RESEARCH ARTICLE OPEN ACCESS

A Method of DTW Based Gait Recognition and Gait Data from Kinect

YiFei Hu*

*(School of Information Science and Technology, Jinan University, Guangzhou, China)

--************************----------------------------------

Abstract:

Gait recognition as a field of biometrics more and more attention. Many people are interested in working with Kinect to study gait recognition. In this paper, we propose a Dynamic Time Warping (DTW) algorithm to classify the features by using the inherent physiological characteristics of human and Kinect's skeletal tracking function, extracting the distance between skeletal and joint points, the height of body mass center, height and centroid to the main skeletal joint points . The experimental results show that this method can obtain more than 90% recognition rate in the UPCV database and has good robustness.

Keywords **—Gait Recognition, Kinect, Centroid, DTW.**

--************************----------------------------------

I. INTRODUCTION

Biometrics is the use of human inherent physiological characteristics or behavioral characteristics of identity, it is also an important field of research in computer vision, pattern recognition, machine learning, and intelligent monitoring. Compared with the traditional identification technology, the use of biometric identity authentication has the advantages of security, accuracy, confidentiality and convenience. Face recognition [1] and fingerprint recognition [2] technologies have been widely used in commercial and forensic fields. However, these biometrics require testers to detect and require high image resolution. As a new biological feature, gait has the advantages of non-intrusiveness, long-range recognition, low resolution and difficult to conceal[3]. It not only overcomes the shortcomings of the traditional biological characteristics, but also requires relatively low equipment, which makes it have extensive economic value and application prospect in the fields of medical diagnosis, video monitoring and human-computer interaction, etc.

Gait recognition is the way to identify people by walking. Its process includes three parts: gait data preprocessing, gait feature extraction and classifier design. Among them, the most important part that affects gait recognition rate is gait feature extraction. Therefore, extracting effective and reliable gait features is crucial in gait recognition.

Currently, the research methods of gait recognition feature extraction can be divided into two types: model method and non-model method. The method of the model is to use the inherent physiological characteristics of the human to model and acquire the important parameters of the model as the gait features. Lee et al[4] divided the human body into seven parts according to the inherent physiological characteristics of the human beings. It uses seven ellipses modeling to represent different parts of the human body and the parameters such as ellipse long axis, short axis and centrifugal rate are used as characteristics. According to anatomical theory, Yoo et al [5] used the physiological structure of human being to abstract the main human joint and connect the joint points to construct 2D human body linear model. The swing angle of each line segment with respect to the vertical direction was taken as the gait feature. The model method is robust to the environment, perspective and occlusion problems. The disadvantage is that the gait sequence requires high definition and high computational complexity, and its disadvantage is that the gait image sequence requires high definition and the computation is too large. The non-model approach acquires the gait features by the statistical properties of spatiotemporal patterns generated by certain parameters (such as contour, texture, velocity, relative position, etc.) in successive image sequences. Han et al[6] proposed the use of gait energy map (GEI) to reflect changes in silhouette. Zhang [7] et al proposed a method of an activity energy map (AEI) established by calculating the frame difference between the contours. Wang Liang et al [8] used the distance from the center of the person's binarized image to the outline boundary as a feature. Non-model methods are simple, easy to extract, and small amount of calculation, but do not perform as well for perspective and occlusion issues.

In 2012, Microsoft released a 3D somatosensory camera--Kinect. The skeleton tracking function provided by the Kinect which can provide 3D coordinate data of the 20 skeletal points in real time. Which can clearly know the information about the skeletal joints of the human body and omit the complicated extraction process of human body model. It provides great convenience for the promotion of gait recognition technology.

