Real Time Gait Recognition System based on Kinect Skeleton Feature

Shuming Jiang[1], Yufei Wang[2], Yuanyuan Zhang[1], and Jiande Sun[2]

[1] Information Research Institute, Shandong Academy of Science, Jinan 250014, China
[2] School of Information Science and Engineering, Shandong University, Jinan 250100, China

Abstract. Gait recognition is a kind of biometric feature recognition technique, which utilizes the pose of walking to recognize the identity. Generally people analyze the normal video data to extract the gait feature. These days, some researchers take advantage of Kinect to get the depth information or the position of joints for recognition. This paper mainly focus on the length of bones namely static feature and the angles of joints namely dynamic feature based on Kinect skeleton information. After preprocessing, we stored the two kinds of feature templates into database which we established for the system. For the static feature, we calculate the distance with Euclidean distance, and we calculated the distance in dynamic time warping algorithm (DTW) for the dynamic distance. We make a feature fusion for the distance between the static and dynamic. At last, we used the nearest neighbor (NN) classifier to finish the classification, and we got a real time recognition system and a good recognition result.

1 Introduction

Gait recognition which uses the walking posture of the people for people identification is a new biometric identification technology. This technology is non-invasive,long-distance identification and it is difficult to hide the biological characteristics, which other biometric identification technologies do not have.Based on these advantages, gait recognition has board application prospects in entrance guard system, security surveillance, human-computer interaction, medical diagnostics, and other fields. Because the research is still in the primary stage, the existing methods are generally utilize all kinds of moving object detection algorithm such as Inter-frame Difference, background subtraction to separate the moving people from the background, and then extract the feature which can best distinguish the person, and finally make the classification.

Lee et al[1]characterize the human body as seven ellipses. Through analyzing the change of centroid and eccentricity, they extract the feature of the gait. Zhang et al[2]propose a model called 3DHW(3-Dimension Human Walking), which abstract the whole body into a connected rigid body. The body is made of 21 points, 14 line segments and 12 joints. Wang et al[3] make use of the distance between centroid and the edge points of silhouette to get the feature of walking people. He et al[4] model for the moving leg of human, and propose a method of gait recognition based on the angles of joints. Bobick et al[5] take the length of body, torso, legs and step to recognize the gait. Kale et al[6]make the extracted moving people into binary image, and use the width of side silhouette as features for classification. In these general methods, is the feature extraction good or not, and is classification successful or not always depends on the gait extraction. However, due to the dynamic changes of background, such as weather, illumination, shadow, etc. the gait extraction becomes a difficult mission. Kinect can be barely affected by these factors. Now some researchers take advantage of Kinect into the research of gait recognition. Sivapalan et al[7]use the deep image information captured by Kinect, and analyze the gait energy image to identify the gait. Gabel et al[8]make use of 20 joints' positions collected by Kinect for gait recognition.

This paper proposed a real time gait recognition system based on Kinect skeleton feature. Kinect can track and extract the joints of human, and provide the 3D coordinates of 20 joints. Because of this, we can complete the gait extraction easily, and can also get static gait feature and dynamic gait feature by calculating the coordinates of joints. After fitting the original data during the preprocessing , we established a system fitted data base. We calculated the distance between each templates and test samples in DTW(Dynamic Time warping), and made a feature fusion in matching layer. Finally we get the classification by using NN classifier.

2 Kinect

Kinect is a line of motion sensing input devices by Microsoft for Xbox 360 and Windows PCs. It has a normal webcam and a depth sensor which can provide RGB-D image.The depth sensor consists of an infrared laser projector combined with a monochrome CMOS sensor, which captures video data in 3D under any ambient light conditions. The detecting range of depth sensor is 0.8m-4.0m . It can utilize the machine learning technique to analyze and track the position of 20 skeleton joints in real time. Each joint includes X, Y, and Z, three coordinates information, whose accuracy can reach the millimeter level. The reasons of taking using Kinect to get through the gait detection and feature extraction are following:

1. Kinect is barely affected by the illumination and background, which often affect the recognition rate in general methods. As shown in Fig.1
2. Kinect is barely affected by wearing. According to the result of experiment, Kinect can even recognize the joints of women's legs under the condition of wearing dress or skirt.
3. Depth information can separate the moving people from background conveniently. When the background includes some moving objects, depth information can also work.

