IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

SPECIAL SECTION ON BIG DATA LEARNING AND DISCOVERY

Received March 12, 2019, accepted April 7, 2019, date of publication April 17, 2019, date of current version April 30, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2911705

Accurate Hierarchical Human Actions Recognition From Kinect Skeleton Data

BENYUE SU[1,2], HUANG WU[1,2], MIN SHENG[2,3], AND CHUANSHENG SHEN[2,3]

[1]School of Computer and Information, Anqing Normal University, Anqing 246133, China
[2]Intelligent Perception and Computing Key Laboratory of Anhui Province, Anhui 246133, China
[3]School of Mathematics and Computational Science, Anqing Normal University, Anqing 246133, China

Corresponding author: Benyue Su (subenyue@sohu.com)

This work was supported in part by the National Nature Science Foundation of China under Grant 61603003 and Grant 11471093, in part by the Funds of Integration of Cloud Computing and Big Data, in part by the Innovation of Science and Education of the Ministry of Education of China under Grant 2017A09116, in part by the Science and Technology Major Project of Anhui Province under Grant 18030901021, in part by the Leading Talent Team Project of Anhui Province, and in part by the Anhui Provincial Department of Education Outstanding Top-Notch Talent-Funded Projects under Grant gxbjZD26. The work of C. Shen was supported by the Project of Coarse-Graining Methods for Multiscale Dynamics on Complex Networks (NSFC) under Grant 11475003.

ABSTRACT Human action recognition has become one of the most active research topics in natural human interaction and artificial intelligence, and has attracted much attention. Human movement ranges from simple to complex, from low-level to advanced, with an increasing degree of complexity and data noise. In other words, there is a complicated hierarchy in movement actions. Hierarchy theory can efficiently describe these complicated hierarchical relationships of human actions. Accordingly, a hierarchical framework for human-action recognition is designed in this paper. Different features are selected according to the level of action, and specific classifiers are selected for different features. In particular, a two-level hierarchical recognition framework is constructed and tested on Kinect skeleton data. At the first level, we use support vector machine for a coarse-grained classification, while at the second level we use a combination of support vector machine and a hidden Markov model for a fine-grained classification. Ten-fold cross-validations are used in our performance evaluation on public and self-built datasets, achieving average recognition rates of 95.69% and 97.64%, respectively. These outstanding results imply that the hierarchical step-wise precise classification can well reflect the inherent process of human action.

INDEX TERMS Activity recognition, statistical learning, supervised learning.

I. INTRODUCTION

Many people learn movements incorrectly in infancy, resulting in problems with posture, bone position, and growth [1], and many suffer from disabilities or paralysis due to accidents and diseases. In fact, tens of thousands of people lose one or more athletic abilities every year, rehabilitation exercise can restore their athletic ability. Currently, there are two methods of medical rehabilitation: artificially- and robot-assisted. Because of the shortage of health-care providers, the expense of intelligent robots, and associated maintenance costs, neither form of rehabilitation can be widely accessed. Therefore, there is strong demand for an affordable, functional, and convenient rehabilitation training method.

Microsoft recently released a low-cost depth-sensing device called Kinect [2]. The user can control the interaction interface through gestures and speech, without a remote or physical control. Its natural somatosensory interaction technology provides a new way to interact with a device. It captures the user's actions and interacts with the device to provide a completely new experience. More importantly, its Software Development Kit (SDK) enables us to obtain information on the human skeleton position [3].

Rehabilitation training systems based on Kinect have emerged in recent years. Chang *et al.* designed the Kinerehab system [4], which uses Kinect-sensor image-processing technology to detect motion information of patients with motion disorders. Lange *et al.* combined virtual reality and video-game technology to develop a Kinect-based game for balance-rehabilitation training for patients with spinal injuries and traumatic brain injuries [5] [6]. Da Gama *et al.*

The associate editor coordinating the review of this manuscript and approving it for publication was Zhong-Ke Gao.

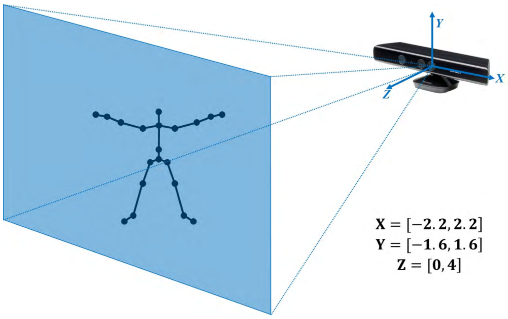


FIGURE 1. 20 skeleton points captured by Kinect.

designed a Kinect-based rehabilitation training system for the shoulder and elbow [7]. Zhao *et al.* proposed a Kinect-based virtual rehabilitation system [8]. Capecci *et al.* proposed an HSMM-based algorithm to monitor and evaluate rehabilitation exercises using Kinect v2.0 [9].

Although Kinect-based rehabilitation training systems were used in these studies [10], most studies use RGB+D data, and few focus on skeleton data. This is because Kinect is a depth sensor based on structured light, which is not accurate in itself, and the internal algorithm obtains skeleton data on the basis of depth data. Therefore, the skeleton data captured with Kinect have low quality and high noise. This is a challenge to the method presented in this article. The Kinect sensor represents a skeleton by capturing the 20 most representative joint points of the human body, as shown in Figure 1.

Since the system obtains the human-skeleton movement-position information through Kinect in this paper. One of the most important parts concerns human-action recognition via skeleton data.

Our work focuses on research of human-action recognition methods. There has been some related work in recent years [11]. Wang *et al.* presented random occupancy pattern features from depth maps [12]. Oreifej *et al.* presented an easily implemented descriptor for activity recognition from depth sequences [13] These studies all used depth-of-field data and focused less on the action of the human skeleton. Later studies gradually moved toward position estimation and motion information based on the joint points of the human body. For example, an actionlet ensemble model was proposed for human-action recognition with depth maps and joint positions [14], and data-mining techniques were applied to spatial-temporal pose structures for action representation using a state-of-the-art pose-estimation algorithm [15]. Rahmani *et al.* presented a new descriptor for action recognition, histogram of oriented principal components (HOPC), which is more robust to action speed and viewpoint variations [16]. Other research concerns skeleton-based human-action recognition. Vemulapalli *et al.* represented a human skeleton as a point in a Lie group and modeled human actions as curves in the group. The approach outperforms various methods of skeleton-based human-action recognition [17]. Chen et al. proposed a two-level hierarchical framework for action recognition with 3D skeleton sequences [18]. A part-based five-dimensional feature vector is defined, then action sequences are clustered by using these features to construct the first level of the hierarchical framework. Motion-feature extraction and action graphs are used at the second level. Approaches based on deep learning are also common in recent research. Du *et al.* proposed an end-to-end hierarchical recurrent neural network (RNN) for skeleton-based action recognition [19]. A large-scale NTU RGB+D for a human-action-recognition dataset was introduced by Shahroudy et al. [20], and they also proposed a part-aware LSTM model to further improve the performance of the LSTM learning framework. Most recently, Two LSTM sub-networks are used by Zhang *et al.* [21] to regress the spatial rotation and translation parameters of the skeleton, and then the skeleton is rotated to an angle suitable for behavior prediction. Song *et al.* [22] use attention mechanism to obtain discriminatory temporal and spatial features for action recognition. Yan *et al.* [23] proposed a spatial temporal graph convolutional network based on GCN model to learn the temporal and spatial characteristics of human skeleton sequences and Li *et al.* [24] use a hierarchical CNN network to learn spatial information and dynamic features between human joints.

