WILEY | Hindawi

Research Article

Evaluation of Motion Standard Based on Kinect Human Bone and Joint Data Acquisition

Xiaoning Zhang (ID)

School of Physical Education, Huaibei Normal University, Huaibei 235000, China

Correspondence should be addressed to Xiaoning Zhang; zhangxn@chnu.edu.cn

In order to improve human bone and joint data, we propose a method to collect data and judge the standard of motion. Kinect is a 3D somatosensory camera released by Microsoft. It has three cameras in total. The middle is a color camera, which can take color images and obtain 30 images per second; on the left is the infrared projector, which irradiates the object to form speckle. On the right is the depth camera to analyze the infrared spectrum. On both sides are two depth sensors to detect the relative position of people. On both sides of Kinect are a set of quaternion linear microphone arrays for speech recognition and filtering background noise, which can locate the sound source. There is also a base with built-in motor below, which can adjust the elevation angle. It can not only complete the collection of color images, but also measure the depth information of objects. The experimental results show that we use MSRAction3D data set and compare the same cross-validation method with other latest research methods in the figures. The highest recognition rate of this method (algorithm 10) is the second, and the lowest and average recognition rates are the highest. The improvement in the lowest recognition rate is obvious, which can show that this method has good recognition performance and better stability than other research methods. Kinect plays a relatively important role in the movement of human bone and joint data acquisition.

1. Introduction

Body motion recognition has always been a hot issue concerned by people. However, there are still many basic problems in the field of computer vision that are not perfect and have not been reasonably solved. We live in the real three-dimensional world, and the images obtained by ordinary cameras are two-dimensional, which leads to the problem of lack of information and inaccurate recognition. Therefore, we use Kinect human bone data to identify human actions, which can reduce the lack of information in data collection. According to the knowledge of three views and projection introduced earlier, if we can use two-dimensional plane human body projection in three different directions to express human action, we can more truly show human motion in the three-dimensional world [1]. This chapter uses the two-dimensional plane projection feature of human bone data. The human action recognition method is to project the three-dimensional joint points into the two-dimensional space, construct the feature vector composed of three upward joint angles, represent the combination of translation and rotation through the changes of 17 joint angles of human body, and select the method of multiclassification support vector machine to classify 20 human actions in MSRAction3d data set [2]. The Kinect somatosensory device developed by Microsoft can directly capture the body movements of patients without requiring patients to wear and operate any peripheral devices. It has a more natural and convenient human-computer interaction mode and is more suitable for the development and adoption of community and family medical rehabilitation platforms [3]. Therefore, this paper selects Kinect somatosensory equipment as the human motion sensing carrier and designs a somatosensory rehabilitation training platform based on Kinect sensor. As shown in Figure 1, the platform integrates the functions of basic motion acquisition and rehabilitation evaluation, pre-sets typical rehabilitation training movements for the shoulder and elbow joints of the left and right limbs, and collects template motion flow data and training motion flow data through somatosensory sensors. Once the data was
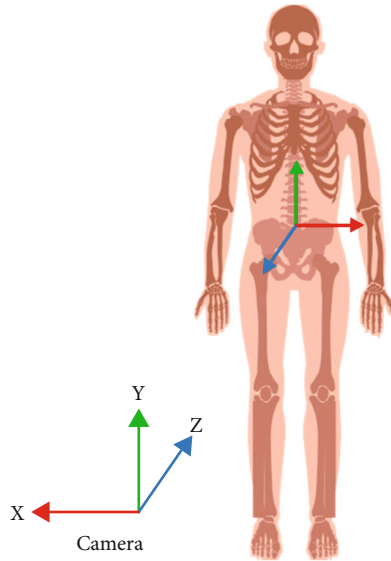
Figure 1: Data acquisition.

processed by the logic of the algorithm, the similarity of the data flows between the two groups of functions was calculated and examined by using the time dynamic control algorithm integration method and Hausdorff distance measurement algorithm. In addition, this paper examines the process of kinematic measurement of rehabilitation treatment to measure the effectiveness of rehabilitation treatment and perform interventions to clarify the algorithm and theoretical in paus [4].

2. Related Works

Since the 1980s, Internet information technology has been greatly developed. A series of human-computer interaction technologies, such as human-computer interaction and intelligent gesture recognition, have emerged one after another. Human-computer interaction means that human and computer user interface interact in some way to produce information data input and output. In real life, people can use gestures and language to complete the information input to the PC, and the computer can complete the data output through pictures, videos, and other ways, so as to realize the information interaction between people and computers. The human-computer somatosensory interaction is based on the acquisition of depth image. The acquisition of depth image mainly includes the following three ways: structure light, time of flight, and multicamera [5]. The Kinect device selected in this design mainly applies structure light detection technology. Prime sense names this depth measurement technology as light coding. Light coding is a kind of structured light technology, but the depth calculation method of light coding is different from that of other structured lights. Light coding is a depth detection algorithm independently developed by Cristache, C. M. It is a structured light technology using special depth calculation method. Compared with the conventional structured light algorithm, the light source of the light coding algorithm is the diffraction spot randomly

