Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1683-1702
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

Research on Key Technology of Multi-View Gait Recognition Based on Neural Network

Chen Ma [1, *], Lei Lei [2]

[1] Research Center for Smart Policing and Big Data Technology, China People's Police University, Langfang, Hebei, 065000, China

[2] School of Fire Protection Engineering, China People's Police University, Langfang, Hebei, 065000, China

*Corresponding Author.

Abstract:

Gait recognition can be performed by taking videos at a distance, without using too many local details, overcoming some of the limitations of current biometrics. Gait recognition technology has obvious advantages, can overcome the shortcomings of current face recognition and other methods, and can be widely used in complex scenes. Improving the accuracy of gait identification will help to improve efficiency and reduce workload, especially in the efficiency of pedestrian identification and authentication at stations. The difficulty to be solved at present is the cross-perspective problem, also known as the multi-perspective problem. This paper will focus on this problem, through deep learning and other methods to carry out research, the main work is as follows: (1) The current research status of gait identification and multi-perspective issues at home and abroad was investigated. (2) Research on perspective conversion model based on generative adversarial network. In this paper, by training a generative adversarial network, the network can convert gait sequences from other perspectives to a unified perspective for recognition. (3) Research on gait recognition model based on human posture. This paper uses the human body joint length, joint angle, angular acceleration and other features extracted from the human body pose coordinates to form a feature matrix for identifying pedestrian identity.

Keywords*: Gait recognition, multi-view, generative adversarial network, neural network, long and short-term memory network.*

I. BACKGROUND AND MEANING

In recent years, due to the great prospect of biometric identification and the great room for improvement in current research, a large number of researchers have devoted themselves to the research direction of biometric identification. Biometric identification technology [1] is a technology that uses the physiological parameter characteristics and habitual gesture characteristics of pedestrians in various aspects to identify the identity of pedestrians. At present, identification systems based on static features such as faces [2] have been applied to station detection, identity verification and mobile payment, etc. However, identification in dynamic situations is very difficult.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1683-1702
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

In some places with complex situations, many people, various scenes, places that may require large-scale verification, etc., gait-based identification can completely identify the identity of pedestrians in distant places without disturbing pedestrians' walking at all. Greatly improve the recognition efficiency. In some places that require strict confidentiality, such as military sites, research institutes, and national projects, it can also quietly complete the task of identifying the identity. In gait identification, only the general posture of pedestrians needs to be seen without facial details, which can greatly save the cost of cameras. To sum up, the advantages of gait identification are mainly reflected in the following aspects:

1 Non-contact: gait identification does not require close contact with people [3], and only needs to shoot walking gait videos of pedestrians.

2. Robustness: The gait identification only needs to complete the pedestrian gait, and the camera distance and local clarity have very little influence on the recognition effect.

3 Safety: The gait feature [4] is determined by the appearance and posture of the human body and natural habits. It is difficult for people to completely disguise the characteristics of walking habits, footsteps, and hand-waving heights. Therefore, it is difficult to completely disguise the gait and it is safer.

II. INTRODUCTION TO BASIC KNOWLEDGE

2.1 Introduction to Image Processing Related Technologies

2.1.1 Image filtering

Image filtering [5], which is to retain the complete feature information on the basis of removing the noise of the image as much as possible, because the general natural image has a little noise, so this is a typical operation in image processing. The filtering effect can sometimes greatly affect the final image quality, and even greatly affect the final result.

The median filter method [6] replaces the gray value of the pixel with the intermediate value of the pixel gray value in the local range of a certain point from small to large. This method has a good effect on filtering high noise.

Mean filtering [7] refers to replacing the gray value of a pixel with the average value of the gray value in the local range of a certain point in the image.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1683-1702
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

2.2 Introduction to Deep Learning and Neural Network Related Technologies

2.2.1 Neural Network

The process of human processing information is to collect information through sensory organs, and then process the corresponding information through neurons, and then pass the processing results to the next neuron. After many repetitions, the information is processed accordingly and finally becomes comprehensible to the brain. information.

The artificial neural network is composed of neurons, each neuron can realize the corresponding function. The artificial neural cell is to imitate the biological neuron, so that the original simple function combination becomes a powerful function. The basic structure of artificial neuron is shown in Figure 1.

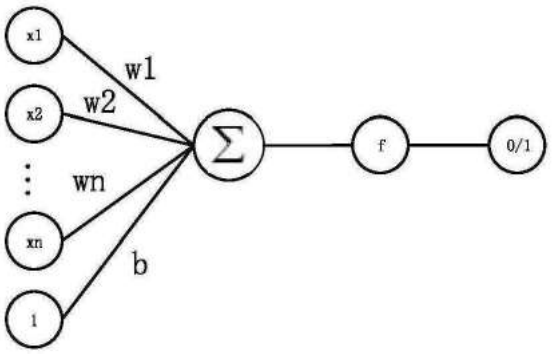

Figure 1 Neuron model

The ability of a neuron is single, and it cannot even solve the XOR problem or complete some complex function mapping. The researchers found that by organically combining multi-layer neural networks, the fitting ability can be greatly increased and the generalization properties of the model can be increased. With the combination and superposition of neurons, the function will become stronger and stronger, and the fitting ability will become stronger and stronger. In theory, if the training samples are rich enough and the network structure is reasonable, most functions can be fitted.

