CrossMark

# View-invariant gait recognition based on kinect skeleton feature

Jiande Sun[1,2] · Yufei Wang[3] · Jing Li[1,4] ·
Wenbo Wan[1,2] · De Cheng[5] · Huaxiang Zhang[1,2]

© Springer Science+Business Media, LLC, part of Springer Nature 2018

**Abstract** Gait recognition is a popular remote biometric identification technology. Its robustness against view variation is one of the challenges in the field of gait recognition. In this paper, the second-generation Kinect (2G–Kinect) is used as a tool to build a 3D–skeleton-based gait dataset, which includes both 2D silhouette images captured by 2G–Kinect and their corresponding 3D coordinates of skeleton joints. Given this dataset, a human walking model is constructed. Referring to the walking model, the length of some specific skeletons is selected as the static features, and the angles of swing limbs as the dynamic features, which are verified to be view-invariant. In addition, the gait recognition abilities of the static and dynamic features are investigated respectively. Given the investigation, a view-invariant gait recognition scheme is proposed based on the matching-level-fusion of the static and dynamic features, and the nearest neighbor (NN) method is used for recognition. Comparison between the existing Kinect-based gait recognition method and the proposed one on different datasets show that the proposed one has better recognition performance.

**Keywords** Gait Recognition · Second Generation Kinect · View-Invariant · 3D Joint Information · Gait Dataset

✉ Jiande Sun
  jiandesun@hotmail.com

1   School of Information Science and Engineering, Shandong Normal University, Jinan, China

2   Institute of Data Science and Technology, Shandong Normal University, Jinan, China

3   School of Information Science and Engineering, Shandong University, Jinan, China

4   School of Mechanical and Electrical Engineering, Shandong Management University, Jinan, China

5   Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an, China

🖄 Springer

# 1 Introduction

Gait Recognition has received increasing attentions as a remote biometric identification technology, because it can achieve identification at a long distance, where other identification technologies can't work, and it needs little subject cooperation. Moreover, differing from other biometric identification technologies such as face and fingerprint recognition [4, 29], 1) gait has high security. It is a kind of behavioral biometrics, which is a kind of temporal sequence verification, so it is hard to imitate. In the movie, Mission: Impossible-Rogue Nation (2015), the gait analysis technology makes the access control system can't be broken except to replace the gait data inside the system. 2) Gait can be captured in a long distance without the subject's realizing. It reduces that probability of the subject changing his gait consciously. 3) Gait is easy to be used in the present surveillance system. Though some psychological studies showed that people can recognize a friend with 70–80% accuracy based on only the gait [20], there have already been many gait recognition applications in the fields of criminal investigation, medical treatment, identity recognition, etc.

Generally, the existing gait recognition methods can be divided into 2D–based and 3D–based methods according to the format of gait. The 2D–based gait recognition methods depend on the human silhouette captured by one 2D camera, which is the most common situation of the video surveillance. The 2D–based gait recognition methods are dominant in this field of gait recognition, and they are usually divided into model-free and model-based methods.

Model-free methods are also known as appearance-based method. They generate gait signatures directly based on the silhouettes, which are extracted from the video sequences with background subtraction. Gait energy image (GEI) [10] is the most popular appearance-based gait representation, and it represents the spatial and temporal gait information in a grey image as shown in Fig. 1a via averaging silhouettes over one gait cycle. Motion silhouette image (MSI) [24] is like GEI, and is a grey image too. The intensity of an MSI is determined by a function of the temporal history of the motion of each pixel, as shown in Fig. 1b. The intensity of an MSI represents motion information during one gait cycle. Shape Variation Based (SVB) Frieze Pattern proposed in [25] projects the silhouettes horizontally and vertically to represent the gait information, and uses keyframe subtraction to reduce the effects of appearance changes on the silhouettes, which is shown in Fig. 1c. Gait Entropy Image (GEnI) [3] encodes the randomness of pixel values in the silhouette images based on Shannon entropy, which makes it robust against appearance changes, such as carrying and clothing, as shown in Fig. 1d. The Chrono-Gait Image (CGI) shown in Fig. 1e utilize a colour mapping function to encode the silhouette images without losing too much temporal relationship between them [44], and it can preserve more temporal information of one image. The model-free methods are



(a)          (b)          (c)          (d)          (e)
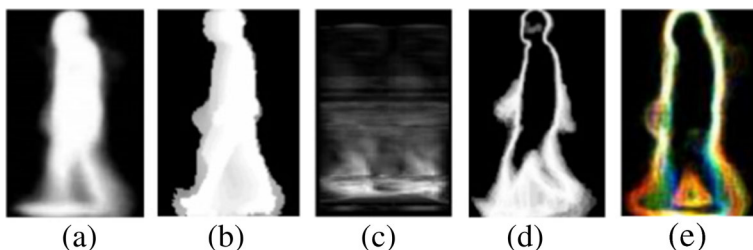
**Fig. 1** Examples of model-free methods

computationally efficient, but their performance will drop due to the appearance changing caused by carrying and clothing.

Comparing with model-free methods, the model-based ones are more robust, but requires high video resolution and computational load. They establish the suitable skeleton or joint model by integrating the shape and dynamics features of the human body from video sequences, and recognize the individuals based on the variation of the parameters in this model. Cunado et al. [5] modelled gait as an articulated pendulum and extracted the line via the dynamic Hough transform to represent the thigh in each frame, as shown in Fig. 2a. Johnson et al. [16] identified the person based on the static body parameters recovered from the walking action across multiple views, which is shown in Fig. 2b. Guo et al. [8] modelled the human body structure from the silhouette by stick figure model as shown in Fig. 2c, which had 10 sticks articulated with 6 joints. Rohr [36] proposed a volumetric model for the analysis of human motion, which used 14 elliptical cylinders to model the human body as shown in Fig. 2d. Wang et al. [26] extracted the dynamic body feature via the model-based approach, which can track the joint-angle trajectories of lower limbs as shown in Fig. 2e, and the static feature by Procrustes shape analysis in a compact representation. They represented the gait via the fusion of the static and dynamic features.

In general, 2D–based methods are easy to implement because they only need a normal camera, while 3D–based methods usually utilize multiple calibrated cameras or the cameras with depth sensors to extract 3D gait features. Zhao et al. [52] proposed to build the 3D skeleton model based on 10 joints and 24 degrees of freedom (DOFs) captured by multiple cameras, which is shown in Fig. 3a. Yamauchi et al. [48] captured the dense 3D range gait from a projector-camera system, which can be used to recognize individuals at different poses, as shown in Fig. 3b. Krzeszowski et al. [21] built a system with 4 calibrated and synchronized cameras to estimate the 3D gait motion using the video sequences, and recognize the view-variant gaits based on marker-less 3D motion tracking, as shown in Fig. 3c.

Both 2D–based and 3D–based methods have their own advantages in gait recognition. However, their robustness against the variation of observation view [22], walking speed [43], clothing [12], and belongs [41] are still challenges, when they will be used in more natural cases.

In this paper, we focus on the view-invariant Kinect-based gait recognition method as it is relatively difficult to derive the information of view from only silhouettes [7]. Kale et al. [18] presented a gait recognition method to identify people based on static body parameters, which are extracted from the walking across multiple views. Jean et al. [15] proposed a framework to
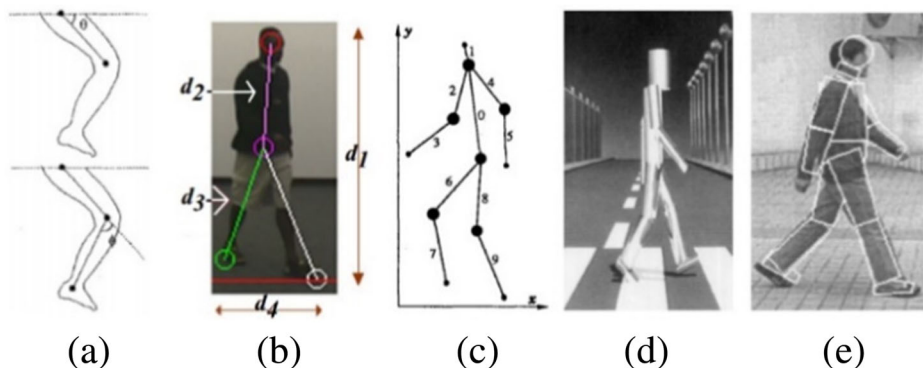


