PAPER

Posture Recognition Technology Based on Kinect

Yan LI[†], **Zhijie CHU**[††], ***Nonmembers***, ***and*** **Yizhong XIN**[††a)], ***Member***

SUMMARY Aiming at the complexity of posture recognition with Kinect, a method of posture recognition using distance characteristics is proposed. Firstly, depth image data was collected by Kinect, and three-dimensional coordinate information of 20 skeleton joints was obtained. Secondly, according to the contribution of joints to posture expression, 60 dimensional Kinect skeleton joint data was transformed into a vector of 24-dimensional distance characteristics which were normalized according to the human body structure. Thirdly, a static posture recognition method of the shortest distance and a dynamic posture recognition method of the minimum accumulative distance with dynamic time warping (DTW) were proposed. The experimental results showed that the recognition rates of static postures, non-cross-subject dynamic postures and cross-subject dynamic postures were 95.9%, 93.6% and 89.8% respectively. Finally, posture selection, Kinect placement, and comparisons with literatures were discussed, which provides a reference for Kinect based posture recognition technology and interaction design.

key words: *Kinect, depth image, distance characteristic, posture recognition*

1. Introduction

There are two methods of human posture recognition based on computer vision: model-free reconstruction and model-based reconstruction [1]. The model-free reconstruction method extracts pictures and video streams, and the human body postures are directly represented with silhouettes or contours from the perspective of images [2]. It does not need to solve human body model parameters, which simplifies the solution of human body posture expression. However, in the case of complex background, the silhouettes or contours are susceptible to noise pollution. In order to reduce the influence of noise, the literatures [3]–[5] used shape context descriptors to represent contour images, but the model-free reconstruction method was still easily affected by the factors such as viewpoint and individual differences of human body. For one case, in different viewpoints, the contour images of human body are quite different. For another case, when different people assume the same postures, the contour images are also quite different, resulting in the unsatisfactory results of posture recognition. The model-based reconstruction method is to recognize the posture according to the built human body model in the computer, which reduces the influence of covariates such as viewpoint [6]. However, two-dimensional model does not contain depth information, and thus it is difficult to recognize the posture changing perpendicularly to the lens [7]. Moreover, visual occlusion also increases the difficulty of human body model building [8].

The emergence of Kinect has changed this phenomenon. Kinect provides three-dimensional position information of 20 skeleton joints of human body. Through the infrared principle, it can identify the skeleton joints without the interference of natural light. However, due to the large amount of three-dimensional coordinate information of skeleton joints formed by Kinect, the time-space complexity of posture recognition increased [1], [9], [10]. As a result, it is necessary to reduce the complexity of posture expression. In addition, with the changes of the viewpoint and the distance between the body and the Kinect, the human skeleton shapes will also affect the posture recognition results. Therefore, it is necessary to explore a characteristic consistent posture recognition method, which is less affected by the viewpoints, distances between the body and the Kinect, and the body heights. Since dynamic posture (motion) can be represented as a series of static postures, static posture recognition lays a foundation for dynamic posture recognition. However, the same motion with different speed might lead to different lengths of the static posture sequences representing the motions, which would affect the accuracy of posture recognition.

In order to simplify the complexity of posture recognition, model-based reconstruction was adopted to recognize the static and dynamic postures. First, three-dimensional coordinate information of skeleton joints is captured, and those joints that contribute less to the postures are discarded. Second, the preserved skeleton joints data is normalized into characteristic consistent distance values that are less affected by the viewpoints, the distances between the body and the Kinect, and the body heights. Third, a shortest distance template matching method is proposed to recognize static postures. Fourth, the dynamic postures are decomposed into several static postures, and a minimum accumulative distance template matching method with dynamic time warping is proposed to recognize dynamic postures. Fifth, the proposed static and dynamic posture recognition methods were empirically evaluated. Last, posture selection, Kinect placement, and comparisons with other methods are discussed.

Manuscript received August 12, 2019.
Manuscript revised November 13, 2019.
Manuscript publicized December 12, 2019.

[†]The author is with Shenyang Sport University, Shenyang, 110102 China.
[††]The authors are with Shenyang University of Technology, Shenyang, 110142 China.
a) E-mail: xyz@sut.edu.cn

DOI: 10.1587/transinf.2019EDP7221

2. Related Work

2.1 Kinect-Based Interaction

Kinect was used in medical, robot and virtual reality fields recently. Lange et al. [11] used Kinect somatosensory technique to provide limb recovery training for sports rehabilitation patients by combining the skeleton data with virtual scenes. Chang et al. [12] designed a fitting mirror system with Kinect and AR technique to permit users to try clothes by gestures. Benko et al. [13] built the DepthTouch platform with Kinect, and users could directly interact with the 3D virtual scene projected on the plane. Cardo et al. [14] designed an application to detect falls with Kinect, which could reduce the risk of falls when left unattended. Tan et al. [15] proposed a posture matching algorithm combining Kinect's color data stream and depth data stream, which realized the high-resolution denoising. Raheja et al. [16] realized hand detection by segmenting Kinect's distance and depth vector. Li et al. [17] recognized sign language with Kinect, and displayed the corresponding meaning of action in the interactive interface. Xia et al. [18] used the Kinect depth image and the Canny operator to extract the data of the human body edge to detect and track human body. Shotton et al. [19] recognized human body parts by using depth image synthesized by pixel difference method and random forest algorithm. Gao et al. [20] captured the hand movement by Kinect, and realized the function of making virtual pottery by transferring the original data to Unity3D through Zigful.

2.2 Posture Recognition

Polana et al. [21] used temporal structure to identify different motions. Zhang et al. [22] proposed a time sequences extraction method, which extracted very short action sequences from long video sequences. Tian et al. [23] combined the locally consistent group sparse representation method with the temporal and spatial structure of each video sequence, and proposed an action recognition framework. Zhu et al. [24] proposed a spatiotemporal descriptor method to detect action events in complex scenes. Sun et al. [25] proposed a real-time sitting posture recognition algorithm based on Index Graph and BLS model, and proposed a double threshold cascade algorithm for the case of too many frames in the video. Jansen et al. [26] designed a disposable stretch skin sensor, which could be used for body position monitoring, rehabilitation feedback and detailed motion monitoring in the process of exercise and fitness. Wang et al. [27] realized the human posture automatic recognition in the process of CT scanning. Tapia et al. [28] placed the wireless accelerometer on the limbs and hips and placed the heart rate monitor on the chest to identify the physical activity and its intensity with the fast decision tree classifier. Bourke et al. [29] used the dual axis gyroscope sensor which was installed on the torso to measure the tilt and rotation speed of human body, and realized the algorithm to distinguish daily life and fall.

