Gait Recognition Based on Joint Distribution of Motion Angles

Wei Lu*, Wei Zong, Weiwei Xing

(School of Software Engineering, Beijing Jiaotong University, Haidian District, Beijing 100044, China)

Abstract—Gait as a biometric trait has the ability to be recognized in remote monitoring. In this paper, a method based on joint distribution of motion angles is proposed for gait recognition. The new feature of the motion angles of lower limbs are defined and extracted from either 2D video database or 3D motion capture database, and the corresponding angles of right leg and left leg are joined together to work out the joint distribution spectrums. Based on the joint distribution of these angles, we build the feature histogram individually. In the stage of distance measurement, three types of distance vector are defined and utilized to measure the similarity between the histograms, and then a classifier is built to implement the classification. Experiments has been carried out both on CASIA Gait Database and CMU motion capture database, which show that our method can achieve a good recognition performance.

Index Terms —biometrics, gait recognition, joint distribution, feature histogram

I. INTRODUCTION

BIOMETRICS refer to the identification of humans by their characteristics or traits[1]. The characteristics include but not limited to face, fingerprint, iris, gait and DNA. However the current recognition methods, such as face, fingerprint or iris based, require a cooperative subject or physical contact. So it is nearly impossible to identify people at a distance by using these methods. However, gait as the way people walk does not have these constraints. In the past decades, many studies have proven that gait has the potential to become a powerful biometric for surveillance and access control, since it has advantages such as noninvasive, hard to conceal and capable of being acquired at a distance [2]. In fact, besides being well-suited to identify people at a distance, gait also have the potential to be applied in the medical field. For example, recognizing changes in walking patterns early can help to identify conditions such as Parkinson’s disease and multiple sclerosis in their earliest stages [3]. Although gait has some limitations, e.g., it may not be as unique as fingerprint or iris, and may be affected by one’s clothing, shoes or physical condition, the inherent gait characteristic of an individual still makes it irreplaceable and useful in visual surveillance.

Nowadays, video is not the only way to collect the gait any more. According to the ways of data collection, gait recognition methods can be divided into three categories: Video Sensor (VS)-based, Floor Sensor (FS)-based and Wearable Sensor (WS)-based [4]. VS-based method collects gait data by video cameras. Without physical contact, VS-based method is the most noninvasive way and can get the most natural way of one’s walking. Moreover, video cameras are widely used in our daily life, so it is quite easy to get the gait data in a variety of occasions. However, the image processing is required, and the images captured from cameras should be preprocessed in order to get the gait information that can be used directly. The most widely used field of VS-based method is remote monitoring. FS-based method is also called footprint method, which puts the sensors on the floor and record the information of one’s footprint such as length and location to be studied. WS-based method needs the subject wear sensors and collects the motion data recorded by them. Different from VS-based method, WS-based method can get the motion data directly and the data is more suitable for gait analysis. WS-based method is popularly used in gait analysis mainly for medical purpose. However, it is not a good choice for remote monitoring as the non-invasive is the key feature. The gait data can be represented in 2D or 3D. 2D data is presented by the sequence of images, which is widely used in early days. While in recent years, gait recognition based on 3D data became a new trend. The 3D data is mainly acquired and calculated by motion capture technology.

In this paper, a method based on joint distribution of motion angles for gait recognition is proposed, which can work on 2D video database or 3D motion capture database. The motion angles of lower limbs from the original data are extracted to propose the new feature, thus the joint distribution spectrums can be work out. Then the feature histogram is built and the distance between the histograms is calculated. Finally, the classification is implemented on the distance vector to recognize the gaits.

As for the experiments data in this paper, we use the CASIA Gait Database [5] from the Institute of Automation, Chinese Academy of Sciences as the 2D data source, and the CMU motion capture database in ASF/AMC format [6] as 3D data. In the stage of motion angle extraction, we only use the motion angles of lower limbs which can work well even when the subject is with half-occlusion in the images. After the motion angles are extracted, we take the result as a time-free model instead of considering them as a time sequence model. The spatial distribution of them, however, is what we concerned. In other words, we only care about the postures of the subject when he is walking. Experimental results are compared with other similar work which demonstrates our method can reach a higher accuracy.

The reminder of the paper is organized as follows: Section

* Corresponding author: luwei@bjtu.edu.cn
2 is the related work. Section 3 shows the feature definition and extraction. The classification is described in Section 4. Section 5 presents the experiments and the results. Section 6 concludes the paper.

