Hybrid approach for human posture recognition using anthropometry and BP neural network based on Kinect V2

CrossMark

Bo Li[1], Cheng Han[1*] and Baoxing Bai[1,2]

Abstract

When it comes to the studies concerning human-computer interaction, human posture recognition which was established on the basis of Kinect is widely acknowledged as a vital studying field. However, there exist some drawbacks in respect of the studying methods nowadays, for instance, limitations, insignificantly postures which can be recognized as well as the recognition rate which is relatively low. This study proposed a brand new approach which is hybrid in order to recognize human postures. The approach synthetically used depth data, skeleton data, knowledge of anthropometry, and backpropagation neural network (BPNN). First, the ratio of the height of the human head and that of body posture is ought to be evaluated. The distinguished four types of postures according to the ratio were standing, sitting or kneeling, sitting cross-legged, and other postures. Second, sitting and kneeling were judged according to the 3D spatial relation of the special points. Finally, feature vectors were extracted, transformed, and input to the BPNN according to the characteristics of the other postures, and bending and lying were recognized. Experiments proved the timeliness and robustness of the hybrid approach. The recognition accuracy was high, in which the average value was 99.09%.

Keywords: Kinect v2, Human posture recognition, Depth image, Skeleton data, Image processing, BP neural network, Head localization

1 Introduction

The human posture can show meaningful expression, and so, a relatively more optimal approach in respect of studying the recognition of the postures of people is necessary. The posture recognition of people based on motion-sensing equipment has gradually been focused on the studying field concerning the interaction between people and computer ever since the last few years. Since the Microsoft Company introduced a low-cost motion-sensing device, Kinect, in 2010, many institutions and scholars have conducted research based on Kinect and many occasions of the interaction between people and computers. In addition, the posture recognition of people has already been applied to this somatosensory device, and good effects have been achieved.

Few researchers have utilized depth image for human posture recognition. A previous study [1] obtained human contour by using depth image and Canny edge detector. After distance transformation, as for the calculation in respect of the head position, an approach which is on the basis of a model was applied. The approach realized the detection of people through resorting to a two-dimensional head contour model as well as a three-dimensional head surface mode for the ultimate goal of realizing the detection of the body of people. A plan for segmentation was put forward for the purpose of segmenting people from the environment around him and exacting the comprehensive entire contour image on the basis of the point of detection. The purposes of human detection and tracking were then achieved. Researchers in [2, 3] succeeded at realizing the prediction of the position of the joint of the human body from a three-dimensional perspective on the basis of a single depth image. In addition, they also realized the designation of an intermediate position

* Correspondence: hancheng@cust.edu.cn
[1]School of Computer Science and Technology, Changchun University of Science and Technology, Room B-1603, Keji Dasha, Weixing Road No.7186, Changchun 130022, Jilin, China
Full list of author information is available at the end of the article

representation for the purpose of mapping the pose estimation which is difficult into a relatively easy problem in respect of the per-pixel classification, estimated body parts by varied training datasets, and through resorting to the re-projecting of the results in respect of re-projecting as well as the finding of local patterns, and 3D confidence scores for multiple body joints can be gained. Paper [4] proposed a hybrid recognition method and together with some techniques in respect of the processing of image, depth image created by the sensors of Kinect are adopted for the purpose of identifying five poses of people which are different. The five poses of people are standing, squatting, sitting, bending, and lying. In [5], depth images were captured by Kinect, and the upper limb of human posture and upper limb motion were estimated.

Several researchers have also utilized skeleton data for human posture recognition. A previous study [6] obtained skeleton data from Kinect for seven different experiments. There are four kinds of characteristics extracted from the skeleton of people which were adopted for the purpose of realizing the recognition of the body postures of people. The four characteristics are bending, sitting, standing, and lying. A previous study [7] proposed an algorithm in terms of the detection of the postures of people as well as the recognition of the postures of people which is multi-class on the basis of characteristics of geometry. A serious of characteristics in respect of angle were converted from Kinect's data concerning three-dimensional skeleton. Through resorting to a support vector machine, or SVM in short which was accompanied with polynomial kernel, the classification of the postures of people was realized. As for the previous research [8], there were four different approaches in respect of the classification of the poses of people which were set in comparison, which are (1) support vector machine, or SVM in short; (2) backpropagation neural network, or BPNN in short; (3) naive Bayes; and (4) decision tree. As for the previous research [8], there were four different approaches in respect of the classification of the poses of people which were set in comparison: (1) support vector machine, or SVM in short, (2) backpropagation neural network, or BPNN in short, (3) naive Bayes, and (4) decision tree. The verification of the four approaches mentioned above was realized, verified by using three postures (standing, sitting, and lying). The conclusion was that the accuracy of the BPNN reached 100%, and the average accuracy of the four methods was 93.72%.

Depth image and skeleton data have also been combined by numerous researchers for human posture recognition. A previous study [9] realized the obtaining of the three-dimensional characteristics of the body on the basis of the coordinate information which is

three-dimensional through the adoption of depth image. In addition, the identification of the postures of people which are three-dimensional was realized on the basis of models in respect of skeleton joint of people and multidimensional dataset. A previous study [10] combined marks concerning the anatomy of the human body as well as the model of the skeleton of people, measured the distances of the body parts by using geodesic distance, and realized the estimation of the posture of people through the adoption of depth image from Kinect.

The Microsoft Company introduced Kinect v2 in 2014, and it has demonstrated significant improvements in contrast to Kinect v1, including various aspects as follows: (1) Kinect v2 is two times more accurate than Kinect v1 in the near range, (2) the accuracy of 3D reconstruction and people tracking is significantly improved in different environments with Kinect v2, (3) Kinect v2 presents an increased robustness to artificial illumination and sunlight, (4) the detection range is further than that of Kinect v1, (5) depth image resolution of Kinect v2 is higher than that of Kinect v1, and (6) Kinect v2 can directly output the depth data of the human body [11]. So, we decided to use Kinect v2 for this study. However, the SDK 2.0 for skeleton recognition is far from completely right, despite the fact that the information concerning the skeleton is correct on the condition that the head of people is situated on the highest place of the body of people and no other body parts overlap the head, as shown in Fig. 1a. When the skeleton joint overlaps another, such as in the bending and lying postures, the skeleton information may be incorrect, as shown in Fig. 1b, c. A previous study [12] proposed a repair method for occlusion of the human single joint point, but not only one joint overlapped in many postures. Studies [1, 5] did not utilize skeleton data and only recognized posture by depth data. In a previous study [4], researchers did not adopt the SDK from Kinect for the purpose of recognizing the postures of people. As for the approached of current literature, there exist some drawbacks, for instance, limitations, insignificantly postures of people which can be recognized, as well as the recognition rate which is low. This study proposed a novel hybrid approach, utilized the Kinect SDK 2.0, obtained depth image and skeleton data, and synthetically used several methods and knowledge of anthropometry to solve problems. There are six different postures of people recognized: (1) standing, (2) sitting, (3) sitting cross-legged, (4) kneeling, (5) lying, and (6) bending.

As for the remaining parts of the paper, the organization can be summarized as follows. Section 2 aims at making a brief introduction of the method and contains Sections 2.1, Section 2.2, Section 2.3, Section

Fig. 1 Skeleton data captured by Kinect v2 and its SDK
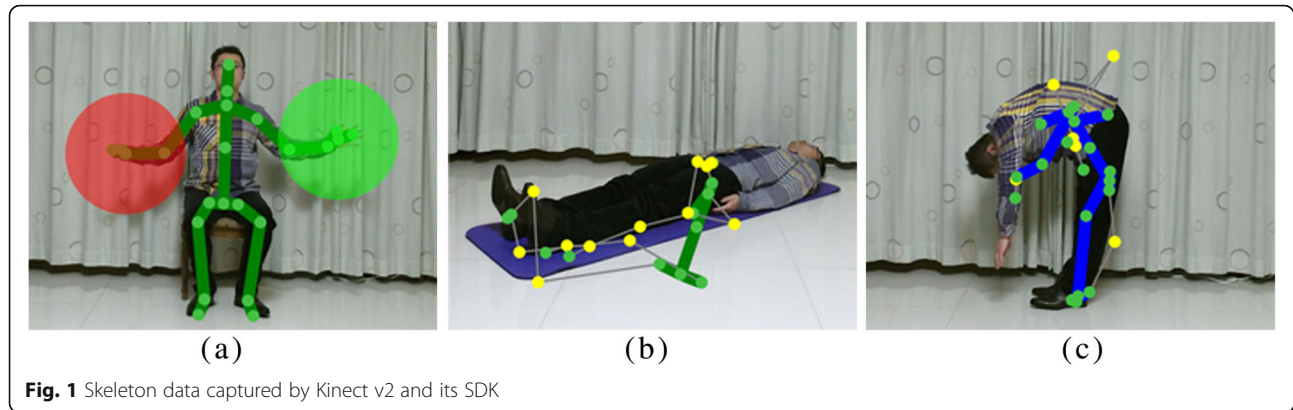
2.4, Section 2.5, and Section 2.6, Section 2.1 shows the entire flow chart of our approach, Section 2.2 presents how to obtain the gravity, the human body height, and contour. Section 2.3 describes the head localization, Section 2.4 explains how to preliminarily distinguish four types of postures according to the ratio of the height of the postures of people and that of the head of people, Section 2.5 describes how to judge sitting and kneeling according to the spatial relation which is three-dimensional in respect to the points which are special, and Section 2.6 applies BPNN to recognize bending and lying. Section 3 elucidates scheme of the experiment, steps of the experiment, experimental content, and the experimental results. In the end, Section 4 concludes the study and suggests further work.

