Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 3876487, 11 pages
https://doi.org/10.1155/2022/3876487

Research Article

Pattern Recognition of Wushu Routine Action Decomposition Process Based on Kinect

Chenxing Cao,[1] Bai Shan,[2] and Haiyan Zhang[3]

[1]*Guangdong Nanhua Vocational College of Industry and Commerce, Sports Art Department, Guangzhou, China*
[2]*Department of Information Engineering, Hebei Agricultural University, Qinhuangdao 066000, Hebei, China*
[3]*Guangdong Nanhua Vocational College of Industry and Commerce, Library, Guangzhou, China*

Correspondence should be addressed to Haiyan Zhang; 1230441236@cjlu.edu.cn

Received 9 June 2022; Revised 10 July 2022; Accepted 13 July 2022; Published 27 August 2022

Academic Editor: Baiyuan Ding

Human action recognition is a hotspot in the fields of computer vision and pattern recognition. Human action recognition technology has created huge social value and considerable economic value for the society. Meeting people's needs and understanding people's expressions are the current research focus. Aiming at the problem that the movement cannot be continuously identified and due to a lack of detailed features in the action decomposition pattern recognition in the traditional Wushu routine decomposition process, it is proposed to use Kinect technology to identify the Wushu routine movement decomposition process in the Wushu routine movement decomposition process. This paper analyzes the principle of skeleton tracking and skeleton extraction performed by the Kinect human sensor and uses the Kinect sensor with the Visual Studio 2015 development platform to collect and process the skeleton data of limb movements and defines eight static limb motion samples and four dynamic limbs. The study uses a deep learning neural network algorithm to train and identify the established database of static body movements and uses the same template matching algorithm and K-NN. The recognition effects of the algorithms were compared and analyzed, and it was concluded that the static body motion recognition rates of the three algorithms were all above 90%. In this paper, recognition experiments are carried out on the MSR action 3D database. The influence of different integrated decision-making methods on the recognition results is further discussed and analyzed, and the average method integrated decision-making, which is most suitable for the algorithm model in this paper, is proposed. The results show that the recognition accuracy of the algorithm reaches 98.1%, which proves the feasibility of the preprocessing algorithm.

1. Introduction

Machines can accurately identify and respond to human body movements, which is a new way of communication between robots and humans [1]. In some relatively harsh environments, the data information and expression effects brought by action recognition are much higher than speech recognition [2]. Using machines to recognize images can be used in various aspects. Figure 1 also shows the development process of image recognition technology. The Kinect camera launched by Microsoft can not only extract human skeleton point data but is also not affected by the lighting environment, which promotes the process of human action recognition [3].

Limb motion recognition refers to the process in which the robot can read and analyze the current body motion of the human body and respond correctly. There are many disciplines related to body recognition [4]. What is a body movement? A body movement is the displacement sequence formed by the position of the same body part of the human body changing with time in space [5]. A well-known psychologist once conducted a survey and researched on "the way of human information communication and expression" and found that only 7% of the information is communicated directly through language, and 38% of it is communicated through the difference in intonation and speed of speech [6]. Convey information, and the proportion of people who convey information through human facial expressions and

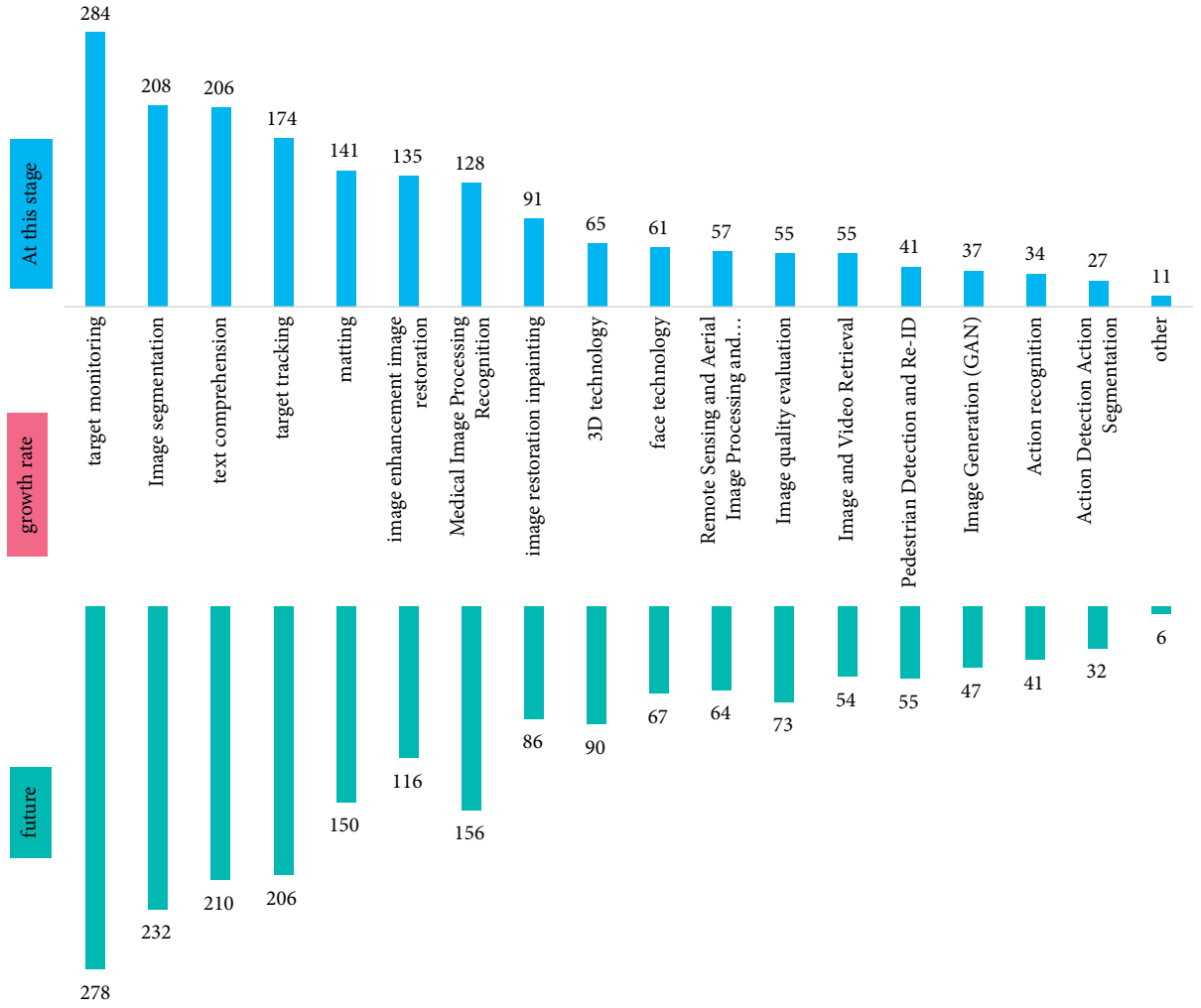

Figure 1: The development history of image recognition technology.

body movements is as high as 55% [7]. It can be seen that in life, the communication of information is more conveyed through human body language. In many cases, although people can disguise themselves through words, it is difficult to disguise their subconscious body movements. Therefore, it is very necessary to study the recognition of human body movements [8]. At present, body motion recognition technology has high research value and application value in this field [9].