II. RELATED WORK

Gait recognition method has been well studied and many methods have been proposed which can be classified into model-free and model-based approaches [7]. In the model-free approaches, moment of shape is one of the most common used features. In addition silhouette and statistical features are widely used in a lot of work. In this approach, the correspondence between successive frames is established based upon the prediction or estimation of features related to position, velocity, shape, texture and color. Alternatively, they assume some implicit notion of what is being observed [8]. In the model-based approaches, the prior information or a known model is needed for fitting human gait. Though the model-based method is more complex, it has some advantages such as immunity to noise [9]. The model-based approaches assume a priori model to match the data extracted from video, and features correspondence is automatically achieved. These two kinds of methods both follow the general framework of features extraction, features correspondence and high-level processing. The essential difference between these two approaches is whether a model is used to fit the bounding contours of the walker.

The model-free approaches gained a rapid development in the early days. L. Wang, et al. [10] put forward a recognition method based on shape analysis, and presented the static pose changes of these silhouettes over time. Then Procrustes shape analysis is implemented to obtain gait signature. R. T. Collins, et al. [11] presented a key frame identification technology which is view point dependent on the basis of the outline template of the human body. N. Cuntoor, et al. [12] studied the dynamic characteristics of the front view and side view, took the motion features such as the motion of arms, legs and body shaking features into consideration for gait information identification, to a certain extent, improved the recognition rate. In [13, 14], based on the appearance and view point, A. Kale, et al. presented the binarization contour as the feature using the Dynamic Time Warping (DTW) to deal with the speed changes in the process of walking, and strengthen the fault tolerance of original data. In addition, A. Kale, et al. [9] use the width of the outer contour of the binarized silhouette as the image feature and built a HMM model to distinguish the dynamic features of the gait. M. Hu, et al. [15] also built a HMM model for gait recognition. And they take the local binary pattern descriptor as the motion feature. J. W. Davis and A. F. Bobick [16] proposed the temporal template first which is for appearance-based action representation. They used the motion energy images (MEIs) and motion history images (MHIs) to represent motion sequences. In recent years, the depth information of the silhouette is also used as the motion feature for gait recognition. P. Chattopadhya, et al. [17] put forward a novel feature known as Pose Depth Volume which is just based on the depth frame.

As a model-based approach, D. Cunado, et al. [18, 19] mapped the leg movement of human body to a pendulum model. It contained a pendulum motion model and a structural model. The lower limb was modeled as two interconnected pendulums. C. Y. Yam, et al. [20] calibrated the rotation of thigh and curves manually, and extracted the transformation of the angles as the feature. A. Yilmaz and M. Shah [21] proposed a model-based method to extract the joints of lower limbs from lateral walking sequences. I. Bouchrika and M. S. Nixon [22] studied the effects of covariates, such as footwear, load carriage, clothing and walking speed for gait recognition. The accuracy of the model-based method is not as high as the one based on the contour, however, the fault tolerance is strengthened. The method proposed in this paper is also model based, and the model we used refers to [23].

In recent years, with the development of 3D technology, more and more research focus on the motion features analysis on 3D data. The 3D data are more accurate than the traditional video data and include more information of human gait. G. Ariyanto and M. S. Nixon [8] used the Soton Multi-biometrics Tunnel to acquire 3D data, and an articulated cylinder model was built to analyze the features such as height, stride and footprint poses. J. Gu and X. Ding [24] addressed a common viewpoint-free framework for action recognition and human identification from gaits. They used a vision-based markerless pose recovery system to extract 3D human joints from 3D reconstructed volume data. D. Gafurov [4] extracted 3D joint information by wearable sensors for gait recognition and gait security research. In this paper, our method has been tested on a 3D database and the result shows a good adaptability of our method on 3D data.

III. FEATURES EXTRACTION

A. Extracting Motion Angles

In this paper, the human model is built according to the Fig.1 [24], and only motion angles of lower limbs are used which affect the gait greatly and are more powerful in distinguishing capability. The thigh and knee angles shown in Fig.1 are used for gait analysis in our method. \( \theta_{\text{thigh1}}(t) \) and \( \theta_{\text{thigh2}}(t) \) are noted as left thigh angle and right thigh angle respectively. \( \theta_{\text{knee1}}(t) \) and \( \theta_{\text{knee2}}(t) \) are noted as left knee angle and right knee angle respectively.

![Fig.1 Gait signature and joint angles](image-url)
In 3D database the angles can be read from the AMC files directly. As for the 2D data, the angles can be extracted from every frame of the gait video by image processing, and the method we used is motivated by [23].

