IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 11, 2019, accepted August 30, 2019, date of publication September 10, 2019,
date of current version September 25, 2019.

*Digital Object Identifier 10.1109/ACCESS.2019.2940291*

# Geometric Algebra Representation and Ensemble Action Classification Method for 3D Skeleton Orientation Data

**WENMING CAO[1,3,4], YITAO LU[1], AND ZHIQUAN HE[1,2,3]**

[1]Shenzhen Key Laboratory of Media Security, Shenzhen University, Shenzhen 518060, China
[2]Guangdong Key Laboratory of Intelligent Information Processing, Shenzhen 518060, China
[3]Guangdong Multimedia Information Service Engineering Technology Research Center, Shenzhen 518060, China
[4]Video Processing and Communication Laboratory, Department of Electrical and Computer Engineering, University of Missouri, Columbia, MO 65211, USA

Corresponding author: Zhiquan He (zhiquan@szu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61771322, 61971290 and 61375015, and in part by the Shenzhen Foundation under Grant JCYJ20160307154630057.

**ABSTRACT** In this paper, we propose a novel human body posture representation based on Geometric Algebra to extract the angles and orientations of the most informative body joints to describe human body postures. As a motion usually consists of a number of postures, which are different even in the same type of motion. We treat the postures of a motion independently. For each posture, a new Geometric Algebra based skeleton posture descriptor is used to construct the feature vectors as the input for the Support Vector Machine classifier to decide its motion type. To get the type of the whole motion, we choose the most frequent class from the sequence of predictions of the motion postures using a simple voting scheme. We have tested the method on a public benchmark SYSU-3D-HIO and an in-house dataset of human exercises. The results have demonstrated the effectiveness of our method.

**INDEX TERMS** Geometric algebra, motion recognition, support vector machine.

## I. INTRODUCTION

### A. INTRODUCTION

Skeleton-based human motion recognition is one of the hottest research topics in computer vision due to its wide range of applications, such as human-computer interaction [1]–[3], surveillance [4], virtual and augmented reality, motion sensing games [5] and humanoid robot control. To tackle the problem of skeleton-based human motion recognition, one critical step is to extract consistent and discerning features from the human body data to describe human body postures or motions. The consistence means the features used to represent human body postures or motions must adapt to different persons, when they are doing the same or similar motions. On the other hand, to be discerning means the feature has distinct characteristics for different human body postures or motions. With the release of Kinect depth sensor, skeleton data provided by Kinect is used in more and more research and applications. Compared to the 2D image data, the 3D skeleton data contains depth information that can help describe human body postures more effectively. The 3D skeleton data in every frame of human motion contains the 3D coordinates of key body points with respect to the Kinect as the coordinate system origin. There exists several different methods that have been proposed for 3D action representation. The survey of [5] categorized the 3D motion representation methods into three categories:

1) joint-based representations that extract feature representations from the skeletons in order to capture the correlation of the body joints. And these methods can further categorized into spatial descriptors, geometric descriptors and key-pose based descriptors.
2) mined joint based descriptors that is trying to discover which body parts are involved to discriminate among actions.
3) dynamics-based descriptors that treat the skeleton sequence as 3D trajectories and model the dynamics of such time series.

In terms of motion recognition or classification, there are plenty of algorithms that have been proposed for this purposes, for example, traditional machine learning methods such as neural network, SVM (Support Vector Machine) [6],

The associate editor coordinating the review of this manuscript and approving it for publication was Guitao Cao.

Random Forest [7] and deep learning methods like stacked Restricted Boltzmann Machines (RBM) [8] and deep convolutional neural networks (DCNN) [9]–[13]. Although these algorithms have significantly improved the accuracy of human motion recognition, there still exists large room to improve. As pointed out in [5], one of the biggest challenges of using posed-based features for action recognition is that semantically similar motions may not necessarily be numerically similar.

In this paper, we propose a novel human body posture representation based on Geometric Algebra using the most informative body joints angles and joints orientations to describe human body postures, and we treat the postures of a motion independently. Specifically, we select the arms and legs which are the Most Informative Parts (MIP) [6] of human body and construct a 24D feature vector for each of the frames in a motion, namely, 16D for joint angles and 8D for bone orientations. Then for each frame, the extracted 24D feature vector is input to an SVM to predict the motion type. By doing so, for a motion, we obtain a sequence of classification results. To decide the type of the whole motion, we adopt a voting scheme to select the most frequent class from the prediction sequence. Fig.1 illustrates the procedure of our skeleton based human motion recognition. On the other hand, public datasets such as MSRAction3D [14], MSRC-12 [15] and UTKinect-Action [16] are almost skeleton data based on Kinect V1, which contains only the position coordinate. To test the method, in this work, we create dataset using Kinect V2 which contains 12 different exercise motions performed by 10 people, each motion sampled 3 times.

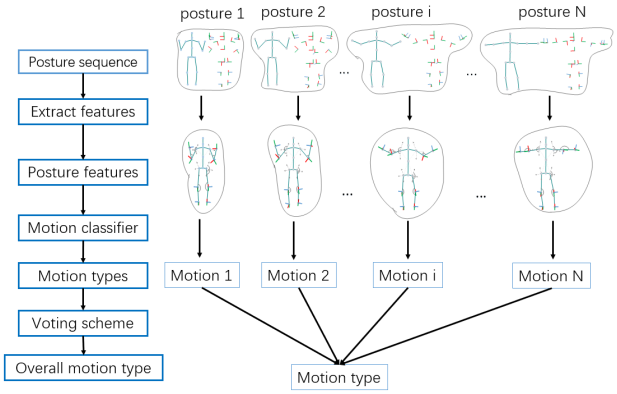


**FIGURE 1.** Method overview. Features are extracted from each posture using the skeleton data provided by Kinect V2. Predicted motion type of the postures are combined by a voting scheme to determine the overall motion type.