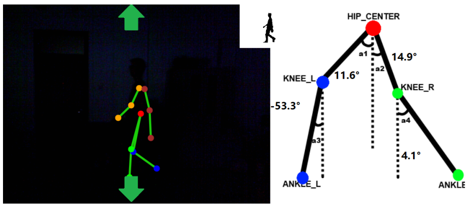

Fig. 1. Experiment in the dark condition.

3 Feature Extraction

3.1 Static Feature

The static feature is a kind of feature that barely changes during the process of walking, such as body height, the length of leg, the length of arm, etc. According to anthropometry [9] , the position of each joint, the measurement of bone, and the scale of skeleton can distinguish each person theoretically. As the above suggest, Kinect is able to get the 3D coordinate of 20 joints, so each segment of skeleton can be used as static feature in theory. This paper selected the static feature according to the following principles:

1. The feature which can distinguish each person effectively.
2. On the premise of the first condition, the dimensions of feature vectors should be as small as possible, so that can reduce the computation complexity.

During the experiment, we perceived that some joint points which are hardly influenced by people's wearing and noise make a bigger contribution to the recognition,so we would like to use these joints. Based on the principles mentioned above, this paper selected following static feature. This paper defined the distance between HIP_RIGHT and KNEE_RIGHT as length of thigh, denoted by d_1, defined the distance between KNEE_RIGHT and ANKLE_RIGHT as length of calf, denoted by d_2, defined the distance between SHOULDER_RIGHT and ELBOW_RIGHT added the distance between ELBOW_RIGHT and WRIST_RIG HT as length of arm, denoted by d_3, defined the distance between HEAD and (FOOT_RIGHT+FOOT_LEFT)/2 as body height, denoted by d_4, defined the radio of thigh and height, namely d_1/d_4, denoted by d_5.As shown in Fig.2. We can get the 3D coordinates of those mentioned joints provided by Kinect SDK, and according to the following formula,

$$d = \sqrt{(x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2} \tag{1}$$

We can calculate the value of d_1,d_2,d_3,d_4,d_5, establishing the feature vector D=(d_1,d_2,d_3,d_4,d_5). We stored this vector into the database as feature template.

Due to the existing gait databases are mostly based on video frame data, such as CASIA [10] , are not suitable for our system, we had to build up our own database. There are 6 tables in the database, respectively are $peopleInfo$,

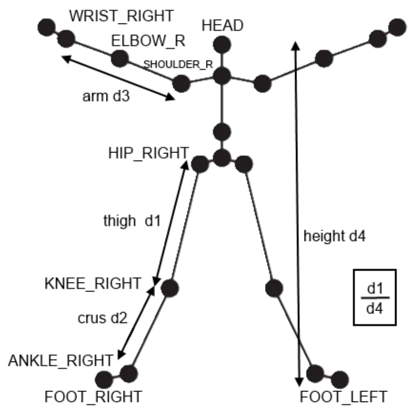

Fig. 2. Schematic of static feature.

staticSkeletonInfo, *angleOne*, *angleTwo*, *angleThree*, *andangleFour*. Here, we stored the feature vector D=(d_1,d_2,d_3,d_4,d_5)into the *staticSkeletonInfo* , and built a connection with *peopleInfo* by key (*peopleID*).

3.2 Dynamic Feature

The dynamic features are changing at any time during the process of walking, and the changing always presents a periodicity. Tian et al [11] analyzed the changing angle of leg during human walking on a treadmill through the video data captured by professional 3D camera VICON MX. The result of experiment provided a theoretical basis for the angle of leg can be used as a kind of dynamic feature. He et al [12] separated the people from the video image, and extracted the leg angle for gait recognition.

In this paper, we mainly selected two kinds of dynamic features, including angles of swinging legs and angles of swinging arms. In the experiment, the Kinect placed on the platform of $1.3m$. Due to the viewing angle of Kinect is $\pm 28.5°$, to make sure that Kinect can record the whole process of walking, test people are suggested to stand $3.3m$ far away from Kinect, and start walking behind optical axis at least $1.8m$. As shown in Fig.3.