1) Estimating motion angles

Let \( F(X, Y, t) \) be the binary frame read from the database directly, and \( \{X_{sil}, Y_{sil}\} \) are the pixels in the area of the human silhouette. The lower limbs positions are estimated following the formulas according to [24]:

\[
\begin{align*}
\theta_{hip} &= \min (Y_{sil}) + 0.50 \cdot H \\
\theta_{knee} &= \min (Y_{sil}) + 0.75 \cdot H \\
\theta_{ankle} &= \min (Y_{sil}) + 0.90 \cdot H
\end{align*}
\]

where \( H \) is the silhouette’s height.

As for the \( r \)th frame, when the legs are not overlapped, there exist two sequence of \( X_{sil} \) with \( Y_{sil} = \theta_{knee}. \) Under this condition the two shins are defined by

\[
\begin{align*}
X_{shinl} &= [x_{1l}, x_{2l}, ..., x_{c}] \\
Y_{shinl} &= [y_{1l}, y_{2l}, ..., y_{c}]
\end{align*}
\]

where \( l \in \{1, 2\} \) refers to left or right shin, and

\[
X_{sl} = \frac{\sum_{j=1}^{c} x_{j} F_{shinl}(x_{j}, y_{l})}{\sum_{j=1}^{c} F_{shinl}(x_{j}, y_{l})}
\]

Then the shins can be linearly approximated by the first order polynomial with coefficients as follows

\[
P_{1}(X_{shinl}, t) = p_{0l}(t) + p_{1l}(t) \cdot X_{shinl}(t)
\]

The angle between the shins and the vertical axis is

\[
\theta_{shinl}(t) = \pi - \tan^{-1} p_{1l}(t)
\]

The hip position is defined by

\[
\begin{align*}
x_{hip} &= \frac{1}{p} \sum_{j=1}^{p} x_{j} \\
y_{hip} &= y_{hip}
\end{align*}
\]

where \( p \) is the number of \( X_{sil} \) with \( Y_{sil} = \theta_{hip}. \)

The coefficients satisfy the equations below

\[
\begin{align*}
\theta_{hip}(t) &= q_{0l}(t) + q_{1l}(t) \cdot x_{hip}(t) \\
\theta_{knee}(t) &= q_{0l}(t) + q_{1l}(t) \cdot x_{knee}(t)
\end{align*}
\]

Finally, we can calculate the motion angles by

\[
\begin{align*}
\theta_{thighl}(t) &= \pi - \tan^{-1} q_{1l}(t) \\
\theta_{kneel}(t) &= \theta_{thighl}(t) + \theta_{shinl}(t)
\end{align*}
\]

When the legs are overlapped, there exist only one sequence of \( X_{sil} \) with \( Y_{sil} = \theta_{knee}. \) In this case, we ignore the value of the pixel, but treat it as a signal to exchange the positions of right leg and left leg.

2) Curve fitting

A sequence of \( \theta_{thighl}(t) \) and \( \theta_{kneel}(t) \) are extracted from the video after motion angles estimation. However the frame rate can’t meet the requirement of our recognition method, and more frames are interpolated for one single gait cycle. The following trigonometric-polynomial functions are used to fit the angles [25]:

\[
\begin{align*}
\theta_{thighl}(t) &= b_{l} + a_{1l} \cdot \sin t + b_{1l} \cdot \cos t \\
\theta_{kneel}(t) &= b_{l} + \sum_{n=1}^{p} [a_{nl} \cdot \sin(nt) + b_{nl} \cdot \cos(nt)]
\end{align*}
\]

Fig.2 and Fig.3 show the fitting results. Then we resample the angles more frequently, and ten times of the original frequency in our experiment.

B. Extracting Feature

Based on the obtained motion angles, the feature will be extracted. In some related work, Fourier Descriptors [26] were used to present the feature of these motion angles. However, the correct classification rate based on Fourier Descriptors is proven to be limited. So in this paper, the feature in time domain is extracted, which is new to gait recognition. To demonstrate the effectiveness of the feature, a series of data are analyzed based on 3D database and 2D database. In the 3D database, all of the data labeled as walk have been analyzed. As for the 2D database we select 50 persons randomly with 6 motion segments each. The analysis method is organized in the following way. The motion angles are extracted as vectors of angle pair

\[
\begin{align*}
\{\theta_{thighl}(t), \theta_{thighl}(t)\} \\
\{\theta_{kneel}(t), \theta_{kneel}(t)\}
\end{align*}
\]

Fig.4 and Fig.5 show the analysis results on motion angles from 3D database, where x axis is right knee angles and y axis is left knee angles.