2.2.2 Convolutional Neural Networks

Convolutional neural network [8] is one of the foundations of deep learning, and most networks in deep learning currently contain convolutional neural networks. The biggest feature of the convolutional neural network is that it can directly input an image matrix, and can automatically extract features without manually extracting features. Its convolution feature can extract small and edge features in the image.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1683-1702
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

Convolutional neural networks have convolutional layers and pooling layers, which give the network more powerful fitting and generalization capabilities. The convolutional layer can abstract the local features in the image for recognition, and the pooling layer can make the model more generalizable. After the calculation of the fully connected layer and the output layer, the process of recognition and classification can be completed.

2.2.3 Long Short-Term Memory Networks

Long short-term memory network [9] (LSTM) has the characteristics of recurrent neural network and can also overcome the shortcomings of traditional recurrent neural network. LSTM network can memorize the calculation result of the previous neuron and act on the current neuron. The general neural network can be used for basic classification, but it is difficult to predict the next information based on the current information, while the recurrent neural network can memorize the previous results and use it to predict the current, which is very suitable for periodic sequence sample prediction. The specific structure of LSTM is shown in Figure 2.

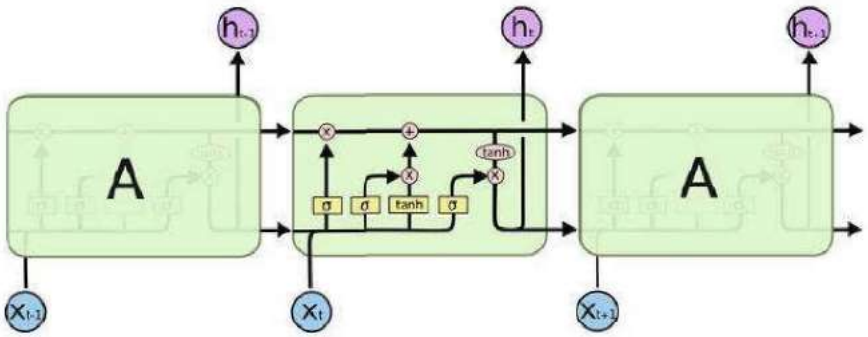

Figure 2. LSTM network diagram

2.3 Introduction to Generative Adversarial Networks

Generative Adversarial Networks (GANs) [10] are relatively new machine learning methods for generating samples, proposed by Ian Goodfellow. Generative Adversarial Networks (GANs) are networks that do not require a large number of annotations to give the network the ability to generate.

2.3.1 Generative Adversarial Network Applications

The principle of Generative Adversarial Network is to give the network a class of standards to allow the network to learn continuously, and finally allow the network to generate samples that meet such standards. In principle, as long as the network is well designed and the samples are sufficient and abundant, the results we want can be generated. At present, it has been well applied in many fields. The application examples of Generative Adversarial Networks are as follows: 1. You can use Generative Adversarial Networks to convert ordinary photos into a fantasy photo, 2. Combine the characteristics of

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1683-1702
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

different objects into the image naturally, 3. Superimpose zebra stripes on wild horses, and natural transitions make wild horses look like zebras , 4 Change the season feature, load the winter snow feature into the spring, you can exchange the season smoothly, and vice versa. The four cases are shown in Figures 3, 4, 5, and 6 respectively.

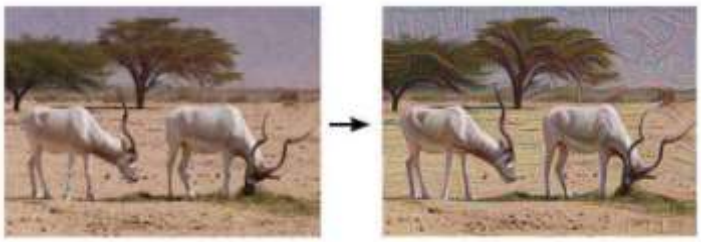

Figure 3 Converting ordinary photos into fantasy photos

Figure 4 Object fusion

Figure 5 Converting a Mustang to a Zebra

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1683-1702
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

Figure 6 Swap season

2.3.2 Generative Adversarial Network Principle

GAN is a kind of competition and confrontation idea. The discriminator keeps making its discriminative ability stronger until it can accurately identify the source of the sample. The generator keeps making its own generation ability stronger, hoping to generate samples that can make the discriminator difficult to distinguish between true and false. . During the iterative training process, both progress and eventually reach a balance.

GAN mainly consists of two adversarially trained networks, which together form the core part of the generative adversarial network and provide core functions. The generator network G uses the mapping ability of the neural network to map the noise to a distribution similar to the real samples, that is, the purpose of the generator network is to generate samples that are closest to the real samples. The discriminator network D is repeatedly trained using both the real distribution and the samples generated by the generator, and finally makes it have strong recognition ability.

The purpose of the discriminative network is to maximize the ability to discriminate whether the input comes from the true sample distribution. The purpose of the generative network is to generate fake samples, so that the discriminant network cannot judge whether the samples are real data or generated by the generator. After repeated iterations, the discriminator becomes more and more accurate, and the forgery ability of the generator becomes stronger and stronger. In this process, the ability of the generator will become stronger and stronger, and finally achieve the perspective conversion effect that this article wants.

