold is greater than a certain pixel, the flag is 0.

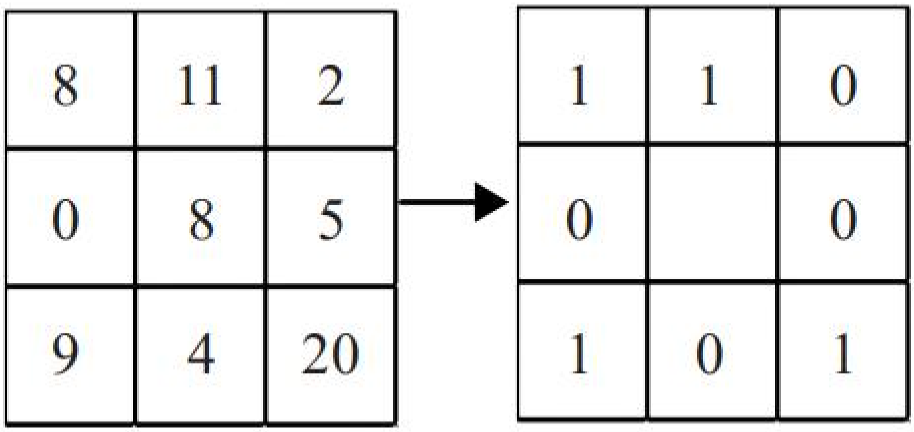


Figure 7. LBP Feature Generation

Thus, in a 3×3 grid, the threshold value is compared with the values of eight adjacent pixel points one by one, and the corresponding eight binary numbers are generated, and eight binary numbers are sequentially connected and converted into a decimal number. It was 11001010 in binary. After the transformation, it is 202.

Given a GEI, the LBP operator calculates the values of all the pixel points of the image to form a series of LBP codes, and then extracts the image features, which are represented by the feature histogram like the graph below.

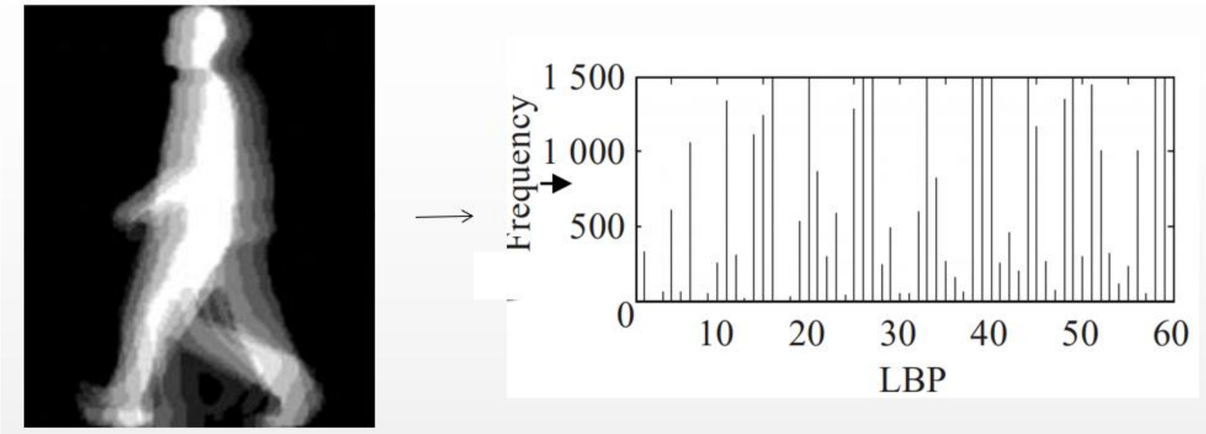


Figure 8. Presentation of LBP Feature

3.3.2 Histogram of Oriented Gradient (HOG)

HOG was first proposed by French researcher Dalal. It is a description operator that reflects the local gradient direction and gradient intensity distribution of the image. The main idea of it is that the distribution of the gradient or edge direction of the image is calculated to reflect the representation and shape of the partial image object. In practical applications, the image is first divided into several small connected regions, each of which is called a cell unit. Gradient size and edge direction of all pixels in each cell would be calculated to generate the histogram. At last the HOG feature of the whole image would be generated by connecting all the histogram that every cell has generated before. Since HOG is calculated on each cell of the image, it maintains good invariance for both geometric and optical deformation.

In the aspect of human target recognition, since the collected gait data is diverse, the gait energy map generated by the human gait image in one cycle is also very different, so the information extracted on the gait energy map is very limited. Edge contour information is very important in gait recognition. HOG is an efficient operator for obtaining the boundary information of the target image, which is also very sensitive to edge shape and edge gradient direction. The relationship between the local pixel points of the image can be better represented by dividing the cell and block methods. It can be suitable and accurate features can be extracted in different GEIs for matching. The HOG operator quantifies the position and direction space and, to a certain extent, eliminates the effects of image position changes. There are four steps for HOG feature extraction:

Step 1- Normalize the color space. It eliminates interference with the captured target image by uncontrollable factors such as the shooting environment, background and device.

Step 2- Calculate the gradient.

We know that there is a gradient on every pixel of the image, both horizontally and vertically. Suppose the gradient in the horizontal direction is Gx and the gradient in the vertical direction is Gy. Gradient can be calculated by the formula below:

$$G_x = f(i, j+1) - f(i, j-1)$$

$$G_y = f(i+1, j) - f(i-1, j)$$

To calculate the orientation of the point, it is needed to first calculate the tangent angle formed by the gradient.

$$tan\theta = G_y/G_x \rightarrow \theta = tan^{-1}(G_y/G_x)$$

Step 3- Divide the image into many cells unit. There is no overlap between adjacent cells. Calculate all the gradients in each cell, divide all the gradient directions into 9 bins, accumulate the gradient magnitudes of each corresponding bin and finally normalize the histogram to get the feature of each cell.

Step 4- Normalize overlapping block histogram. Combine several cell units into one block. The feature histograms of all the cells in a block are connected in series to form a characteristic histogram of a block, and all block feature histograms are connected in order to obtain the HOG feature of the entire image.

3.2.3 Hierarchical Feature

The LBP feature operates on the values of grayscale image pixels. The HOG feature mainly uses the gradient size and direction of the pixel. Therefore, after first round feature extraction applying LBP and HOG operator, there is still an image with a gray scale change value. In order to obtain richer and more useful texture information and edge shape information from the grayscale image, it can be implement that a hierarchical LBP and HOG can be generate like the picture below.