2 Method

2.1 General flow chart of our hybrid approach
As described in the last paragraph of Section 1, the general flow chart in respect of the hybrid approach which is put forward in this study can be seen in Fig. 2.

2.2 Generation of the body center of gravity, contour, and height
Due to the technical limitations of Kinect v2, it has to work in a simple environment. However, by adopting Kinect v2, the identification well as the output in a direct way of depth data in respect of the humanoid area can be realized. In addition, the differentiation between the foreground and background became needless and the feet and the ground in the same area are also distinguished relatively accurately [13]. We utilize Kinect v2 and Kinect SDK 2.0 for Windows to extract the humanoid area, but some noises occur due to the reflection on the ground. Low-pass filter was used to remove noises, but this study does not focus on the low-pass filter. Thus, the low-pass filter is not described further.

After the fixed humanoid area is obtained, the body center of gravity (x_c, y_c) x_c, y_c is calculated by (1).

$$\begin{cases} x_c = \dfrac{1}{S_c} \sum_{i=1}^{S_c} x_i \\ y_c = \dfrac{1}{S_c} \sum_{i=1}^{S_c} y_i \end{cases} \qquad (1)$$

As for the equation set above, (x_c, y_c) can be taken as the coordinate of the gravity center; S_c is the total of white pixels in the human body area; x_i can be viewed as the x-axis of the ith pixel while the $y_i y_i$ represents the y-axis values of that. The human contour is obtained by Canny Operator according to the human body area, as shown in Fig. 3. In this figure, as for the red rectangle which can be seen in (a), it can be taken as gravity center in respect of the body area of people; the human contour is shown in (b), and the red rectangle can be viewed as the gravity center of the contour of people.

Two situations occur. The correct one is the center of gravity in respect of the body area of people. In addition, the incorrect one can be taken as the gravity center outside the body area of people because the feature vectors (will be discussed in Section 2.6) are incorrect. At this incorrect situation, the center of gravity is made as the origin; horizontal and vertical lines are drawn; the line which exactly matches the two intersection points on the contour line of the body is selected. The new gravity center changes into the center position in respect of the two points which are crossing, as shown in Fig. 4.

2.3 Head localization
In this study, postures are preliminarily judged by the head position and the head height. Head localization is thus important. Compared with the other parts of the human body, the head is seldom occluded and is more conducive to be obtained. The correlation of its feature and posture is high. Therefore, the algorithm of head localization is simple with inconsiderable computation. During the postures of people when they stand, sit,
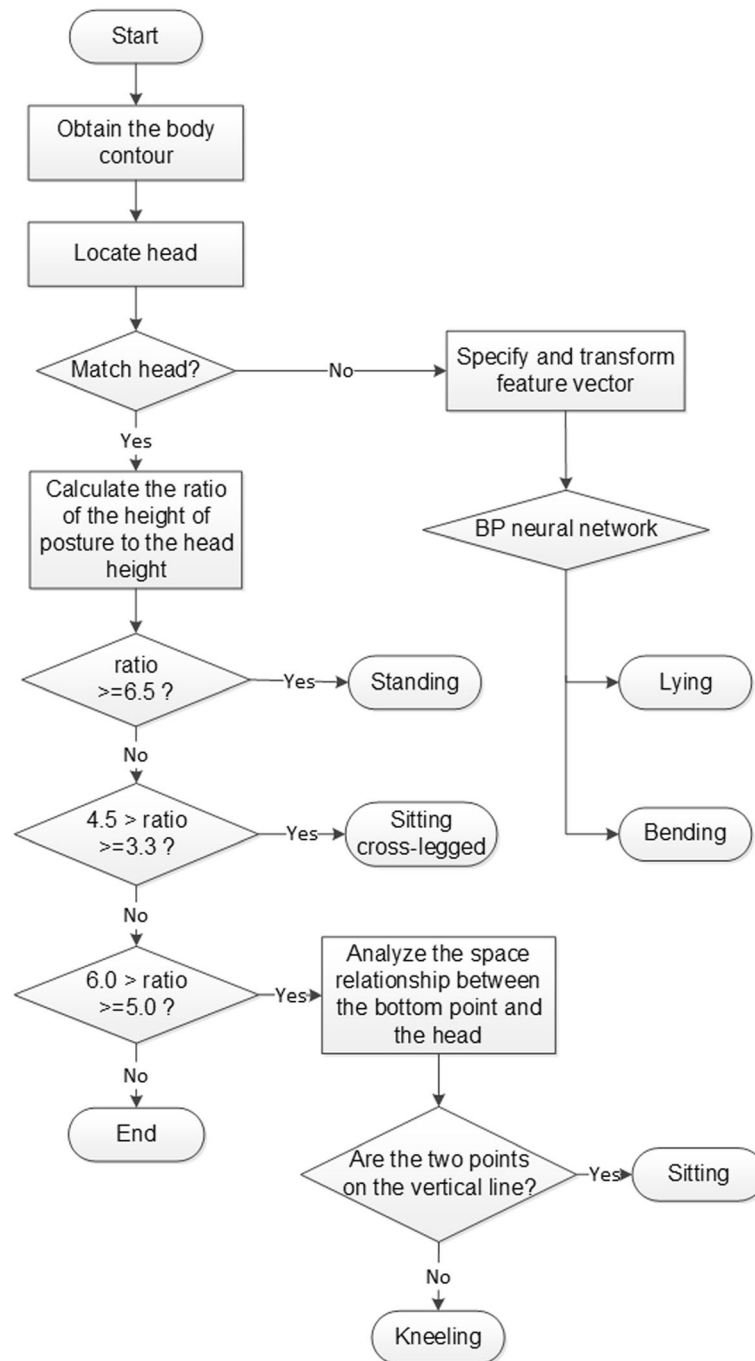
Fig. 2 General flow chart of our hybrid approach

kneel, and sit cross-legged, the head is not occluded by the other parts of the body. Head localization based on skeleton data is accurate. The use of skeleton images from Kinect SDK can hence obtain accurate head information. On the contrary, on condition that the head of people is not situated in the highest place of people, or the other body parts occlude the head, skeleton data may be inaccurate or incorrect, as shown in Fig. 1b, c. Our method utilizes depth image and skeleton data in an integrated manner to position the head. The general process is as follows:

The head coordinate from the skeleton image is judged whether the head is situated in the body area of people from the depth image.
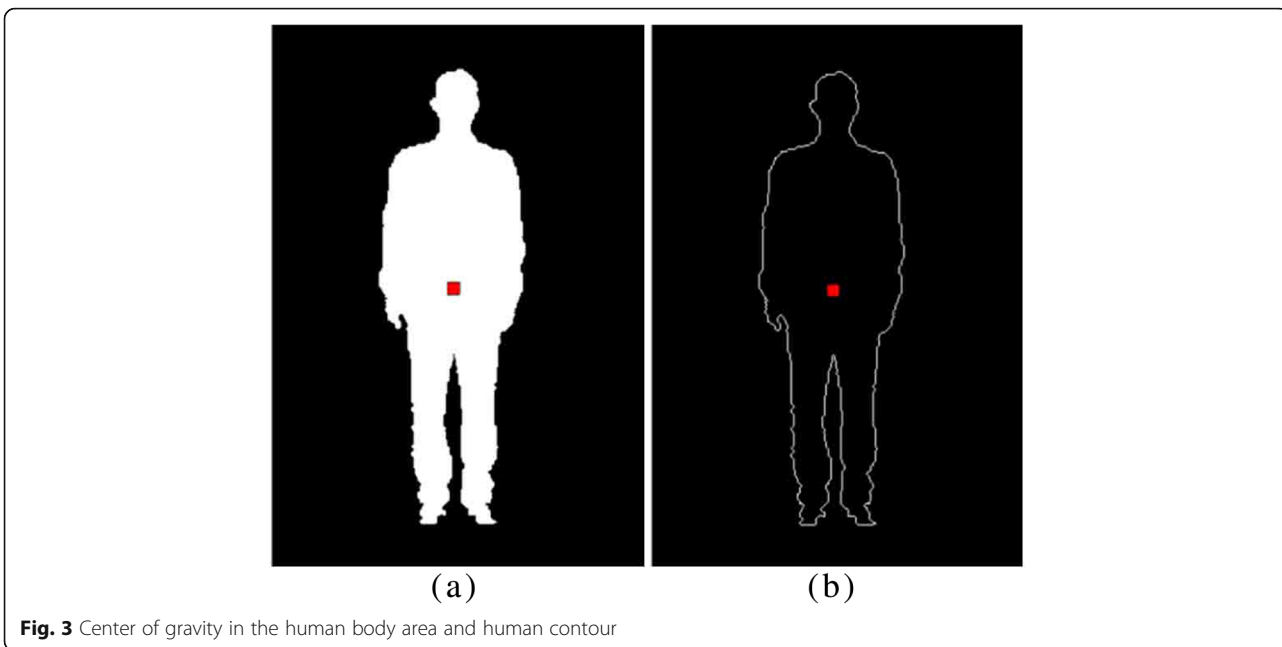
Fig. 3 Center of gravity in the human body area and human contour

1. If the head coordinate is not situated in the body area of people from depth image, the posture is not one of standing, sitting, kneeling, and sitting cross-legged. Then the recognition in respect of the postures of people is realized through resorting to BPNN (see Section 2.6).