Fig. 2 Examples of model-based methods
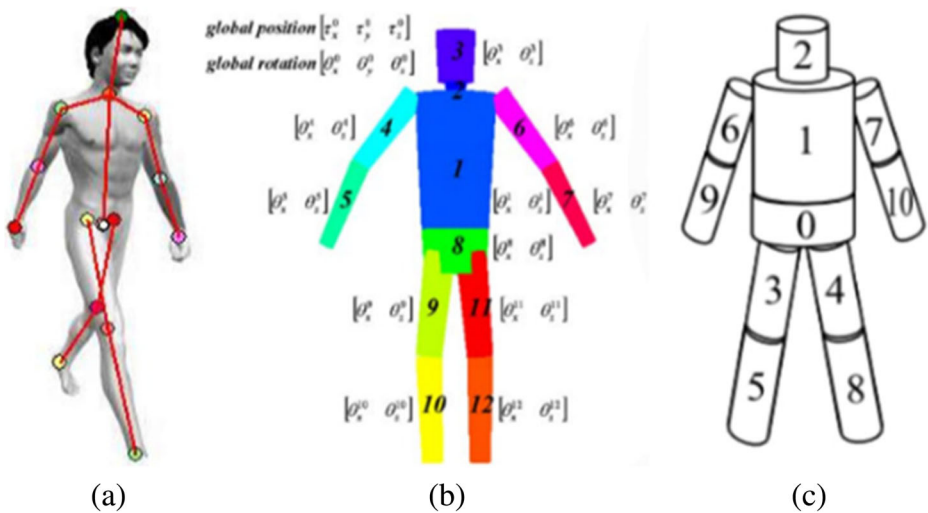
**Fig. 3** Examples of 3D–based methods

compute and evaluate view normalized trajectories of feet and head obtained from monocular videos. They calculated the homography matrix by constructing a linear system using the correspondences between the corners of walking planes. Hu et al. [13] proposed to project the original gait features extracted from any view(s) into a low dimensional feature space, and improve the discriminative ability of multi-view gait features by a unitary linear projection. Makihara et al. [27] proposed a View Transformation Model (VTM) to transform correlated gait information from one view (i.e. source view) onto another view (i.e. target view). And Muramatsu et al. [31] further improved the VTM by incorporating a score normalization framework with quality measures, where the quality measures are calculated from a pair of gait features, and used to calculate the posterior probability that both gait features originate from the same subjects with the biased dissimilarity score. Kusakunniran et al. [23] proposed a novel motion co-clustering to partition the most related parts of gaits from different views into the same group, and described the relationships between gaits from different views based on multiple groups of the motion co-clustering instead of a single correlation descriptor. Kale et al. [17] showed that if the person is far enough from the camera, it is possible to synthesize a side view from any of the other arbitrary views using a single camera. Goffredo et al. [7] used the human silhouette and human body anthropometric proportions to estimate the pose of lower limbs in the image reference system with low computational cost. After a marker-less motion estimation, the trends of the obtained angles were corrected by the viewpoint independent gait reconstruction algorithm, which can reconstruct the pose of limbs in the sagittal plane for identification. Daigo et al. [32] proposed an arbitrary view transformation model (AVTM) for cross-view gait matching. 3D gait volume sequences of training subjects were constructed, and then 2D gait silhouette sequences of the training subjects were generated by projecting the 3D gait volume sequences onto the same views as the target views. Finally, the AVTM was trained with gait features extracted from the 2D sequences. In the latest work, Wu et al. [47] proposed the first deep CNN-based on gait recognition, which took advantage of labeled cross-view pairs (identical or not) and amounts to directly predicting the similarity given a pair of samples in an end-to-end manner. The method showed its outstanding recognition ability in several gait recognition datasets, when the cross-view angle was no less than 36°.

At the same time, Kinect has been used as a popular tool in gait recognition since its release in 2010, because it can capture skeletons with less influence caused by illumination and clothing, and depth with less computation and expense than previous depth sensors. Sivapalan et al. [39] extended the concept of the GEI from 2D to 3D with the depth images captured by Kinect, and averaged the registered 3D volumes to form the gait energy volume (GEV). Araujo et al. [2] calculated the length of the body parts derived from joint points as the static anthropometric features, and used them for gait recognition. Milos et al. [30] used the coordinates of all the joints captured by Kinect to generate a RGB image, combined such RGB images into a video to represent the walking sequence, and identified the gait based on the spirit of Content Based Image Retrieval. Preis et al. [35] selected 11 skeleton features captured by Kinect as the static feature, used the step length and speed as dynamic feature, and integrated both static and dynamic features for recognition. Yang et al. [50] proposed a relative distance-based gait features, which can reserve the periodic characteristics of gait better than anthropometric features. Ahmed et al. [6] generated a gait signature from a sequence of joint relative angles (JRA) over a complete gait cycle, and introduced a dynamic time warping (DTW)-based kernel to measure the distance between two JRA sequences. Kastaniotis et al. [19] proposed a framework for gait-based recognition using Kinect. The captured pose sequences were expressed as Euler angular vectors of eight selected limbs. Then these angular vectors were mapped into the dissimilarity space and transferred into a vector of dissimilarities. Finally, the dissimilarity vectors of pose-sequences are modeled via sparse representation.

In this paper, we adopt the spirits of model-based gait recognition methods, and extract the view-invariant gait features via 2G–Kinect to identify the person from his walking. Our contributions are summarized as follows.

1) We build a Kinect-based gait dataset. There are 52 subjects in this dataset. Each subject has 20 gait sequences, and walks in 6 fixed and 2 arbitrary walking directions. It contains both the 3D coordinates of skeleton joints captured by 2G–Kinect and the corresponding 2D silhouettes. We aim to use it to bridge 2D–based and 3D–based gait recognition methods. The dataset will be introduced in Section 3, and it can be accessed at https://sites.google.com/site/sdugait/.

2) We combine the hand-crafted static and dynamic features in the proposed gait recognition method. The features that are related to the body and barely change during walking are selected as static features, and the ones that can represent the changes along with time during walking are selected as dynamic features. The stability of 8 static and 4 pairs of dynamic features are analyzed, and the stable ones are fused for gait recognition.

3) We compare the proposed method with some existing 2D–based and 3D–based methods from both the joint- and Kinect-based perspectives in our and other Kinect-based datasets. It shows that our method works well in view-invariant gait recognition even when there is no prior information of the view angle.

In the remaining part of this paper, we will introduce our Kinect-based gait dataset in Section 2, and then present the details of our proposed method in Section 3. Experimental results are given in Section 4, and conclusion of this paper is given in Section 5. Comparing with the conference version [45], the proposed method in this paper uses two angle pairs as the dynamic feature instead of only one angle pair, which brings about performance improvement. And the comparison experiments are carried on more dataset and on various view angles to verify the view-invariance performance of the proposed method entirely.

## 2 Gait dataset

Gait dataset is important to the performance improvement and evaluation of gait recognition. There are lots of gait datasets in the current academia. Here we give a brief introduction about several popular gait datasets.

### 2.1 Previous gait dataset

SOTON Large Database [38] is a classical gait database containing 115 subjects, who are observed from side view and oblique view, and walk in different environments, including indoor, treadmill and outdoor.

SOTON Temporal Database [9] contains the largest time variations. The gait sequences are captured monthly during one year with controlled and uncontrolled clothing condition. It is suitable for investigating the time elapse effect on the gait recognition without regarding clothing.

USF HumanID [41] is one of the most frequently used gait datasets. It contains 122 subjects, who walk along an ellipsoidal path outdoor, as well as contains a variety of covariates, including view, surface, shoes, bag, and time elapse. This database is suitable for investigating the influence of each covariate on the gait recognition.

CASIA Gait Database contains three sets, i.e., A, B and C. Set A, also known as NLPR, is composed by 20 subjects, and each subject contains 12 sequences, which includes three walking directions, i.e., 0°, 45° and 90°. Set B [51] contains large view variations from the front view to the rear view with 18° interval. There are 10 sequences for each subject, which are six normal sequences, two sequences with a long coat, and two sequences with a backpack. Set B is suitable for evaluating cross-view gait recognition. Set C contains the infrared gait data of 153 subjects captured by infrared camera at night under 4 walking conditions, which are walking with normal speed, walk fast, walk slow, and walk with carrying backpack.

OU-ISIR LP [14] contains the largest number of subjects, i.e., over 4000, with a wide age range from 1 year old to 94 years old and with an almost balanced gender ratio, although it does not contain any covariate. It is suitable for evaluating gait-based age estimation.

TUM-GAID [11] is the first multi-model gait database, which contains gait audio signals, RGB gait images, and depth body images obtained by Kinect.

KinectREID [33] is a Kinect-based dataset which has 483 video sequences of 71 individuals under different illumination conditions and three view directions (i.e., frontal, rear, and lateral). Its original motivation is person reidentification, and it provides all what Kinect can captured and is convenient for other Kinect SDK based applications.

### 2.2 Our gait dataset

Referring to the overview of the gait dataset in [28], most datasets are based on 2D videos or based on 3D motion data captured by professional camera, such as VICON. To our best knowledge, there are few gait datasets contain both 2D silhouette images and 3D coordinates of joints. Therefore, a novel dataset based on Kinect is built, and it can be accessed at the website, https://sites.google.com/site/sdugait/. Such a dataset can make the joint-based methods. For instance, the method in [7] can directly use the joint coordinates captured by Kinect, which can make use of both advantages of 2D–based and 3D–based methods, and bring improvement to the recognition performance. Meanwhile, the Kinect-based method such as [6, 30, 35, 50] will have a uniform platform to compare with each other. Its characteristics are summarized as follows.

1. The dataset contains the 3D coordinates of 21 joints (excluding 4 finger joints) and the corresponding silhouette images of each subject, as shown in Fig. 4. The dataset makes it possible to compare the 2D–based and 3D–based methods in the same dataset, and provides a bridge between the research of the 2D–based and 3D–based gait recognition.

2. The data acquisition environment is shown in Fig. 5. Two Kinects are located mutually perpendicular at the distance of 2.5 m to form the biggest visual field, i.e., walking area. Considering the angle of view, we put two Kinects at 1 m height on the tripod. The red dash lines are the maximum and minimum deep that Kinect can probe. The area enclosed by the black solid lines is the available walking area. The reason we use two Kinects is that we can one walking from two different views simultaneously because the two Kinects can be considered as same.

3. Each subject has in 6 fixed and 2 arbitrary walking directions as shown in Fig. 6, which is used to investigate the influence of view variation on the performance of gait recognition. Each subject walks several times on the predefined directions shown as the arrows ①-⑤ in Fig. 5. Particularly, ③ is defined as the main direction of Kinect A, and ⑤ means the subjects walk on an arbitrary direction.