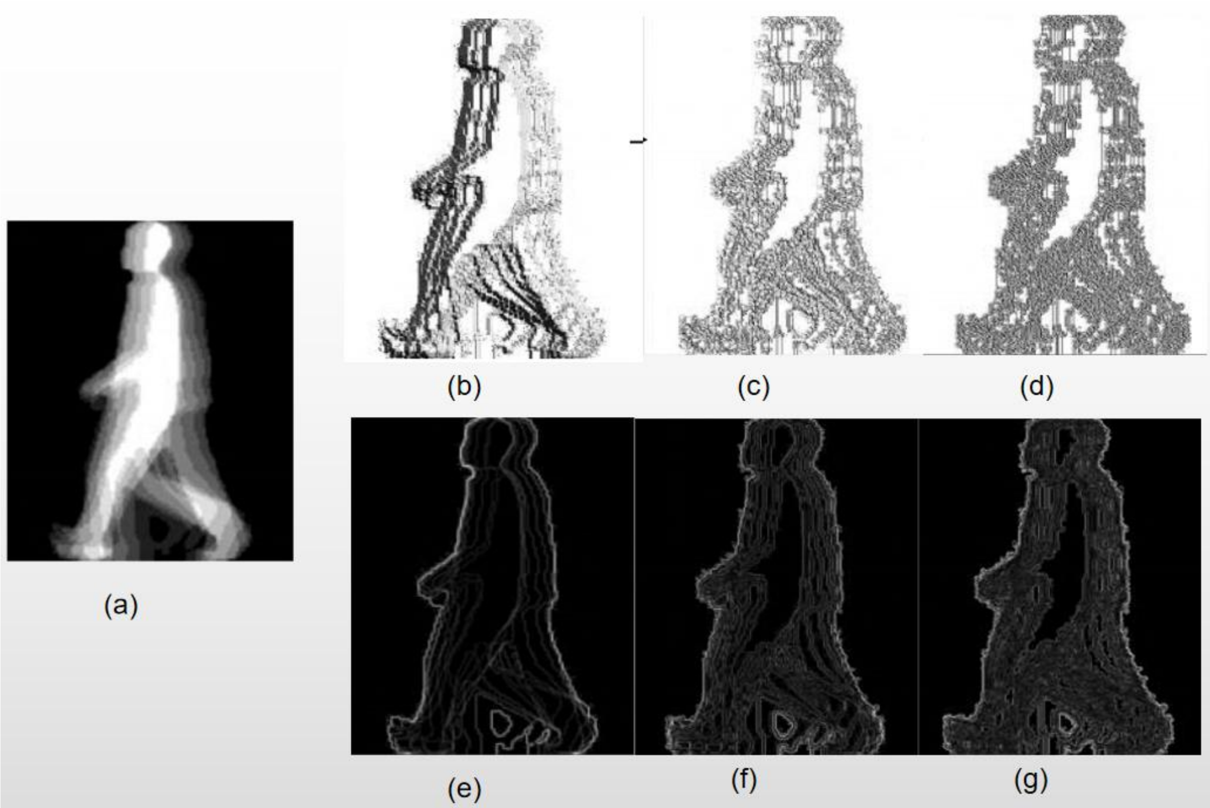


Figure 9. Hierarchical LBP and HOG Feature

As it shows in the picture, all of them contain useful information for further process. (b) can clearly express the human gait texture structure with strong texture information and contour information. (e) Clearly indicate the edge information of the human gait. (c) and (f) are weaker than (b) and (e), but there are still some useful information that can be utilized. Therefore, the hierarchical feature can obviously provide more information.

3.2.4 Combination of Features

The traditional gait recognition algorithm extracts gait features are relatively simple. Besides, due to the influence of occlusion, light and other interference factors, the gait recognition rate is usually low. Therefore, information fusion theory is widely used in gait recognition. Choudhury combines Platts analysis with elliptical Fourier descriptors to match human contours for gait recognition. The method combines the temporal and spatial motion characteristics, statistical and physical parameters of the human body to analyze the contour shape of the human body. However, it only considers the contour features of the human body and ignores the internal features. Derlatka proposes a biometric system based on different gait data from the human body, combining dynamic information (ground reaction) and static information (human measurement data such as:width of the trunk and buttocks, length of the thigh and height). This method requires a cumbersome feature extraction process and complex information input,which leads to a complicated calculation process and cannot meet real-time needs. The combination method in this paper can describe not only local texture information but also the edge contour information and it is simple and fast to calculate.

The LBP feature image and the HOG feature image of the first two layers contain relatively clear texture features and contour edge features. As the number of extracted layers increases, less and less useful information is available in the image. The information of the third layer is obviously blurred. Therefore, this paper divides the LBP and HOG feature images extracted by 3 layers, each layer contains useful information. If the information of different layers is synthesized into new features, the extracted information would be more complete and effective. Therefore, this paper uses feature fusion methods, namely hierarchical LBP and layered LBP-based hierarchical HOG feature fusion. The principle of the method is to first extract the LBP features three times for the GEI map, and obtain the LBP histogram of each layer and the LBP image of each layer, and then extract the corresponding HOG features on the LBP image of each layer to obtain each layer. Based on the HOG histogram and the HOG feature image of the LBP image, the LBP feature histogram of each layer and the HOG feature histogram based on the LBP image are sequentially connected in series to obtain a fusion feature histogram of each layer, and finally the fusion of the three layers. The feature histograms are sequentially cascaded to form the final fusion feature.

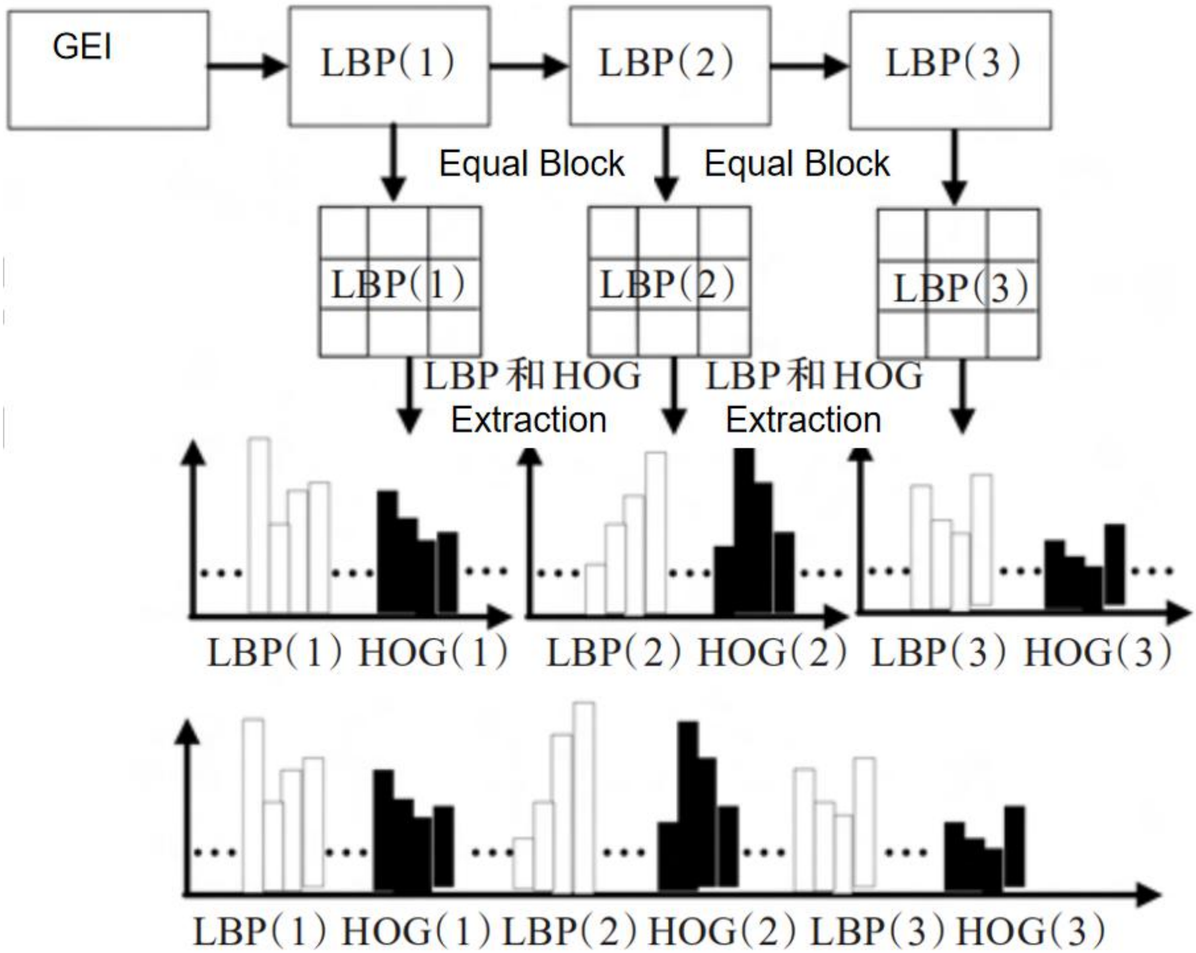


Figure 10. Flow Chart of Combination LBP and HOG Feature Extraction

4. Result

4.1 Database

Table 1 shows some common databases that have been established in the research field. Experimental researches need to select different databases according to different research directions.

Database	Subject Number	View Angel	Clothing Change	Carrying Change
OU-ISIR MVLP	10307	14	NO	NO
OU-ISIR LP	4016	4	NO	NO
OU-ISIR LP-Bag	2070	1	NO	YES
CASIA--B	124	11	YES	YES
USF	122	2	YES	YES

Table1. Existing Gait Recognition Database

At present, these databases generally possess a small collection range, a lack of universality and wide applicability. The current large gait database is the Soton Big Gait Database created by the University of Southampton in the UK and the Multi-view gait database (CASIA-B) and infrared gait database (CASIA-C) created by the Institute of Automation, Chinese Academy of Sciences. In addition, the CASIA-C database is currently the only publicly available large-scale infrared gait database. The USMT gait database is one of the few 3D gait libraries that currently available.