generated by the laser passing through the ground glass, which has high randomness and is related to the pattern and distance [6]. Of course, the prior light source calibration process cannot be omitted. Compared with the traditional structured light algorithm, light coding algorithm is a kind of three-dimensional space coding, which is controlled by the chip product ps1080 of Kulczyk, T. when calibrating the light source [7]. Moreover, the measurement accuracy is only affected by the density of the calibrated reference plane and has nothing to do with the spatial geometric position of the reference object. The first generation of Kinect somatosensory devices adopts light coding algorithm technology to collect the depth information of three-dimensional space and compare it with the previously saved speckle reference image to obtain the distance between the target object in the Kinect field of view scene and the Kinect camera. However, according to Ma et al.'s paper on Kinect depth data accuracy analysis, with the increase of the distance between the object and the sensor, the random error of depth measurement will also increase, and increase from a few millimeters to 4 meters (when the maximum range of the sensor is reached) [8, 9]. For improving this accuracy, Naufal, A., Anam, C., Widodo, C. E., and Dougherty, G. put forward a theoretical error analysis, which can clarify what factors affect the accuracy of data. Vignesh et al. studied and developed a new somatosensory rehabilitation system, which aims to improve the enthusiasm of patients' rehabilitation training and improve the efficiency of patients' rehabilitation training [10]. In this system, the motion data of patients can be recorded in real time, and the obtained motion data can be compared with the corresponding standard motion data in the database, so as to judge the recovery of patients' condition. In the system Kinerehab, patients are reflected on the interface with the image of virtual animation, which has a good and natural human-computer interaction atmosphere, enhances the interest of rehabilitation training, makes patients interested in rehabilitation training, and stimulates patients to train independently. Fajri et al. studied and developed a rehabilitation training system based on human tracking technology, which is mainly aimed at the recovery of the shoulder and elbow [11]. The system takes the position information of the patient's key joint points through the sensor, modifies the effective position information of the key joint points according to the standard action position information, and then imports the modified standard action data information into the system model. The system can effectively reduce the wrong actions of patients in rehabilitation training and avoid the damage of wrong actions to human body. At the same time, the system also has a scoring mechanism to evaluate the recovery of patients. Kim designed and studied a rehabilitation training system aimed at improving patients' balance ability. In this system, sensors are used to obtain depth image information, reshape the 3D model of human body, and conduct human-computer interaction in the form of simple games through virtual reality scenes, so as to achieve the purpose of sports rehabilitation [12]. As a new rehabilitation training platform, robot is an effective clinical intervention means to assist doctors in the reconstruction of patients' motor

function, represented by the upper limb rehabilitation training system developed by Zhou, S. and others [13]. In the early stage, the motion rehabilitation robot provided resistance and active force through the spring support of free inertia balance. Wang et al. also used the human bone detection ability of Kinect somatosensory equipment to realize the imitation control of 16 joint humanoid robot for human action [14]. Their experimental results show that the time lag of the developed control system is only a short 200 ms. In addition, Afrieda. N. has developed humanoid robots that can recognize human joints in the whole body. They also use Kinect somatosensory equipment to solve the joint angle of the robot by using analytical geometry method, which can solve the kinematics of the robot more quickly [15].

3. Method

With the support of Kinect for windows SDK, it is possible to monitor the work of one or two human bones moving into the Kinect vision by controlling the bones, receiving data from human bone marrow, and then get a triangle integration of all joints [16]. The human skeleton frame design in this article is the first generation of Kinect. The number of each joint is shown in Figure 2 to support the description of the human skeleton.

According to the sequence of Kinect bone data acquisition, number 20 human bone joint points from a to t, as shown in Figure 2. Each number corresponds to the position coordinates of a human bone joint point, which represents different parts of the human body.

Table 1 is the naming of joint points corresponding to the number. 20 joint points represent the complete structure of human bones. Through the analysis of human motion, using the coordinate data of human joint points, every two adjacent joint points are formed into a joint vector, which contains the motion information of the joint [17]. It is also easy to construct the two-dimensional joint vector of the human body. Assuming that the coordinate data of the two-dimensional space of two adjacent joint points of the human body are $U(x_1, y_1)$ and $V(x_2, y_2)$, respectively, the joint vector of the two-dimensional space composed of them is

$$UV = (x_2 - x_1, y_2 - y_1). \tag{1}$$

According to formula (1), set point $U(x_1, y_1)$ and point $V(x_2, y_2)$ to represent the joint points of the left ankle and left foot, respectively, and the vector in the formula represents the activity state of the end of the left foot in two-dimensional space, covering the details of the movement of the left leg. A total of 19 human two-dimensional joint vectors are composed of 20 human bone joint points.

If you look at different human functions in three-dimensional space from an angle, you will not be able to tell the truth of the order, because it will create a local block and covers some of the characteristics of movement. Therefore, we use three images of the human body and the projection method to reflect objects in three-dimensional space in two-dimensional planes, find the objects of motion at differ-
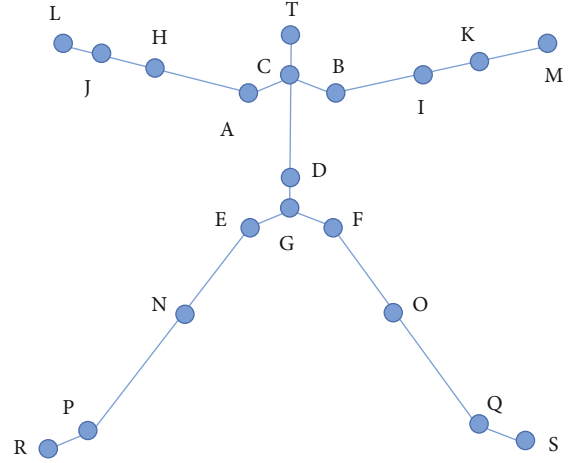

FIGURE 2: Numbering of 20 joint points.

ent plane angles, to vary class, and pay for defects. Consider only the features of motion in a triangle [18]. Because changes in the angle of the human joint affect the nature of translation and rotation, angular data can be used to determine the location of various movements. We consider the angle of the joints of the human body in three places inside the two planes, operate the plane from three different angles of view, and create the angles comfort in three different planes: the angle between the joints and the joints in the big picture and the anterior-posterior angular projection of the major plane of the human body. The angle of integration on the left watch is the projection of the joint from left to right of the left plane of the human body. The angle of inclination at the top is the approximate angle of inclination in the lower plane of the human body from the top to the bottom [19]. The angle of the selected human bone marrow is shown in Figure 3.