Zhang et al. [4] used the image representation of motion context to represent actions as 3D descriptors, which reduced the influence of noise points on body posture. Yao et al. [30] designed a robust vision system to detect the action of raising hands to ask questions through time and space segmentation, skin color recognition and other technologies, and solved the problem of false recognition caused by the change of light, the number of matching objects and other factors. Jiang et al. [31] put forward a motion recognition method based on depth video sequence to generate fuzzy amount of motion sequence by accumulating equal amount of motion to get different length subsequences and controlling fuzzy amount of motion to capture boundary information. Qi et al. [32] made up for the shortcoming that the static objects were often ignored in tracking-based recognition method through extracting video report, object information and tracking information at the same time.

Gianaria et al. [33] proposed a method to describe walking gait by using the three-dimensional skeleton information obtained by Kinect sensor, and found that some dynamic parameters related to knee, elbow and head motions are good candidates for robust gait characteristics. Wang et al. [34] extracted part characteristics from action set and expressed them with sparse matrix to decrease the noise and occlusion. Kim et al. [9] proposed the DMA (dynamic motion appearance) and the DMH (deep motion history) methods to identify human behavior by simply using depth map other than joints information. Chen et al. [35] proposed a GLAC (gradient local autocorrelation) method to extract characteristics from DMM (depth motion map). Chen et al. [36] proposed a human motion recognition method that utilized distance-weighted Tikhonov matrix with L2 regular cooperative suppression classifier to recognize motion from DMM. Chen et al. [37] proposed a method to obtain compact characteristics representation from DMM using LBP (local binary pattern). Yang et al. [38] proposed a characteristics representation method by gathering the local normal vectors of hypersurface in the depth sequence to form a Polynormal combining the local shape and motion information of human body. Bari et al. [39] designed a neural network framework for gait recognition and optimized the traditional machine learning model so that the accuracy rate of gait recognition reached 93.73%. Ali et al. [10] proposed an action recognition framework based on Kinect to detect human skeleton joints. The framework captured human actions in all directions by utilizing deep motion images so as to classify human behaviors. Deng et al. [1] developed a gait recognition system by using the deterministic learning algorithm and Kinect to improve the gait recognition accuracy. Sun et al. [40] proposed a model to recognize sign language and find out the difference frame and representative frame in videos. Ordóñez et al. [41] proposed a general depth framework based on CNN (Convolutional Neural Network) and LSTM (Long Short-Term Memory) recursive unit for human activity recognition.

Although Kinect interaction and posture recognition were studied in the above literatures, some focused on the posture recognition of partial human body [23], [33]–[35], [38], [39], and some did not consider the human body structures [1], [21], [30], [34]–[36], [40]. Moreover, some models were complicated or required an amount of computation [9], [31], [33], [34]. And what's more, most did not explore the posture characteristics from static and dynamic perspectives [1], [9], [10], [21]–[23], [25], [30], [31], [33]–[36], [39]–[41]. In order to reduce the complexity of posture recognition, the skeleton model was simplified, and characteristic consistent distance values were generated for the static and dynamic posture recognitions.

3. Static Posture Recognition

3.1 Depth Image Extraction

Kinect can extract three kinds of original data streams including (1) depth data stream that is extracted by depth sensor composed of infrared CMOS camera and infrared transmitters, (2) color video stream that is extracted by color camera, and (3) audio data stream that is extracted by microphone. The depth image formed by Kinect depth data stream reflects the distance between object and camera in the visible range.

Kinect uses infrared transmitters to project structured light into the whole space. Due to the different roughness of object surface in the space, different shapes of highly random speckles are formed. The sizes and shapes of the speckles at any two positions in the space are different. According to the speckle image collected by the infrared CMOS camera, the light source is calibrated. The shorter calibration interval, the higher accuracy achieves (Fig. 1). Through the correlation operation between the speckle image to be measured and the pre-stored reference image, the correlation image is obtained. The peak position of the correlation image is used to measure the position of the object in the space, and then the three-dimensional shape of the scene to be measured is obtained by the interpolation operation and the superposition of the peak value.

3.2 Skeleton Characteristics Extraction

The depth image captured by Kinect is used to track and locate the human skeleton joints. Kinect can recognize the whole body skeleton data of two people in the measurement space, generate about 30 frames of skeleton data per second, and provide three-dimensional coordinate information of 20 human skeleton joints. These joints are head (*H*), shoulder center (*SC*), shoulder left (*SL*) / right (*SR*), elbow left (*EL*) / right (*ER*), wrist left (*WL*) / right (*WR*), hand left (*HL*) / right (*HR*), spine (*S*), hip center (*IC*), hip left (*IL*) / right (*IR*), knee left (*KL*) / right (*KR*), ankle left (*AL*) / right (*AR*), and foot left (*FL*) / right (*FR*) (Fig. 2). The origin of Kinect skeleton tracking coordinate system is infrared camera. The positive directions of *X*, *Y* and *Z* are the left direction, the upward direction and the facing direction of the camera.

Thanh et al. [42] proposed the method of transforming joints information into three-dimensional skeleton histogram, and Sempena et al. [43] proposed the method of the quaternion characteristics extraction, whose purpose was to transform the three-dimensional coordinate information into the characteristics quantity for recognition. However, they did not consider the internal structure of human body. Considering the human body structure, this paper extracted the distance characteristics from the three-dimensional skeleton joints coordinate data and used them to represent the posture. First, the posture template database was established according to the collected posture samples. Second, the posture was defined through characteristic extraction, and the characteristic template database was formed, which is regarded as the modeling process. Third, the characteristics of postures to be identified were extracted. Last, postures were identified and classified by the classifier, which is regarded as the recognition process. The whole process of modeling and recognition is shown in Fig. 3.

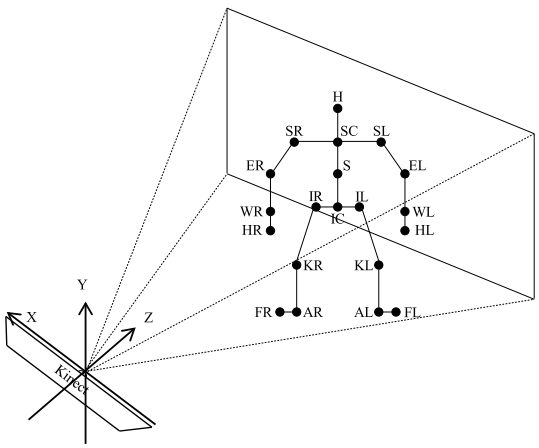

Fig. 2 Diagram of skeleton 3D tracking coordinate system.

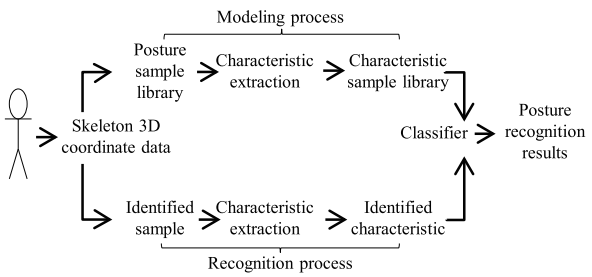

Fig. 3 Diagram of modeling and recognition process.