2. If the head coordinate is situated in the body area of people from depth image, the head coordinate is credible. Then, the head is positioned according to the positional relation of the coordinates between the head and the neck of people.

a. If the head coordinate is lower than all of the coordinates of the other body parts, or the connection line slope of the head coordinate and the neck coordinate is greater than -1 (or less than 1). Then, posture recognition is processed by BPNN (see Section 2.6).

b. If the head coordinate is higher than all of the coordinates of the other body parts and the connection line slope of the coordinates of the neck and head of people is less than -1 (or greater than 1), then it is confirmed that the posture is one of standing, sitting, kneeling, and sitting cross-legged. The posture is judged by the relation between the heights of contour and the head of people (see Section 2.4) and the spatial relation among those feature points which are three-dimensional (see Section 2.5).
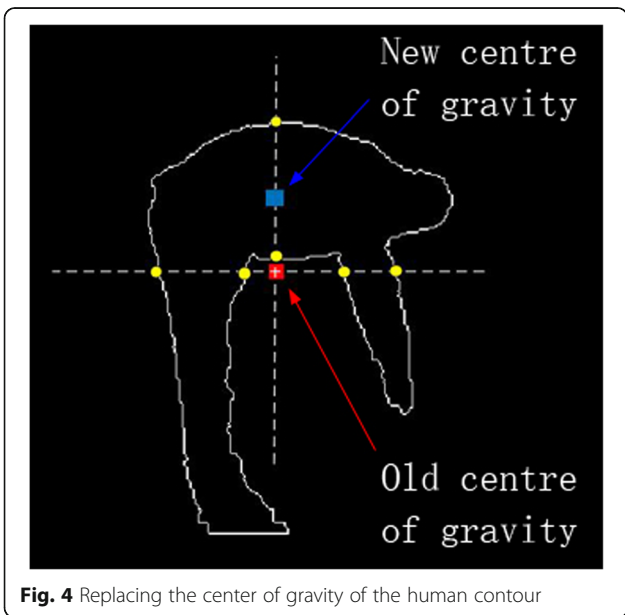
We need the accurate relationship between two data, namely, by resorting to the depth image, the head height of the human contour, and the difference in terms of height between the head and neck node, and the bone image is seen. We set the resolution of the depth and skeleton images as 512×424.

Forty (20 males, 20 females) healthy persons (18 years old to 25 years old) from a university are recruited. Their body sizes when they are facing the Kinect and their profile facing the Kinect are measured. The body sizes of 160 (80 males aged 26 years old to 60 years old and 80 females aged 26 years old to 55 years old) healthy persons from a medical center when they are facing the Kinect are also measured. The head height based on statistical analysis is as follows:

Fig. 4 Replacing the center of gravity of the human contour

$$H_h = 1.8726 L_{hn} \qquad (2)$$

in the depth image, where HhH_h can be taken as the height of the head and $LhnL_{hn}$ can be taken as the difference in terms of the height between the head and neck nodes in the skeleton image, as shown in Fig. 5.

2.4 Estimation of human posture according to head height and the height of the human contour

Ergonomics and anthropometry indicate that all body parts of people and the structure of the human body satisfy a certain proportion of natural features. The adult body size in the Chinese National Standard GB10000-88 [14] and the relevant literature about standardization present that the human body size about height has minimal difference. The difference ratio between the measurements of the human height (data from the China National Institute of Standardization in 2009) and the record of human height (data from the Chinese National Standard [14]) is below 0.864%. The Chinese National Standard distinguishes the human body sizes of different genders and ages according to percentile and divides certain groups of people into seven percentiles (1%, 5%, 10%, 50%, 90%, 95%, and 99%). Any percentile denotes the percentage of the persons whose size is no more than the measured values. For instance, the 50% percentile means that the percentage of the persons whose size is not greater than the measured values is 50%, and the 50% percentile indicates the standard size of the persons of medium size in that age range. Table 1 shows a part of the specific data of the Chinese National Standard.

A study [15] compared the relevant standards and literature of western countries [16–20] and concluded that the Westerners are taller and stronger than the Orientals, with the characteristics of long arms, long legs, and big hands and feet. In terms of anthropometry dimensions, such as the height, surrounded degree, and width size, the Westerners
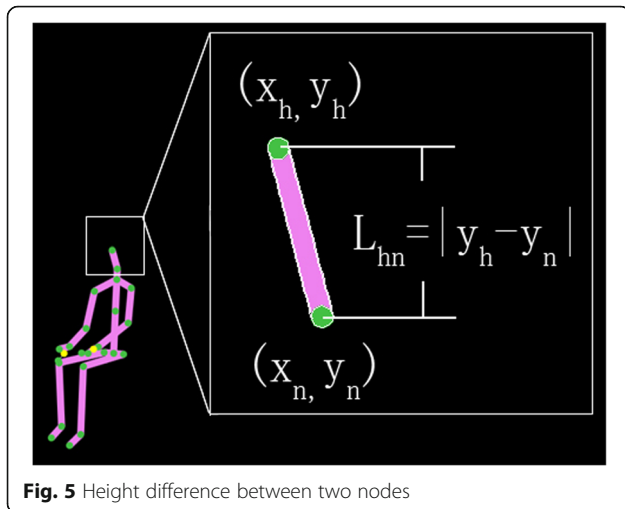

Fig. 5 Height difference between two nodes

are bigger than the Orientals. In terms of head, neck, and the length of the upper body, the difference between the sizes of the Westerners and the Orientals is insignificant.

The ratio of the height in respect of postures of people to that of the head of adult men and women is calculated in every percentile, as shown in Table 2. The ratio of the 99 percentile posture height to the 1 percentile head height is greater than the other percentiles. While the ratio of body height in the 1 percentile to the head height in the 99 percentile is less than the other percentages. We conclude that both of these cases are rare, and the persons are unhealthy. For the stronger suitability of our method, we put the abovementioned two extreme values into the statistical scope. The ratio of the height in respect of postures of people to that of the head for male and female are 7.5516 and 7.3222, respectively; the ratios of the sitting height to the head height for male and female are 5.9638 and 5.7265, respectively; and the ratios of the height of sitting cross-legged to the head height for male and female are 4.0845 and 3.9817, respectively, as shown in Table 2. We can summarize that the difference among the three ratios is relatively large. Thus, we can distinguish three postures (standing, sitting, and sitting cross-legged). However, the thigh length is similar to the lower leg length plus foot height, and the sitting and kneeling heights are difficult to distinguish. Therefore, when $\text{Ratio} \geq 6.5$ $\text{Ratio} \geq 6.5$, the posture refers to the standing one; when $6 > \text{Ratio} \geq 5.0$, the posture refers to the sitting or kneeling one; when $4.5 > \text{Ratio} \geq 3.3$, the posture refers to sitting cross-legged one.

In Table 2, A denotes the average value of all of the ratios of the height in respect of postures of people to that of the head in each percentile (column 2–8 in the corresponding row). H represents the ratio of posture height (99%) to head height (1%). L indicates the ratio of posture height (1%) to head height (99%). P represents the average of all the ratio values in the row (column 2–11). H/H indicates the ratio of the vertical standing height to the head height. S/H stands for the ratio of sitting height to head height. C/H stands for the ratio of sitting cross-legged height to head height.