Figure 3 shows 17 human joint angles selected by us. According to the 19 human two-dimensional joint vectors constructed above, the size of human bone joint angle can be calculated by using cosine similarity formula, as shown in the formula

$$\cos \theta(t) = \frac{U(t).V(t)}{\|U(t)\|.\|V(t)\|}. \tag{2}$$

θ is the angle of phase t of each bone data, and $U(t)$ and $V(t)$ are both vectors of phase t. Since the collected kinect bone data is in the form of three-dimensional coordinates (x, y, z), first, reduce the dimensionality of the three-dimensional bone data to make it become the bone data on the two-dimensional projection plane of the XOY surface, and then use the two-dimensional coordinates to Calculate the size of the joint angle. For example, during T, calculate the angle of inclination of the joint from the left wrist in both planes. The joint of the left wrist at time t is 11 $(x(t), y(t))$, the joint of the left elbow of t is 22 $(x(t)), y$

TABLE 1: Naming of 20 individual joint points.

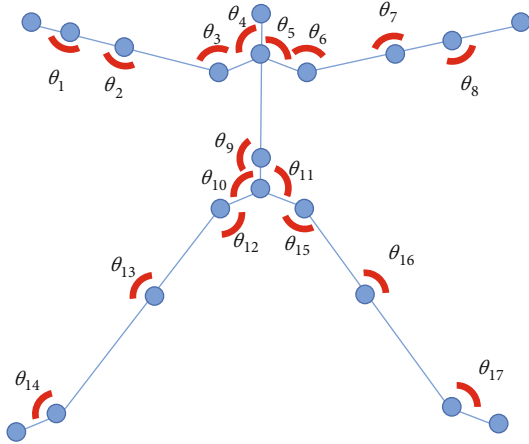| Number	Name	Number	Name	Number	Name	Number	Name
A	Left shoulder	F	Right hip	K	Right wrist	P	Left ankle
B	Right shoulder	G	Hips	L	Left hand	Q	Right ankle
C	Neck	H	Left elbow	M	Right hand	R	Left foot
D	Spine	I	Right elbow	N	Left knee	S	Right foot
E	Left hip	J	Left wrist	O	Right knee	T	Head

FIGURE 3: 17 joint angles of human body.

$(t))$, and the left wrist of t is 33 $(x(t), y(t))$ are shown as

$$JL^{\{2\}}(t) = \left(x_1^{(t)}.y_1^{(t)}\right) - \left(x_3^{(t)}.y_3^{(t)}\right), \tag{3}$$

$$JH^{\{2\}}(t) = \left(x_2^{(t)}.y_2^{(t)}\right) - \left(x_3^{(t)}.y_3^{(t)}\right), \tag{4}$$

$$\cos \theta^{\{2\}}(t) = \frac{JL^{\{2\}}(t).JH^{\{2\}}(t)}{\left\|JL^{\{2\}}(t)\right\|.\left\|JH^{\{2\}}(t)\right\|}. \tag{5}$$

4. Experimental Results and Discussion

On this basis, aiming at the problem of low accuracy and stability of human action recognition in complex environment with high noise, a method of human action recognition based on hierarchical feature fusion was proposed, which divided different parts of human body according to the composition of human body structure system. The layered strategy is adopted, which is conducive to the decomposition of complex human movements. Firstly, according to the bone joint coordinates obtained by Kinect, the features of human joint angles in two-dimensional space are extracted, and the actions are roughly classified by the support vector machine (SVM) method. Then, the body vector, angular velocity, and acceleration in 3d space were extracted, and the movements were classified by HMM.

With the knowledge of human models, the human body can be divided into five sections. Part I: physical. The body includes the head, neck, back, and hips. Part II: left hand. The left hand includes the forearm, left wrist, left elbow, and left shoulder. Part III: right hand. The right hand

includes the arm, wrist, elbow, and shoulder. Part IV: left foot. The left leg includes the left foot, the ankle, the left knee, and the left thigh. Section V: right. Right foot includes the right foot, right ankle, right knee, and right foot. The body is an essential part of the human body. The motion feature information of some waists of the human body comes from this part of the joint points, and the motion feature information of the hands and feet comes from the joint points of the limbs [20].

By dividing the structure of the human body, it is possible to combine these five components to represent some of the most important functions of the human body. Therefore, in terms of proportions, this paper adopts a hierarchical concept. In step 1, first, divide the above five functions into categories with the same type of combination. For example, the two hands are a combination of the second and third sections. This is a rough proportion. The second layer redivides the functions of the same type of combination and filters the specific functions, which are the details of the process. We take the vector of the joint angle, including the projection of 17 joints on both sides of the three planes included in Chapter 3, according to the characteristics of the first rough decision and distribution of human activities. When distinguishing human actions with the same combination mode, we extract features from kinematics theory. A complete human action can be divided into main action and auxiliary action. The main action reflects the global state of the motion mode, and the auxiliary action reflects the local state of the motion mode. Only by combining the characteristics of the main and auxiliary actions can this action be expressed more accurately [21]. For the five parts of the human body divided above, we construct their limb vectors in three-dimensional space, respectively, shown as

$$GT^{\{3\}}{}_{(t)} = T^{\{3\}}{}_{(t)} - G^{\{3\}}{}_{(t)}, \tag{6}$$

$$AJ^{\{3\}}{}_{(t)} = J^{\{3\}}{}_{(t)} - A^{\{3\}}{}_{(t)}, \tag{7}$$

$$BK^{\{3\}}{}_{(t)} = K^{\{3\}}{}_{(t)} - B^{\{3\}}{}_{(t)}, \tag{8}$$

$$EP^{\{3\}}{}_{(t)} = P^{\{3\}}{}_{(t)} - E^{\{3\}}{}_{(t)}, \tag{9}$$

$$FQ^{\{3\}}{}_{(t)} = Q^{\{3\}}{}_{(t)} - F^{\{3\}}{}_{(t)}. \tag{10}$$

$\{3\}$ represents in the three-dimensional space, t represents a certain time, and the joint points that are prone to drift at the end points of hands and feet are temporarily rounded off. Finally, the limb vectors in the three-dimensional space at all times of the five parts are