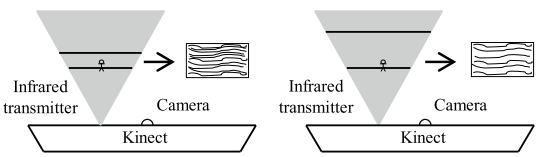

Fig. 1 Comparison diagram of calibration distance.

Through the observation of three-dimensional Kinect skeleton joints data, it is found that the head joint (*H* in Fig. 2) and 7 joints (*SR, SC, SL, S, IR, IC, IL* in Fig. 2) of the human torso part are relatively scale consistent. Variations of postures have little influence on these joints data which can be discarded in posture expression and only used as a reference in skeleton coordinate system. Furthermore, the human body postures are mainly represented by the positional changes of the limbs especially the *WL/WR, EL/ER, KL/KR, AL/AR* joints who contribute more to the posture expression. Therefore, these eight joints data are selected as the main parameters of posture recognition.

Considering that the distance between hand and wrist and the distance between foot and ankle were very short, it was not essential to choose both hand and wrist or both foot and ankle in posture recognition so that the computation amount could be decreased. Thus, left/right hand and left/right foot joints (*HL, HR, FL, FR* in Fig. 2) are discarded too. After selecting the spine joint (*S*) as the reference and calculating the distance between the 8 joints (*WL, WR, EL, ER, KL, KR, AL, AR*) and the spine joint (*S*) in three dimensional coordination, a total of 24 distance characteristic values are achieved. The distance characteristic vector set F_d of the selected joints of the limbs is represented as Eq. (1).

$$\begin{aligned} F_d = \{&d^X_{El-S}, d^Y_{El-S}, d^Z_{El-S}, d^X_{ER-S}, d^Y_{ER-S}, d^Z_{ER-S}, \\ &d^X_{WL-S}, d^Y_{WL-S}, d^Z_{WL-S}, d^X_{WR-S}, d^Y_{WR-S}, d^Z_{WR-S}, \\ &d^X_{KL-S}, d^Y_{KL-S}, d^Z_{KL-S}, d^X_{KR-S}, d^Y_{KR-S}, d^Z_{KR-S}, \\ &d^X_{AL-S}, d^Y_{AL-S}, d^Z_{AL-S}, d^X_{AR-S}, d^Y_{AR-S}, d^Z_{AR-S}\} \end{aligned} \quad (1)$$

For example, if the three-dimensional coordinates (X_{EL}, Y_{EL}, Z_{EL}) of the left elbow joint *EL* are (2.630015, −0.3500767, 0.3967017) extracted by the Kinect, and the three-dimensional coordinates (X_S, Y_S, Z_S) of the spine joint *S* are (2.676984, −0.1217509, 0.4159655), $d^X_{El-S} = X_{EL} - X_S = -0.046969$, $d^Y_{El-S} = Y_{EL} - Y_S = -0.2283258$, and $d^Z_{El-S} = Z_{EL} - Z_S = -0.0192638$ represent the distances from *EL* to *S* on *X*-axis, *Y*-axis and *Z*-axis.

Considering that the skeleton model may vary when the distance between human and Kinect or the shooting angle of Kinect camera changes, a reference distance between *SC* to *S* on *Y*-axis was selected to normalize the distance characteristics to obtain scale consistent model. A new set of distance characteristics is formed after normalization and expressed as Eq. (2).

Again, if the three-dimensional coordinates (X_{SC}, Y_{SC}, Z_{SC}) of the shoulder center joint *SC* are (2.585047, −0.1356404, 0.8063439) extracted by the Kinect, and $d^Y_{SC-S} = Y_{SC} - Y_S = -0.0138895$ is the distances from *SC* to *S* on *Y*-axis, $d^X_{El-S}/d^Y_{SC-S} = (-0.046969)/(-0.0138895) = 3.381619209$, $d^Y_{El-S}/d^Y_{SC-S} = (-0.2283258)/(-0.0138895) = 16.4387343$, and $d^Z_{El-S}/d^Y_{SC-S} = (-0.0192638)/(-0.0138895) = 1.386932575$ are normalized distance characteristic values of joint *EL* on *X*-axis, *Y*-axis and *Z*-axis. Although the values of d^X_{El-S} and d^Y_{SC-S} will change with the distance between the human body and Kinect, the value of d^X_{El-S}/d^Y_{SC-S} will remain unchanged. Similarly, the other 23 values of F_D will all remain unchanged, which provides the solution of characteristic consistence.

$$\begin{aligned} F_D = \{&d^X_{El-S}/d^Y_{SC-S}, d^Y_{El-S}/d^Y_{SC-S}, d^Z_{El-S}/d^Y_{SC-S}, \\ &d^X_{ER-S}/d^Y_{SC-S}, d^Y_{ER-S}/d^Y_{SC-S}, d^Z_{ER-S}/d^Y_{SC-S}, \\ &d^X_{WL-S}/d^Y_{SC-S}, d^Y_{WL-S}/d^Y_{SC-S}, d^Z_{WL-S}/d^Y_{SC-S}, \\ &d^X_{WR-S}/d^Y_{SC-S}, d^Y_{WR-S}/d^Y_{SC-S}, d^Z_{WR-S}/d^Y_{SC-S}, \\ &d^X_{KL-S}/d^Y_{SC-S}, d^Y_{KL-S}/d^Y_{SC-S}, d^Z_{KL-S}/d^Y_{SC-S}, \\ &d^X_{KR-S}/d^Y_{SC-S}, d^Y_{KR-S}/d^Y_{SC-S}, d^Z_{KR-S}/d^Y_{SC-S}, \\ &d^X_{AL-S}/d^Y_{SC-S}, d^Y_{AL-S}/d^Y_{SC-S}, d^Z_{AL-S}/d^Y_{SC-S}, \\ &d^X_{AR-S}/d^Y_{SC-S}, d^Y_{AR-S}/d^Y_{SC-S}, d^Z_{AR-S}/d^Y_{SC-S}\} \end{aligned} \quad (2)$$

3.3 Static Posture Classification

The three-dimensional coordinates of the skeleton joints vary little when the subjects remain still. To reduce the unnecessary calculation, the averages of 30-frame three-dimensional skeleton data extracted by Kinect per second were converted to distance characteristic values to represent the static posture *G* which can be expressed as Eq. (3).

$$G = F = (F_1, F_2, \cdots, F_i, \cdots, F_{24}) \quad (3)$$

The *F* here represents the set of distance characteristics vector, and the F_i represents one of the distance characteristic values e.g. d^X_{El-S}/d^Y_{SC-S} in Eq. (2).