People walk along the negative X-axis of Kinect. The leg far away from Kinect would be covered by another leg, causing inaccuracy in data collecting, so we decided to get the joint angle data from the near leg. As shown in Fig.4, this paper defined the angle between the right thigh and vertical direction as a_1, defined the angle between the right calf and vertical direction as a_2, defined the angle between the right arm and vertical direction as a_3,and defined the angle between the right forearm and vertical direction as a_4. Assuming the coordinate of HIP_CENTER is (x, y), and the coordinate of KNEE_RIGHT is (x_1, y_1), so

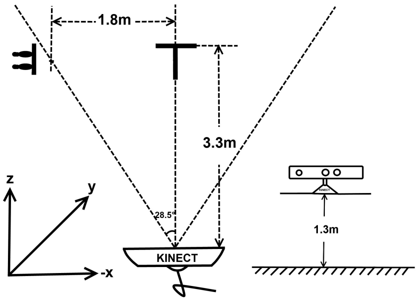

Fig. 3. Experiment environment.

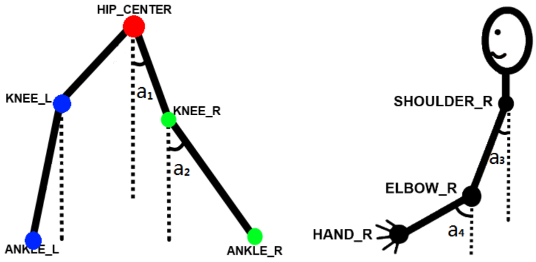

Fig. 4. Schematic of dynamic feature.

we can get a_1 according to the following formula:

$$\tan\angle a_1 = \left(\frac{x - x_1}{y - y_1}\right) \qquad \Rightarrow \qquad a_1 = \arctan\left(\frac{x - x_1}{y - y_1}\right), \tag{2}$$

We can also get a_2,a_3,a_4 in the same way. These four angles can made up dynamic feature vector a=(a_1,a_2,a_3,a_4). We built up four arrays in order to add corresponding angle in each frame, so we got four sequences of angles, $a_1(t)$,$a_2(t)$,$a_3(t)$,$a_4(t)$. The dynamic feature vector is finally expressed by $A[a(t)]$.

4 Data Processing

4.1 Gait Cycle Analysis

Human's gait information shows some certain periodicity.A gait cycle contains a large number of human gait features [13] . Fig.5 shows one person's gait cycle which is defined as the gait motion set that the same foot touches the ground twice adjacently. Since the gait is periodic, the gait cycle feature could be expressed through extracting the feature vector of one gait cycle. So we should extract one gait cycle from the data we got.

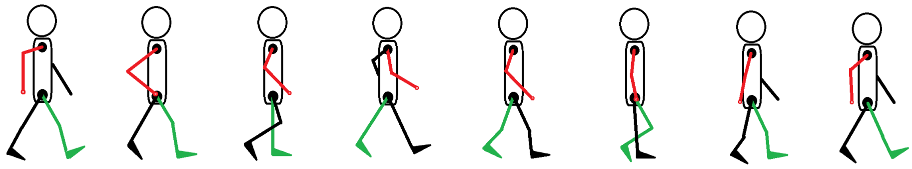

Fig. 5. One gait cycle, we can conspicuously notice the periodism from the leg and the arm separately marked by the green and the red.

Through analyzing the data of the video and angle collected by the Kinect, the paper regard the state when two feet get the farthest distance with each other and meanwhile the right foot is forward as the starting point of the gait cycle. Also, the paper regard the next similar state as the ending point of the gait cycle. We separately record the frame number of the two points and get the data between the starting point and ending point from the original data as a gait cycle. The 4 angle sequences $a_1(t)$,$a_2(t)$,$a_3(t)$,$a_4(t)$ turn into $a_1(T)$,$a_2(T)$,$a_3(T)$,$a_4(T)$. As shown in Fig.6.