There are many methods of human-action recognition, each with its own advantages. However, current work has not yet fully utilized the human hierarchical structure for action recognition, thus we propose a method to build upon a two-level hierarchical recognition framework. The contributions and highlights of this paper are as follows.

- Based on the human biological structure and the multi-granularity of human action, we design a multi-level hierarchical rehabilitation-action recognition model based on hierarchical theory.
- We use a combination classifiers to adapt hierarchical different actions and features.
- We present a low-cost method to monitor human actions recognition via Kinect skeleton data.

The paper is organized as follows. Section 1 introduces the background and significance of Kinect-based rehabilitation training systems and related work on skeleton-based human-action recognition. Section 2 explains the hierarchical recognition model, including feature extraction and classifier construction. We verify our method by experiment in section 3. Conclusions and ideas for future work are summarized in section 4.

II. BIOLOGY-BASED HIERARCHICAL MODEL

A. HUMAN PARTS BASED ON HUMAN ANATOMY

The human motion system consists of bones, joints, and skeletal muscle. The bones of the body are connected by joints to form the skeleton; skeletal muscles adhere to the bones and cross the joints. Skeletal muscle contraction, using the joint as a fulcrum, leads to bone-position changes based on skeletal traction, which is movement. In movement, skeletal muscle is a dynamic organ. Therefore, skeletal muscle is

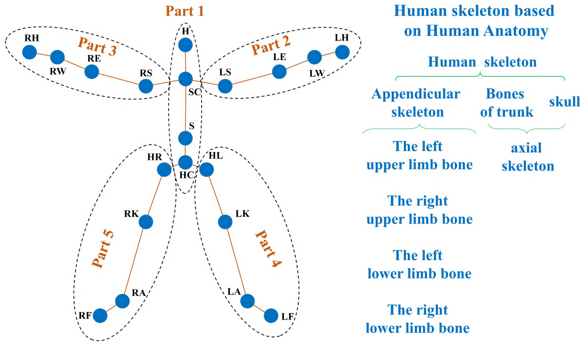


FIGURE 2. Skeleton classification based on human anatomy.

the active part in the human motion system, and the bones and joints are the passive parts. In this paper, one can recognize human action by changing the position of the bone. Since the understanding of the human skeleton is of high importance, we begin with its analysis.

Most adults have 206 bones. These are divided into five categories: axial skeleton, left-upper-extremity skeleton, right-upper-extremity skeleton, left-lower-extremity skeleton, and right-lower-extremity skeleton. A person is represented by 20 skeleton points in Kinect: ***HipCenter (HC), Spine (S), ShoulderCenter (SC), Head (H), Left Shoulder (LS), Left Elbow (LE), Left Wrist (LW), Left Hand (LH), Right Shoulder (RS), Right Elbow (RE), Right Wrist (RW), Right Hand (RH), HipLeft (HL), Left Knee (LK), Left Ankle (LA), Left Foot (LF), HipRight (HR), Right Knee (RK), Right Ankle (RA), Right Foot (RF)***. The five parts of the human body based on Kinect skeleton data are shown in Figure 2.

We categorize these as shown below.

$$\begin{cases} Part\ 1: \{HC, S, SC, H\} & (Waist\&Head) \\ Part\ 2: \{LS, LE, LW, LH\} & (Left\ Arm) \\ Part\ 3: \{RS, RE, RW, RH\} & (Right\ Arm) \\ Part\ 4: \{HL, LK, LA, LF\} & (Left\ Leg) \\ Part\ 5: \{HR, RK, RA, RF\} & (Right\ Leg) \end{cases}$$

B. ACTION FEATURES BASED ON HUMAN KINEMATICS

Kinect (Version 1.0) can obtain 20 pieces of key joint-position information to represent human movement. Suppose that $P_i^{(t)} = (x_i^{(t)}, y_i^{(t)}, z_i^{(t)})$ $(i = 1, 2, 3, \ldots, I; t = 1, 2, 3, \ldots, T)$ indicates the location of the i^{th} joint point at frame t, i.e., $P_i^{(t)}$ indicates the position of the point on the X-, Y-, and Z-axes. To enhance readability, we replace i with the joint abbreviation. For example, we can express the right hand as $P_{RH}^{(t)} = (x_{RH}^{(t)}, y_{RH}^{(t)}, z_{RH}^{(t)})$.

The skeleton data obtained by Kinect is the position information in the world coordinate system, and we have already divided the human body into five parts based on human anatomy. Each part must be studied in its local coordinate system. We define HC, LS, RS, HL, HR as the origins of the local coordinate systems for parts 1-5, respectively, of the human body. We take the right arm (RS, RE, RW, RH) as an example in Figure 3.

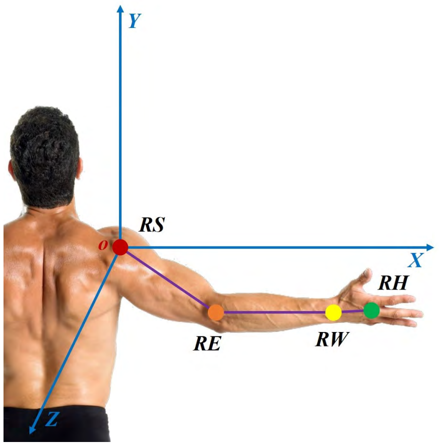


FIGURE 3. Constructing local coordinate system of right arm.

The skeleton data obtained by Kinect do not include the movement of the fingers; these actions can be recognized by a two-level hierarchical recognition model from our previous work [25].

On the first level, we take the part of the human body involved in an action, such as the action of the right hand alone or both hands simultaneously, as a class. We use the barycenter of each part of the first three joint points to extract action features. Let us take the right arm as an example. To calculate $C_3^{(t)}$, the supposed barycenter coordinates of RE, RW, and RH in the third part of the body at frame t, we propose the following formula:

$$\begin{aligned} C_3^{(t)} &= \frac{P_{RE}^{(t)} + P_{RW}^{(t)} + P_{RH}^{(t)}}{3} - P_{RS}^{(t)} \\ &\Longrightarrow C_3^{(t)} = (x_{C_3}^{(t)}, y_{C_3}^{(t)}, z_{C_3}^{(t)}) \end{aligned} \tag{1}$$