In summary, the major contributions of our work are the following three aspects:

1) We propose a novel skeleton-based human body posture representation based on geometric algebra using the most informative body joints angles and joints orientations.
2) We propose a new approach to recognize human motions by predicting the motion type for each posture using SVM, and then select the most frequent motion class as the class for the entire motion by a simple voting strategy.
3) We create a motion dataset using Kinect V2, which contains 12 different human exercise motions by 10 people.

The rest of this paper is organized as follows. Section II presents the related work of invariant features or descriptors of human motions and skeleton-based human body motion recognition. In Section III, we explain how to describe human body postures in GA theoretic framework and the the skeleton-based human motion recognition procedure. Section V presents the experimental results on two datasets. And lastly, Section VI concludes this paper.

## II. RELATED WORK

### A. INVARIANT FEATURES/DESCRIPTORS OF MOTION

The skeleton data provided by Kinect V2 contains 3D coordinate of 25 key joints with respect to the Kinect coordinate system, in which the X axis is pointing right, Y axis is pointing up, and Z axis is pointing towards the tester standing facing the Kinect. Fig.2 shows the 25 key joints of a human body. As the joint position coordinates are absolutely values in the unit of meters, the coordinates will be different when the tester stands at different angle or distance to the Kinect sensor. Therefore, it is necessary to extract invariant features from the raw skeleton data to represent human body posture or motion.

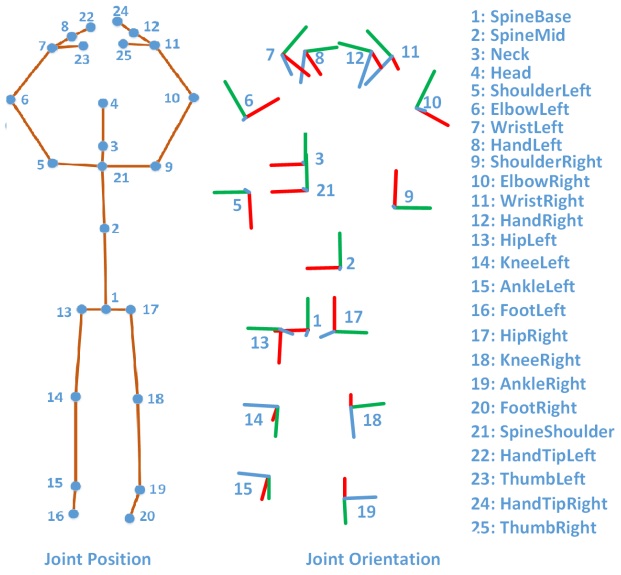


**FIGURE 2.** The visualization of human body joint positions and the corresponding orientation.

#### 1) SPATIAL MOTION INVARIANT DESCRIPTOR

Some researchers simply transform the raw absolute joint position coordinate into the distances between joints. Such kind of representation discards all temporal and orientation information, which may lead to inaccurate descriptions of

the motions. To solve the problem, in [3], Ellis *et al*. represented the body postures using three types of distances: the pairwise distance between all joints; the distance between the joints in the current frame and the ones in the previous frame, the distance between the joints in the current frame and in a neutral or initial posture. Akhter *et al*. also used angle angles to conduct 3D human pose reconstruction from 2D joints locations [17]. and the work of [18] extracted the joint angles relative to the torso. In the work of [19], the authors used the joint angle features as well as the HOG [20] image features for human action recognition. The joint angles and the distance between the joints are intuitively simple and easy to implement.

#### 2) SPATIAL-TEMPORAL INVARIANT DESCRIPTOR

As a motion lasts for a period of time. Spatial motion descriptors alone are unable to describe the motions accurately. Spatial and temporal features are often used together represent human motion in space and time. Shao *et al*. proposed invariant descriptor for multiple motion trajectories based on the kinematic relation among multiple moving parts where the trajectories are defined based on orientation and distance changes [21]. In [6], Guo *et al*. described human motions as the joint trajectories of the Most Informative Parts (MIP) of a human body. Instead of using the entire skeleton data, it decomposed the skeleton into five MIP parts, i.e. left arm, right arm, left leg, right leg, and torso.

Inspired by the concept of MIP, in our work, we extract features from human arms and legs. Motions of different types have different postures. However, we also observe that even the posture frames of the same type of motion are different if the motion is done differently or by different person. Based on this consideration, in our work, we treat each frame in a motion independently.

#### 3) ORIENTATION-BASED INVARIANT DESCRIPTOR

Kinect V2 can capture the 3D coordinates of 25 key joints and the orientation of body bones can be therefore derived. Fig.2 shows the 25 key joint orientations provided by Kinect V2. The joint orientations are invariant to human body size, the relative position and angle to the camera [22]. Therefore, human body posture representation based on joint orientations are more robust and widely used in human motion analysis. In [23], each human joint orientation with respect to the camera was computed and transformed to the joint rotation matrix with respect to the person's torso. In our procedure, the raw joint orientation with respect to the camera is converted into the Euler angles which is easier to understand than the rotation matrix.

### B. SKELETON BASED HUMAN MOTION RECOGNITION

With the descriptor defined, we can extract effective features for human body posture to differentiate different type of motions by machine learning algorithms such as hidden markov model (HMM) and SVM. In [24], Zhang *et al*. proposed to use the Dual Square Root Function (DSRF) descriptors calculated from the raw joint position data. An SVM classifier was trained to recognize the motion type. In [23], the joint orientation features were input to a hierarchical maximum entropy Markov model, which considered a person's activity as composed of a set of sub-activities.

Most of the above classification methods extracted features from the entire motion sequence and the classifier determined the label for the whole motion. Although these methods achieved good results, the input features were complicated as they combined both spatial and temporal information from the full length of the motion. In our proposed method, we extract features from a single posture or a frame of a motion and use the classifier to decide the motion type of the frame. Therefore, a motion will have a sequence of motion labels for each frame. The final motion type is decided by a simple voting scheme which is the most frequent one in the label sequence.