The method of gait recognition based on Kinect can also be roughly divided into two categories. One is gait recognition based on Kinect depth image. Sivapalan et al. [9] proposed the use of Kinect depth images to obtain gait energy volume (GEV). Sabesan et al. [10] used Kinect's depth image to extend the gait energy map into a three-dimensional space, and achieved good recognition results. The other is gait recognition based on Kinect skeletal data. Ball et al.[11] proposed changes in the angle of the joints of the human lower extremity, focusing only on changes in the angle of the lower extremities, ignoring other important features. Preis et al. [12] extracted the gait and velocity as the dynamic gait feature. Because walking and walking speed are constantly changing when walking, the effect of using only gait and gait will not be very good. Ahmed et al. [13] used the change of relative spatio-temporal angle of different joint points and specific reference points in the skeletal model as the gait feature, so that the gait dynamic characteristics were relatively effectively expressed.

In summary, this paper uses Kinect skeletal data to study gait recognition. Unlike previous image gait recognition processing, skeletal data not only reduces the image processing work, but also extracts features more easily and conveniently. In this paper, we propose gait recognition based on the characteristics of human joint length, centroid height, body height and centroid to the key point. Experiments using a public database: UPCV [14], in the MATLAB environment simulation experiments. Experimental results show that the proposed method has achieved an ideal recognition rate.

II. OUR METHOD

This section mainly introduces the proposed method, which is divided into four parts: Kinect skeleton data flow, period detection, gait feature extraction and recognition classification.

As mentioned in Section 1, this paper uses the Kinect's skeletal tracking function and the inherent physiological characteristics of the human body to obtain the mean and standard deviation of the distances between the joint points as a feature. At the same time we also calculate the centroid, taking the height of the centroid and the distance from the centroid to the knuckle point as another feature. According to the extracted features, K-nearest neighbor classifier is used to complete the classification process. We have evaluated our method on a public database UPCV skeleton dataset collected using the Kinect v1 device .

A. Introducing Kinect

Kinect was an XBOX-360 accessory by Microsoft in 2010. It is a 3D motion-sensing camera with instant capture, image recognition, voice recognition and other functions, because it can capture color images, depth images, skeletal data stream and voice data stream of these data. The overall structure of the Kinect has three cameras, of which the middle of the lens is RGB color camera, used to obtain color images at 30 frames per second image, left and right lens are respectively infrared emitters and infrared CMOS cameras, which form a 3D structure of light depth sensing (3D Depth), the use of infrared detection of the relative position of players, the same principle with the human eye three-dimensional imaging. At the same time, Kinect has an array of microphones for voice recognition and an angle adjustment motor that adjusts the camera's pitch angle to increase its field of view.

The proposed method only uses skeletal data streams. Kinect V1 captures 3D coordinates of the 20 human bones at 30 fps. The proposed method uses only skeletal data streams. Kinect V1 captures 3D coordinates of 20 human bones at 30 fps. The coordinate system is based on the depth sensor center, as shown in Figure 1 to establish the Cartesian coordinate system. The x-axis is parallel to the Kinect, the y-axis is perpendicular to the bottom of the Kinect and the z-axis is the direction normal to the sensor, as shown in Figure 1 [13]. Figure 2 shows the human body model consisting of these 20 body joints.

Fig.1 Kinect sensor

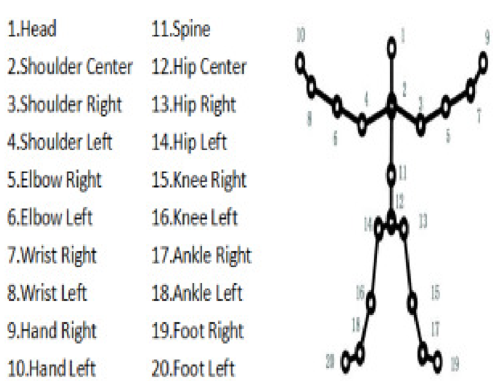

Fig.2 Human skeleton Coordinates

B. Periodic Inspection

According to a large number of past studies have shown that people walking under normal circumstances will show a regular change is a cyclical movement. It has a variety of forms of expression, including width-high ratio, step size, heart rate fluctuations and a series of variables can illustrate the cyclical changes. Among them, the movement state of the limbs presents a more obvious cyclical change. The most prominent periodical change is the distance between the two feet (ie, the distance between the feet or the step length)[15]. In this paper, we define the gait cycle as the periodic variation of the selected foot spacing. The data between the three wave crests with continuous foot spacing is taken as a gait cycle sequence. That is to say, when the foot spacing is the largest which is taken as the starting point and walked normally until Restore the original state, as shown in Figure 3.