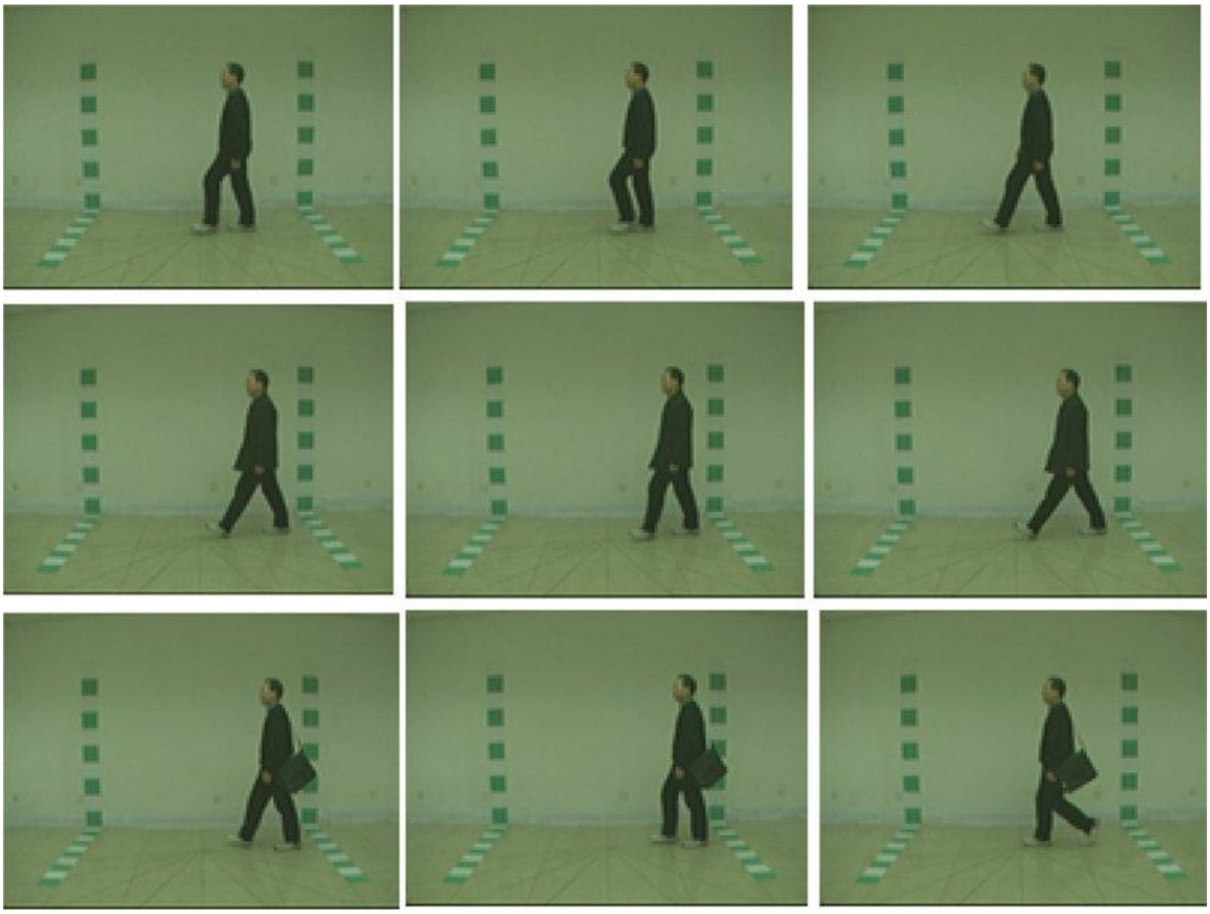


Figure 11. Samples of CASIA-B Database Adopted in This Paper

In this paper, the experiments are mainly processed on CASIA-B dataset. It is a big dataset with 124 subjects in 11 view angel. Every subject provides 3 walking condition in 10 video with 6 normal, 2 clothes changed and 2 walking with a bag. The resolution of each video is 320 by 240, 25 frames per second.

The working environment is Intel(R) Core(TM) i7-8750H CPU @2.20 GHz, 32.0 GB RAM, Nvidia 1070Ti GPU, Windows 10 operating system, Python 3.6 + OpenCV v2.4, Pycharm. In addition, some gait video of researchers were taken for test and verification.

Figure 12. Sample of Self-taking dataset Adopted in This Paper

4.2 Human Silhouette Segmentation and GEI Generation

This paper adopts frame difference method to generate silhouette images. After the closure operation, it can achieve a better standard showing in picture 13.

(a) Grey Scale of Human Gait Image

(b) Image of FDM extraction

(c) Silhouette Image after Closure Operation

Figure 13. Human Silhouette Generation

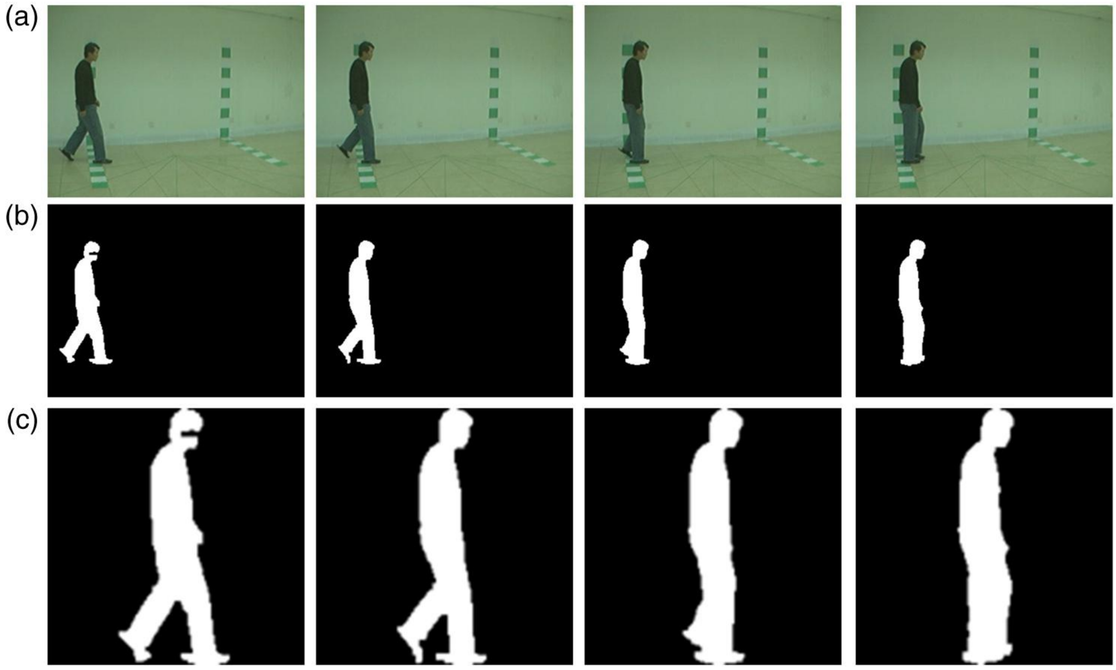


Figure 14. Samples of Human gait silhouette Generation

4.3 GEI Generation

As it mentioned in section 2, GEI image can be generated using segmented silhouette images, the GEI result is showed in picture 15. Each line represents GEI in the same perspective of different people. Each column represents the same person from different perspectives of GEI.

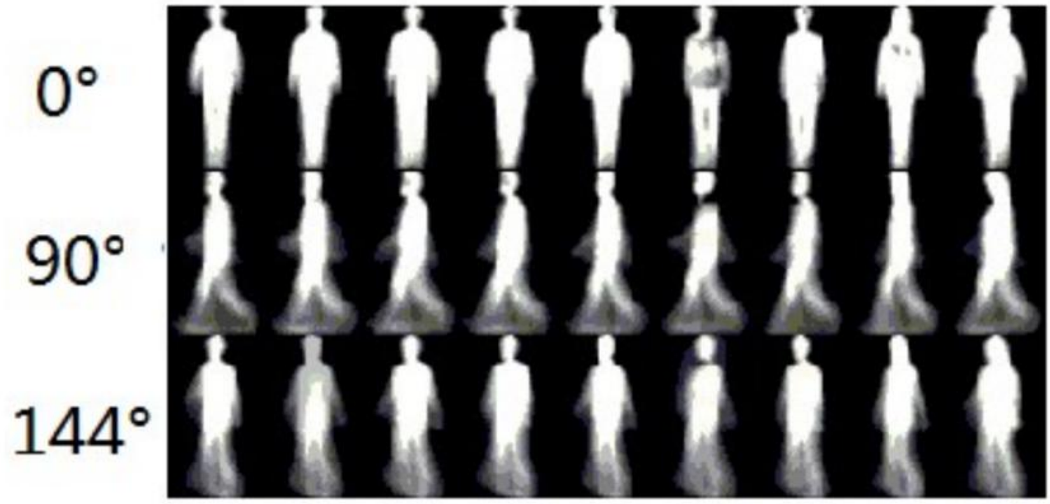


Figure 15. GEI Results

4.4 Feature Extraction

As it mentioned in section 3, a feature that combined with HOG feature and LBP feature is adopted in this paper.