represented by $GT^{\{3\}}$, $AJ^{\{3\}}$, $BK^{\{3\}}$, $EP^{\{3\}}$, and $FQ^{\{3\}}$, respectively. According to the different contribution to the expression of human action, two joint angles are selected from each part, which are called the main action joint angle. The torso part selects angle θ_4 and angle θ_9, the left arm part selects angle θ_3 and angle θ_2, the right arm part selects angle θ_6 and angle θ_7, the left leg part selects angle θ_{12} and angle θ_{13}, the right leg part selects angle θ_{15} and angle θ_{16}, and the angular velocity of the joint angle is shown as

$$\omega(t) = \theta(t + 1) - \theta(t). \tag{11}$$

The human action sequence is continuous and changes with time. The change of joint angle before and after forms the value of angular velocity ω. the limb vector and the angular velocity of active joint angle are the characteristics of active action, representing the overall movement of human limbs and trunk [22]. The bending of human limbs and trunk is reflected by the change of the distance between joint points. The human body is projected into the YOZ side plane from the left view direction. The distance between joint points of five bones is shown as

$$d_{GT}(y, z) = \sqrt{\left(y_G^{(t)} - y_T^{(t)}\right)^2 + \left(Z_G^{(T)} - Z_T^{(t)}\right)}, \tag{12}$$

$$d_{GT}(y, s) = \sqrt{\left(y_G^{(t)} - y_T^{(t)}\right)^2 + \left(s_G^{(t)} - s_T^{(t)}\right)^2},$$

$$d_{CL}(y, s) = \sqrt{\left(y_C^{(t)} - y_L^{(t)}\right)^2 + \left(s_C^{(t)} - s_L^{(t)}\right)^2},$$

$$d_{CM}(y, s) = \sqrt{\left(y_C^{(t)} - y_M^{(t)}\right)^2 + \left(s_C^{(t)} - s_M^{(t)}\right)^2}, \tag{13}$$

$$d_{GN}(y, s) = \sqrt{\left(y_G^{(t)} - y_N^{(t)}\right)^2 + \left(s_G^{(t)} - s_N^{(t)}\right)^2},$$

$$d_{GS}(y, s) = \sqrt{\left(y_G^{(t)} - y_S^{(t)}\right)^2 + \left(s_G^{(t)} - s_S^{(t)}\right)^2},$$

where $d(y, z)$ represents the Euclidean distance between the two joints in the lateral plane and t represents the distance between the head and end joint points of the five parts of the human body at a certain time, reflecting the bending of the limbs and trunk in motion. The change of the distance before and after the time forms the speed v, and the acceleration is a physical quantity describing the speed of the change of human motion, as shown in

$$v(t) = d(t + 1) - d(t), \\ \partial(t) = [v(t)]'. \tag{14}$$

The distance between five joints in the lateral plane and the acceleration of motion are regarded as the characteristics of auxiliary action. The accelerations of the five parts are ∂_1, ∂_2, ∂_3, ∂_4, and ∂_5. The characteristics of main action and auxiliary action together constitute the characteristics of the second level fine classification of human action. Because

everyone's height and arm length are different, even if two people make the same action posture, there will be some errors. In order to eliminate individual differences, divide the items in the formula by the shoulder width d_{AB} and the mean value $\bar{d}$ of the Euclidean distance between joints in the plane of YOZ side, as shown in

$$d_{AE}(x, y, s) = \sqrt{(x_A - x_E)^2 + (y_A - y_E)^2 + (s_A - s_E)^2}, \tag{15}$$

$$\bar{d}(y, s) = \frac{\sum_{t=1}^{n} d(t)}{n}, \tag{16}$$

where d_{AB} represents the width of each person's shoulder and D represents the average distance between the five major joints of the human body at all times in the YOZ side plane. The feature vector of the final rough classification is expressed as $[\theta_1^{\{2\}}, \theta_2^{\{2\}}, \theta_3^{\{2\}} \cdots \theta_{17}^{\{2\}}]$. The matrix composed of feature vectors of fine classification is expressed as

$$\begin{bmatrix} GT^{\{3\}} & \omega_4 & \omega_9 & \partial_1 \\ AJ^{\{3\}} & \omega_3 & \omega_2 & \partial_2 \\ BK^{\{3\}} & \omega_6 & \omega_7 & \partial_3 \\ EP^{\{3\}} & \omega_{12} & \omega_{13} & \partial_4 \\ FQ^{\{3\}} & \omega_{15} & \omega_{16} & \partial_5 \end{bmatrix}. \tag{17}$$

Each row in the matrix represents a set of eigenvectors.

In this system, we will compare the similarity between the standard action data sequence template and the action data sequence collected in real time, so as to achieve the effect of rehabilitation evaluation. In reality, the time taken by the patient to complete a set of rehabilitation exercises is usually inconsistent with the time taken by the standard action template. Some patients with serious injuries take several times as long as the standard action template. For two sets of motion data sequences with different lengths, the operation of corresponding time points alone cannot meet the accuracy requirements of the current system. DTW, namely dynamic time warping algorithm, is a widely used speech recognition algorithm. At first, it is to solve the error caused by different tones in speech recognition. Compared with other algorithms, dynamic time warping algorithm expresses the relationship between two sequences with inconsistent length through the time warping function under certain conditions. At present, it is widely used in gesture recognition, language recognition, and other fields. The central idea of DTW is to extend or shorten two data sequences with different time lengths to ensure the consistency of the length of the two data sequences to be tested and select an appropriate path from the constructed distance matrix to minimize the sum of the distances obtained by the two sequences. In the recognition of human action sequence, it is actually to find the minimum distortion distance between the current sequence and the standard template.