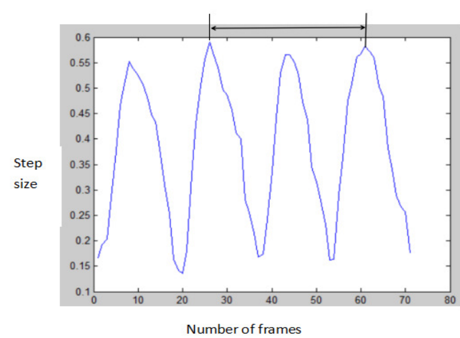

Fig.3 Gait Cycle

C. Feature Extraction

Gait recognition as one of the pattern recognition,one of the focuses of his research is how to choose and extract the characteristics that are effective and can represent the only individual. In the process of walking, no matter how to go, the overall physical structure of a person is always the same. Kinect obtains the 3D skeleton model of human body through depth image, using image processing, machine learning and other techniques to obtain skeleton coordinate data. Compared with other models, the most obvious advantage of 3D human skeleton model lies in that the skeleton model has known the three-dimensional coordinates of the skeletal joint points and can measure and calculate the relative distance between the various joint points of the human body, so that the three-dimensional human body model is intuitively displayed.

According to the inherent physiological characteristics of human and Kinect skeletal data to build a 3D human skeleton model, this paper from the human body overall skeletal structure considerations, the extraction of human skeletal joint distance between the points. Select several individuals with more significant differences in the length of the skeleton structure as a parameter. Respectively, head-neck length, shoulder width, hip width, upper torso length, arm length (forearm and back arm), leg length (thigh and calf), taking into account the complexity of hands and feet and joints compared to other shorter distances and other factors, Ignore the joints of hands and feet. Since the normal person has the symmetry, this paper selects one of the left and right sides of the arm and the leg as its length value.

The chosen distance formula is the Euclidean distance between two points, which is calculated as follows (Equation 1):

$$\mathcal{D}(\mathrm{i},\mathrm{j}) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$$

Where i and j are the coordinate values of the $i^{th}$and j^{th}skeleton points, respectively.

In addition to extracting the distance between the bones and joints as a feature, we also calculate the centroid of human beings and extract the distance from the center of mass to the key nodes as gait features. The research shows that the whole model can be divided into 15 parts, the method of division and its parameters (Fig.4) [16], and the centroid (PCOM) of each part can be calculated separately. The calculation formula is as follows:

$$\begin{cases} X_i = x_p w_p + x_d w_d \\ Y_i = y_p w_p + y_d w_d \\ Z_i = z_p w_p + z_d w_d \end{cases}$$

Where (X_i, Y_i, Z_i) is the coordinate value of the centroid PCOM of each part of the human body, $i = 1,2,\ldots,15$; (x_p, y_p, z_p) is the coordinate value of the proximal endpoint, (x_d, y_d, z_d) is the coordinate value of the distal endpoint, and (w_p, w_d) is the coordinate value percentage coefficient of the proximal end point and the distal endpoint, respectively. Then by the following part of PCOM synthesis, to find the body's total mass center (COM), calculated as follows:

$$\begin{cases} X_{com} = \dfrac{\sum_{i=1}^{15} m_i x_i}{M} = \sum_{i=1}^{15} l_i x_i \\ Y_{com} = \dfrac{\sum_{i=1}^{15} m_i y_i}{M} = \sum_{i=1}^{15} l_i y_i \\ Z_{com} = \dfrac{\sum_{i=1}^{15} m_i z_i}{M} = \sum_{i=1}^{15} l_i z_i \end{cases}$$

Where $(X_{com}, Y_{com}, Z_{com})$ is the coordinate value of the body's total centroid COM, (x_i, y_i, z_i) is the coordinate value of the PCOM of the first part, m is the mass of the i^{th} body partition, M is the total mass of the body model, l_i is the body part mass coefficient, ie m_i/M .