In the process of decomposing martial arts routines, traditional action pattern recognition methods use the form of RGB images to identify, and the identification content includes action information and martial arts routines, but traditional methods are very sensitive to changes in light, viewing angle, frequency of actions, and other factors [10]. When the chain structure is used, the movement characteristics of Wushu routines are often manifested in the change of joint positions. Due to the limitation of the recognition method, the continuous movement cannot be recognized to a large extent [11]. The traditional recognition method is very difficult to recognize the human action in the low-dimensional motion space; the effectiveness of pattern recognition is less than 45%. Although complex dynamic data are required to describe human motion in low-dimensional motion space, martial arts routines are limited by the influence of kinematic mechanics [12]. This paper proposes a decomposition process of martial arts routines. It can effectively solve the limitations of angle, light perception, and other factors in traditional pattern recognition methods. Using the Kinect technology as the expression method of human body details features, it can effectively change the details of changes in the human body [13]. Maintain a high sensitivity and reoptimize the embedding method to achieve the continuity of low-dimensional motion space mapping, so as to solve the problem that the traditional method cannot continuously identify [14]. Therefore, this paper uses the human body sensor Kinect with the Visual Studio 2015 development platform for limbs. The collection and processing of action data, the use of deep learning neural network algorithms to train and recognize body movements, the design of a remote-control system for robots based on martial arts action recognition, the combination of martial arts action recognition and mobile robots, and finally the realization of the real-time control of remote robots [15].

2. Methodology

2.1. Introduction of the Kinect Device. Human action recognition is generally based on pictures or videos to carry out related research, aiming to classify various actions that occur in them [16]. With the successive advent of depth sensors represented by Kinect, classification through bone data and depth image database has become an important means of human behavior recognition research. The current into three processes: feature extraction, feature fusion, and behavior classification [17]. Liu et al. systematically studied the modeling and online updating methods of two-dimensional principal component analysis. The two-dimensional principal component analysis algorithm is applied to the foreground segmentation of overlapping blocks. Then, using the research methods of background difference, optical flow, and frame difference, the research model of the adaptive recognition method of martial arts decomposition VR image based on feature extraction is established [18]. As a successful image classification technology, deep learning has significant advantages in feature extraction. It can automatically extract features from preprocessed image data and has better robustness than artificially constructed feature vectors [19]. Therefore, deep learning-related technologies have been concerned by some scholars and gradually applied to human behavior recognition [20].

Kinect is a game somatosensory peripheral launched by Microsoft. It can enable players to get rid of interface devices such as mouse and keyboards and control the application by the player's body to achieve a natural human-computer interaction experience. This experience is due to Kinect. The principle of depth imaging and bone tracking technology. Kinect has its own development tools, which can not only obtain image and audio data but also develop related programs. Based on its powerful functions, Kinect has also been innovatively used in medicine, scientific research, monitoring, and other fields. There have been two generations of Kinect since its release: Kinect V1 and Kinect V2, which were released in 2012 and 2014, respectively. Compared with the Kinect V1, the properties of the Kinect V2 have been further improved, and some configuration functions have been enhanced.

The first camera on the left is an infrared emitting device, which is used to emit infrared light; the second camera in the middle is an RGB camera, which is used to capture color images; the third camera on the right is an infrared depth camera, which is used to receive infrared light reflection signals, and for depth imaging; the microphone array at the bottom is a plurality of microphone holes, which are used to collect sound signals within a specific distance and perform noise reduction processing. Therefore, the Kinect device can obtain three raw data streams, namely, color RGB image, depth data, and audio data. The establishment of a human behavior database promotes the research of behavior recognition algorithm, which can provide a common platform for many scholars to verify the recognition rate of different algorithms, which will undoubtedly become an important basis for comparing the performance of each algorithm. Kinect devices have established four common human behavior databases, including the MSR Action 3D database, MSR Daily Activity 3D database, UTD-MHAD database, and NTU RGB-D database. Among them, the UTD-MHAD database contains four-modal data: inertial sensor data, depth image sequence, RGB video, and skeletal motion sequence. These advantages of depth image and bone data enable researchers and developers of recognition system to pay more attention to the research of pattern recognition algorithm. There is no need to put too much energy into some front-end tasks such as image preprocessing. This greatly reduces the development time and difficulty of human motion recognition system. All these advantages make depth data have more application space in motion recognition than RGB images. Because the actions in this database have different execution speeds and large intraclass differences, there is a great deal of recognition difficulty. The four-modal data on the UTD-MHAD database are shown in Figure 2.

For the replacement of Kinect, the appearance of the two generations of products has not changed much, but the hardware performance has been improved. For example, the resolution of the RGB camera increased from $640 * 480$ to $1920 * 1080$, the resolution of the infrared depth camera increased from $320 * 240$ to $512 * 424$, and the transmission capacity of the USB connector increased from 60 MB/s to 500 MB/s; at the same time, the KinectV2, the skeletal tracking ability of the system, has been further enhanced. The number of skeletal nodes recognized by KinectVl has been increased from 20 skeletal nodes per person to 25 skeletal nodes per person. The system structure of Kinect is shown in Figure 3.