III. DATA PROCESSING AND HUMAN POSE KEY POINT EXTRACTION

3.1 Introduction of Gait Database

The gait database [11] refers to a sample library that contains gait data, usually including gait contours or original gait recognition. The quality and abundance of database samples have a greater

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1683-1702
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

impact on the experimental results, and the type of data samples determines the choice of identification methods. Using deep learning methods can be counterproductive when the sample size is small. At present, major scientific research institutions have published some gait databases, such as the CASIA database published by the Institute of Automation, Chinese Academy of Sciences. This database is currently widely used, including multi-view gait samples, face samples, and so on.

The CASIA-B gait dataset [12] is provided by the Institute of Automation, Chinese Academy of Sciences and contains data from multiple perspectives. This dataset controls for some relevant variables and reduces many of the confounding factors that affect gait recognition studies. This dataset was born in 2005, avoiding the complex changes of the open air environment, minimizing the noise generated by natural light and removing background interference factors.

3.2 Human pose key point extraction

Human pose key points are a collection of main skeletal joint points of the human body, usually including coordinate points or images. Human posture key points have great advantages in human action recognition, and it is especially convenient to track and monitor scenes. Human posture key points also have good practicability in sports motion detection and human motion observation.

With the progress in the field of deep learning, in recent years, some scholars have begun to use the key points of human posture for gait identification. Human pose estimation is making continuous progress, and it has been possible to estimate the poses of many people at the same time. In 2017, cao [13] and others proposed a 2d pose estimation method, as shown in Figure 7 below, and its accuracy has exceeded most algorithms. The method performs joint prediction based on the confidence map, and simultaneously obtains the position and direction of the limbs, jointly learns and predicts in both directions, and finally converts the connection of related nodes into a graph theory problem.

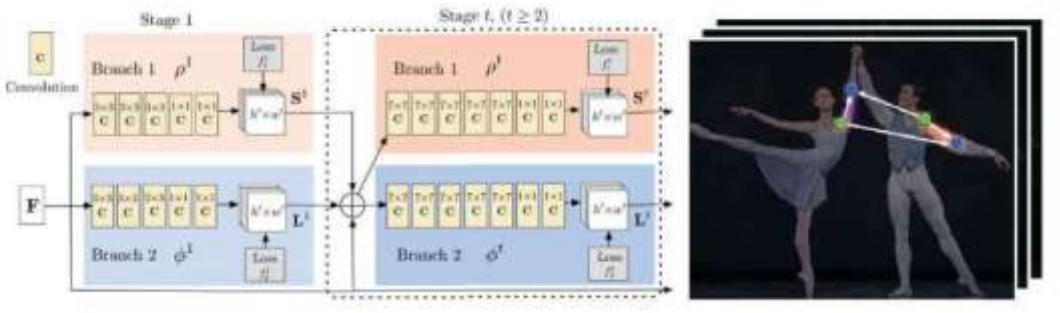

Figure 7. Human Pose Estimation

The human pose key points are the markers of the main joint points of the human body. From the above figure, it can be clearly seen that the key points of human posture fully represent the main joint points, which are enough to simulate human movement. Gait identification is to identify people's identity through human motion posture. The key points of human posture fully meet the requirements of

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1683-1702
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

gait identification, and the use of key points of human posture can avoid the interference of external features. Therefore, this paper adopts the human pose key point coordinate sequence as the main recognition data. This paper only uses 14 key points without using left eye, right eye, right ear and left ear. This is the ideal x and y coordinates of the 14 key points are the main data used in this article, and the robustness feature is also extracted based on the coordinates of the 14 key points. The specific joint points are shown in Figure 8 below.

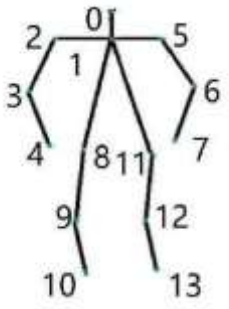

Figure 8. 14 key points used in this paper

This paper extracts human pose keypoints for each frame from pedestrian videos. The advantage of this approach is that the input to the network is no longer an image but a sequence of coordinates. Pedestrians carrying backpacks or changing clothes have little impact on key points of human posture, which naturally avoids the drawbacks of human contours on gait recognition.

3.3 Normalization processing

In the actual scene, it is not known where the pedestrian is and how far it is from the camera, but the difference in the distance of the camera will cause huge changes in the human body posture coordinates, thereby reducing the recognition accuracy. Therefore, normalization should be performed before using the key points of human pose.

The coordinates of the key points of the human body pose are regular 14 joint points, and there is a corresponding positional relationship between these joint points. The general normalization method is to directly divide these joint points by a certain number to reduce these numbers to between (0-1), this method will make the sample data unbalanced and not make good use of the characteristics of the human body pose itself.