A posture template set is constructed by recording the subject's posture in advance and calculating all the 24 distance characteristic values of the posture. Given a posture sample to be identified, the Euclidean distances between the sample and all the postures in the template set are calculated. And the identified sample is classified into the category of the sample in the template set that Euclidean distance is shortest to the identified sample. For a given posture sample G_q to be identified, and a posture template set $\{G_1, G_2, \cdots, G_i, \cdots, G_P\}$, the Euclidean distance $d(G_q, G_i)$ between G_q and G_i is calculated as Eq. (4).

$$d(G_q, G_i) = \sqrt{\sum_{\omega=1}^{24} (F_{q\omega} - F_{i\omega})^2} \quad (4)$$

P is the size of the posture template set. The F_q and the F_i are distance characteristics vectors of the sample and one posture in the template set. The $F_{q\omega}$ and the $F_{i\omega}$ are a certain distance characteristic value of the sample and one posture in the template set. The larger $d(G_q, G_i)$, the lower similarity between the two samples is.

The posture classification label of identified posture G_q is assigned as Eq. (5).

$$\begin{aligned} &lable(G_q) = lable(G_m), \\ &m = arg \min_{i=1,\cdots,n} d(G_q, G_i) \end{aligned} \quad (5)$$

m is the number of a sample in the template set who is

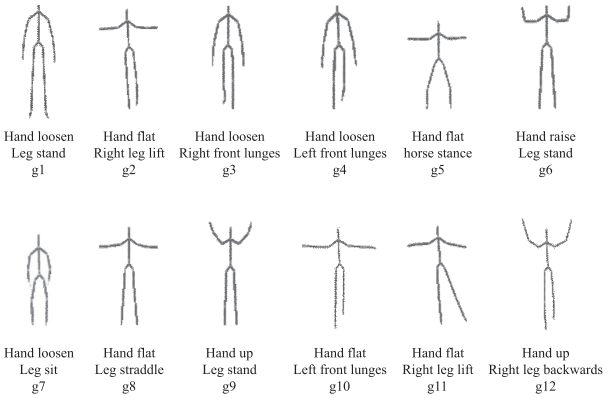

Fig. 4 The static postures to be identified.

Table 1 Static posture recognition results.

Action	Precision	Recall	F_1-score
P1	0.923	0.960	0.941
P2	1.000	0.940	0.969
P3	0.961	0.980	0.970
P4	1.000	0.940	0.969
P5	0.980	0.980	0.980
P6	1.000	0.920	0.958
P7	1.000	1.000	1.000
P8	0.932	0.960	0.946
P9	0.902	0.920	0.911
P10	0.926	1.000	0.962
P11	1.000	0.950	0.974
P12	0.906	0.960	0.932

the most similar to the identified sample, and the classification label of G_q is assigned to that of G_m. *arg* means to get the number.

3.4 Subjects, Experimental Conditions and Results

10 subjects (4 females, 6 males) participated in the experiment. Their average age was 24.7 ranging from 18 to 43, and their average height was 1.69m ranging from 1.59m to 1.78m. At all time, one subject was selected as a identified sample to recognize his/her posture, and the other subjects' postures were used as the template set, which is arranged as leave-one-out cross validation (LOOCV). All the subjects were selected as identified sample once. 12 postures were identified in the experiment (Fig. 4).

A Microsoft Kinect for Windows was used in the experiment. The Kinect was placed 1.2m from the ground. The subjects were asked to stand 2.4m away from Kinect with deflection angle less than 20°. The experiment was carried out in natural light environment. The experimental program ran on an Acer Aspire v5-473g laptop computer.

The experimental results showed that the average recognition rate was 95.9%, which indicated that the proposed posture recognition method is feasible and effective. Detailed results are shown in Table 1.

The postures g7 and g10 were recognized completely, which was mainly because these two postures are quite different from other postures. Moreover, the bending of legs resulted in greater differentiation from other postures so that the classifier could distinguish them accurately. For the g6 and g9 case, their lower limb postures are same, and their upper limb postures are less different, so that misrecognitions often occur, which resulted in the lowest recognition rate.

4. Dynamic Posture Recognition

A key matter for the dynamic posture recognition is to express the dynamic posture. Considering that dynamic posture can be represented with the combination of a series of skeleton frame data, and each frame is equivalent to a static posture, the motion (dynamic posture) can be regarded as a combination of static postures. For example, a lifting motion may be regarded as a sequence of static posture g7, g8 and g9 in Fig. 4. Therefore, static posture recognition lays a foundation of dynamic posture recognition. Dynamic postures are represented with a continuous static posture sequences over a period of time. Based on the characteristic values, a dynamic posture M is expressed as Eq. (6).

$$M = (G_1, G_2, \cdots, G_i, \cdots, G_n) \tag{6}$$

The G_i is the i^{th} static posture distance characteristic vector which is a 24-dimensional distance characteristic value set described as Eq. (2). The n is the total number of static postures contained in a dynamic posture. Thus, the characteristic vector of a motion is $24 \times n$-dimensional.

4.1 Dynamic Time Warping

Dynamic posture recognition has to construct the model both temporally and spatially. When a person makes the same motion with different speed, a time shifting will be found in the curve of motion on the time axis. If a motion is completed quickly, the posture sequence representing the motion will be correspondingly short, and vice versa.

In order to judge whether the curves of two samples belong to the same motion, the Euclidean distance is used to calculate the distance between the characteristic values of two samples. However, in order to meet the conditions of Euclidean distance algorithm, the sequence lengths of the two curves are required to be consistent, and the corresponding characteristic values on the curves are aligned on the time axis. Sakoe et al. [44] proposed an optimization algorithm of dynamic time warping to solve the irregularity on the time axis in speech recognition. Considering that this algorithm reduced distortion and maximizes overlap between two sequences (Fig. 5), the dynamic time warping (DTW) algorithm is also used to match two characteristic value curves according to the minimum accumulative distance.

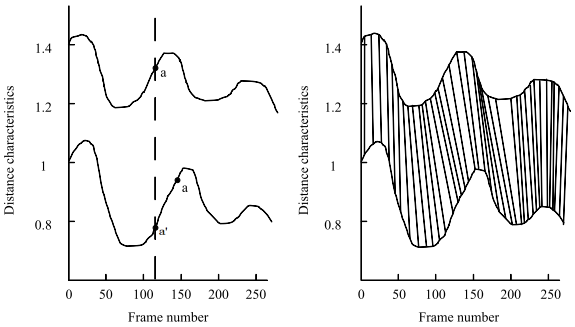

Fig. 5 The mapping relationship after dynamic time warping.

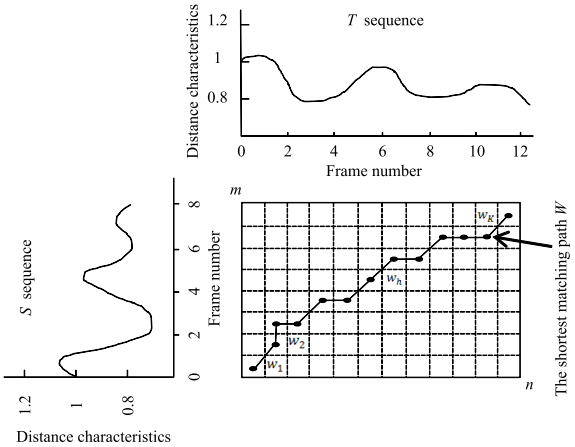

Fig. 6 The shortest matching path between dynamic posture T and S.