![Fig.2 Curve for thigh angle](image)

![Fig.3 Curve for knee angle](image)

![Fig.4 Knee angles trajectory 1](image)

![Fig.5 Knee angles trajectory 2](image)

The two trajectories in Fig.4 are formed by the knee angles which belong to one person but different motion segments.
Each point represents one knee angle pair of one frame, and the trajectory represents the spatial distribution of the knee angles. Fig. 5 corresponds to another person, where the trajectory is totally different from Fig. 4.

Regarding to the motion angles analysis on 2D data, we select 4 out of 50 persons randomly with 5 motion segments each and the corresponding results are shown in Fig. 6(a)-Fig. 6(d). Due to the raw data noises and computation tolerance, some trajectories of the same person seem a little different, however, most of the data show pattern aggregation.

4. Feature Extraction

Based on the data analysis, the trajectory is a two-value image. In this paper, a histogram is presented and built based on variance of the motion angles to describe the characteristics of the trajectory. The histograms construction is described below:

Step 1: Obtain the motion angle pairs in Eq. (10).

Step 2: Build motion angles datasets by

\[
\begin{align*}
&S_{\text{thigh}}(i) = \{\theta_{\text{thigh}}(i)\mid \theta_{\text{thigh}}(i)'\} = \{\theta_{\text{thigh}}(i)\}\nonumber \\
&S_{\text{knee}}(i) = \{\theta_{\text{knee}}(i)\mid \theta_{\text{knee}}(i)'\} = \{\theta_{\text{knee}}(i)\} \tag{11}
\end{align*}
\]

where \(t'\) and \(i\) are the index of the interpolated frames. For example in \(S_{\text{thigh}}(i)\) for every \(\theta_{\text{thigh}}(i)\) we can get a list of \(\theta_{\text{thigh}}(i)\).

Step 3: Calculate the variance of each dataset. For example, as for each list of \(S_{\text{thigh}}(i)\), the variance is calculated by

\[
\text{Var} = \frac{\sum_{i=1}^{n}(\theta_{\text{thigh}}(i) - \bar{\theta}_{\text{thigh}})^2}{n}
\]

where \(\bar{\theta}_{\text{thigh}} = \frac{\sum_{i=1}^{n}\theta_{\text{thigh}}(i)}{n}\) and \(n\) is the number of data in \(S_{\text{thigh}}(i)\).

Therefore the four histograms based on left knee, right knee, left thigh and right thigh are obtained, which present the postures of legs during walking and are used as the features for gait recognition. Fig. 7 is an example of the obtained histogram.

IV. CLASSIFICATION

A. Distance Measurement

The gait is represented by a feature histogram, so the distance of two gaits can be defined as the similarity between two histograms. Many suitable functions can be used to calculate the similarity. In this paper, the correlation coefficient function, \(\chi^2\) function and \(L_1\) distance are chosen, which have been extensively used for histogram comparison.

1) Correlation coefficient function

In statistics the correlation coefficient is a measure of the linear correlation between two variables, and belongs to \([-1, +1]\) where +1 is total positive correlation, 0 is no correlation, and -1 is total negative correlation. We let \(H_1\) and \(H_1'\) denote the two corresponding histograms and \(I\) is the index of the bins. The formula for correlation coefficient is

\[
r(H_1, H_1') = \frac{\sum(H_1(i)-\bar{H}_1)(H_1'(i)-\bar{H}_1')}{\sqrt{\sum(H_1(i)-\bar{H}_1)^2 \sum(H_1'(i)-\bar{H}_1')^2}}
\]

where \(\bar{H}_1 = \frac{1}{N} \sum_{i} H_1(i)\) and \(\bar{H}_1' = \frac{1}{N} \sum_{i} H_1'(i)\), \(N\) is the number of bins.

The distance between two histograms \(d_r(H_1, H_1')\) is defined as

\[
d_r(H_1, H_1') = \frac{2}{r(H_1, H_1')} - 1
\]

The value of \(d_r(H_1, H_1')\) is between \([0, +\infty)\).