We can similarly obtain $C_1^{(t)}, C_2^{(t)}, C_4^{(t)}, C_5^{(t)}$ by calculating the distance between the frame and the initial frame on the X-, Y-, and Z-axes from the second frame. The distance change of the barycenter in the third part is shown in Figure 4. Taking the X-axis as an example, suppose that $\tilde{x}_{C_3}^{(t)}$ represents the distance change on the X-axis at frame t:

$$\begin{aligned} \tilde{x}_{C_3} &= \left\{\tilde{x}_{C_3}^{(1)}, \tilde{x}_{C_3}^{(2)}, \cdots, \tilde{x}_{C_3}^{(T-1)}\right\} \\ where\ \tilde{x}_{C_3}^{(t-1)} &= ||x_{C_3}^{(t)} - x_{C_3}^{(1)}||, t = 2, 3, \cdots, T. \end{aligned} \tag{2}$$

We calculate the respective range, mean, variance, and skewness of $\tilde{x}_{C_3}$ as

$$\begin{aligned} R(\tilde{x}_{C_3}) &= Max(\tilde{x}_{C_3}) - Min(\tilde{x}_{C_3}) \\ M(\tilde{x}_{C_3}) &= Mean(\tilde{x}_{C_3}) \\ V(\tilde{x}_{C_3}) &= Variance(\tilde{x}_{C_3}) \\ S(\tilde{x}_{C_3}) &= Skewness(\tilde{x}_{C_3}). \end{aligned} \tag{3}$$

Note that $x_{C_3}^{(1)}$ in formula (2) represents the calibration frame of the action. We can similarly determine the statistical

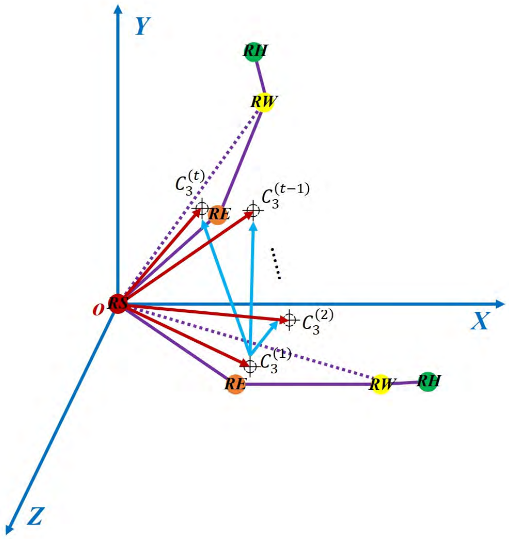


FIGURE 4. First-level features of right arm.

features of $\tilde{y}_{C_3}$, $\tilde{z}_{C_3}$ to obtain:

$$\begin{aligned} R(\tilde{C}_3) &= (R(\tilde{x}_{C_3}), R(\tilde{y}_{C_3}), R(\tilde{z}_{C_3})) \\ M(\tilde{C}_3) &= (M(\tilde{x}_{C_3}), M(\tilde{y}_{C_3}), M(\tilde{z}_{C_3})) \\ V(\tilde{C}_3) &= (V(\tilde{x}_{C_3}), V(\tilde{y}_{C_3}), V(\tilde{z}_{C_3})) \\ S(\tilde{C}_3) &= (S(\tilde{x}_{C_3}), S(\tilde{y}_{C_3}), S(\tilde{z}_{C_3})). \end{aligned} \tag{4}$$

In the same way, we can obtain $R(\tilde{C}_1)$, $R(\tilde{C}_2)$, $R(\tilde{C}_4)$, $R(\tilde{C}_5)$, $M(\tilde{C}_1)$, $M(\tilde{C}_2)$, $M(\tilde{C}_4)$, $M(\tilde{C}_5)$, $V(\tilde{C}_1)$, $V(\tilde{C}_2)$, $V(\tilde{C}_4)$, $V(\tilde{C}_5)$, $S(\tilde{C}_1)$, $S(\tilde{C}_2)$, $S(\tilde{C}_4)$, $S(\tilde{C}_5)$.

On the second level, human actions tend to be fine-grained, so we must classify similar actions that involve the same parts. The fourth joint point in each part is called the end-effector of the human-body action. Its movement trajectory can best reflect the features of human action. In human kinematics, movement contains a uniform relationship in time and space. We take the relative position of the end-effector to represent its spatial motion. We take the relative distance of the end-effector in the previous frame and the last frame (i.e., speed) to indicate its temporal motion. To better describe the motion characteristics of the end-effector, we construct local and global coordinate systems to describe the respective local and global information of the motion. The global coordinate system is constructed with the HipCenter (HC) as the origin.

We use L and G, respectively, to represent the end-effectors in local and global coordinates for reading and writing. Therefore, the third-part end-effector (RH) at the current frame t can be expressed as $L_3^{(t)}$ and $G_3^{(t)}$, and so on. The equations for the local and global positions of the third-part end-effector are

$$\begin{aligned} L_3^{(t)} &= P_{RH}^{(t)} - P_{RS}^{(t)} = (x_{L_3}^{(t)}, y_{L_3}^{(t)}, z_{L_3}^{(t)}) \\ G_3^{(t)} &= P_{RH}^{(t)} - P_{HC}^{(t)} = (x_{G_3}^{(t)}, y_{G_3}^{(t)}, z_{G_3}^{(t)}). \end{aligned} \tag{5}$$

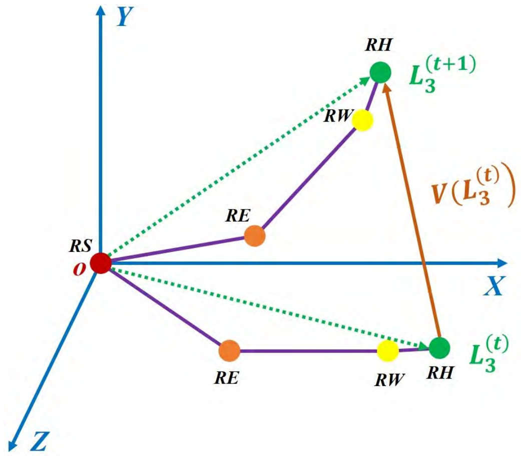


FIGURE 5. Second-level features of right arm.

We calculate the local relative offset distance on the X-, Y-, and Z-axes between the next frame t+1 and the current frame t as:

$$\begin{aligned} v(L_3^{(t)}) &= (v(x_{L_3}^{(t)}), v(y_{L_3}^{(t)}), v(z_{L_3}^{(t)})) \\ v(x_{L_3}^{(t)}) &= x_{L_3}^{(t+1)} - x_{L_3}^{(t)} \\ v(y_{L_3}^{(t)}) &= y_{L_3}^{(t+1)} - y_{L_3}^{(t)} \\ v(z_{L_3}^{(t)}) &= z_{L_3}^{(t+1)} - z_{L_3}^{(t)} \\ where\ t &= 1, 2, \cdots, T-1. \end{aligned} \tag{6}$$

Similarly, we can obtain the global relative offset distance $v(G_3^{(t)})$. Taking the third part (RS, RE, RW, RH) as an example, as shown in Figure 5, we use L_3, $v(L_3)$ as local motion features of the third part end-effector (RH). In the same way, we obtain other end-effector features, i.e., L_1, L_2, L_4, L_5, $v(L_1)$, $v(L_2)$, $v(L_4)$, $v(L_5)$,G_1, G_2, G_4, G_5, $v(G_1)$, $v(G_2)$, $v(G_4)$, $v(G_5)$.