4. There are 52 subjects in the dataset, who are 28 males and 24 females with average age of 22. The height of subjects is from 1.5 to 1.9 m. Each subject has 20 walking sequences from two Kinects so that there are totally 1040 gait sequences. Most subjects wear shorts



Fig. 4　3D coordinates of 21 joints (in the upper area) and the corresponding silhouette images (in the lower area) in our dataset

**Fig. 5** The top view of the data acquisition environment
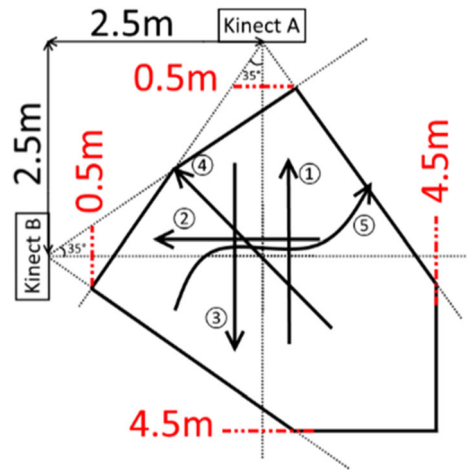


and T-shirts, and few females wear dress and high-heeled shoes. The attributes of each subject, such as name, sex, age, height, wearing (e.g., high-heeled shoes, long skirt, etc.), and so on, are provided for other potential analysis and data mining.

# 3 Proposed method

Unlike the traditional 2D–based methods mostly extract the gait feature from the silhouette captured from the lateral view, the proposed method captured the data of joints from the front view because the Kinect can't capture the accurate joint data of the limb covered by the near one, i.e., from the lateral view. Therefore, we take the front direction as our main walking direction in the proposed method and dataset. Figure 7 demonstrates the frame of the proposed method. The static/dynamic

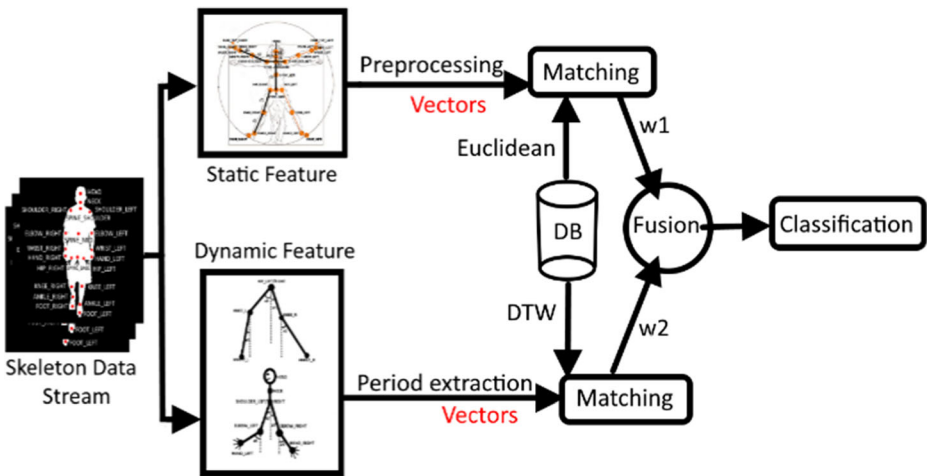**Fig. 6** Walking directions and the corresponding walking times
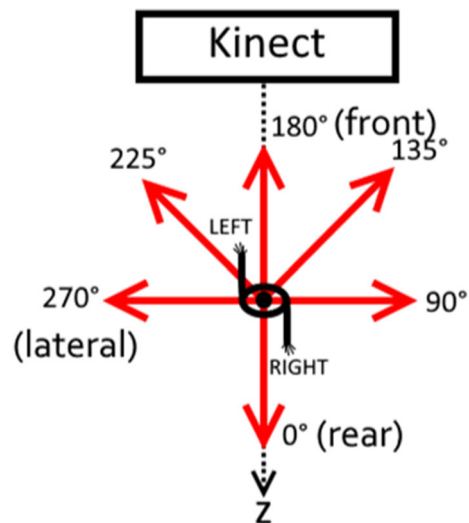
**Fig. 7** The frame of the proposed method

feature is extracted respectively according to the 3D joint position captured by Kinect. After the preprocessing and period extraction, the static/dynamic feature vector is finally obtained. The templates stored in the database are calculated similarity with static/dynamic feature vector respectively in the different way. After the feature fusion, the final result will be acquired from the classifier. In this section, each part of this frame will be introduced with details, and the view invariance of static/dynamic is investigated.

### 3.1 Static gait feature

Static feature is defined as the feature that can barely change during walking. Given the knowledge of anthropometry, the person can be recognized based on static body parameters to some extent [42]. Here we take the length of some skeletons as the static gait features. And the 2G–Kinect can recognize 25 joints of human as shown in Fig. 8, and can get more accurate coordinates of the joints than its first generation.

Considering the symmetry of human body, the length of limbs on both sides are usually treated to be equal. Although the length of skeleton won't change during walking theoretically, the estimation accuracy of Kinect on the joints of the swing limbs will decline, because it is easily affected by the walking style. And the absolute coordinates of the joints are not as stable as the relative distance between the joints. Consequently, we choose the joints which could be monotonous in the deep direction (Z direction) of Kinect to representative the body, and use the distance between them as the static feature. In this paper, the static feature is defined as $F_s = (d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8)$, where $d_i$ is the distance between Joint_1 and Joint_2 listed in Table 1. We can acquire the 3D coordinates of the joints listed in Table 1 in each frame, and the Euclidean distance is chosen to measure the space distance referring to [28] and [1].

$$d_i = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2} \tag{1}$$

where $(x_1, y_1, z_1)$ and $(x_2, y_2, z_2)$ represent 3D coordinates of Joint_1 and Joint_2, respectively.

In addition, the estimation accuracy of Kinect is low at the quite near and far distances, and it is high and stable at the distance from 1.8 m to 3.0 m. Hence, we use the depth of HEAD
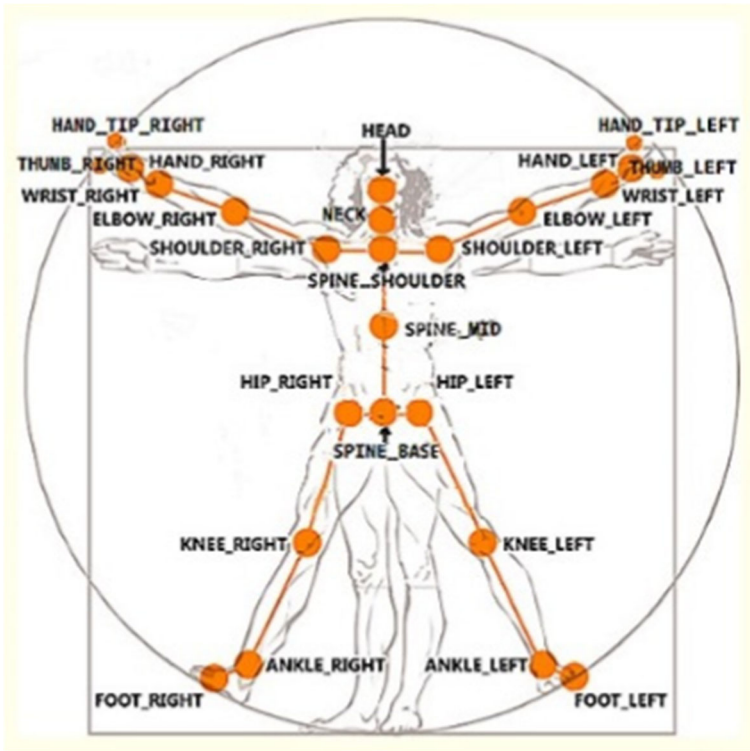
**Fig. 8** The joints recognized by 2G–Kinect

joint as a reference, set the upper and lower bounds in the depth direction to form the stable range, and use only the data captured in the stable range for gait recognition.

$$\{f_s\} = \{f | f_{H.z > 1.8} \cap f_{H.z < 3.0}\} \tag{2}$$

where $f$ represents the all frames. $f_s$ is the frames in the stable range, i.e., the frames with the z axis coordinate of the HEAD joint larger than 1.8 m, $f_{H.\,z > 1.8}$, and smaller than 3.0 m, $f_{H.\,z < 3.0}$. We average data in these stable frames to obtain the final components of the static feature vector. In Fig. 9, we select $d_1$ component as an example to show the comparison on stability between the static features obtained from all frames and from the ones within the stable range.