Normalize the coordinates of all human joint points to (-1, +1), so that the recognition accuracy will not be reduced due to different shooting distances. Among the various joint points of the human body, the center of the human neck and the buttocks belong to two relatively stable joint points, so this paper takes the center of the buttocks as the origin of the coordinate plane, and takes the distance between the center of the ankle and the buttocks as the reduction ratio, Normalize the coordinates to be between -1 and +1.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1683-1702
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

3.4 Data set division

The CASIA-B data set is currently the most authoritative gait data set, which is highly recognized in the field of gait recognition, and is often used as a standard data set to evaluate the pros and cons of algorithm models. In order to more effectively verify the effect of the method in this paper on gait recognition, and also to ensure the reliability of the comparison experiment. The methods proposed in this paper are all tested on the CASIA-B dataset. In this dataset, 11 cameras are used to capture gait videos of 124 pedestrians at the same time, and each person has 11 perspectives, which is very suitable for this paper to explore the multi-perspective problem.

In the field of gait recognition, the information of a single frame of gait image is too small to recognize the identity of a pedestrian, and at least a complete gait sequence can complete the task of identity recognition. Some literatures use correlation analysis and other methods to divide the gait cycle during identification, and then use a cycle of gait sequence to identify and judge. This paper directly selects a continuous 32-frame video as a gait sequence sample. When the model generalization is good and the samples are relatively sufficient, the problem of inaccurate cycle division is avoided.

IV. PERSPECTIVE TRANSFORMATION MODEL

4.1 Common Generative Adversarial Networks for Perspective Transformation

4.1.1 Traditional Generative Adversarial Networks for Perspective Transformation

The traditional generative adversarial network [10] consists of two core parts, one is a generator used to generate fake distribution samples so that the discriminator is difficult to distinguish between true and false, and the other is a discriminator used to identify the source of the sample. The characteristics of the distribution of the samples greatly affect the generation effect. For example, the real sample is a handwritten digit. After training, the generator has the ability to generate this handwritten digit. If the real sample is multiple handwritten digits, these will be generated after training. One of the handwritten numbers. Although Generative Adversarial Networks can make the generated samples more and more like real samples, it is impossible to predict which type of samples will be generated when the real samples contain multiple categories, that is, the generated samples are similar to the real samples but the generated categories are unpredictable.

The data set used in this article contains the gait sequences of 124 people, and there are 124 types of gait sequences. The purpose of this paper is to allow the other perspectives of each person to generate a unified perspective of this person instead of the unified perspective of others. The phenomenon that the generation results of traditional generative adversarial networks are unpredictable is reflected in this paper. After training by generative adversarial networks, a very realistic unified perspective can be generated from other perspectives, but it is not known which person will be generated.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1683-1702
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

In this paper, the 36-degree angle of view is used as the unified angle of view, the samples of the unified angle of view are used as the distribution of real samples, and the samples of other angles of view are used as the distribution of fake samples. After several iterations of training, the generator has the ability to generate fake samples that are extremely similar to real samples. In this way, the generator has the ability to generate a unified perspective from other perspectives, that is, the perspective conversion ability.

4.1.2 Improved Generative Adversarial Networks for Perspective Transformation

The traditional generative adversarial network cannot predict which kind of samples will be generated, which is extremely unfavorable for a certain type of samples input from other perspectives, such as gait perspective conversion, which needs to generate samples from other perspectives of the same type. Traditional Generative Adversarial Networks are completely incompetent for the task of gait perspective conversion. The GaitGAN [14] network proposed by S Yu et al. adds a domain discriminator to the traditional generative adversarial network, and the domain discriminator can better guide the generator to convert the input gait samples from other perspectives into the same unified perspective samples. This network is used to convert gait energy maps from other perspectives to a unified perspective energy map.

Based on this idea, this paper also tries to use this network to convert the perspective of gait sequences. The improved generative adversarial network is shown in Figure 9. Like the traditional generative adversarial network, the real/fake discriminator in the network is used to judge that the sample belongs to the real sample and is also generated by the generator. The added domain discriminator is to generate samples of the same category from a unified perspective in the generator from a certain class of samples from other perspectives. Compared with the traditional generative adversarial network, this method can constrain the generation direction of the generator, so as to make the generator have a better effect as much as possible.

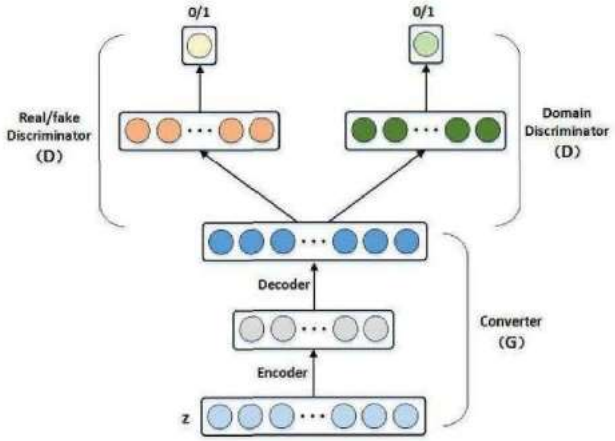

Figure 9 Improved Generative Adversarial Network

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1683-1702
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

4.2 View conversion model based on key points of human posture

4.2.1 Input data of the perspective transformation model

The key points of human posture include most of the main joints of the human body, and these key points are enough to reflect the gait of pedestrians, so the gait recognition in this paper is based on these key points. The perspective transformation model is also the key point to transform the key points of other perspectives into a unified perspective.