C. FEATURE SELECTION AND COMBINATION CLASSIFIER

The complexity of human-body movement makes human-action recognition a challenge [26]. To overcome this difficulty, we use knowledge of the human anatomy and human kinematics to design a two-level hierarchical recognition model. The algorithm flow-process diagram is shown in Figure 6.

On the first level, we divide all action categories into several major categories based on the five parts of the human body. These major categories have the feature that the parts of the human body involved in the action are the same. For example, if only the third part of the human body (RS, RE, RW, RH) is active and the other parts remain still, this is called a right-arm action. Similarly, if both the second (LS, LE, LW, LH) and third part (RS, RE, RW, RH) are active throughout the course of the movement, this is called a double-arm action. One of our experiments is based on the MSRAction3D dataset, whose details are shown in Table 1 (specifically, side-boxing with single hand and with both

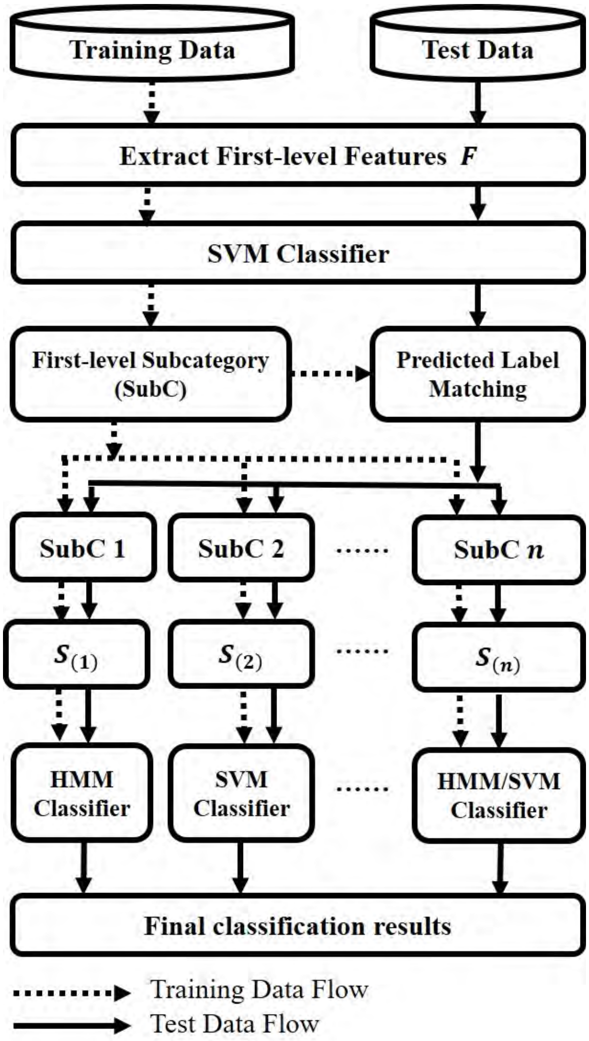


FIGURE 6. Flow-process diagram of two-level recognition model.

hands are the same action in MSRAction3D, and are collectively abbreviated as SB).

Therefore, we use ***F*** [see formula (7)] to represent the first-level feature vectors. The features of the first-level classification are as follows:

$$\begin{aligned} F &= (F_{C_1}, F_{C_2}, F_{C_3}, F_{C_4}, F_{C_5}), \\ F_{C_i} &= (R(\tilde{C}_i), M(\tilde{C}_i), V(\tilde{C}_i), S(\tilde{C}_i)), \\ where\ i &= 1, 2, 3, 4, 5 \end{aligned} \tag{7}$$

F, the distance between each frame and the calibration frame, is time-independent. A support vector machine (SVM) is a strong classifier. We use SVM as the first-level classifier.

On the second level, reclassification of the first-level classification results is needed. If the number of subcategories in the first-level classification results is unique, then it directly outputs the final result. If not, then further classification is needed. As the second-level becomes fine-grained, features that contain temporal information must be represented. We use ***S*** [see formula (8)] to represent the second-level feature vectors, and S_{P_j} to represent the second-level features of the j^{th} part of the human body. The features of the second-level classification are as follows:

TABLE 1. First-level classification target on MSRAction3D.

First-level categories	Original categories
Actions with right arm	***high arm wave (HiW), horizontal arm wave (HoW), hammer (H), hand catch (HCa), forward punch (FP), high throw (HT), draw X (DX), draw tick (DT), draw circle (DC), tennis swing (TSw), side-boxing with single hand (SB).***
Actions with both arms	***hand clap (HCl), tennis serve (TSe), two-hand wave (TH), side-boxing with both hands (SB).***
Actions with right leg	***forward kick (FK), side kick (SK).***
Actions with waist	***bend (B).***
Actions with arms and legs	***jogging (J).***
Actions with waist and both arms	***golf swing (GS).***
Actions with waist and right arm	***pick up & throw (P&T).***

$$\begin{aligned} S &= (S_{P_1} \oplus S_{P_2} \oplus S_{P_3} \oplus S_{P_4} \oplus S_{P_5}), \\ S_{P_j} &= S_{E_j} \mid S_{C_j}, \\ S_{E_j} &= (L_j, v(L_j), G_j, v(G_j)), \\ S_{C_j} &= (R(\tilde{C}_i), M(\tilde{C}_i), V(\tilde{C}_i), S(\tilde{C}_i)), \\ where\ j &= 1, 2, 3, 4, 5 \end{aligned} \tag{8}$$

Note: Symbols $\oplus$ and $\mid$ in formula (8) represent that $a \oplus b := a\ or\ b\ or\ (a, b)$, $A \mid B := A\ or\ B$.

For better understanding, we give the following definition.

For $a \oplus b := a\ or\ b\ or\ (a, b)$, we may wish to assume that a and b are the feature vectors of the second part (left arm) and third part (right arm), respectively. Then

$$\begin{cases} a \oplus b := a & given\ Condition\ 1 \\ a \oplus b := b & given\ Condition\ 2 \\ a \oplus b := (a, b) & given\ Condition\ 3 \end{cases}$$

Condition 1: If and only if the second part is involved in the action.
Condition 2: If and only if the third part is involved in the action.
Condition 3: Both the second and third part are involved in the action.

As the human body has five major parts, so the rest of the situation and so on.

For $A \mid B := A \text{ or } B$, we may wish to assume that A and B are the feature vectors of the first and second level, respectively. Then

$$\begin{cases} A \mid B := A & \textit{given Condition } 4 \\ A \mid B := B & \textit{given Condition } 5 \end{cases}$$

Condition 4: At least two parts are involved in the action. Moreover, we select SVM as the classifier for the current feature.

Condition 5: If and only if only one part is involved in the action. Moreover, we select HMM as the classifier for the current feature.