**Table 1** Components of the static feature vector

| Component | Joint_1 | Joint_2 |
|---|---|---|
| $d_1$ | HIP_RIGHT | KNEE_RIGHT |
| $d_2$ | KNEE_RIGHT | ANKLE_RIGHT |
| $d_3$ | SHOULDER_RIGHT | ELBOW_RIGHT |
| $d_4$ | ELBOW_RIGHT | WRIST_RIGHT |
| $d_5$ | SPINE_SHOULDER | SPINE_BASE |
| $d_6$ | SHOULDER_RIGHT | SHOULDER_LEFT |
| $d_7$ | SPINE_SHOULDER | NECK |
| $d_8$ | NECK | HEAD |
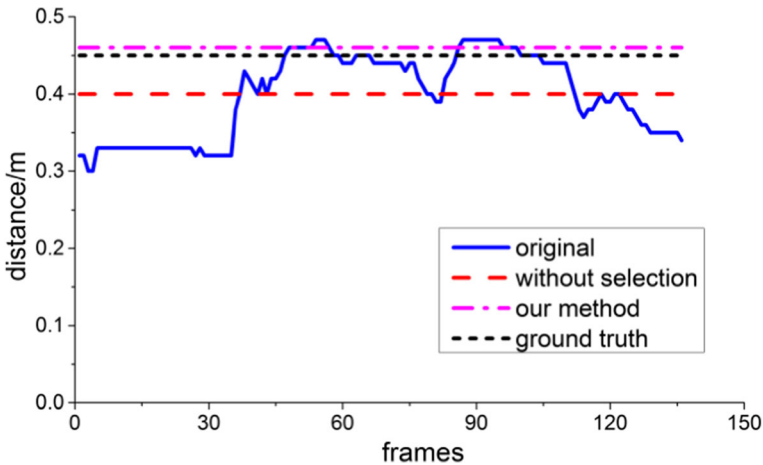
**Fig. 9** The stability comparison between the static features from all frames and only the frames in the stable range

In Fig. 9, the blue solid curve shows the original $d_1$ extracted each frame, the black dot line shows the ground truth of $d_1$, the red dash line shows the mean of all original $d_1$, and the pink dash dot line shows the mean of $d_1$ in the range from 1.8 m to 3.0 m, i.e. our method. It can be seen that our method has much less errors according to the ground truth.

Given the stable static feature, $F_s$, along the walking, we investigate its stability to the variance of view angles. We let the subject walk along the main direction at first, and make the Kinect turn in clockwise and anti-clockwise directions from 0 degree to 15 degree with 5-degree interval respectively. The clockwise and anti-clockwise directions are defined as the positive and negative directions respectively, which are denoted as $p$ and $n$ for short. At each interval, the subject still walks along the previous direction, and then we can get the static gait vectors on $p5$, $p10$, $p15$, $n5$, $n10$, and $n15$ directions. We randomly select 15 subjects and investigate the stability of their static features to the variance of view angles. We show the components of the static feature on the 15 view angles in Fig. 10. We show only one of the subjects in Fig. 10, because the static features of all subjects have the same invariance to view angle. From Fig. 10, we can see that the static feature barely changes while the view changes.

### 3.2 Dynamic gait feature

The dynamic feature is a kind of feature that any change along with time during walking, such as speed, stride, barycenter, etc. Given previous researches in [40, 42, 49], the angles of swing limbs during walking are proved to be remarkable dynamic gait features. For this reason, four groups of swing angles of upper limbs, i.e., arm and forearm, and lower limbs, i.e., thigh and crus, are defined as shown in Fig. 11, and denoted as a1, ..., a8. Here, a1 is taken as the example for demonstration, and the other angles of swing limbs can be in the same way as a1. If the coordinate at HIP_LEFT is $(x, y, z)$, and the coordinate at KNEE_LEFT is $(x_1, y_1, z_1)$, a1 can be calculated as follows.

$$\tan\angle a1 = \left(\frac{x-x1}{y-y1}\right) a1 = tan^{-1}\left(\frac{x-x1}{y-y1}\right) \tag{3}$$
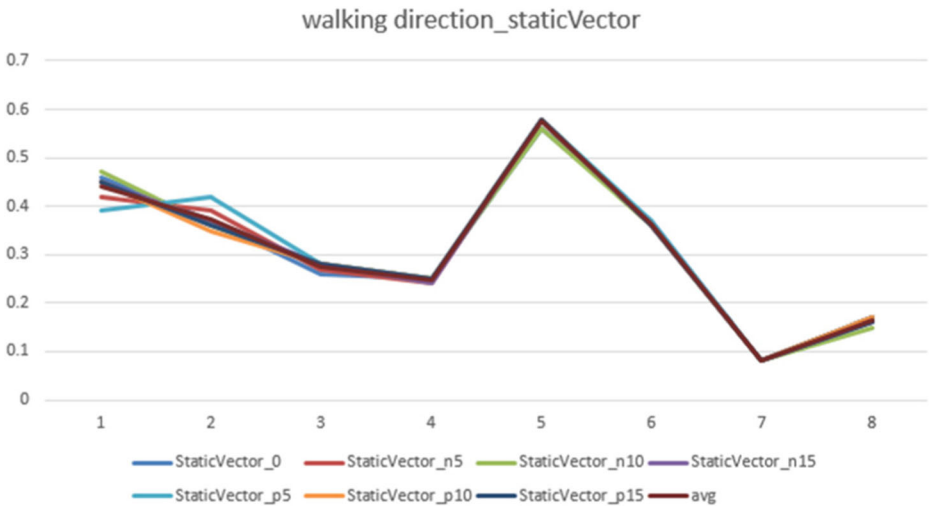
**Fig. 10** The stability of the static features on 7 directions

Each dynamic angle can be regarded as an independent dynamic feature for recognition. Given the symmetry of human body, we divide a1, …, a8 into 4 pairs, which are a1a3 pair, a2a4 pair, a5a7 pair, and a6a8 pair. Figure 12 shows the 4 angle pairs in a walking period.

Like the stability investigation of the static feature, we investigate the stability of the dynamic feature to the variation of view angles in the same way. Here, we take a1 (R_hip_knee) as an example to show the stability of the dynamic feature to the variation of view angles in Fig. 13.

It can be seen from Fig. 13a that there is a temporal translation between views due to the different starting time. After the synchronization, the dynamic feature is stable to the variance of views as shown in Fig. 13b.

### 3.3 Gait period extraction

As we all know, gait is a periodic signal, and gait period extraction is an important step in gait analysis. In existing 2D–based gait recognition methods, the gait period is usually extracted by
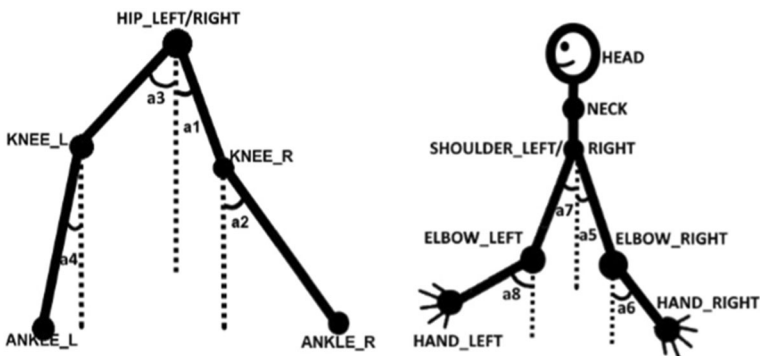


**Fig. 11** The side view of walking model
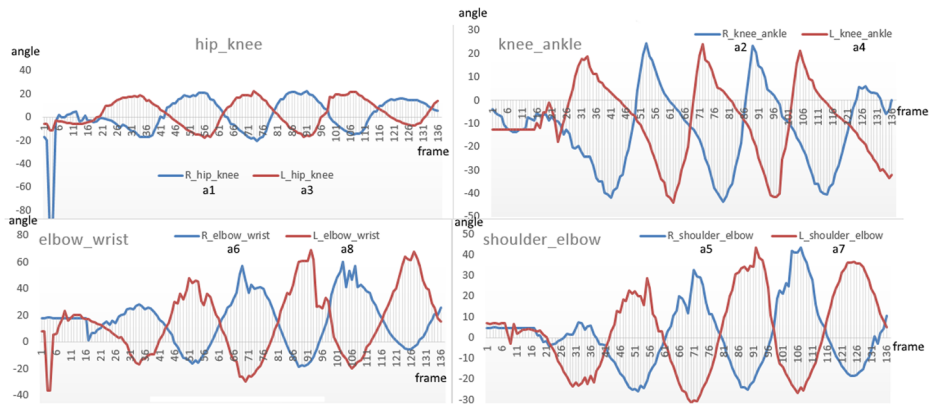
**Fig. 12** Four pairs of dynamic features

peak-valley-detection-based methods, such as peak-valley-detection variation of the number of foreground pixels in [41].

In this paper, we propose a crossing point-based method to extract the gait period by combining the signals of left and right limbs. In this method, we cut off the beginning of the signals of the left and right limbs to preserve their stable part, detect out their crossing points, and take the period between two crossing points as the gait period. Figure 14 shows the results of period extraction on a1a3 and a2a4 pairs, where the black dash lines mark the beginning and end points of the detected-out gait period.

### 3.4 Solution to occlusion

The data of the occluded limbs can't be captured accurately by Kinect, which will result in poor static and dynamic features. Figure 15 shows an example of the subject's walking along with the 90° direction, where the left side of the body is near to the Kinect. It can be observed from Fig. 15 that the point cloud on the left side is distributed densely and the coordinates of the joints are estimated appropriately, while the point cloud is barely distributed on the right side and the coordinates of the joints can't be estimated correctly. To reduce such influence caused by occlusion, we only use the static and dynamic features from the side near to the Kinect for gait recognition.

### 3.5 Dynamic feature matching

We can't use Euclidean distance for the dynamic ones, because even if it is the same subject, there are still some tiny differences in gait periods or styles as shown in Fig. 16, which makes Euclidean distance not work well. However, it can be seen from Fig. 16 that the three dynamic features have high correlation with each other, which is like the case in speech recognition. Therefore, we adopt the Dynamic Time Warping (DTW) method to calculate the distance between dynamic features, and the ability of DTW method has been verified in speech recognition [37]. Suppose that there are two dynamic feature vectors with different dimensions.