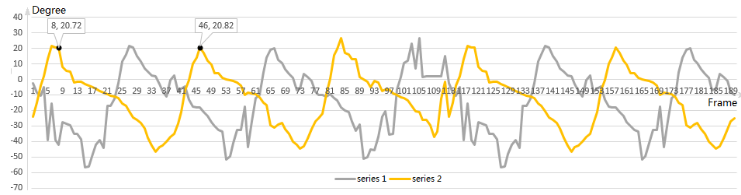

Fig. 6. Gait cycle analysis, two marked points are separately the starting point and ending point.

4.2 Curve Fitting

In this paper, the original discrete data will be using least squares polynomial fitting method. Fitting function is $f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$.According to the experiments, when $n = 7$, the fitting result would be the best. Fitting result is shown in Fig.7. The 4 angle sequences $a_1(T)$,$a_2(T)$,$a_3(T)$,$a_4(T)$, after the curve fitting, they would turn into $a_1(T_{fitted})$,$a_2(T_{fitted})$,$a_3(T_{fitted})$,$a_4(T_{fitted})$. We respectively store these four sequences as template in the database.

4.3 Dynamic Time Warping (DTW)

Since we had already gotten the template, we want to calculate the distance between the tester and templates. However, the walking cycle is not entirely

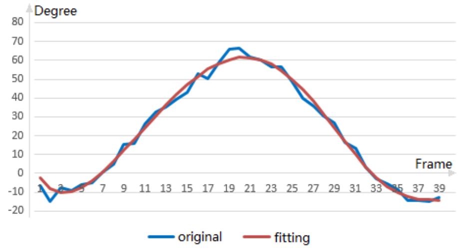

Fig. 7. Curve fitting.

consistent even for the same person. Since the length of the sequence is inconsistent, we could not directly calculate the Euclidean distance. As shown in the Fig.8. We can see that the curves are not completely the same even through the same person. But we can also notice that the tendency of three curves are the same.

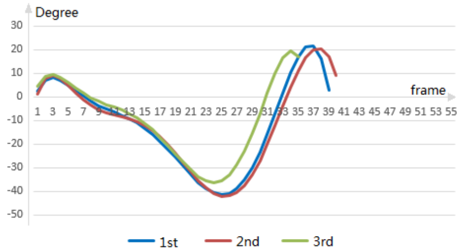

Fig. 8. Three times walking result of the same person.

The paper utilize the DTW which is widely applied in the speech recognition. DTW is able to compare the relationship between two templates which have the different time scale.Finally, the shortest cumulative distance obtained is the DTW distance between two templates, shown in Fig.9.

5 Classification and Recognition

5.1 Recognition of Static Features

As mentioned above, Kinect can be used to extract static features based on bone length and store static feature vectors as templates in the user database. When tester is passing in front of Kinect, the same method can be used to extract static feature vectors. In this paper, we used NN classifier to perform the final classification. Here we selected Euclidean distance as the criteria for

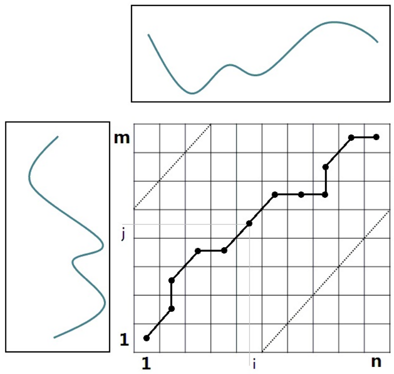

Fig. 9. Dynamic Time Warping schematics, the black segment is the path with the shortest cumulative distance .

calculating the distance between sample and template. Templates for each row in the database and the test samples calculated Euclidean distance and returned the minimum distance corresponding to *peopleID*, then utilized *peopleID* to retrieve the corresponding user name.

5.2 Recognition of Dynamic Features

For the recognition of dynamic features, we still used NN classifier for classification. However, unlike the static feature classification, when calculating the distance we used the aforementioned DTW algorithm to calculate the sum of cumulative minimum distance between sample and template, namely the DTW distance, and this distance is set as the distance criteria between sample and template.