S_{C_j} and $\boldsymbol{F}$ are similar, and we use SVM as their classifier. Because S_{E_j} of second-level human action features includes temporal information, SVM is not suitable for its classifier. The hidden Markov model (HMM) can express a transition between states, and the human-action details can be expressed as a state-to-state change. For S_{E_j}, HMM is a more appropriate classifier. At the same time, according to human behavioral habits, when humans use only one part of the body, they tend to move more carefully. On the choice of features, S_{E_j} can better describe detailed human actions, such as moving a single arm. For a coordinated action of a number of parts, such as a double arm movement, the type of action range is wide, and not very fine, and S_{C_j} can better distinguish this type of action. Therefore, we define that the action type is only a single part, and we use S_{E_j} + HMM for classification. For other multi-part actions, we use S_{C_j} + SVM for classification.

One of our experiments is based on the MSRAction3D dataset. Taking second-level features as an example, suppose that the first subcategory (*SubC* 1; see Figure 6) is the first major subcategory, actions with the right arm (see Table 1). Its feature vector can be represented by $S_{(1)}$, as follows:

$$S_{(1)} = S_{P_3} = S_{E_3} = (L_3, v(L_3), G_3, v(G_3)). \tag{9}$$

Similarly, the second major subcategory is actions with both arms (see Table 1). Its feature vector can be represented by $S_{(2)}$, as follows:

$$\begin{aligned} S_{(2)} &= (S_{P_3}, S_{P_3}) = (S_{C_2}, S_{C_3}) \\ S_{(2)} &= (R(\tilde{C}_2), M(\tilde{C}_2), V(\tilde{C}_2), S(\tilde{C}_2), R(\tilde{C}_3), M(\tilde{C}_3), V(\tilde{C}_3), S(\tilde{C}_3)) \end{aligned} \tag{10}$$

Based on the above definition, $S_{(1)}$ and $S_{(2)}$ respectively use the SVM and HMM classifiers for recognition.

III. METHOD VALIDATION

For validation, we used an Intel Core i5 4210M CPU @ 2.6 GHz, 8 GB RAM, with a Windows 10 64-bit operating system and MATLAB R2015b. We used the MSRAction3D (MSRAction3D Skeleton Real3D) public dataset and a self-built dataset collected from the actions in [27].

MSRAction3D consists of 20 action types of 10 subjects, who perform each action two or three times. There are 567 action-sequence files in total. The self-built dataset includes the same 20 action types. We used Kinect with its own SDK to directly obtain the skeleton-movement data. Ten subjects, including five males and five females, were included in the self-built dataset. Each action was performed three times, for a total of 600 action-sequence files. To better understand the 20 actions, we visualize the skeleton-action data, as shown in Figure 7.

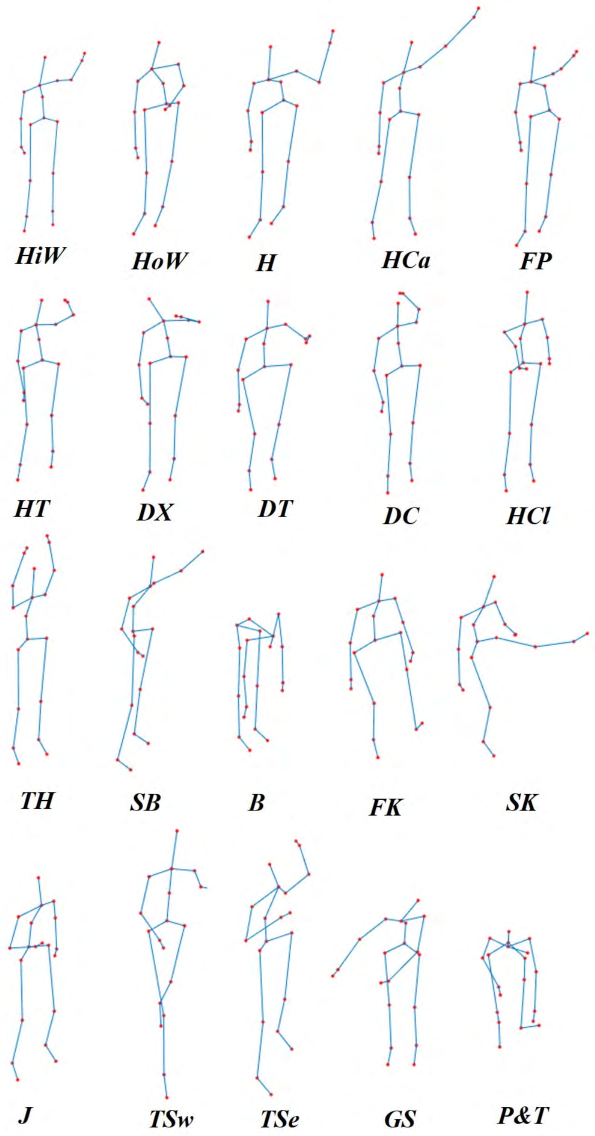


FIGURE 7. Action types in MSRAction3D and self-built dataset.

Cross-validation is a practical method of statistically cutting data samples into smaller subsets, and it is used to prevent overfitting. A round of cross-validation involves dividing a data sample into complementary subsets, analyzing one subset (the training set), and validating the analysis on the other subset (the validation set or test set). To reduce variability, multiple rounds of cross-validation are performed using different partitions, and the verification results (e.g., averages) are combined to estimate the final prediction model. Cross-validation is a common algorithm used in machine learning,

TABLE 2. Actions in each of the MSRAction3D subsets.

AS1		AS2		AS3	
Label	Action	Label	Action	Label	Action
a02	***HoW***	a01	***HiW***	a06	***HT***
a03	***H***	a04	***HCa***	a14	***FK***
a05	***FP***	a07	***DX***	a15	***SK***
a06	***HT***	a08	***DT***	a16	***J***
a10	***HCl***	a09	***DC***	a17	***TSw***
a13	***B***	a11	***TH***	a18	***TSe***
a18	***TSe***	a14	***FK***	a19	***GS***
a20	***P&T***	a12	***SB***	a20	***P&T***

evaluation, and statistical analysis. To ensure reliable results, we used several cross-validation methods.

Hold-Out Method: The original data are randomly divided into training and test sets. The training set is used as the training classifier, and the test set is used to verify the model. It is common practice to use approximately $2/3 \sim 4/5$ of the data for training, and the rest for testing. In this article, we selected $2/3$ of the data for training.

K-fold Cross-validation (K-CV): The original data are divided into K groups (usually equalized). Each subset of the data is a test set, and the remaining $K-1$ groups are a training set, which will get K models. The average test-set classification accuracy of these K models is used as the performance indicator of the classifier under K-CV. In general, $K = 10$ (as an empirical parameter).

Leave-One-Out Cross-validation (LOO-CV): If the original data consists of N samples, then LOO-CV is N-CV, i.e., each sample alone is a test set and the remaining $N-1$ samples are a training set, so LOO-CV will include N models. The average test-set classification accuracy of the N models is the performance indicator of the classifier.

Since user differences can affect the recognition results of actions, to reflect the robustness of the verification, we include a user-independent verification method. Following the standard experimental protocol of MSRAction3D, **Leave-One-Actor-Out Cross Validation (LOAO-CV)** and **Random Selection Cross Validation (RS-CV)** were used in our experiment. With LOAO-CV training with all the actor sequences in the dataset except one user for testing, 10 subjects perform all actions in the datasets, hence we tested 10 times in total. RS-CV randomly chooses half the subjects for training, and the remaining ones for testing, all the possible combinations (252 tests in total) are in our experiment.