## III. METHOD

### A. GEOMETRIC ALGEBRA: AN OUTLINE

#### 1) THE BASICS OF GEOMETRIC ALGEBRA

Geometric Algebra (GA) was proposed by William K Clifford in 1873, it is also called Clifford Algebra. GA is a powerful mathematic language due to its universality and convenience and has been successfully applied in both theoretical researches such as theoretical physics and practical engineering applications, such as computer vision and the inverse kinematics of robotics. For example, GA theories have been widely used in feature extraction and representation of images [25]–[29]. In [27], GA was used to extract features to register two multimodal medical images.

GA defines a new product called geometric product that unites the Grassmann and Hamilton algebras into one single structure. Let $\mathcal{G}(\mathbb{R}^n)$ represent an n-dimensions GA graded linear space, and $\boldsymbol{u}, \boldsymbol{v} \in \mathcal{G}(\mathbb{R}^n)$ be two vectors. The geometric product is defined as

$$\boldsymbol{u}\boldsymbol{v} = \boldsymbol{u} \cdot \boldsymbol{v} + \boldsymbol{u} \wedge \boldsymbol{v} \tag{1}$$

where $\boldsymbol{u} \cdot \boldsymbol{v}$ is the vector inner product or dot product and $\boldsymbol{u} \wedge \boldsymbol{v}$ is the outer product or wedge product.

The outer product of two linearly independent vectors is a bi-vector, or a 2-blade, which can be regarded as an oriented plane containing $\boldsymbol{u}$ and $\boldsymbol{v}$. The orientation of a bi-vector is clockwise along $\boldsymbol{u}$ and $\boldsymbol{v}$. Outer product can be straightforwardly generalized to higher dimensions. A k-vector or a k-blade is the outer product of k linearly independent vectors, which can be written as

$$A_{\langle k \rangle} = a_1 \wedge a_2 \wedge \cdots \wedge a_k = \Lambda_{i=1}^{k} a_i \tag{2}$$

The grade of a blade is the number of vectors. Therefore, the outer product can be considered as a grade-increasing operation. Conversely, the inner product can be considered as a grade-decreasing operation. Given two blades $A_{\langle k \rangle}, B_{\langle l \rangle} \in \mathcal{G}(\mathbb{R}^n)$, $0 < k \leq l \leq n$, we have the inner product of the two

blades as

$$A_{\langle k\rangle} \cdot B_{\langle l\rangle} = \boldsymbol{a}_1 \cdot \left(\boldsymbol{a}_2 \cdot \left(\ldots \cdot \left(\boldsymbol{a}_k \cdot B_{\langle l\rangle}\right)\right)\right) \tag{3}$$

which is a $(l-k)$-blade.

In N-dimensional GA space $\mathcal{G}(\mathbb{R}^n)$, any element of $\mathcal{G}(\mathbb{R}^n)$ can be represented by a set of orthogonal basis $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n\}$ with the following properties

$$\boldsymbol{e}_1^2 = \boldsymbol{e}_2^2 = \ldots = \boldsymbol{e}_n^2 = -1 \tag{4}$$

$$\boldsymbol{e}_i \cdot \boldsymbol{e}_j = 0, \quad i \neq j \text{ and } i, j \in [1, n] \tag{5}$$

$$\boldsymbol{e}_i\boldsymbol{e}_j = \boldsymbol{e}_i \wedge \boldsymbol{e}_j, \quad i \neq j \text{ and } i, j \in [1, n] \tag{6}$$

$$\boldsymbol{e}_i\boldsymbol{e}_j = -\boldsymbol{e}_i\boldsymbol{e}_i, \quad i \neq j \text{ and } i, j \in [1, n] \tag{7}$$

Let A be a multi-vector $A \in \mathcal{G}(\mathbb{R}^n)$, A can be represented by a linear combination of one scalar, n vectors $\boldsymbol{e}_i$, $n(n-1)/2$ bi-vectors $\boldsymbol{e}_{ij} = \boldsymbol{e}_i\boldsymbol{e}_j$, and higher dimensional vectors until to one n-vector $I = \boldsymbol{e}_1\boldsymbol{e}_2\cdots\boldsymbol{e}_n = \boldsymbol{e}_{12\ldots n}$ as Eq. (8) shows.

$$A = a_0 + \sum_{i=1}^{C_n^1} a_i\boldsymbol{e}_i + \sum_{k=1}^{C_n^2} a_k\boldsymbol{e}_{ij} + \cdots + a_I\boldsymbol{e}_{12\ldots n} \tag{8}$$

where $a_0, a_i, a_k, \cdots, a_I \in \mathbb{R}$, $C_n^i$ is the combination of order n and $I = \boldsymbol{e}_{12\ldots n}$ is the unit pseudo-scalar of $\mathcal{G}(\mathbb{R}^n)$.

#### 2) REFLECTION: GRADE PRESERVING OPERATION

As mentioned earlier, both the inner and outer product change the grade of the subspace of GA. Reflection is a grade preserving operation that does not change the grade of a blade. Let $\boldsymbol{a}, \boldsymbol{n} \in \mathcal{G}(\mathbb{R}^n)$ be two vectors, and $\|\boldsymbol{n}\| = 1$, $\boldsymbol{a} = \boldsymbol{a}^{\parallel} + \boldsymbol{a}^{\perp}$, where $\boldsymbol{a}^{\parallel}$ is parallel to $\boldsymbol{n}$ and $\boldsymbol{a}^{\perp}$ is perpendicular to $\boldsymbol{n}$, then we have

$$\boldsymbol{nan} = \boldsymbol{a}^{\parallel} - \boldsymbol{a}^{\perp} \tag{9}$$

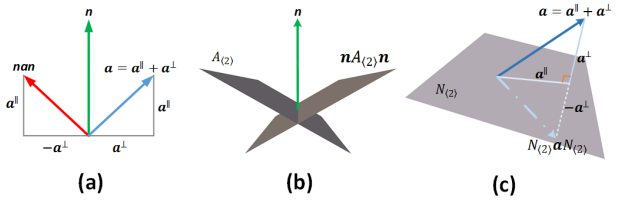


**FIGURE 3.** Illustration of reflection operation in GA. (a) Reflection of vector $a$ on vector $n$. (b) Reflection of bi-vector $A_{\langle 2\rangle}$ on vector $n$. (c) Reflection of vector $a$ on bi-vector $N_{\langle 2\rangle}$.