Extracting feature from GEI for *m* times can obtain *m* level LBP feature images like picture 16. LBP(*m*),*m*=1,2,3.

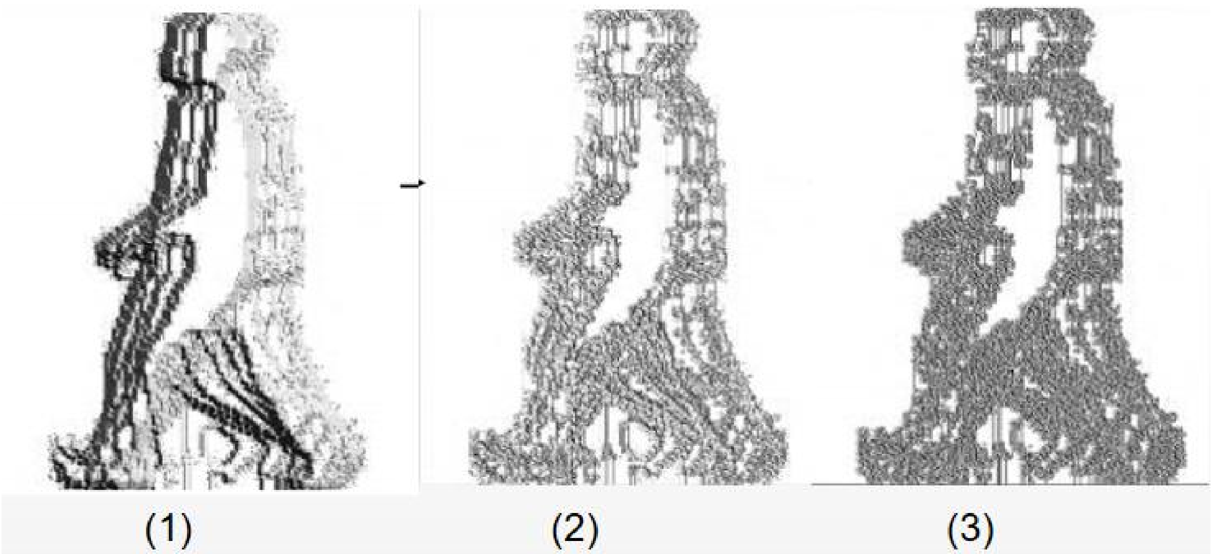


Figure 16. LBP feature Extracted for GEI

Dividing LBP feature into same size of block without overlapping and calculating the LBP feature histogram of all sub-blocks and sequentially concatenating all histograms to obtain the final histogram of each LBP feature image, which is, the final feature. Extracting all the HOG of the sub-block and sequentially concatenating all histograms to obtain the final histogram of each HOG feature image,which is, the final feature. Fusing each level of the LBP feature with HOG feature could obtain the final fusion feature. Sequentially concatenating all the combined feature together provides the final feature of this GEI which shows in picture 17.

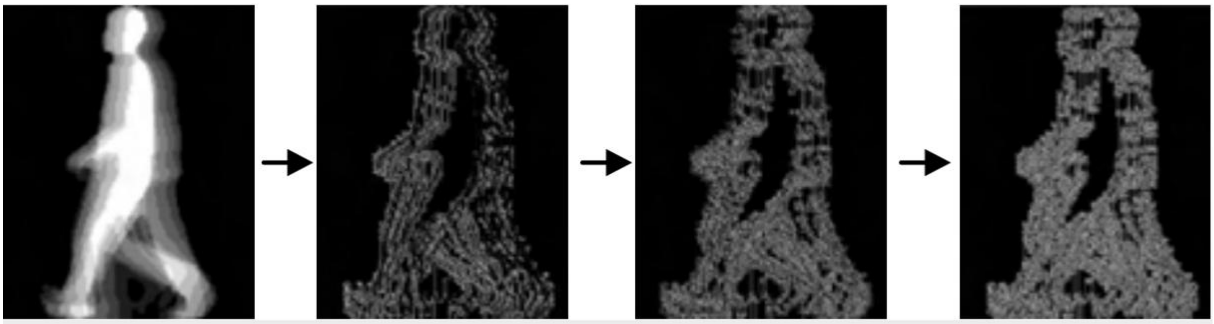


Figure 17. Final Fusion Feature

4.5 Recognition and Classification

After feature extraction, the extracted feature will be sent into CNNs for train and test. Picture 18 shows the framework of the method.

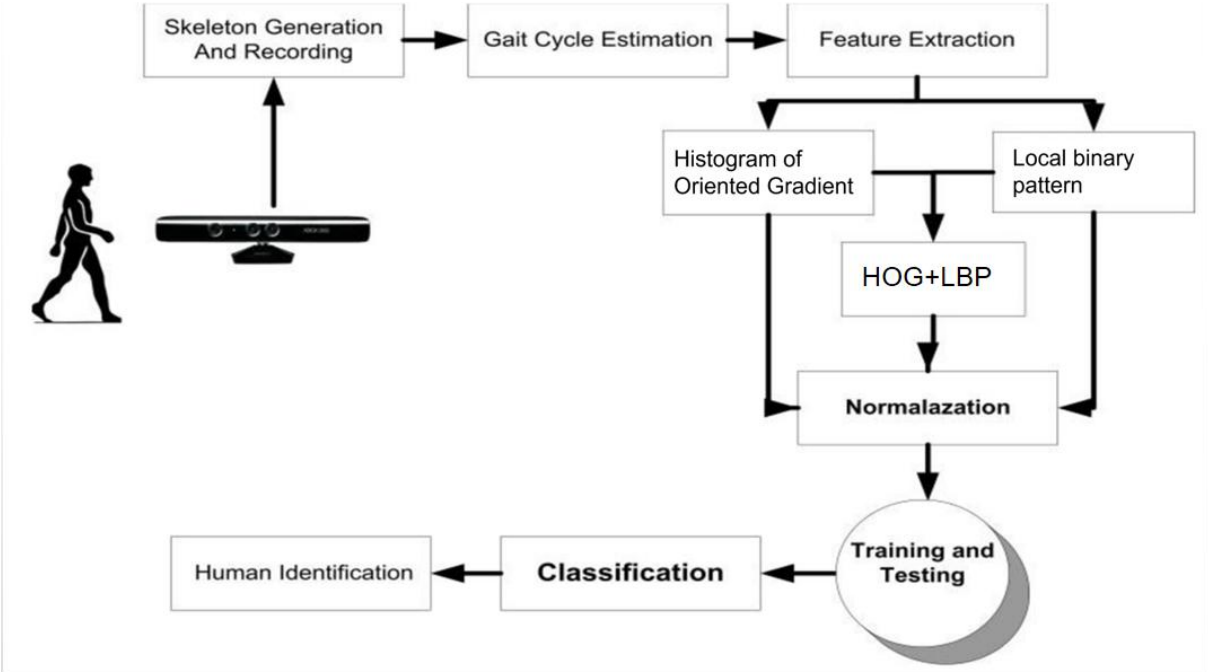


Figure 18. Frame work of This Method

The feature learning network of this paper consists of a convolutionally constrained Boltzmann machine and a fully connected layer network. Different from the supervised learning method of convolutional neural networks, the convolution-limited Boltzmann machine is an unsupervised learning model, which can learn the network parameters by comparing the divergence algorithm, and complete the feature extraction of the input image for identification and classification. The network proposed in this paper uses the GEI+LBP+HOG feature as input and convolves it with 3*3 and 7*7 convolution kernels to obtain the asynchronous features at different scales. At the same time, after the convolution-limited Boltzmann machine obtains the gait feature, the features acquired by the two different convolution kernels are merged using the fully connected network for feature learning and the feature is used at the end of the fully connected network using softmax for classification and identification as it shows in picture 19.