In a gait sequence, 14 key points are extracted from each video frame after the preprocessing stage; the x, y coordinates, and 32 consecutive frames are intercepted as a gait sequence, that is, a sample is the key point sequence after preprocessing. The processed 32*28=896 numbers, and then normalized again is the input of the generative adversarial network. The samples from other perspectives are input into the generative adversarial network to generate samples from the same perspective.

4.2.2 Perspective transformation model structure

Generative adversarial networks can generate the results we want unsupervised, but we cannot fully control the generative adversarial network to generate only the results we want without producing other interference results.

There are also two problems when using the generative adversarial network to convert perspectives: 1) Other perspectives of different people generate a unified perspective of the same person, that is, mode collapse occurs. 2) It is difficult to preserve the complete identity information unique to each person while completing the perspective conversion. In Generative Adversarial Networks, the discriminator determines the generation effect of the generator. Aiming at these two problems, this paper mainly realizes perspective conversion by setting three discriminators and one generator together. This paper uses these three discriminators to make the final generation effect of the generator better.

In this model, the real samples and the generator are jointly trained with samples generated by false distributions. When the discriminator is trained by backpropagation, the discriminator's ability to judge the true and false distributions will become stronger and stronger. When training the generator, the generator generates The sample will become more and more like the real sample. After iterative optimization of the discriminator and the generator, the ability of the generator to generate a unified perspective is gradually improved.

4.3 Evaluation and experimental comparison of perspective conversion model

In this chapter, the gait sequences of all perspectives without interference items such as backpacks in the dataset are selected for experiments. All other perspectives are converted to 36-degree perspectives with the trained perspective transformation model of this paper. When comparing, the view generated by

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1683-1702
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

someone's generative network is compared to the native 36-degree view.

4.3.1 Experimental comparison

Cross-view gait identification is a difficult problem in the field of gait recognition, and perspective switching is one of the main solutions to this problem. At present, the most commonly used method of perspective transformation is to use the gait energy map to perform singular value decomposition to establish a perspective transformation model, and to study the transformation between different perspectives as a regression problem. However, the effect of the angle conversion of this method is very limited, and the conversion of the angle of view by matrix decomposition will cause more noise in the gait energy map.

The generative adversarial network used in this paper can more accurately convert other perspectives to a unified perspective, and can also save the gait information of human posture. When the perspective conversion is completed, it can be understood that there is only one unified perspective, and then this unified perspective is used for training and recognition, thus solving the multi-perspective problem. The accuracy of the converted perspective cannot be distinguished by the naked eye, so this article uses other identification methods to distinguish.

If a simple identity discriminator can have a good identity discrimination effect, it is enough to show that the gait sequence converted by the perspective conversion model in this paper retains enough pedestrian identity information, and the effect of the perspective conversion model is good.

There is currently no internationally recognized indicator to measure the effect of perspective conversion, but the ultimate purpose of perspective conversion is to make cross-view gait identification more accurate, that is, to allow the converted gait sequence to be accurately identified. Therefore, this paper uses the pre-trained viewpoint discriminator and identity discriminator to discriminate the converted gait sequence. If the converted gait sequence can be better judged as the correct perspective and identity by the two discriminators, it means that the converted gait perspective retains relatively complete gait information and is successfully converted to a unified perspective.

In order to verify that the perspective conversion model based on human pose key points proposed in this paper is better than traditional generative adversarial networks and improved generative adversarial networks, this paper conducts comparative experiments in the same environment, and proves that the perspective conversion model in this paper can effectively convert Gait perspective.

The accuracy rate of the perspective conversion model in this paper is higher than the other two models, indicating that the perspective conversion model in this paper has a better effect on perspective conversion; the recognition accuracy of the identity discriminator is much higher than that of the traditional generative adversarial network and improved generation. The adversarial network shows that the perspective transformation model in this paper preserves more complete identity information. The recognition accuracy of both perspective discrimination and identity discrimination has reached more

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1683-1702
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

than 78%, which is enough to prove that after the perspective conversion model in this paper, the complete identity information is preserved and the perspective is converted.

V. POSE-BASED GAIT RECOGNITION

This paper proposes a method of identifying the identity by using the key points of human pose as input and combining the parameters of human joint motion. Human body motion can be visualized by using human body posture, and the key points of human body posture can be extracted from the video frame by using the human body posture estimation algorithm. This method can directly extract human pose key points from video frames for identification, or can use the unified perspective pose key points converted from the method in the previous chapter for identification. Therefore, this paper uses the combination of static features such as human body pose coordinates, joint lengths, and joint angles, and dynamic features such as joint angular velocity and acceleration to build a network. In order to make the model more generalizable and robust, this paper makes the features of different synchrony more discriminative, and tries to select two loss functions (softmax and center loss function) to train the model to achieve better results.

5.1 Gait characteristics

Machine learning and pattern recognition have gone through a long period of time, and their applications in image recognition, speech recognition and other fields have also achieved great results, especially after deep learning became popular, features can be automatically extracted by directly relying on deep networks. There is no need to manually extract features, which greatly promotes the development of the recognition field. However, in practical application scenarios, there are often not enough samples to train a very deep network, so that it is impossible to automatically extract features through the deep network to complete the recognition.