$\boldsymbol{nan}$ is the reflection of the vector $\boldsymbol{a}$ on the line through the origin with direction $\boldsymbol{n}$. Fig. 3(a) illustrates the reflection of vector $\boldsymbol{a}$ on vector $\boldsymbol{n}$. Reflection operation can be applied to any blade with dimensions greater than 2. For example, a bi-vector $A_{\langle 2\rangle} \in \mathcal{G}(\mathbb{R}^3)$ can be reflected on a normalized vector $\boldsymbol{n} \in \mathbb{R}^3$ via evaluating $\boldsymbol{n}A_{\langle 2\rangle}\boldsymbol{n}$. Refection has the following property

$$\boldsymbol{n}A_{\langle 2\rangle}\boldsymbol{n} = (\boldsymbol{n}\boldsymbol{a}_1\boldsymbol{n}) \wedge (\boldsymbol{n}\boldsymbol{a}_2\boldsymbol{n}) \tag{10}$$

which means the reflection of the outer product of two vectors equals the outer product of vectors after reflection. Fig. 3(b) illustrates the reflection of bi-vector $A_{\langle 2\rangle}$ on vector $\boldsymbol{n}$. And Fig. 3(c) shows the reflection of vector $\boldsymbol{a}$ on bi-vector $N_{\langle 2\rangle}$.

#### 3) ROTATION

In GA, two consecutive reflections on normalized vectors $\boldsymbol{n}$ and $\boldsymbol{m}$ are equivalent to a rotation of $\boldsymbol{a}$ with the angle of $2\theta$, where $\theta = \angle\boldsymbol{mn}$. Conversely, the rotation of vector $\boldsymbol{a}$ in the plane $\boldsymbol{m} \wedge \boldsymbol{n}$ by angle $2\theta$ gives

$$\boldsymbol{b} = \boldsymbol{mnanm} \tag{11}$$

Fig. 4 shows the rotation of vector $\boldsymbol{a}$ by consecutive reflections of $\boldsymbol{a}$ on $\boldsymbol{n}$ and $\boldsymbol{m}$.

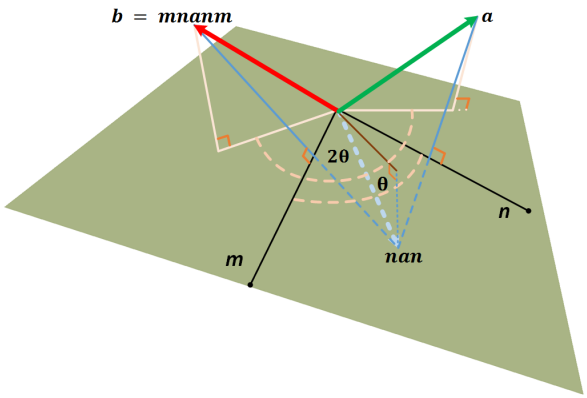


**FIGURE 4.** Rotation of vector $a$ by consecutive reflections of $a$ on $n$ and $m$.

Let $R = \boldsymbol{mn}$, then have

$$\boldsymbol{b} = R\boldsymbol{a}\tilde{R} = e^{-\theta\boldsymbol{mn}}\boldsymbol{a}e^{\theta\boldsymbol{mn}} \tag{12}$$

where R is called rotor and the condition $R\tilde{R} = \tilde{R}R = 1$ holds. In GA, the rotor R can describe a blade rotation with respect to another blade.

With the GA theories introduced above, we can employ GA rotor to represent the rotations of human body bones in 3D GA space. In 3D GA space $\mathcal{G}(\mathbb{R}^3)$, the orthogonal basis is $(\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3)$, and the rotor for vector $v = \sum_{i=1}^{3} \boldsymbol{a}_i\boldsymbol{e}_i \in \mathcal{G}(\mathbb{R}^3)$ can be written as

$$R = w + x(\boldsymbol{e}_2\boldsymbol{e}_3) + y(\boldsymbol{e}_3\boldsymbol{e}_1) + z(\boldsymbol{e}_1\boldsymbol{e}_2) \tag{13}$$

where $w, x, y, z \in \mathbb{R}$. The rotor in the 3D GA space consists of four components $(1, \boldsymbol{e}_2\boldsymbol{e}_3, \boldsymbol{e}_3\boldsymbol{e}_1, \boldsymbol{e}_1\boldsymbol{e}_2)$. So we have the rotated vector $\boldsymbol{v} \mapsto R\boldsymbol{v}\tilde{R} = \boldsymbol{v}'$.

### B. JOINTS ANGLE AND ORIENTATION HUMAN POSTURE DESCRIPTOR BASED GEOMETRIC ALGEBRA

As mentioned earlier, in order to better recognize the human motions, it is critical to ensure the extracted features are consistent and discriminative. We propose to use the angles and orientations of the Most Informative Parts of human body to describe the body posture. Joint orientations and angles are consistent because it is invariant to human position, body size, and relative angle to the camera. On the other hand, joint