$P = (p_1, p_2, \ldots, p_i, \ldots, p_n)$ and $Q = (q_1, q_2, \ldots, q_j, \ldots, q_m)$, the dimensions of P and Q are n and m. Construct a $n \times m$ matrix, and its element $(i, j)$ represents the distance between $p_i$ and $q_j$, which using Euclidean distance. The DTW can be described as searching a path through several elements of this matrix which has the minimum accumulated distance.
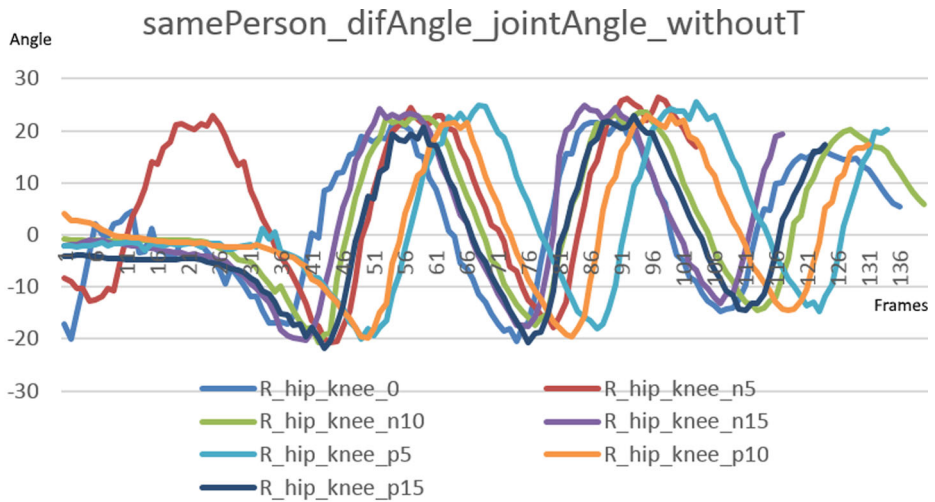
(a) comparison without synchronization
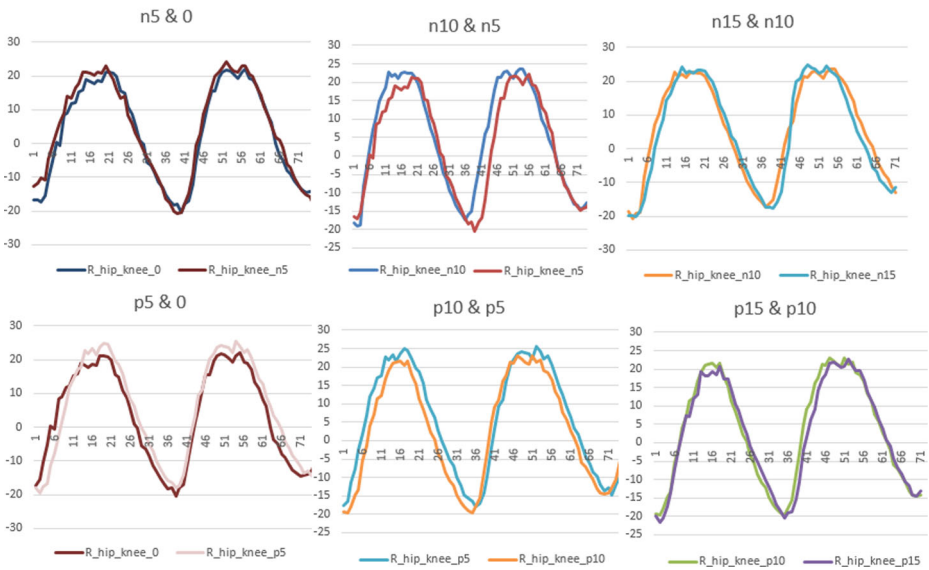


(b) comparison with synchronization

**Fig. 13** Comparison of a1 on 7 view angles (**a**) without synchronization and (**b**) with synchronization

### 3.6 Dynamic feature selection

Referring to [34] and our experiment, the dynamic features from the upper limbs are easily affected by external factors, like clothing, carrying, etc. Hence, we carry out a verification experiment on these swing angles to investigate their recognition ability via Nearest-Neighbor (NN) classification. The result is shown in Fig. 17 using CMC (Cumulative Matching Characteristics) curve, where the horizontal axis is the rank, and the vertical axis is the Correct Classification Rate (CCR)@K. Given Fig. 17, the angle pairs of the lower limbs, i.e., a1a3 and a2a4, perform better in recognition, so we select them as the final dynamic features.
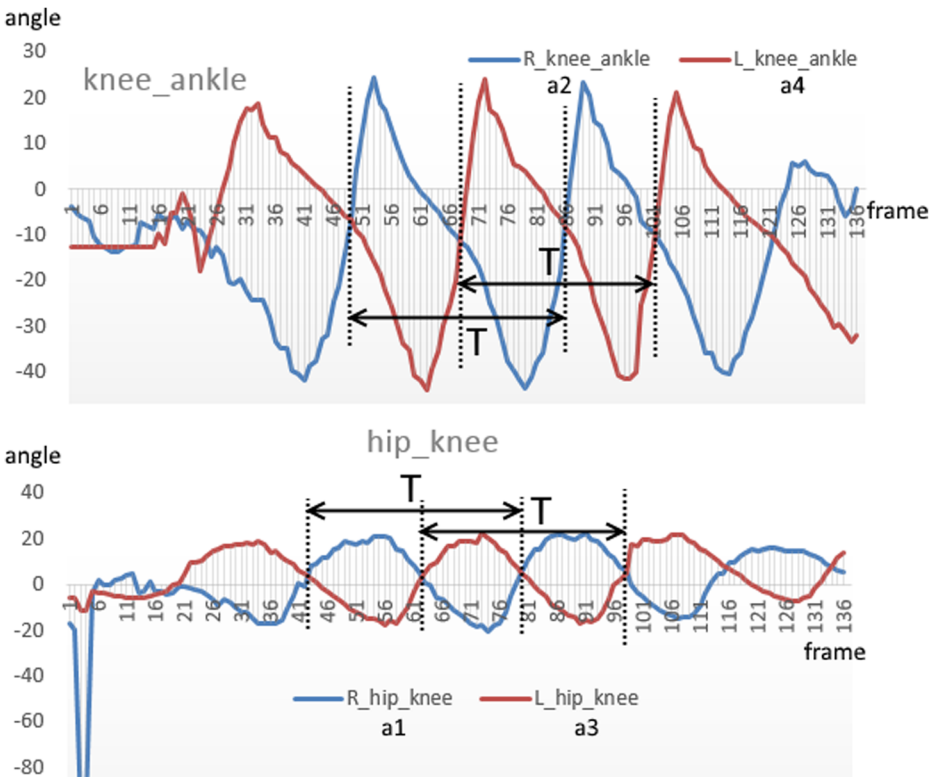
Fig. 14 Crossing point-based period extraction

Furthermore, we find that these two pairs of features are complementary in gait recognition, so we select the nearest neighbor from a1a3 and a2a4 at the same time.

### 3.7 Feature fusion

The static and the dynamic features have their own pros and cons. The static feature is more stable than the dynamic feature, but its discrimination is not as good as that of the dynamic
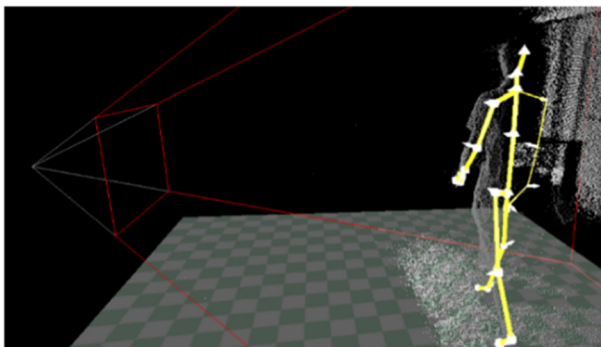


Fig. 15 Demonstration of occluded limbs during the walking in the 90° direction
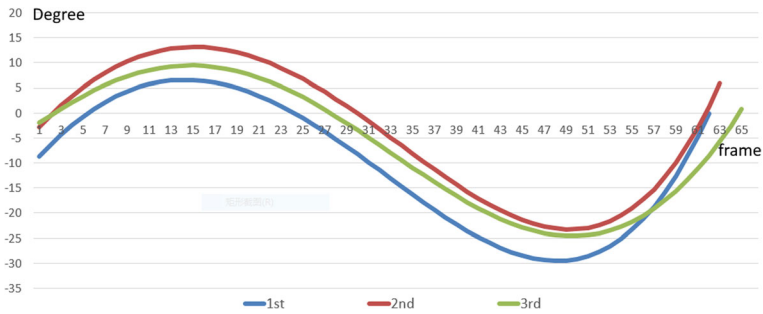
**Fig. 16** The dynamic feature extracted from the walking data of the same subject on three times

feature. Inspired by [36], we make a fusion of the static and dynamic features in the score-level, and recognize gaits based on their fusion. Two different kinds of matching scores are linearly normalized onto the closed interval [0,1].