The recognition effect is not ideal when only the coordinates of the gait sequence are used in the gait recognition process. We can understand that when only the coordinates of the human body pose are used as features, the neural network does not fully utilize the gait information for classification, and it can also be understood that the deep information between each skeleton point, such as the geometric relationship, is not portrayed. Hasan et al. proposed to identify identities by extracting multiple features that are robust to perspective and then using _ network. This paper also draws inspiration from it, and improves the recognition accuracy by extracting more effective features from human poses for gait recognition. The difference between Hasan's method in this paper is that the extracted features are different, and the extracted features are more suitable for the identification of the unified perspective gait sequence converted by the perspective conversion model in this paper.

In order to better describe gait features and make better use of gait information, we extract joint length, joint angle, angular acceleration, angular velocity and coordinates from the coordinates to form a feature matrix together as feature input. The coordinates of joint points already contain information such

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1683-1702
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

as joint length, angle, etc. However, if the automatic extraction of features such as convolutional networks is used directly, the internal information of human body posture cannot be guaranteed to be utilized. For example, the human body joint angle is a very important information, but automatic Extracting features may not make good use of this information, so we should actively extract features with high recognition for identification and classification.

5.1.1 Angle feature

The foot span of the human body changes the most when walking, as shown in Figure 10. The joint angle is a key feature in gait recognition, and it is also a gait feature with obvious changes. Some literatures even directly use the gait energy map of the lower body of the human body to identify. The angle information is shown in Figure 10. This paper mainly extracts 7 key angles of each frame (bladder angle, right knee angle, left knee angle, right shoulder angle, left shoulder angle, right elbow angle, left elbow angle). The span angle is formed by the center of the hip and the left and right knees as endpoints, the right knee angle is formed by the right knee, right ankle and right hip, the left knee angle is formed by the left knee, left ankle and left hip, and the right shoulder angle is formed by the right shoulder, neck and right Elbow point is formed, left shoulder angle is formed by left shoulder, neck and left elbow point, right elbow angle is formed by right elbow, right shoulder and right wrist, left elbow angle is formed by left elbow, left shoulder and left wrist.

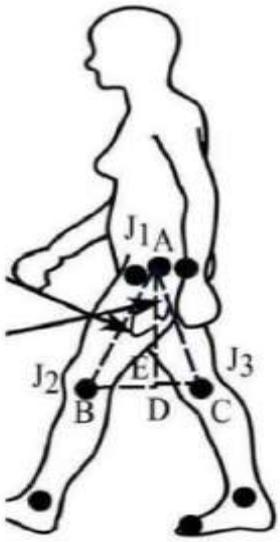

Figure 10 Human body angle identification diagram

5.1.2 Joint Length Features

From a distance, the most striking feature of the human body is its proportions, so joint length is one of the main features of the human body. The human body pose coordinates obtained in this article are

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1683-1702
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

normalized to be between -1 and 1. Although the absolute length cannot be used for comparison after normalization, the joint length ratio of each person is different. Therefore, joints can be used. length as a feature.

5.1.3 Velocity and acceleration of joint angle

The joint length and joint angle can only represent the characteristics of a static pedestrian, and cannot describe the angular change relationship between frames, let alone the dynamic characteristics of a gait sequence. Therefore, this paper uses the angular velocity and acceleration between two consecutive frames of gait as dynamic features. Due to the different frame rates of cameras and the deviation of data collected by different cameras, this paper directly uses the time of two adjacent frames as the unit time.

In this paper, a continuous 32-frame gait sequence is used, and 7 velocity values and 7 acceleration values can be calculated from 7 angles of each frame.

5.1.4 Coordinates of key points of human posture

The joint length, joint angle and other features used in this article are all calculated from the coordinates of the key points of the human body posture. The information related to the human body posture is hidden in the key coordinates of the human body posture. More useful features are extracted from point coordinates for identification. In this paper, the abscissa and ordinate coordinates of 14 key points in the human body pose key points are selected as part of the feature matrix.

5.2 Gait recognition model selection

5.2.1 Gait recognition based on convolutional neural network

Traditional machine learning commonly used models to solve classification problems include support vector machines, multilayer perceptrons, Bayesian discrimination, and so on. These classification models work very well when the input features have distinct categorical properties and minimize structural risk.