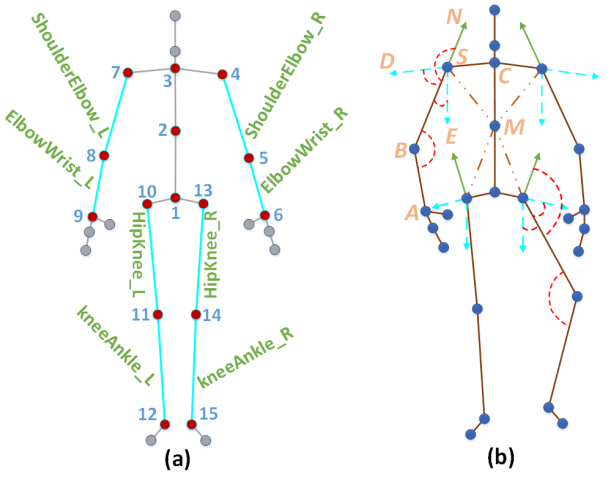


**FIGURE 5.** **(a) The selected body bones and joints. Bones are named by the texts and the joints are denoted by the numbers. (b) Illustration of the joint angles of the skeleton.**

orientations and angles are discriminative as different body postures have different joint orientations and angles. In this paper, we use the angles with respect to torso, and the orientations of human body MIP bones to describe skeleton-based human body postures. And we assume that all human body bones have same original state being straight up in Kinect coordinate system, and all considered bones have the same length namely unit vector $v = (0, 1, 0)$. As for a certain body posture, all considered bones are rotated from the original state. Fig. 5(a) shows the bones and joints in arms and legs considered in this work. The descriptor we use to describe the human body posture is a 24D vector, i.e. 16 dimensions for joint angles and 8 dimensions for bone orientations. Fig. 5(b) illustrates the joint angles we use for the descriptor. For instance, the four joint angles we calculate for left arm are $\theta_{ABS}, \theta_{DSE}, \theta_{BSN}, \theta_{BSE}$. The angle $\theta_{ABS}$ satisfies the following equation

$$\tan\theta_{ABS} = \frac{\text{rej}\left(v^{BA}, v^{BS}\right)}{\text{proj}\left(v^{BA}, v^{BS}\right)} \tag{14}$$

where $\text{proj}\left(v^{BA}, v^{BS}\right)$ is the projection of $v^{BA}$ on $v^{BS}$, and $rej\left(v^{BA}, v^{BS}\right)$ means the rejection of $v^{BA}$ on $v^{BS}$.

$$\text{proj}\left(v^{BA}, v^{BS}\right) = \frac{v^{BA} \cdot v^{BS}}{v^{BS}} = \left(v^{BA} \cdot v^{BS}\right) v^{BS^{-1}} \tag{15}$$

$$\begin{aligned} rej\left(v^{BA}, v^{BS}\right) &= v^{BA} - \text{proj}\left(v^{BA}, v^{BS}\right) \\ &= \left(v^{BA} v^{BS} - v^{BA} \cdot v^{BS}\right) v^{BS^{-1}} \end{aligned} \tag{16}$$

According to Eq. (15) and Eq. (16), we have

$$\begin{aligned} \tan\theta_{ABS} &= \frac{\left(v^{BA} v^{BS} - v^{BA} \cdot v^{BS}\right) v^{BS^{-1}}}{\left(v^{BA} \cdot v^{BS}\right) v^{BS^{-1}}} \\ &= \frac{v^{BA} \cdot v^{BS} + v^{BA} \wedge v^{BS} - v^{BA} \cdot v^{BS}}{v^{BA} \cdot v^{BS}} \\ &= \frac{v^{BA} \wedge v^{BS}}{v^{BA} \cdot v^{BS}} \end{aligned} \tag{17}$$

So the angle $\theta_{ABS}$ can be calculated as:

$$\theta_{ABS} = 180^\circ \times \text{atan}\,2\frac{v^{BA} \wedge v^{BS}}{v^{BA} \cdot v^{BS}}/\pi + 180^\circ \tag{18}$$

The rest of joint angles are calculated similarly. We can see that for each arm and leg we have 4 angles, giving 16 joint angles in total for a specific body posture.

## IV. SKELETON-BASED ENSEMBLE HUMAN MOTION RECOGNITION

In the recognition stage, we extract the features mentioned above from the input joint position and orientation data captured by Kinect V2. Thus we have 25 key joint positions $J_{p_i} = (x_i, y_i, z_i)$, $i \in [1, 25]$ and 25 key orientations $J_{o_i} = (w_i, x_i, y_i, z_i)$, $i \in [1, 25]$ as shown in Fig. 2. So, the feature data for body posture of a motion is

$$J_{pos} = \left[J_{p_{x,y,z}}(n)^T, J_{o_{w,x,y,z}}(n)^T\right]^T, \quad n \in [1, 25] \tag{19}$$

With these features, we train a SVM to learn the motion type of each body posture in a motion, which contains a series of postures or frames. And we treat these postures independently. So for a motion, we will have a sequence of predictions from the classifier. To decide the motion type of the entire motion, we adopt a simple voting scheme to choose the most frequent class label. The procedure is summarized in **Algorithm 1** listed below.

## V. EXPERIMENT

To demonstrate the effectiveness of our method, we test the method on SYSU-3D-HIO [30] which is a public skeleton-based dataset and in-house dataset, called SZU-3D-SOEARD.

### A. SZU 3D SKELETON AND ORIENTATION EXERCISE ACTION RECOGNITION DATASET

SZU-3D-SOEARD is collected to study the action recognition of human exercises based on 3D skeleton and orientation data. We utilize Kinect V2 to capture 12 different exercise motions of 10 people. The frame rate is set to 30 fps. Each exercise motion is tried and captured 3 times. So, in total, there are $12 \times 3 \times 10 = 360$ motion samples in the databases. The 12 different motions are: squat, jumping jack, arm turning, walk in-place, arm pendulum, TW stretching, right leg side lift, left leg side lift, bilateral leg press, shoulder turning, shoulder right angle movement, and wave right hand.