2) \(\chi^2\) function

\(\chi^2\) function is used for statistical test and evaluate the difference between two data sets. We use here a symmetrized approximation of \(\chi^2\):

\[
d_{\chi^2}(H_1, H_1') = \sum_i \frac{(H_1(i)-H_1'(i))^2}{H_1(i)+H_1'(i)}
\]

When \(H_1(i)\) and \(H_1'(i)\) are both equal to 0, \(d_{\chi^2}(H_1, H_1') = 0\).

3) \(L_1\) distance

The \(L_1\) distance is also known as Manhattan distance and defined as:

\[
L_1(H_1, H_1') = \sum_i |H_1(i) - H_1'(i)|
\]

As the distribution interval of the bin-values is not fixed, the normalization is needed for evaluating the distance between two histograms. The normalized formula is

\[
d_{L_1}(H_1, H_1') = \sum_i \frac{|H_1(i)-H_1'(i)|}{\max(H_1(i), H_1'(i))}
\]

When \(H_1(i)\) and \(H_1'(i)\) are both equal to 0, the value is equal to 0.
In order to compare similarity of the motion segments, the corresponding three distance vectors should be calculated by
\[
\begin{align*}
D_x(H,H') &= \left(d_x(H_{\text{right}},H_{\text{right}}), d_x(H_{\text{left}},H_{\text{left}})\right) \\
D_y(H,H') &= \left(d_y(H_{\text{right}},H_{\text{right}}), d_y(H_{\text{left}},H_{\text{left}})\right) \\
D_r(H,H') &= \left(d_r(H_{\text{right}},H_{\text{left}}), d_r(H_{\text{left}},H_{\text{right}})\right)
\end{align*}
\]

One of the three vectors or the combination of them with a Boolean flag will be taken as the input data of the next classification, where the Boolean flag represents whether the two motion segments come from the same person.

**B. Classification**

In [27], SVM classifier is proven to be superior in solving the histogram-based classification problem. The SVM is a supervised learning model with associated learning algorithms that analyze data and recognize patterns, and be used for binary classification. The key of using SVM is to select a kernel carefully as an inappropriate kernel may lead to poor performance. In our experiments four types of kernel are selected. \(K_{\text{poly}}\) stands for polynomial kernel, \(K_{\text{RBF}}\) for a radial basis function (RBF), \(K_{\text{sig}}\) for the sigmoid kernel, and \(K_{\text{line}}\) for the linear kernel. We assume \((a_x, b_x)\) to be the input sample, where \(b_x \in \{0,1\}\) is the flag used for labeling the category, 0 stands for different persons and 1 for same person. In this paper, the experiments on the regular expression \(a_i = (D_x^2, D_y^2, D_r^2)\) and \(a_i \neq \emptyset\) are implemented. Besides the kernel type and the input vector, the different training sets are also discussed. The classification experiments will attempt to achieve the best combination with the highest correct classification rate.

**V. EXPERIMENTS**

**A. Database**

2D database and 3D database are used for our experiments.

1) 2D Database

The 2D gait database we used is from CASIA Gait Database Dataset B [5], which was created in 2005 and used a lot in gait recognition research in recent years. There are 124 subjects and the data was captured from 11 views. The angle interval between two adjacent views is 18°, and the view angle is 0°, 18°, 36°, 54°, 72°, 90°, 108°, 126°, 144°, 162° and 180° from left to right. And there are three variations, i.e. view angle, clothing and carrying condition, which are separately considered. The captured sample frames are shown in Fig.8. In this paper, we use the side view frame and the view angle is 90° in the database.

2) 3D Database

3D Data used in our experiment is from CMU motion capture database [6]. The data is captured using 12 Vicon infrared MX-40 cameras, each of which is capable of recording 120 Hz with images of 4 megapixel resolutions. The cameras are placed around a 3m × 8m rectangular area. Humans wear a black jumpsuit and have 41 markers shown in Fig.9. The skeleton model used in the system is shown in Fig.10 to build 3D capture data. The data is provided in ASF/AMC format with a motion type label, such as walk, run and jump. Our experiments use the motion of walk for gait recognition. The index number of data can tell us which person the motion belongs to.
1) Experiments on the normal walking data

These experiments are designed to demonstrate the validity of our method on normal walking data, meanwhile how the three factors (SVM kernel, distance function and training set) affect CCR will be discussed. Here a total of 124 persons with 6 motion segments each are used for feature extraction. Each pair of them is made comparison calculation, and a total of $C_{124}^2$ comparisons are obtained. In [28], C. Chen discussed several recognition methods and the recognition rate achieved using the same data. The result is shown in Table 1.