2.5 Distinguishing of sitting or kneeling by depth data

The depth data are mapped to the Kinect coordinate system by using Kinect SDK. The coordinate of $P_e(x_e, y_e, z_e)$ corresponds to the bottom point $D_e D_e$ in respect of the contour of people, and the coordinate of $P_h(x_h, y_h, z_h)$ corresponds to the central point $D_h D_h$ in respect of the head of people. The sitting or kneeling posture is distinguished according to the revelation of the two coordinates. Head can be approximately considered the shape of a sphere, and its radius is set as

$$R_h = \frac{1}{2} \times \frac{1}{3}(H_h + L_h + W_h) \qquad (3)$$

Table 1 Chinese adult body size (mm)

	Male, 18 years old to 60 years old						
	1%	5%	10%	50%	90%	95%	99%
Height	1543	1583	1604	1678	1754	1775	1814
Thigh length	413	428	436	465	496	505	523
Sitting height	836	858	870	908	947	958	979
Lower leg length + foot height	372	383	389	413	439	448	463
Head height	199	206	210	223	237	241	249
Head maximum width	141	145	146	154	162	164	168
Head maximum length	168	173	175	184	192	195	200
	Female, 18 years old to 55 years old						
	1%	5%	10%	50%	90%	95%	99%
Height	1449	1484	1503	1570	1640	1659	1697
Thigh length	387	402	410	438	467	476	494
Sitting height	789	809	819	855	891	901	920
Lower leg length + foot height	331	342	350	382	399	405	417
Head height	193	200	203	216	228	232	239
Head maximum width	137	141	143	149	156	158	162
Head maximum length	161	165	167	176	184	187	191

where H_h can be viewed as the height in respect of the head, L_h can be taken as the length of the head of people, and W_h refers to the width of the head of people, as shown in Fig. 6.

As shown in Fig. 7a, P_e can be viewed as the point in three-dimensional space corresponding to the bottom point in respect of the body area of people which can be seen in the two-dimensional image. $P'_h(x_h, y_e, z_h)$ refers to the projected point in terms of $P_h P_h$ which is in X–Z plane, and the distance between $P_e P_h$ and $P'_h P_{h'}$ is calculated by the following:

$$L_{eh}' = \sqrt{(x_e - x_h)^2 + (z_e - z_h)^2} \qquad (5)$$

If L_{eh}' exceeds a threshold, then P_e can be confirmed as the foot. The sitting posture may be not standard. Thus, the threshold is set as one head height H_h,

namely, when $L_{eh}' > H_h$ $L_{eh'} > H_h$, the foot is outside of the head projection. However, the posture in which the foot is outside the projection may be the posture of sitting or kneeling, which can be seen in Fig. 7b, c. There are two types of judgment way are put forward as follows:

1. A circle is drawn, and the center of the circle is situated at the upper right $P'_h P_{h'}$. The distance between the circle center and P'_h is H_h, and the radius is H_h. On condition that there is no body parts of people found in the circle, then the posture can be judged as the sitting one; on condition that there exist some body parts of people found in the circle, then the posture can be judged as the kneeling one.
2. A circle is drawn, and the center of the circle is situated at the upper right P_e. The distance between

Table 2 Ratio of the height in respect of postures of people to that of the head

| Male, 18 years old to 60 years old											
	1%	5%	10%	50%	90%	95%	99%	A	H	L	P
H/H	7.7538	7.6844	7.6381	7.5247	7.4009	7.3651	7.2851	7.5217	9.1156	6.1968	7.5516
S/H	6.0704	6.0243	5.9952	5.9238	5.8481	5.8340	5.7912	5.9267	7.2462	4.8514	5.9538
C/H	4.2010	4.1650	4.1429	4.0717	3.9959	3.9751	3.9317	4.0690	4.9196	3.3574	4.0845
Female, 18 years old to 55 years old											
	1%	5%	10%	50%	90%	95%	99%	A	H	L	P
H/H	7.5078	7.4200	7.4039	7.2685	7.1930	7.1509	7.1004	7.2921	8.7927	6.0628	7.3222
S/H	5.8031	5.7550	5.7586	5.7269	5.6579	5.6293	5.5941	5.7036	6.9275	4.6862	5.7265
C/H	4.0881	4.0450	4.0345	3.9583	3.9079	3.8836	3.8494	3.9667	4.7668	3.3013	3.9817
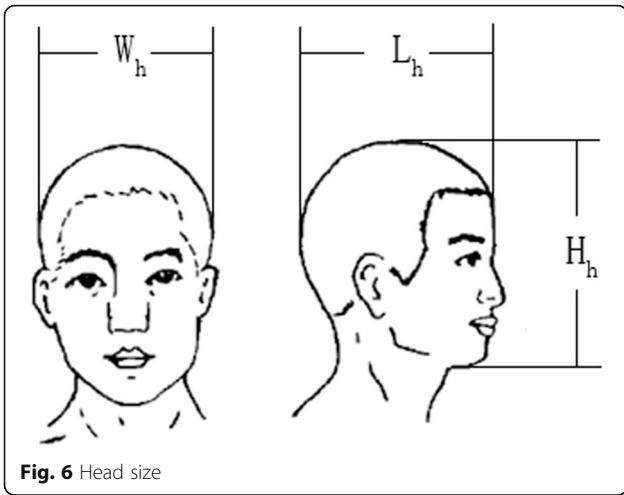
Fig. 6 Head size

the circle center and P_e is H_h, and the radius is also H_h. On condition that there is no body parts of people found in the circle, then the posture can be judged as the kneeling one; on condition that there exist some body parts of people found in the circle, then the posture can be judged as the sitting one.

Figure 7b, c depict that one of the abovementioned two judgment ways can confirm the sitting or kneeling posture. The point (i, j) in the 2D image corresponding to P_{he} (or $P_{hh}{}'$) in the 3D environment can be calculated. Several points (x, j) in the human body area are then utilized to confirm whether a human body part exists in the circle.

2.6 Recognition of bending and lying postures by using BPNN
2.6.1 BPNN
BPNN refers to a feed forward network which is of multi-layers established on the basis of Error Back

Propagation (or BP in short) algorithm. As for the neural network, it utilizes the difference between the actual and desired outputs to correct the connection rights of network layers and correct the layer by layer from back to forward. BPNN has great advantages in solving nonlinear problems or nonlinear structural problems: it can make use of the input and output variables of the neural network for training network, to achieve non-linear calibration purposes. A single sample has m input nodes, n output nodes, and hidden nodes in one or more hidden layers. A considerable amount of training time is required by many hidden layers. According to Kolmogorov's theory, a three-layer BP network can approach any kind of network under a reasonable structure as well as proper weights. As for any continuous function, it can be approximated by a BP network which has three layers [21, 22]. These three layers refer to the input layer, hidden layer as well as the output layer. As a result, the BP network which is endowed with three layers that has a relatively simple structure was selected by us, as shown in Figure 8.

2.6.2 Extraction of feature vectors
The gravity center of the body area of people is given information. The distance between the contour in respect of the body of people and the gravity center can be calculated as follows:

$$d_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2} \qquad (6)$$

As for the equation above, d_i refers to the distance value between the center of gravity (x_c, y_c) and any point on the contour (x_i, y_i). The start is from the left-most pixel, and the movement is clockwise to the end pixel, as shown in Fig. 9.
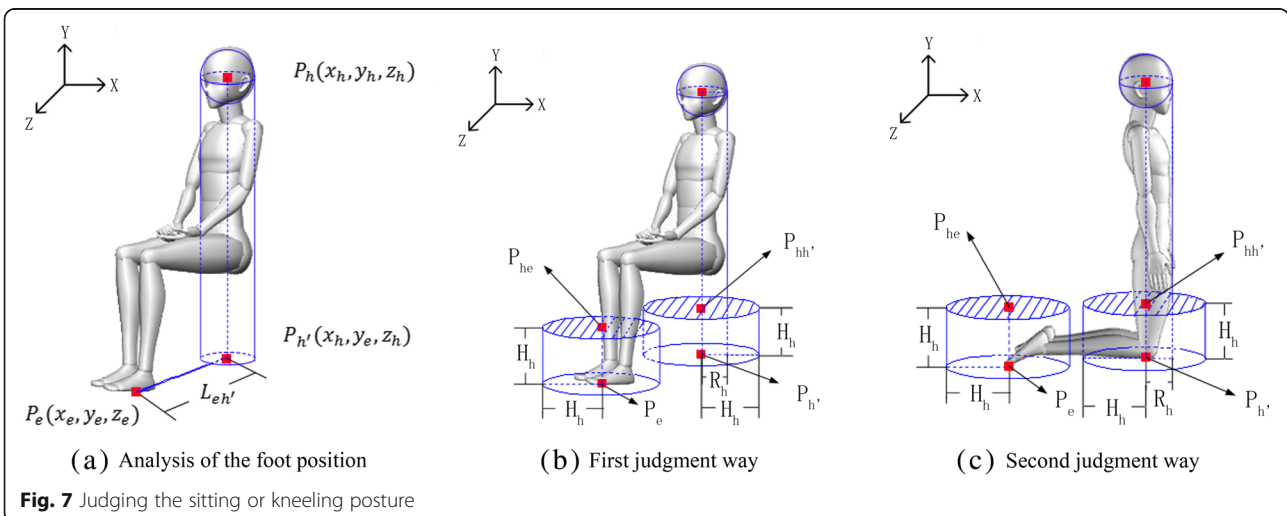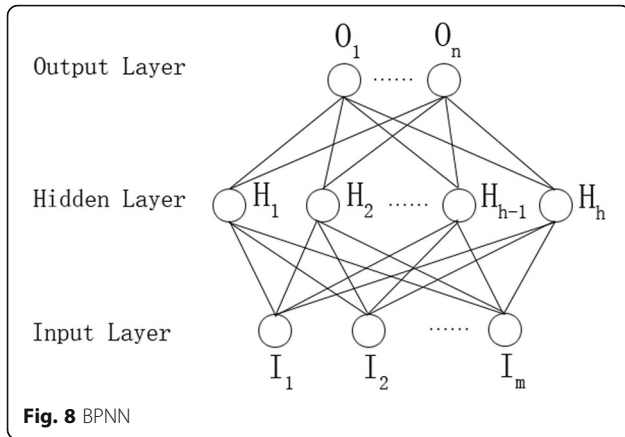

(a) Analysis of the foot position   (b) First judgment way   (c) Second judgment way

Fig. 7 Judging the sitting or kneeling posture

Fig. 8 BPNN

The distance between the pixel and the center of gravity is calculated. A curve in respect of the distance value is then gained and filtered by a low-pass filter, which can be seen in Fig. 10. The peak points in the curve correspond to the smaller red points in Fig. 9. The smaller red points are the feature point, and the feature vectors are the lines from the center of gravity to the feature points, as shown in Fig. 13.