It is assumed that the two sequences of the standard template and the template to be tested are R and T, and the length is m frames and N frames, respectively [23]

shown as

$$R = (R_1, R_2 \cdots R_{m-1}, R_m)\{R_m, m = 1, 2, \cdots M\},$$
$$T = (T_1, T_2 \cdots T_{n-1}, T_n)\{T_n, n = 1, 2, \cdots N\}. \tag{18}$$

In general, for two sequences with different lengths, it is necessary to construct a $m \times n$ matrix grid, and the matrix element (i, j) represents the distance $d(R_i, T_j)$ between R_i and T_j. The matrix is shown as

$$D = \begin{bmatrix} d(R_1, T_1) & d(R_1, T_2) & \cdots & d(R_1, T_n) \\ d(R_2, T_1) & d(R_2, T_2) & \cdots & d(R_2, T_n) \\ \vdots & \vdots & & \vdots \\ d(R_m, T_1) & d(R_m, T_2) & \cdots & d(R_m, T_n) \end{bmatrix}. \tag{19}$$

DTW algorithm is to find an optimal path through the grid matrix, and the grid points passed by this path are the points that need to be processed and calculated in the two sequences. Let this path be the regular path W, then the K-th element of W is defined as $W_k = (I, j)_k$. That is, the regular path is shown as

$$W = \{w_1, w_2, w_3 \cdots w_k\}(\max(m, n) \le K \le m + n - 1). \tag{20}$$

According to the constraints of continuity and monotonicity, the path of each grid point has only three directions. If the grid point is $(i.j)$ at this time, the path of the next grid point is only $(i + 1, j + 1), (i, j + 1), (i + 1, j)$. There are many regular paths that meet this condition, but the formula of minimum distortion distance is shown as

$$\text{DTW}(R, T) = \min \left\{ \frac{\sqrt{\sum_{k=1}^{k} w_k}}{k} \right\}. \tag{21}$$

The path meeting the minimum distortion distance needs to meet three conditions:

(1) The path must meet the requirements from the beginning $W(1, 1)$ of the sequence to the end $W_k(m, n)$ of the sequence

(2) The path needs to meet the sequence of time, namely $m + 1 \ge m \& n + \ge n$

(3) The path selection needs to meet the monotonicity, and m and n increase by 0 or 1 in turn. That is, if $W_{k-1} = (i, j)$, W_k has three choices $(i + 1, j + 1)$, $(i + 1, j)$, and $(i, j + 1)$

The path with the smallest sum of cumulative distances of adjacent elements is the optimal regular path. According to the above formula, the DTW cumulative distance formula

can be deduced as

$$D(i, j) = d(r_i, t_j) + \min \{D(i - 1, j - 1), D(i - 1, j), D(i, j - 1)\}. \tag{22}$$

A Kionix KXSD9 three-axis accelerometer is included in the Kinect internal structure to prevent the error caused by Kinect being placed on an uneven plane and improve the stability of Kinect's depth image acquisition. Among them, the camera in Kinect can be adjusted according to the needs of users. The moving touch drive motor in Kinect can adjust the elevation of the camera to match the user's position change. Kinect also has a focusing system. If the user exceeds the field of vision, Kinect can automatically drive the base motor to adjust vertically by ±28° [24]. The field of view of Kinect imaging is 43.5° vertical and 57.5° horizontal.

The maximum distance range that Kinect sensor can track and recognize is 0.8 m to 4 m, but in practice, in order to ensure accurate data, the recognition distance is 1.2 m to 3.5 m as shown in Table 2:

Kinect is a "pipeline" system functional architecture, as shown in Figure 4. The original sensor data stream includes depth data stream, color image data stream, and audio data stream. Researchers can directly develop applications based on the original data stream information obtained by Kinect SDK.

Kinect uses depth measurement technology in acquiring depth images, which is also known as light coding. That is to mark and code the space to be measured with light source. This technology belongs to structured light technology. Compared with other structured light technologies, the depth calculation method of this technology is different from other technologies. The light source of light coding is called "laser speckle." Laser speckle is the diffraction spot image formed after the light source irradiates the nonsmooth object or passes through the ground glass. These speckle images will constantly change with the distance of the light source, with a high degree of random variation. Generally speaking, the speckle images at any two places in the spatial environment are inconsistent. All speckle images in the whole space are saved and recorded, and then the structured light is irradiated into the space for marking. At this time, put a new object into the space and only view the speckle image generated by the object to obtain the specific position of the object [25].

Today, with the rapid development of science and technology, somatosensory technology has gradually matured. At present, in addition to being mainly used in entertainment games, there are also various human-computer interaction systems developed by many researchers. Therefore, at present, many companies at home and abroad are carrying out research and development in the field of motion sensing, designing, and manufacturing their own motion sensing equipment. Three of the most popular products on the market are the following.

Xtion, a motion sensing device designed and produced by ASUS, mainly uses the Structure Light detection technology to obtain depth images. The Structure Light is used to

TABLE 2: Effective sight distance of Kinect sensor.

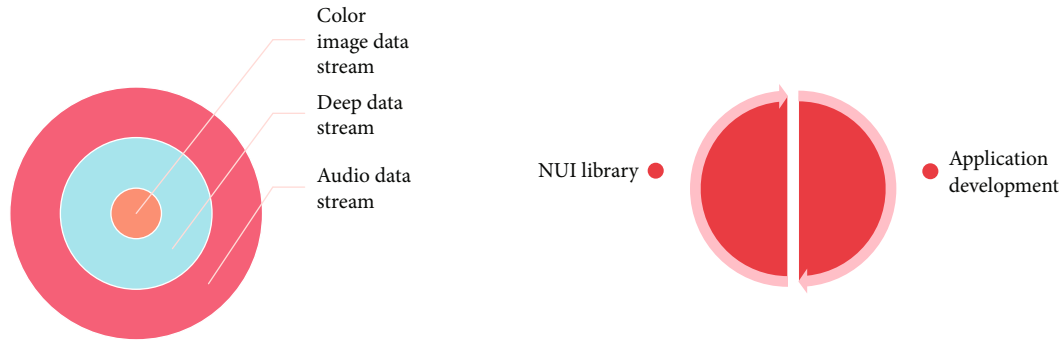| Sensor characteristics	Effective sight distance
Color and depth	Default mode (0.8-4.0), close range mode (0.4-3.5)
Bone tracking	Default mode (0.8-4.0), close range mode (0.4-3.5)

FIGURE 4: Interaction between application and Kinect sensor.

obtain spatial data and calculate other data, such as depth image and object skeleton. In the working process of the equipment, an infrared laser emitter emits the encoded near infrared light source. After the infrared light is reflected, it is captured by the infrared camera, and then the corresponding feature coding correspondence is calculated to determine the depth information.