In terms of computer vision and image recognition, convolutional neural networks are very popular because they can directly input an image matrix to complete classification and recognition without additional feature extraction. Deep information in human poses can be extracted using convolutional neural networks. Use the convolutional neural network to build the model as shown in Figure 11 below.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1683-1702
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

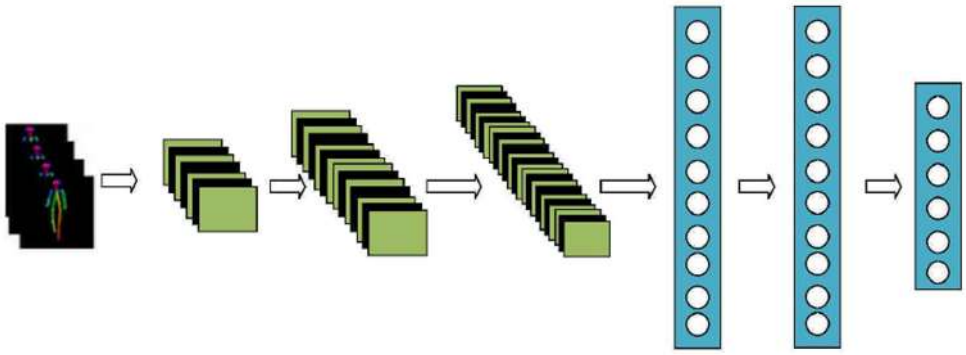

Figure 11 Convolutional Neural Network

The convolutional neural network can obtain more powerful power to express features by modifying the structure of the network, which can better extract more useful features from the input data, so as to achieve a better classification effect. The input data used in this paper is a sequence of human pose key points extracted from walking videos, and each frame contains human pose information. In this paper, the perspective conversion model based on key points of human pose is used to convert other perspectives into a unified perspective, and then use convolutional neural network to identify.

5.2.2 Gait identification based on long short-term memory network

There is a very large correlation between the frames of the gait sequence. For example, the change of the footstep span during walking is continuous, and the change of the human body posture in each frame is also related. Many current methods identify pedestrians through gait energy maps, etc., but this method loses the correlation information between frames, that is, the dynamic information of the gait sequence is lost. The current gait identification mainly focuses on the use of static information such as height, body shape, aspect ratio, etc. At present, these methods have achieved good recognition results in the case of undisturbed walking, but they have not been developed further. How to obtain better temporal information features will be an important breakthrough in gait identification.

In this paper, a long short-term memory network is used to identify the transformed view samples with a view-based translation model.

5.2.3 Combination of Convolutional Neural Networks and Long Short-Term Memory Networks

The commonly used models for classification and recognition problems are mainly convolutional neural networks and long short-term memory networks. Convolutional neural network has higher recognition accuracy for image samples with static feature structure. Long short-term memory network is more suitable for sample recognition containing temporal features. The gait sequence has typical temporal characteristics, and the human posture also has static characteristics such as structural relationship and geometric relationship. Therefore, this paper attempts to organically combine the convolutional neural network and the long short-term memory network to extract more essential gait

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1683-1702
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

features from the sequence of human pose key points, so as to better complete the gait identification, hoping to make good use of the static state in the gait sequence at the same time. characteristics and dynamic characteristics.

The combination of convolutional neural network and long short-term memory network can be divided into two kinds of parallel and series connection of two networks. When the convolutional neural network and the long short-term memory network are placed in series, there are two situations. The experiments in this paper compare these three cases with the case of using a single neural network.

5.3 Construction of gait recognition network based on human posture

Liao et al. [15] proposed to use convolutional neural network and long short-term memory network to extract the spatiotemporal information in pedestrian gait for identification of pedestrians. Inspired by this paper, it is also a combination of convolutional neural networks and long short-term memory networks for gait recognition. The difference between the method used in this paper and the method proposed by Liao et al. is: 1. This paper does not completely rely on the network to extract features from the key points of the human body pose, but also manually calculates the robust features such as angle and speed from the human body pose as input features. 2. The influence of the combination of long short-term memory network and convolutional neural network on the recognition accuracy is compared through experiments, and then the parameters and specific structure of the network structure are determined through experiments. 3. The model in this paper is more to cooperate with and better solve the problem of cross-view, and for this purpose, the model of this paper is more suitable for the problem of cross-view recognition.

At present, the key to gait recognition is to better extract the spatial and temporal features in the gait sequence. In the field of gait recognition, many people use the gait energy map to identify, because the gait energy map is a cycle of gait contour overlay, which contains a cycle of gait information, and then uses a deep neural network to extract feature and identify. The feature matrix used in this paper is a human pose feature matrix of a gait sequence, which is formed by the combination of 32 consecutive frames of gait data. The use of convolutional neural networks can better extract spatial features. A single network cannot achieve higher accuracy. The fusion of the two networks makes full use of dynamic features and static features to make the network have powerful representation capabilities, thereby improving the recognition accuracy.

The difference between gait identification and biometric identification such as face and iris is that the identity cannot be identified through a single frame of image. Biometrics such as a face can express relatively complete identity information in one image, and the information is mainly the shape, structure, skin color and other characteristics of the face, so it can be identified by the same single image, and dynamic features are not required. Gait recognition cannot identify a person's identity through a single frame of image, at least through a period of video frames to identify a person's identity.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1683-1702
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

In this paper, each frame of gait should be related to the images of the two adjacent frames, such as continuous changes in angle, roughly constant frequency, uniform speed, and so on. The LSTM can accept the influence of the previous neuron as if the current gait video frame is related to the previous gait video frame.

Since the neural network can theoretically map most of the systems, it should be able to fit almost most of the systems when the network is large enough and the training samples are rich enough, so the neural network is widely used in the field of recognition. However, the disadvantage of neural network is that the structure is uncertain and the number of network layers is uncertain, so it is necessary to obtain a good network through many experiments and parameter adjustment. This paper combines LSTM and CNN to form a network. In order to make the network play a better role, this paper has determined the network structure after many experiments.