To carry out the experiments, we split the dataset into two parts, one for training and the other one for testing. Specifically, data from person 1, 3, 5, 7 and 9 are used for training, and person 2, 4, 6, 8 and 10 for testing. In this work, we train the SVM classifier using a linear kernel with tolerance for stopping criterion being 0.001. The result of our method on this dataset is shown in Fig. 6. In Fig. 6(a), the confuse matrix shows the average motion recognition accuracy of body postures. The overall prediction accuracy is 77.2%. We can see that for some motions such as squat, arm pendulum, bilateral leg press and shoulder turning, the recognition

**Algorithm 1** Human Motion Recognition Method

input body motion sequence: (N is the number of body posture in motion sequence)

$$\boldsymbol{M} = \left[\boldsymbol{J}_{pos}(\boldsymbol{t})\right]^{\boldsymbol{T}}, \boldsymbol{t} \in [1, \boldsymbol{N}], \boldsymbol{M} \in \mathbb{R}^{(3+4)*25*\boldsymbol{N}}$$

fetch joint angles and joint Euler angles :

foreach $\boldsymbol{J}_{\text{pos}}$ in $\boldsymbol{M}$

for the $\boldsymbol{J}_o$ in $\boldsymbol{J}_{\text{pos}}$, convert it into Euler angle:

$\boldsymbol{J}_o \in \mathbb{R}^{4*25} \rightarrow \boldsymbol{J}_{e_{x,y,z}} \in \mathbb{R}^{3*25}$

for the $\boldsymbol{J}_p$ in $\boldsymbol{J}_{\text{pos}}$, calculate body joint angle:

$\boldsymbol{J}_p \in \mathbb{R}^{3*25} \rightarrow \boldsymbol{J}_\alpha \in \mathbb{R}^{16}$

the body posture descriptor of $\boldsymbol{J}_{\text{pos}}$ :

$\boldsymbol{J}_f = \left[\boldsymbol{J}_{e_y}^T, \boldsymbol{J}_a^T\right]^T, \boldsymbol{J}_f \in \mathbb{R}^{(8+16)}$

the body posture descriptor of $\boldsymbol{M}$ :

$\mathbf{F} = \left[J_f(\boldsymbol{t})\right]^T, \boldsymbol{t} \in [\mathbf{1}, \boldsymbol{N}], \boldsymbol{F} \in \mathbb{R}^{(8+16)*N}$

foreach $\boldsymbol{J}_f$ in $\boldsymbol{F}$ :

feed $\boldsymbol{J}_f$ into trained SVM and get the reference motion:( $\boldsymbol{C}$ is the number of motion type )

$\boldsymbol{J}_f \rightarrow SVM \rightarrow$reference motion type $\widehat{\boldsymbol{M}} \in [1, \boldsymbol{C}]$

the corresponding results sequence :

$\boldsymbol{R} = [\widehat{\boldsymbol{M}}(\boldsymbol{t})]^T, \boldsymbol{t} \in [\mathbf{1}, \boldsymbol{N}], \boldsymbol{R} \in \mathbb{R}^N$

assemble the results sequence by vote : ( $t_i$ is the number of the motion type i )

$\boldsymbol{R} \xrightarrow{vote} \boldsymbol{R}_v = [\boldsymbol{\varepsilon_1}, \boldsymbol{\varepsilon_2}, \cdots, \boldsymbol{\varepsilon_C}]$

output the motion recognition result:

$\widetilde{M} = \arg\max_i \boldsymbol{\varepsilon_i}$

accuracy is over 93%. However, for motions of jumping jack, walk in-place, TW stretching and left leg side, the recognition accuracy is only about 60%. One of the reasons for this is these motions overlap each other to some extent in terms of body postures. The second reason is the size of the training dataset. We believe that the performance can be significantly improved by adding more training data.

In Fig. 6(b), the confuse matrix shows the motion recognition accuracy for motions. Although average recognition accuracy at posture level is not high (77.2%), the motion average recognition accuracy achieves 94.4%, which is the mean of Fig. 6(b). All motions are predicted with average accuracy over 80%. Except for arm turning and left leg side, all motions are correctly predicted with accuracy are over 93%. Especially, motions such as squat, arm pendulum, right leg side lift, bilateral leg press and shoulder turning are correctly predicted with 100% accuracy.

### B. SYSU 3D HUMAN-OBJECT INTERACTION DATASET

In order to further evaluate the effectiveness of our method, we use the public dataset of SYSU 3D Human-Object Interaction Dataset (SYSU-3D-HOI) to test and compare to other existing methods. SYSU-3D-HOI is a collection of 3D Human Object Interaction (HOI) data collected by Hu Jianfang from Sun Yat-sen University (SYSU) iSEE Intelligent Science and Systems Laboratory. In the dataset, there are 12 actions performed by 40 different people interacting with the mobile phone, chair, backpack, wallet, broom, and mop, 480 action sequences in total. The 12 actions are drinking, pouring, calling phone, playing phone, wearing backpacks, packing backpacks, sitting chair, moving chair, taking out wallet, taking from wallet, mopping and sweeping.

In this experiment, we just extract the joint angle feature as the data does not contain joint orientation information. To carry out the experiments, we split the dataset into two parts, 2/3 used for training, and the rest 1/3 for testing. The SVM classifier is trained with a polynomial kernel. The parameter gamma is set to 0.00001 and tolerance for stopping criterion is set to 0.0001. The corresponding result is shown in Fig. 7. In Fig. 7(a), the average recognition accuracy of body postures is 55.5%. For the three actions of pouring, packing and sweeping, the posture classification accuracy is less than 37%, which is relatively low. But for the rest of the actions, single posture classification accuracy is above 53%, even above 60%. Fig. 7 (b) shows the average motion recognition of entire action samples in SYSU-3D-HIO. The average recognition accuracy is 84.62%. As we can see in Table 1, the performance is comparable to the best result of [30].