<table>
<thead>
<tr>
<th>Angle</th>
<th>PHNM</th>
<th>Frieze feature</th>
<th>Wavelet feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>90</td>
<td>90.3%</td>
<td>58.1</td>
<td>91.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>95.2</td>
<td>61.3</td>
</tr>
</tbody>
</table>

In the first experiment, one motion segment is selected randomly and compared with other 457 motion segments in the dataset for analyzing whether they belong to the same person. The regularization parameter (C) is set to 400. The performance under different SVM kernels and distance functions are evaluated in order to achieve the highest CCR.

Table 2 shows the CCR results under the same training set and testing set, and different SVM kernel and distance function. It is noted that the highest CCR of 99.12% is reached under the combination of $K_{sig}$ and $D_x^2$, which performs better than 95.20% in Table 1. The second highest CCR achieved by the single distance is up to 96.06%. The high CCR shows the advantage of our method working on normal walking data. Meanwhile the results in Table 2 also show the fact that increasing input dimensions does not improve the CCR but increase the computing complexity. So combination of the distance input cannot obtain a better performance. As for the normal walking data, $D_x^2$ and $D_{L_1}$ are proven to be more effective than other distance vectors. Moreover among all of the kernels, $K_{sig}$ achieves the highest average CCR. So the next experiments will be designed on the combination of $K_{sig}$ with $D_x^2$ and $D_{L_1}$.

<table>
<thead>
<tr>
<th>Distance vector</th>
<th>SVM Kernel</th>
<th>$K_{poly}$</th>
<th>$K_{RBF}$</th>
<th>$K_{sig}$</th>
<th>$K_{line}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_{x^2}$</td>
<td>81.40%</td>
<td>78.34%</td>
<td>99.12%</td>
<td>84.25%</td>
<td></td>
</tr>
<tr>
<td>$D_{L_1}$</td>
<td>84.46%</td>
<td>90.59%</td>
<td>96.06%</td>
<td>70.20%</td>
<td></td>
</tr>
<tr>
<td>$D_{x^2}$</td>
<td>67.40%</td>
<td>77.68%</td>
<td>86.21%</td>
<td>91.03%</td>
<td></td>
</tr>
<tr>
<td>$D_{x^2}$, $D_{L_1}$</td>
<td>89.93%</td>
<td>77.02%</td>
<td>84.68%</td>
<td>86.87%</td>
<td></td>
</tr>
<tr>
<td>$D_{x^2}$, $D_{x^2}$</td>
<td>62.80%</td>
<td>76.15%</td>
<td>99.12%</td>
<td>73.93%</td>
<td></td>
</tr>
<tr>
<td>$D_{x^2}$, $D_{L_1}$, $D_{L_1}$</td>
<td>84.46%</td>
<td>93.22%</td>
<td>82.06%</td>
<td>96.06%</td>
<td></td>
</tr>
<tr>
<td>$D_{x^2}$, $D_{x^2}$, $D_{L_1}$, $D_{L_1}$</td>
<td>83.59%</td>
<td>77.46%</td>
<td>98.25%</td>
<td>77.02%</td>
<td></td>
</tr>
</tbody>
</table>

In order to explore that how the training set will influence the CCR, the second experiment is carried out. We select one motion segment randomly and compare with all of the motion segments in the dataset including itself, and obtain a total of $124 \times 6=744$ comparisons. These comparisons form the testing set in the experiment. The regularization parameter (C) is set to 400. $K_{sig}$ is taken as the kernel, $D_x^2$ and $D_{L_1}$ are used to measure the distance. Four training sets are selected from all the comparisons and follow the rules below.

- $t_s_1$: 60 comparisons including 30 data with $b_1 = 0$.
- $t_s_2$: 60 comparisons including 40 data with $b_1 = 0$.
- $t_s_3$: 5000 comparisons including 2500 data with $b_1 = 0$.
- $t_s_4$: 5000 comparisons including 3400 data with $b_1 = 0$.

The classification results are presented in Table 3. Under the same condition the CCR varies with the training set, which means the quality of training set affects CCR. Among all the training sets, $t_s_3$ achieves the lowest CCR, which indicates that the amount of training set samples is not a necessary factor for a high quality training set. In addition, although the scale of $t_s_4$ and $t_s_2$ different greatly, the two CCRs are close to each other. The CCR of $t_s_4$ is higher than $t_s_2$ under $D_x^2$, however, the case is the contrary as for $D_{L_1}$.