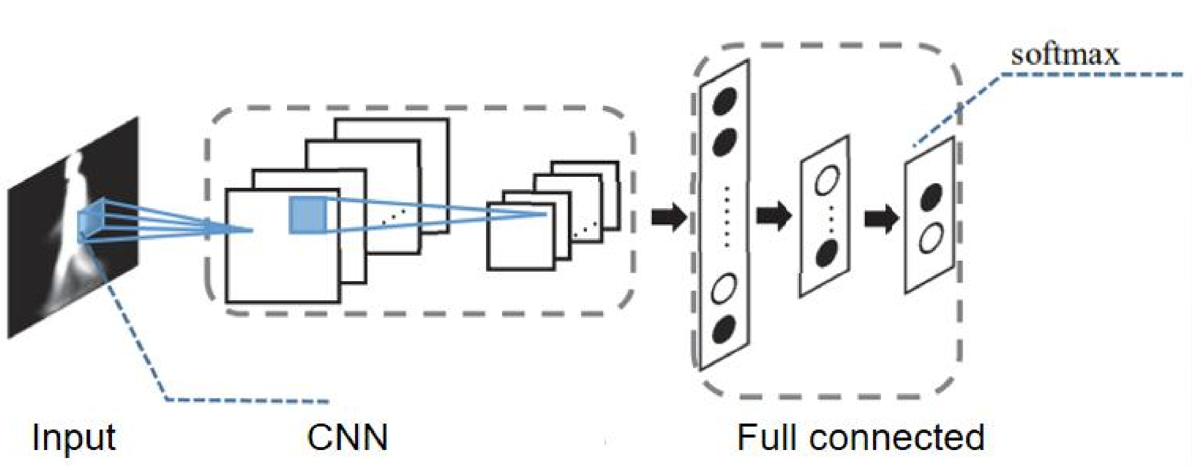


Figure 19. Network Structure

First, feature extraction is performed using a convolutionally constrained Boltzmann machine; the extracted gait features are trained using a fully connected network and the features are finally classified using the softmax layer to obtain prediction results; then the prediction results are compared with the real results. Calculate the cross entropy loss, which is the error between the two. At the same time, the back propagation algorithm is used to optimize the network parameters according to the error of the two.

4.6 Experimental Strategy

The same state experiment and cross state experiment were set in the experiment to study the effectiveness of the algorithm. In the same state experiment, the walking posture of the training data is the same as the walking posture of the test data, and the set experiment (training set-test set) includes (nm-nm), (bg-bg), (cl-cl). In the cross-state experiment, the walking posture of the training data is different from the walking posture of the test data, and the set experiment (training set-test set) includes (nm-bg), (nm-cl). Four video sequences in the nm data set are used in training (nm-nm) and other data is used in the test. In (bg-bg) and (cl-cl), half of the data is used as a training set and half is used as a test set. In the cross-state experiments (nm-bg) and (nm-cl), four video sequences in nm were taken as training sets, and all data in bg and cl were used as test sets. The following table show the experiment strategy above.

Train-Test	Training Set	Testing Set
nm-nm	124*4	124*2
bg-bg	124*1	124*1
cl-cl	124*1	124*1
nm-bg	124*4	124*2
nm-cl	124*4	124*2

Table 2. Experiment Strategy

Two convolution kernels of different sizes were set in the convolution-limited Boltzmann model training, which are 3*3 and 7*7, respectively, and there are 32 convolution kernels of each size. By this, kernel size impact on the performance would be discovered.

The calculated gait energy map is input to the neural network for learning training, and the reconstruction error of the convolution-limited Boltzmann machine is shown in the picture 20.

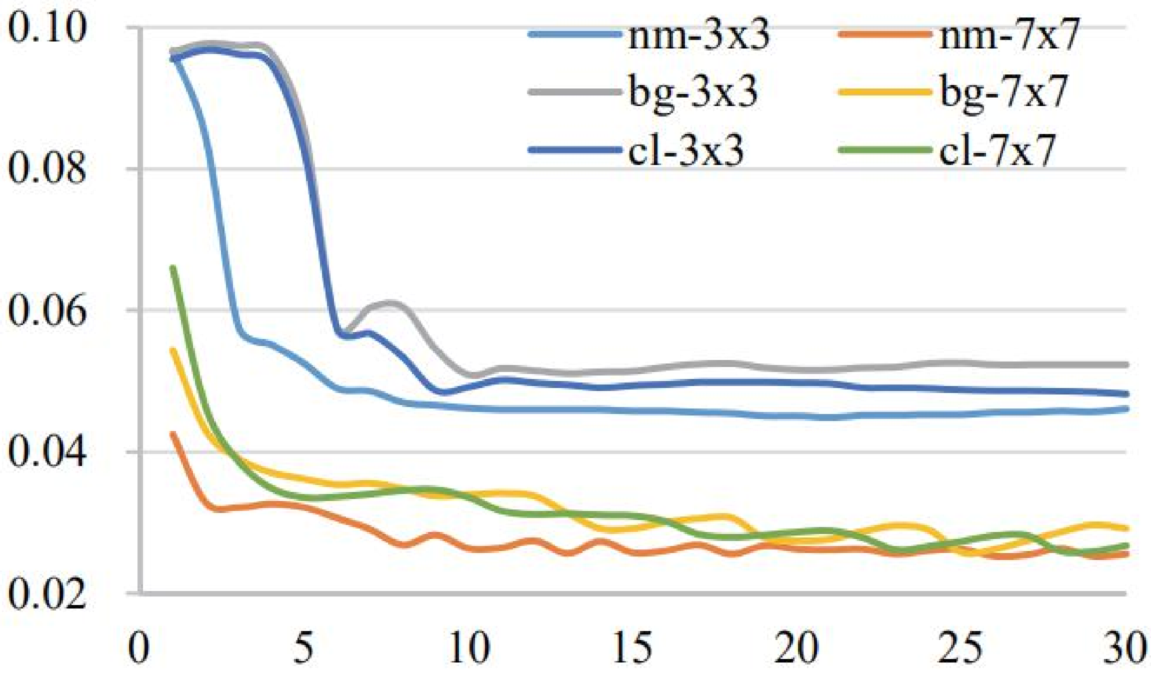


Figure 20. Reconstruction Error

It can be seen from the figure that the reconstruction error is large at the beginning because the parameters are randomly initialized. After several iterations of parameter updating, the error rapidly decreases and gradually tends to a certain value or fluctuates within a small range.

The training of the fully connected layer network is based on the cross entropy loss to iteratively update the network parameters to obtain an optimized network model. During the training process, the training error curve is shown in the figure 21. After several iterations, the training error gradually decreases and tends to be stable, indicating that the model converges.

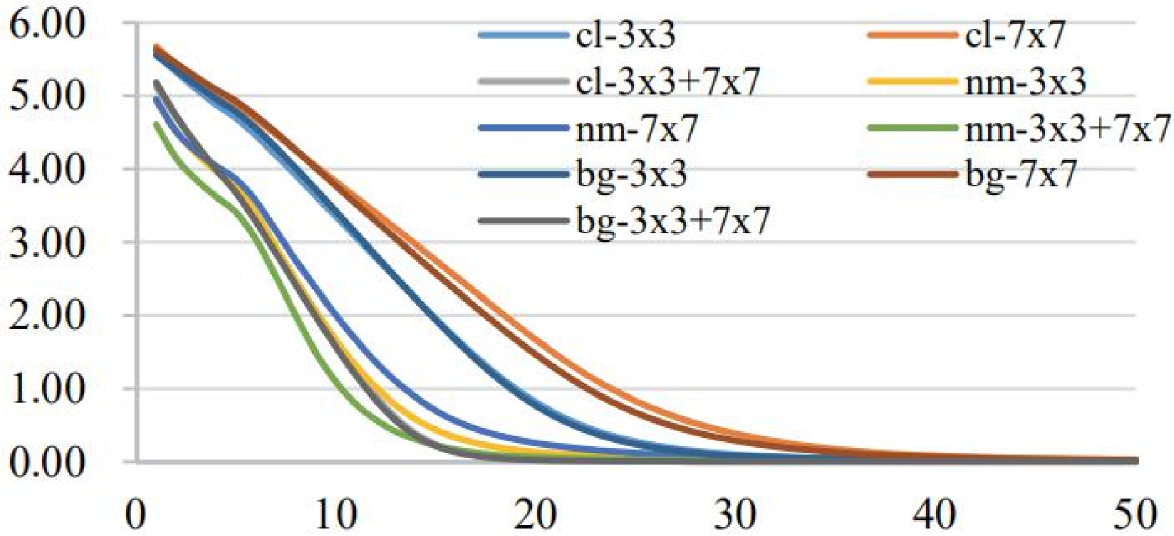


Figure 21. Training Error Curve

Verification of this method had also be tested on the self taking database. Figure 22 shows some of the correct result and figure 23 shows some error result of this method.