2.6.3 Standardization of feature vectors

Inspired by [4], we obtained the angle information of feature vector according to the characteristics of bending and lying postures to distinguish them. Hence, the feature vector in the Cartesian coordinate system is transformed into the feature vector in the polar coordinate system.

$$R_i = \sqrt{x_{pi}{}^2 + y_{pi}{}^2} \tag{7}$$

$$\theta_i = \tan^{-1}\frac{y_{pi}}{x_{pi}} = \begin{cases} \theta_i, \text{when } \theta_i \text{ is postive} \\ \theta_i + 360, \text{when } \theta_i \text{ is negative} \end{cases} \tag{8}$$

As for the equations above, $i = 1, 2, \ldots mi = 1, 2, \ldots m$. m refer to the number in respect of feature vectors, and the maximum

Fig. 9 Calculation of the distance value $d_i d_i$
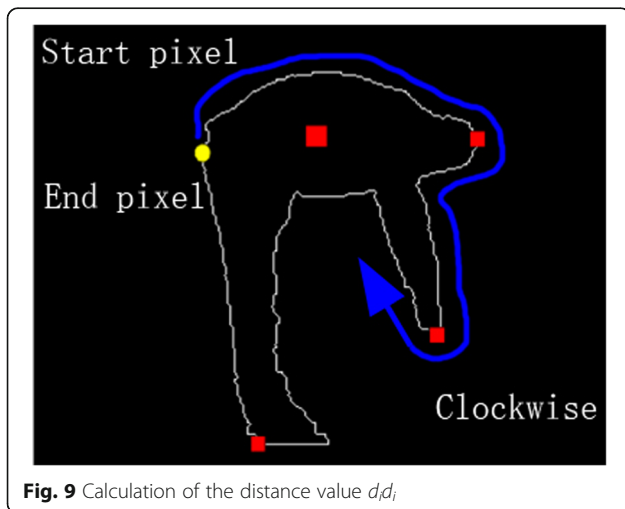
value of m is 4, (x_{pi}, y_{pi}) is the coordinate of the feature point in the Cartesian coordinate system, as shown in Fig. 11.

It is necessary for us to specify the order in respect of the feature vectors for the purpose of distinguishing them better. Therefore, a disk with four regions is set, as shown in Fig. 12. The feature vector of the lying or bending must be in regions 1, 2, and 3 and region 4. Thus, regions 1 and 2 and regions 3 and 4 are symmetrical regions. The order of the feature vectors in these four regions can be specified as follows:

1. On condition that the angle of the feature vector is closer to 180° when set in comparison with other feature vectors in area 1, the feature vector can be called as $V_1 V_1$. At this time, we set $L_1 = R_1 L_1 = R_1$ and $T_1 = \theta_1 T_1 = \theta_1$. Otherwise, we set $L_1 = 0$ and $T_1 = 0$.
1. If an angle of the feature vector is closer to 0°/360° when set in comparison with other feature vectors in region 2, then this feature vector is called V_2. At this time, we set $L_2 = R_2 L_1 = R_1$ and $T_2 = \theta_2 T_1 = \theta_1$. Otherwise, we set $L_2 = 0$ and $T_2 = 0$.
2. If an angle of the feature vector more verges on 315° when set in comparison with other feature in region 3, then this feature vector is called V_3. At this time, we set $L_3 = R_3 L_1 = R_1$ and $T_3 = \theta_3 T_1 = \theta_1$. Otherwise, we set $L_3 = 0$ and $T_3 = 0$.
3. If an angle of the feature vector more verges on 225° when set in comparison with other feature in region 4, then this feature vector is able to be called as V_4. At this time, we set $L_4 = R_4 L_1 = R_1$ and $T_4 = \theta_4 T_1 = \theta_1$. Otherwise, we set $L_4 = 0$ and $T_4 = 0$. The specifying process in respect of the feature vector sequence come to an end.

The specified order of the feature vectors can be seen in Fig. 13.

The difference in the distance between the Kinect and the body of people and the difference in the human height causes a difference in the height of the human body area, which may cause errors in posture recognition. Therefore, the feature values should be normalized. We set $\overline{R}_i = L_i / R_{\max} \overline{R}_i = L_i / R_{\max}$, and $\overline{\theta}_i = T_i / 360 \overline{\theta}_i = T_i / 360$, where $R_{\max} = \max(R_i), i = 1, 2, \ldots m. R_{\max} = \max_i (R_i), i = 1, 2, \ldots m$. In this case, the feature values are ratio values regardless of the height regarding the body are of people. The posture recognition approach is able to then be adopted in respect of various persons with different heights. There are eight feature values showed in the Table 3; the eight feature vectors also function as the input neurons regarding the BPNN.

2.6.4 Training of the BPNN

In the BPNN, it is easy to make clear of the number in respect of the input layer neurons as well as the output

Fig. 10 Curve of the distance value $d_i d_i$
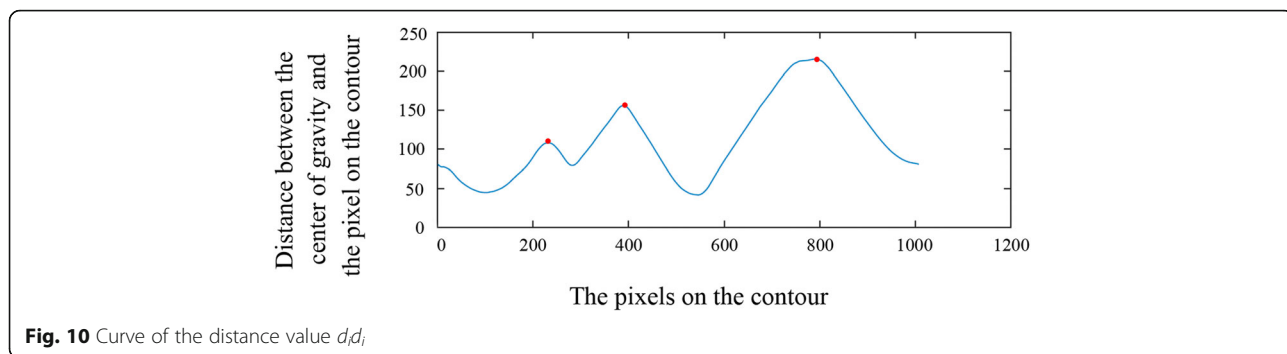
layer neurons, but that of the hidden layer neurons has an impact on the performance of the BPNN. As shown in Fig. 14, eight types of postural samples are input into the BPNN in this study, and each sample has eight feature vectors. The BPNN outputs two judged results of the posture, that is, there are eight input layer neurons as well as two output layer neurons, but the exact number in respect of the hidden layer neurons needs to be determined. Three empirical formulas can be used. The first formula is

$$h = \sqrt{m + n} + a \qquad (9)$$

where h refers to there are h hidden layer neurons, m refers to there are m input layer neurons, n refers to there are n output layer neurons, and a can be taken as constant that is applied to adjust h. The scope of a is from 1 to 10.

The second formula is

$$h = \sqrt{m \times n} \qquad (10)$$

where h refers to there are h hidden layer neurons, m

refers to there are m input layer neurons, and n refers to there are n output layer neurons.