4.2 Calculation of the Accumulative Distance

The motions were classified through the minimum accumulative distance of the characteristic vectors between a dynamic posture to be identified and a certain dynamic posture in the template set.

Given two dynamic postures T and S which are composed of static postures $(T_1, T_2, \cdots, T_n)$ and $(S_1, S_2, \cdots, S_m)$, if $n = m$, the accumulative distance is the value of $\sum_{i=1}^{m} d(T_i, S_j)$ which is calculated as Eq. (4).

If $n \neq m$, the accumulative distance is calculated through the shortest matching path algorithm with dynamic time warping which is illustrated as following.

Construct an $n \times m$ Matrix as shown in Fig. 6. Dynamic posture T which is a static sequence $(T_1, T_2, \cdots, T_n)$ is represented on the horizontal axis. Dynamic posture S which is a static sequence $(S_1, S_2, \cdots, S_m)$ is represented on the vertical axis. The matrix element (i, j) represents the distance $d(T_i, S_j)$ between the two static postures T_i and S_j. $d(T_i, S_j)$ is calculate as Eq. (7) that is consistent with Eq. (4).

$$d(T_i, S_j) = \sqrt{\sum_{\omega=1}^{24} (T_{i\omega} - S_{j\omega})^2} \tag{7}$$

$T_{i\omega}$ is the ω^{th} distance characteristic value of the i^{th} static posture in dynamic posture T. $S_{j\omega}$ is the ω^{th} distance characteristic value of the j^{th} static posture in dynamic posture S. T_i is the i^{th} static posture in dynamic posture T. S_j the j^{th} static posture in dynamic posture S.

Construct a shortest matching path W from (T_1, S_1) to (T_n, S_m) to align the two dynamic postures T and S on the time axis. The point w_h through which the path W passes indicates that two static postures T_i and S_j are the h^{th} aligned on the time axis, and the align relationship is expressed as $w_h(i, j)$. The shortest matching path is expressed as Eq. (8).

$$\begin{aligned} &W = (w_1, w_2, \cdots, w_h, \cdots, w_K), \\ &\quad max(m, n) \leq K \leq m + n - 1 \end{aligned} \tag{8}$$

K here is the total number of points in the path.

The shortest matching path has three constraints:

(1) Continuity: given an align relationship $w_h(i, j)$, when the next align relationship $w_{h+1}(i', j')$ is looked for, i' and j' have to meet the conditions of $i' - i \leq 1$ and $j' - j \leq 1$.

(2) Monotonicity: given an align relationship $w_h(i, j)$, when the next align relationship $w_{h+1}(i', j')$ is looked for, i' and j' have to meet the conditions of $i' - i \geq 0$ and $j' - j \geq 0$.

(3) Boundary: $w_1(1, 1)$ and $w_K(n, m)$ are the two boundaries.

With these constraints, the next align relationship to $w_h(i, j)$ in the shortest matching path can only be $w_{h+1}(i + 1, j)$, $w_{h+1}(i, j + 1)$, or $w_{h+1}(i + 1, j + 1)$.

The accumulative distance $\gamma(n, m)$ is calculated as Eq. (9).

$$\begin{aligned} \gamma(i, j) &= d(T_i, S_j) + \\ &\quad min\{\gamma(i-1, j-1), \gamma(i-1, j), \gamma(i, j-1)\} \end{aligned} \tag{9}$$

The accumulative distance $\gamma(n, m)$ indicates the similarity of dynamic postures T and S. The smaller $\gamma(n, m)$, the more similar the two dynamic postures T and S are.

4.3 Dynamic Posture Classification

Given a dynamic posture T to be identified and a posture template set of m dynamic postures $\{S_1, S_2, \cdots, S_i, \cdots, S_m\}$, calculate the accumulative distances between the sample and all the postures in the template set in turn, so that the minimum accumulative distance can be obtained. Then, the sample is assigned to the category of the posture in the template set who has the minimum accumulative distance to the sample. The classification result of posture T is expressed as Eq. (10).

$$\begin{aligned} lable(T) &= lable(S_c), c = arg(min\gamma(T, S_i)), \\ i &= 1, 2, \cdots, m \end{aligned} \tag{10}$$

c here is the number of the posture in the template set who has the minimum accumulative distance from the identified posture T.

In order to prevent the postures which were not in the template set from misidentification, the similarity threshold τ was set to identify the postures that were quite different

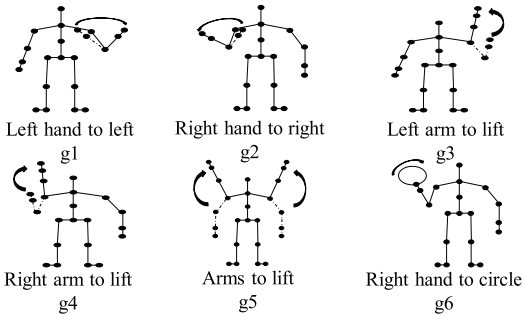

Fig. 7 The dynamic postures to be identified.

Table 2 Dynamic posture recognition results in non-cross-subject part.

Action	Precision	Recall	F_1-score
g1	1.000	0.940	0.969
g2	0.949	0.925	0.937
g3	1.000	0.967	0.983
g4	1.000	0.975	0.987
g5	1.000	0.917	0.957
g6	0.964	0.892	0.927

from the postures in the template set as rejected. Threshold τ is calculated as Eq. (11).

$$\tau = max(\gamma(S_i, S_j)), i, j = 1, 2, \cdots, M; i \neq j \tag{11}$$

$\gamma(S_i, S_j)$ here is the accumulative distance of dynamic posture S_i to S_j in the template set. The threshold τ is set to the maximum accumulative distance among all postures in the posture set.

4.4 Experimental Results

6 subjects (3 females, 3males) participated in the experiment. Their average age was 25.5 ranging from 18 to 43, and their average height was 1.71m ranging from 1.59m to 1.78m. 6 kinds of postures as shown in Fig. 7 were identified in the experiment. The experimental equipment and conditions were consistent with those in Sect. 3.4.

The experiment was divided into non-cross-subject and cross-subject parts. In non-cross-subject part, each subject was asked to record 6 kinds of postures of himself in advance as templates, and then the subject were asked to make the postures same as the record ones for identification. Every posture was performed 20 times. In total, the experiment consisted of 6 subjects × 6 postures × 20 times = 720 trials.

The experimental results showed that the average recognition rate of the six postures was 93.6%. The speed of each posture made by the subjects was not limited, which indicated that the method was robust and effective for motion recognition. Detailed results are shown in Table 2.