In order for the model to efficiently achieve the purpose of identifying the identity in this paper, the network structure and loss function must be carefully designed. The loss function determines how the network model converges. This paper uses two loss functions, the first is the classic softmax loss function, which has sufficient performance on multi-classification problems. The other is the center loss function. This loss function allows the model to separate the classification hyperplanes between various categories more clearly, and the distribution of similar samples is more concentrated.

VI. CONCLUSION

This paper firstly proposes a perspective transformation model that uses a generative adversarial network to transform the gait sequence formed by human pose key points into a unified perspective. A gait recognition method is built, which takes the key points of human posture and the motion features extracted from it as input, and then establishes the recognition network. The gait recognition network constructed in this paper not only uses the key point coordinate sequence, but also calculates information such as the joint angle of human motion from it. Experiments show that these features effectively improve the recognition accuracy. In addition, the method in this paper can make better use of the time-series features and motion pose features of human poses, and organically combine the two to play a greater role. The perspective conversion model is a method based on the biological adversarial network. The advantage is that it still has a good perspective conversion effect in the case of a large perspective span; the second advantage is that it is different from the perspective conversion of the singular value decomposition method. This method directly generates a unified perspective. The key points of human body posture reduce the interference of noise.

In addition, the method in this paper still faces challenges in the case of multiple perspectives. For the research content in this paper, the following aspects need to be further improved:

(1) Optimization of gait perspective conversion. The current perspective conversion model mainly completes the process of converting from other perspectives to a unified perspective by building a

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1683-1702
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

generative adversarial network. The generation direction of the adversarial generative network depends on the discriminative strategy, which is not easy to control.

(2) The effectiveness of gait feature extraction. The main reason is that the positioning of the key points of the human body posture is not accurate enough. If a special institution can study this problem, the positioning accuracy can be further improved. I believe this method will have more possibilities.

(3) Algorithm optimization of gait recognition. The algorithm in this paper has relatively good results in terms of cross-view, but the recognition accuracy in ordinary situations is not as good as that of the model dedicated to ordinary situation recognition, and this aspect needs to be improved.

ACKNOWLEDGEMENTS

National key R & D Program "Research on innovative business operation and strategic technology system of invigorating police by science and technology" (No. 2020YFC0832604): Ministry of Public Security Technical Research Project “Research on the key technology and application of rapid and intelligent security check based on beijing-tianjin-hebei Transportation Integration” (No. 2021JSYJC26)；Key Scientific Research Projects of the China People's Police University "Intelligent fire risk assessment and early warning technology based on big data analysis" (No. ZDZX202102).

REFERENCES

[1] G. Johansson. Visual perception of biological motion and a model for its analysis[J]. Perception& Psychophysics, 1973, 14(2): 201-211.
[2] Goudail F, Lange E, Iwamoto T, et al. Face recognition system using local autocorrelations and multiscale integration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,1996, 18(10):1024-1028.
[3] H. Murase, R. Sakai. Moving object recognition in eigen space representation: Gait analysis andlip reading[J]. Pattern Recognition Letters, 1996, 17(2): 155-162.
[4] Xin C, Yang T. Extraction Method of Gait Feature Based on Human Centroid Trajectory[J] Springer Iteational Publishing, 2014.
[5] Kamei T. Image Filter Design for Fingerprint Enhancement[J]. Proc. of ISCV'95-Florida, 1995.
[6] Chen, Tao, Ma, et al. Tri State Median Fiter for Image Denoising. [J]. IEEE Transactions on Image Processing, 1999.
[7] Slot K, Kowalski J, Napieralski A, et al. Analogue median/average image filter based on cellular neural network paradigm[J]. Electronics Letters, 1999, 35(19):1619-1620.
[8] MD Zeiler, Fergus R. Visualizing and Understanding Convolutional Neural Networks[J]. Springer International Publishing, 2013.
[9] Greff K, Srivastava R K, J Koutnik, et al. LSTM: A Search Space Odyssey[J]. IEEE Transactions on Neural Networks and Learning Systems","pubMedId":"27411231, 2017.
[10] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Networks[J]. Advances in Neural Information Processing Systems, 2014, 3:2672-2680.
[11] Wang Kejun, Ding Xinnan, Xing Xianglei, et al. Review of multi-view gait recognition [J]. Journal of Automation, 2019, 45(005):841-852.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1683-1702
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

[12] Liu C L, F Yin, Wang D H, et al. CASIA Online and Offline Chinese Handwriting Databases[C]// International Conference on Document Analysis & Recognition. IEEE Computer Society, 2011.
[13] CAO Z, SIMON T, WEI S-E, et al. Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields[C] // CVPR.2017.
[14] Yu S, Chen H, Reyes E, et al. GaitGAN: Invariant Gait Feature Extraction Using Generative Adversarial Networks[C]// Computer Vision & Pattern Recognition Workshops. IEEE, 2017.
[15] HAN J, BHANU B. Individual recognition using gait energy image[J]. IEEE transactions on pattern analysis and machine intelligence, 2006, 28(2): 316- 322.