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

where $\mathbf{S}$ is the matrix before normalization, whose component is s, i.e., the score. $\hat{\mathbf{S}}$ is the normalized matrix, whose component is $\hat{s}$. The fusion of static and dynamic features is weighted as:

$$\mathbf{F} = \sum_{i=1}^{R} \omega_i \, \hat{s}_i, \quad \omega_i = \frac{C_i}{\sum_{j=1}^{R} C_i} \tag{5}$$

where F is the score after fusion, R is the number of features used for fusion, $\omega_i$ is the weight of $i$ th classifier, $\hat{s}_i$ is the score of $i$ th classifier, which is our distance. $C_i$
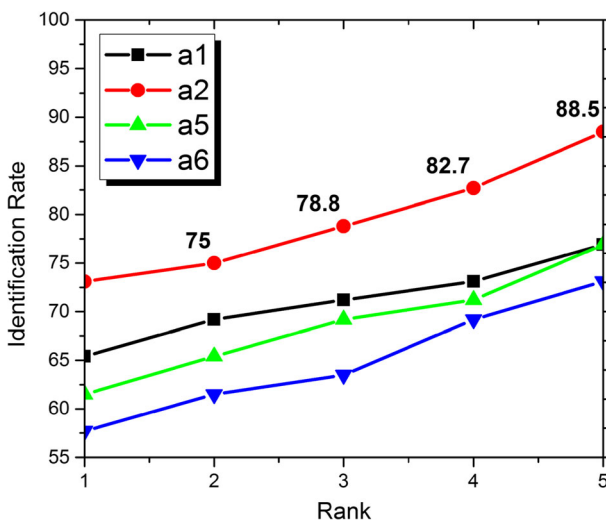


**Fig. 17** The comparison on recognition ability between various dynamic features

is the Correct Classification Rate (CCR) of $i$th feature used for recognition. The weight is set according to CCR.

The comparison between static, dynamic and fusion features is shown in Fig. 18. It demonstrates that the fusion feature has the best performance, and the evaluation at rank 3 has good tradeoff so that we select the performance at rank 3 in the following experiments.

# 4 Experiment result and analysis

The experiment environment is shown in Fig. 5. The data captured by two Kinects are considered as the data captured by one Kinect from different views. There are 52 subjects, and 20 sequences for each subject, and each subject has gait sequences on 7 view angles, i.e., 0°, 90°, 135°, 180°, 225°, 270°, and arbitrary angle (i.e. Arb for short).

## 4.1 Performance of the proposed method

### 4.1.1 Performance on our dataset

We demonstrate the recognition performance of the static, dynamic and fusion features cross view angles in Tables 2, 3, and 4 respectively. We select the data from one view angle as the sample data and the data from the other view angles as the test data. We investigate the best sample views angle for each test view angle, and marked the best CCR corresponding to each testing view angle in bold. From Table 2, we can find that 90° is the best sample view angle for static feature, and 180° is the best sample view angle for dynamic and fusion features. It means that the data on these two view angles are most important during data collection. The shadow part in Table 3 shows the worst recognition performance of the dynamic feature, because the sample and testing data are from two opposite sides of the body and the dynamic feature is not symmetric. And the corresponding performance is improved by fusing the static feature in the fusion feature, which is shown in Table 4.
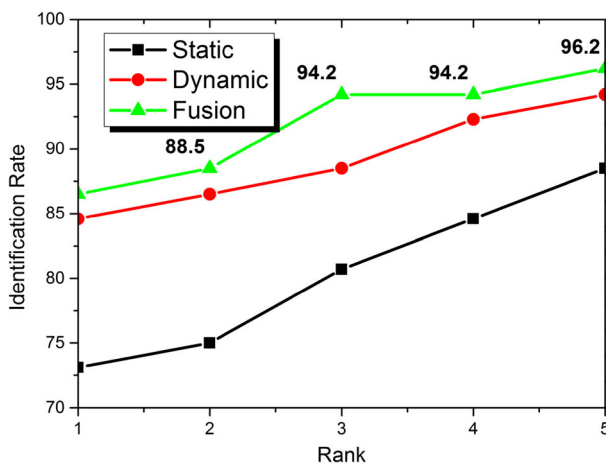


Fig. 18 Performance comparison between different features

**Table 2** CCR of Static Feature @ Rank3 (%), where the best performance is in bold

| Test Sample | 0° | 90° | 135° | 180° | 225° | 270° | Arb |
|---|---|---|---|---|---|---|---|
| 0° | – | 73.1 | 76.9 | 75 | 75 | 75 | 63.5 |
| 90° | **80.7** | – | **78.8** | **78.8** | **76.9** | **80.7** | **67.3** |
| 135° | **80.7** | **78.8** | – | 75 | 75 | 78.8 | **67.3** |
| 180° | **80.7** | **78.8** | 76.9 | – | 75 | 78.8 | 69.2 |
| 225° | 71.2 | 76.9 | 76.9 | 73.1 | – | 71.2 | **67.3** |
| 270° | 78.8 | 76.9 | 76.9 | 76.9 | 75 | – | 65.4 |
| Arb | 65.4 | 67.3 | 67.3 | 65.4 | 67.3 | 63.5 | – |

**Table 3** CCR of Dynamic Feature @ Rank3 (%), where, the best performance is in bold

| Test Sample | 0° | 90° | 135° | 180° | 225° | 270° | Arb |
|---|---|---|---|---|---|---|---|
| 0° | – | 73.1 | 73.1 | 76.9 | 75 | 73.1 | 73.1 |
| 90° | 76.9 | – | 76.9 | 78.8 | 28.8 | 17.3 | 48.1 |
| 135° | 76.9 | 78.8 | – | 78.8 | 48.1 | 17.3 | 48.1 |
| 180° | **78.8** | **82.7** | **78.8** | – | **76.9** | **76.9** | **76.9** |
| 225° | 73.1 | 21.2 | 48.1 | 76.9 | – | 75 | 46.2 |
| 270° | 75 | 21.2 | 42.3 | 76.9 | **76.9** | – | 48.1 |
| Arb | 57.7 | 48.1 | 48.1 | 59.6 | 48.1 | 48.1 | – |

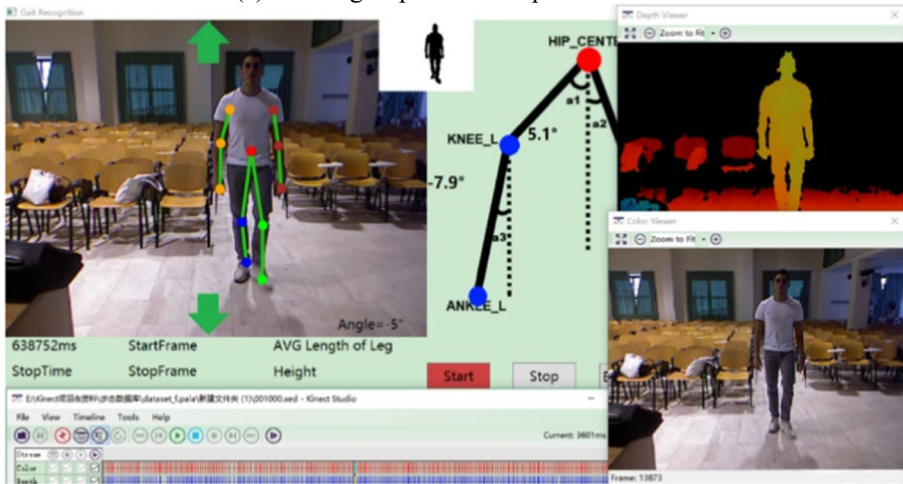### 4.1.2 Performance on KinectREID dataset

We carry out the proposed method on KinectREID dataset because we can extract the proposed gait features from the data of this dataset by using Kinect Studio as shown in Fig. 19. There are three view directions, i.e., frontal, rear, and lateral, in KinectREID dataset. We select 52 subjects out from KinectREID dataset to keep it have the same number of subjects as our dataset. The related data contains 350 walking sequences, which include 150 front, 150 rear and 50 lateral ones. We use data on one of the view directions as the sample data, and the others as the test data. From this point of view, the gait recognition can be taken as person reidentification. The CCR performance at Rank 1 (R1), Rank 3 (R3) and Rank 5 (R5) is listed in Table 5.

**Table 4** CCR of Fusion Feature @ Rank3 (%), where the best performance is in bold

| Test Sample | 0° | 90° | 135° | 180° | 225° | 270° | Arb |
|---|---|---|---|---|---|---|---|
| 0° | – | 78.8 | 82.7 | 83.6 | 80.7 | 78.8 | 76.9 |
| 90° | 82.7 | – | 83.6 | 83.6 | 78.8 | 82.7 | 71.1 |
| 135° | 83.6 | 86.5 | – | 86.5 | 76.9 | 80.7 | 71.1 |
| 180° | **86.5** | **88.5** | **86.5** | – | **83.6** | **86.5** | **82.7** |
| 225° | 78.8 | 78.8 | 78.8 | 82.7 | – | 80.7 | 69.2 |
| 270° | 82.7 | 78.8 | 78.8 | 83.6 | 82.7 | – | 69.2 |
| Arb | 69.2 | 71.1 | 71.1 | 69.2 | 69.2 | 69.2 | – |

(a) Walking sequence examples.



(b)

**Fig. 19** **a** The 7 walking sequence examples of each subject in KinectREID Dataset. **b** Extraction of the proposed gait features from the walking data in KinectREID by replaying the data using Kinect Studio

From Table 5 we can see that the performance of the proposed method on KinectREID dataset is worse than that on our dataset. The reasons are 1) KinectREID is collected by the first-generation Kinect (1G–Kinect), whose joint data is not as stable as that captured by 2G–Kinect. 2) There are obvious hesitancy and halt during some walking, which makes the walking not natural and lowers the recognition performance. Even though the recognition performance of our proposed method is comparable to that of multimodal method published in [11].