2.2. Obtaining Human Behavior Data Based on Kinect

2.2.1. Kinect Obtains Depth Image. Kinect obtains depth images according to its infrared emission device and infrared depth camera, and its specific imaging principle is shown in Figure 4. When Kinect's infrared emitting device emits infrared light onto the surface of the scene, the rough surface or transparent diffuser will scatter the light into randomly distributed light and dark spots, that is, the phenomenon of light interference. Because these light spots are independent of each other, and scattered points at different distances in the same scene will form different light spot patterns, this feature can be used to encode the location information of the scene. Then, through the CMOS sensor included in the Kinect infrared depth camera, the light spot pattern of each scattering point is collected, and then the depth image can be obtained through the relevant decoding operation. If the scene in the field of view moves, it will not affect the acquisition of position information, but correspondingly, the depth video will be generated at a rated frequency of 30 frames/s. The above is actually the depth imaging principle of KinectVl, and the depth imaging principle of Kinect V2 is different. Although KinectV2 still transmits and receives infrared light through the infrared emission devices and infrared depth cameras on both sides, it is based on the reflection principle of light and uses the built-in

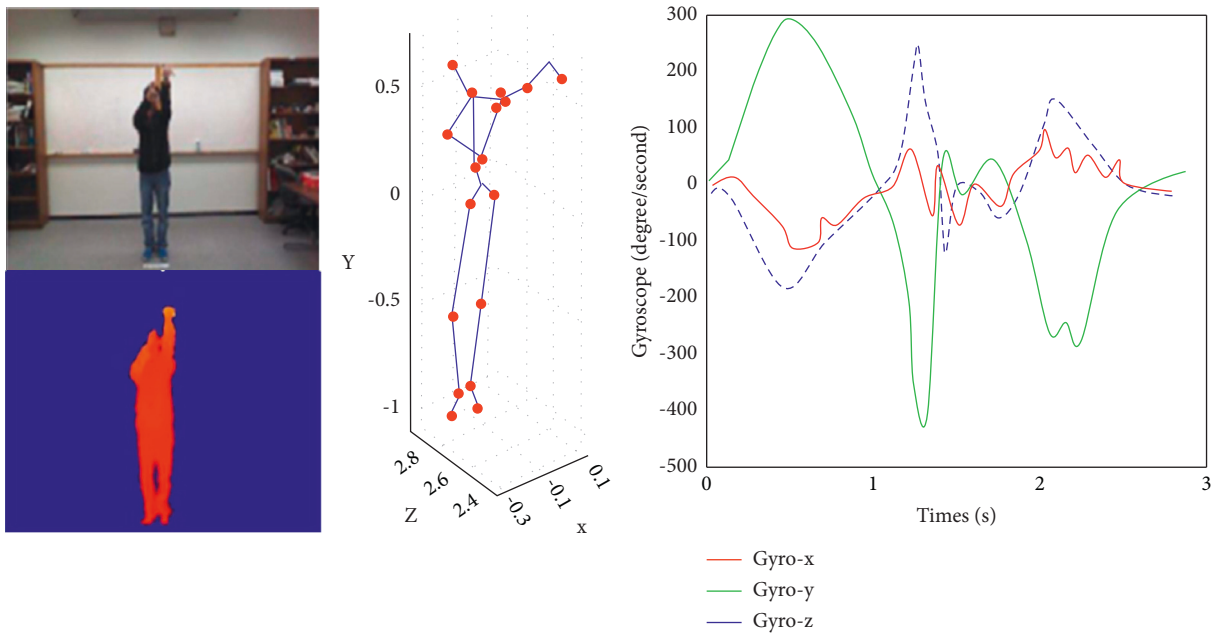

Figure 2: Four-modal data of UTD-MHAD database.

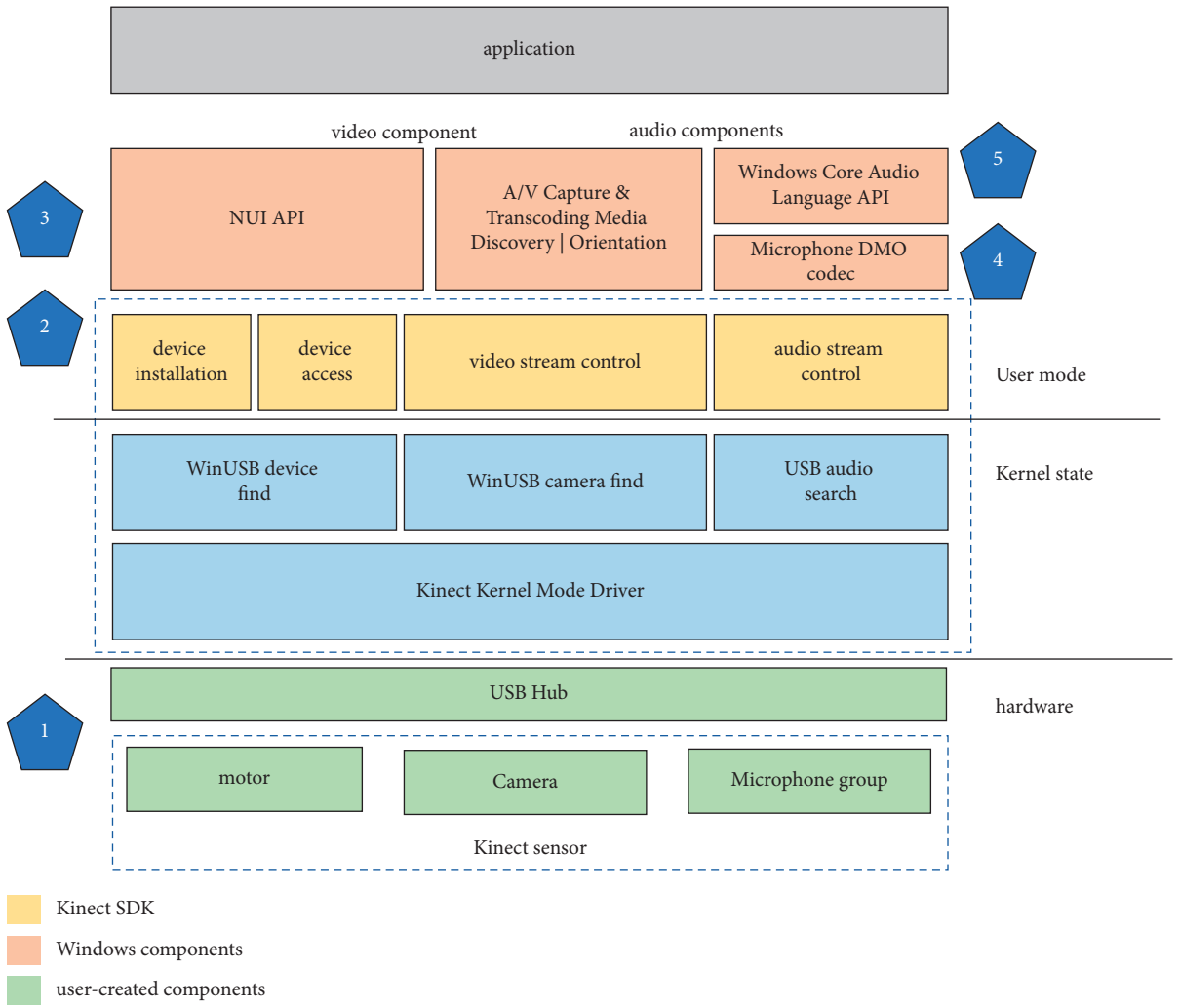

Figure 3: Kinect system structure.