Table 3 Classification results on different training set

<table>
<thead>
<tr>
<th>Distance vector</th>
<th>$t_s_1$</th>
<th>$t_s_2$</th>
<th>$t_s_3$</th>
<th>$t_s_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_x^2$</td>
<td>93.03%</td>
<td>98.87%</td>
<td>89.45%</td>
<td>99.44%</td>
</tr>
<tr>
<td>$D_{L_1}$</td>
<td>96.05%</td>
<td>99.62%</td>
<td>90.96%</td>
<td>96.23%</td>
</tr>
</tbody>
</table>

2) Experiments on the mixed data

The main purpose of this experiment is to discuss how the clothes affect the recognition results. In previous work clothing types are always treated as an important factor, as they can influence the overall body shape and some certain types of clothing can even affect the way a person walks. In [22] a significant drop in performance (87% to 60%) was reported when the person wore a coat on top of the normal clothes. So that evaluating whether our method is clothes sensitive or not is necessary. CASIA Gait Database Dataset B contains three types of data, i.e. normal walking data, walking data with a bag and walking data with a coat, and a total of $124 \times 10=1240$ motion segments. Here all of these data are mixed together to evaluate our method. Some previous work also made effort on the clothes problem, in [29] several methods were discussed on the CASIA Gait database with clothes type considered. The results in [29] are shown in Table 4.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Recognition accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CASIA dataset B</td>
<td>91.50</td>
</tr>
</tbody>
</table>

Table 4 Recognition accuracy using different approaches
Our experimental results are shown in Table 5, where the training dataset is fixed to 60 motion segments randomly selected from the database. The regularization parameter (C) is set to 400. The testing data are the mixed data including the motion segments randomly selected from the normal walking data, from the walking data with a coat, and from the walking data with a bag respectively. Then all of these three motion segments are compared with whole dataset. So as for each element in Table 5, which is implemented on a total of 124 × 10=1240 comparisons.

Table 5 Experimental results on mixed data

<table>
<thead>
<tr>
<th>Distance vector</th>
<th>SVM Kernel</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>K_poly</td>
</tr>
<tr>
<td>D_x^2</td>
<td>95.94%</td>
</tr>
<tr>
<td>D_1^k</td>
<td>77.66%</td>
</tr>
<tr>
<td>D_r</td>
<td>52.10%</td>
</tr>
</tbody>
</table>

Table 5 shows our experimental results. It is noted that by using the high quality training set, the CCRs under D_x^2 all achieve to 95% and above. The best CCR is 99.44% by D_1^k and K_sig. All of these results under D_x^2 are much higher than the CCR presented in Table 4. The results demonstrate that our method works well on the mixed data and is not clothes sensitive. The experimental results also show that D_1^k and D_r perform worse than D_x^2 on the mixed data. Based on the experiments on the 2D database, D_x^2 is supposed to be used as the distance measurement and can obtain better recognition result.

C. Experiments on 3D database

In this section the 3D database is used to evaluate the performance of our method. We select 54 motion segments labeled as walk from the database which come from 8 persons. Every pair of the motion segments are compared since it is a small dataset, make it a total of 1458 comparisons in this experiment. The experiments are divided into 2 parts. The first aim to evaluate the effectiveness of our method on the 3D data, and the second are designed to show how the training set would influence the CCR.

1) Experiments of performance on 3D data

A training set containing 150 comparisons is used for these experiments, which is selected randomly from the dataset. And the rest of the 1308 comparisons are treated as the testing set. Four kinds of kernels and seven kinds of distance vectors are implemented as the elements in the experiment for achieving the highest CCR. The results are shown in Table 6.

Table 6 Experimental results on 3D Data

<table>
<thead>
<tr>
<th>Distance vector</th>
<th>SVM Kernel</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>K_poly</td>
</tr>
<tr>
<td>D_x^2</td>
<td>94.46%</td>
</tr>
<tr>
<td>D_1^k</td>
<td>93.50%</td>
</tr>
<tr>
<td>D_r</td>
<td>66.44%</td>
</tr>
<tr>
<td>(D_x^2, D_1^k)</td>
<td>94.61%</td>
</tr>
<tr>
<td>(D_1^k, D_r)</td>
<td>72.55%</td>
</tr>
<tr>
<td>(D_x^2, D_r)</td>
<td>68.31%</td>
</tr>
<tr>
<td>(D_x^2, D_1^k, D_r)</td>
<td>71.79%</td>
</tr>
</tbody>
</table>

Table 6 shows a quite different pattern compared with the result on 2D database. The situation of dataset (amount of persons and motion segments), the raw data error and the data format may lead to this difference. However the CCR on 3D data achieves up to 95.6% under the distance vector of (D_x^2, D_1^k, D_r) with K_RBF and (D_x^2, D_1^k) with K_line. K_sig performs the best on 2D data but not on 3D data. Instead of K_sig, K_RBF become a proper choice for gait recognition in the 3D environment. G. Ariyanto and his fellows achieved to 79.4% correct classification rate in their paper [8]. But the data we used in the experiment is different, while the data collection way is the same.