A previous study [23] provided the formula for determining the upper bound of how many hidden layer neurons there exist. The third formula is as follows:
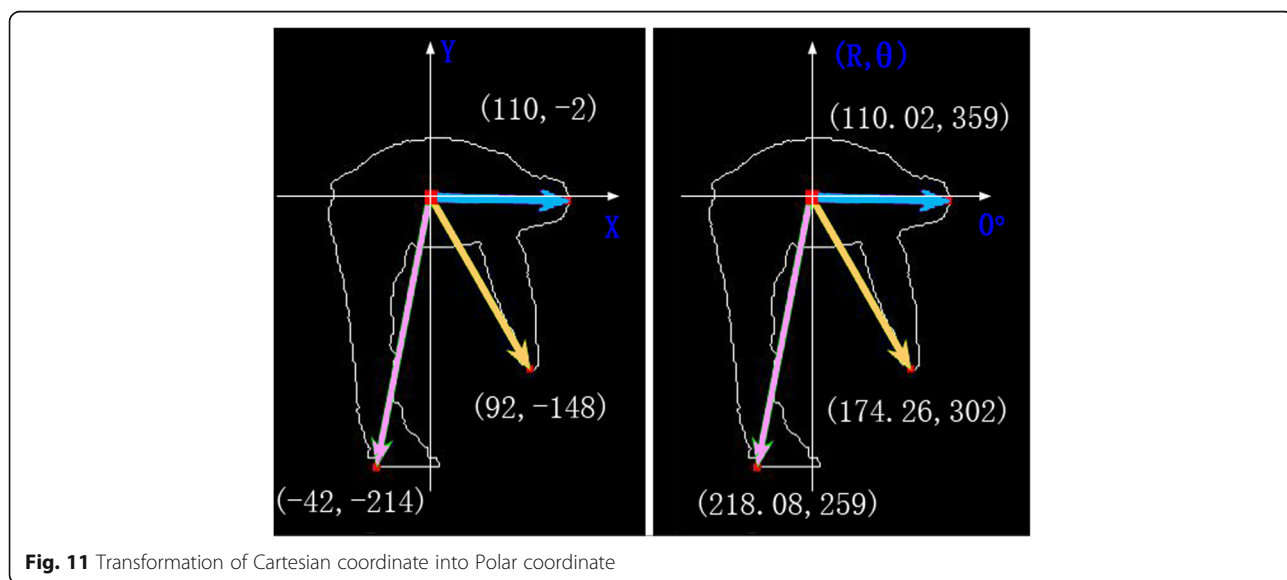
$$N_{\text{hid}} \leq \frac{N_{\text{train}}}{R \times (N_{\text{in}} + N_{\text{out}})} \qquad (11)$$

where N_{in} refers to there are N_{in} hidden layer neurons, N_{out} refers to there are N_{out} input layer neurons, and N_{train} refers to there are N_{train} training samples, $5 \leq R \leq 10$.

The three empirical formulas confirm there exist five hidden layer neurons after many experiments are conducted, as shown in Fig. 8. m, h, and n are confirmed as 8, 5, and 2, respectively.

Eight types of postural samples are adopted for training, as shown in Fig. 14.

Feature vectors of bending differ from those of lying significantly. Therefore, in this study, BPNN was applied in a simple way with inputting several types of samples into the BPNN, and each type of samples was arranged with the same number approximately. It is
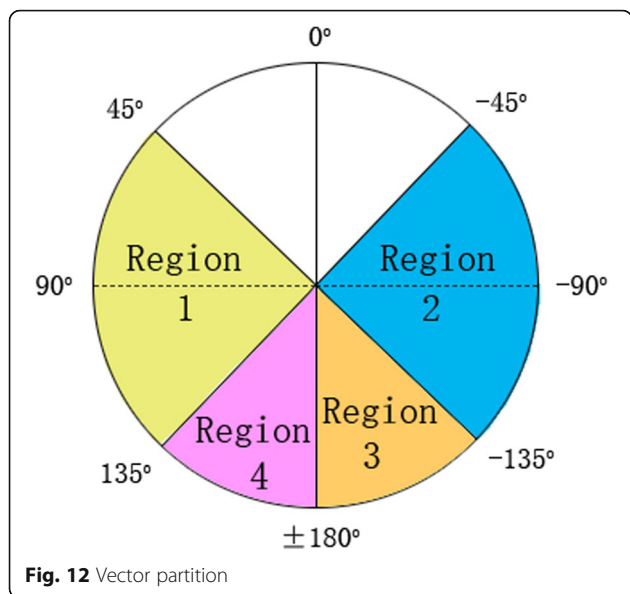

Fig. 11 Transformation of Cartesian coordinate into Polar coordinate

Fig. 12 Vector partition

Table 3 Final feature values

N_1	N_2	N_3	N_4	N_5	N_6	N_7	N_8
$\overline{R}_1$	$\overline{\theta}_1$	$\overline{R}_2$	$\overline{\theta}_2$	$\overline{R}_3$	$\overline{\theta}_3$	$\overline{R}_4$	$\overline{\theta}_4$

64-bit Windows 10 Enterprise Edition; IDE environment: Visual Studio.net 2015 and Kinect SDK 2.0.

Forty-six adults with healthy bodies (25 males and 21 females) participate in the experiments. Among them, 10 males and 10 females are not only the models in terms of the training samples in respect of BPNN but also the participants of all experiments. All the participants have different heights (the highest height is 1.88 m, and the shortest height is 1.57 m) and weights (the maximum weight is 84 kg, and the minimum weight is 45 kg). The distances of the participant and the Kinect v2 are between 2.0 and 4.2 m, and the distance of the Kinect v2 and the ground is 70 or 95 cm.

Before running the entire program, the training samples of bending and lying postures need to be collected. The samples are used to train the BPNN. There are ten male and women respectively from the participants are chosen, and they own various height as well as weight. In addition, they were required to pose two postures which were bending, lying oriented to the Kinect in four directions ($\pm 45°$ and $\pm 90°$). The feature vectors of each posture as well as eight kinds of input samples were taken as the data of training are obtained, as shown in Fig. 14. Each participant is examined five times in every direction. A total of 800 samples are thus collected. All of the samples are input to the BPNN for training (see Section 2.6), and the set in respect of the results after training was conserved later.

As for the entire participants, they are classified into three sets of samples. Set 1 refers to the model set that undergo training samples of the BPNN, set 2 refers to
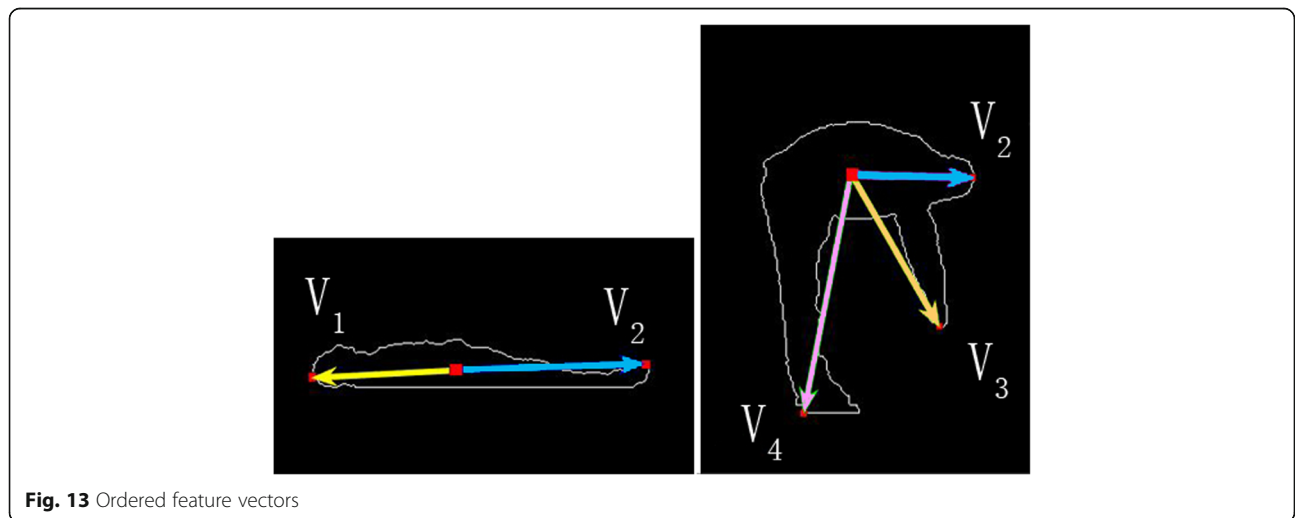
considered that this arrangement will not result in over-training problem. Although the selected samples were representative, and almost no sample noise, early stopping technique [24] was adopted in this study to avoid over-fitting problem. In this study, the accuracy of the neural-network-training target is determined to be 0.001, and as for the learning step size, it is determined to be 0.01. Figure 15 presents that as for the optimal training performance, it is realized at epoch 67.