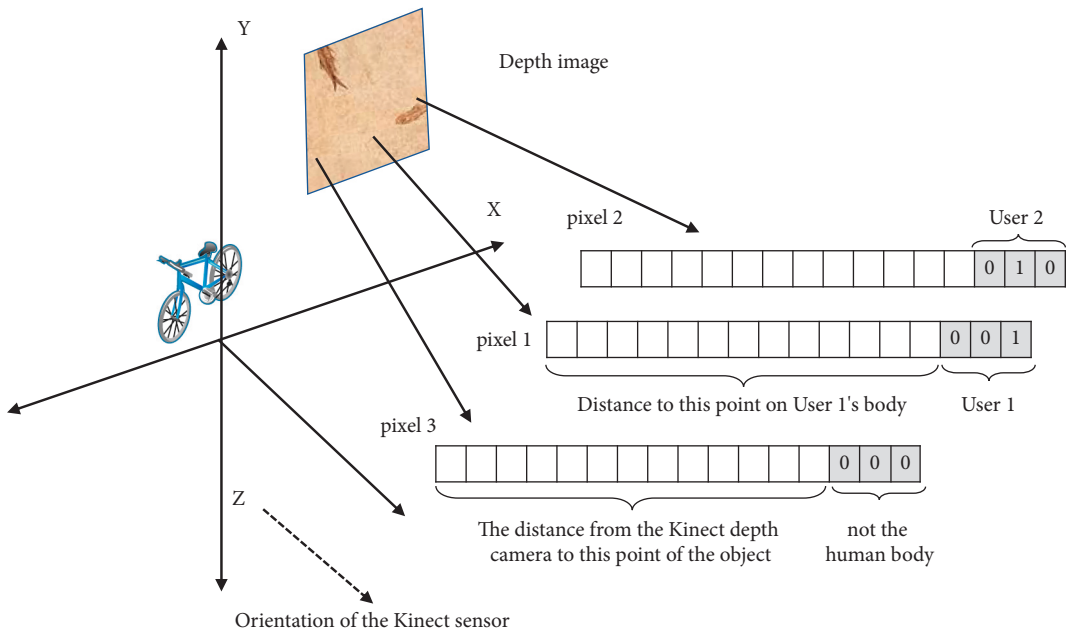

Figure 4: Kinect depth imaging principle.

photosensitive device to record the round-trip time difference of light so as to calculate the position information of the scene and obtain the depth image.

2.2.2. Kinect Obtains Bone Data. Bone positioning based on the depth image of human behavior can obtain human bone data, which should be attributed to the Kinect bone tracking technology. Bone tracking systems usually use depth cameras to obtain the most reliable real-time results, but at the same time, 2D cameras with open-source software can be used to track bones at a lower frame rate. In short, bone tracking algorithms can recognize the presence of one or more people, as well as the location of their heads, bodies, and limbs. Some systems can track hands or specific gestures at the same time but not all bone tracking systems. The principle of bone tracking technology is based on the method of computer vision, and its specific implementation is as follows: firstly, the obtained depth image of human behavior is image processed, and the distance information is used to detect the edge of human behavior; then, the detected target human body is imaged for the purpose of the behavior contour, which is extracted from the image background; secondly, the key parts of the human body and main skeletal nodes, such as head, limbs, and body are identified by machine learning methods; finally, the human skeleton model is established based on the Kinect coordinate system, and the three-dimensional coordinates of the skeletal nodes are generated. Figure 5 shows the human skeleton extracted by Kinect.

2.3. Deep Learning Technology. This method can be used to analyze the characteristic patterns of target objects. Compared with traditional machine learning, deep learning can avoid the trouble of manually designing features. The robustness of deep learning is that its performance is very stable, with more data. The more stable it is, the less reliable it will be. Secondly, versatility is also surprising. Many of the same business processes will do it, especially in the medical field. Many scholars have begun to refer to its algorithm to do some tumor analysis. Finally, scalability, because when doing data analysis, each data sample is independent of each other. These independent data can be used for training, and distributed clusters can also be used for parallelization of models and data. So as to quickly help model training and get a better accuracy. It adopts an end-to-end method to convert data input into target output. The intermediate process automatically collects low-level features of data to represent high-level features of the target. Deep learning corresponds to the deep-level structure in machine learning. Compared with the shallow-level structure, deep learning can better extract the high-dimensional features of the target, and is more suitable for processing data such as images and videos. Since the development of deep learning, many learning models have emerged, but they can be roughly divided into three types: discriminative models, generative models, and hybrid models. Among them, the discriminant model directly models the conditional probability $p(y|x)$, that is, it first learns the experience from the sample, and then extracts the corresponding features of the target to predict the label. Common discriminant models are: CNN and conditional random field (CRF) wait. The generative model models $p(x,y)$ then uses the Bayesian formula to find $p(y\,|\,x)$, and finally selects the one that maximizes $p(y\,|\,x)$. The calculation process is expressed as follows:

$$y = \operatorname{argmax}_y p(y\,|\,x) = \operatorname{argmax}_y \frac{p(x\,|\,y)p(y)}{p(x)} = \operatorname{argmax}_y p(x, y). \tag{1}$$