We adopted two types of experiments. In the first type, all 567 sequences of the dataset were split into three subsets such as Action Set1 (AS1), Action Set2 (AS2), and Action Set3 (AS3) (see Table 2) [27], each with eight actions [28]. We used five validation methods for this experiment. The results are given in Table 3, and we compare our work with others with or without the MSRAction3D experimental protocol in Table 4.

TABLE 3. The rate of the first experiment on three subsets, where the overall recognition rate is calculated by averaging the results over subsets.

CV-Methods	AS1		AS2		AS3		Overall	
	best	avg	best	avg	best	avg	best	avg
Hold-Out	—	96.00	—	96.10	—	97.33	—	96.48
10-fold CV	96.67	91.89	100	94.17	100	96.87	98.89	94.31
LOO-CV	—	93.36	—	94.37	—	97.79	—	95.17
LOAO-CV	100	87.42	100	84.89	100	94.86	100	89.06
252 tests	94.44	81.33	94.64	82.64	100	92.66	96.36	85.54

TABLE 4. Comparison of the first experiment results on subsets, where the recognition rate is in terms of the best results.

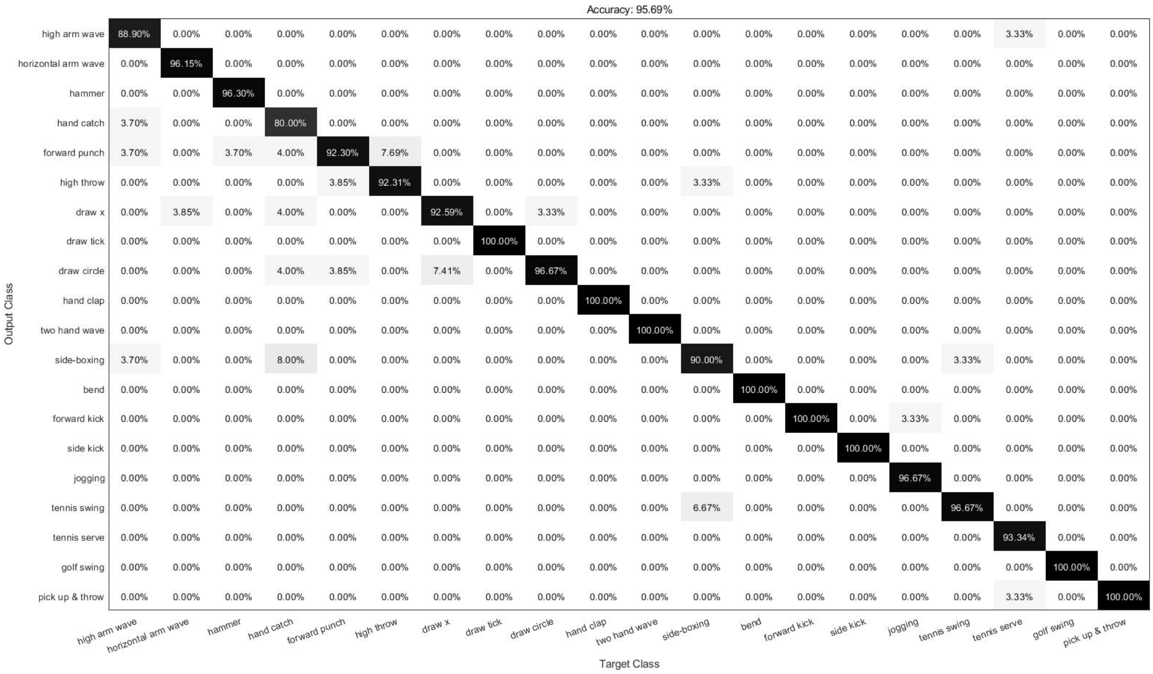
Methods	Best recognition rate
Bag of 3D points [27]	74.7
Histograms of 3D joints [29]	78.97
EigenJoints [30]	82.3
EigenJoints + Hierarchical [31]	90.3
Histograms of oriented displacements [32]	91.26
Random forests [33]	94.3
Space-time pose [34]	92.77
Points in a Lie group [17]	92.46
Dynemes and forward differences [35]	93.6
Part-based feature vector [18]	96.1
Ensemble TS-LSTM [36]	97.22
Ours (252 tests)	96.36
Ours (LOAO-CV)	100
Ours (10-fold CV)	98.89
Ours (Hold-Out)	96.48
Ours (LOO-CV)	95.17

TABLE 5. The rate of the second experiment on MSRAction3D.

CV-Methods	Best rate	Worst rate	Average rate
Hold-Out	—	—	96.76
10-fold CV	98.48	86.79	95.69
LOO-CV	—	—	96.05
LOAO-CV	96.55	73.58	89.62
252 tests	95.19	72.56	84.72

From Table 4, we can observe that the results of our method and the state of the art are close with the same cross-subject tests (252 tests) for this experiment. It is noteworthy that our method outperforms the state-of-the-art method in non-user independent crossover experiments (10-fold CV).

In the second type of experiment, 10 samples with serious data loss were removed, and the rest of the sequences



	high arm wave	horizontal arm wave	hammer	hand catch	forward punch	high throw	draw x	draw tick	draw circle	hand clap	two hand wave	side-boxing	bend	forward kick	side kick	jogging	tennis swing	tennis serve	golf swing	pick up & throw
high arm wave	88.90%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.33%	0.00%	0.00%
horizontal arm wave	0.00%	96.15%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
hammer	0.00%	0.00%	96.30%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
hand catch	3.70%	0.00%	0.00%	80.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
forward punch	3.70%	0.00%	3.70%	4.00%	92.30%	7.69%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
high throw	0.00%	0.00%	0.00%	0.00%	3.85%	92.31%	0.00%	0.00%	0.00%	0.00%	0.00%	3.33%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
draw x	0.00%	3.85%	0.00%	4.00%	0.00%	0.00%	92.59%	0.00%	3.33%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
draw tick	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
draw circle	0.00%	0.00%	0.00%	4.00%	3.85%	0.00%	7.41%	0.00%	96.67%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
hand clap	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
two hand wave	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
side-boxing	3.70%	0.00%	0.00%	8.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	90.00%	0.00%	0.00%	0.00%	0.00%	3.33%	0.00%	0.00%	0.00%
bend	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
forward kick	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	3.33%	0.00%	0.00%	0.00%	0.00%
side kick	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%
jogging	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	96.67%	0.00%	0.00%	0.00%	0.00%
tennis swing	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	6.67%	0.00%	0.00%	0.00%	0.00%	96.67%	0.00%	0.00%	0.00%
tennis serve	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	93.34%	0.00%	0.00%
golf swing	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%
pick up & throw	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.33%	0.00%	100.00%

FIGURE 8. Confusion matrix of the 10-fold CV result. The vertical coordinate "Output Class" represents the predicted label of an action, and the horizontal coordinate "Target Class" represents its true label. The value of each coordinate grid represents the accuracy rate of the predicted action label recognized as the true action label.