Leap-motion, a motion-sensing device developed by Leap (USA), controls a person's hands by detecting their movements. Leap-motion uses multicamera detection technology. It has two infrared cameras and uses powerful data processing chips to quickly process the image data to detect the hand movements of the target. The leap-motion sensor's field of view detects the spatial shape of an inverted quad-pyramid with effective detection range of 25 mm to 600 mm. The product also has an open SDK (Software Development Kit), which can meet the needs of developers to develop and research on Windows, Linux, and Mac, the three mainstream operating system platforms.

Kinect for Xbox 360 is Microsoft's external motion sensing device that was officially unveiled on June 2, 2009. Structure Light detection technology is mainly used in obtaining depth images, which is the same as Xtion's principle. In terms of open source development, Microsoft has also designed SDK tools containing rich API interfaces, so developers can combine various languages to develop Kinect programming.

By comparing the above three kinds of motion sensing devices, it can be found that Kinect for Xbox 360 motion sensing devices is more convenient and high-precision for controlling people. Moreover, Microsoft also provides a large number of API interfaces in Kinect SDK, which is more advanced, so that Kinect can not only get the original depth data in the work, but also get the target object bone node data. Therefore, Kinect for Xbox 360 is selected as the somatosensory device in this design.

In order to obtain the coordinate information of human key joint points in real time to judge the similarity algorithm in the next step, Kinect sensor is used to track human joint points to obtain the motion data sequence in real time. The human skeleton is composed of human skeleton joint points, and each joint point has a relative position and direction. Kinect bone tracking module first detects the human body through the depth image information technology in Kinect and then corrects the human body posture. When the correction is successful, the measured human body can be tracked in real time, and the relevant bone information can be obtained in real time. When the correction fails, it will enter, while the cycle processes to calibrate again.

In the example provided by Kinect, the programs of two technologies of depth image data and bone tracking are provided, respectively, but there is no relevant program of data extraction. After learning the relevant application software, the two technologies of depth image and bone tracking are combined to successfully record and extract the data information of key bone joint points of human body. The base note is as follows: firstly, initialize the device environment, create new objects and create new user generators to store relevant data information, and facilitate subsequent calls. Secondly, register relevant callback functions and calibrate the skeleton. The functions to be called include the generation of new users and the detection of bone posture. Finally, bone tracking is performed, and relevant bone information is updated and read in real time. The flow chart of obtaining human bone information is shown in Figure 5.

The hardware setting of the simulation environment for the experiment in this chapter is Intel(R) Core(TM)i5-4210M CPU @2.60GHZ, 8GB RAM. Software settings are Windows 10, 64 bit operating system, and MATLAB R2017B. The MSRAction3D public data set is used. The data set contains 567 samples. Each action category is repeated by 10 different male and female subjects for 2~3 times. We take
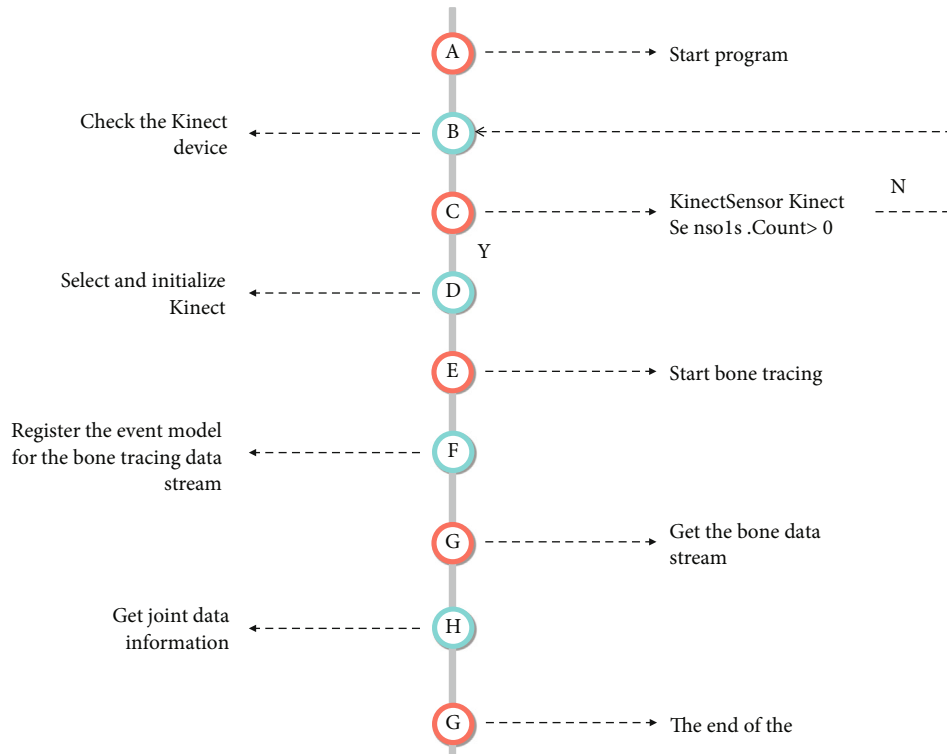
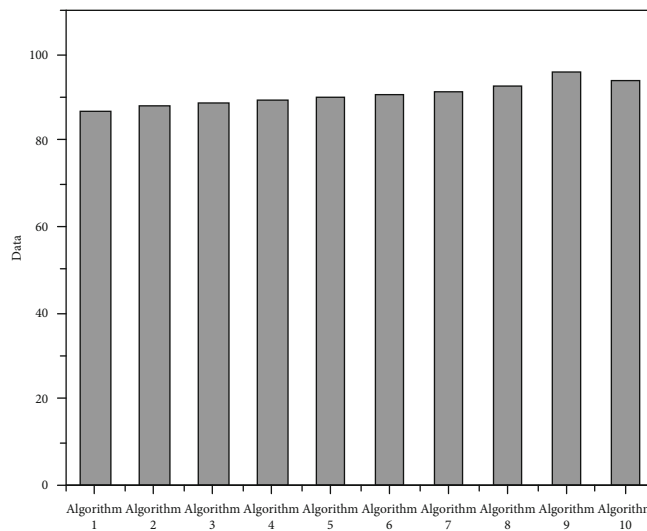FIGURE 5: Flow chart of obtaining human body data information.