Body segment	Segment definitions (proximal end, distal end)	Segmental mass (m)/ Total body mass (M)	CM/Segment length	
			Proximal	Distal
Head and neck	Mid-pt of L/R temporal ridges, Mid-pt of L/R acromion processes	0.081	1.000	N/A
Thorax and abdomen	Mid-pt of L/R acromion processes, Mid-pt of L/R iliac crests	0.355	0.500	0.500
Upper arms	1-inch medial to shoulder joint, Mid-pt of L/M elbow joints	0.028	0.436	0.564
Forearms	Mid-pt of L/M elbow joints, Mid-pt of L/M wrist joints	0.016	0.430	0.570
Hands	Mid-pt of L/M wrist joints, 3rd metacarpal articulate	0.006	0.506	0.494
Pelvis	Mid-pt of L/R iliac crests Mid-pt of L/R greater trochanters	0.142	0.105	0.895
Thighs	3-inch medial to greater trochanter, 2-inch medial to femoral condyles	0.100	0.433	0.567
Legs	2-inch medial to femoral condyles, 1-inch medial to malleolus condyles	0.0465	0.433	0.567
Feet	Heel, 3rd metatarsal articulate	0.0145	0.500	0.500

Fig.4 The division and coefficient of human section

Calculate the center of mass of the human skeleton according to the above calculation formula, and then calculate the height, height, and the distance from the center of mass to the key nodes of the human skeleton (head, left foot and right foot)

III. PAGE STYLE

There are many common classifier algorithms, such as KNN, SVM, BP, Naïve Bayes and decision tree and some other machine learning algorithms. In the previous work related to gait recognition, some people chose KNN and DTW algorithm as classifier[17]. Taking into account each person walking speed are inconsistent, a gait cycle of the number of frames are inconsistent. In order to solve this problem, we choose to use dynamic time warping algorithm (DTW) [18].

Dynamic Time Warping Algorithm is a flexible pattern matching algorithm based on dynamic programming idea. By comparing the similarity between two time-varying feature sequences (the lengths may not be equal), it can find the least distance similarity path. Dynamic time warping algorithm has been widely used in the fields of speech recognition, gesture recognition, data mining and information retrieval.

Suppose gait feature sequences Q and C are M and N in length respectively: (one sequence is a reference template and one sequence is a test template, the value of each point in the sequence is the eigenvalue of each frame in the sequence)

$$Q = q_1, q_2, \dots q_m \dots, q_M; \quad C = c_1, c_2, \dots c_n \dots, c_N;$$

If N = M, the distance between two sequences is calculated directly.

If N ≠ M, then you need to align, in order to align these two sequences, the idea of using Dynamic Programming. You need to construct a M × N matrix grid. The matrix element (m, n) represents the distance $\mathrm{d}(q_m, c_n)$ between two points q_m and c_n and the two points are aligned. The common European distance is as follows: $\mathrm{d}(q_m, c_n) = \|q_m - c_n\|$. The dynamic time warping algorithm can be attributed to finding an optimal path through several grid points in the grid such that the sum of the distances between all the grid points in the path is the smallest. This path is defined as a warping path. The grid passed by the path is the point at which the alignment of the two sequences is calculated