In cross-subject part, one of the six subjects was randomly selected to record his/her postures as template in advance. Then, the postures made by the other five subjects were identified according to the template. Each posture was identified 20 times. In total, the experiment consisted of 5 subjects × 6 postures × 20 times = 600 trials.

Table 3 Dynamic posture recognition results in cross-subject part.

Action	Precision	Recall	F_1-score
g1	1.000	0.891	0.942
g2	0.945	0.862	0.902
g3	1.000	0.936	0.967
g4	1.000	0.951	0.975
g5	1.000	0.897	0.896
g6	0.944	0.852	0.938

The experimental results showed that the average recognition rate of the six postures was 89.8%. Compared to non-cross-subject part, there are two possible reasons for the decline of recognition rate. For one hand, the postures made by the subjects didn't reach the designated position, which is different from the template. For the other hand, some postures may be rejected due to the differences of skeleton profiles between subjects. Detailed results are shown in Table 3.

The recognition rate of postures g3 and g4 were higher than others, which was mainly because these two postures were quite different from other postures, and the single hand lifting motion was significantly different from other motions. However, the recognition rate of postures g2 and g6 were lower than the other, which was mainly because these two postures were quite similar, resulting in a misidentification of each other.

In order to add new postures not existing in the posture template set, the threshold τ was set. However, the accuracy of posture recognition will be affected by the threshold too. Without the threshold, all postures will be assigned an existing category even mistakenly for the reason that the minimum accumulative distance will be obtained inevitably in the recognition process, and the posture label will be assigned according to the minimum accumulative distance. When the threshold τ is set, some postures may not meet the threshold requirements and be rejected. The rejected postures constituted a candidate set of new postures that is not found in the template set.

When the threshold was set to the longest accumulative distance among postures in the template set, the rejection rates of the non-cross-subject and the cross-subject gesture recognitions were 5.0% and 8.5% respectively.

5. Discussion

5.1 Collection of Postures

For the reason that the structure of human body is complex, various postures can be made through the changes of skeleton position. In order to make the subjects understand postures accurately in experiment, daily and typical postures were selected and expressed in appropriate ways.

Since a dynamic posture can be regarded as a sequence of static postures, static posture recognition lays a foundation for dynamic posture recognition. Static postures were selected carefully and cautiously. In order that the static pos-

tures used in the experiments can be made accurately by subjects, they were collected through questionnaire, gymnastics, comics and on-site observation for ease of understand. Moreover, priorities were given to those postures with less occlusion among the 20 joints extracted by Kinect. As a result, three types of static postures which were standing, sitting and squatting were collected. Moreover, standing postures were divided into standing, golden chicken independence, lunge, straddling, forward leaning, backward leaning, tiptoeing and other postures. Sitting postures were divided into cross legged sitting, sitting, kneeling, meditation and other postures. Squatting postures were divided into deep-squatting, semi-squatting, horse-squatting, leg pressing-squatting and other postures. After comprehensive consideration of typicality, differentiation, less occlusion and less difficulty, the static postures in Fig. 4 and the dynamic postures in Fig. 8 were finally confirmed.

It is essential to make sure that the experimenter and the subjects reach a consensus on postures understanding. Different people may understand the postures with same name quite differently. In order to make sure that the subjects make accurate and appropriate postures, a "graphic + text + oral interpretation" posture conveying method was used.

5.2 Kinect Placement

Kenect placement is quite important. According to the Kinect parameter specification, its viewing angles are 43.5° in the vertical direction and 57.5° in the horizontal direction, and its camera can be adjusted within 28° upward and downward. The effective viewing distance for skeleton data capture is from 0.8m to 4m by default. Nathan et al. [45] found that as the distance between the object and Kinect increased, the deviation between the distance in the depth image and the actual distance increased. On the other hand, when the distance between the subject and Kinect is 1m, Kinect cannot capture the whole body skeleton data because of too short distance. Moreover, the height of Kinect placement might also affect the number of joints identified. In order to analyze the postures accurately, it is necessary to make clear and set up the Kinect placement height and the appropriate distance range between the subjects and the Kinect.

A pilot experiment of Kinect placement was performed. Kinect was placed at a height of 0.4m, 0.6m, 0.8m, 1.0m, 1.2m, 1.4m and 1.6m from the ground respectively, and then the shortest and longest distances that Kinect could capture all the 20 skeleton joints of an upright subject 1.78m high were measured. It was found that the proper distance between human body and Kinect was from 1.8m to 3.6m, and the deflection angle between human body and Kinect should not exceed 45°. The height of Kinect from the ground had no significant effect on the proper distance between human body and Kinect.

A further pilot experiment of Kinect placement was performed. Kinect was placed 1.8m, 2.4m, 3m and 3.6m away from the subjects and at a height of 0.4m, 0.8m, 1.2m and 1.6m from the ground respectively. The 7200 frames of skeleton data collected in the experiment were converted into 24-dimensional distance characteristic values as described in Eq. (2). It was found that these distance characteristic values were not affected by the height of Kinect and the distance between the subject and Kinect, which indicated that the 24-dimensional distance characteristic vector F_D described in Sect. 3.2 could remain consistence. When the subjects turned left or right more than 45° on the basis of facing Kinect, the distance characteristic values changed significantly. Thus, the subjects were asked to face Kinect as they can in the formal experiments.

Table 4 Comparisons with other posture recognition works.

Method	Accuracy(%)
STOP feature[30]	84.8
ROPs[31]	85.9
DMA+DMH+HOG[32]	90.5
DMMs-based GLAC[33]	90.5
DMM-l_2+regularized[34]	90.5
DMM-LBP-FF[35]	91.9
Polynormals[36]	92.7
MLP-RMSProp-tanh[37]	93.7
DMM[38]	94.6
Deterministic Learning[39]	97.7
Latent SVM[40]	84.7
Deep learning[41]	95.8
Proposed method for Static posture	95.9
Proposed method for dynamic posture recognition (non-cross-subject)	93.6
Proposed method for dynamic posture recognition (cross-subject)	89.8

5.3 Comparisons to Other Literatures and Significance

Through literature review, posture recognition methods were compared. The comparison results are shown in Table 4.

The Stop feature [33] and ROPS [34] modelled in high-dimensional situation, which increased the computational complexity. The DMA + DMH + HOG [9] modelled by means of establishing multi perspective camera, which was time-consuming. The DMMs-based GLAC [35] and the Polynormals [38] modelled with partial information, which had some one-sided posture recognition. The DMM-l2-regularized method [36] might implement unreliably due to large intra-class changes. The MLP-RMSProp-tanh [39] and the Deterministic Learning [1] methods were only applied in dynamic gait recognition, and static postures and systemic postures were not examined. When SVM algorithm [40] was used in posture recognition, the meaningless transition between two sign language postures often reduced recognition accuracy. Deep learning [41] method was difficult to distinguish the overlapping of motions, and the recognition accuracy was often affected for reciprocating motions.