FIGURE 6: The highest recognition rate of this method and other methods.

547 samples for the test and remove 20 missing data samples. In the previous chapter, we directly classified 20 different human actions at one time. This chapter first roughly classifies these 20 actions into seven categories and then subdivides them: class I: single arm action; the second category: the movement of both arms; category III: single leg movements; the fourth category: trunk movement; category 5: the movement of both arms and legs; category 6: trunk plus arms; and category 7: trunk plus legs plus one arm. Because the number of subjects with different height and weight is

not included in the study set, only the proportion of male and female subjects is included in the study set. We use MSRAction3D data set and compare the same cross-validation method with other latest research methods in Figures 6, 7, and 8. The highest recognition rate of this method (algorithm 10) is the second, and the lowest and average recognition rates are the highest. The improvement in the lowest recognition rate is obvious, which can show that this method has good recognition performance and better stability than other research methods.
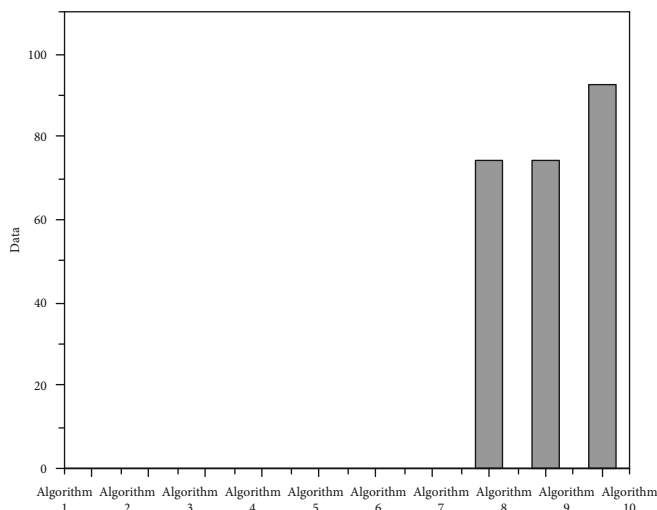
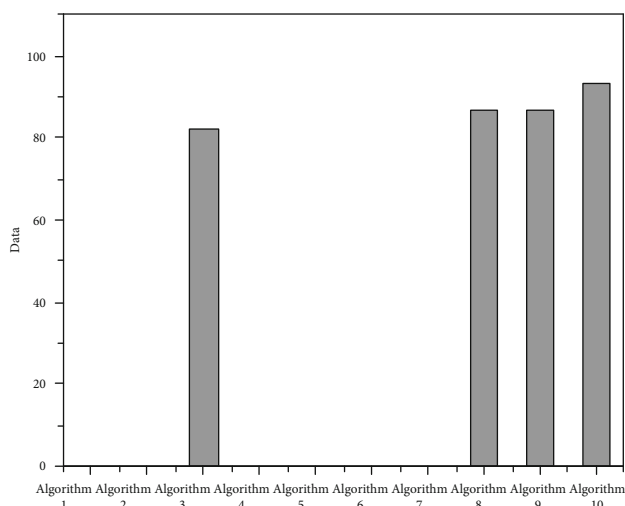FIGURE 7: Minimum recognition rate of this method and other methods.

FIGURE 8: Average recognition rate of this method and other methods.

5. Conclusion

Based on Kinect human skeleton and joint data collection, the expression of human motion is completed through limbs, and all human bones together constitute human limbs. Rehabilitation exercise training is mainly to help people with movement disorders gradually recover their limb motor function through existing medical technologies and means. The traditional training methods are not suitable for patients to carry out rehabilitation training in family or community environment, regardless of the use cost or the complexity of operation. At the same time, the complex and cumbersome process of sports data acquisition often affects the rehabilitation doctors' judgment of patients' limb recovery. Therefore, this paper studies and designs a set of motion obstacle auxiliary evaluation training system based on Kinect. The system combines Kinect somatosensory sensor and virtual reality technology. By capturing and col-

lecting human motion data information in real time and calculating relevant algorithms, patients can carry out rehabilitation training and complete the evaluation of rehabilitation training actions independently, so as to improve the rehabilitation efficiency of patients.

Data Availability

The data sets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

It is declared by the author that this article is free of conflict of interest.

Acknowledgments

References

[1] R. Li, "Evaluation and simulation of medical sports health equipment multimedia image based on information asymmetry theory," *Multimedia Tools and Applications*, vol. 79, no. 15-16, pp. 9957–9976, 2020.

[2] A. N. Smirnov, N. V. Ababkov, V. V. Murav'Ev, and S. V. Fol'Mer, "Criteria for the evaluation of the technical state of the long-lived metal of HPP equipment based on acoustic structuroscopy," *Russian Journal of Nondestructive Testing*, vol. 51, no. 2, pp. 94–100, 2015.

[3] S. J. Seo, J. H. Lim, Y. Lee, C. W. Bark, and Y. G. Kim, "High-throughput analysis of new bone formation and bone substitutes after maxillary sinus floor elevation using synchrotron radiation micro-computed tomography," *Journal of*

Nanoscience and Nanotechnology, vol. 19, no. 2, pp. 680–686, 2019.