3 Discussion and experimental results

The hardware environment in respect of the study in this paper is as follows: computer with Intel Xeon(R) CPU E5-2650 and 32G memory, Nvidia Quadro K5000 GPU, Kinect v2 for Windows; System environment:
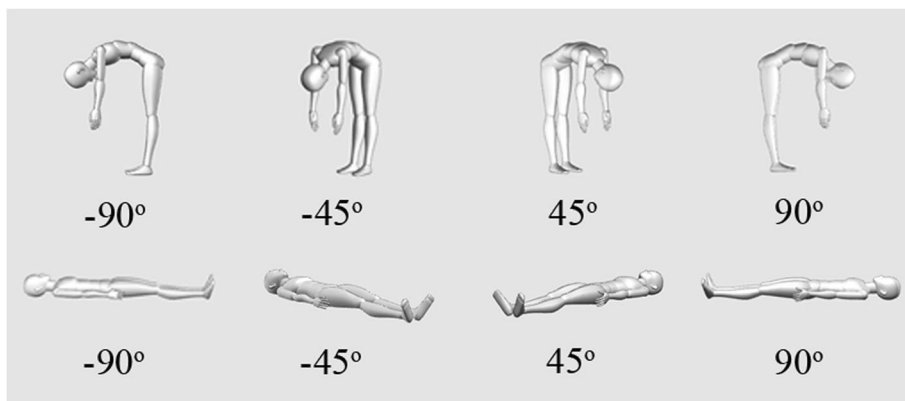

Fig. 13 Ordered feature vectors

Fig. 14 Eight types of postural samples for training

the participants set who do not undergo samples from the perspective of a sample of BPNN, and set 3 refers to the entire participants set. There are six postures of people recognized, and the six postures are (1) standing, (2) sitting, (3) kneeling, (4) sitting cross-legged, (5) bending, and (6) lying. In addition, the six postures of each participant are made in three different experimental environments (indoor with light, indoor with natural light, and outdoor with natural light). Each participant poses six postures several times when he/she orients to the Kinect v2 in various directions. In addition, he/she is then recognized by the program. The directions of standing and sitting cross-legged postures are $0°\sim\pm180°$, and they are posed 30 and 24 times, respectively; the directions of sitting and kneeling postures are $0°\sim\pm90°$, and they are posed 28 and 20 times, respectively; The directions of lying and bending postures are $\pm45°\sim\pm90°$, and they are posed 24 and 20 times, respectively. Figure 16 shows the real scene of human posture. Table 4 shows the experimental statistics for the successful times of posture recognition and recognition accuracy.

The running in respect of our program comprises five stages, stage A is the obtainment of human body area and contour; stage B is the head localization; stage C is the estimation of the ratio of human contour height to head height and the distinguishing of standing, sitting, or kneeling, sitting cross-legged, and other postures; stage D is the judgment of sitting and kneeling according to the relation among the feature points which are three-dimensional spatial; and stage E is the distinguishing bending from lying based on BPNN.

Table 5 deals with the running time in respect of each stage of each posture. The running time of each stage is short on average, that is, the running time of each posture is shorter or equals to 15 ms, which can satisfy the real-time requirement.

A previous study [4] judged kneeling posture by the human contour and the ratio of the width of the upper body to that of the lower body. The accuracy was 99.69%, but the experiment had a precondition that the person must orient some fixed direction ($\pm90°$, $\pm45°$) to the Kinect. The lower body width larger than the upper body width, if the person is facing the Kinect, might cause misjudgment. A previous study [8] utilized four methods to recognize human postures. The best one was the method based on BPNN. However, although its recognition accuracy was 100%, the method could recognize only three postures (standing, sitting, and lying). Its feature values were extracted by skeleton data. Consequently, the method could not extract effective feature values for bending and kneeling postures. Similar to the sitting posture in [4], that in [8] required the person to sit with a fixed posture. The sitting postures of men and women are different. Women often sit elegantly, i.e., sit with legs closed together and sit with the ankle on the knee. Our approach first judges kneeling and sitting postures according to the ratio of human contour height to head height, and then, it distinguishes the two postures according to the relation among the feature points which are three-dimensional spatial. This approach can judge
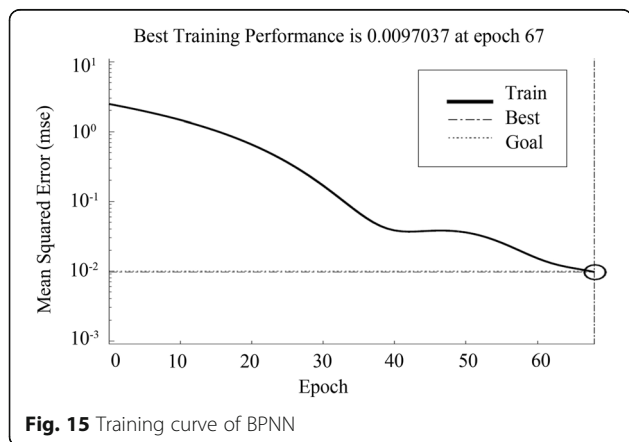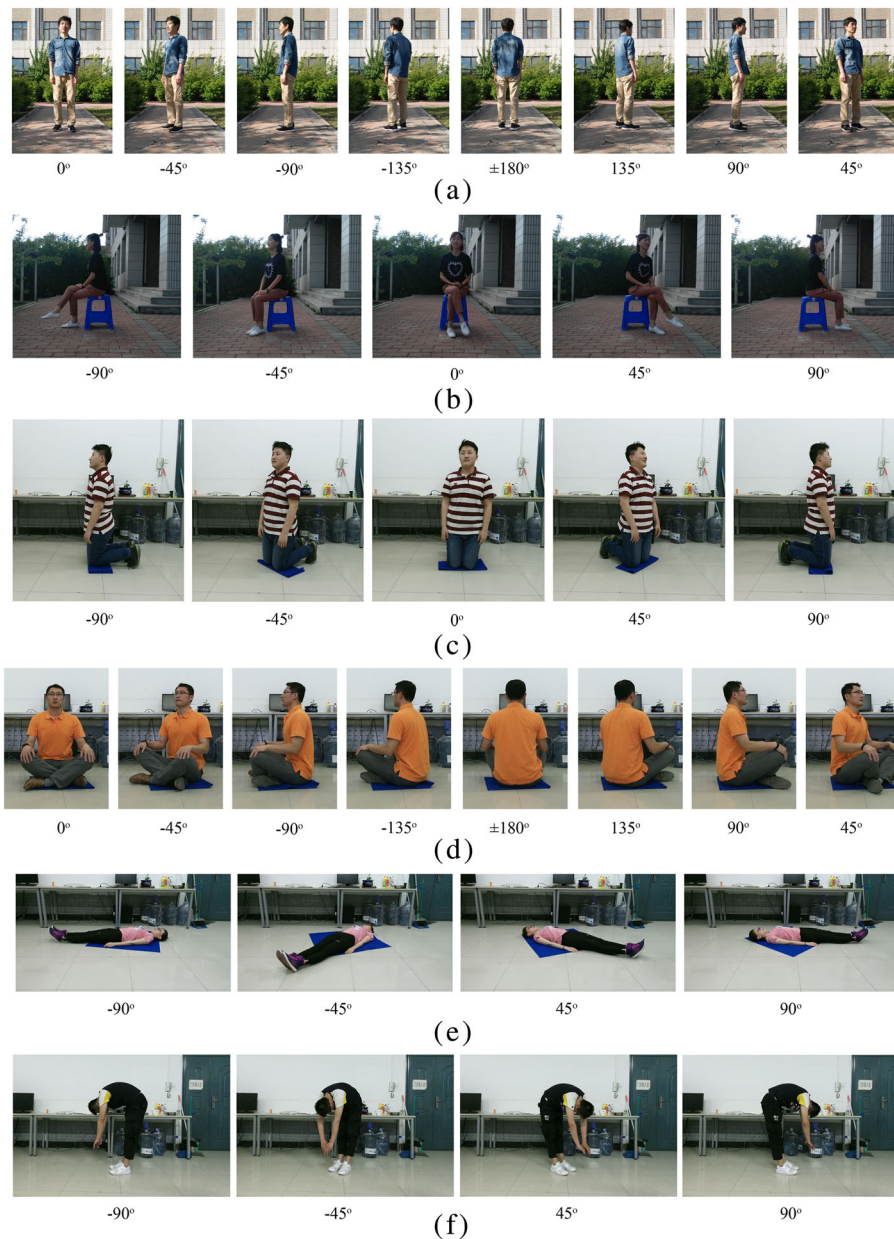

Fig. 15 Training curve of BPNN

Fig. 16 Real depiction of human posture

Table 4 Accuracy of human posture recognition

	A	B	C	D	E	F	Average
Set 1	600/600	544/560	391/400	480/480	480/480	400/400	2895/2920
Acc1	100%	97.14%	97.75%	100%	100%	100%	99.14%
Set 2	780/780	719/728	505/520	624/624	621/624	511/520	3760/3796
Acc 2	100%	98.76%	97.12%	100%	99.52%	98.27%	99.05%
Set 3	1380/1380	1263/1288	896/920	1104/1104	1101/1104	911/920	6655/6716
Acc 3	100%	98.06%	97.39%	100%	99.73%	99.02%	99.09%

A denotes standing, B denotes sitting; C denotes kneeling; D denotes sitting cross-legged; E denotes lying; and F denotes bending; the average represents the average precision of six postures which are made in one experiment; Acc_i (i = 1, 2, 3) represents the accuracy of one pose in the Set_i group

Table 5 Running time of stages of each posture (unit: ms)

	A	B	C	D	E	F
Stage A	10.218	9.151	8.936	9.178	11.106	9.216
Stage B	0.028	0.032	0.032	0.028	0.022	0.011
Stage C	0.299	0.309	0.328	0.309	0	0
Stage D	0	0.045	0.047	0	0	0
Stage E	0	0	0	0	3.812	3.207
Total	10.545	9.537	9.343	9.515	14.94	12.434

A denotes standing, B denotes sitting, C denotes kneeling, D denotes sitting cross-legged, E denotes lying, and F denotes bending

more kneeling postures on condition that people are in various directions of the Kinect (from − 90° to 90°), including the person facing the Kinect, and even elegant sitting posture of women.