The comparison results show that the recognition accuracies of our methods are in the middle level to the existing posture recognition literature. However, the method proposed in this paper has the following advantages and sig-

nificance. (1) The model construction was relatively simple so that the calculation amount was reduced, resulting in a method ease of implement. (2) A characteristic consistent model that was not influenced by the camera distance was proposed according to the human skeleton structure. The robustness of the recognition was improved. (3) The same 24-dimensional distance characteristics were used for static and dynamic posture expression. The shortest distance and the minimum accumulative distance with dynamic time warping were used to match the static and dynamic postures respectively, which assured the consistency of static and dynamic posture recognitions. (4) The threshold τ was used to control the strictness of posture recognition. According to the actual requirements and the differentiation between postures, recognition level can be adjusted and the posture template set can be expanded. (5) The proposed posture recognition methods provide a methodology for reference.

6. Conclusion

In order to reduce the complexity of posture recognition and the influence of human internal structure on posture recognition, the posture recognition methods based on Kinect in static and dynamic conditions were proposed. First, the depth image data extracted by Kinect was normalized to characteristic consistent distances which were used to identify and classify the static postures according to the shortest distance. Next, the dynamic time warping (DTW) was used to solve the posture sequence inconsistency on the time axis. And then, the minimum accumulative distance was used to identify the dynamic postures. The experimental results showed that the recognition rates of static postures, non-cross-subject dynamic postures and cross-subject dynamic postures were 95.9%, 93.6% and 89.8% respectively. The experimental results verified the robustness and effectiveness of the proposed method. The proposed method and experimental results have reference significance for the posture recognition.

Limitations and future work: For the reason that the shortest distance was used to identify the samples, a posture can be classified to only one posture category, which may reject some postures or misidentify some postures as others. In the future work, k-Nearest Neighbors (KNN) method might be used to further improve the recognition effect by constructing algorithms and finding more suitable k values. In addition, the proposed method can be further applied to posture authentication and abnormal posture detection in future work.

Acknowledgments

This study has been partially supported by National Natural Science Foundation of China (No. 61100091), Program for Liaoning Excellent Talents in University (LJQ2012007), SRF for ROCS, SEM ([2013]693), Shenyang University of Technology Research Start-up Foundation for Doctors, and Shenyang University of Technology Young Key Academic Teacher Foundation.

References

[1] M. Deng and C. Wang, "Human gait recognition based on deterministic learning and data stream of Microsoft Kinect," IEEE Trans. Circuits Syst. Video Technol., 2018, pp.3636–3645, 2018. DOI: 10.1109/TCSVT.2018.2883449.

[2] J. Yamato, J. Ohya, and K. Ishii, "Recognizing human action in time-sequential images using hidden Markov model," Proc. CVPR 1992, pp.379–385, 1992. DOI: 10.1109/CVPR.1992.223161.

[3] F. Lv and R. Nevatia, "Single view human action recognition using key pose matching and Viterbi path searching," Proc. CVPR 2007, pp.17–22, 2007. DOI: 10.1109/CVPR.2007.383131.

[4] Z. Zhang, Y. Hu, S. Chan, and L.-T. Chia, "Motion context: A new representation for human action recognition," Proc. ECCV 2008, vol.5305, pp.817–829, 2008. DOI: 10.1007/978-3-540-88693-8_60.

[5] S.S. Beauchemin and J.L. Barron, "The computation of optical flow," ACM Computing Surveys, vol.27, no.3, pp.433–466, 1995. DOI: 10.1145/212094.212141.

[6] K. Yang, Y. Dou, S. Lv, F. Zhang, and Q. Lv, "Relative distance features for gait recognition with Kinect," Journal of Visual Communication and Image Representation, vol.39, pp.209–217, 2016. DOI: 10.1016/j.jvcir.2016.05.020.

[7] P. Turaga, R. Chellappa, V.S. Subrahmanian, and O. Udrea, "Machine recognition of human activities: A survey," IEEE Trans. Circuits Syst. Video Technol., vol.18, no.11, pp.1473–1488, 2008. DOI: 10.1109/TCSVT.2008.2005594.

[8] L. Torresani, A. Hertzmann, and C. Bregler, "Learning non-rigid 3D shape from 2D motion," Proc. NIPS'03, pp.1555–1562, 2003.

[9] D. Kim, W. Yun, H. Yoon, and J. Kim, "Action recognition with depth maps using HOG descriptors of multi-view motion appearance and history," Proc. UBICOMM 2014, pp.126–130, 2014.

[10] H.H. Ali, H.M. Moftah, and A.A.A. Youssif, "Real-time framework for human action recognition," Proc. ICBET'18, pp.55–60, 2018. DOI: 10.1145/3208955.3208972.

[11] B. Lange, S. Koenig, E. McConnell, C.-Y. Chang, R. Juang, E. Suma, M. Bolas, and A. Rizzo, "Interactive game-based rehabilitation using the Microsoft Kinect," Proc. VR'12, pp.171–172, 2012. DOI: 10.1109/VR.2012.6180935.

[12] H.-T. Chang, Y.-W. Li, H.-T. Chen, S.-Y. Feng, and T.-T. Chien, "A dynamic fitting room based on Microsoft Kinect and augmented reality technologies," Proc. HCI'13, vol.8007, pp.177–185, 2013. DOI: 10.1007/978-3-642-39330-3_19.

[13] H. Benko and A. Wilson, "Depth-Touch: Using depth-sensing camera to enable freehand interactions on and above the interactive surface," IEEE Workshop on Tabletops and Interactive Surfaces, vol.8, pp.1–7, 2009.

[14] A.L. Cardo, V.M.R. Penichet, M.D. Lozano, and J.E. Garrido, "Falls and fainting detection at home through movement-based interaction," Proc. Interacción 2018, pp.1–2, 2018. DOI: 10.1145/3233824.3233857.

[15] F. Tan, X. Feng, and Z. Xia, "An efficient algorithm for human body matting with RGB-D data," Proc. ICVR 2018, pp.40–43, 2018. DOI: 10.1145/3198910.3198912.

[16] J.L. Raheja, A. Chaudhary, and K. Singal, "Tracking of fingertips and centers of palm using Kinect," Proc. CIMSim'11, pp.248–252, 2011. DOI: 10.1109/CIMSim.2011.51.

[17] K.F. Li, K. Lothrop, E. Gill, and S. Lau, "A web-based sign language translator using 3D video processing," Proc. NBiS'11, pp.356–361, 2011. DOI: 10.1109/NBiS.2011.60.

[18] L. Xia, C.-C. Chen, and J. Aggarwal, "Human detection using depth information by Kinect," Proc. CVPR 2011, pp.15–22, 2011. DOI: 10.1109/CVPRW.2011.5981811.