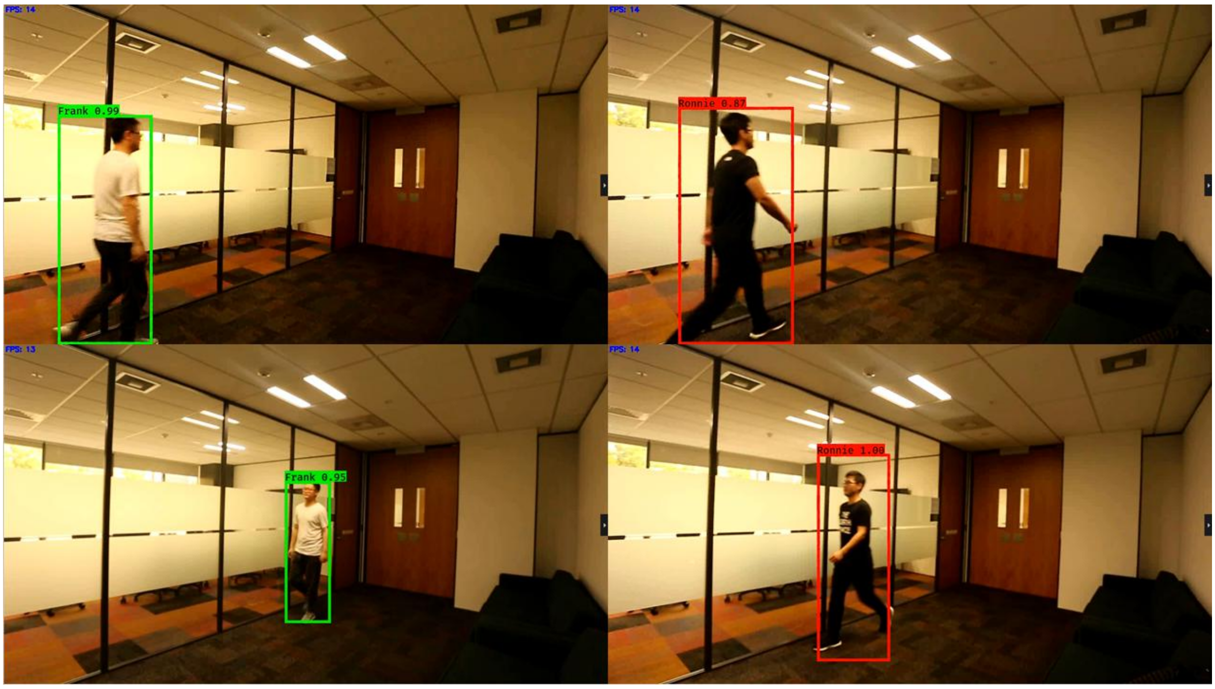


Figure 22. Some of Correct Result

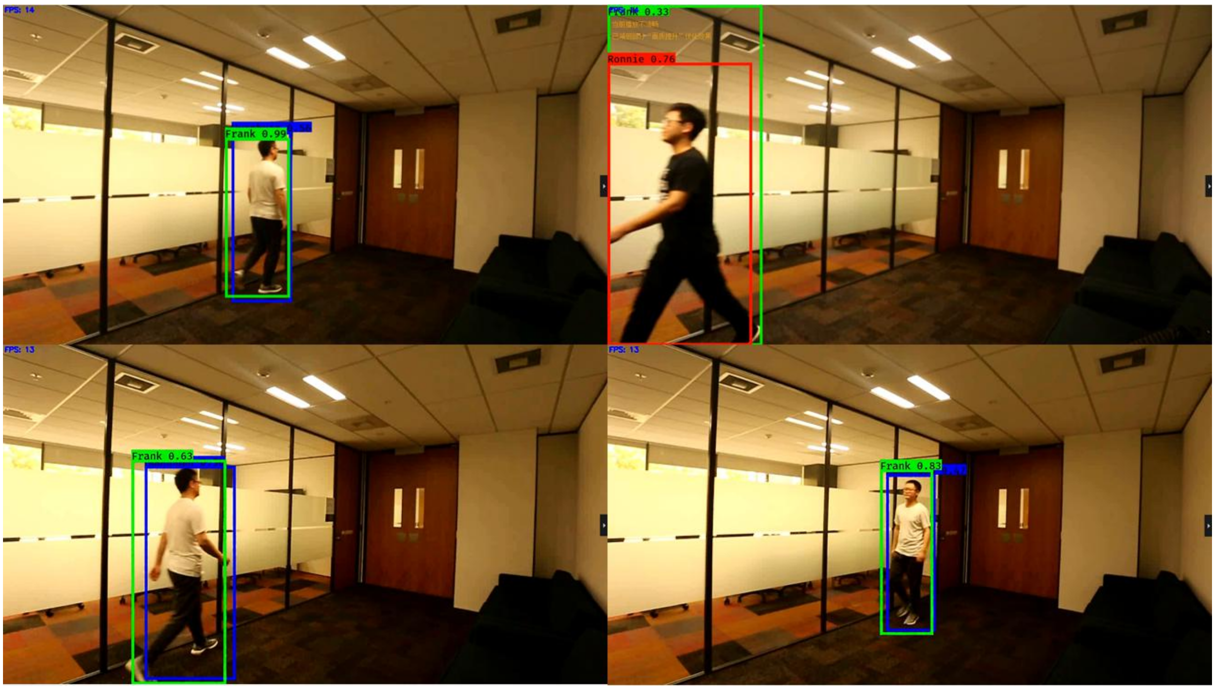


Figure 23. Some of the incorrect Result

5. Analysis and Discussion

5.1 Feature Extraction

The training parameters and direct results have been introduced in the last section. Some of the video result of test were also provided above. In the fusion feature algorithm, the LBP(m), m = 1, 2, 3 feature images are equally divided into several sub-blocks of the same size and without overlapping. In this experiment, the recognition rates of the algorithm on the three gait data sets are compared under the 2×4, 4×4, 4×8, 8×8, 8×16, 16×16, 16×32 blocks. The best block is obtained and the results are shown in the figure.

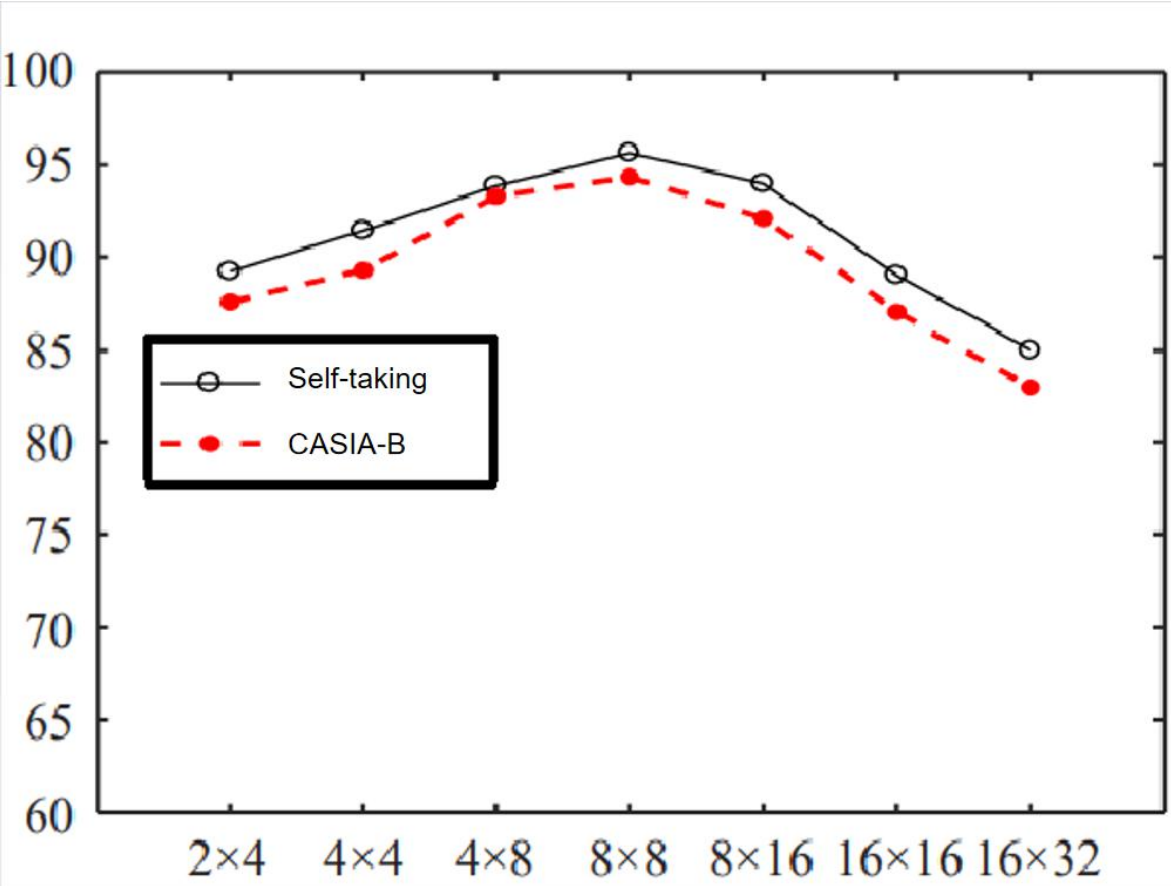


Figure 24. The Effect of Block Size on Accuracy

It can be seen from the figure that the optimal block is 8×8. Blocking larger or smaller will affect the recognition rate. Because oversized blocks cannot extract accurate and appropriate local features. Meanwhile, too small blocks are susceptible to image registration and human motion, making the extracted global and local features inaccurate.