TABLE 6. Comparison of the second experiment results on MSRAction3D.

Methods	Best	Worst	Avg
Random occupancy patterns [12]	86.5	—	—
Actionlet ensemble [14]	88.2	—	—
HON4D + D$_{disc}$ [13]	88.89	—	82.15
Spatial and temporal part sets [15]	90.22	—	—
HOPC of 3D pointclouds [16]	92.39	74.36	86.49
Points in a Lie group [17]	89.48	—	—
Histograms of action poses + DTW [37]	90.56	—	—
Dynemes and forward differences [35]	91.94	—	—
Part-based feature vector [18]	95.56	74.39	87.05
Ours (252 tests)	95.19	72.56	84.72
Ours (LOAO-CV)	96.55	73.58	89.62
Ours (10-fold CV)	98.48	86.79	95.69
Ours (Hold-Out)	—	—	96.76
Ours (LOO-CV)	—	—	96.05

were used. Therefore, all 20 actions with 557 samples were applied, as in [18]. This experiment is obviously more challenging. We used cross-validation, as with the first experiment, and the results and comparison are shown in Tables 5 and 6.

As shown in Table 6, with the same cross-subject tests (252 tests) for this experiment, the result of our method is similar to that of the state of the art, and even decreases slightly in general. However, it is certain that our method outperforms other methods in terms of the best, worst, or average result in non-user independent crossover experiments (10-fold CV, hold-out, and LOO-CV). With another cross-subject test (LOAO-CV), which is commonly used for the MSRAction3D dataset [38], our method outperformed the state-of-the-art method, which means that our method needs enough training data to train a better model. Although LOAO-CV has shown excellent performance, the range of 10-fold CV recognition rates is smaller, therefore, its result is more stable and reliable; its confusion matrix is shown in Fig. 8. From the confusion matrix, we can see that eight of 20 actions achieve 100%, and most recognition rates exceed 90%.

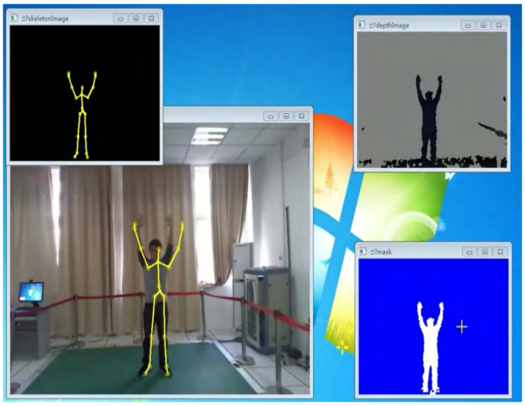


FIGURE 9. Collected skeleton action data via Kinect.

Furthermore, to ensure the practicability of the method, we used a self-built dataset to verify its feasibility in the real world. We collected human actions, as shown in Figure 9. The

TABLE 7. Rate of second experiment on self-built dataset.

CV-Methods	Best rate	Worst rate	Average rate
Hold-Out	—	—	97.00
10-fold CV	100	95.16	97.64
LOO-CV	—	—	96.50
LOAO-CV	100	85.00	93.33
252 tests	97.33	76.33	90.53

environment of the second type of experiments was applied to the self-built dataset, with recognition results as shown in Table 7. We found the performance on the self-built dataset to be better than on the MSRAction3D dataset.

Comparing the above experimental results, we find that our method has good performance on recognition rate both in the public and self-built datasets. The experimental results on the public dataset show that our method has its advantages over other methods. It is close or superior to the state-of-the-art method, with or without the user-independent experiment, and the case on the self-built dataset shows that our method has practical significance.

IV. CONCLUSIONS AND DISCUSSIONS

We propose a low-cost human action recognition method, based on a biology-based hierarchical model using Kinect skeleton data. Under the theory of biology, the hierarchical structure of the human body is based on human anatomy, while feature extraction is based on human kinematics, and the theory of hierarchy is applied to a two-level hierarchical action-recognition model in the present work. We selected different features at various levels and applied an appropriate combination of classifiers. The experimental results show promising recognition rate with self-built datasets and the recognition ability no less than the existing studies on the public datasets.

Deep learning has made remarkable achievements in recent years, including applications to human-action recognition. We have introduced a deep learning method with improved performance in a comparison of experimental results.

We think the relationship between traditional machine learning and the case based on deep learning should be complementary, based on their respective advantages. Therefore, future work will have two directions. One is the application of our method to other types of datasets, such as MSR Daily Activity, MSRC-Kinect12, HDM-05, and NTU-RGB+D. The other is to combine deep learning with our proposed hierarchical strategy, enhancing the stability and effectiveness of human-action recognition.

CONFLICTS OF INTEREST

The authors declare that there are no competing interests.

DATA AVAILABILITY STATEMENT

Datasets are available. The public dataset can be found at https://www.uow.edu.au/%7ewanqing/#Datasets, and the self-built dataset at https://github.com/bysu2017/DataSets.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Ali for his linguistic assistance during the preparation of this manuscript.

REFERENCES

[1] G. T. O'connor, J. E. Buring, and S. Yusuf, "An overview of randomized trials of rehabilitation with exercise after myocardial infarction," *Circulation*, vol. 80, no. 2, pp. 234–244, Aug. 1989.

[2] Z. Zhang, "Microsoft Kinect sensor and its effect," *IEEE MultiMedia*, vol. 19, no. 2, pp. 4–10, Feb. 2012.

[3] G. T. Papadopoulos, A. Axenopoulos, and P. Daras, "Real-time skeleton-tracking-based human action recognition using Kinect data," in *Proc. Int. Conf. Multimedia Model.*, Jan. 2014, pp. 473–483.

[4] Y. J. Chang, S. F. Chen, and J. D. Huang, "A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities," *Res. Developmental Disabilities*, vol. 32, no. 6, pp. 2566–2570, Dec. 2011.

[5] B. Lange *et al.*, "Development and evaluation of low cost game-based balance rehabilitation tool using the microsoft Kinect sensor," in *Proc. Annu. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2011, pp. 1831–1834.

[6] B. Lange *et al.*, "Interactive game-based rehabilitation using the microsoft Kinect," in *Proc. IEEE Virtual Reality*, Feb. 2012, pp. 171–172.

[7] G. A. Da *et al.*, "Poster: Improving motor rehabilitation process through a natural interaction based system using Kinect sensor," in *Proc. IEEE Symp. 3D User Inter.*, Mar. 2012, pp. 145–146.

[8] L. Zhao *et al.*, "A Kinect-based virtual rehabilitation system through gesture recognition," in *Proc. Int. Conf. Virtual Reality Visualizat.*, Sep. 2017, pp. 380–384.

[9] M. Capecci *et al.*, "Physical rehabilitation exercises assessment based on hidden semi-Markov model by Kinect v2," in *Proc. IEEE-EMBS Int. Conf. Biomed. Health Informat. (BHI)*, Feb. 2016, pp. 256–259.

[10] Z. K. Gao *et al.*, "An adaptive optimal-Kernel time-frequency representation-based complex network method for characterizing fatigued behavior using the SSVEP-based BCI system," *Knowl.-Based Syst.*, vol. 152, pp. 163–171, Jul. 2018.