In this test, we only use the skeleton data, but the other methods utilized the RGB data and/or depth data. However, our method still achieves almost the best performance. In addition, our proposed motion classification method can be used to classify human motions online in realtime manner. Our method predicts the motion type of each body posture during

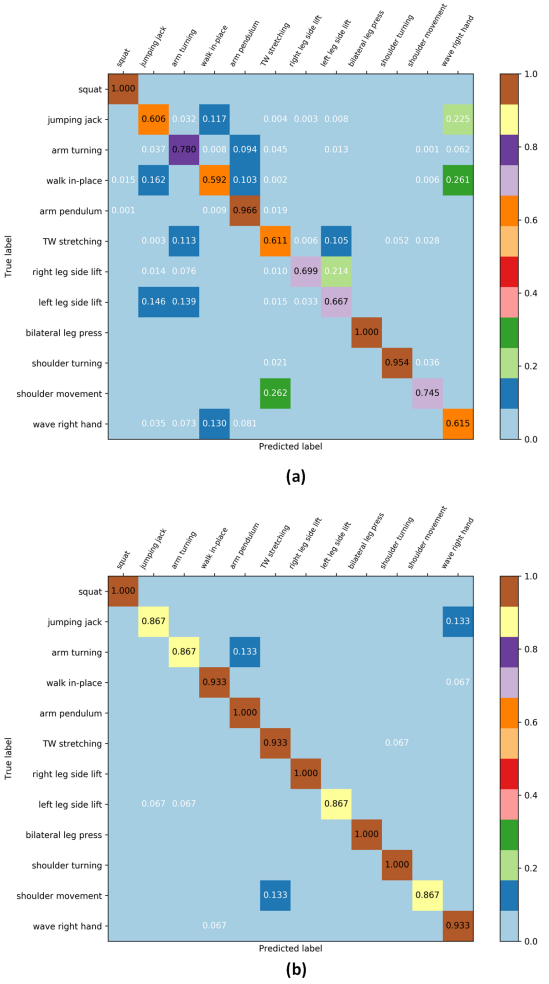


**FIGURE 6.** The recognition performance on dataset SZU-3D-SOEARD. (a) Average recognition accuracy of body postures. (b) Average recognition accuracy of motions.

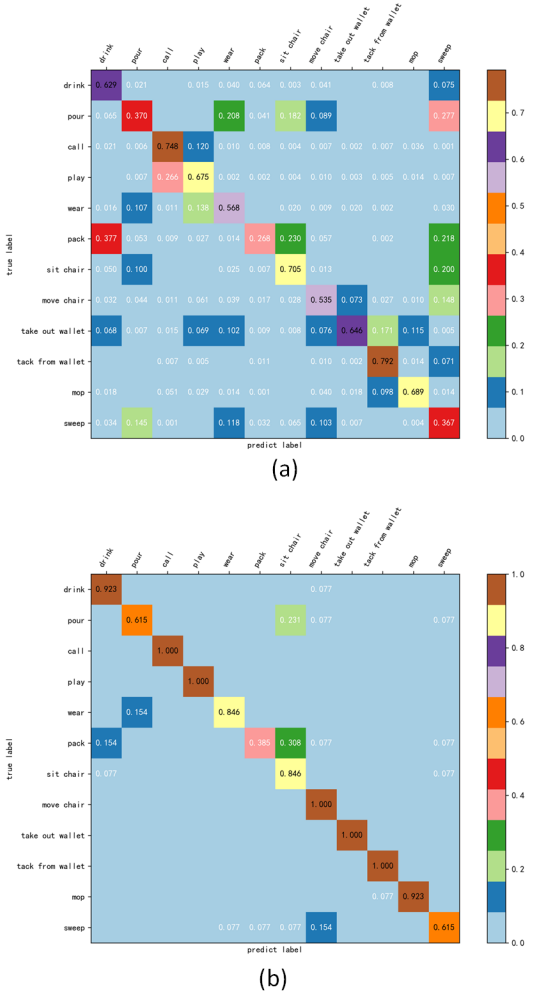


**FIGURE 7.** The recognition performance on dataset SYSU-3D-HIO. (a) Average recognition accuracy of body postures. (b) Average recognition accuracy of motions.

**TABLE 1.** Comparison results with the existing skeleton based methods on SYSU-3D-HIO.

| method | accuracy(%) |
|---|---|
| HON4D [32] | 79.22 |
| DS+SVM [31] | 75.53 |
| DCP+SVM [31] | 77.32 |
| DDP+SVM [31] | 78.29 |
| DS+DCP+DDP+SVM [31] | 82.78 |
| DS+DCP+DDP+MTDA [24] | 84.21 |
| DS+DCP+DDP+JOULE-Score [31] | 84.24 |
| DS+DCP+DDP+JOULE-SVM [31] | 84.89 |
| GA + SVM(Propose) | 84.62 |

the motion, and decide the overall motion type at the end of the motion. The recognition delay is thus greatly reduced. To further improve the recognition accuracy, besides the predictions of each body postures, we can utilize the dependency between adjacent postures during a motion to improve the overall motion recognition performance, using methods like Hidden Markov Model or even more advanced machine learning algorithms.

## VI. CONCLUSION

We have proposed a novel method for human motion recognition based 3D body skeleton data. In the method, we utilize Clifford Algebra to represent the joint angles and orientations of the most informative body parts. With the proposed skeleton posture descriptors, each posture is represented by a feature vector calculated from 3D skeleton data provided by Kinect V2. As a motion consists of a sequence of different body postures, we use a trained SVM classifier to decide the motion type of each posture and then decide the motion type for the whole motion by a simple voting scheme which selects the most frequent class label. The experiments on two datasets have demonstrated the effectiveness of the method.

## REFERENCES

[1] A. Dix, "Human–computer interaction, foundations and new paradigms," *J. Vis. Lang. Comput.*, vol. 42, pp. 122–134, Oct. 2017.