5.2 Correct Classification Result

In this paper, when evaluating the performance of the algorithm, the "Leave a Cross-Validation Method" is used to obtain an unbiased estimate of the Correct Classification Rate (CCR), that is, one sample sequence is reserved for each experiment as the test sample, and the remaining sample sequence is used. As a training sample of NN, it is classified according to the similarity of training samples. In order to compare the advantages of the fusion features of this paper, the recognition rates of Original-GEI, GEI+LBP+PCA, Bayes+HMM, and Integrated-HOG were compared. The results are shown in the table.

Method	Correct Classification Rate
Original-GEI	74.62
GEI+LBP+PCA	82.21
Bayes+HMM	87.34
Intergrated-HOG	87.7
This Method	92.36

Table 3 CCR of Different Feature Extraction Methods

Besides, Cumulative Match Score is also applied for evaluation. The CMS is defined as the cumulative probability of the actual category of a test metric between its previous Rank match value (here Rank = 10). CCR is the cumulative recognition rate when Rank is equal to 1. The CMS for the five methods are shown separately.

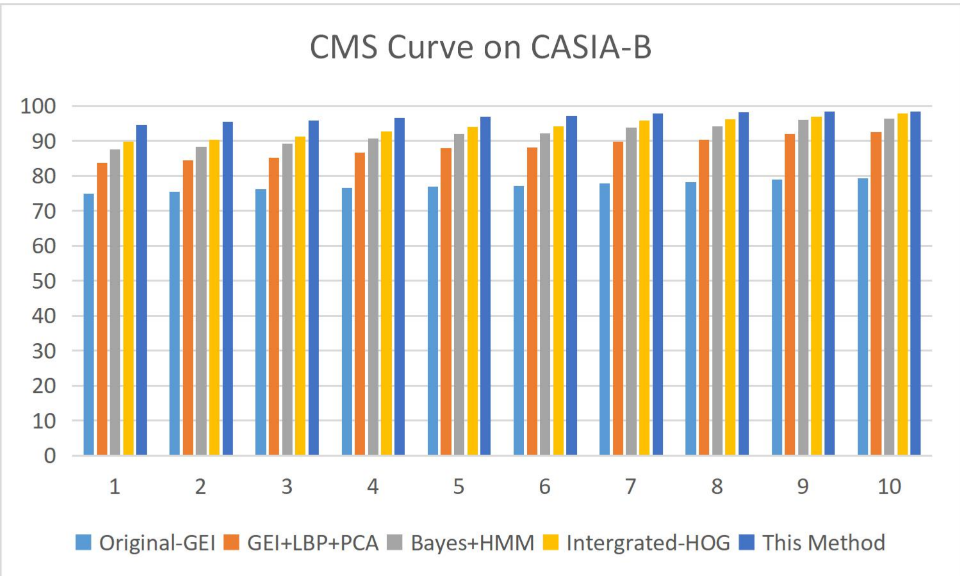


Table 4. CMS curve of Different Method

Train-Test	nm-nm	bg-bg	cl-cl	nm-bg	nm-cl	Average
GPPE	93.4	62.2	55.1	56.1	22.4	57.8
GEnI	92.3	65.3	55.1	56.1	26.5	59.1
STIPs	95.4	73	70.6	60.9	52	70.4
CNN	100	89.1	95.9	30.6	12.3	65.6
3*3	97	84.5	87.9	37.5	16.4	64.7
7*7	97.8	89.7	95.7	40.5	17.2	68.2
3*3+7*7	98.3	90.5	95.7	40.5	18.1	68.6

Table 5. Cross-State Test Results

As can be seen from the above chart, the recognition result based on the convolution kernel of the CASIA-B data set is better than the recognition result of the 3*3 convolution kernel, and the feature fusion of the 3*3 and 7*7 convolution kernels. The subsequent recognition accuracy is better than or equal to the recognition accuracy of a single convolution kernel. At the same time, this paper compares the proposed method with other existing methods GPPE, GEnI, STIPs and deep learning CNN based methods. It can be found from the table that the proposed method is slightly lower than the existing deep learning-based methods in the same state experiments (nm-nm) and (cl-cl), mainly because it uses a deeper network. Structure, which is somewhat better at characterizing. However, in the experiment of cross-state, the method is better than it, indicating that the method of this paper is better than its robustness. In addition, the experimental accuracy of cross-state is lower than other existing methods, mainly because traditional methods can be identified by targeted design features, while deep learning requires more walking posture during training to increase the generalization ability of the network. . However, only one kind of walking state in the experimental training set leads to poor generalization ability of the network, and thus the experimental result of the cross state is not better than the existing method.

6. Conclusion and Future Work

6.1 Conclusion

This paper proposes an identifiable algorithm based on gait energy map for layered fusion of LBP and HOG features. The method uses the layering idea to overcome the defects of the gait energy map, not only captures the local information of the human target image, but also extracts its contour and shape information. The fusion of the two kinds of information is on the CASIA and USF gait data sets. Verification, the recognition algorithm of this paper has achieved a high recognition rate.

And this paper proposes a feature learning network consisting of a convolutionally constrained Boltzmann machine and a fully connected layer network. The network first uses feature-constrained Boltzmann machines with different convolution kernels to extract features from gait images, because convolution-limited Boltzmann machines are unsupervised, so that they can learn more extensively and more abundantly. Gait information. Although rich gait feature information is learned, there must be redundant and useless information. Therefore, this paper uses a supervised all-connected layer network to train and learn these features deeper and use the softmax layer for classification decisions.

6.2 Future Work

(1) In the single-person, single-scene laboratory conditions, the detection and

extraction of gait is relatively simple. However, in complex practical scenarios, the complexity of factors such as background and multiple individuals makes it difficult to accurately acquire gait. Inaccurate gait information directly affects the final recognition efficiency and accuracy. Therefore, how to accurately and efficiently obtain gait images from complex scenes is the basis of gait recognition research.

(2) Gait recognition mainly uses different postures of each individual to perform identity authentication. In the existing research, the full-cycle gait images are used for feature extraction, but there are problems such as occlusion and frame loss in the actual scene. It is not easy to obtain a full-cycle gait image. Therefore, how to construct a highly efficient gait recognition algorithm framework based on a small number of gait images to improve the efficiency and accuracy of recognition has become the focus of gait recognition research.

(3) With many biometrics being applied in daily life, such as iris recognition technology, fingerprint recognition technology, voiceprint recognition technology, face recognition technology, etc., these biometric recognition technologies bring many things to people's daily lives. Convenience and security. However, any biometric technology can have certain disadvantages and applicable scenarios. Therefore, how to integrate multiple biometric technologies to adapt to different identification needs will also be one of the research contents.

Reference

Ahonen, T., Hadid, A., & Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. IEEE Transactions on Pattern Analysis & Machine Intelligence, (12), 2037-2041.

Bashir, K., Xiang, T., & Gong, S. (2009). Gait recognition using gait entropy image.

Begg, R. K., Palaniswami, M., & Owen, B. (2005). Support vector machines for automated gait classification. IEEE transactions on Biomedical Engineering, 52(5), 828-838.

Boykov, Y. Y., & Jolly, M. P. (2001). Interactive graph cuts for optimal boundary & region segmentation of objects in ND images. In Proceedings eighth IEEE international conference on computer vision. ICCV 2001 (Vol. 1, pp. 105-112). IEEE.

Bulat, A., & Tzimiropoulos, G. (2016, October). Human pose estimation via convolutional part heatmap regression. In European Conference on Computer Vision (pp. 717-732). Springer, Cham.

Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 7291-7299).