[11] L. L. Presti and M. L. Cascia, "3D skeleton-based human action classification: A survey," *Pattern Recognit.*, vol. 53, pp. 130–147, May 2016.

[12] J. Wang *et al.*, "Robust 3d action recognition with random occupancy patterns," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Oct. 2012, pp. 872–885.

[13] O. Oreifej and Z. Liu, "Hon4d: Histogram of oriented 4d normals for activity recognition from depth sequences," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, Aug. 2013, pp. 716–723.

[14] J. Wang *et al.*, "Mining actionlet ensemble for action recognition with depth cameras," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 1290–1297.

[15] C. Wang, Y. Wang, and A. L. Yuille, "An approach to pose-based action recognition," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 915–922.

[16] H. Rahmani *et al.*, "HOPC: Histogram of oriented principal components of 3D pointclouds for action recognition," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, May 2014, pp. 742–757.

[17] R. Vemulapalli, F. Arrate, and R. Chellappa, "Human action recognition by representing 3d skeletons as points in a lie group," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, Aug. 2014, pp. 588–595.

[18] H. Chen, G. Wang, J.-H. Xue, and L. He, "A novel hierarchical framework for human action recognition," *Pattern Recognit.*, vol. 55, pp. 148–159, Jul. 2016.

[19] Y. Du, W. Wang, and L. Wang, "Hierarchical recurrent neural network for skeleton based action recognition," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1110–1118.

[20] A. Shahroudy *et al.*, "NTU RGB+ D: A large scale dataset for 3D human activity analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1010–1019.

[21] P. Zhang *et al.*, "View adaptive recurrent neural networks for high performance human action recognition from skeleton data," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Aug. 2017, pp. 2136–2145.

[22] S. Song *et al.*, "An end-to-end spatio-temporal attention model for human action recognition from skeleton data," in *Proc. AAAI Conf. Artif. Intell.*, vol. 1, no. 2, 2017, pp. 4263–4270.

[23] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," in *Proc. AAAI Conf. Artif. Intell.*, Aug. 2018, pp. 1–9.

[24] C. Li *et al.*, "Co-occurrence feature learning from skeleton data for action recognition and detection with hierarchical aggregation," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, 2018, pp. 786–792.

[25] B. Su, H. Wu, and M. Sheng, "Human action recognition method based on hierarchical framework via Kinect skeleton data," in *Proc. Int. Conf. Mach. Learn. Cybern. (ICMLC)*, Jul. 2017, pp. 83–90.

[26] L. Chen, H. Wei, and J. Ferryman, "A survey of human motion analysis using depth imagery," *Pattern Recognit. Lett.*, vol. 34, no. 15, pp. 1995–2006, Nov. 2013.

[27] W. Li, Z. Zhang, and Z. Liu, "Action recognition based on a bag of 3D points," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, pp. 9–14.

[28] M. E. Hussein *et al.*, "Human action recognition using a temporal hierarchy of covariance descriptors on 3D joint locations," in *Proc. 23rd Int. Joint Conf. Artif. Intell. (IJCAI)*, Jun. 2013, pp. 2466–2472.

[29] L. Xia, C. C. Chen, and J. K. Aggarwal, "View invariant human action recognition using histograms of 3D joints," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, May 2012, pp. 20–27.

[30] X. Yang and Y. L. Tian, "EigenJoints-based action recognition using naive-bayes-nearest-neighbor," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, Mar. 2012, pp. 14–19.

[31] H. Chen, G. Wang, and L. He, "Accurate and real-time human action recognition based on 3D skeleton," in *Proc. Int. Conf. Opt. Instrum. Technol.*, Dec. 2013, p. 744,.

[32] M. A. Gowayyed *et al.*, "Histogram of oriented displacements (HOD): Describing trajectories of human joints for action recognition," in *Proc. Int. Joint Conf. Artif. Intell.*, Jun. 2013, pp. 1351–1357.

[33] Y. Zhu, W. Chen, and G. Guo, "Fusing spatiotemporal features and joints for 3D action recognition," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 486–491.

[34] M. Devanne *et al.*, "Space-time pose representation for 3D human action recognition," in *Proc. Int. Conf. Image Anal. Process.*, Jun. 2013, pp. 456–464.

[35] I. Kapsouras and N. Nikolaidis, "Action recognition on motion capture data using a dynemes and forward differences representation," *J. Vis. Comun. Image Represent.*, vol. 25, no. 6, pp. 1432–1445, Aug. 2014.

[36] I. Lee *et al.*, "Ensemble deep learning for skeleton-based action recognition using temporal sliding LSTM networks," in *Proc. Int. Conf. Comput. Vis. (ICCV)*, Feb. 2017, pp. 1012–1020.

[37] M. Barnachon, S. Bouakaz, B. Boufama, and E. Guillou, "Ongoing human action recognition with motion capture," *Pattern Recognit.*, vol. 47, no. 1, pp. 238–247, 2014.

[38] J. R. Padilla-López, A. A. Chaaraoui, F. A. Fllórez-Revuelta, "Discussion on the validation tests employed to compare human action recognition methods using the MSR Action3D dataset," *Comput. Sci.*, to be published.

BENYUE SU received the Ph.D. degree from the School of Computer and Information, Hefei University of Technology, China, in 2007. He held a postdoctoral position with the Hefei Institute of Intelligent Machines, Chinese Academy of Sciences, China, from 2009 to 2012. He is currently a Professor with the School of Computer and Information, Anqing Normal University, China. He has published about 30 high quality referred papers in international conferences and journals. His current research interests include intelligent perception and computing, visual computing, machine learning with applications to exercise rehabilitation and training, and computer vision. He currently serves as the Council Member for the Special Committee of Digital Entertainment and Simulation, Chinese Simulation Federation and of the Technical Committee of Geometric Design and Computing, and the China Society for Industrial and Applied Mathematics. He also is Anhui Province's leading technology talents in China.

HUANG WU received the master's degree in statistics from Anqing Normal University. He is currently working on artificial intelligence in education. His current research interests include computer vision, human action recognition, and deep learning.

MIN SHENG received the Ph.D. degree from the School of Computer and Information, Hefei University of Technology, China, in 2009. She is currently a Professor of mathematics and computational science with the Anqing Normal University, China. She has published about 20 high quality referred papers in international conferences and journals. Her current research interests include image processing, visual computing, and machine learning.

CHUANSHENG SHEN received the Ph.D. degree from the University of Science and Technology of China, in 2012. He was a Visiting Scholar with the Humboldt University of Berlin and also with the Potsdam-Institute for Climate Impact Research, sponsored by the China Scholarship Council, from 2016 to 2017. He is currently a Professor of physics and applied mathematics, and the Dean of the School of Mathematics and Computational Science, Anqing Normal University, Anhui, China. He has authored more than 30 scientific papers. His current research interests include the mesoscopic methods based on statistical mechanics in networked systems, and structure, dynamics, and function of complex networks. He has been a Member Fellow with the China Society for Industrial and Applied Mathematics, since 2017.