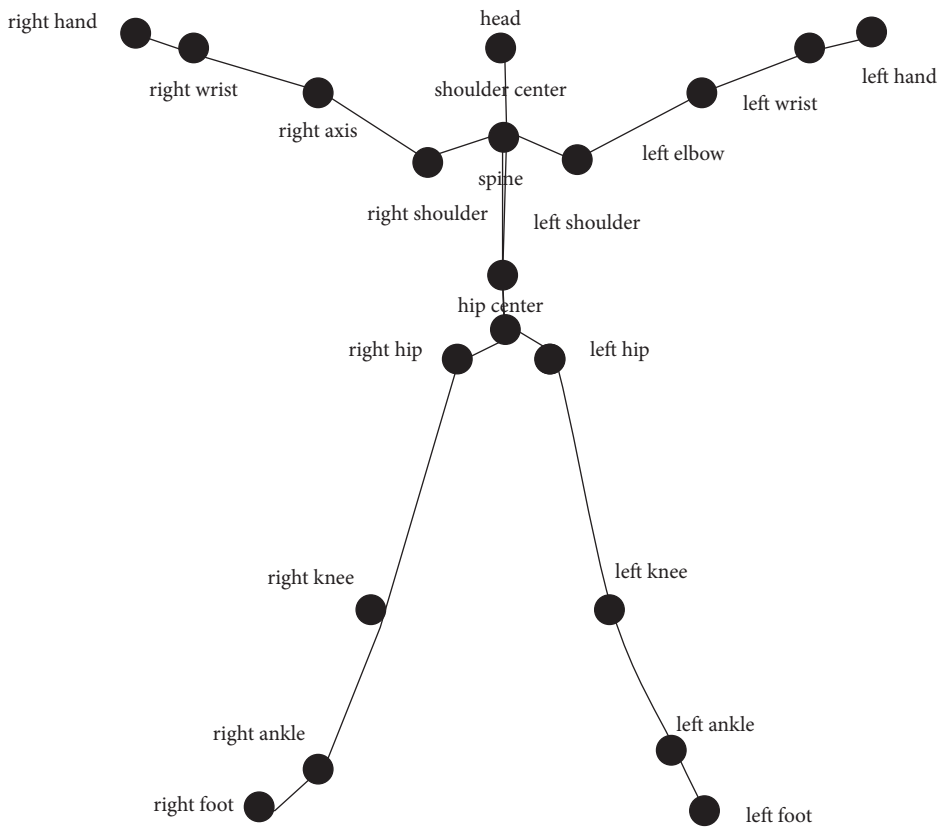

Figure 5: Kinect extracts skeleton diagram.

That is, first learn multiple experiences based on the characteristics of multiple samples, and then select the maximum value of the probability that the corresponding features of the target belong to each sample to predict the label. Regarding the comparison between the two, as shown in Figure 6, the hybrid model is the combination of the discriminative model and the generative model.

3. Methodology

3.1. Wushu Action Feature Extraction Based on Neural Network. Different neural network technologies have different accuracy in extracting martial arts movements. Figure 7 shows the accuracy results of neural network technologies including MLP and SVM in image feature extraction.

3.1.1. Preprocessing of Martial Arts Action Images. The depth image sequence is the depth image arranged in time sequence, which contains rich spatial structure and temporal information. The depth image at a certain moment is called the frame or the i-th frame of the depth image sequence. In order to construct each frame into a two-dimensional structure similar to a slice, this paper adopts a projection algorithm to obtain the three-dimensional information of the behavioral depth image. Specifically, each frame of the depth image Mapi (i represents the i-th frame) of the depth image sequence map is projected onto three orthogonal Cartesian plane coordinate systems to obtain the front view Map_f^i, side view Map_s^i and the top view Map_t^i is $\text{Map}_v^i (v \in (f, s, t))$. The main steps to obtain Map_v^i in this article are as follows:

(1) Calculate the maximum pixel value of each frame of the depth image sequence:

$$\text{Max} = \max\left(\text{Map}_{a*b*n}\right), \tag{2}$$

where $a * b$ is the size of the frame and n is the total number of frames in the sequence;

(2) Determine the dimensions of the three projection views as

$$\begin{aligned} \text{Size}\left(Map_f^i\right) &= a * b, \\ \text{Size}\left(\text{Map}_s^i\right) &= a * \text{Max}. \end{aligned} \tag{3}$$

(3) If $\text{Map}_f^i\ (x,y)$ represents the pixel in the xth row and the yth column of Map_f^i, when $\text{Map}_f^i\ (x,y) \neq 0$, the projection formula of this pixel on Map_s^i is

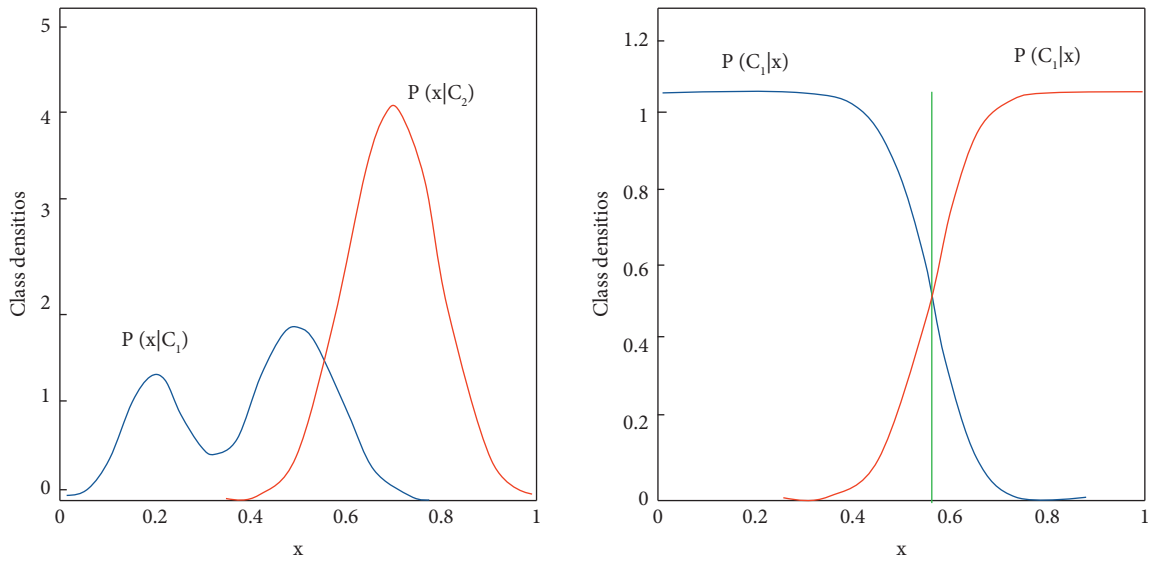

Figure 6: Comparison of generative models and discriminative models.