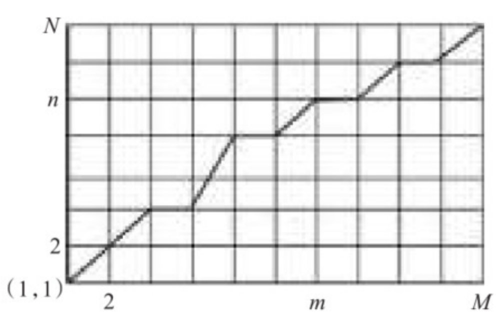

Fig.5 DTW sample grid diagram

In order to represent the lattice points in the DTW path, $\alpha(q_m, c_n)$ is defined as the previous lattice point of point (q_m, c_n), and $\alpha(q_m, c_n)$ must satisfy one of the following conditions:

$$\alpha(q_m, c_n) = (q_{m-1}, c_{n-1}) = (q_m - 1, c_n)$$

$$\alpha(q_m, c_n) = (q_{m-1}, c_{n-1}) = (q_m, c_n - 1)$$

$$\alpha(q_m, c_n) = (q_{m-1}, c_{n-1}) = (q_m - 1, c_n - 1)$$

Define the sum $\lambda(q_m, c_n)$ of the path distances from the starting point to the point (m,n). Based on the above constraints, the following formula is obtained:

$$\lambda(1,1) = \mathrm{d}(q_1, c_1)$$

$$\lambda(q_m, c_n) = \mathrm{d}(q_m, c_n) + \min\{\lambda(q_m - 1, c_n), \lambda(q_m, c_n - 1), \lambda(q_m - 1, c_n - 1)\}$$

Through the use of these formulas can calculate the sum of the path distance, the minimum distance is the best path.

IV. CONCLUSIONS

The paper carries on the simulation experiment in the Matlab environment, the simulation interface like chart 9 shows. The skeletal gait data set used is the UPCV gait skeletal database published by Kastaniotis D et al. There are 30 people in the UPCV database and each contains 5 gait sequences. After the database research, we found that due to the lack of data, lack of cycles and other factors, a small part of the data is invalid, so the final sample of 30 people selected data, each sample contains four gait sequences, a total of 120 gait sequences . In this paper, three sample sequences of each sample are selected as the training set and the other sample sequence as the test set.

In order to evaluate the classification result, Correct Classification Rate (CCR) was introduced as the evaluation index. Equation can define the correct classification rate CCR . The dynamic time warping algorithm is used to match the training set and the test set according to the above mentioned method to obtain the feature with the smallest distance as the test category.

$$\text{CCR} = \frac{Correct\ number}{Number\ of\ test\ samples} \times 100\%$$

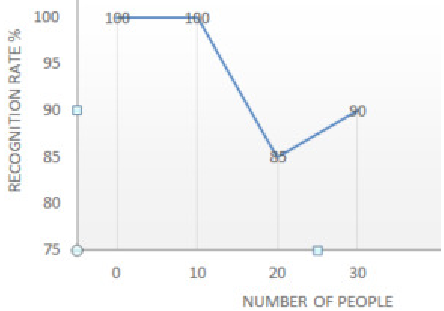

Fig. 6 correct recognition rate (CCR)

The results obtained by the simulation experiment shown in Figure 6, we can see that there is fluctuation in recognition rate, the reason is that the person who misidentified is 10-20 among them. Overall, our method works well in the UPCV database. In the meantime, to illustrate our method, some simple comparisons with other methods are given in Table 1. Compared with the methods provided by other papers, we can find that using Kinect skeleton data does not require image processing, the feature extraction is simple and convenient, and the recognition effect is also ideal.