## 4.2 Comparisons

### 4.2.1 Comparison with kinect-based method

Preis et al. [30] proposed to use 11 kinds of skeleton lengths acquired by Kinect as the static features, and took the stride and speed as two dynamic features. Their method was tested on a dataset including 9 persons, where its CCR can reach 90% without view variation. We carry out the comparison between out method and theirs on our database as the dataset used in [30] can't be

**Table 5** CCR @ Rank1, Rank3, Rank5 (%) on KinectREID

| Test<br>Train | rear (0°) | | | front (180°) | | | lateral (270°) | | |
|---|---|---|---|---|---|---|---|---|---|
| | R1 | R3 | R5 | R1 | R3 | R5 | R1 | R3 | R5 |
| Rear (0°) | – | – | – | 56 | 60 | 64 | 54 | 58 | 60 |
| Front (180°) | 58 | 62 | 68 | – | – | – | 64 | 66 | 70 |
| Lateral (270°) | 54 | 58 | 60 | 60 | 64 | 70 | – | – | – |

**Table 6** CCR (%) comparison with the method in [30]

|                | Static | Dynamic | Fusion |
|----------------|--------|---------|--------|
| Method in [30] | 80.7%  | 51.9%   | 82.7%  |
| Our method     | 78.8%  | 83.6%   | 92.3%  |



**Fig. 20** Positions of the joints estimated by the method in [45] from 2D silhouette (blue points) and positions projected from the 3D coordinates captured by Kinect (red points)

**Table 7** The deviation between the blue and red points

| Joint point | Hip_left | Hip_right | Knee_left |
|---|---|---|---|
| Deviation | 12.1% | 13.78% | 52.8% |
| Joint point | Knee_right | Ankle_left | Ankle_right |
| Deviation | 23.36% | 48.3% | 8.2% |

accessed publicly. We randomly choose 3 walking sequences on one view angle as the sample data, and the rest are used as testing data. The average comparison results are listed in Table 6, which demonstrates that the proposed method performs much better in dynamic and fusion features.

### 4.2.2 Comparison with joint-based method

The static and dynamic relationship between the joints are important features for gait recognition. In previous 2D based gait recognition schemes, various methods were used to estimate the position of joints from 2D videos. For instance, Goffredo et al. [45] proposed to estimate the positions of joints according to the geometrical characteristics of the silhouette, calculate the angle between the shins and the vertical, and the angle between thigh and the vertical as the dynamic features, and project these features into the sagittal plane according to their viewpoint rectification algorithm. Because our dataset not only has the 3D coordinates of joints, but also has the 2D silhouette at each frame, it is possible to compare it with our proposed method on our dataset directly. The positions of joints estimated by the method in [45] (blue points) and projected from the 3D coordinates captured by Kinect (red points) are shown in Fig. 20. We use the positions of the red points in Fig. 20 as the standard, and the deviation between the blue and red points of all the subjects in our dataset can be calculated by:

$$\delta = \frac{|O_{blue} - O_{red}|}{R} \tag{6}$$

where $\delta$ denotes the deviation between the two positions $O_{blue}$ and $O_{red}$ in Fig. 20. $R$ is the predefined deviation radius, which is 10. The average deviations on the joints are shown in Table 7.

**Table 8** The performance of the method in [45] with two kinds of joint positions

| Test Sample | 0° | 90° | 135° | 180° | 225° | 270° |
|---|---|---|---|---|---|---|
| 0° | – | 63.4 | 59.6 | 55.7 | 63.4 | 61.5 |
|  | – | 67.3 | 69.2 | 71.2 | 65.4 | 69.2 |
| 90° | 57.7 | – | 65.4 | 61.5 | 63.4 | 65.4 |
|  | 71.2 | – | 71.2 | 69.2 | 69.2 | 71.2 |
| 135° | 55.7 | 63.4 | – | 65.4 | 67.3 | 67.3 |
|  | 73.1 | 75 | – | 75 | 71.2 | 75 |
| 180° | 51.9 | 59.6 | 63.4 | – | 59.6 | 61.5 |
|  | 71.2 | 69.2 | 71.2 | – | 71.2 | 73.1 |
| 225° | 69.2 | 65.4 | 69.2 | 61.5 | – | 73.1 |
|  | 69.2 | 69.2 | 73.1 | 71.2 | – | 67.3 |
| 270° | 61.5 | 67.3 | 71.2 | 65.4 | 76.9 | – |
|  | 69.2 | 71.2 | 73.1 | 75 | 71.2 | – |

We carry out the method in [45] with the joint positions estimated from 2D silhouette and those projected from the 3D coordinates captured by Kinect respectively on our dataset and the performance is listed in Table 8.

In Table 8, the first row corresponding to each sample view angle shows the test results with the joint positions estimated from 2D silhouette, and the second row shows the results with the joint positions projected from the 3D coordinates captured by Kinect, which is better than the results in the first row. The results show that our method is much better than the method in [45] with the same joint positions as ours, which means that the effectiveness of the features used in our proposed method. It demonstrates that the proposed gait feature has better performance in recognition. Certainly, the performance of our proposed method is also based on the accuracy of joint estimation, so better joint estimation method will bring better performance.

## 5 Conclusion

In this paper, the 2G Kinect is used as a tool to establish a 3D–skeleton-based gait database, which includes both the 3D position information of skeleton joints collected by the Kinect and the corresponding 2D silhouette images. Given this database, we build up a human walking model and extract the static and dynamic features, which are verified to be view-invariant. Referring to the walking model, a gait recognition scheme based on the matching-level-fusion of these two features is proposed, in which the recognition is achieved by the K-Nearest-Neighbor (KNN) classification method. Experiment results show that the proposed scheme has better performance on cross-view gait recognition. In the future work, we will focus on the adaptive scheme to improve the recognition performance. Furthermore, as the fundamental of our proposed method is the joints, the joints estimated based on the other devices, such as RealSense and web camera [46] will be tested in our proposed method to improve its generalization.

## References

1. Andersson VO, Araujo RM (2015) Person identification using anthropometric and gait data from kinect sensor. In: Proc. AAAI Conference on Artificial Intelligence, 425–431
2. Araujo RM, Graña G, Andersson V (2013) Towards skeleton biometric identification using the microsoft kinect sensor. In Proceedings of the 28th Annual ACM Symposium onApplied Computing (SAC '13). ACM, New York, USA, 21–26. https://doi.org/10.1145/2480362.2480369
3. Bashir K, Xiang T, Gong S (2009) Gait recognition using gait entropy image. 3rd International Conference on Imaging for Crime Detection and Prevention (ICDP 2009), London, 1–6. https://doi.org/10.1049/ic.2009.0230
4. Cappelli R, Ferrara M, Maltoni D Minutia Cylinder-Code: A New Representation and Matching Technique for Fingerprint Recognition. IEEE Trans Pattern Anal Mach Intell 32(12):2128–2141
5. Cunado D, Nixon MS, Carter JN (1997) Using gait as a biometric, via phase-weighted magnitude spectra. Bigun J, Chollet G and Borgefors G (eds.). At Proceedings of 1st Int. Conf. on Audio and Video-Based Biometric Person Authentication, 95–102
6. Faisal A (2015) Polash Paul Padma, and Gavrilova Marina L, Kinect-Based Gait Recognition Using Sequence of the Most Relevant Joint Relative Angles. Journal of WSCG 23(2):147–156

7.  Goffredo M, Bouchrika I, Carter JN, Nixon MS (2010) Self-Calibrating View-Invariant Gait Biometrics. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 40(4):997–1008

8.  Guo Y, Xu G, Saburo T (1994) Understanding Human Motion Patterns. Proc IEEE Int'l Conf Pattern Recognition 2:325–329

9.  Hadid A, Ghahramani M, Bustard J, Nixon MS (2013) Improving Gait Biometrics under Spoofing Attacks, Proc. Image Analysis and Processing 8157:1–10

10. Han J, Bhanu B (2006) Individual Recognition Using Gait Energy Image. IEEE Trans Pattern Anal Mach Intell 28(2):316–322

11. Hofmann M, Geiger J, Bachmann S, Schuller B, Rigoll G (2014) The TUM Gait from Audio, Image and Depth (GAID) Database: Multimodal Recognition of Subjects and Traits. J Vis Commun Image Represent 25:195–206

12. Hossain MA, Makihara Y, Wang J, Yagi Y (2010) Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control. Proc Int'l Conf Pattern Recognition 43(6):2281–2291

13. Hu M, Wang Y, Zhang Z, Little JJ, Huang D (2013) View-Invariant Discriminative Projection for Multi-View Gait-Based Human Identification. IEEE Transactions on Information Forensics and Security 8(12):2034–2045

14. Iwama H, Okumura M, Makihara Y (2012) The OU-ISIR Gait Database Comprising the Large Population Dataset and Performance Evaluation of Gait Recognition. IEEE Transactions on Information Forensics and Security 7(5):1511–1521

15. Jean F (2007) Bergevin R, Albu AB (2007) Computing view-normalized body parts trajectories. In: Proc. Canadian Conference on Computer and Robot Vision, pp. 89-96

16. Johnson AY, Bobick AF (2001) A multi-view method for gait recognition using static body parameters. In: Bigun J, Smeraldi F (eds). Audio and Video-Based Biometric Person Authentication. AVBPA 2001. Lecture notes in computer science, vol 2091. Springer, Berlin, Heidelberg

17. Kale A, Chowdhury AKR, Chellappa R (2003) Towards a view invariant gait recognition algorithm. Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance, 143–150. https://doi.org/10.1109/AVSS.2003.1217914

18. Kale A, Sundaresan A, Rajagopalan AN, Cuntoor NP, Roy-Chowdhury AK, Krüger V, Chellappa R (2004) Identification of Humans Using Gait. IEEE Trans Image Process 13(9):1163–1173