[4] M. Meglioli, A. Naveau, G. M. Macaluso, and S. Catros, "3D printed bone models in oral and cranio-maxillofacial surgery: a systematic review," *3D Printing in Medicine*, vol. 6, no. 1, p. 30, 2020.

[5] H. Lung, J. T. Hsu, Y. J. Wu, and H. L. Huang, "Biomechanical effects of diameters of implant body and implant platform in bone strain around an immediately loaded dental implant with platform switching concept," *Applied Sciences*, vol. 9, no. 10, p. 1998, 2019.

[6] C. M. Cristache, I. Tudor, L. Moraru, G. Cristache, and M. Burlibasa, "Digital workflow in maxillofacial prosthodontics—an update on defect data acquisition, editing and design using open-source and commercial available software," *Applied Sciences*, vol. 11, no. 3, p. 973, 2021.

[7] L. Zhang, "Research on the evaluation of sports events based on the concept of green environmental protection," *IOP Conference Series Earth and Environmental Science*, vol. 651, no. 4, p. 042028, 2021.

[8] Y. Ma, D. Liu, and L. Cai, "Deep learning-based upper limb functional assessment using a single kinect v2 sensor," *Sensors*, vol. 20, no. 7, p. 1903, 2020.

[9] A. Naufal, C. Anam, C. E. Widodo, and G. Dougherty, "Automate the calculation of human body height using a MATLAB-based Kinect camera for estimating body size: a pilot study," *AIP Conference Proceedings*, vol. 2346, no. 1, p. 040002, 2021.

[10] T. Vignesh, R. Mahadevan, and S. P. Harish, "Adherence to Kinect adventure games in individuals with Parkinson's disease – an experimental study from Mysore, India," *Journal of Evolution of Medical and Dental Sciences*, vol. 10, no. 33, pp. 2754–2760, 2021.

[11] B. R. Fajri, A. D. Samala, and F. Ranuharja, "Perancangan media interaktif gerak tari topeng patih pada wayang topeng malangan menggunakan sensor kinect," *INVOTEK Jurnal Inovasi Vokasional dan Teknologi*, vol. 20, no. 2, pp. 75–88, 2020.

[12] H. N. Kim, "Ambient intelligence: placement of Kinect sensors in the home of older adults with visual disabilities," *Technology and Disability*, vol. 32, no. 4, pp. 271–283, 2020.

[13] S. Zhou, F. Kang, W. Li, J. Kan, and Y. Zheng, "Point cloud registration for agriculture and forestry crops based on calibration balls using Kinect v2," *International Journal of Agricultural and Biological Engineering*, vol. 13, no. 1, pp. 198–205, 2020.

[14] Y. Wang, K. Junghyun, and S. G. Jahng, "Interactive research based on Kinect - focused on the work <Dream>," *TECHART Journal of Arts and Imaging Science*, vol. 7, no. 3, pp. 17–21, 2020.

[15] A. Procházka, O. Vyšata, M. Vališ, O. Ťupa, M. Schätz, and V. Mařík, "Use of the image and depth sensors of the Microsoft Kinect for the detection of gait disorders," *Neural Computing and Applications*, vol. 26, no. 7, pp. 1621–1629, 2015.

[16] Z. Feyziolu, S. Diner, A. Akan, and Z. C. Algun, "Is Xbox 360 Kinect-based virtual reality training as effective as standard physiotherapy in patients undergoing breast cancer surgery?," *Supportive Care in Cancer*, vol. 28, no. 9, pp. 4295–4303, 2020.

[17] Y. Chen, B. Zhang, J. Zhou, and K. Wang, "Real-time 3D unstructured environment reconstruction utilizing VR and Kinect-based immersive teleoperation for agricultural field robots," *Computers and Electronics in Agriculture*, vol. 175, no. 2020, p. 105579, 2020.

[18] N. Takeshima, T. Kohama, M. Kusunoki et al., "Development of simple, objective chair-standing assessment of physical function in older individuals using a Kinect^TM Sensor," *The Journal of Frailty & Aging*, vol. 8, no. 4, pp. 186–191, 2019.

[19] L. E. N. Díaz, D. Pool, and M. Torres, "Development of an educational software as support for learning English and Mayan languages through the use of the Kinect sensor," *IARJSET*, vol. 6, no. 11, pp. 57–63, 2019.

[20] M. Mansoor, R. Amin, Z. Mustafa, S. Sengan, H. Aldabbas, and M. T. Alharbi, "A machine learning approach for non-invasive fall detection using Kinect," *Multimedia Tools and Applications*, vol. 81, no. 11, pp. 15491–15519, 2022.

[21] L. Wang, J. Liu, and J. Lan, "Feature evaluation of upper limb exercise rehabilitation interactive system based on Kinect," *IEEE Access*, vol. 7, pp. 165985–165996, 2019.

[22] R. Lozada-Yánez, N. La-Serna-Palomino, and F. Molina-Granja, "Augmented reality and MS-Kinect in the learning of basic mathematics: KARMLS case," *International Education Studies*, vol. 12, no. 9, p. 54, 2019.

[23] Y. S. Kang, S. T. Lu, C. C. Chiu, C. C. Tu, and Y. J. Chang, "Using Kinect v2 combined with Unity3D to design an agility training game," *International Journal of Computer Theory and Engineering*, vol. 11, no. 3, pp. 56–60, 2019.

[24] R. Bianchi, B. Yyelland, J. Yang, and M. McHarg, "Avatar Kinect: drama in the virtual classroom among L2 learners of English," *The Qualitative Report*, vol. 24, no. 13, pp. 6–6, 2019.

[25] M.-H. Nguyen, C.-C. Hsiao, W.-H. Cheng, and C.-C. Huang, "Practical 3D human skeleton tracking based on multi-view and multi-Kinect fusion," *Multimedia Systems*, vol. 28, no. 2, pp. 529–552, 2022.