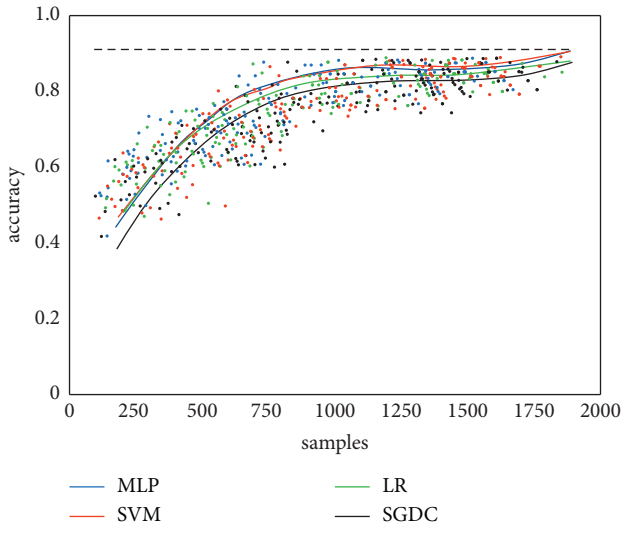

Figure 7: Accuracy of different neural network techniques for extracting martial arts action features.

$$\mathrm{JMap}_s^i\left(x, \mathrm{Map}_f^i(x, y)\right) = y. \tag{4}$$

The projection formula on Map_t^i is as follows:

$$\mathrm{Map}_t^i\left(\mathrm{Map}_f(x, y), y\right) = x. \tag{5}$$

After the above algorithm steps, the projected view Map_v^i of each frame of Map^i in each map can be obtained. Map_v^i contains the three-dimensional structure of the depth image sequence and extracts the spatial characteristics of the behavior, but lacks the temporal information of the behavior, so the acquired Map_v^i needs to be further processed.

The combined morphological operation formula can be expressed as follows:

$$G(x, y) = F(x, y).B(x, y) - F(x, y). \tag{6}$$

$G(x, y)$ represents the image processed by the combined operation; $F(x, y)$ represents a frame of image; $B(x, y)$ represents the structural element; and ● represents the closing operation. Complete target extraction. A frame of the video images cannot fully describe an action. Due to differences in motion rates, the number of frames per video image may be different even for the same motion. In order to deal with the changes of these two rates, the grayscale features of it on this into the image, and this are from the established images.

The operation process of the accumulated edge image is as follows: a frame of image processed by it in the video image is by $G(x, y)$; the edge of it by using this on $G(x, y)$. $E(x, y)$ indicates that this image; it by multiplying $G(x, y)$ and $E(x, y)$ on it is $I(x, y)$, and the grayscale information is on the edge point. If the pixel point is this, the grayscale value is 0; the accumulated edge image is by $H(x, y, t)$. The scale is of $G(x, y)$, and $H(x, y, t)$ is obtained to the $I(x, y)$ in a certain time window in the video image to one image.

Initialize $H(x, y, t)$, set all pixels to 0, and set the time condition to $t = 0$; based on edge detection, the first frame is the $G(x, y)$ of it can be obtained. The grayscale image $I(x, y)$ is the $G(x, y)$ and the $E(x, y)$; compare $I(x$ on all pixels, y) and the accumulated $H(x, y, t{-}1)$ obtained in the previous frame, the gray value of it with a larger gray value will be used of $H(x, y, t)$; repeat Edge detection step until all image operations are completed. The information content in this is huge, and the formula for it at point (x, y) at time t is

$$\begin{aligned} I(x, y) &= G(x, y)E(x, y), \\ H(x, y, t) &= \max\left(H(x, y, t-1), I(x, y)\right). \end{aligned} \tag{7}$$

Image into it, not accumulating each frame of binary image into one image. 0 and 1 are the only two values o fit of the $E(x, y)$, if this in the $E(x, y)$ and the $I(x, y)$ is 1. If it is

accumulated for it, it contains this of more frame images, the directional gradient can be directly solved these.

If this is accumulated for it, the information center already contains it, and the directional gradient histogram can be directly solved all problems.

4. Result Analysis and Discussion

4.1. Experimental Data Setup. Simple descriptive statistics can only make a superficial description and a display of statistical data. In order to explore the regularity, we need to infer the statistical method. Inferential statistics is to infer the relevant population on the basis of collecting and sorting out the data of observation samples. According to the random observation sample data and the conditions and assumptions of the problem, an inference in the form of a probability is made for the unknown. That is the content of probability theory and mathematical statistics. To use descriptive data, it is necessary to set the data standard. In this paper, the human body is decomposed and set. The experimental data are shown in Table 1.

4.2. Human Wushu Action Recognition Based on Dynamic Time Regularization. The continuity, that is, an action can be a collection of it. The human body can reflect the change trend of it of this, and the angle change curve of it can be called it. The human motion features are joint angle time series. If the duration of a martial arts action is set to T, the motion features can be defined as follows:

$$\text{action feature} = \{A_1, A_2, \ldots, A_M T\}. \tag{8}$$

Here, vector A is the row; M is the number of motions, and the range is $1 \le M \le 16$.

The comparison of the focus of this, which will lead to it of the data, so the following formula is

$$x_i = \frac{x_1 + x_2 + \cdots + x_n + x_{n+1}}{n}. \tag{9}$$

X_i is the joint angle value at the i-th time; x_n, x_n+1 are the values of n and n+1 orders, respectively; n is an integer greater than 0.

The theory is based on the idea of it to find the distance between two test samples. The time is set to $R = \{r_1, r_2, \ldots, r_i, \ldots, r_{L_1}\}$, and the test sample is set to $T = \{t_1, t_2, \ldots, t_j, \ldots, t_{L_2}\}$.

The values at time i and j are r_i and t_j; L_1 and L_2 represent the lengths. The distance matrix $D(i, j)$ can be described as follows:

$$D(i, j) = \min \begin{Bmatrix} D(i, j-1) \\ D(i-1, j) \\ D(i-1, j-1) \end{Bmatrix} + \mathrm{d}(r_i, t_j), \quad i = 1, 2, \ldots, L_1;\ j = 1, 2, \ldots, L_2. \tag{10}$$

Here, $d(r_i,t_j)$ represents the distance function of r_i and t_j; $D(i,j-1)$, $D(i-1,j)$, $D(i-1,j-1)$ are the distance elements.

To make points r and t on the different angle Y-axis values, it is necessary to construct a three-dimensional vector based on points r_i and t_j to redefine $d(r_i, t_j)$ to replace the distance, that is, $r_i = [r_i, \dot{r}_i, \ddot{r}_i]$ and $t_j = [t_j, \dot{t}_j, \ddot{t}_j]$, the first derivative $\dot{r}_i$ of it and the $\ddot{r}_i$ of the reference sequence are described in turn as follows:

$$\dot{r}_i = \frac{(r_i - r_{i-1}) + ((r_{i+1} - r_{i-1})/2)}{2}, \qquad \ddot{r}_i = r_{i+1} + r_{i-1} - 2r_i, \tag{11}$$

where r_i-1 is value at the $i-1$th time; r_i+1 is the value in $i+1$th time point. Since the vector is beneficial to the accuracy of the mapping, $d(r_i, t_j)$ can be defined as follows:

$$\mathrm{d}(\ddot{r}_i - \ddot{t}_j) = w_1(\ddot{r}_i - \ddot{t}_j)^2 + w_2(\ddot{r}_i - \ddot{t}_j)^2 + w_3(\ddot{r}_i - \ddot{t}_j)^2. \tag{12}$$

Here, $\dot{t}_j$ represents the first-order derivative value of it of this; $\ddot{t}_j$ is the second-order value of it of this; w_1, w_2, and w_2 represent, respectively. Adjust the weight of it of the value, and adjust the first-order value of angle. The shortest distance weight and the shortest distance weight of the second derivative of the adjustment it.