19. Kastaniotis D, Theodorakopoulos I, Theoharatos C, Economou G, Fotopoulos S (2015) A Framework for Gait-Based Recognition Using Kinect. Pattern Recogn Lett 68:327–335

20. Kozlowski LT, James E (1977) Cutting, Recognizing the Sex of a Walker from a Dynamic Point-Light Display. Percept Psychophys 21(6):575–580

21. Krzeszowski T, Michalczuk A, Kwolek B, Switonski A, Josinski H (2013) Gait Recognition Based on Marker-Less 3D Motion Capture. 10th IEEE International Conference on Advanced Video and Signal Based Surveillance, Krakow, 232–237. https://doi.org/10.1109/AVSS.2013.6636645

22. Kusakunniran W, Wu Q, Zhang J, Li H (2012) Gait Recognition under Various Viewing Angles Based on Correlated Motion Regression. IEEE Transactions on Circuits and Systems for Video Technology 22(6):966–980

23. Kusakunniran W, Wu Q, Zhang J, Li H, Liang W (2014) Recognizing Gaits Across Views Through Correlated Motion Co-Clustering. IEEE Transaction on Image Processing 23(2):696–709

24. Lam THW, Lee RST (2006) A New Representation for Human Gait Recognition: Motion Silhouettes Image (MSI). Proc Int'l Conf Advances in Biometrics 3832:612–618

25. Lee S, Liu Y, Collins R (2007) Shape variation-based frieze pattern for robust gait recognition. IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, 1–8. https://doi.org/10.1109/CVPR.2007.383138

26. Liang W, Ning H, Tan T, Hu W (2004) Fusion of Static and Dynamic Body Biometrics for Gait Recognition. IEEE Transactions on Circuits and Systems for Video Technology 14(2):149–158

27. Makihara Y, Sagawa R, Mukaigawa Y, Echigo T, Yagi Y (2006) Gait recognition using a view transformation model in the frequency domain. In: Proc. European Conference on Computer Vision, 151-163

28. Makihara Y, Matovski DS, Nixon MS, Carter JN, and Yagi Y (2015) Gait recognition: databases, representations, and applications. Wiley Encyclopedia of Electrical and Electronics Engineering. 1–15. https://doi.org/10.1002/047134608X.W8261

29. Marcolin F, Vezzetti E (2017) Novel Descriptors for Geometrical 3D Face Analysis. Multimedia Tools and Applications 76(12):13805–13834

30. Milovanovic M, Minovic M, Starcevic D (2013) Walking in Colors: Human Gait Recognition Using Kinect and CBIR. IEEE MultiMedia 20(4):28–36

31. Muramatsu D, Makihara Y, Yagi Y (2006) View Transformation Model Incorporating Quality Measures for Cross-View Gait Recognition. IEEE Transactions on Cybernetics 46(7):1602–1615

32. Muramatsu D, Shiraishi A, Makihara Y, Uddin MZ, Yagi Y (2015) Gait-Based Person Recognition Using Arbitrary View Transformation Model. IEEE Trans Image Process 24(1):140–154

33. Pala F, Satta R, Fumera G, Roli F (2016) Multimodal Person Re-Identification Using RGB-D Cameras. IEEE Transactions on Circuits and Systems for Video Technology 26(4):788–799

34. Pfister A, West AM, Bronner S, Noah JA (2014) Comparative Abilities of Microsoft Kinect and Vicon 3D Motion Capture for Gait Analysis. J Med Eng Technol 38:274–280
35. Preis J, Kessel M, Werner M, Linnhoff-Popien C (2012) Gait recognition with kinect. In: Proc. the First Workshop on Kinect in Pervasive Computing, 1-4
36. Rohr K (1994) Towards Models-Based Recognition of Human Movements in Image Sequences. Computer Vision, Graphics, and Image Processing (CVGIP): Image Understanding 59(1):94–115
37. Sakoe H, Chiba S (1978) Dynamic Programming Algorithm Optimization for Spoken Word Recognition. IEEE Trans Acoust Speech Signal Process 26(1):43–49
38. Shutler J, Grant M, Nixon MS, Carter JN (2002) On A Large Sequence-Based Human Gait Database. Applications and Science in Soft Computing 24:339–346
39. Sivapalan S, Chen D, Denman S, Sridharan S, Fookes C (2011) Gait energy volumes and frontal gait recognition using depth images. International Joint Conference on Biometrics (IJCB), Washington, DC, 1–6. https://doi.org/10.1109/IJCB.2011.6117504
40. Stoia DI, Toth-Tascau M (2011) Influence of treadmill velocity on joint angles of lower limbs during human gait. E-Health and Bioengineering Conference (EHB), Iasi, 1–4
41. Sudeep Sarkar P, Phillips J, Liu Z, Vega IR, Grother P, Bowyer KW (2005) The HumanID Gait Challenge Problem: Data Sets, Performance, and Analysis. IEEE Trans Pattern Anal Mach Intell 27(2):162–177
42. Tanawongsuwan R, Bobick A (2001) Gait Recognition from Time-Normalized Joint-Angle Trajectories in the Walking Plane. Proc IEEE Int'l Conf Computer Vision and Pattern Recognition 2:726
43. Tsuji A, Makihara Y, Yagi Y (2010) Silhouette transformation based on walking speed for gait identification. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, 717–722. https://doi.org/10.1109/CVPR.2010.5540144
44. Wang C, Zhang J, Jian P, Yuan X, Liang W, Image C-G (2010) A Novel Temporal Template for Gait Recognition. European Conference on Computer Vision 6311:257–270
45. Wang Y, Sun J, Li J, Zhao D (2016) Gait recognition based on 3D skeleton joints captured by Kinect. IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, 3151–3155. https://doi.org/10.1109/ICIP.2016.7532940
46. Wei SE, Ramakrishna V, Kanade T, Sheikh Y. (2016) Convolutional Pose Machines, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 4724–4732. https://doi.org/10.1109/CVPR.2016.511
47. Wu Z, Huang Y, Wang L, Wang X, Tan T (2017) A Comprehensive Study on Cross-View Gait Based Human Identification with Deep CNNs. IEEE Transactions on Pattern Analysis & Machine Intelligence 39(2):209–226
48. Yamauchi K, Bhanu B, Saito H (2009) Recognition of Walking Humans in 3D: Initial Results. In: Proc. IEEE Int'l Conf. Computer Vision & Pattern Recognition Workshops, pp. 45–52
49. Yang SXM, Larsen PK, Alkjær T, Simonsen EB, Lynnerup N (2014) Variability and Similarity of Gait as Evaluated by Joint Angles: Implications for Forensic Gait Analysis. J Forensic Sci 59:494–504
50. Yang K, Dou Y, Lv S, Zhang F, Lv Q (2016) Relative distance features for gait recognition with kinect. J Vis Commun Image Represent 39:209–217
51. Yu S, Tan D, Tan T (2006) A Framework for Evaluating the Effect of View Angle, Clothing and Carrying Condition on Gait Recognition. Int'l Conf Pattern Recognition 4:441–444
52. Zhao G, Liu G, Li H, Pietikainen M (2006) 3D gait recognition using multiple cameras. 7th International Conference on Automatic Face and Gesture Recognition (FGR06), Southampton, 529–534. https://doi.org/10.1109/FGR.2006.2

**Jiande Sun** received the Ph.D. degree in communication and information system from Shandong University, Jinan, China, in 2000 and 2005, respectively. From September 2008 to August 2009, he was a Visiting Researcher with the Institute of Telecommunications System, Technical University of Berlin, Berlin, Germany. From October 2010 to December 2012, he was a Post-Doctoral Researcher with the Institute of Digital Media, Peking University, Beijing, China, and with the State Key Laboratory of Digital-Media Technology, Hisense Group, respectively. He is currently a Professor with the School of Information Science and Engineering, Shandong Normal University. He has published more than 60 journal and conference papers. He is the co-author of two books. His current research interests include video content analysis, video hashing, image/video watermarking, 2D to 3D conversion.
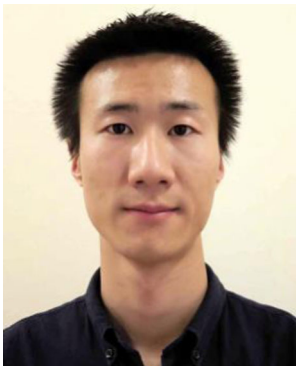


**Yufei Wang** is a master student of School of Information Science and Engineering, Shandong University. His research interests include gait recognition, Kinect-based application, etc.

**Jing Li** is a Ph.D. student in Shandong Normal University, Jinan, China. She is also a lecturer with School of Mechanical and Electrical Engineering, Shandong Management University. Her research interests include machine learning, multimedia processing and retrieval, etc.



**Wenbo Wan** received Ph.D. degree in Shandong Normal University, Jinan, China in 2007 and 2010. Now he is a Ph.D. candidate with the school of information science and engineering, Shandong University. His research interests include image processing and image watermarking.

**De Cheng** received the B.S. degree in automation control from Xi'an Jiaotong University, Xi'an, China, in 2011, and is currently a Ph.D. candidate in the Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University. His research interests include pattern recognition, machine learning and tracking.



**Huaxiang Zhang** received his Ph.D. from Shanghai Jiaotong University, in 2004, and now is a professor and Ph. D supervisor at the Department of Computer Science, Shandong Normal University, Shandong, China. His research interests include machine learning, pattern recognition, evolutionary computation, and web information processing.