[19] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," Proc. CVPR 2011,

pp.1297–1304, 2011. DOI: 10.1109/CVPR.2011.5995316.
[20] Z. Gao, J. Li, H. Wang, and G. Feng, "DigiClay: An interactive installation for virtual pottery using motion sensing technology," Proc. ICVR 2018, pp.126–132, 2018. DOI: 10.1145/3198910.3234659.
[21] R. Polana and R. Nelson, "Recognition of motion from temporal texture," Proc. CVPR 1992, pp.129–134, 1992. DOI: 10.1109/CVPR.1992.223216.
[22] Z. Zhang, Z. Kuang, P. Luo, L. Feng, and W. Zhang, "Temporal sequence distillation: Towards few-frame action recognition in videos," Proc. MM'18, pp.257–264, 2018. DOI: 10.1145/3240508.3240534
[23] Y. Tian, Q. Ruan, G. An, and Y. Fu, "Action recognition using local consistent group sparse coding with spatio-temporal structure," Proc. MM'16, pp.317–321, 2016. DOI: 10.1145/2964284.2967234.
[24] G. Zhu, M. Yang, K. Yu, W. Xu, and Y. Gong, "Detecting video events based on action recognition in complex scenes using spatio-temporal descriptor," Proc. MM'09, pp.165–174, 2009. DOI: 10.1145/1631272.1631297.
[25] W. Sun, Z. Zhou, and H. Li, "Sitting posture recognition in real-time combined with index map and BLS," Proc. ICIAI 2019, pp.101–105, 2019. DOI: 10.1145/3319921.3319955.
[26] K. Jansen, B. Tarren, and M. Slingerland, "Disposable, stretchable on-skin sensors for posture monitoring," Proc. WearSys'18, pp.1–4, 2018. DOI: 10.1145/3211960.3211969.
[27] W. Wang, F. Zhang, and L. Geng, "Posture recognition in CT scanning based on HOG feature and mixture-of-parts model," Proc. IMIP'19, pp.62–66, 2019. DOI: 10.1145/3332340.3332351.
[28] E.M. Tapia, S.S. Intille, W. Haskell, K. Larson, J. Wright, A. King, and R. Friedman, "Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor," Proc. ISWC'07, pp.37–40, 2007. DOI: 10.1109/ISWC.2007.4373774.
[29] A.K. Bourke and G.M. Lyons, "A threshold-based fall-detection algorithm using a bi-axial gyroscope sensor," Medical Engineering & Physics, vol.30, no.1, pp.84–90, 2008. DOI: 10.1016/j.medengphy.2006.12.001.
[30] J. Yao and J.R. Cooperstock, "Arm gesture detection in a classroom environment," Proc. WACV'02, pp.153–157, 2002. DOI: 10.1109/ACV.2002.1182174.
[31] M. Jiang, K. Jin, and J. Kong, "Action recognition using multi-temporal DMMs based on adaptive vague division," Proc. ICIGP 2018, pp.8–13, 2018. DOI: 10.1145/3191442.3191462.
[32] L. Qi, X. Lu, and X. Li, "Action recognition by jointly using video proposal and trajectory," Proc. ICVISP 2018, pp.1–7, 2018. DOI: 10.1145/3271553.3271563.
[33] E. Gianaria, N. Balossino, M. Grangetto, and M. Lucenteforte, "Gait characterization using dynamic skeleton acquisition," Proc. MMSP'13, pp.440–445, 2013. DOI: 10.1109/MMSP.2013.6659329.
[34] J. Wang, Z. Liu, J. Chorowski, Z. Chen, and Y. Wu, "Robust 3D action recognition with random occupancy patterns," Proc. ECCV 2012, pp.872–885, 2012. DOI: 10.1007/978-3-642-33709-3_62.
[35] C. Chen, Z. Hou, B. Zhang, J. Jiang, and Y. Yang, "Gradient local auto-correlations and extreme learning machine for depth-based activity recognition," Proc. ISVC 2015, pp.613–623, 2015. DOI: 10.1007/978-3-319-27857-5_55.
[36] C. Chen, K. Liu, and N. Kehtarnavaz, "Real-time human action recognition based on depth motion maps," Journal of Real-Time Image Processing, vol.12, no.1, pp.155–163, 2016. DOI: 10.1007/s11554-013-0370-1.
[37] C. Chen, R. Jafari, and N. Kehtarnavaz, "Action recognition from depth sequences using depth motion maps-based local binary patterns," Proc. WACV 2015, pp.1092–1099, 2015. DOI: 10.1109/WACV.2015.150.
[38] X. Yang and Y. Tian, "Polynormal Fisher vector for activity recognition from depth sequences," Proc. SA'14, pp.1–4, 2014. DOI: 10.1145/2668956.2668962.
[39] A.S.M.H. Bari and M.L. Gavrilova, "Multi-layer perceptron architecture for Kinect-based gait recognition," Proc. CGI 2019, pp.356–363, 2019. DOI: 10.1007/978-3-030-22514-8_31.
[40] C. Sun, T. Zhang, and C. Xu, "Latent support vector machine modeling for sign language recognition with Kinect," ACM Transactions on Intelligent Systems and Technology, vol.6, no.2, pp.1–20, 2015. DOI: 10.1145/2629481.
[41] F. Ordóñez and D. Roggen, "Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition," Sensors, vol.16, no.1, pp.115–140, 2016. DOI: 10.3390/s16010115.
[42] T.T. Thanh, F. Chen, K. Kotani, and B. Le, "Extraction of discriminative patterns from skeleton sequences for accurate action recognition," Fundamenta Informaticae, vol.130, no.2, pp.247–261, 2014. DOI: 10.3233/FI-2014-991.
[43] S. Sempena, N.U. Maulidevi, and P.R. Aryan, "Human action recognition using Dynamic Time Warping," Proc. ICEEI 2011, pp.1–5, 2011. DOI: 10.1109/ICEEI.2011.6021605.
[44] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol.26, no.1, pp.43–49, 1978. DOI: 10.1109/TASSP.1978.1163055.
[45] C. Nathan, "Kinect depth vs. actual distance," http://mathnathan.com/2011/02/depthvsdistance/, 2011.

Yan Li received B.S. and M.S. degrees in Computer Science from Shenyang University of Technology, China in 1998 and 2007, respectively. She is now a Lecturer in the School of Management and News Dissemination at Shenyang Sport University.

Zhijie Chu received B.S. degree in Computer Science from Liaoning Shihua University, China in 2017. She is now a Master candidate of Shenyang University of Technology, China.

Yizhong Xin received B.S. and M.S. degrees in Computer Science from Shenyang University of Technology, China in 1997 and 2004, respectively and received Ph.D. degree in Computer Science from Kochi University of Technology, Japan in 2010. In 2005–2006, he stayed at Aalen University, Germany to do research on Information Security. He is now a professor and doctoral advisor in the School of Information Science and Engineering at Shenyang University of Technology, China.