Its extraction of this in the image is carried out by the Kinect technology, and then the time sequence of it is calculated by it. Martial arts action decomposition and identification process.

4.3. Analysis of Experimental Results

4.3.1. Analysis of Experimental Results Based on NReJ3D Technology. During the experiment, two pieces of muscle pattern recognition data and two pieces of joint transition recognition data were selected, respectively. The experimental recognition results are shown in Figure 8. The basic idea of action recognition is to use the sample data in the action data set. Nrej3d technology is used to match the angle features extracted from the sample training set with the angle features in the test set. It can be seen that, at the beginning, the traditional method is slightly better than the recognition method. But in the subsequent experiments, it is not higher than this method. Therefore, the recognition method proposed in this paper has stronger recognition ability. Therefore, the angle between bones can be used to replace the changes of bones, that is, the changes of actions, so as to recognize actions. Not all the joint points in the action data set can be used, so unnecessary joint points can be filtered out to realize the dimension reduction processing of the data, making the data more reliable and universal.

4.3.2. Method Comparison Based on MSR Action 3D Database. After discussing and analyzing the influence of different ensemble decision methods on the behavior recognition results, the average value method ensemble decision that is most suitable for the algorithm model in this paper is proposed. The experimental results show that the average of the algorithm on the MSR Action 3D database

Table 1: Experimental parameter settings.

Superficial muscle movement	Large joint movement
Vastus rectus, VR	Flexion, leg lift, and kick
Tibialis anterior, TA	Dorsiflexion, inversion, and adduction
Peroneus longus, PL	Use group eversion and plantar flexion to flex the forearm, lift the heel, and fix the joint
Gastrocnemius, Gs	Upper body leaning forward (similar to the movement of the lateral gastrocnemius muscle on the inside)
So leus musc, SM	Fixed joints, back support, and half body rotation
Tuorp lrgsiing, TL	Stabilize the half body and rotate the lower body ankle joints

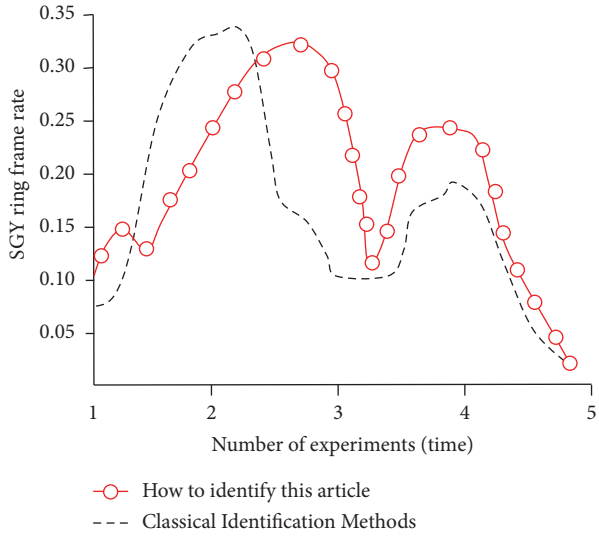

Figure 8: Comparison of experimental results.

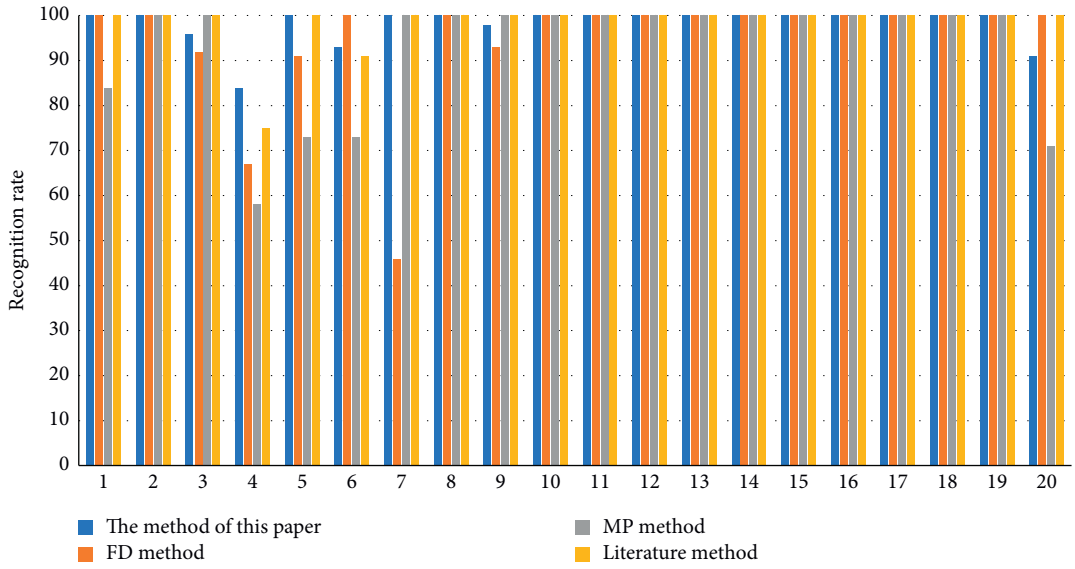

Figure 9: Classification and recognition performance of this paper and the other three methods.

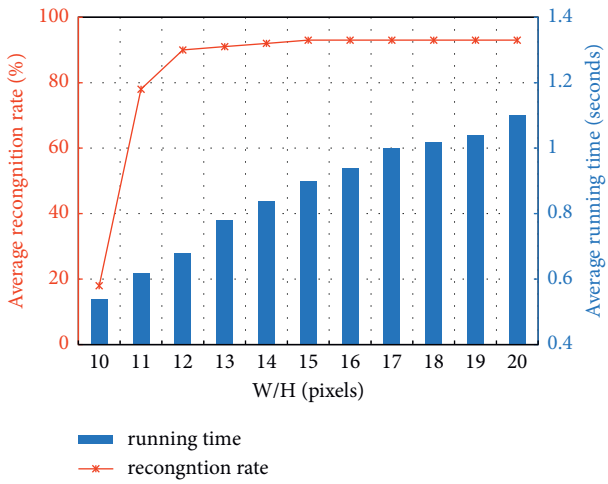

Figure 10: Accuracy and speed of the recognition algorithm in this paper.

reaches 98.1%, which verifies the effectiveness of the 5C–CNN model construction, which was further analyzed through the confusion matrix, and the recognition performance of the 5C–CNN model based on the mean value method ensemble decision on the MSR Action 3D database test set was further analyzed. This paper will compare other literature algorithms using the same experimental settings on the same database. The experimental results are compared in Figures 9 and 10.