2) Experiments on different training sets

The experiments on 2D database show that the CCR get influenced greatly by the quality of training set. This experiment aims to evaluate the performance on 3D data of different training set. 1000 comparisons are selected from the dataset randomly as the testing data, and the subset of the rest data forms the training set. The rules below are followed for the four training sets.

- t_s1: 150 comparisons, where the data is from 8 persons.
- t_s2: 150 comparisons, where the data is from 8 persons but the motion segments are different from t_s1.
- t_s3: 150 comparisons, where the data is from 4 persons.
- t_s4: 300 comparisons, where the data is from 4 persons.

In the experiment K_RBF is chosen to be the kernel as its performance is the best according to the former experiments. The results on different training sets are shown in Table 7.

Table 7 Experimental results on different training sets

<table>
<thead>
<tr>
<th>Distance vector</th>
<th>t_s1</th>
<th>t_s2</th>
<th>t_s3</th>
<th>t_s4</th>
</tr>
</thead>
<tbody>
<tr>
<td>D_x^2</td>
<td>94.61%</td>
<td>94.53%</td>
<td>84.25%</td>
<td>82.30%</td>
</tr>
<tr>
<td>D_1^k</td>
<td>92.89%</td>
<td>93.23%</td>
<td>77.91%</td>
<td>72.71%</td>
</tr>
<tr>
<td>D_r</td>
<td>69.57%</td>
<td>67.55%</td>
<td>56.50%</td>
<td>58.12%</td>
</tr>
<tr>
<td>(D_x^2, D_1^k)</td>
<td>93.12%</td>
<td>94.84%</td>
<td>84.94%</td>
<td>81.95%</td>
</tr>
<tr>
<td>(D_1^k, D_r)</td>
<td>93.27%</td>
<td>95.07%</td>
<td>76.22%</td>
<td>72.71%</td>
</tr>
<tr>
<td>(D_x^2, D_r)</td>
<td>71.10%</td>
<td>70.72%</td>
<td>60.63%</td>
<td>62.44%</td>
</tr>
<tr>
<td>(D_x^2, D_1^k, D_r)</td>
<td>93.43%</td>
<td>95.68%</td>
<td>78.98%</td>
<td>75.82%</td>
</tr>
</tbody>
</table>

The results show that the number of the class in training set is more important than the number of motion segments. More segments from different persons can achieve a higher CCR. The accuracy of D_r is much less than others, moreover, D_x^2 and (D_x^2, D_1^k) perform the best in average. The highest CCR is achieved by (D_x^2, D_1^k, D_r) under the second training set t_s2, however the low quality of the training set can decrease the CCR more than D_x^2 and (D_x^2, D_1^k). Therefore D_x^2 and (D_x^2, D_1^k) are the better distance vector used in the 3D environment and K_RBF would be the choice of kernel. As for the training set, it should cover as many persons as possible to guarantee the recognition accuracy.
VI. CONCLUSION

In this paper, a gait recognition method based on the joint distribution of motion angles is proposed. The distribution characteristic of the data in time domain is presented and the feature histograms are built for gait recognition. Three distance measurements are implemented as the input vectors and then a SVM classifier is built to perform the classification. Experiments are conducted both on 2D video gait database and 3D motion capture database to evaluate the performance of our method. The results show a high CCR and the effectiveness of the proposed method both on 2D and 3D databases. In addition, the quality of training set is analyzed and discussed. The results show the importance of proportional relationship of different classes in the training set as for 2D database and the importance of covering more persons on 3D database. This paper only considers the lateral view data in the database, and extending the method to multiple view angles will be our further work.

ACKNOWLEDGMENTS

This work is supported in part by the National Natural Science Foundations of China (No.61100143, 61370128, and 61272353), Program for New Century Excellent Talents in University (NCET-13-0659), Fundamental Research Funds for the Central Universities (2014JBZ004), Beijing Higher Education Young Elite Teacher Project (YETP0583).

REFERENCES