We used the four angle sequences generated by test samples and the statistics in corresponding angle table in database to calculate DTW distance, separately denoted as d_{a1},d_{a2},d_{a3},d_{a4}, then linearly normalized these four distances, a.k.a. formula :

$$\hat{d} = \frac{d}{max(D)}, \tag{3}$$

Where $\hat{d}$ is normalized value, d is the original distance to be normalized, D is the set of all the calculated distance for each row of corresponding angle table in test samples and database. $\hat{d}_{a1}, \hat{d}_{a2}, \hat{d}_{a3}, \hat{d}_{a4}$ denote the normalized distance, D_{total} denotes the sum, i.e. $D_{total} = \hat{d}_{a1} + \hat{d}_{a2} + \hat{d}_{a3} + \hat{d}_{a4}$. And it was characterized as final distance and then sorted ascendingly. Then categorized the test sample into the training sample set of minimum distance.

5.3 Recognition of Feature Fusion

In this paper, we extracted static feature based on bone length and dynamic feature based on angle change of lower limb and arm. As indicated by the experiments, using these two features separately to perform gait recognition is not satisfactory, thus we use feature fusion technique to classify and recognize these two features after feature fusion.

Biometric recognition can be broadly divided into three procedures: feature extraction, template matching and classification. According to these three procedures, multi-biometric features fusion can be correspondingly divided into three levels [14] : feature level fusion, matching level fusion, decision-making level fusion.

Distance calculated by different methods using static and dynamic features is used as matching score for each part respectively, attempting to perform feature fusion of static and dynamic features in matching level.

Normalization of matching scores of each component is cardinal to feature fusion in matching level, here we use linear normalization method.

$$\hat{s} = \frac{s - min(S)}{max(S) - min(S)}, \tag{4}$$

Where S is the matching score matrix before normalization, and its elements are s. Scoring matrix after normalization is $\hat{S}$, and its element is $\hat{s}$. Matching scores after normalization are mapped to the range $[0, 1]$, enabling the later fusion of the static and dynamic skeletal features. There are many rules in data fusion, whose purpose is to perform calculation using matching score of each feature according to certain fusion rules, to obtain a matching score with higher separability. In this paper, the addition rule is used to add up normalized scores to get the final score, and perform classification. The addition rules:

$$F = \frac{1}{R}\sum_{i=1}^{R} s_i^n, \tag{5}$$

Where F represents the score after fusion, s_i^n denotes the ith normalized matching score, $i = 1, 2, \ldots, R$. We then reused NN classifier to classify and recognize after getting the matching score F.

6 Experiment and Result

The experiment was proceeded in the environment mentioned above, as shown in Fig.3. The paper established a Kinect fitted database which included 10 persons' gait information. Each person walks for 5 times as training data, in other words, there are 50 series of gait templates in the database. When the test begin, each person also walks in front of Kinect for 5 times as testing data. In order to estimate the performance of system,the paper made a statistics about the CCR (Correct Classification Rate) of using static feature (STA) alone, using dynamic

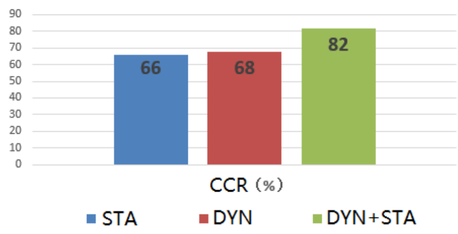

Fig. 10. CCR of 3 different features.

feature (DYN) alone, and after feature fusion (DYN+STA). The results are shown in Fig.10.

In the figure we can obviously find that the CCR of feature fusion (DYN+STA) is higher than STA and DYN. In order to compare the verification performance of three features, paper also made the ROC of three features under the same classification. As shown in Fig.11.

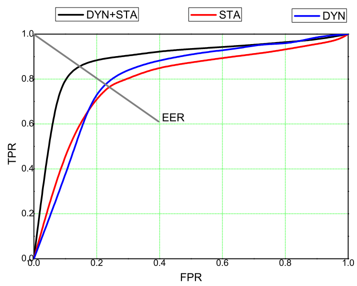

Fig. 11. ROC of three different features under the same classification.

In the ROC, the line which is closer to the upper left corner has the better classification method. Due to using the same classification, we can indirectly proved that the line is closer, the feature is better. In the figure we can see that the line of feature fusion (DYN+STA) is much closer than STA and DYN.