TABLE I
QUALITATIVE COMPARISONS WITH OTHER METHODS

方法	数据库大小	识别率/%
Ball et al.[11] paper	4	66.7 100.00
Preis et al.[12] paper	9	85.1 100.00
Kastaniotis et al. [14] paper	30	77.27 90.00

V. CONCLUSIONS

In this paper, we use skeleton data from Kinect skeletal tracking to extract the distance between skeletal and joint points using the physiological characteristics of the human body. Meanwhile, we calculate the center of gravity of the human body and obtain the height of the centroid, the height and the distance from the center of mass to the key skeletal points Gait characteristics go to other people's identity, in the simulation has been a good result. The method has the advantages of simple extraction features and good results in the simulation. Kinect can acquire not only the bone data, but also the depth image data. The next step is to combine the depth image data with the bone data for those studies on gait recognition. And the improved algorithm can be applied to the real-time system.

REFERENCES

[1] Xiaofei He; Shuicheng Yan; Yuxiao Hu; Niyogi, P.; Hong-Jiang Zhang, "Face recognition using Laplacianfaces," in Pattern Analysis and Machine Intelligence, IEEE Transactions on ,vol.27, no.3, pp.328-340, March 2005.

[2] Jain, A.K.; Hong, L.; Pankanti, S.; Bolle, R., "An identity-authentication system using fingerprints," in Proceedings of the IEEE , vol.85, no.9, pp.1365-1388, Sep 1997.

[3] Lee T K M, Belkhatir M, Sanei S. A comprehensive review of past and present vision-based techniques for gait recognition[J]. Multimed Tools Appl: 2014, 72(3):2833-2869.

[4] Lee L, Grimson W E L. Gait analysis for recognition and classification[C]// Proceeding of Automatic Face and Gesture Recognition, the Fifth IEEE International Conference on. 2002: 148-155.

[5] Yoo J H, Nixon M S, Harris C J. Extracting gait signatures based on anatomical knowledge[C]// Proceeding of The British Machine Vision Association and Society for Pattern Recognition, London, 2002.

[6] Han J, Bhanu B. Individual recognition using gait energy image[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence: 2006, 28(2): 316-322.

[7] Zhang E H, Zhao Y W, Xiong W. Active energy image plus 2DLPP for gait recognition[J]. Signal Processing: 2010, 90 (7): 2295-2302.

[8] Wang Liang ,HU Wei Ming ,TAN Tie Niu .Gait-Based Human Identi fi cati 0n [J].CHINESE JOURNAL OF COM PUTER, 2003,26(3):353-360.

[9] Sivapalan S, Chen D, Denman S, et al. Gait energy volumes and frontal gait recognition using depth images[C]// Proceeding of International Joint Conference on Biometrics. IEEE Computer Society, 2011: 1-6.

[10] Sivapalan S, Chen D, Denman S, et al. Gait energy volumes and frontal gait recognition using depth images, F, 2011 [C].

[11] Ball A, Rye D, Ramos F, et al. Unsupervised clustering of people from 'skeleton' data[C]. Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction. USA: ACM, 2012:225-226.

[12] Preis J, Kessel M, Linnhoff-Popien C, et al. Gait recognition with Kinect[C]// 1st International Workshop on Kinect in Pervasive Computing, 2012.

[13] F Ahmed，PP Paul，M Gavrilova. Kinect-Based Gait Recognition Using Sequences of the Most Relevant Joint Relative Angles[J].Journal of Wscg, 2015, 23(2):147-156.

[14] Kastaniotis D, Theodorakopoulos I, Theoharatos C, et al. A framework for gait-based recognition using Kinect[J]. Pattern Recognition Letters: 2015, 68: 327-335.

[15] AK Sadhu,S Saha,A Konar,R Janarthanan. Person Identification Using Kinect Sensor [J].International Conference on Control, 2014(4):214-218

[16] Yeung L F, Cheng K C, Fong C H, et al. Evaluation of the Microsoft Kinect as a clinical assessment tool of body sway[J]. Gait & Posture: 2014, 40(4): 532-8.

[17] Andersson V O, Araujo R M. Full Body person identification using the Kinect sensor[C]// Proceedings of IEEE International Conference on Tools with Artificial Intelligence. 2014: 627-633.

[18] Dynamic time warping[M]. Berlin Heidelberg: Springer, 2007.