[2] A. Haria, A. Subramanian, N. Asokkumar, S. Poddar, and J. S. Nayak, "Hand gesture recognition for human computer interaction," *Procedia Comput. Sci.*, vol. 115, pp. 367–374, Jan. 2017.
[3] C. Ellis, S. Z. Masood, M. F. Tappen, J. J. LaViola, Jr., and R. Sukthankar, "Exploring the trade-off between accuracy and observational latency in action recognition," *Int. J. Comput. Vis.*, vol. 101, no. 3, pp. 420–436, 2013.
[4] W. Lao, J. Han, and P. H. N. De With, "Automatic video-based human motion analyzer for consumer surveillance system," *IEEE Trans. Consum. Electron.*, vol. 55, no. 2, pp. 591–598, May 2009.
[5] L. L. Presti and M. La Cascia, "3D skeleton-based human action classification: A survey," *Pattern Recognit.*, vol. 53, pp. 130–147, May 2016.
[6] Y. Guo, Y. Li, and Z. Shao, "DSRF: A flexible trajectory descriptor for articulated human action recognition," *Pattern Recognit.*, vol. 76, pp. 137–148, Apr. 2018.
[7] M. Huang, G.-R. Cai, H.-B. Zhang, S. Yu, D.-Y. Gong, D.-L. Cao, S. Li, and S.-Z. Su, "Discriminative parts learning for 3D human action recognition," *Neurocomputing*, vol. 291, pp. 84–96, May 2018.
[8] C. Zhang, N. Ji, and G. Wang, "Restricted Boltzmann machines," *Chin. J. Eng. Math.*, vol. 32, no. 2, pp. 159–173, 2015.
[9] D. Mehta, S. Sridhar, O. Sotnychenko, H. Rhodin, M. Shafiei, H.-P. Seidel, W. Xu, D. Casas, and C. Theobalt, "VNect: Real-time 3D human pose estimation with a single RGB camera," *ACM Trans. Graph.*, vol. 36, no. 4, p. 44, 2017.
[10] W. Cao, Q. Lin, Z. He, and Z. He, "Hybrid representation learning for cross-modal retrieval," *Neurocomputing*, vol. 345, pp. 45–57, Jun. 2019.
[11] W. Cao, J. Yuan, Z. He, Z. Zhang, and Z. He, "Fast deep neural networks with knowledge guided training and predicted regions of interests for real-time video object detection," *IEEE Access*, vol. 6, pp. 8990–8999, 2018.
[12] M. Dan, L. Zhang, G. Cao, W. Cao, G. Zhang, and H. Bing, "Liver fibrosis classification based on transfer learning and FCNet for ultrasound images," *IEEE Access*, vol. 5, pp. 5804–5810, 2017.
[13] D. Meng, G. Cao, Y. Duan, M. Zhu, L. Tu, D. Xu, and J. Xu, "Tongue images classification based on constrained high dispersal network," *Evidence-Based Complementary Alternative Med.*, vol. 2017, no. 4, Mar. 2017, Art. no. 7452427.
[14] W. Li, Z. Zhang, and Z. Liu, "Action recognition based on a bag of 3D points," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.-Workshops*, Jun. 2010, pp. 9–14.
[15] S. Fothergill, H. Mentis, P. Kohli, and S. Nowozin, "Instructing people for training gestural interactive systems," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, May 2012, pp. 1737–1746.
[16] L. Xia, C.-C. Chen, and J. K. Aggarwal, "View invariant human action recognition using histograms of 3D joints," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2012, pp. 20–27.
[17] I. Akhter and M. J. Black, "Pose-conditioned joint angle limits for 3D human pose reconstruction," in *Proc. CVPR*, Jun. 2015, pp. 1446–1455.
[18] Y. Gu, H. Do, Y. Ou, and W. Sheng, "Human gesture recognition through a Kinect sensor," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2012, pp. 1379–1384.
[19] E. Ohn-Bar and M. M. Trivedi, "Joint angles similarities and HOG2 for action recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2013, pp. 465–470.
[20] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, Jun. 2005, pp. 886–893.
[21] Z. Shao and Y. F. Li, "A new descriptor for multiple 3D motion trajectories recognition," in *Proc. IEEE Int. Conf. Robot. Automat.*, May 2013, pp. 4749–4754.
[22] F. Han, B. Reily, W. Hoff, and H. Zhang, "Space-time representation of people based on 3D skeletal data: A review," *Comput. Vis. Image Understand.*, vol. 158, pp. 85–105, May 2017.
[23] J. Sung, C. Ponce, B. Selman, and A. Saxena, "Unstructured human activity detection from RGBD images," in *Proc. IEEE Int. Conf. Robot. Automat.*, May 2012, pp. 842–849.
[24] Y. Zhang and D.-Y. Yeung, "Multi-task learning in heterogeneous feature spaces," in *Proc. 25th AAAI Conf. Artif. Intell.*, Aug. 2011, pp. 574–579.
[25] R. Wang, W. Zhang, Y. Shi, X. Wang, and W. Cao, "GA-ORB: A new efficient feature extraction algorithm for multispectral images based on geometric algebra," *IEEE Access*, vol. 7, pp. 71235–71244, 2019.
[26] R. Wang, M. Shen, and W. Cao, "Multivector sparse representation for multispectral images using geometric algebra," *IEEE Access*, vol. 7, pp. 12755–12767, 2019.
[27] W. Cao, F. Lyu, Z. He, G. Cao, and Z. He, "Multimodal medical image registration based on feature spheres in geometric algebra," *IEEE Access*, vol. 6, pp. 21164–21172, 2018.
[28] M. Shen, R. Wang, and W. Cao, "Joint sparse representation model for multi-channel image based on reduced geometric algebra," *IEEE Access*, vol. 6, pp. 24213–24223, 2018.
[29] R. Wang, Y. He, C. Huang, X. Wang, and W. Cao, "A novel least-mean kurtosis adaptive filtering algorithm based on geometric algebra," *IEEE Access*, vol. 7, pp. 78298–78310, 2019.
[30] J.-F. Hu, W.-S. Zheng, J. Lai, and J. Zhang, "Jointly learning heterogeneous features for RGB-D activity recognition," in *Proc. CVPR*, Jun. 2015, pp. 5344–5352.
[31] O. Oreifej and Z. Liu, "HON4D: Histogram of oriented 4D normals for activity recognition from depth sequences," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 716–723.