In the research of gait recognition, different gait database would get the different recognition rate. Because of using our own database, it's hard to compare with other methods. In order to compare with others, this paper learned the holistic thinking and algorithm flow of paper [15, 16] , and transplanted their methods into our database for comparison experiment. These two papers are all based on the general video methods. Paper [15] proposed that make use of leg angles, transform the angle sequence into Fourier series, search the coefficient of each harmonic by genetic algorithm, and generate the feature vector. Paper [16] extracted the key frame in the period of gait, locate the position of leg joints, and calculate the angle of leg joints. Table.1 shows the comparison of CCR.

Table 1. The comparison of three papers in CCR.

algorithm	CCR
Weihua H et al [15]	78.5%
AI-HUA WANG et al [16]	80%
This paper	82%

The experiment result shows that the CCR of this paper is better because of adding static feature and making a feature fusion.

7 Conclusion

People has been paying more attention to the biometric recognition. It has many kinds of methods to work things out, and no need to say which method is the best because each method has its own merits and faults . This paper focus on the gait recognition which is frontier area in biometric recognition sector. Other than general video processing methods, the paper took advantage of Kinect to extract the gait feature. The paper extracted the length of skeleton as the static feature, and extracted the angle of swing legs and arms as the dynamic feature by Kinect. On the basis of this, we made a feature fusion and stored the feature vector into the database which was established by ourselves. The paper proved the CCR is much higher after the feature fusion, reached 82%. The paper also compared with other two methods to prove the advantage of our feature extracted by Kinect.

References

1. Lee L, Grimson W E L. Gait analysis for recognition and classification[C]//Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on. IEEE, 2002: 148-155.
2. Bofeng Z, Jingru Z, Ke Y, et al. Research on Gait Feature Extracting Methods Based on Human Walking Model[J]. Computer Application and Software, 2009, 26(5): 198-201.
3. Wang L, Tan T, Ning H, et al. Silhouette analysis-based gait recognition for human identification[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2003, 25(12): 1505-1518.
4. Weihua H, Ping L, Haijun Y. Gait Recognition Using the Information About Crura's Joint Angle[J]. 1994.
5. Bobick A F, Davis J W. The recognition of human movement using temporal templates[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2001, 23(3): 257-267.

6. Kale A, Cuntoor N, Chellappa R. A framework for activity-specific human identification[C]//Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on. IEEE, 2002, 4: IV-3660-IV-3663.
7. Sivapalan S, Chen D, Denman S, et al. Gait energy volumes and frontal gait recognition using depth images[C]//Biometrics (IJCB), 2011 International Joint Conference on. IEEE, 2011: 1-6.
8. Gabel M, Gilad-Bachrach R, Renshaw E, et al. Full body gait analysis with Kinect[C]//Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE. IEEE, 2012: 1964-1967.
9. Wang Guilun. Anthropometric and application The Anthropology Department of Biology Department of Fudan University.
10. http://www.cbsr.ia.ac.cn/china/Gait%20Databases%20CH.asp,Powered by Institute of Automation, Chinese Academy of Sciences.
11. Tian W, Cong Q, Yan Z, et al. Spatio-temporal characteristics of human gaits based on joint angle analysis[C]//Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on. IEEE, 2010, 6: 439-442.]
12. He W, Li P. Gait recognition using the temporal information of leg angles[C]//Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on. IEEE, 2010, 5: 78-83.
13. Makihara Y, Mannami H, Yagi Y. Gait analysis of gender and age using a large-scale multi-view gait database[M]//Computer Vision-ACCV 2010. Springer Berlin Heidelberg, 2011: 440-451.
14. Zhigang Fang, Weizhong Ye. A Survey on Multi-biometrics. Computer Engineering.2003,29(9):140-142
15. Weihua H, Ping L, Yumei W et al. Gait Recognition Using the Information About Crura's Joint Angle [C] 2006.
16. AI-HUA WANG, JI-WEILIU. Gait recognition method based on position human-body joints[J].Proceedings of the 2007 International Conference on Wavelet Analysis and Pattern Recognition, Beijing, China,2-4 Nov.2007