Chen, C., Liang, J., Zhao, H., Hu, H., & Tian, J. (2009). Frame difference energy image for gait recognition with incomplete silhouettes. Pattern Recognition Letters, 30(11), 977-984.

Cunado, D., Nixon, M. S., & Carter, J. N. (1997, March). Using gait as a biometric, via phase-weighted magnitude spectra. In International Conference on Audio-and Video-Based Biometric Person Authentication (pp. 93-102). Springer, Berlin, Heidelberg.

Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In international Conference on computer vision & Pattern Recognition (CVPR'05) (Vol. 1, pp. 886-893). IEEE Computer Society.

Derlatka, M., & Bogdan, M. (2015, June). Fusion of static and dynamic parameters at decision level in human gait recognition. In International Conference on Pattern Recognition and Machine Intelligence (pp. 515-524). Springer, Cham.

Foresti, G. L. (1999). Object recognition and tracking for remote video surveillance. IEEE Transactions on circuits and systems for video technology, 9(7), 1045-1062.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 2961-2969).

Jamieson, K., Balakrishnan, H., & Tay, Y. C. (2006, February). Sift: A MAC protocol for event-driven wireless sensor networks. In European workshop on wireless sensor networks(pp. 260-275). Springer, Berlin, Heidelberg.

Jeevan, M., Jain, N., Hanmandlu, M., & Chetty, G. (2013, September). Gait recognition based on gait pal and pal entropy image. In 2013 IEEE International Conference on Image Processing (pp. 4195-4199). IEEE.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. Perception & psychophysics, 14(2), 201-211.

Kozlowski, L. T., & Cutting, J. E. (1977). Recognizing the sex of a walker from a dynamic point-light display. Perception & psychophysics, 21(6), 575-580.

Kusakunniran, W. (2014). Attribute-based learning for gait recognition using spatio-temporal interest points. Image and Vision Computing, 32(12), 1117-1126.

Lam, T. H., Cheung, K. H., & Liu, J. N. (2011). Gait flow image: A silhouette-based gait representation for human identification. Pattern recognition, 44(4), 973-987.

Lee, C. S., & Elgammal, A. (2004, May). Gait style and gait content: bilinear models for gait recognition using gait re-sampling. In Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings.(pp. 147-152). IEEE.

Lee, H., Grosse, R., Ranganath, R., & Ng, A. Y. (2009, June). Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In Proceedings of the 26th annual international conference on machine learning (pp. 609-616). ACM.

Lee, L., & Grimson, W. E. L. (2002, May). Gait analysis for recognition and classification. In Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition (pp. 155-162). IEEE.

Little, J., & Boyd, J. (1998). Recognizing people by their gait: the shape of motion. Videre: Journal of computer vision research, 1(2), 1-32.

Liu, Z., Malave, L., Osuntogun, A., Sudhakar, P., & Sarkar, S. (2004, August).

Toward understanding the limits of gait recognition. In Biometric Technology for Human Identification(Vol. 5404, pp. 195-206). International Society for Optics and Photonics.

Ma, G., Wu, L., & Wang, Y. (2017). A general subspace ensemble learning framework via totally-corrective boosting and tensor-based and local patch-based extensions for gait recognition. Pattern Recognition, 66, 280-294.

Murase, H., & Sakai, R. (1996). Moving object recognition in eigenspace representation: gait analysis and lip reading. Pattern recognition letters, 17(2), 155-162.

Muro-De-La-Herran, A., Garcia-Zapirain, B., & Mendez-Zorrilla, A. (2014). Gait analysis methods: An overview of wearable and non-wearable systems, highlighting clinical applications. Sensors, 14(2), 3362-3394.

Niyogi, S. A., & Adelson, E. H. (1994, November). Analyzing gait with spatiotemporal surfaces. In Proceedings of 1994 IEEE Workshop on Motion of Non-rigid and Articulated Objects (pp. 64-69). IEEE.

Ojala, T., Pietikäinen, M., & Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. Pattern recognition, 29(1), 51-59.

Rida, I., Boubchir, L., Al-Maadeed, N., Al-Maadeed, S., & Bouridane, A. (2016, June). Robust model-free gait recognition by statistical dependency feature selection and Globality-Locality Preserving Projections. In 2016 39th International Conference on Telecommunications and Signal Processing (TSP) (pp. 652-655). IEEE.

Sarkar, S., Phillips, P. J., Liu, Z., Vega, I. R., Grother, P., & Bowyer, K. W. (2005). The humanid gait challenge problem: Data sets, performance, and analysis. IEEE transactions on pattern analysis and machine intelligence, 27(2), 162-177.

Shin, J., Kim, S., Kang, S., Lee, S. W., Paik, J., Abidi, B., & Abidi, M. (2005). Optical flow-based real-time object tracking using non-prior training active feature model. Real-Time Imaging, 11(3), 204-218.

Shiraga, K., Makihara, Y., Muramatsu, D., Echigo, T., & Yagi, Y. (2016, June). Geinet: View-invariant gait recognition using a convolutional neural network. In 2016 international conference on biometrics (ICB) (pp. 1-8). IEEE.

Stringa, E., & Regazzoni, C. S. (2000). Real-time video-shot detection for scene surveillance applications. IEEE Transactions on Image Processing, 9(1), 69-79.

Tao, D., Li, X., Wu, X., & Maybank, S. J. (2007). General tensor discriminant analysis and gabor features for gait recognition. IEEE transactions on pattern analysis and machine intelligence, 29(10), 1700-1715.

Tan, D., Huang, K., Yu, S., & Tan, T. (2006, August). Efficient night gait recognition based on template matching. In 18th International Conference on Pattern Recognition (ICPR'06)(Vol. 3, pp. 1000-1003). IEEE.

Thomaz, C. E., Gillies, D. F., & Feitosa, R. Q. (2004). A new covariance estimate for Bayesian classifiers in biometric recognition. IEEE Transactions on circuits and systems for video technology, 14(2), 214-223.

Urtasun, R., & Fua, P. (2004, May). 3D tracking for gait characterization and recognition. In Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings. (pp. 17-22). IEEE.

Viola, P., Jones, M. J., & Snow, D. (2005). Detecting pedestrians using patterns of motion and appearance. International Journal of Computer Vision, 63(2), 153-161.

Wang, L., Ning, H., Tan, T., & Hu, W. (2003). Gait recognition by combining static and dynamic body biometrics.

Wang, L., Ning, H., Tan, T., & Hu, W. (2004). Fusion of static and dynamic body biometrics for gait recognition. IEEE Transactions on circuits and systems for video technology, 14(2), 149-158.

Wang, L., Tan, T., Ning, H., & Hu, W. (2003). Silhouette analysis-based gait recognition for human identification. IEEE transactions on pattern analysis and machine intelligence, 25(12), 1505-1518.

Xu, D., Huang, Y., Zeng, Z., & Xu, X. (2011). Human gait recognition using patch distribution feature and locality-constrained group sparse representation. IEEE Transactions on Image Processing, 21(1), 316-326.

Yan, C., Zhang, B., & Coenen, F. (2015, October). Multi-attributes gait identification by convolutional neural networks. In 2015 8th International Congress on Image and Signal Processing (CISP) (pp. 642-647). IEEE.

Yu, S., Tan, D., & Tan, T. (2006, January). Modelling the effect of view angle variation on appearance-based gait recognition. In Asian Conference on Computer Vision (pp. 807-816). Springer, Berlin, Heidelberg.

Yu, S., Tan, D., & Tan, T. (2006, August). 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) (Vol. 4, pp. 441-444). IEEE.

Yu, S., Tan, T., Huang, K., Jia, K., & Wu, X. (2009). A study on gait-based gender classification. IEEE Transactions on image processing, 18(8), 1905-1910.

Zhao, N., Zhang, L., Du, B., Zhang, L., Tao, D., & You, J. (2016, July). Sparse tensor discriminative locality alignment for gait recognition. In 2016 International Joint Conference on Neural Networks (IJCNN) (pp. 4489-4495). IEEE.

Zheng, S., Huang, K., & Tan, T. (2010). Translation-invariant representation for cumulative foot pressure images. arXiv preprint arXiv:1010.5426.

Zheng, S., Huang, K., & Tan, T. (2011, September). Evaluation framework on translation-invariant representation for cumulative foot pressure image. In 2011 18th IEEE International Conference on Image Processing (pp. 201-204). IEEE.

Zheng, S., Zhang, J., Huang, K., He, R., & Tan, T. (2011, September). Robust view transformation model for gait recognition. In 2011 18th IEEE International Conference on Image Processing (pp. 2073-2076). IEEE.