5. Conclusion

As a research hotspot at this stage, human behavior recognition has been successfully applied to various fields of life and technology. With the emergence of Kinect devices, the research object of human behavior recognition has gradually changed from traditional RGB images to depth data that is not easily disturbed by noise. Based on the research status at home and abroad, depth image-based, bone data-based, and deep learning-based have become the three major directions of human behavior recognition in recent years, but the above studies still have their own problems: human behavior recognition based on depth images. Higher actions lead to misjudgment, and this based on skeleton data has serious self-occlusion problems. Human behavior recognition based on deep learning requires the model to make corresponding structural adjustments to different data. We propose a deep learning algorithm for human action recognition based on Kinect multiview features. Behavior recognition combined with depth images and skeleton data can better make up for the lack of single data and enrich the details of behavior to a greater extent. In this paper, the recognition experiment is carried out on the MSR action 3D database, the influence of different ensemble decision methods on the recognition results is further discussed and analyzed, and the average value method ensemble decision that is most suitable for the algorithm model of this paper is proposed. The results show that the recognition accuracy of the algorithm reaches 98.1%, proving the feasibility of the preprocessing algorithm.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

References

[1] T. Lei, "Intelligent recognition method of human motion global features based on Kinect skeleton information[C]// 2020 IEEE international conference on industrial application of artificial intelligence (IAAI)," *IEEE*, vol. 11, no. 4, pp. 99–105, 2020.

[2] P. Cao and S. Zhang, "Research on image recognition of Wushu action based on remote sensing image and embedded system," *Microprocessors and Microsystems*, vol. 82, no. 3, Article ID 103841, 2021.

[3] L. Wang, "Analysis and evaluation of Kinect-based action recognition algorithms[J]," vol. 4, no. 3, pp. 17–24, 2021, https://arxiv.org/abs/2112.08626.

[4] Y. Liu, J. Chen, J. Jiang et al., "Recognition of jet modes in electrohydrodynamic direct-writing based on image segmentation[J]," *Modern Physics Letters B*, vol. 36, no. 08, pp. 55–61, 2022.

[5] F. Farsian, N. Krachmalnicoff, and C. Baccigalupi, "Foreground model recognition through Neural Networks for CMB *B*-mode observations," *Journal of Cosmology and Astroparticle Physics*, vol. 2020, no. 7, p. 17, 2020.

[6] Z. Wu, T. Weise, L. Zou, F. Sun, and M. Tan, "Skeleton based action recognition using a stacked denoising autoencoder with constraints of privileged," *Information[J]*, vol. 16, no. 9, pp. 87–95, 2020.

[7] M. Tabejamaat and H. Mohammadzade, "Contributive representation based reconstruction for online 3D action recognition[J]," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 4, no. 7, pp. 55–63, 2020.

[8] T. Hussain, N. Iqbal, H. F. Maqbool, M. Khan, and M. Tahir, "Amputee walking mode recognition based on mel frequency cepstral coefficients using surface electromyography sensor," *International Journal of Sensor Networks*, vol. 32, no. 3, p. 139, 2020.

[9] T. Karacsony, A. Mira Loesch-Biffar, C. Vollmar, S. Noachtar, and J. Paulo Silva Cunha, "A deep learning architecture for epileptic seizure class ification based on object and action recognition[C]//ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP)," *IEEE*, vol. 15, no. 4, pp. 33–45, 2020.

[10] A. Ma, F. Khan, M. J. Khan, Y. Amin, and A. Akram, "EEG-Based emotion recognition for multi-channel fast empirical mode decomposition using VGG-16[C]//2020 international conference on engineering and emerging technologies (ICEET)," vol. 9, no. 1, pp. 78–84, 2020.

[11] R. Rajeswari, T. Devi, and S. Shalini, "Dysarthric speech recognition using variational mode decomposition and convolutional neural networks," *Wireless Personal Communications*, vol. 122, no. 1, pp. 293–307, 2022.

[12] H. Ali, M. S. Z. Azalan, A. F. A. Zaidi, T. S. T. Amran, M. R. Ahmad, and M. Elshaikh, "Feature extraction based on empirical mode decomposition for shapes recognition of

buried objects by ground penetrating radar," *Journal of Physics: Conference Series*, vol. 1878, no. 1, Article ID 012022, 2021.

[13] I. Bulugu, "Sign language recognition using Kinect sensor based on color stream and skeleton points," *Tanzania Journal of Science*, vol. 47, no. 2, pp. 769–778, 2021.

[14] J. Ragot, "Active mode recognition of dynamic systems," *International Journal of Systems Science*, vol. 51, no. 13, pp. 2500–2519, 2020.

[15] W. Kim, Y. Kim, and K. Y. Lee, "Human gait recognition based on integrated gait features using Kinect depth cameras [C]//2020 IEEE 44th annual computers, software, and applications conference (COMPSAC)," *IEEE*, vol. 2, no. 6, pp. 77–84, 2020.

[16] Y. Na and D. K. Ko, "Deep-learning-based high-resolution recognition of fractional-spatial-mode-encoded data for free-space optical communications[J]," *Scientific Reports*, vol. 11, no. 1, pp. 112–123, 2021.

[17] F. Sherratt, A. Plummer, and P. Iravani, "Understanding LSTM network behaviour of IMU-based locomotion mode recognition for applications in prostheses and wearables," *Sensors*, vol. 21, no. 4, p. 1264, 2021.

[18] W. Liu, X. Han, and M. M. Kamruzzaman, "Adaptive recognition method for VR image of Wushu decomposition based on feature extraction," *IEEE Access*, vol. 8, no. 99, p. 1, 2020.

[19] 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.

[20] M. Song, J. Wang, H. Zhao, and Y. Li, "Bearing failure of reciprocating compressor sub-health recognition based on CAGOA-VMD and GRCMDE," *Advances in Mechanical Engineering*, vol. 14, no. 3, 2022.