The researchers of [7, 25] proposed methods for the purpose of realizing the recognition of ten postures. The three postures in [7] (which are (1) standing, (2) leaning forward, (3) sitting on a chair) and the three postures in [25] (which are (1) standing, (2) sitting, (3) picking up) bear some similarities with standing, sitting, and bending adopted in the study of nowadays, respectively. The method used in [7] was based on the characteristics of angle in respect of the postures as well as polynomial kernel SVM classifier. In addition, the recognition rate in respect of the ten postures was 95.87% on average. In [25], the authors used three-dimensional joint histograms of depth images as well as hidden Markov models for the purpose of classifying postures. We compare these methods with that used in the current study; comparisons can be seen in Table 6.

Our hybrid approach has limitations. On the one hand, bending and lying postures cannot be recognized in more directions when the person orients to the Kinect because the feature value is unobvious and cannot extract a better feature value when the person orients to the Kinect at 0°. On the other hand, when the person laterally sits (± 90° oriented to the Kinect), the chair and the stool may be recognized as parts of the human body because of the reflective light of the chair or the stool, which may in turn lead to misjudgment about sitting posture.

Table 6 Comparison of the accuracy of several methods (%)

	A	B	C	D	E	F
Current paper	100	98.06	97.39	100	99.73	99.02
Paper [4]	99.38	97.25	99.69	–	99.22	99.53
Paper [8]	100	100	–	–	–	100
Paper [7]	95.87	95.87	–	–	95.87	–
Paper [25]	93.5	91.5	–	–	97.5	–

A denotes standing, B denotes sitting, C denotes kneeling, D denotes sitting cross-legged, E denotes lying, F denotes bending

4 Conclusions

In this study, we propose a novel hybrid approach for human body recognition. This approach can recognize six postures, namely, standing, sitting, kneeling, sitting cross-legged, lying, and bending. Different from other studies, we innovatively use the relevant knowledge of anthropometry and preliminarily judge four postures according to the natural ratio of the human body parts. The average recognition rate of all of the six postures is more than 99%. The standing and sitting cross-legged postures can be recognized when the person orients to the Kinect in any direction (from 0° to ±180°), and the recognition rate is 100%; the kneeling and sitting postures can be recognized when the person orients to the Kinect in the direction from 0° to ± 90°, and the recognition rate is more than 97%; and the bending and lying postures are recognized by BPNN, and the recognition rate is more than 99%. However, the bending and lying postures can only be recognized when the person orients to the Kinect in the direction from ± 45° to ± 90°. A part of the body joints is obscured, because we use only one Kinect. Feature vectors are consequently difficult to extract. In the further work, we will use multiple Kinect and aim to recognize human posture in all directions when the person orients to any Kinect.

Availability of data and materials
Please contact author for data requests.

Authors' contributions
BL and BB conceived and designed the experiments. BL performed the experiments. CH analyzed the data and contributed materials and tools. BL wrote the paper. All authors took part in the discussion of the work described in this paper. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]School of Computer Science and Technology, Changchun University of Science and Technology, Room B-1603, Keji Dasha, Weixing Road No.7186, Changchun 130022, Jilin, China. [2]College of Optical and Electronical Information, Changchun University of Science and Technology, Room B-1603, Keji Dasha, Weixing Road No.7186, Changchun 130022, Jilin, China.

References

1. L. Xia, C.C. Chen, J.K. Aggarwal, in *Computer Vision and Pattern Recognition Workshops IEEE*. Human detection using depth information by Kinect (2011), pp. 15–22
2. J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, A. Blake, in *Computer Vision and Pattern Recognition (CVPR) IEEE*. Real-time human pose recognition in parts from single depth images (2011), pp. 1297–1304
3. J. Shotton, T. Sharp, A. Kipman, et al., Real-time human pose recognition in parts from single depth images. Commun. ACM **56**(1), 116–124 (2013)
4. W.J. Wang, J.W. Chang, S.F. Haung, et al., Human posture recognition based on images captured by the Kinect sensor. Int. J. Adv. Robot. Syst. **13**(2), 1 (2016)
5. S.C. Hsu, J.Y. Huang, W.C. Kao, et al., Human body motion parameters capturing using kinect. Mach. Vis. Appl. **26**(7–8), 919–932 (2015)
6. T.L. Le, M.Q. Nguyen, T.T.M. Nguyen, in *International Conference on Computing, Management and Telecommunications IEEE*. Human posture recognition using human skeleton provided by Kinect (2013), pp. 340–345
7. P.K. Pisharady, Kinect based body posture detection and recognition system. Proc. SPIE Int. Soc. Opt. Eng. **8768**(1), 87687F (2013) -87687F-5
8. O. Patsadu, C. Nukoolkit, B. Watanapa, in *International Joint Conference on Computer Science and Software Engineering IEEE*. Human gesture recognition using Kinect camera (2012), pp. 28–32
9. Z. Xiao, M. Fu, Y. Yi, et al., in *International Conference on Intelligent Human-Machine Systems and Cybernetics IEEE*. 3D human postures recognition using Kinect (2012), pp. 344–347
10. L.A. Schwarz, A. Mkhitaryan, D. Mateus, et al., Human skeleton tracking from depth data using geodesic distances and optical flow ☆. Image Vis. Comput. **30**(3), 217–226 (2012)
11. O. Wasenmüller, D. Stricker, in *Asian Conference on Computer Vision*. Comparison of Kinect V1 and V2 depth images in terms of accuracy and precision (Springer, Cham, 2016), pp. 34–45
12. X. Li, Y. Wang, Y. He, G. Zhu, Research on the algorithm of human single joint point repair based on Kinect (a Chinese paper). Tech. Autom. Appl. **35**(4), 96–98 (2016)
13. H. Li, C. Zhang, W. Quan, C. Han, H. Zhai, T. Liu, An automatic matting algorithms of human figure based on Kinect depth image. J Chang Univ Sci Technol **39**(6), 81–84 (2016)
14. Human dimensions of Chinese adults, National standards of P. R. China GB10000–88, 1988
15. Y. Yin, J. Yang, C.S. Ma, J.H. Zhang, Analysis on difference in anthropomety dimenions between East and West human bodies (a Chinese paper). Standard Sci **7**, 10–14 (2015)
16. Blackwell S, Robinette K M, Boehmer M, et al. Civilian American and European Surface Anthropometry Resource (CAESAR). Volume 2: Descriptions[J]. Civilian American & European Surface Anthropometry Resource, 2002
17. J. Bougourd, P.U.K. Treleaven, in *International Conference on 3d Body Scanning Technologies, Lugano, Switzerland, 19–20 October*. National Sizing Survey – Size UK (2010), pp. 327–337
18. C.D. Fryar, Q. Gu, C.L. Ogden, Anthropometric reference data for children and adults: United States, 2007-2010[J]. Vital Health Stat. **252**, 1–48 (2012)
19. The full results of SizeUK. Database. 2010. [Online]. Available: https://www.arts.ac.uk/__data/assets/pdf_file/0024/70098/SizeUK-Results-Full.pdf.
20. A. Seidl, R. Trieb, H.J. Wirsching, in *World Congress on Ergonomics*. SizeGERMANY-the new German anthropometric survey conceptual design, implementation and results (2009)
21. Han, *Artificial Neural Network Tutorial, 1st Ed* (Beijing University of Posts and Telecommunications Press, Beijing, 2006), pp. 47–78
22. H.Y. Shen, Determining the number of BP neural network hidden layer units. J Tianjin Univ Technol **5**, 13–15 (2008)
23. J.M. Nasser, D.R. Fairbairn, The application of neural network techniques to structural analysis by implementing an adaptive finite-element mesh generation. AI EDAM **8**(3), 177–191 (1994)
24. S. Thawornwong, D. Enke, The adaptive selection of financial and economic variables for use with artificial neural networks. Aichi Gakuin Daigaku Shigakkai Shi **56**(10), 205–232 (2004)
25. L. Xia, C.C. Chen, J.K. Aggarwal, in *Computer Vision and Pattern Recognition Workshops. IEEE*. View invariant human action recognition using histograms of 3D joints (2012), pp. 20–27